Parameter estimation algorithm for multivariable controlled autoregressive autoregressive moving average systems

Qinyao Liu[a], Feng Ding[a,b,*], Erfu Yang[c]

[a]*Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, PR China*
[b]*College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, PR China*
[c]*Space Mechatronic Systems Technology Laboratory, University of Strathclyde, Glasgow G1 1XJ, Scotland, United Kingdom*

Abstract

This paper investigates parameter estimation problems for multivariable controlled autoregressive autoregressive moving average (M-CARARMA) systems. In order to improve the performance of the standard multivariable generalized extended stochastic gradient (M-GESG) algorithm, we derive a partially coupled generalized extended stochastic gradient algorithm by using the auxiliary model. In particular, we divide the identification model into several subsystems based on the hierarchical identification principle and estimate the parameters using the coupled relationship between these subsystems. The simulation results show that the new algorithm can give more accurate parameter estimates of the M-CARARMA system than those of the M-GESG algorithm.

Keywords: Coupling identification, Parameter estimation, Stochastic gradient, Auxiliary model, Multivariable system

1. Introduction

Parameter estimation has wide application in many areas such as controller designs [1, 2] and stochastic systems [3, 4], signal processing [5, 6, 7, 8] and other practical projects [9, 10]. Exploring valid methods is the eternal theme of the parameter estimation [11, 12, 13, 14]. Recently, Torres et al. proposed an approach based on state observers to identify the parameters of an unknown periodic force exerted on a mechanical system [15]. The approach can be described as two stages, the one is to obtain the estimates of the coefficients of a Fourier series that approximates the periodic force, and the other one is to get the frequencies of the signal. By eliminating the linear parameters through the orthogonal projection, Gan et al. presented a variable projection algorithm for the radial basis function network-based autoregressive with exogenous inputs model [16].

In recent years, multivariable systems, i.e., MIMO systems, have garnered widespread attentions [17, 18], resulting in a rich collection of studies [19, 20] and further interests in system identification for MIMO systems [21, 22, 23]. For MIMO systems with unknown inner variables, an auxiliary model based algorithm was studied by means of the iterative search principle [24]. Like the least squares approach [25, 26, 27], the stochastic gradient (SG) has been widely used in the parameter estimation [28, 29]. Recently, Cheng et al. studied the identification problem for Hammerstein nonlinear ARMAX systems and proposed a multi-innovation fractional order stochastic gradient algorithm [30]. In addition, they introduced a forgetting factor on the step size and a variable gradient order for the purpose of improving the convergence performance. Since the SG algorithm is a kind of recursive algorithm, it will produce the data saturation with the time length increasing [31, 32]. Parameter estimation methods can be applied to many areas [33, 34, 35, 36], including bilinear systems [37] and bilinear-parameter systems [38].

This paper studies the parameter estimation problems for multivariable systems. As we all know, multivariable systems are high-dimensional and have many parameter matrices to be dealt with [39]. Generally, we divide a multivariable system into several subsystems according to the number of the outputs by using the hierarchical

*Corresponding author
Email addresses:* `qyliu12@126.com` (Qinyao Liu), `fding@jiangnan.edu.cn` (Feng Ding), `erfu.yang@strath.ac.uk` (Erfu Yang)

identification principle [40, 41]. However, each subsystem contains a same parameter vector after the decomposition. In order to cut down the redundant computation, the coupling identification concept is employed to estimate the parameters [42]. With regard to the unknown variables, we establish the auxiliary models and replace the unknown variables with the outputs of the auxiliary models [43, 44]. The main contributions of this paper lie in the following aspects.

- This paper decomposes a multivariable controlled autoregressive autoregressive moving average (M-CARARMA) system into several subsystems by means of the hierarchical identification principle.

- A partially coupled generalized extended stochastic gradient (PC-GESG) algorithm is derived for M-CARARMA systems by using the coupling concept and the auxiliary model.

- The proposed PC-GESG algorithm has a higher computational efficiency than the multivariable generalized extended stochastic gradient (M-GESG) algorithm.

- The parameter estimates given by the PC-GESG algorithm are more accurate than those given by the M-GESG algorithm under the same noise level. In addition, the forgetting factor partially generalized extended stochastic gradient (FF-PC-GESG) algorithm can achieve better performance than the PC-GESG algorithm.

The rest of this paper is organized as follows. In Section 2, we give some definitions and the identification model for the M-CARARMA system. Section 3 proposes the M-GESG algorithm for comparison. Section 4 divides the system into several subsystems and presents the PC-GESG algorithm by employing the coupling concept. Section 5 shows the results of numerical simulations to assess the performance of the proposed methods. Finally, we offer some concluding remarks in Section 6.

2. Problem formulation

Let us introduce some symbols. "$A =: X$" 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 sizes $m \times m$; $\mathbf{1}_m$ stands for an m-dimensional column vector whose elements are 1; $\mathbf{1}_{m \times n}$ 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 the vector formed by all columns of matrix $\boldsymbol{X}$ arranged in order, for example, $\boldsymbol{X} := [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n] \in \mathbb{R}^{m \times n}$, $\boldsymbol{x}_i \in \mathbb{R}^m$ $(i = 1, 2, \cdots, n)$, $\mathrm{col}[\boldsymbol{X}] := [\boldsymbol{x}_1^\mathrm{T}, \boldsymbol{x}_2^\mathrm{T}, \cdots, \boldsymbol{x}_n^\mathrm{T}]^\mathrm{T} \in \mathbb{R}^{mn}$; $\hat{\boldsymbol{\theta}}(t)$ denotes the estimate of $\boldsymbol{\theta}$ at time t; the norm of a matrix (or a column vector) $\boldsymbol{X}$ is defined by $\|\boldsymbol{X}\|^2 := \mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^\mathrm{T}]$.

A multiple-input multiple-output (MIMO) system with the colored noise is given by

$$\boldsymbol{A}(z)\boldsymbol{y}(t) = \boldsymbol{B}(z)\boldsymbol{u}(t) + \boldsymbol{w}(t), \tag{1}$$

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, $\boldsymbol{A}(z)$ and $\boldsymbol{B}(z)$ are matrix polynomials in the unit backward shift operator z^{-1} $[z^{-1}\boldsymbol{u}(t) = \boldsymbol{u}(t-1)]$, and

$$\boldsymbol{A}(z) := \boldsymbol{I}_m + \boldsymbol{A}_1 z^{-1} + \boldsymbol{A}_2 z^{-2} + \cdots + \boldsymbol{A}_{n_a} z^{-n_a}, \ \boldsymbol{A}_i \in \mathbb{R}^{m \times m},$$
$$\boldsymbol{B}(z) := \boldsymbol{B}_1 z^{-1} + \boldsymbol{B}_2 z^{-2} + \cdots + \boldsymbol{B}_{n_b} z^{-n_b}, \ \boldsymbol{B}_i \in \mathbb{R}^{m \times r},$$

$\boldsymbol{w}(t) := [w_1(t), w_2(t), \cdots, w_m(t)]^\mathrm{T} \in \mathbb{R}^m$ is taken as an autoregressive moving average (ARMA) process of a white noise vector $\boldsymbol{v}(t) := [v_1(t), v_2(t), \cdots, v_m(t)]^\mathrm{T} \in \mathbb{R}^m$.

When the $\boldsymbol{w}(t)$ is an ARMA process, there are four cases available for the description of the noise term.

1) The first form is $\boldsymbol{w}(t) = \boldsymbol{C}^{-1}(z)\boldsymbol{D}(z)\boldsymbol{v}(t)$, where $\boldsymbol{C}(z)$ and $\boldsymbol{D}(z)$ are matrix polynomials in z^{-1}, and they are defined as

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

2) The second form is

$$\boldsymbol{w}(t) = \frac{d(z)}{c(z)} \boldsymbol{v}(t), \qquad (2)$$

where $c(z)$ and $d(z)$ are scalar polynomials in z^{-1}, and they are expressed as

$$c(z) := 1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}, \ c_i \in \mathbb{R},$$
$$d(z) := 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots + d_{n_d} z^{-n_d}, \ d_i \in \mathbb{R}.$$

3) The third form is $\boldsymbol{w}(t) = d(z)\boldsymbol{C}^{-1}(z)\boldsymbol{v}(t)$, where $d(z)$ is a scalar polynomial and $\boldsymbol{C}(z)$ is a matrix polynomial.

4) The last form is $\boldsymbol{w}(t) = \frac{\boldsymbol{D}(z)}{c(z)}\boldsymbol{v}(t)$, where $\boldsymbol{D}(z)$ is a matrix polynomial while $c(z)$ is a scalar polynomial. Substituting (2) into (1) gives

$$\boldsymbol{A}(z)\boldsymbol{y}(t) = \boldsymbol{B}(z)\boldsymbol{u}(t) + \frac{d(z)}{c(z)}\boldsymbol{v}(t). \qquad (3)$$

Without loss of generality, assume that 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 $t \leqslant 0$.

