Journals

The Institution of
Engineering and Technology

Partially coupled gradient estimation algorithm for multivariable equation-error autoregressive moving average systems using the data filtering technique

Qinyao Liu[1], Feng Ding[1,2,3,], Ling Xu[1], Erfu Yang[4]*

[1] *Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, People's Republic of China*
[2] *College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, People's Republic of China*
[3] *Department of Mathematics, King Abdulaziz University, Jeddah 21589, Saudi Arabia*
[4] *Department of Design, Manufacture and Engineering Management, Space Mechatronic Systems Technology Laboratory, Strathclyde Space Institute, University of Strathclyde, Glasgow G1 1XJ, Scotland, United Kingdom*
* E-mail: fding@jiangnan.edu.cn

Abstract: System identification provides many convenient and useful methods for engineering modeling. This paper targets the parameter identification problems for multivariable equation-error autoregressive moving average systems. To reduce the influence of the colored noises on the parameter estimation, the data filtering technique is adopted to filter the input and output data, and to transform the original system into a filtered system with white noises. Then we decompose the filtered system into several subsystems and develop a filtering based partially-coupled generalized extended stochastic gradient algorithm via the coupling concept. In contrast to the multivariable generalized extended stochastic gradient algorithm, the proposed algorithm can give more accurate parameter estimates. Finally, the effectiveness of the proposed algorithm is well demonstrated by simulation examples.

1 Introduction

Mathematical models can describe the dynamic behavior of a system as a function of time and exist in various fields, such as fault diagnosis [1, 2], telecommunication transmission [3–5], control system [6, 7] and signal processing [8, 9]. In most cases, it is not easy to analyse the whole system and to construct the mathematical model based on the analytic approach. Therefore, system identification becomes the first choice when people model a system [10–12], and has a wide application in both linear [13] and nonlinear systems [14–17]. For the multiple-input and multiple-output Box-Jenkins model with disturbances, a two-stage identification method was developed by using the residual model of Kalman filter [18]. Separating the parameters to be estimated into a linear part and a nonlinear part is a common idea to solve the nonlinear least squares problems. Based on this method, Gan et al. eliminated the linear parameters through the orthogonal projection and presented a variable projection algorithm for the radial basis function network-based autoregressive with exogenous inputs model [19].

During the past decade, a great deal of attention has been given to multivariable system identification [20, 21] for the reason that many modern industrial processes are multivariable systems [22]. Applying the scalar system identification methods to multivariable systems may give poor performances, because multivariable systems have high-dimensional variables, complicated structures and many uncertain disturbances [23]. It attracts an increasing interest for researchers to explore more valid methods for multivariable systems [24, 25]. When dealing with the multivariable systems with colored noises, the data filtering technique can be applied to reduce the influence of the noise and improve the estimation accuracy [26–28]. In [29], a filtering based multi-innovation extended stochastic gradient algorithm was developed for improving the parameter estimation accuracy of the multivariable system with the moving average noise.

To simplify the identification model for multivariable systems, the hierarchical identification principle provides a new idea to this problem [30, 31]. The hierarchical identification, which is also known as

the decomposition identification, is to decompose a large-scale system into several small size subsystems with fewer variables. However, after the decomposition of some multivariable systems, there are some parameter coupled relations between the subsystems. In this case, we link the parameter estimates inside the subsystems and identify each subsystem based on the coupling identification concept [32]. The coupling identification concept can avoid the redundant estimates and improve the computational efficiency for multivariable systems. In this literature, Ding et al. combined the coupling concept and the gradient search to estimate the non-uniformly sampled systems and obtained a highly computational efficiency algorithm [33]. For nonlinear multivariable output-error moving average systems, a partially coupled extended stochastic gradient algorithm was presented by using the decomposition technique and the coupling concept [34]. In the previous work [35], we applied the hierarchical identification principle to multivariable system and used the coupling concept to solve the redundant computation.

This work studies the recursive identification methods of multivariable equation-error autoregressive moving average systems by using the data filtering technique and the coupling identification concept. In view of the colored noises, we first design a filter according to the model structure. By filtering the input-output data, the multivariable system is divided into a noise model and a filtered system model. Next, the filtered system model is decomposed into several sub-models on the basis of the number of outputs. However, a major challenge encountered in estimating the parameters is that there are some unmeasurable variables inside the noise system and the sub-models. To obtain the feasible algorithm, we establish some auxiliary models and replace the unknown variables with the outputs of these models. Then, a filtering based partially-coupled generalized extended stochastic gradient (F-PC-GESG) algorithm is derived by using the coupling concept and the auxiliary model. Compared with the multivariable generalized extended stochastic gradient (M-GESG) algorithm, the F-PC-GESG algorithm gives more accurate parameter estimates under the same noise level. In addition, we

introduce a forgetting factor to improve the performance of the F-PC-GESG algorithm.

The outline of this paper is as follows. The symbols and the identification model for the multivariable equation-error autoregressive moving average system are given in Section 2. Section 3 derives the filtered identification models and presents the F-PC-GESG algorithm based on the data filtering technique and the coupling concept. Section 4 proposes the M-GESG algorithm and gives some comparisons with the F-PC-GESG algorithm. The numerical simulations are shown to verify the effectiveness of the proposed algorithms in Section 5. Section 6 offers some concluding remarks to end the paper.

2 Problem formulation

In this section, we state the identification problems and derive the hierarchical identification model for multivariable equation-error autoregressive moving average systems. Let us start by introducing some symbols used in this paper. "$A =: X$" or "$X := A$" stands for "A is defined as X"; the superscript T stands for the vector/matrix transpose; the symbol $\boldsymbol{I}_m$ denotes an identity matrix of appropriate size ($m \times m$); $\boldsymbol{1}_m$ stands for an m-dimensional column vector whose elements are 1; $\boldsymbol{1}_{m \times n}$ represents a matrix of size $m \times n$ whose elements are 1; the symbol $\otimes$ represents the 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, $\boldsymbol{A} \otimes \boldsymbol{B} \neq \boldsymbol{B} \otimes \boldsymbol{A}$; $\mathrm{col}[\boldsymbol{X}]$ is defined as a vector formed by all columns of matrix $\boldsymbol{X}$ and arranged in order, for example, $\boldsymbol{X} := [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n] \in \mathbb{R}^{m \times n}$, $\boldsymbol{x}_i \in \mathbb{R}^m (i = 1, 2, \cdots, n)$, $\mathrm{col}[\boldsymbol{X}] := [\boldsymbol{x}_1^{\mathrm{T}}, \boldsymbol{x}_2^{\mathrm{T}}, \cdots, \boldsymbol{x}_n^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{mn}$; $\hat{\boldsymbol{\theta}}(t)$ denotes the estimate of $\boldsymbol{\theta}$ at time t; the norm of a matrix (or a column vector) $\boldsymbol{X}$ is defined by $\|\boldsymbol{X}\|^2 := \mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$.

Consider the following multivariable equation-error autoregressive moving average system which is also called the multivariable controlled autoregressive autoregressive moving average (CARARMA) system:

$$\boldsymbol{A}(z)\boldsymbol{y}(t) = \boldsymbol{B}(z)\boldsymbol{u}(t) + \frac{D(z)}{C(z)}\boldsymbol{v}(t), \qquad (1)$$

where $\boldsymbol{y}(t) := [y_1(t), y_2(t), \cdots, y_m(t)]^{\mathrm{T}} \in \mathbb{R}^m$ refers to the m-dimensional output vector, $\boldsymbol{u}(t) := [u_1(t), u_2(t), \cdots, u_r(t)]^{\mathrm{T}} \in \mathbb{R}^r$ denotes the r-dimensional input vector, $\boldsymbol{v}(t) := [v_1(t), v_2(t), \cdots, v_m(t)]^{\mathrm{T}} \in \mathbb{R}^m$ is a white noise vector, $\boldsymbol{A}(z)$ and $\boldsymbol{B}(z)$ are matrix polynomials in the unit backward shift operator z^{-1} $[z^{-1}u(t) = u(t-1)]$, $C(z)$ and $D(z)$ are scalar polynomials in z^{-1}, and they are defined as

$$\boldsymbol{A}(z) := \boldsymbol{I}_m + \boldsymbol{A}_1 z^{-1} + \boldsymbol{A}_2 z^{-2} + \cdots + \boldsymbol{A}_{n_a} z^{-n_a},$$
$$\boldsymbol{B}(z) := \boldsymbol{B}_1 z^{-1} + \boldsymbol{B}_2 z^{-2} + \cdots + \boldsymbol{B}_{n_b} z^{-n_b},$$
$$C(z) := 1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c},$$
$$D(z) := 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots + d_{n_d} z^{-n_d}.$$

Define an intermediate variable

$$\boldsymbol{w}(t) := \frac{D(z)}{C(z)}\boldsymbol{v}(t). \qquad (2)$$

Let $n := mn_a + rn_b$, define the parameter matrix $\boldsymbol{\theta}$, the parameter vector $\boldsymbol{\rho}$, the information vector $\boldsymbol{\varphi}(t)$ and the information matrix $\boldsymbol{\psi}(t)$ as

$$\boldsymbol{\theta}^{\mathrm{T}} := [\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_{n_a}, \boldsymbol{B}_1, \boldsymbol{B}_2, \cdots, \boldsymbol{B}_{n_b}] \in \mathbb{R}^{m \times n},$$
$$\boldsymbol{\rho} := [c_1, c_2, \cdots, c_{n_c}, d_1, d_2, \cdots, d_{n_d}]^{\mathrm{T}} \in \mathbb{R}^{n_c + n_d},$$
$$\boldsymbol{\varphi}(t) := [-\boldsymbol{y}^{\mathrm{T}}(t-1), -\boldsymbol{y}^{\mathrm{T}}(t-2), \cdots, -\boldsymbol{y}^{\mathrm{T}}(t-n_a),$$
$$\qquad \boldsymbol{u}^{\mathrm{T}}(t-1), \boldsymbol{u}^{\mathrm{T}}(t-2), \cdots, \boldsymbol{u}^{\mathrm{T}}(t-n_b)]^{\mathrm{T}} \in \mathbb{R}^n,$$
$$\boldsymbol{\psi}(t) := [-\boldsymbol{w}(t-1), -\boldsymbol{w}(t-2), \cdots, -\boldsymbol{w}(t-n_c),$$
$$\qquad \boldsymbol{v}(t-1), \boldsymbol{v}(t-2), \cdots, \boldsymbol{v}(t-n_d)] \in \mathbb{R}^{m \times (n_c + n_d)}.$$

Through the above definitions, $\boldsymbol{w}(t)$ can be expressed as different forms,

$$\boldsymbol{w}(t) = [1 - C(z)]\boldsymbol{w}(t) + D(z)\boldsymbol{v}(t)$$
$$= -\sum_{i=1}^{n_c} c_i \boldsymbol{w}(t-i) + \sum_{i=1}^{n_d} d_i \boldsymbol{v}(t-i) + \boldsymbol{v}(t)$$
$$= \boldsymbol{\psi}(t)\boldsymbol{\rho} + \boldsymbol{v}(t). \qquad (3)$$

According to (1) and (2), $\boldsymbol{w}(t)$ can also be described as

$$\boldsymbol{w}(t) = \boldsymbol{A}(z)\boldsymbol{y}(t) - \boldsymbol{B}(z)\boldsymbol{u}(t)$$
$$= \boldsymbol{y}(t) - \boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(t). \qquad (4)$$

Substituting (2)–(4) into (1), we can obtain the following hierarchical identification model,

$$\boldsymbol{y}(t) = [\boldsymbol{I}_m - \boldsymbol{A}(z)]\boldsymbol{y}(t) + \boldsymbol{B}(z)\boldsymbol{u}(t) + \boldsymbol{w}(t)$$
$$= \boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(t) + \boldsymbol{w}(t)$$
$$= \boldsymbol{\psi}(t)\boldsymbol{\rho} + \boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(t) + \boldsymbol{v}(t). \qquad (5)$$

Equation (5) has a parameter vector $\boldsymbol{\rho}$ and a parameter matrix $\boldsymbol{\theta}$. In order to simplify the model in (5), combine the information vector $\boldsymbol{\varphi}(t)$ with the information matrix $\boldsymbol{\psi}(t)$ to construct an information matrix $\boldsymbol{\Phi}(t)$ by means of the Kronecker product:

$$\boldsymbol{\Phi}(t) := [\boldsymbol{\psi}(t), \boldsymbol{\varphi}^{\mathrm{T}}(t) \otimes \boldsymbol{I}_m] \in \mathbb{R}^{m \times n_0}, \quad n_0 := n_c + n_d + mn.$$

Define a new parameter vector:

$$\boldsymbol{\vartheta} := \begin{bmatrix} \boldsymbol{\rho} \\ \mathrm{col}[\boldsymbol{\theta}^{\mathrm{T}}] \end{bmatrix} \in \mathbb{R}^{n_0}.$$

Then Equation (5) can be rewritten as a pseudo-linear regressive model:

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}(t)\boldsymbol{\vartheta} + \boldsymbol{v}(t). \qquad (6)$$

Equation (6) is the identification model for the multivariable CARARMA system in (1).

Remark 1: For the identification model in (6), the parameter vector $\boldsymbol{\vartheta}$ contains all the parameters to be estimated. Although some algorithms can estimate the parameter vector $\boldsymbol{\vartheta}$, the high-dimensional parameter vector and information matrix result in a heavy computational burden and poor performance of these algorithms. On the other hand, the information matrix $\boldsymbol{\Phi}(t)$ consists of the unknown intermediate variables $\boldsymbol{w}(t-j)$ and the noise terms $\boldsymbol{v}(t-j)$. The objective of this paper is to present a new effective algorithm using the observation data $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$. By employing the coupling identification concept and the data filtering technique, the new algorithm can deal with the unknown variables and give more accurate parameter estimates for the multivariable CARARMA system in (1). Since this paper focuses on the parameter estimation, we assume that the orders and the initial values are known, that is to say, the orders m, r, n_a, n_b, n_c and n_d are known and $\boldsymbol{y}(t) = \boldsymbol{0}$, $\boldsymbol{u}(t) = \boldsymbol{0}$ and $\boldsymbol{v}(t) = \boldsymbol{0}$ for $t \leqslant 0$.

3 The filtering based partially-coupled generalized extended stochastic gradient algorithm

It is worth noting that the system considered in this paper is disturbed by the autoregressive moving average noise (i.e., colored noise). In order to reduce the impact of the noise, we introduce the data filtering technique here to obtain more accurate estimates. The basic idea is to use a filter to filter the input-output data and to decompose the original identification model into two models which contain a filtered system model and a noise model. Then we further divide the system model into a series of subsystems according to the number of

outputs and identify each subsystem based on the coupled relations in part of the parameters between subsystems. From the above-mentioned idea, we deduce the filtering based partially-coupled generalized extended stochastic gradient (F-PC-GESG) algorithm in this section.

Define the filtered input vector $\boldsymbol{u}_{\mathrm{f}}(t)$ and the filtered output vector $\boldsymbol{y}_{\mathrm{f}}(t)$ as

$$\boldsymbol{u}_{\mathrm{f}}(t) := C(z)\boldsymbol{u}(t) \in \mathbb{R}^r,$$
$$\boldsymbol{y}_{\mathrm{f}}(t) := C(z)\boldsymbol{y}(t) \in \mathbb{R}^m.$$

Multiplying both sides of (1) by $C(z)$ gives

$$\boldsymbol{A}(z)C(z)\boldsymbol{y}(t) = \boldsymbol{B}(z)C(z)\boldsymbol{u}(t) + D(z)\boldsymbol{v}(t),$$

or

$$\boldsymbol{A}(z)\boldsymbol{y}_{\mathrm{f}}(t) = \boldsymbol{B}(z)\boldsymbol{u}_{\mathrm{f}}(t) + D(z)\boldsymbol{v}(t). \tag{7}$$

Define the noise parameter vector $\boldsymbol{d}$, the information matrix $\boldsymbol{\phi}(t)$ and the filtered information vector $\boldsymbol{\varphi}_{\mathrm{f}}(t)$:

$$\boldsymbol{d} := [d_1, d_2, \cdots, d_{n_d}]^{\mathrm{T}} \in \mathbb{R}^{n_d},$$
$$\boldsymbol{\phi}(t) := [\boldsymbol{v}(t-1), \boldsymbol{v}(t-2), \cdots, \boldsymbol{v}(t-n_d)] \in \mathbb{R}^{m \times n_d},$$
$$\boldsymbol{\varphi}_{\mathrm{f}}(t) := [-\boldsymbol{y}_{\mathrm{f}}^{\mathrm{T}}(t-1), -\boldsymbol{y}_{\mathrm{f}}^{\mathrm{T}}(t-2), \cdots, -\boldsymbol{y}_{\mathrm{f}}^{\mathrm{T}}(t-n_a),$$
$$\boldsymbol{u}_{\mathrm{f}}^{\mathrm{T}}(t-1), \boldsymbol{u}_{\mathrm{f}}^{\mathrm{T}}(t-2), \cdots, \boldsymbol{u}_{\mathrm{f}}^{\mathrm{T}}(t-n_b)]^{\mathrm{T}} \in \mathbb{R}^n.$$

Then Equation (7) can be modified as

$$\boldsymbol{y}_{\mathrm{f}}(t) = [\boldsymbol{I}_m - \boldsymbol{A}(z)]\boldsymbol{y}_{\mathrm{f}}(t) + \boldsymbol{B}(z)\boldsymbol{u}_{\mathrm{f}}(t) + D(z)\boldsymbol{v}(t)$$
$$= \boldsymbol{\phi}(t)\boldsymbol{d} + \boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}_{\mathrm{f}}(t) + \boldsymbol{v}(t). \tag{8}$$

Let $\boldsymbol{\phi}_i^{\mathrm{T}}(t) \in \mathbb{R}^{1 \times n_d}$ be the ith row of the information matrix $\boldsymbol{\phi}(t)$:

$$\boldsymbol{\phi}(t) := [\boldsymbol{\phi}_1(t), \boldsymbol{\phi}_2(t), \cdots, \boldsymbol{\phi}_m(t)]^{\mathrm{T}} \in \mathbb{R}^{m \times n_d}.$$

Similarly, let $\boldsymbol{\theta}_i(t) \in \mathbb{R}^n$ be the ith column of the parameter matrix $\boldsymbol{\theta}$ and $y_{\mathrm{f}i}(t)$ be the ith element of the filtered output vector $\boldsymbol{y}_{\mathrm{f}}(t)$, that is

$$\boldsymbol{\theta} := [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_m] \in \mathbb{R}^{n \times m},$$
$$\boldsymbol{y}_{\mathrm{f}}(t) := [y_{\mathrm{f}1}(t), y_{\mathrm{f}2}(t), \cdots, y_{\mathrm{f}m}(t)]^{\mathrm{T}} \in \mathbb{R}^m.$$

Equation (8) can be explicitly written as

$$\begin{bmatrix} y_{\mathrm{f}1}(t) \\ y_{\mathrm{f}2}(t) \\ \vdots \\ y_{\mathrm{f}m}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}_1^{\mathrm{T}}(t) \\ \boldsymbol{\phi}_2^{\mathrm{T}}(t) \\ \vdots \\ \boldsymbol{\phi}_m^{\mathrm{T}}(t) \end{bmatrix} \boldsymbol{d} + \begin{bmatrix} \boldsymbol{\theta}_1^{\mathrm{T}} \\ \boldsymbol{\theta}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\theta}_m^{\mathrm{T}} \end{bmatrix} \boldsymbol{\varphi}_{\mathrm{f}}(t) + \begin{bmatrix} v_1(t) \\ v_2(t) \\ \vdots \\ v_m(t) \end{bmatrix}. \tag{9}$$

Then we can decompose the filtered system in (9) into m filtered subsystem identification models

$$y_{\mathrm{f}i}(t) = \boldsymbol{\phi}_i^{\mathrm{T}}(t)\boldsymbol{d} + \boldsymbol{\theta}_i^{\mathrm{T}}\boldsymbol{\varphi}_{\mathrm{f}}(t) + v_i(t)$$
$$= \boldsymbol{\phi}_i^{\mathrm{T}}(t)\boldsymbol{d} + \boldsymbol{\varphi}_{\mathrm{f}}^{\mathrm{T}}(t)\boldsymbol{\theta}_i + v_i(t)$$
$$= [\boldsymbol{\phi}_i^{\mathrm{T}}(t), \boldsymbol{\varphi}_{\mathrm{f}}^{\mathrm{T}}(t)] \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{\theta}_i \end{bmatrix} + v_i(t), \ i = 1, 2, \cdots, m. \tag{10}$$

Define the subsystem information vector

$$\boldsymbol{\psi}_i(t) := \begin{bmatrix} \boldsymbol{\phi}_i(t) \\ \boldsymbol{\varphi}_{\mathrm{f}}(t) \end{bmatrix} \in \mathbb{R}^{n+n_d}.$$

The filtered subsystems in (10) can be expressed as

$$y_{\mathrm{f}i}(t) = \boldsymbol{\psi}_i^{\mathrm{T}}(t) \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{\theta}_i \end{bmatrix} + v_i(t). \tag{11}$$