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

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

Notice that $\boldsymbol{w}(t)$ can be expressed as different forms,

$$\begin{aligned}
\boldsymbol{w}(t) &= \frac{d(z)}{c(z)}\boldsymbol{v}(t) \\
&= \boldsymbol{A}(z)\boldsymbol{y}(t) - \boldsymbol{B}(z)\boldsymbol{u}(t) \\
&= \boldsymbol{y}(t) - \boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\varphi}(t) \qquad (4) \\
&= \boldsymbol{\phi}(t)\boldsymbol{\beta} + \boldsymbol{v}(t). \qquad (5)
\end{aligned}$$

Substituting (5) into (3) and applying the definitions, we can obtain the hierarchical identification model

$$\begin{aligned}
\boldsymbol{y}(t) &= [\boldsymbol{I}_m - \boldsymbol{A}(z)]\boldsymbol{y}(t) + \boldsymbol{B}(z)\boldsymbol{u}(t) + \boldsymbol{w}(t) \\
&= \boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\varphi}(t) + \boldsymbol{w}(t) \\
&= \boldsymbol{\phi}(t)\boldsymbol{\beta} + \boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\varphi}(t) + \boldsymbol{v}(t). \qquad (6)
\end{aligned}$$

Observing (6), we can see that there is not only a parameter vector $\boldsymbol{\beta}$ to be identified, and also a parameter matrix $\boldsymbol{\rho}$ to be identified, which makes the model structure complex. Moreover, the input $\boldsymbol{u}(t)$ and output $\boldsymbol{y}(t)$ are available, that is, only $\boldsymbol{y}(t)$ and $\boldsymbol{\varphi}(t)$ in (6) are known. The objective of this paper is to find a way to handle the unknown variables and to present a highly efficient algorithm for the M-CARARMA system in (3) by using the auxiliary model and the coupling concept.

3. The multivariable generalized extended stochastic gradient algorithm

In order to provide a comparison, we drive the basic multivariable generalized extended stochastic gradient (M-GESG) algorithm for the M-CARARMA system in (3).

For convenience, combine the information vector $\boldsymbol{\varphi}(t)$ with the information matrix $\boldsymbol{\phi}(t)$ to construct an information matrix $\boldsymbol{\Phi}(t)$ by means of the Kronecker product:

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

3

The parameter matrix $\boldsymbol{\rho}$ and the parameter vector $\boldsymbol{\beta}$ can be constituted as a high-dimensional parameter vector:

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

Then rewrite (6) as the pseudo-linear regressive model

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}(t)\boldsymbol{\vartheta} + \boldsymbol{v}(t). \tag{7}$$

The new parameter vector $\boldsymbol{\vartheta}$ contains all parameters of the M-CARARMA system in (3). Applying the stochastic gradient method to the identification model in (7), we can derive the M-GESG algorithm.

In consideration of the unknown variables $\boldsymbol{w}(t)$ and $\boldsymbol{v}(t)$, we employed the auxiliary model method to solve this problem. Let $\hat{\boldsymbol{w}}(t)$ and $\hat{\boldsymbol{v}}(t)$ be the outputs of the auxiliary models, and define the estimate of $\boldsymbol{\phi}(t)$:

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

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

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

According to (4) and (7), replacing $\boldsymbol{\Phi}(t)$, $\boldsymbol{\rho}$ and $\boldsymbol{\vartheta}$ with their estimates $\hat{\boldsymbol{\Phi}}(t)$, $\hat{\boldsymbol{\rho}}(t)$ and $\hat{\boldsymbol{\vartheta}}(t)$, the outputs $\hat{\boldsymbol{w}}(t)$ and $\hat{\boldsymbol{v}}(t)$ of the auxiliary models can be computed by

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

Based on the auxiliary model method, replacing $\boldsymbol{\Phi}(t)$ in (7) with its estimate $\hat{\boldsymbol{\Phi}}(t)$ and using the negative gradient search, we can obtain the following M-GESG algorithm:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)}{r(t)}[\boldsymbol{y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\vartheta}}(t-1)], \tag{8}$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\Phi}}(t)\|^2, \tag{9}$$

$$\hat{\boldsymbol{\Phi}}(t) = [\hat{\boldsymbol{\phi}}(t), \boldsymbol{\varphi}^{\mathrm{T}}(t) \otimes \boldsymbol{I}_m], \tag{10}$$

$$\boldsymbol{\varphi}(t) = [-\boldsymbol{y}^{\mathrm{T}}(t-1), -\boldsymbol{y}^{\mathrm{T}}(t-2), \cdots, -\boldsymbol{y}^{\mathrm{T}}(t-n_a), \boldsymbol{u}^{\mathrm{T}}(t-1), \boldsymbol{u}^{\mathrm{T}}(t-2), \cdots, \boldsymbol{u}^{\mathrm{T}}(t-n_b)]^{\mathrm{T}}, \tag{11}$$

$$\hat{\boldsymbol{\phi}}(t) = [-\hat{\boldsymbol{w}}(t-1), -\hat{\boldsymbol{w}}(t-2), \cdots, -\hat{\boldsymbol{w}}(t-n_c), \hat{\boldsymbol{v}}(t-1), \hat{\boldsymbol{v}}(t-2), \cdots, \hat{\boldsymbol{v}}(t-n_d)], \tag{12}$$

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\rho}}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t), \tag{13}$$

$$\hat{\boldsymbol{v}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\vartheta}}(t), \tag{14}$$

$$\hat{\boldsymbol{\vartheta}}(t) = \begin{bmatrix} \hat{\boldsymbol{\beta}}(t) \\ \mathrm{col}[\hat{\boldsymbol{\rho}}^{\mathrm{T}}(t)] \end{bmatrix}. \tag{15}$$

The procedure involved in the M-GESG algorithm in (8)–(15) is listed as follows.

1. Set the initial values: let $t = 1$, $\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_{n_0}/p_0$, $r(0) = 1$, $\hat{\boldsymbol{v}}(t-j) = \mathbf{1}_m/p_0$, $\hat{\boldsymbol{w}}(t-j) = \mathbf{1}_m/p_0$, $i = 1, 2$, $\cdots$, $\max[n_c, n_d]$, $p_0 = 10^6$ and set a small positive number ϵ.

2. Collect the observation data $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$, and construct the information vector and matrices $\boldsymbol{\varphi}(t)$, $\boldsymbol{\phi}(t)$ and $\hat{\boldsymbol{\Phi}}(t)$ using (11)–(12) and (10).

3. Compute $r(t)$ using (9) and update the parameter estimation vector $\hat{\boldsymbol{\vartheta}}(t)$ by (8).

4. Read $\hat{\boldsymbol{\rho}}(t)$ and $\hat{\boldsymbol{\beta}}(t)$ from $\hat{\boldsymbol{\vartheta}}(t)$ by (15) and compute $\hat{\boldsymbol{w}}(t)$ and $\hat{\boldsymbol{v}}(t)$ by (13)–(14).

5. Compare $\hat{\boldsymbol{\vartheta}}(t)$ with $\hat{\boldsymbol{\vartheta}}(t-1)$: if $\|\hat{\boldsymbol{\vartheta}}(t) - \hat{\boldsymbol{\vartheta}}(t-1)\| > \epsilon$, increase t by 1 and go to Step 2; otherwise, terminate recursive calculation procedure and obtain $\hat{\boldsymbol{\vartheta}}(t)$.

Remark 1: The M-GESG algorithm in (8)–(15) is an extension of the scalar stochastic gradient algorithm. By means of the auxiliary model identification idea, the unknown information matrix $\boldsymbol{\Phi}(t)$ is replaced by its estimate $\hat{\boldsymbol{\Phi}}(t)$ in order to guarantee the realization of the algorithm.

Remark 2: For improving the performance of the M-GESG algorithm, a forgetting factor (FF) λ can be introduced in (9):

$$r(t) = \lambda r(t-1) + \|\hat{\boldsymbol{\Phi}}(t)\|^2, \quad 0 \le \lambda < 1. \tag{16}$$

Another way is to modify (8) as

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)}{r^\varepsilon(t)}[\boldsymbol{y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\vartheta}}(t-1)], \quad \frac{1}{2} < \varepsilon \le 1, \tag{17}$$

where ε is the convergence index. Introducing the forgetting factor and the convergence index in (16)–(17) can improve the convergence rate and the parameter estimation accuracy of the M-GESG algorithm.

Remark 3: Although the M-GESG algorithm in (8)–(15) can produce the parameter estimation vector $\hat{\boldsymbol{\vartheta}}(t)$, the weakness is that $\hat{\boldsymbol{\Phi}}(t)$ is a high-dimensional informational matrix ($n_0 \times n_0$, $n_0 = n_c + n_d + mn$), which gives rise to a heavy computational burden. This motivates us to study some new identification algorithms to decrease the computational cost. For ease of comparison, the computational efficiency of the M-GESG algorithm at each recursive step is shown in Table 1, where flops represent the floating point operations.

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
$\hat{\boldsymbol{\Phi}}(t) = [\hat{\boldsymbol{\phi}}(t), \boldsymbol{\varphi}^{\mathrm{T}}(t) \otimes \boldsymbol{I}_m] \in \mathbb{R}^{m \times n_0}$	n	
$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \boldsymbol{\rho}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t) \in \mathbb{R}^m$	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	$N_1 = 8mn_0 + 2mn + m + n$	

4. The partially coupled generalized extended stochastic gradient algorithm

In this section, a partially coupled generalized extended stochastic gradient (PC-GESG) algorithm is derived by employing the decomposition technique and the coupling concept. The basic idea is to divide the multivariable identification model in (6) into m subsystems, and to identify each subsystem based on the coupled relations in the part of parameters between subsystems.