From the definition of $\boldsymbol{w}(t)$, we can obtain

$$C(z)\boldsymbol{w}(t) = D(z)\boldsymbol{v}(t). \tag{12}$$

Rewrite Equation (12) as the following form

$$\boldsymbol{w}(t) = \boldsymbol{\Omega}(t)\boldsymbol{c} + \boldsymbol{\phi}(t)\boldsymbol{d} + \boldsymbol{v}(t),$$

where $\boldsymbol{\Omega}(t)$ is the noise information matrix and $\boldsymbol{c}$ is the noise parameter vector:

$$\boldsymbol{\Omega}(t) := [-\boldsymbol{w}(t-1), -\boldsymbol{w}(t-2), \cdots, -\boldsymbol{w}(t-n_c)] \in \mathbb{R}^{m \times n_c},$$
$$\boldsymbol{c} := [c_1, c_2, \cdots, c_{n_c}]^{\mathrm{T}} \in \mathbb{R}^{n_c}.$$

Define an intermediate variable $\boldsymbol{w}_{\mathrm{n}}(t) := \boldsymbol{w}(t) - \boldsymbol{\phi}(t)\boldsymbol{d} \in \mathbb{R}^m$. Then another similar formula can be derived for $\boldsymbol{w}_{\mathrm{n}}(t)$:

$$\boldsymbol{w}_{\mathrm{n}}(t) = \boldsymbol{\Omega}(t)\boldsymbol{c} + \boldsymbol{v}(t). \tag{13}$$

Equations (11) and (13) are the filter hierarchical subsystem identification model and the noise model for the multivariable CARARMA system in (1). Compared with the identification model in (6), the filter identification models divide the system parameters into two parts, Equation (11) contains the parameter vectors $\boldsymbol{\theta}_i$ and $\boldsymbol{d}$, and the noise vector $\boldsymbol{c}$ is in the noise model (13). Note that the filtered subsystem identification model is disturbed by the white noise $\boldsymbol{v}(t)$, in other word, we transform the origin system in (1) which has colored noise to m filtered subsystem models with white noise and a noise model.

Based on the filtered subsystem identification model (11) and the noise model (13), using the negative gradient search principle gives

$$\begin{bmatrix} \hat{\boldsymbol{d}}(t) \\ \hat{\boldsymbol{\theta}}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{d}}(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} + \frac{\boldsymbol{\psi}_i(t)}{r_1(t)}$$
$$\times \left\{ y_{\mathrm{f}i}(t) - \boldsymbol{\psi}_i^{\mathrm{T}}(t) \begin{bmatrix} \hat{\boldsymbol{d}}(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} \right\}, \tag{14}$$

$$r_{1,i}(t) = r_{1,i}(t-1) + \|\boldsymbol{\psi}_i(t)\|^2, \tag{15}$$

$$\hat{\boldsymbol{c}}(t) = \hat{\boldsymbol{c}}(t-1) + \frac{\boldsymbol{\Omega}^{\mathrm{T}}(t)}{r_2(t)}[\boldsymbol{w}_{\mathrm{n}}(t) - \boldsymbol{\Omega}(t)\hat{\boldsymbol{c}}(t-1)], \tag{16}$$

$$r_2(t) = r_2(t-1) + \|\boldsymbol{\Omega}(t)\|^2. \tag{17}$$

Note that the information vector $\boldsymbol{\psi}_i(t)$ $(i = 1, 2, \cdots, m)$ contains the unknown variable $\boldsymbol{v}(t-j)$, and the information matrix $\boldsymbol{\Omega}(t)$ is made of the unmeasured intermediate variable $\boldsymbol{w}(t-j)$. Moreover, since the parameter vector $\boldsymbol{c}$ is unknown, we cannot obtain the filtered input vector $\boldsymbol{u}_{\mathrm{f}}(t)$ and the filtered output vector $\boldsymbol{y}_{\mathrm{f}}(t)$. All these imply that the above relations are impossible to generate the estimates $\hat{\boldsymbol{d}}(t)$, $\hat{\boldsymbol{\theta}}_i(t)$ and $\hat{\boldsymbol{c}}(t)$ directly. The solution here is to replace the unknown variables $\boldsymbol{w}(t-j)$ and $\boldsymbol{v}(t-j)$ with the outputs $\hat{\boldsymbol{w}}(t-j)$ and $\hat{\boldsymbol{v}}(t-j)$ of the auxiliary models. Then we can obtain the estimates of $\boldsymbol{\Omega}(t)$ and $\boldsymbol{\phi}(t)$:

$$\hat{\boldsymbol{\Omega}}(t) := [-\hat{\boldsymbol{w}}(t-1), -\hat{\boldsymbol{w}}(t-2), \cdots, -\hat{\boldsymbol{w}}(t-n_c)] \in \mathbb{R}^{m \times n_c},$$
$$\hat{\boldsymbol{\phi}}(t) := [\hat{\boldsymbol{v}}(t-1), \hat{\boldsymbol{v}}(t-2), \cdots, \hat{\boldsymbol{v}}(t-n_d)]$$
$$= [\hat{\boldsymbol{\phi}}_1(t), \hat{\boldsymbol{\phi}}_2(t), \cdots, \hat{\boldsymbol{\phi}}_m(t)]^{\mathrm{T}} \in \mathbb{R}^{m \times n_d}.$$

According to Equation (4), replacing $\boldsymbol{\theta}$ with its estimate $\hat{\boldsymbol{\theta}}(t)$ gives

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t).$$

From the definition of the intermediate variable $\boldsymbol{w}_{\mathrm{n}}(t)$, replacing $\boldsymbol{w}(t)$ and $\boldsymbol{d}$ with $\hat{\boldsymbol{w}}(t-1)$ and $\hat{\boldsymbol{d}}(t-1)$ gives

$$\hat{\boldsymbol{w}}_{\mathrm{n}}(t) = \hat{\boldsymbol{w}}(t-1) - \hat{\boldsymbol{\phi}}(t)\hat{\boldsymbol{d}}(t-1)$$
$$= \boldsymbol{y}(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t) - \hat{\boldsymbol{\phi}}(t)\hat{\boldsymbol{d}}(t-1).$$

Use the parameter estimate

$$\hat{\boldsymbol{c}}(t) := [\hat{c}_1(t), \hat{c}_2(t), \cdots, \hat{c}_{n_c}(t)]^{\mathrm{T}} \in \mathbb{R}^{n_c \times 1}$$

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

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

Here, we use the filter $\hat{C}(t,z)$ to filter the input vector and the output vector, and obtain the estimates of the filtered input vector $\boldsymbol{u}_{\mathrm{f}}(t)$ and the filtered output vector $\boldsymbol{y}_{\mathrm{f}}(t)$:

$$\begin{aligned}
\hat{\boldsymbol{u}}_{\mathrm{f}}(t) &= \hat{C}(t,z)\boldsymbol{u}(t) \\
&= \boldsymbol{u}(t) + \hat{c}_1(t)\boldsymbol{u}(t-1) + \cdots + \hat{c}_{n_c}(t)\boldsymbol{u}(t-n_c) \\
&= \boldsymbol{u}(t) + [\boldsymbol{u}(t-1), \boldsymbol{u}(t-2), \cdots, \boldsymbol{u}(t-n_c)]\hat{\boldsymbol{c}}(t), \\
\hat{\boldsymbol{y}}_{\mathrm{f}}(t) &= \hat{C}(t,z)\boldsymbol{y}(t) \\
&= \boldsymbol{y}(t) + \hat{c}_1(t)\boldsymbol{y}(t-1) + \cdots + \hat{c}_{n_c}(t)\boldsymbol{y}(t-n_c) \\
&= \boldsymbol{y}(t) + [\boldsymbol{y}(t-1), \boldsymbol{y}(t-2), \cdots, \boldsymbol{y}(t-n_c)]\hat{\boldsymbol{c}}(t) \\
&= [\hat{y}_{\mathrm{f}1}(t), \hat{y}_{\mathrm{f}2}(t), \cdots, \hat{y}_{\mathrm{f}m}(t)]^{\mathrm{T}}.
\end{aligned}$$

Then the estimate of $\boldsymbol{\varphi}_{\mathrm{f}}(t)$ can be defined as

$$\begin{aligned}
\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t) := &[-\hat{\boldsymbol{y}}_{\mathrm{f}}^{\mathrm{T}}(t-1), -\hat{\boldsymbol{y}}_{\mathrm{f}}^{\mathrm{T}}(t-2), \cdots, -\hat{\boldsymbol{y}}_{\mathrm{f}}^{\mathrm{T}}(t-n_a), \\
&\hat{\boldsymbol{u}}_{\mathrm{f}}^{\mathrm{T}}(t-1), \hat{\boldsymbol{u}}_{\mathrm{f}}^{\mathrm{T}}(t-2), \cdots, \hat{\boldsymbol{u}}_{\mathrm{f}}^{\mathrm{T}}(t-n_b)]^{\mathrm{T}} \in \mathbb{R}^n.
\end{aligned}$$

Subsequently, the information vector $\hat{\boldsymbol{\psi}}_i(t)$ is constructed by using $\hat{\boldsymbol{\phi}}_i(t)$ and $\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t)$:

$$\hat{\boldsymbol{\psi}}_i(t) = \begin{bmatrix} \hat{\boldsymbol{\phi}}_i(t) \\ \hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t) \end{bmatrix} \in \mathbb{R}^{n+n_d}.$$

According to Equation (8), replacing $\boldsymbol{y}_{\mathrm{f}}(t)$, $\boldsymbol{\phi}(t)$ and $\boldsymbol{\varphi}_{\mathrm{f}}(t)$ with their estimates $\hat{\boldsymbol{y}}_{\mathrm{f}}(t)$, $\hat{\boldsymbol{\phi}}(t)$ and $\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t)$, we can get the estimate of $\boldsymbol{v}(t)$:

$$\hat{\boldsymbol{v}}(t) = \hat{\boldsymbol{y}}_{\mathrm{f}}(t) - \hat{\boldsymbol{\phi}}(t)\hat{\boldsymbol{d}}(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t)\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t). \tag{18}$$

Besides the unknown variables involved in (14)–(17), there is still a problem that each subsystem will produce a parameter estimation vector $\hat{\boldsymbol{d}}(t)$ of the same parameter vector $\boldsymbol{d}$ at each time t. In order to make it clear, we use $\hat{\boldsymbol{d}}_i(t)$ to represent the estimate of Subsystem i. Then, replace $y_{\mathrm{f}i}(t)$ and $\boldsymbol{\psi}_i(t)$ in (14)–(15) with their estimates $\hat{y}_{\mathrm{f}i}(t)$ and $\hat{\boldsymbol{\psi}}_i(t)$, and replace $\boldsymbol{w}_{\mathrm{n}}(t)$ and $\boldsymbol{\Omega}(t)$ in (16)–(17) with their estimates $\hat{\boldsymbol{w}}_{\mathrm{n}}(t)$ and $\hat{\boldsymbol{\Omega}}(t)$. Through these replacement, we modify Equations (14)–(17) as

$$\begin{aligned}
\begin{bmatrix} \hat{\boldsymbol{d}}_i(t) \\ \hat{\boldsymbol{\theta}}_i(t) \end{bmatrix} &= \begin{bmatrix} \hat{\boldsymbol{d}}_i(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} + \frac{\hat{\boldsymbol{\psi}}_i(t)}{r_{1,i}(t)} \\
&\times \left\{ \hat{y}_{\mathrm{f}i}(t) - \hat{\boldsymbol{\psi}}_i^{\mathrm{T}}(t) \begin{bmatrix} \hat{\boldsymbol{d}}_i(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} \right\},
\end{aligned} \tag{19}$$

$$r_{1,i}(t) = r_{1,i}(t-1) + \|\hat{\boldsymbol{\psi}}_i(t)\|^2, \ i = 1, 2, \cdots, m, \tag{20}$$

$$\hat{\boldsymbol{c}}(t) = \hat{\boldsymbol{c}}(t-1) + \frac{\hat{\boldsymbol{\Omega}}^{\mathrm{T}}(t)}{r_2(t)}[\hat{\boldsymbol{w}}_{\mathrm{n}}(t) - \hat{\boldsymbol{\Omega}}(t)\hat{\boldsymbol{c}}(t-1)], \tag{21}$$

$$r_2(t) = r_2(t-1) + \|\hat{\boldsymbol{\Omega}}(t)\|^2. \tag{22}$$

It is worth noting that there are $\hat{\boldsymbol{d}}_1(t)$, $\hat{\boldsymbol{d}}_2(t)$, $\cdots$, $\hat{\boldsymbol{d}}_m(t)$ for $i = 1, 2, \cdots, m$ in (19)–(22), and this leads to many redundant parameter estimates. However, cutting down the redundant parameter estimates and improving the parameter estimation accuracy are our aims to explore new identification methods. In general, we usually desire that the proposed algorithm is convergent, that is to say, the parameter estimates approach their true values as the time t increases. Therefore, we can assume that the parameter estimate $\hat{\boldsymbol{d}}_{i-1}(t)$ is closer to the true value than the parameter estimate $\hat{\boldsymbol{d}}_i(t-1)$. Based on the coupling identification concept, for $i = 2, 3, \cdots, m$, use $\hat{\boldsymbol{d}}_{i-1}(t)$ to replace $\hat{\boldsymbol{d}}_i(t-1)$ and for $i = 1$, use $\hat{\boldsymbol{d}}_m(t-1)$ to replace $\hat{\boldsymbol{d}}_1(t-1)$. Thus, we can obtain the following F-PC-GESG algorithm:

$$\begin{aligned}
\begin{bmatrix} \hat{\boldsymbol{d}}_i(t) \\ \hat{\boldsymbol{\theta}}_i(t) \end{bmatrix} &= \begin{bmatrix} \hat{\boldsymbol{d}}_{i-1}(t) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} + \frac{\hat{\boldsymbol{\psi}}_i(t)}{r_{1,i}(t)} \\
&\times \left\{ \hat{y}_{\mathrm{f}i}(t) - \hat{\boldsymbol{\psi}}_i^{\mathrm{T}}(t) \begin{bmatrix} \hat{\boldsymbol{d}}_{i-1}(t) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} \right\},
\end{aligned} \tag{23}$$

$$r_{1,i}(t) = r_{1,i}(t-1) + \|\hat{\boldsymbol{\psi}}_i(t)\|^2, \ i = 2, 3, \cdots, m, \tag{24}$$

$$\begin{aligned}
\begin{bmatrix} \hat{\boldsymbol{d}}_1(t) \\ \hat{\boldsymbol{\theta}}_1(t) \end{bmatrix} &= \begin{bmatrix} \hat{\boldsymbol{d}}_m(t-1) \\ \hat{\boldsymbol{\theta}}_1(t-1) \end{bmatrix} + \frac{\hat{\boldsymbol{\psi}}_1(t)}{r_{1,1}(t)} \\
&\times \left\{ \hat{y}_{\mathrm{f}1}(t) - \hat{\boldsymbol{\psi}}_1^{\mathrm{T}}(t) \begin{bmatrix} \hat{\boldsymbol{d}}_m(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} \right\},
\end{aligned} \tag{25}$$

$$r_{1,1}(t) = r_{1,1}(t-1) + \|\hat{\boldsymbol{\psi}}_1(t)\|^2, \tag{26}$$

$$\hat{\boldsymbol{c}}(t) = \hat{\boldsymbol{c}}(t-1) + \frac{\hat{\boldsymbol{\Omega}}^{\mathrm{T}}(t)}{r_2(t)}[\hat{\boldsymbol{w}}_{\mathrm{n}}(t) - \hat{\boldsymbol{\Omega}}(t)\hat{\boldsymbol{c}}(t-1)], \tag{27}$$

$$r_2(t) = r_2(t-1) + \|\hat{\boldsymbol{\Omega}}(t)\|^2, \tag{28}$$

$$\hat{\boldsymbol{y}}_{\mathrm{f}}(t) = \boldsymbol{y}(t) + [\boldsymbol{y}(t-1), \cdots, \boldsymbol{y}(t-n_c)]\hat{\boldsymbol{c}}(t) \tag{29}$$

$$= [\hat{y}_{\mathrm{f}1}(t), \hat{y}_{\mathrm{f}2}(t), \cdots, \hat{y}_{\mathrm{f}m}(t)]^{\mathrm{T}}, \tag{30}$$

$$\hat{\boldsymbol{u}}_{\mathrm{f}}(t) = \boldsymbol{u}(t) + [\boldsymbol{u}(t-1), \cdots, \boldsymbol{u}(t-n_c)]\hat{\boldsymbol{c}}(t), \tag{31}$$

$$\hat{\boldsymbol{\psi}}_i(t) = [\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(t), \hat{\boldsymbol{\varphi}}_{\mathrm{f}}^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{32}$$

$$\begin{aligned}
\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t) = &[-\hat{\boldsymbol{y}}_{\mathrm{f}}^{\mathrm{T}}(t-1), -\hat{\boldsymbol{y}}_{\mathrm{f}}^{\mathrm{T}}(t-2), \cdots, -\hat{\boldsymbol{y}}_{\mathrm{f}}^{\mathrm{T}}(t-n_a), \\
&\boldsymbol{u}_{\mathrm{f}}^{\mathrm{T}}(t-1), \boldsymbol{u}_{\mathrm{f}}^{\mathrm{T}}(t-2), \cdots, \boldsymbol{u}_{\mathrm{f}}^{\mathrm{T}}(t-n_b)]^{\mathrm{T}}, \tag{33}
\end{aligned}$$

$$\hat{\boldsymbol{\phi}}(t) = [\hat{\boldsymbol{v}}(t-1), \hat{\boldsymbol{v}}(t-2), \cdots, \hat{\boldsymbol{v}}(t-n_d)] \tag{34}$$

$$= [\hat{\boldsymbol{\phi}}_1(t), \hat{\boldsymbol{\phi}}_2(t), \cdots, \hat{\boldsymbol{\phi}}_m(t)]^{\mathrm{T}}, \tag{35}$$

$$\hat{\boldsymbol{\Omega}}(t) = [-\hat{\boldsymbol{w}}(t-1), -\hat{\boldsymbol{w}}(t-2), \cdots, -\hat{\boldsymbol{w}}(t-n_c)], \tag{36}$$

$$\begin{aligned}
\boldsymbol{\varphi}(t) = &[-\boldsymbol{y}^{\mathrm{T}}(t-1), -\boldsymbol{y}^{\mathrm{T}}(t-2), \cdots, -\boldsymbol{y}^{\mathrm{T}}(t-n_a), \\
&\boldsymbol{u}^{\mathrm{T}}(t-1), \boldsymbol{u}^{\mathrm{T}}(t-2), \cdots, \boldsymbol{u}^{\mathrm{T}}(t-n_b)]^{\mathrm{T}}, \tag{37}
\end{aligned}$$

$$\hat{\boldsymbol{w}}_{\mathrm{n}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t) - \hat{\boldsymbol{\phi}}(t)\hat{\boldsymbol{d}}_m(t-1), \tag{38}$$

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t), \tag{39}$$

$$\hat{\boldsymbol{v}}(t) = \hat{\boldsymbol{y}}_{\mathrm{f}}(t) - \hat{\boldsymbol{\phi}}(t)\hat{\boldsymbol{d}}_m(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t)\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t), \tag{40}$$

$$\hat{\boldsymbol{\theta}}(t) = [\hat{\boldsymbol{\theta}}_1(t), \hat{\boldsymbol{\theta}}_2(t), \cdots, \hat{\boldsymbol{\theta}}_m(t)]. \tag{41}$$
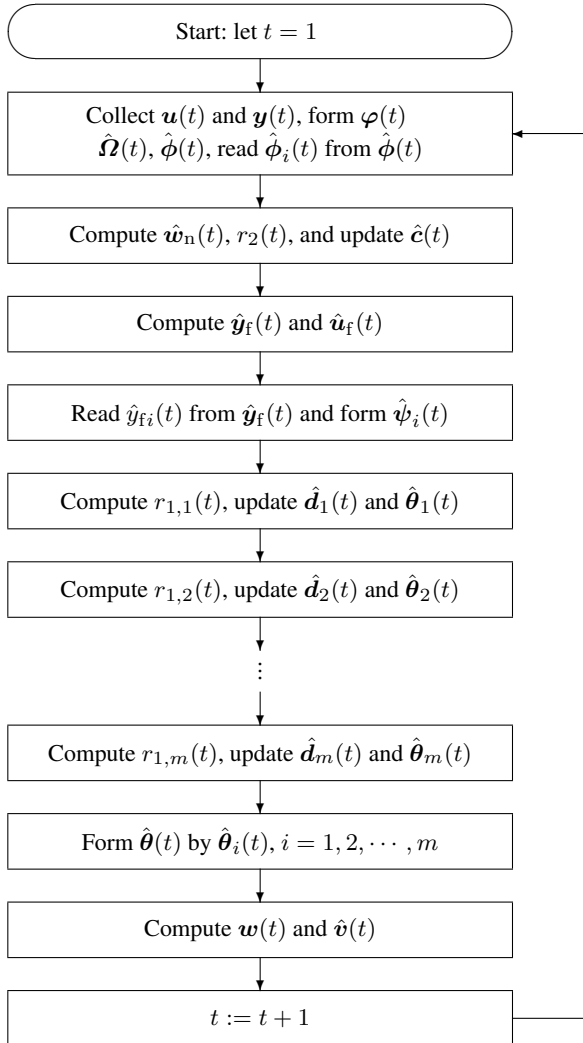
Through the F-PC-GESG algorithm in (23)–(41), we can get the estimates $\hat{\boldsymbol{c}}(t)$, $\hat{\boldsymbol{d}}(t) := \hat{\boldsymbol{d}}_m(t)$ and $\hat{\boldsymbol{\theta}}(t) = [\hat{\boldsymbol{\theta}}_1(t), \hat{\boldsymbol{\theta}}_2(t), \cdots, \hat{\boldsymbol{\theta}}_m(t)]$. So $\hat{\boldsymbol{d}}(t)$ in (18) is modified as $\hat{\boldsymbol{d}}_m(t)$ when we calculate $\hat{\boldsymbol{v}}(t)$.