Rewrite the hierarchical identification model in (6) for the M-CARARMA system in (3) as

$$\boldsymbol{y}(t) = \boldsymbol{\phi}(t)\boldsymbol{\beta} + \boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\varphi}(t) + \boldsymbol{v}(t). \tag{18}$$

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

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

Similarly, let $\boldsymbol{\rho}_i(t) \in \mathbb{R}^n$ be the ith column of the parameter matrix $\boldsymbol{\rho}$, that is

$$\boldsymbol{\rho} := [\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \cdots, \boldsymbol{\rho}_m] \in \mathbb{R}^{n \times m}.$$

By using the above definitions, Equation (18) can be rewritten as

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}_1^{\mathrm{T}}(t) \\ \boldsymbol{\phi}_2^{\mathrm{T}}(t) \\ \vdots \\ \boldsymbol{\phi}_m^{\mathrm{T}}(t) \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\rho}_1^{\mathrm{T}} \\ \boldsymbol{\rho}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\rho}_m^{\mathrm{T}} \end{bmatrix} \boldsymbol{\varphi}(t) + \begin{bmatrix} v_1(t) \\ v_2(t) \\ \vdots \\ v_m(t) \end{bmatrix}. \tag{19}$$

Then we can decompose (19) into m subsystem identification models:

$$
\begin{aligned}
y_i(t) &= \boldsymbol{\phi}_i^{\mathrm{T}}(t)\boldsymbol{\beta} + \boldsymbol{\rho}_i^{\mathrm{T}}\boldsymbol{\varphi}(t) + v_i(t) \\
&= \boldsymbol{\phi}_i^{\mathrm{T}}(t)\boldsymbol{\beta} + \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\rho}_i + v_i(t) \\
&= [\boldsymbol{\phi}_i^{\mathrm{T}}(t), \boldsymbol{\varphi}^{\mathrm{T}}(t)] \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\rho}_i \end{bmatrix} + v_i(t), \quad i = 1, 2, \cdots, m.
\end{aligned} \tag{20}
$$

Define the subsystem information vector

$$
\boldsymbol{\psi}_i(t) := \begin{bmatrix} \boldsymbol{\phi}_i(t) \\ \boldsymbol{\varphi}(t) \end{bmatrix} \in \mathbb{R}^{n_1}, \quad n_1 := n + n_c + n_d.
$$

Equation (20) can be expressed as

$$
y_i(t) = \boldsymbol{\psi}_i^{\mathrm{T}}(t) \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\rho}_i \end{bmatrix} + v_i(t), \ i = 1, 2, \cdots, m. \tag{21}
$$

This is the subsystem identification model. From (21), we can see that each subsystem identification model contains the same parameter vector $\boldsymbol{\beta}$ and the same information vector $\boldsymbol{\varphi}(t)$.

Based on the subsystem identification models in (21), using the negative gradient search gives

$$
\begin{bmatrix} \hat{\boldsymbol{\beta}}(t) \\ \hat{\boldsymbol{\rho}}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}(t-1) \\ \hat{\boldsymbol{\rho}}_i(t-1) \end{bmatrix} + \frac{\boldsymbol{\psi}_i(t)}{r_i(t)} \left\{ y_i(t) - \boldsymbol{\psi}_i^{\mathrm{T}}(t) \begin{bmatrix} \hat{\boldsymbol{\beta}}(t-1) \\ \hat{\boldsymbol{\rho}}_i(t-1) \end{bmatrix} \right\}, \tag{22}
$$

$$
r_i(t) = r_i(t-1) + \|\boldsymbol{\psi}_i(t)\|^2, \ r_i(0) = 1, \quad i = 1, 2, \cdots, m. \tag{23}
$$

Here, some problems arise. The first problem is that the subsystem information vector $\boldsymbol{\psi}_i(t)$, $i = 1, 2, \cdots, m$, contains the unknown variables $\boldsymbol{w}(t-j)$ and $\boldsymbol{v}(t-j)$. Similarly, we use the auxiliary models' outputs $\hat{\boldsymbol{w}}(t-j)$ and $\hat{\boldsymbol{v}}(t-j)$ to replace $\boldsymbol{w}(t-j)$ and $\boldsymbol{v}(t-j)$ and construct the estimate of $\boldsymbol{\phi}(t)$:

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

Define the estimate of $\boldsymbol{\psi}_i(t)$ by using $\hat{\boldsymbol{\phi}}_i(t)$ and $\boldsymbol{\varphi}(t)$:

$$
\hat{\boldsymbol{\psi}}_i(t) := \begin{bmatrix} \hat{\boldsymbol{\phi}}_i(t) \\ \boldsymbol{\varphi}(t) \end{bmatrix} \in \mathbb{R}^{n_1 \times 1}.
$$

Based on (4)–(5), the estimates $\hat{\boldsymbol{w}}(t)$ and $\hat{\boldsymbol{v}}(t)$ can be calculated by

$$
\begin{aligned}
\hat{\boldsymbol{w}}(t) &= \boldsymbol{y}(t) - \hat{\boldsymbol{\rho}}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t), \\
\hat{\boldsymbol{v}}(t) &= \boldsymbol{w}(t) - \hat{\boldsymbol{\phi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}(t) \\
&= \boldsymbol{y}(t) - \hat{\boldsymbol{\rho}}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t) - \hat{\boldsymbol{\phi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}(t).
\end{aligned}
$$

The second problem is that each subsystem has the same parameter vector $\boldsymbol{\beta}$. That is to say, each subsystem can obtain the estimate $\hat{\boldsymbol{\beta}}(t)$ of $\boldsymbol{\beta}$. In order to make it clear, $\hat{\boldsymbol{\beta}}_i(t)$ represents the estimate of Subsystem i. Replacing the unknown information vector $\boldsymbol{\psi}_i(t)$ with its estimate $\hat{\boldsymbol{\psi}}_i(t)$ and the common parameter estimation vector $\hat{\boldsymbol{\beta}}(t)$ with $\hat{\boldsymbol{\beta}}_i(t)$ in (22)–(23), we have

$$
\begin{bmatrix} \hat{\boldsymbol{\beta}}_i(t) \\ \hat{\boldsymbol{\rho}}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_i(t-1) \\ \hat{\boldsymbol{\rho}}_i(t-1) \end{bmatrix} + \frac{\hat{\boldsymbol{\psi}}_i(t)}{r_i(t)} \left\{ y_i(t) - \hat{\boldsymbol{\psi}}_i^{\mathrm{T}}(t) \begin{bmatrix} \hat{\boldsymbol{\beta}}_i(t-1) \\ \hat{\boldsymbol{\rho}}_i(t-1) \end{bmatrix} \right\}, \tag{24}
$$

$$
r_i(t) = r_i(t-1) + \|\hat{\boldsymbol{\psi}}_i(t)\|^2. \tag{25}
$$

As we can see, there are estimates $\hat{\boldsymbol{\beta}}_1(t)$, $\hat{\boldsymbol{\beta}}_2(t)$, $\cdots$, $\hat{\boldsymbol{\beta}}_m(t)$ for $i = 1, 2, \cdots, m$. In fact, we only need one parameter estimate of $\boldsymbol{\beta}$. In general, the parameter estimates approach their true values with the time t increasing. One may consider that the parameter estimate $\hat{\boldsymbol{\beta}}_{i-1}(t)$ is closer to the true parameter than the parameter estimate $\hat{\boldsymbol{\beta}}_i(t-1)$. Based on the coupling identification concept, for $i = 1$, use $\hat{\boldsymbol{\beta}}_m(t-1)$ to replace $\hat{\boldsymbol{\beta}}_1(t-1)$, for $i = 2, 3, \cdots, m$, use $\hat{\boldsymbol{\beta}}_{i-1}(t)$ to replace $\hat{\boldsymbol{\beta}}_i(t-1)$. Through the substitutions, we can obtain the PC-GESG algorithm:

$$
\begin{bmatrix} \hat{\boldsymbol{\beta}}_i(t) \\ \hat{\boldsymbol{\rho}}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{i-1}(t) \\ \hat{\boldsymbol{\rho}}_i(t-1) \end{bmatrix} + \frac{\hat{\boldsymbol{\psi}}_i(t)}{r_i(t)} \left\{ y_i(t) - \hat{\boldsymbol{\psi}}_i^{\mathrm{T}}(t) \begin{bmatrix} \hat{\boldsymbol{\beta}}_{i-1}(t) \\ \hat{\boldsymbol{\rho}}_i(t-1) \end{bmatrix} \right\}, \tag{26}
$$

$$r_i(t) = r_i(t-1) + \|\hat{\boldsymbol{\psi}}_i(t)\|^2, \ i = 2, 3, \cdots, m, \tag{27}$$