To state the algorithm clearly, we list the steps involved in the F-PC-GESG algorithm in (23)–(41) as follows.

1. Set the initial values: let $t = 1$, $\hat{\boldsymbol{c}}(0) = \boldsymbol{1}_{n_c}/p_0$, $\hat{\boldsymbol{d}}_m(0) = \boldsymbol{1}_{n_d}/p_0$, $\hat{\boldsymbol{\theta}}_{i,0} = \boldsymbol{1}_n/p_0$, $r_{1,i}(0) = 1$ $(i = 1, 2, \cdots, m)$, $r_2(0) = 1$, $\hat{\boldsymbol{y}}_{\mathrm{f}}(t-j) = \boldsymbol{1}_m/p_0$, $\hat{\boldsymbol{u}}_{\mathrm{f}}(t-j) = \boldsymbol{1}_r/p_0$, $\hat{\boldsymbol{w}}(0) = \boldsymbol{1}_m/p_0$, $\hat{\boldsymbol{v}}(0) = \boldsymbol{1}_m/p_0$, $p_0 = 10^6$.
2. Collect the input and output data $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$, and construct the information vector $\boldsymbol{\varphi}(t)$ using (37), form $\hat{\boldsymbol{\phi}}(t)$ using (34), read $\hat{\boldsymbol{\phi}}_i(t)$ from (35), and construct $\hat{\boldsymbol{\Omega}}(t)$ by (36).
3. Compute $\hat{\boldsymbol{w}}_{\mathrm{n}}(t)$ and $r_2(t)$ using (38) and (28), update the parameter estimate $\hat{\boldsymbol{c}}(t)$ using (27).
4. Compute $\hat{\boldsymbol{y}}_{\mathrm{f}}(t)$ by (29), read $\hat{y}_{\mathrm{f}i}(t)$ from (30), compute $\hat{\boldsymbol{u}}_{\mathrm{f}}(t)$ by (31), form $\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t)$ using (33), then construct $\hat{\boldsymbol{\psi}}_i(t)$ by (32).
5. Compute $r_{1,1}(t)$ by (26), update $\hat{\boldsymbol{d}}_1(t)$ and $\hat{\boldsymbol{\theta}}_1(t)$ using (25).
6. For $i = 2, 3, \cdots, m$, compute $r_{1,i}(t)$ using (24) and update $\hat{\boldsymbol{d}}_i(t)$ and $\hat{\boldsymbol{\theta}}_i(t)$ using (23).

7. Form $\hat{\boldsymbol{\theta}}(t)$ by (41), compute $\hat{\boldsymbol{w}}(t)$ and $\hat{\boldsymbol{v}}(t)$ using (39)–(40).
8. Increase t by 1 and go to Step 2.

The flowchart of computing $\hat{\boldsymbol{c}}(t)$, $\hat{\boldsymbol{\theta}}(t)$ and $\hat{\boldsymbol{d}}_m(t)$ in the F-PC-GESG algorithm is shown in Figure 1.

Fig. 1: The flowchart of computing the F-PC-GESG parameter estimates $\hat{\boldsymbol{\theta}}(t)$, $\hat{\boldsymbol{d}}_m(t)$ and $\hat{\boldsymbol{c}}(t)$.

Remark 2: From the F-PC-GESG algorithm in (23)–(41), we can see that only part of the parameters are coupled. To be more specific, only $\hat{\boldsymbol{d}}_i(t)$ are coupled between the filtered subsystems. The parameter vectors $\hat{\boldsymbol{\theta}}_i(t)$ are independent, because every subsystem has a corresponding $\hat{\boldsymbol{\theta}}_i(t)$. This is the meaning of the partially-coupled algorithm.

Remark 3: To improve the performance of the F-PC-GESG algorithm, we introduce forgetting factors $\lambda_{1,i}$ and λ_2 to the F-PC-GESG algorithm. Replace (24), (26) and (28) with Equations (42)–(44), and remain other formulas unchanged for the F-PC-GESG algorithm in (23)–(41):

$$r_{1,i}(t) = \lambda_{1,i} r_{1,i}(t-1) + \|\hat{\boldsymbol{\psi}}_i(t)\|^2, \quad i = 2, 3, \cdots, m, \quad (42)$$

$$r_{1,1}(t) = \lambda_{1,1} r_{1,1}(t-1) + \|\hat{\boldsymbol{\psi}}_1(t)\|^2, \quad 0 \le \lambda_{1,i} < 1, \quad (43)$$

$$r_2(t) = \lambda_2 r_2(t-1) + \|\hat{\boldsymbol{\Omega}}(t)\|^2, \quad 0 \le \lambda_2 < 1. \quad (44)$$

Since the F-PC-GESG algorithm produces the data saturation problem with the time length increasing. The forgetting factors can

reduce the weight of the past data and improve the estimation accuracy. The forgetting factors $\lambda_{1,i}$ and λ_2 can be the same or different in (42)–(44).

Remark 4: Compared with the algorithms in the previous work [33] and [34], this paper considers on the parameter identification problems for multivariable equation-error system with the autoregressive moving average noises. In order to eliminate the effect of the colored noises, the F-PC-GESG algorithm uses the data filtering technique to transform the system into a filtered system and a noise system. Then we decompose the filtered system into m subsystems and estimate the parameters by the coupling identification concept.

4 The multivariable generalized extended stochastic gradient algorithm

In the following, we discuss the multivariable generalized extended stochastic gradient (M-GESG) algorithm for comparison. As we can see from the identification model (6), the parameter vector $\boldsymbol{\vartheta}$ contains parameters $\boldsymbol{\theta}$, $c_i, (i = 1, 2, \cdots, n_c)$ and $d_i, (i = 1, 2, \cdots, n_d)$ of System (1).

In consideration of the unknown variables $\boldsymbol{w}(t - j)$ and $\boldsymbol{v}(t - j)$, we also employ the auxiliary model method as we do in the F-PC-GESG algorithm. Let $\hat{\boldsymbol{w}}(t - j)$ and $\hat{\boldsymbol{v}}(t - j)$ be the outputs of the auxiliary models, and define the estimate of $\boldsymbol{\psi}(t)$:

$$\hat{\boldsymbol{\psi}}(t) := [-\hat{\boldsymbol{w}}(t - 1), -\hat{\boldsymbol{w}}(t - 2), \cdots, -\hat{\boldsymbol{w}}(t - n_c),$$
$$\hat{\boldsymbol{v}}(t - 1), \hat{\boldsymbol{v}}(t - 2), \cdots, \hat{\boldsymbol{v}}(t - n_d)] \in \mathbb{R}^{m \times (n_c + n_d)}.$$

Then, the estimate of the information matrix $\boldsymbol{\Phi}(t)$ can be constructed by $\boldsymbol{\varphi}(t)$ and $\hat{\boldsymbol{\phi}}(t)$:

$$\hat{\boldsymbol{\Phi}}(t) := [\hat{\boldsymbol{\psi}}(t), \boldsymbol{\varphi}^{\mathrm{T}}(t) \otimes \boldsymbol{I}_m] \in \mathbb{R}^{m \times n_0}.$$

Replacing $\boldsymbol{\Phi}(t)$, $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ with their estimates $\hat{\boldsymbol{\Phi}}(t)$, $\hat{\boldsymbol{\theta}}(t)$ and $\hat{\boldsymbol{\vartheta}}(t)$ in (4) and (6), $\hat{\boldsymbol{w}}(t)$ and $\hat{\boldsymbol{v}}(t)$ can be computed by

$$\hat{\boldsymbol{w}}(t) := \boldsymbol{y}(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t),$$
$$\hat{\boldsymbol{v}}(t) := \boldsymbol{y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\vartheta}}(t).$$

Replacing $\boldsymbol{\Phi}(t)$ in (6) with its estimate $\hat{\boldsymbol{\Phi}}(t)$ and using the negative gradient search principle, we can obtain the following M-GESG algorithm:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)}{r(t)}[\boldsymbol{y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\vartheta}}(t-1)], \quad (45)$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\Phi}}(t)\|^2, \quad (46)$$

$$\hat{\boldsymbol{\Phi}}(t) = [\hat{\boldsymbol{\psi}}(t), \boldsymbol{\varphi}^{\mathrm{T}}(t) \otimes \boldsymbol{I}_m], \quad (47)$$

$$\boldsymbol{\varphi}(t) = [-\boldsymbol{y}^{\mathrm{T}}(t - 1), -\boldsymbol{y}^{\mathrm{T}}(t - 2), \cdots, -\boldsymbol{y}^{\mathrm{T}}(t - n_a),$$
$$\boldsymbol{u}^{\mathrm{T}}(t - 1), \boldsymbol{u}^{\mathrm{T}}(t - 2), \cdots, \boldsymbol{u}^{\mathrm{T}}(t - n_b)]^{\mathrm{T}}, \quad (48)$$

$$\hat{\boldsymbol{\psi}}(t) = [-\hat{\boldsymbol{w}}(t - 1), -\hat{\boldsymbol{w}}(t - 2), \cdots, -\hat{\boldsymbol{w}}(t - n_c),$$
$$\hat{\boldsymbol{v}}(t - 1), \hat{\boldsymbol{v}}(t - 2), \cdots, \hat{\boldsymbol{v}}(t - n_d)], \quad (49)$$

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t), \quad (50)$$

$$\hat{\boldsymbol{v}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\vartheta}}(t), \quad (51)$$

$$\hat{\boldsymbol{\vartheta}}(t) = \begin{bmatrix} \hat{\boldsymbol{\rho}}(t) \\ \mathrm{col}[\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t)] \end{bmatrix}. \quad (52)$$

The procedures involved in the M-GESG algorithm in (45)–(52) are listed as follows.

1. Set the initial values: let $t = 1$, $\hat{\boldsymbol{\vartheta}}(0) = \boldsymbol{1}_{n_0}/p_0$, $r(0) = 1$, $\hat{\boldsymbol{v}}(-j) = \boldsymbol{1}_m/p_0$, $\hat{\boldsymbol{w}}(-j) = \boldsymbol{1}_m/p_0$, $j \geqslant 0$, $p_0 = 10^6$.

2. Collect the observation data $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$, and construct the information vector and matrices $\boldsymbol{\varphi}(t)$, $\hat{\boldsymbol{\psi}}(t)$ and $\hat{\boldsymbol{\Phi}}(t)$ using (48)–(49) and (47).
3. Compute $r(t)$ using (46) and update the parameter estimation vector $\hat{\boldsymbol{\vartheta}}(t)$ by (45).
4. Read $\hat{\boldsymbol{\theta}}(t)$ and $\hat{\boldsymbol{\rho}}(t)$ from $\hat{\boldsymbol{\vartheta}}(t)$ by (52), and compute $\hat{\boldsymbol{w}}(t)$ and $\hat{\boldsymbol{v}}(t)$ by (50)–(51).
5. Increase t by 1 and go to Step 2.

Remark 5: By means of the auxiliary model identification idea, the M-GESG algorithm in (45)–(52) handles the unknown information matrix $\boldsymbol{\Phi}(t)$ by replacing it with its estimate $\hat{\boldsymbol{\Phi}}(t)$ in order to guarantee the realization of the algorithm.

Remark 6: Although the M-GESG algorithm in (45)–(52) can produce the parameter estimation vector $\hat{\boldsymbol{\vartheta}}(t)$, the weakness is that $\hat{\boldsymbol{\Phi}}(t)$ is a large dimension informational matrix, which gives rise to heavy computational burden. Differing from the M-GESG algorithm, the F-PC-GESG algorithm in (23)–(41) divides the system in (1) into one noise model and m filtered subsystems. Moreover, the F-PC-GESG algorithm uses the data filtering to reduce the influence of the noise and improve the estimation accuracy. For every t, the F-PC-GESG algorithm first gets the noise parameter vector $\hat{\boldsymbol{c}}(t)$, and uses it to calculate the filtered input and output, then identifies the parameters $\hat{\boldsymbol{\theta}}_i(t)$ and $\hat{\boldsymbol{d}}_i(t)$ based on the coupled relations between these filtered subsystems. This property also can be seen in Figure 1.

Remark 7: The computational efficiency of an algorithm can be measured by using flops. The computational costs of the M-GESG algorithm and the F-PC-GESG algorithm are listed in Tables 1–2, where $n_0 = n_c + n_d + mn$, $n_1 = n + n_d$ and $n = mn_a + rn_b$. In order to make it clear, we take an example. Assume that $m = 10$, $r = 10$, $n_a = 10$, $n_b = 10$, $n_c = 10$ and $n_d = 10$. Then we can calculate $N_1 - N_2 = 167630 - 17811 = 149819$. It shows that the F-PC-GESG algorithm has a higher computational efficiency than the M-GESG algorithm. For the system with high orders, the advantage of the F-PC-GESG algorithm becomes more obvious.

The proposed algorithms in this paper can be developed to multivariable bilinear systems with colored noises [36–38], and can combine the neural network methods [39] and the kernel methods [40, 41] to study parameter identification of different systems [42–44].

5 Examples

In this section, we give two numerical simulations to show the effectiveness of the proposed algorithm.

Example 1. Consider the following multivariable system with two-input two-output:

$$\boldsymbol{A}(z)\boldsymbol{y}(t) = \boldsymbol{B}(z)\boldsymbol{u}(t) + \frac{D(z)}{C(z)}\boldsymbol{v}(t),$$

$$\boldsymbol{A}(z) = \boldsymbol{I}_2 + \left[\begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array}\right] z^{-1}$$

$$= \left[\begin{array}{cc} 1 + 0.24z^{-1} & 0.94z^{-1} \\ -0.80z^{-1} & 1 + 1.05z^{-1} \end{array}\right],$$

$$\boldsymbol{B}(z) = \left[\begin{array}{cc} b_{11} & b_{12} \\ b_{21} & b_{22} \end{array}\right] z^{-1} = \left[\begin{array}{cc} 0.10z^{-1} & 0.15z^{-1} \\ 0.12z^{-1} & -0.10z^{-1} \end{array}\right],$$

$$C(z) = 1 + c_1 z^{-1} = 1 + 0.14z^{-1},$$

$$D(z) = 1 + d_1 z^{-1} = 1 - 0.80z^{-1},$$

$$\boldsymbol{\theta}^{\mathrm{T}} = \left[\begin{array}{cccc} 0.24 & 0.94 & 0.10 & 0.15 \\ -0.80 & 1.05 & 0.12 & -0.10 \end{array}\right],$$

$$\boldsymbol{\rho} = [c_1, d_1]^{\mathrm{T}} = [0.14, -0.80]^{\mathrm{T}},$$

$$\boldsymbol{\vartheta} = \left[\begin{array}{c} \boldsymbol{\rho} \\ \mathrm{col}[\boldsymbol{\theta}^{\mathrm{T}}] \end{array}\right].$$

To apply the proposed method, it is necessary to collect the input and output data. Here, we take the inputs $\{u_1(t)\}$ and $\{u_2(t)\}$ as two independent persistent excitation signal sequences with zero mean and unit variances, and $\{v_1(t)\}$ and $\{v_2(t)\}$ are taken as two white noise sequences with zero mean and variances $\sigma_1^2 = 0.20^2$ for $v_1(t)$ and $\sigma_2^2 = 0.30^2$ for $v_2(t)$. Then, we can compute the output vector $\boldsymbol{y}(t) = [y_1(t), y_2(t)]^{\mathrm{T}}$ based on the given input signals, the model and the simulation condition. After obtaining the input and output data, apply the M-GESG algorithm in (45)–(52) and the F-PC-GESG algorithm in (23)–(41) to estimate the parameters of this system. In addition, we add a forgetting factor $\lambda = 0.99$ to the F-PC-GESG algorithm to estimate the parameters. The parameter estimates and errors are shown in Tables 3–5. The parameter estimation errors $\delta := \|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}\| / \|\boldsymbol{\vartheta}\|$ versus t and the parameter estimates versus t are shown in Figure 2–4.
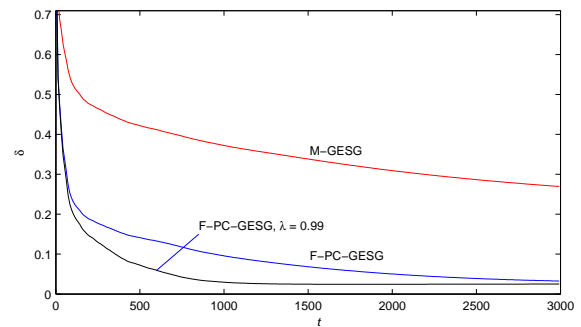
Fig. 2: The parameter estimation errors versus t
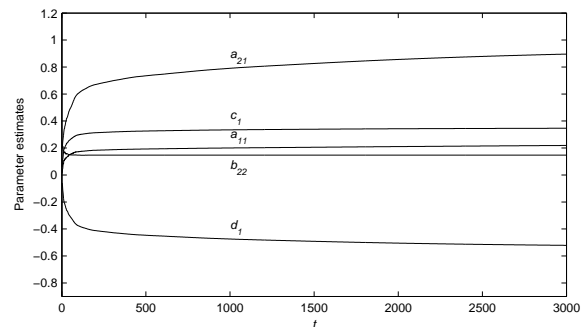
Fig. 3: The M-GESG parameter estimates $\hat{c}_1(t)$, $\hat{d}_1(t)$, $\hat{a}_{11}(t)$, $\hat{a}_{21}(t)$, $\hat{b}_{22}(t)$ versus t
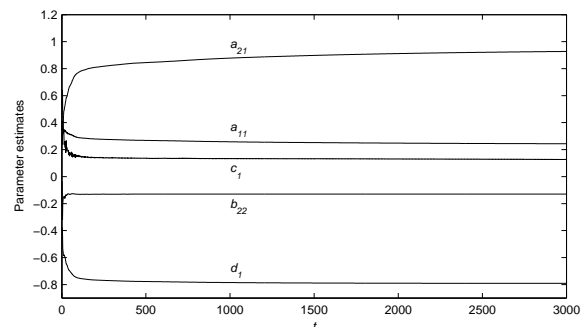
Fig. 4: The F-PC-GESG parameter estimates $\hat{c}_1(t)$, $\hat{d}_1(t)$, $\hat{a}_{11}(t)$, $\hat{a}_{21}(t)$, $\hat{b}_{22}(t)$ versus t

Table 1 The computational efficiency of the M-GESG algorithm

Expressions	Number of multiplications	Number of additions
$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)}{r(t)}\boldsymbol{e}(t) \in \mathbb{R}^{n_0}$	$mn_0 + m$	mn_0
$\boldsymbol{e}(t) := \boldsymbol{y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\vartheta}}(t-1) \in \mathbb{R}^m$	mn_0	mn_0
$r(t) = r(t-1) + \|\hat{\boldsymbol{\Phi}}(t)\|^2 \in \mathbb{R}$	mn_0	mn_0
$\hat{\boldsymbol{\Phi}}(t) = [\hat{\boldsymbol{\psi}}(t), \boldsymbol{\varphi}^{\mathrm{T}}(t) \otimes \boldsymbol{I}_m] \in \mathbb{R}^{m \times n_0}$	n	
$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \boldsymbol{\theta}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t) \in \mathbb{R}^m$	mn	mn
$\hat{\boldsymbol{v}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\vartheta}}(t) \in \mathbb{R}^m$	mn_0	mn_0
Sum	$4mn_0 + mn + m + n$	$4mn_0 + mn$
Total flops	$N_1 = 8mn_0 + 2mn + m + n$	