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_1(t) \\ \hat{\boldsymbol{\rho}}_1(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_m(t-1) \\ \hat{\boldsymbol{\rho}}_1(t-1) \end{bmatrix} + \frac{\hat{\boldsymbol{\psi}}_1(t)}{r_1(t)} \left\{ y_1(t) - \hat{\boldsymbol{\psi}}_1^{\mathrm{T}}(t) \begin{bmatrix} \hat{\boldsymbol{\beta}}_m(t-1) \\ \hat{\boldsymbol{\rho}}_1(t-1) \end{bmatrix} \right\}, \tag{28}$$

$$r_1(t) = r_1(t-1) + \|\hat{\boldsymbol{\psi}}_1(t)\|^2, \tag{29}$$

$$\hat{\boldsymbol{\psi}}_i(t) = [\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(t), \boldsymbol{\varphi}^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{30}$$

$$\boldsymbol{\varphi}(t) = [-\boldsymbol{y}^{\mathrm{T}}(t-1), -\boldsymbol{y}^{\mathrm{T}}(t-2), \cdots, -\boldsymbol{y}^{\mathrm{T}}(t-n_a), \boldsymbol{u}^{\mathrm{T}}(t-1), \boldsymbol{u}^{\mathrm{T}}(t-2), \cdots, \boldsymbol{u}^{\mathrm{T}}(t-n_b)]^{\mathrm{T}}, \tag{31}$$

$$\hat{\boldsymbol{\phi}}(t) = [-\hat{\boldsymbol{w}}(t-1), -\hat{\boldsymbol{w}}(t-2), \cdots, -\hat{\boldsymbol{w}}(t-n_c), \hat{\boldsymbol{v}}(t-1), \hat{\boldsymbol{v}}(t-2), \cdots, \hat{\boldsymbol{v}}(t-n_d)] \tag{32}$$

$$= [\hat{\boldsymbol{\phi}}_1(t), \hat{\boldsymbol{\phi}}_2(t), \cdots, \hat{\boldsymbol{\phi}}_m(t)]^{\mathrm{T}}, \tag{33}$$

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\rho}}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t), \tag{34}$$

$$\hat{\boldsymbol{v}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\phi}}(t)\hat{\boldsymbol{\beta}}_m(t) - \hat{\boldsymbol{\rho}}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t), \tag{35}$$

$$\hat{\boldsymbol{\rho}}(t) = [\hat{\boldsymbol{\rho}}_1(t), \hat{\boldsymbol{\rho}}_2(t), \cdots, \hat{\boldsymbol{\rho}}_m(t)]. \tag{36}$$

The PC-GESG estimates are $\hat{\boldsymbol{\beta}}(t) = \hat{\boldsymbol{\beta}}_m(t)$ and $\hat{\boldsymbol{\rho}}(t)=[\hat{\boldsymbol{\rho}}_1(t), \hat{\boldsymbol{\rho}}_2(t), \cdots, \hat{\boldsymbol{\rho}}_m(t)]$. So $\hat{\boldsymbol{\beta}}(t)$ in the calculation of $\hat{\boldsymbol{v}}(t)$ should be modified as $\hat{\boldsymbol{\beta}}_m(t)$.

The steps involved in the PC-GESG algorithm in (26)–(36) are listed as follows.

1. Set the initial values: let $t = 1$, $\hat{\boldsymbol{\beta}}_m(0) = \mathbf{1}_{n_c+n_d}/p_0$, $\hat{\boldsymbol{\rho}}_{i,0} = \mathbf{1}_n/p_0$, $r_i(0) = 1$ $(i = 1, 2, \cdots, m)$, $\hat{\boldsymbol{w}}(0) = \mathbf{1}_m/p_0$, $\hat{\boldsymbol{v}}(0) = \mathbf{1}_m/p_0$, $p_0 = 10^6$ and set a small positive number ϵ.
2. Collect the observation data $\boldsymbol{y}(t)$ and $\boldsymbol{u}(t)$, and construct the information matrix $\boldsymbol{\varphi}(t)$ using (31).
3. Form $\hat{\boldsymbol{\phi}}(t)$ using (32), read $\hat{\boldsymbol{\phi}}_i(t)$ from (33), then construct $\hat{\boldsymbol{\psi}}_i(t)$ by (30).
4. Compute $r_1(t)$ using (29) and update the parameter estimates $\hat{\boldsymbol{\beta}}_1(t)$ and $\hat{\boldsymbol{\rho}}_1(t)$ using (28).
5. For $i = 2, 3, \cdots, m$, compute $r_i(t)$ using (27) and update $\hat{\boldsymbol{\beta}}_i(t)$ and $\hat{\boldsymbol{\rho}}_i(t)$ using (26).
6. Form $\hat{\boldsymbol{\rho}}(t)$ by (36), compute $\hat{\boldsymbol{w}}(t)$ and $\hat{\boldsymbol{v}}(t)$ using (34)–(35).
7. Compare $\hat{\boldsymbol{\rho}}(t)$ with $\hat{\boldsymbol{\rho}}(t-1)$ and compare $\hat{\boldsymbol{\beta}}_m(t)$ with $\hat{\boldsymbol{\beta}}_m(t-1)$: if $\|\hat{\boldsymbol{\vartheta}}(t)-\hat{\boldsymbol{\vartheta}}(t-1)\| > \epsilon$ and $\|\hat{\boldsymbol{\vartheta}}(t)-\hat{\boldsymbol{\vartheta}}(t-1)\| > \epsilon$, increase t by 1 and go to Step 2; otherwise, terminate recursive calculation procedure and obtain $\hat{\boldsymbol{\rho}}(t)$ and $\hat{\boldsymbol{\beta}}(t) = \hat{\boldsymbol{\beta}}_m(t)$.

The schematic diagram of the PC-GESG algorithm in (26)–(36) is shown in Figure 1.
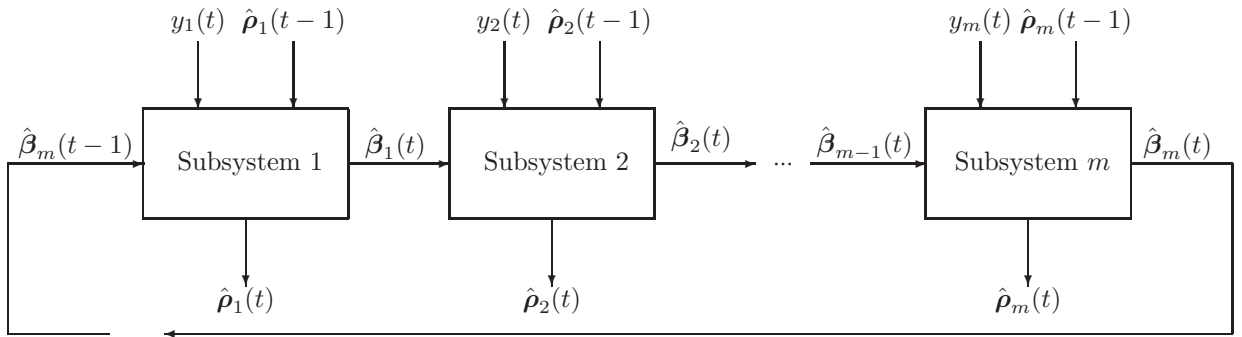

Figure 1: The schematic diagram of the PC-GESG algorithm

From Figure 1, we can see that only part parameters are coupled, that is to say, only $\hat{\boldsymbol{\beta}}_i(t)$ are coupled between the subsystems, the parameter vectors $\hat{\boldsymbol{\rho}}_i(t)$ are independent because each subsystem has a corresponding $\hat{\boldsymbol{\rho}}_i(t)$. This is the reason that we call this algorithm the partially coupled GESG algorithm.

Remark 4: Similarly, it is also reasonable to replace $r_1(t-1)$ with $r_m(t-1)$ and replace $r_i(t-1)$ $(i = 2, 3, \cdots, m)$ with $r_{i-1}(t)$. Then Equations (27) and (29) can be modified as

$$r_i(t) = r_{i-1}(t) + \|\hat{\boldsymbol{\psi}}_i(t)\|^2, \ i = 2, 3, \cdots, m, \tag{37}$$

$$r_1(t) = r_m(t-1) + \|\hat{\boldsymbol{\psi}}_1(t)\|^2. \tag{38}$$

Then (26), (40), (28), (42) and (30)–(36) form another more complex PC-GESG algorithm. Here, the parameter $\boldsymbol{\beta}$ and the variable $r(t)$ are coupled in the algorithm.