Table 2 The computational efficiency of the F-PC-GESG algorithm

Expressions	Number of multiplications	Number of additions
$\begin{bmatrix} \hat{\boldsymbol{d}}_i(t) \\ \hat{\boldsymbol{\theta}}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{d}}_{i-1}(t) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} + \frac{\hat{\boldsymbol{\psi}}_i(t)}{r_{1,i}(t)}x_{1,i}(t) \in \mathbb{R}^{n_1}$	$mn_1 + m$	mn_1
$x_i(t) := y_i(t) - \hat{\boldsymbol{\psi}}_i^{\mathrm{T}}(t)\begin{bmatrix} \hat{\boldsymbol{d}}_{i-1}(t) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} \in \mathbb{R}$	mn_1	mn_1
$r_{1,i}(t) = r_{1,i}(t-1) + \|\hat{\boldsymbol{\psi}}_i(t)\|^2 \in \mathbb{R}$	mn_1	mn_1
$\hat{\boldsymbol{c}}(t) = \hat{\boldsymbol{c}}(t-1) + \frac{\hat{\boldsymbol{\Omega}}^{\mathrm{T}}(t)}{r_2(t)}\boldsymbol{x}_2(t) \in \mathbb{R}^{n_c}$	$mn_c + 1$	mn_c
$\boldsymbol{x}_2(t) := \hat{\boldsymbol{w}}_{\mathrm{n}}(t) - \hat{\boldsymbol{\Omega}}(t)\hat{\boldsymbol{c}}(t-1) \in \mathbb{R}^m$	mn_c	mn_c
$r_2(t) = r_2(t-1) + \|\hat{\boldsymbol{\Omega}}(t)\|^2 \in \mathbb{R}$	mn_c	mn_c
$\hat{\boldsymbol{y}}_{\mathrm{f}}(t) = \boldsymbol{y}(t) + [\boldsymbol{y}(t-1), \boldsymbol{y}(t-2), \cdots, \boldsymbol{y}(t-n_c)]\hat{\boldsymbol{c}}(t) \in \mathbb{R}^m$	mn_c	mn_c
$\hat{\boldsymbol{u}}_{\mathrm{f}}(t) = \boldsymbol{u}(t) + [\boldsymbol{u}(t-1), \boldsymbol{u}(t-2), \cdots, \boldsymbol{u}(t-n_c)]\hat{\boldsymbol{c}}(t) \in \mathbb{R}^r$	rn_c	rn_c
$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \boldsymbol{\theta}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t) \in \mathbb{R}^m$	mn	mn
$\hat{\boldsymbol{v}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\phi}}(t)\hat{\boldsymbol{d}}_m(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t) \in \mathbb{R}^m$	$m(n_c + n_d)$	$m(n_c + n_d)$
Sum	$4mn_c + 3mn_1 + rn_c$ $+mn + m + 1$	$4mn_c + 3mn_1 + rn_c$ $+m(n + n_c + n_d)$
Total flops	$N_2 = 9mn_c + 6mn_1 + 2rn_c + m(2n + n_d) + m + 1$	

Table 3 The M-GESG parameter estimates and errors

t	c_1	d_1	a_{11}	a_{12}	a_{21}	a_{22}	b_{11}	b_{12}	b_{21}	b_{22}	$\delta(\%)$
100	0.29814	-0.37722	0.17325	0.60704	-0.44106	0.39523	0.09329	0.19856	0.15585	0.14712	52.61754
200	0.31325	-0.41248	0.18312	0.67058	-0.51004	0.44544	0.09371	0.19839	0.15499	0.14735	47.59884
500	0.32579	-0.44659	0.19271	0.73519	-0.58466	0.51074	0.09380	0.19812	0.15509	0.14702	42.19534
1000	0.33453	-0.47454	0.20027	0.79069	-0.65147	0.58459	0.09380	0.19811	0.15518	0.14697	37.21425
2000	0.34254	-0.50497	0.21045	0.85604	-0.72945	0.70419	0.09378	0.19812	0.15511	0.14698	30.92303
3000	0.34628	-0.52138	0.21874	0.89587	-0.77404	0.80734	0.09378	0.19812	0.15511	0.14698	26.93724
True values	0.14000	-0.80000	0.24000	0.94000	-0.80000	1.05000	0.10000	0.15000	0.12000	-0.10000	

Table 4 The F-PC-GESG parameter estimates and errors

t	c_1	d_1	a_{11}	a_{12}	a_{21}	a_{22}	b_{11}	b_{12}	b_{21}	b_{22}	$\delta(\%)$
100	0.14689	-0.75226	0.28804	0.77230	-0.65336	0.68846	0.08574	0.13932	0.09144	-0.13092	23.45508
200	0.13990	-0.76652	0.27860	0.80983	-0.68485	0.75869	0.08627	0.13935	0.09232	-0.13025	18.75964
500	0.13497	-0.77753	0.26826	0.84469	-0.71807	0.82916	0.08624	0.13931	0.09195	-0.12974	14.12815
1000	0.13303	-0.78535	0.25797	0.87810	-0.74940	0.90104	0.08616	0.13928	0.09178	-0.12966	9.57124
2000	0.13030	-0.79015	0.24796	0.91167	-0.77756	0.97892	0.08617	0.13927	0.09181	-0.12965	5.02683
3000	0.12748	-0.79124	0.24341	0.92736	-0.78984	1.01712	0.08617	0.13927	0.09181	-0.12965	3.23992
True values	0.14000	-0.80000	0.24000	0.94000	-0.80000	1.05000	0.10000	0.15000	0.12000	-0.10000	

Example 2. Consider the following multivariable system:

$$\boldsymbol{A}(z)\boldsymbol{y}(t) = \boldsymbol{B}(z)\boldsymbol{u}(t) + \frac{d(z)}{c(z)}\boldsymbol{v}(t),$$

$$\boldsymbol{A}(z) = \boldsymbol{I}_3 + \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} z^{-1}$$

$$= \begin{bmatrix} -0.55 & 0.45 & -0.15 \\ -0.63 & -0.64 & 0.56 \\ 0.69 & 0.32 & -0.41 \end{bmatrix} z^{-1},$$

$$\boldsymbol{B}(z) = \begin{bmatrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_7 & b_8 & b_9 \end{bmatrix} z^{-1}$$

Here, we take the inputs $\{u_1(t)\}$, $\{u_2(t)\}$ and $\{u_3(t)\}$ as three independent persistent excitation signal sequences with zero mean and unit variances, and $\{v_1(t)\}$, $\{v_2(t)\}$ and $\{v_3(t)\}$ are taken as three white noise sequences with zero mean and variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0.10^2$ for $v_1(t)$, $v_2(t)$ and $v_3(t)$. We use the M-GESG algorithm and the PC-GESG algorithm to estimate the parameters of this multivariable system, respectively. Since there are 20 parameters, we do not give the table for parameter estimates and errors.

Table 5 The F-PC-GESG parameter estimates and errors ($\lambda = 0.99$)

t	c_1	d_1	a_{11}	a_{12}	a_{21}	a_{22}	b_{11}	b_{12}	b_{21}	b_{22}	$\delta(\%)$
100	0.14682	-0.76037	0.28162	0.79617	-0.67552	0.73238	0.08518	0.13966	0.09120	-0.12830	20.46260
200	0.13898	-0.77519	0.26953	0.84149	-0.71302	0.82292	0.08582	0.13969	0.09222	-0.12729	14.56687
500	0.13454	-0.78673	0.25350	0.89464	-0.76236	0.93976	0.08579	0.13949	0.09144	-0.12640	7.24352
1000	0.13318	-0.78989	0.24296	0.92922	-0.79208	1.02175	0.08569	0.13935	0.09123	-0.12619	2.95373
2000	0.12958	-0.79009	0.24026	0.93935	-0.79938	1.04802	0.08569	0.13933	0.09124	-0.12615	2.45025
3000	0.12587	-0.79010	0.24001	0.93995	-0.79996	1.04986	0.08569	0.13933	0.09124	-0.12615	2.50150
True values	0.14000	-0.80000	0.24000	0.94000	-0.80000	1.05000	0.10000	0.15000	0.12000	-0.10000	

Here, we use the parameter estimation errors curve to show the performance of the algorithms. The parameter estimation errors versus t are shown in Figure 5.
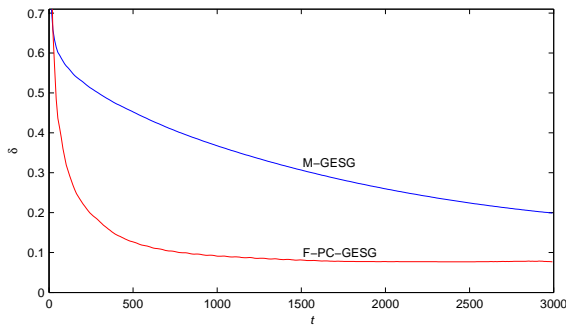
Fig. 5: The parameter estimation errors versus t

From Tables 3–5 and Figures 2–5, we can draw the following conclusions.

1. The parameter estimation errors of the M-GESG and the F-PC-GESG algorithms become smaller with the data length t increasing – see the estimation errors of the last columns in Tables 3–5.

2. Under the same noise level, the PC-GESG algorithm has a higher parameter estimation accuracy than the M-GESG algorithm – see Tables 3–5 and Figures 2–4. This demonstrates that the proposed algorithm is effective.

3. Introducing a forgetting factor can improve the parameter estimation accuracy of the F-PC-GESG algorithm – see Table 5 and Figure 2.

6 Conclusions

In this paper, the parameter estimation problem has been investigated for multivariable CARARMA systems. An F-PC-GESG algorithm is derived by adopting the filtering technique and the coupling concept. In order to reduce the influence of the noise, we construct a filter to filter the input and output data and transform the system into two parts, including a noise model and a filtered system model. In addition, the filtered system model is divided into several subsystems and identified by the coupling concept. According to the computational comparison, the F-PC-GESG algorithm has less computational burden than the M-GESG algorithm. The simulation results indicate that the F-PC-GESG algorithm can generate more accurate parameter estimates. Moreover, introducing the forgetting factor can improve the performance of the F-PC-GESG algorithm. The basic idea of the proposed method in this paper can be used to study the parameter identification problems of other multivariable systems with different structures and disturbance noises.

The identification method presented in this paper can combine the multi-innovation methods [45] and some mathematical skills [46–48] and statistical methods [49–51] can be used to study the performances of parameter estimation algorithms and can be applied other fields [52–55].