Remark 5: In order to improve the parameter estimation accuracy of the PC-GESG algorithm in (26)–(36), we can introduce the forgetting factor λ in (27) and (29), or add the convergence index ε in (26) and (28):

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_i(t) \\ \hat{\boldsymbol{\rho}}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{i-1}(t) \\ \hat{\boldsymbol{\rho}}_i(t-1) \end{bmatrix} + \frac{\hat{\boldsymbol{\psi}}_i(t)}{r_i^\varepsilon(t)} \left\{ y_i(t) - \hat{\boldsymbol{\psi}}_i^{\mathrm{T}}(t) \begin{bmatrix} \hat{\boldsymbol{\beta}}_{i-1}(t) \\ \hat{\boldsymbol{\rho}}_i(t-1) \end{bmatrix} \right\}, \quad \frac{1}{2} < \varepsilon \leqslant 1, \tag{39}$$

$$r_i(t) = \lambda r_i(t-1) + \|\hat{\boldsymbol{\psi}}_i(t)\|^2, \quad 0 \leq \lambda < 1, \ i = 2, 3, \cdots, m, \tag{40}$$

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_1(t) \\ \hat{\boldsymbol{\rho}}_1(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_m(t-1) \\ \hat{\boldsymbol{\rho}}_1(t-1) \end{bmatrix} + \frac{\hat{\boldsymbol{\psi}}_1(t)}{r_1^\varepsilon(t)} \left\{ y_1(t) - \hat{\boldsymbol{\psi}}_1^{\mathrm{T}}(t) \begin{bmatrix} \hat{\boldsymbol{\beta}}_m(t-1) \\ \hat{\boldsymbol{\rho}}_1(t-1) \end{bmatrix} \right\}, \tag{41}$$

$$r_1(t) = \lambda r_1(t-1) + \|\hat{\boldsymbol{\psi}}_1(t)\|^2. \tag{42}$$

Equations (26), (40), (28), (42) and (30)–(36) form the forgetting factor partially generalized extended stochastic gradient (FF-PC-GESG) algorithm. Combining (39)–(42) with (30)–(36), we can obtain the modified FF-PC-GESG (M-FF-PC-GESG) algorithm. When $\lambda = 1$ and $\varepsilon = 1$, the M-FF-PC-GESG algorithm reduces to the PC-GESG algorithm in (26)–(36).

Remark 6: The PC-GESG algorithm in (26)–(36) divides the MIMO system in (3) into m subsystems, and identifies the parameters based on the coupled relations between the subsystems. The computational efficiency of the PC-GESG algorithm is shown in Table 2.

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

Expressions	Number of multiplications	Number of additions
$\begin{bmatrix} \hat{\boldsymbol{\beta}}_i(t) \\ \hat{\boldsymbol{\rho}}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{i-1}(t) \\ \hat{\boldsymbol{\rho}}_i(t-1) \end{bmatrix} + \frac{\hat{\boldsymbol{\psi}}_i(t)}{r_i(t)} x_i(t) \in \mathbb{R}^{n_1}$	$mn_1 + m$	mn_1
$x_i(t) := y_i(t) - \hat{\boldsymbol{\psi}}_i^{\mathrm{T}}(t) \begin{bmatrix} \hat{\boldsymbol{\beta}}_{i-1}(t) \\ \hat{\boldsymbol{\rho}}_i(t-1) \end{bmatrix} \in \mathbb{R}$	mn_1	mn_1
$r_i(t) = r_i(t-1) + \|\hat{\boldsymbol{\psi}}_i(t)\|^2 \in \mathbb{R}$	mn_1	mn_1
$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \boldsymbol{\rho}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t) \in \mathbb{R}^m$	mn	mn
$\hat{\boldsymbol{v}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\phi}}(t)\hat{\boldsymbol{\beta}}_m(t) - \hat{\boldsymbol{\rho}}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t) \in \mathbb{R}^m$	$m(n_c + n_d)$	$m(n_c + n_d)$
Sum	$4mn_1 + m$	$4mn_1$
Total flops	$N_2 = 8mn_1 + m$	

In the M-GESG algorithm in (8)–(15), the dimension of the information matrix $\hat{\boldsymbol{\Phi}}(t)$ is $m \times n_0$ ($n_0 = n_c + n_d + mn$). In the PC-GESG algorithm in (26)–(36), the dimension of the information matrix $\hat{\boldsymbol{\psi}}_i(t)$ is $n_1 \times 1$ ($n_1 = n_c + n_d + n$). It is apparent from Tables 1–2 that the computational burden of the PC-GESG algorithm is less than the M-GESG algorithm, that is to say, $N_1 \gg N_2$. Compared with the M-GESG algorithm, the PC-GESG algorithm avoids many redundant parameter estimates and improves the computational efficiency.

Remark 7: This paper focuses on the parameter identification problems for multivariable CARARMA systems, which can be seen as an expansion to multivariable case of the CARARMA model in [26]. The multivariable CARMA-like model in [20] and [21] is another type of MIMO systems, which has the scalar polynomial in front of the output vector $\boldsymbol{y}(t)$. The PC-GESG algorithm proposed in this paper decompose the multivariable CARARMA system into m subsystems with the single output and identify each subsystem using the coupled relations between these subsystems. The work in [25] and [21] also used the hierarchical identification principle, but they divided the system into two parts and applied the least squares based iterative method to deal with the two parts. As for the estimation procedure, for every i, we get the estimate $\hat{\boldsymbol{\rho}}_i(t)$. In order to obtain the $\hat{\boldsymbol{\beta}}_i(t)$, we use the estimate $\hat{\boldsymbol{\beta}}_{i-1}(t)$ of Subsystem $i-1$ to replace the estimate $\hat{\boldsymbol{\beta}}_i(t-1)$ of Subsystem i. Finally, we can obtain $\hat{\boldsymbol{\rho}}(t) = [\hat{\boldsymbol{\rho}}_1(t), \hat{\boldsymbol{\rho}}_2(t), \cdots, \hat{\boldsymbol{\rho}}_m(t)]$ and $\hat{\boldsymbol{\beta}}(t) = \hat{\boldsymbol{\beta}}_m(t)$.

5. Examples

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

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

$$\boldsymbol{A}(z) = \boldsymbol{I}_2 + \left[\begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array}\right] z^{-1} = \left[\begin{array}{cc} 1 + 0.25z^{-1} & 0.40z^{-1} \\ -0.43z^{-1} & 1 - 0.76z^{-1} \end{array}\right],$$

$$\boldsymbol{B}(z) = \left[\begin{array}{cc} b_{11} & b_{12} \\ b_{21} & b_{22} \end{array}\right] z^{-1} = \left[\begin{array}{cc} 0.50z^{-1} & 0.45z^{-1} \\ 0.80z^{-1} & 0.42z^{-1} \end{array}\right],$$

$$c(z) = 1 + c_1 z^{-1} = 1 - 0.17z^{-1},$$

$$d(z) = 1 + d_1 z^{-1} = 1 + 0.12z^{-1},$$

$$\boldsymbol{\rho}^{\mathrm{T}} = \left[\begin{array}{cccc} 0.25 & 0.40 & 0.50 & 0.45 \\ -0.43 & -0.76 & 0.80 & 0.42 \end{array}\right],$$

$$\boldsymbol{\beta} = [c_1, d_1]^{\mathrm{T}} = [-0.17, 0.12]^{\mathrm{T}},$$

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

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 $\sigma_1^2 = 0.20^2$, $\sigma_2^2 = 0.30^2$, respectively, we use them to generate the output vector $\boldsymbol{y}(t) = [y_1(t), y_2(t)]^{\mathrm{T}}$. The parameter estimation error is in direct proportional to the number of the parameters and the noise variance, and is in inverse proportion to the data length. Here, we choose the data length $t = 3000$ to estimate the parameters. To apply the M-GESG algorithm in (8)–(15), the PC-GESG algorithm in (26)–(36) and the FF-PC-GESG algorithm to estimate the parameters of this system, all the initial values of the parameters are set to be zero. 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 are shown in Figures 2–5.

Table 3: The M-GESG 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	-0.06867	0.06610	0.05685	0.19915	-0.29108	-0.31399	0.30361	0.23460	0.37417	0.17876	52.51553
200	-0.08817	0.08414	0.06589	0.22001	-0.29961	-0.35736	0.31480	0.24555	0.39796	0.18911	48.68301
500	-0.10174	0.09510	0.08670	0.24658	-0.30398	-0.40209	0.33002	0.26416	0.42648	0.20727	44.14468
1000	-0.11188	0.10385	0.09795	0.26501	-0.30945	-0.43253	0.33982	0.27402	0.44760	0.22024	41.06893
2000	-0.12179	0.11170	0.10835	0.27978	-0.31164	-0.46424	0.34754	0.28500	0.46732	0.23182	38.18475
3000	-0.12704	0.11609	0.11397	0.28862	-0.31442	-0.48180	0.35245	0.29060	0.47988	0.23845	36.50561
True values	-0.17000	0.12000	0.25000	0.40000	-0.43000	-0.76000	0.50000	0.45000	0.80000	0.42000	

Table 4: The PC-GESG 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	-0.12069	0.11019	0.12141	0.31966	-0.39648	-0.49910	0.46218	0.38248	0.52012	0.29075	29.19004
200	-0.13543	0.12283	0.13629	0.33369	-0.39808	-0.54744	0.46702	0.38959	0.55218	0.29628	25.22640
500	-0.14584	0.13040	0.16156	0.35161	-0.39697	-0.59295	0.47234	0.40259	0.58890	0.31626	20.67429
1000	-0.15145	0.13462	0.17440	0.36258	-0.39834	-0.61841	0.47483	0.40752	0.61324	0.32896	17.98801
2000	-0.15714	0.13883	0.18394	0.37035	-0.39634	-0.64536	0.47677	0.41339	0.63350	0.33989	15.62146
3000	-0.15941	0.14063	0.18898	0.37435	-0.39722	-0.65767	0.47811	0.41597	0.64644	0.34621	14.32294
True values	-0.17000	0.12000	0.25000	0.40000	-0.43000	-0.76000	0.50000	0.45000	0.80000	0.42000	

Table 5: The FF-PC-GESG 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	-0.13296	0.12033	0.13689	0.33229	-0.39721	-0.53841	0.47208	0.38619	0.54370	0.30444	25.79253
200	-0.15048	0.13510	0.16087	0.35064	-0.39951	-0.60660	0.47642	0.39826	0.59605	0.31322	19.97721
500	-0.16221	0.14222	0.21880	0.38224	-0.40065	-0.68726	0.48550	0.42655	0.68732	0.35863	10.47748
1000	-0.16548	0.14256	0.24410	0.40267	-0.41147	-0.72488	0.48764	0.43626	0.75133	0.39656	4.85150
2000	-0.17035	0.14634	0.24776	0.39674	-0.41453	-0.75299	0.49516	0.44617	0.78449	0.41986	2.35778
3000	-0.16843	0.14501	0.24977	0.39864	-0.42984	-0.76080	0.50236	0.44720	0.80790	0.42432	1.77632
True values	-0.17000	0.12000	0.25000	0.40000	-0.43000	-0.76000	0.50000	0.45000	0.80000	0.42000	

Example 2. Consider a following multivariable system:

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

$$\boldsymbol{A}(z) = \boldsymbol{I}_2 + \left[\begin{array}{cc} a_1 & a_2 \\ a_3 & a_4 \end{array}\right] z^{-1} + \left[\begin{array}{cc} a_5 & a_6 \\ a_7 & a_8 \end{array}\right] z^{-2}$$

$$= \left[\begin{array}{cc} 1 + 0.52z^{-1} - 0.43z^{-2} & 0.45z^{-1} + 0.29z^{-2} \\ -0.32z^{-1} - 0.24z^{-2} & 1 - 0.52z^{-1} + 0.48z^{-2} \end{array}\right],$$
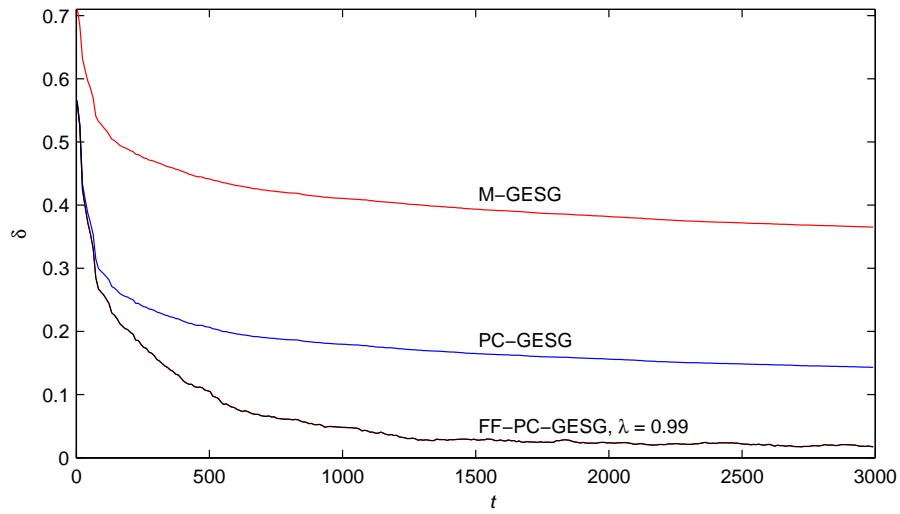
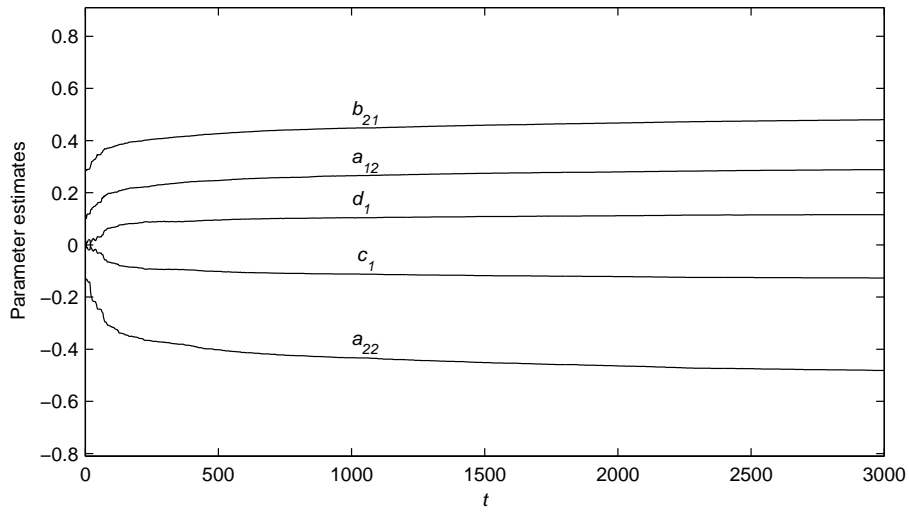Figure 2: The parameter estimation errors versus t

Figure 3: The M-GESG estimates $\hat{c}_1(t)$, $\hat{d}_1(t)$, $\hat{a}_{12}(t)$, $\hat{a}_{22}(t)$, $\hat{b}_{21}(t)$ versus t
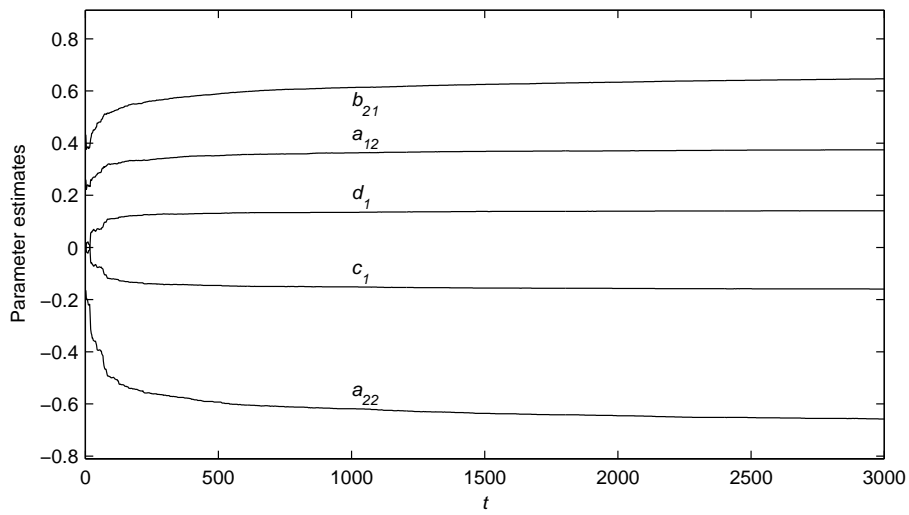

Figure 4: The PC-GESG estimates $\hat{c}_1(t)$, $\hat{d}_1(t)$, $\hat{a}_{12}(t)$, $\hat{a}_{22}(t)$, $\hat{b}_{21}(t)$ versus t
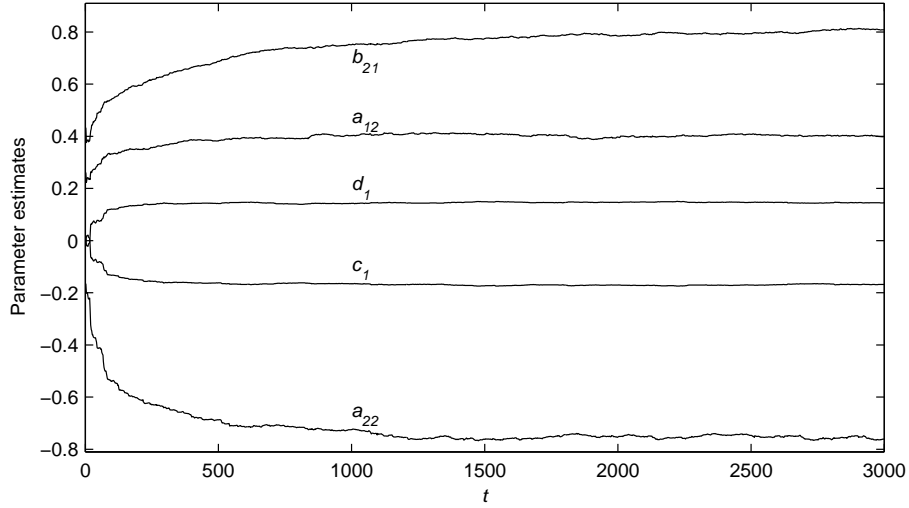
Figure 5: The FF-PC-GESG estimates $\hat{c}_1(t)$, $\hat{d}_1(t)$, $\hat{a}_{12}(t)$, $\hat{a}_{22}(t)$, $\hat{b}_{21}(t)$ versus t

$$\boldsymbol{B}(z) = \left[\begin{array}{cc} b_1 & b_2 \\ b_3 & b_4 \end{array} \right] z^{-1} + \left[\begin{array}{cc} b_5 & b_6 \\ b_7 & b_8 \end{array} \right] z^{-2} = \left[\begin{array}{cc} -0.10z^{-1} - 0.24z^{-2} & 0.26z^{-1} - 0.25z^{-2} \\ 0.67z^{-1} + 0.43z^{-2} & 0.26z^{-1} + 0.51z^{-2} \end{array} \right],$$