8 References

1 Cao, Y., Li, P., Zhang, Y.: 'Parallel processing algorithm for railway signal fault diagnosis data based on cloud computing', *Future Generation Computer Systems*, 2018, **88**, pp. 279-283

2 Zhou, Z.P., Liu, X.F.: 'State and fault estimation of sandwich systems with hysteresis', *International Journal of Robust and Nonlinear Control*, 2018, **28**, (13), pp. 3974-3986

3 Cao, Y., Ma, L.C., Xiao, S., *et al.*: 'Standard analysis for transfer delay in CTCS-3', *Chinese Journal of Electronics*, 2017, **26**, (5), pp. 1057-1063

4 Zhang, Y.Z., Cao, Y., Wen, Y.H., *et al.*: 'Optimization of information interaction protocols in cooperative vehicle-infrastructure systems', *Chinese Journal of Electronics*, 2018, **27**, (2), pp. 439-444

5 Cao, Y., Ma, L.C., Xiao, S., *et al.*: 'Standard analysis for transfer delay in CTCS-3', *Chinese Journal of Electronics*, 2017, **26**, (5), pp. 1057-1063

6 Wang, Y., Zhang, H., Wei, S., *et al.*: 'Control performance assessment for ILC-controlled batch processes in two-dimensional system framework', *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, **48**, (9), pp. 1493-1504

7 Zhang, W.H., Xue, L., Jiang, X.: 'Global stabilization for a class of stochastic nonlinear systems with SISS-like conditions and time delay', *International Journal of Robust and Nonlinear Control*, 2018, **28**, (13), pp. 3909-3926

8 Xu, L.: 'The parameter estimation algorithms based on the dynamical response measurement data', *Advances in Mechanical Engineering*, 2017, **9**, (11), pp. 1-12, doi: 10.1177/1687814017730003

9 Xu, L., Ding, F.: 'Iterative parameter estimation for signal models based on measured data', *Circuits Systems and Signal Processing*, 2018, **37**, (7), pp. 3046-3069

10 Dong, S.J., Liu, T., Wang, Q.G.: 'Identification of dual-rate sampled systems with time delay subject to load disturbance', *IET Control Theory and Applications*, 2017, **11**, (9), pp. 1404-1413

11 Na, J., Yang, J., Wu, X., *et al.*: 'Robust adaptive parameter estimation of sinusoidal signals', *Automatica*, 2015, **53**, pp. 376-384

12 Xu, L., Ding, F.: 'Parameter estimation for control systems based on impulse responses', *International Journal of Control, Automation, and Systems*, 2017, **15**, (6), pp. 2471-2479

13 Bottegal, G., Aravkin, A.Y., Hjalmarsson, H., Pillonetto, G.: 'Robust EM kernel-based methods for linear system identification', *Automatica*, 2016, **67**, pp. 114-126

14 Gan, M., Chen, C.L.P., Chen, G.Y., Chen, L: 'On some separated algorithms for separable nonlinear squares problems', *IEEE Transactions on Cybernetics*, 2018, **48**, (10), pp. 2866-2874

15 Na, J., Yang, J., Ren, X.M., *et al.*: 'Robust adaptive estimation of nonlinear system with time-varying parameters', *International Journal of Adaptive Control and Signal Processing*, 2015, **29**, (8), pp. 1055-1072

16 Mattsson, P., Zachariah, D., Stoica, P.: 'Recursive nonlinear-system identification using latent variables', *Automatica*, 2018, **93**, pp. 343-351

17 Li, J.H., Zheng, W.X., Gu, J.P., Hua, L.: 'A recursive identification algorithm for Wiener nonlinear systems with linear state-space subsystem', *Circuits Systems and Signal Processing*, 2018, **37**, (6), pp. 2374-2393

18 Doraiswami, R., Cheded, L.: 'Robust Kalman filter-based least squares identification of a multivariable system', *IET Control Theory and Applications*, 2018, **12**, (8), pp. 1064-1074

19 Gan, M., Li, H.X., Peng, H.: 'A variable projection approach for efficient estimation of RBF-ARX model', *IEEE Transactions on Cybernetics*, 2015, **45**, (3) pp. 462-471

20 Yu, C.P., Verhaegen, M.: 'Blind multivariable ARMA subspace identification', *Automatica*, 2016, **66**, pp. 3-14

21 Cham, C.L., Tan, A.H., Tan, W.H.: 'Identification of a multivariable nonlinear and time-varying mist reactor system', *Control Engineering Practice*, 2017, **63**, pp. 13-23

22 Ding, S., Dang, Y.G., Li, X.M., *et al.*: 'Forecasting Chinese CO_2 emissions from fuel combustion using a novel grey multivariable model', *Journal of Cleaner Production*, 2017, **162**, pp. 1527-1538

23 Mobayen, S.: 'Robust tracking controller for multivariable delayed systems with input saturation via composite nonlinear feedback', *Nonlinear Dynamics*, 2014, **76**, (1), pp. 827-838

24   Zhang, E.L., Pintelon, R.: 'Identification of multivariable dynamic errors-in-variables system with arbitrary inputs', *Automatica*, 2017, **82**, pp. 69-78

25   Ding, J.L.: 'Recursive and iterative least squares parameter estimation algorithms for multiple-input-output-error systems with autoregressive noise', *Circuits Systems and Signal Processing*, 2018, **37**, (5), pp. 1884-1906

26   Li, M.H., Liu, X.M.: 'The least squares based iterative algorithms for parameter estimation of a bilinear system with autoregressive noise using the data filtering technique', *Signal Processing*, 2018, **147**, pp. 23-34

27   Pan, J., Ma, H., Jiang, X., *et al.*: 'Adaptive gradient-based iterative algorithm for multivariate controlled autoregressive moving average systems using the data filtering technique', *Complexity*, 2018, Article ID 9598307. https://doi.org/10.1155/2018/9598307

28   Afshari, H.H., Gadsden, S.A., Habibi, S.: 'Gaussian filters for parameter and state estimation: A general review of theory and recent trends', *Signal Processing*, 2017, **135**, pp. 218-238

29   Pan, J., Jiang, X., Wan, X.K., Ding, W.F.: 'A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems', *International Journal of Control, Automation, and Systems*, 2017, **15**, (3), pp. 1189-1197

30   Xu, L., Xiong, W.L., Alsaedi, A., Hayat, T.: 'Hierarchical parameter estimation for the frequency response based on the dynamical window data', *International Journal of Control, Automation and Systems*, 2018, **16**, (4), pp. 1756-1764

31   Ding, J.L.: 'The hierarchical iterative identification algorithm for multi-input-output-error systems with autoregressive noise', *Complexity*, 2017, pp. 1-11. Article ID 5292894. https://doi.org/10.1155/2017/5292894

32   Ding, F.: 'Coupled-least-squares identification for multivariable systems', *IET Control Theory and Applications*, 2013, **7**, (1), pp. 68-79

33   Ding, F., Liu, G., Liu, X.P.: 'Partially coupled stochastic gradient identification methods for non-uniformly sampled systems', *IEEE Transactions on Automatic Control*, 2010, **55**, (8), pp. 1976-1981

34   Wang, X.H., Ding, F.: 'Partially coupled extended stochastic gradient algorithm for nonlinear multivariable output error moving average systems', *Engineering Computations*, 2017, **34**, (2), pp. 629-647

35   Liu, Q.Y., Ding, F., Yang, E.F.: 'Parameter estimation algorithm for multivariable controlled autoregressive autoregressive moving average systems', *Digital Signal Processing*, 2018, **83**, pp. 323-331

36   Zhang, X., Xu, L., Ding, F., *et al.*: 'Combined state and parameter estimation for a bilinear state space system with moving average noise', *J. Franklin Inst.*, 2018, **355**, (6), pp. 3079-3103.

37   Zhang, X., Ding, F., Xu, L., *et al.*: 'State filtering-based least squares parameter estimation for bilinear systems using the hierarchical identification principle', *IET Control Theory Appl.*, 2018, **12**, (12), pp. 1704-1713.

38   Zhang, X., Ding, F., Alsaadi, F.E., Hayat, T.: 'Recursive parameter identification of the dynamical models for bilinear state space systems', *Nonlinear Dyn.*, 2017, **89**, (4), pp. 2415-2429

39   Li, X., Zhu, D.Q. : 'An improved SOM neural network method to adaptive leader-follower formation control of AUVs', *IEEE Transactions on Industrial Electronics*, 2018, **65**, (10), pp. 8260-8270

40   Geng, F.Z., Qian, S.P.: 'An optimal reproducing kernel method for linear nonlocal boundary value problems', *Appl. Math. Lett.*, 2018, **77**, pp. 49-56.

41   Li, X.Y., Wu, B.Y.: 'A new reproducing kernel collocation method for nonlocal fractional boundary value problems with non-smooth solutions', *Appl. Math. Lett.*, 2018, **86**, pp. 194-199. December śżÒ ýčž0

42   Ding, F., Xu, L., Alsaadi, F.E., *et al.*: 'Iterative parameter identification for pseudo-linear systems with ARMA noise using the filtering technique', *IET Control Theory Appl.*, 2018, **12**, (7), pp. 892-899.

43   Ding, F., Chen, H.B., Xu, L., *et al.*: Dai, J.Y., Li, Q.S., Hayat, T.: 'A hierarchical least squares identification algorithm for Hammerstein nonlinear systems using the key term separation', *J. Franklin Inst.*, 2018, **355**, (8), pp. 3737-3752.

44   Wang, Y.J., Ding, F., Xu, L.: 'Some new results of designing an IIR filter with colored noise for signal processing', *Digit. Signal Process.*, 2018, **72**, pp. 44-58

45   Ding, F.: 'Several multi-innovation identification methods', *Digital Signal Processing*, **20**, (4), pp. (2010) 1027-1039.

46   Liu, F., Wu, H.X.: 'Singular integrals related to homogeneous mappings in triebel-lizorkin spaces', *Journal of Mathematical Inequalities*, 2017, **11**, (4), pp. 1075-1097 Cited 12

47   Liu, F., Wu, H.X.: 'Regularity of discrete multisublinear fractional maximal functions', *Science China – Mathematics*, 2017, **60**, (8), pp. 1461-1476 Cited 14

48   Liu, F., Wu, H.X.: 'On the regularity of maximal operators supported by submanifolds', *Journal of Mathematical Analysis and Applications*, 2017, **453**, (1), pp. 144-158 Cited 13

49   Yin, C.C., Wen, Y.Z.: 'Exit problems for jump processes with applications to dividend problems', *J. Comput. Appl. Math.*, 2013, **245**, pp. 30-52

50   Yin, C.C., Wen, Y.Z.: 'Optimal dividend problem with a terminal value for spectrally positive Levy processes', *Insurance Mathematics & Economics*, 2013, **53**, (3), pp. 769-773

51   Yin, C.C., Wang, C.W.: 'The perturbed compound Poisson risk process with investment and debit interest', *Methodology and Computing in Applied Probability*, 2010, **12**, (3), pp. 391-413

52   Gong, P.C., Wang, W.Q., Li, F.C., Cheung, H.: 'Sparsity-aware transmit beamspace design for FDA-MIMO radar', *Signal Process.*, 2018, **144**, pp. 99-103.

53   Rao, Z.H., Zeng, C.Y., Wu, M.H. Wu, et al.: 'Research on a handwritten character recognition algorithm based on an extended nonlinear kernel residual network', *KSII Transactions on Internet and Information Systems*, 2018, **12**, (1), pp. 413-435.

54   Zhao, N., Liu, R., Chen, Y., Wu, M., et al.: 'Contract design for relay incentive mechanism under dual asymmetric information in cooperative networks', *Wireless Networks*, 2018, **24**, (8), pp. 3029-3044.

55   Pan, J., Li, W., Zhang, H.P.: 'Control algorithms of magnetic suspension systems based on the improved double exponential reaching law of sliding mode control', *International Journal of Control, Automation and Systems*, 2018.

http://dx.doi.org/10.1007/s12555-017-0616-y