$$c(z) = 1 + c_1 z^{-1} = 1 + 0.22z^{-1},$$

$$d(z) = 1 + d_1 z^{-1} = 1 - 0.17z^{-1},$$

$$\boldsymbol{\rho}^{\mathrm{T}} = \left[\begin{array}{cccccccc} 0.52 & 0.45 & -0.43 & 0.29 & -0.10 & 0.26 & -0.24 & -0.25 \\ -0.32 & -0.52 & -0.24 & 0.48 & 0.67 & 0.26 & 0.43 & 0.51 \end{array} \right],$$

$$\boldsymbol{\beta} = [c_1, d_1]^{\mathrm{T}} = [0.22, -0.17]^{\mathrm{T}},$$

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

Take the same simulation conditions as those in Example 1. Similarly, use the M-GESG algorithm, the PC-GESG algorithm and the FF-PC-GESG algorithm to estimate the parameters of this multivariable system, respectively. Since there are 18 parameters, we do not give the table for parameter estimates and errors. Here, we use the parameter estimation error curves to show the performance of the algorithms. The parameter estimation errors versus t are shown in Figure 6.
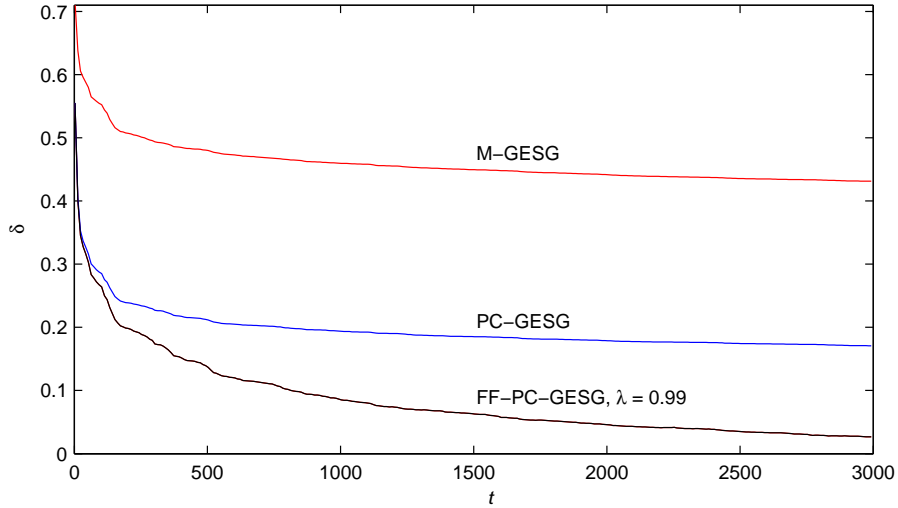
Figure 6: The parameter estimation errors versus t

Example 3. The paper-making process can be viewed as a multi-input and multi-output stochastic system and can be modeled by a multivariable controlled autoregressive system. A simple sketch of the paper-making process is given in Figure 7. The thick paper fiber materials with a certain concentration from the pulp workshop

11

are pumped into a flow box to dilute them with water. After discharging from the flow box, the next step is to drain water from the materials by the braid fabric. Then the paper is formed by the pressing machine. The paper sheet goes through the dryer section to remove the remaining water by steam heating. Finally, workers check the moisture and basis weight of the manufactured paper and obtain the qualified products. By regulating the amount of the thick paper fiber materials and the steam pressure in the dryer section, the moisture and basis weight of the production can be controlled. In Figure 7, $u_1(t)$ and $u_2(t)$ are the amount of the materials and the steam pressure, $y_1(t)$ and $y_2(t)$ are the moisture and the basis weight.
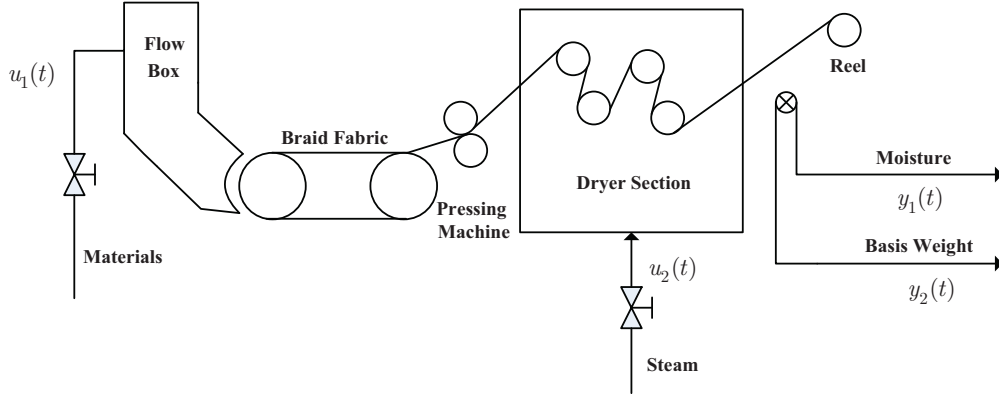
Figure 7: A simple sketch of the paper-making process

Assume that the system from $\boldsymbol{u}(t)$ to $\boldsymbol{y}(t)$ be modeled by the following multivariable model:

$$\boldsymbol{y}(t) + \begin{bmatrix} -0.62 & 0.45 \\ -0.58 & -0.32 \end{bmatrix} z^{-1}\boldsymbol{y}(t) + \begin{bmatrix} -0.33 & 0.29 \\ -0.24 & 0.48 \end{bmatrix} z^{-2}\boldsymbol{y}(t) =$$
$$\begin{bmatrix} -0.86 & 0.16 \\ 0.68 & -0.44 \end{bmatrix} z^{-1}\boldsymbol{u}(t) + \begin{bmatrix} -0.84 & -0.65 \\ 0.55 & 0.61 \end{bmatrix} z^{-2}\boldsymbol{u}(t) + \boldsymbol{w}(t),$$

where $\boldsymbol{w}(t)$ is taken as an autoregressive moving average process of the white noise vector $\boldsymbol{v}(t)$:

$$\boldsymbol{w}(t) = \frac{1 + 0.15z^{-1} - 0.12z^{-2}}{1 - 0.12z^{-1} + 0.11z^{-2}} \boldsymbol{v}(t).$$

The simulation conditions and steps are similar to those in Example 1. Take $\lambda = 0.9$ and the data set length $t = 3000$. Apply the M-GESG, PC-GESG and FF-PC-GESG algorithms to estimate the parameters of this example system. Their estimation errors δ versus t are shown in Figure 8.
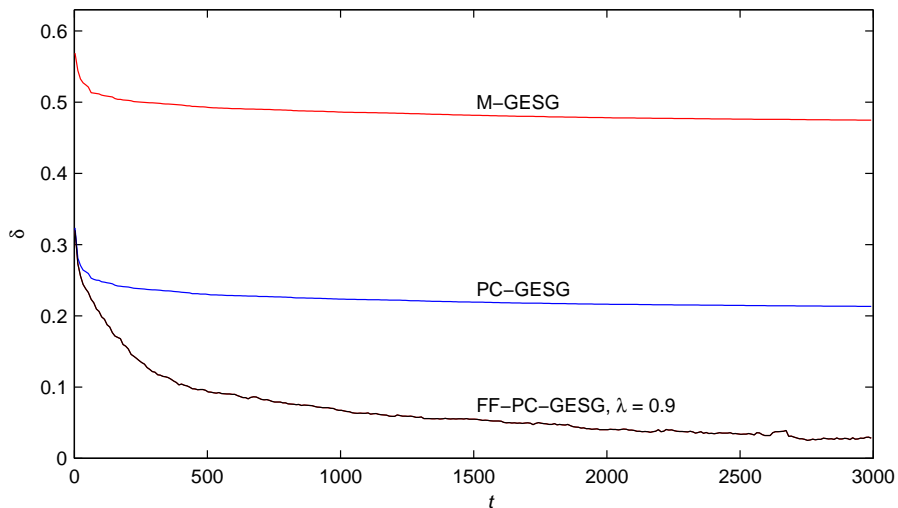

Figure 8: The parameter estimation errors versus t

From Tables 3–5 and Figures 2–8, we can draw the following conclusions.

1. The parameter estimation errors of the M-GESG, the PC-GESG and the FF-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 levels, the PC-GESG algorithm can give more accurate parameter estimates than the M-GESG algorithm – see Tables 3–4 and Figures 2–6 and 8.

3. The FF-PC-GESG algorithm can improve the parameter estimation accuracy – see Table 5 and Figures 2, 6 and 8.

6. Conclusions

In this paper, we have presented a recursive identification algorithm of multivariable systems. This identification algorithm has been developed for multivariable systems with autoregressive moving average noise. Differing from the M-GESG algorithm, the PC-GESG algorithm decomposes the original system into several subsystems according to the number of the outputs. The basic idea is to use the coupling concept to reduce the redundant parameter estimates of the coupled parameter vectors between subsystems. So the PC-GESG algorithm avoids the redundant computation and can improve the calculation efficiency. Through the numerical simulations, we can conclude that the PC-GESG algorithm has more accurate parameter estimates compared with the M-GESG algorithm. In addition, the FF-PC-GESG algorithm has smaller parameter estimation errors than the PC-GESG algorithm.

The differences between the previous identification algorithms and the PC-GESG algorithm are discussed. The presented algorithm is based on the decomposition technique and aimed to solve the coupled items between subsystems. The proposed coupling based identification method can be applied to other multivariable systems with different structures and disturbance noises. Moreover, the idea of this algorithm can be adopted when the concerned system model has the coupled terms. Applying the proposed algorithm to a real project will be further investigated in our future work. The proposed approaches in the paper can combine other mathematical tools [45, 46, 47] and statistical strategies [48, 49, 50, 51] to study the performances of some parameter estimation algorithms and can be applied to other multivariable systems with different structures and disturbance noises and other literature [52, 53, 54].

References

[1] Y. Wang, D. Zhao, Y. Li, S.X. Ding, Unbiased minimum variance fault and state estimation for linear discrete time-varying two-dimensional systems, IEEE Transactions on Automatic Control 62 (10) (2017) 5463-5469.

[2] Y. Wang, H. Zhang, S. Wei, D. Zhou, B. Huang, Control performance assessment for ILC-controlled batch processes in two-dimensional system framework, IEEE Transactions on Systems, Man and Cybernetics: Systems 48 (2018). doi: 10.1109/TSMC.2017.2672563

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

[4] Y. Li, W.H. Zhang, X.K. Liu, H-index for discrete-time stochastic systems with Markovian jump and multiplicative noise, Automatica 90 (2018) 286-293.

[5] S. Wu, W. Cui, Q. Shen, et al., Efficient parameter estimation method for maneuvering targets in discrete randomly-modulated radar, Digital Signal Processing 67 (2017) 91-106.

[6] L. Xu, F. Ding, Iterative parameter estimation for signal models based on measured data, Circuits, Systems and Signal Processing 37 (7) (2018) 3046-3069.

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

[8] Y.J. Wang, F. Ding, L. Xu, Some new results of designing an IIR filter with colored noise for signal processing, Digital Signal Processing 72 (2018) 44-58.

[9] M.M. Mansouri, H.N. Nounou, M.N. Nounou, A.A. Datta, State and parameter estimation for nonlinear biological phenomena modeled by S-systems, Digital Signal Processing 28 (2014) 1-17.

[10] J. Na, G. Herrmann, K.Q. Zhang, Improving transient performance of adaptive control via a modified reference model and novel adaptation, International Journal of Robust and Nonlinear Control 27 (8) (2017) 1351-1372.

[11] M. Gan, C.L.P. Chen, G.Y. Chen, L. Chen, On some separated algorithms for separable nonlinear squares problems, IEEE Transactions on Cybernetics (2018). doi: 10.1109/TCYB.2017.2751558

[12] J. Na, J. Yang, X. Wu, et al, Robust adaptive parameter estimation of sinusoidal signals, Automatica 53 (2015) 376-384.

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

[14] L. Xu, F. Ding. Parameter estimation for control systems based on impulse responses, International Journal of Control, Automation, and Systems 15 (6) (2017) 2471-2479.

[15] L. Torres, J.F. Gómez-Aguilar, J. Jiménez, et al., Parameter identification of periodical signals: Application to measurement and analysis of ocean wave forces, Digital Signal Processing 69 (2017) 59-69.

[16] M. Gan, H.X. Li, H. Peng, A variable projection approach for efficient estimation of RBF-ARX model, IEEE Transactions on Cybernetics 45 (3) (2015) 462-471.

[17] X.R. Jing, A. Li, H.Q. Liu, A low-complexity Lanczos-algorithm-based detector with soft-output for multiuser massive MIMO systems, Digital Signal Processing 69 (2017) 41-49.

[18] J. Pan, H. Ma, X. Jiang, 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

[19] C.W.R. Chiong, Y. Rong, Y. Xiang, Blind channel estimation and signal retrieving for MIMO relay systems, Digital Signal Processing 52 (2016) 35-44.

[20] D.Q. Wang, Y.P. Gao, Recursive maximum likelihood identification method for a multivariable controlled autoregressive moving average system, IMA Journal of Mathematical Control and Information 33 (4) (2016) 1015-1031.

[21] H.Q. Han, L. Xie, F. Ding, X.G. Liu, Hierarchical least squares based iterative identification for multivariable systems with moving average noises, Mathematical and Computer Modelling 51 (9-10) (2010) 1213-1220.

[22] P. Ma, F. Ding, Q.M. Zhu, Decomposition-based recursive least squares identification methods for multivariate pseudolinear systems using the multi-innovation, International Journal of Systems Science 49 (5) (2018) 920-928.

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

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

[25] G.Y. Yao, R. Ding, Two-stage least squares based iterative identification algorithm for controlled autoregressive moving average (CARMA) systems, Computers & Mathematics with Applications 63 (5) (2012) 975-984.

[26] W. Wang, F. Ding, J.Y. Dai, Maximum likelihood least squares identification for systems with autoregressive moving average noise, Applied Mathematical Modelling 36 (5) (2012) 1842-1853.

[27] F. Ding, Y.J. Wang, J.Y. Dai, Q.S. Li, Q.J. Chen, A recursive least squares parameter estimation algorithm for output nonlinear autoregressive systems using the input-output data filtering, Journal of the Franklin Institute 354 (15) (2017) 6938-6955.

[28] X. Zhang, L. Xu, F. Ding, T. Hayat, Combined state and parameter estimation for a bilinear state space system with moving average noise, Journal of the Franklin Institute 355 (6) (2018) 3079-3103.

[29] L. Xu, W.L. Xiong, A. Alsaedi, T. Hayat, Hierarchical parameter estimation for the frequency response based on the dynamical window data, International Journal of Control, Automation and Systems 16 (4) (2018) 1756-1764.

[30] S.S. Cheng, Y.H. Wei, D. Sheng, et al., Identification for Hammerstein nonlinear ARMAX systems based on multi-innovation fractional order stochastic gradient, Signal Processing 142 (2018) 1-10.

[31] Y.J. Wang, F. Ding, A filtering based multi-innovation gradient estimation algorithm and performance analysis for nonlinear dynamical systems, IMA Journal of Applied Mathematics 82 (6) (2017) 1171-1191.

[32] L. Xu, F. Ding, Y. Gu, A. Alsaedi, T. Hayat, A multi-innovation state and parameter estimation algorithm for a state space system with d-step state-delay, Signal Processing 140 (2017) 97-103.

[33] Y. Cao, P. Li, Y. Zhang, Parallel processing algorithm for railway signal fault diagnosis data based on cloud computing, Future Generation Computer Systems 88 (2018) 279-283.

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

[35] Y. Cao, Y. Wen, X. Meng, W. Xu, Performance evaluation with improved receiver design for asynchronous coordinated multipoint transmissions, Chinese Journal of Electronics 25 (2) (2016) 372-378.

[36] Y.Z. Zhang, Y. Cao, Y.H. Wen, L. Liang, F. Zou, Optimization of information interaction protocols in cooperative vehicle-infrastructure systems, Chinese Journal of Electronics 27 (2) (2018) 439-444.

[37] X. Zhang, F. Ding, F.E. Alsaadi, T. Hayat, Recursive parameter identification of the dynamical models for bilinear state space systems, Nonlinear Dynamics 89 (4) (2017) 2415-2429.

[38] M.T. Chen, F. Ding, L. Xu, T. Hayat, A. Alsaedi, Iterative identification algorithms for bilinear-in-parameter systems with autoregressive moving average noise, Journal of the Franklin Institute 354 (17) (2017) 7885-7898.

[39] J. Pan, X. Jiang, X.K. Wan, et al., A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems, International Journal of Control, Automation, and Systems 15 (3) (2017) 1189-1197.

[40] X. Zhang, F. Ding, L. Xu, E.F. Yang, State filtering-based least squares parameter estimation for bilinear systems using the hierarchical identification principle, IET Control Theory and Applications 12 (12) 1704-1713.

[41] F. Ding, H.B. Chen, L. Xu, J.Y. Dai, Q.S. Li, T. Hayat, A hierarchical least squares identification algorithm for Hammerstein nonlinear systems using the key term separation, Journal of the Franklin Institute 355 (8) (2018) 3737-3752.

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

[43] F. Ding, D.D. Meng, J.Y. Dai, Q.S. Li, A. Alsaedi, T. Hayat, Least squares based iterative parameter estimation algorithm for stochastic dynamical systems with ARMA noise using the model equivalence, International Journal of Control, Automation and Systems 16 (2) (2018) 630-639.

[44] F. Ding, L. Xu, F.E. Alsaadi, T. Hayat, Iterative parameter identification for pseudo-linear systems with ARMA noise using the filtering technique, IET Control Theory and Applications 12 (7) (2018) 892-899.

[45] F. Liu, H.X. Wu, Singular integrals related to homogeneous mappings in triebel-lizorkin spaces, Journal of Mathematical Inequalities 11 (4) (2017) 1075-1097.

[46] F. Liu, H.X. Wu, Regularity of discrete multisublinear fractional maximal functions, Science China–Mathematics 60 (8) (2017) 1461-1476.

[47] F. Liu, H.X. Wu, On the regularity of maximal operators supported by submanifolds, Journal of Mathematical Analysis and Applications 453 (1) (2017) 144-158.

[48] C.C. Yin, C.W. Wang, The perturbed compound Poisson risk process with investment and debit interest, Methodology and Computing in Applied Probability 12 (3) (2010) 391-413.

[49] C.C. Yin, K.C. Yuen, Optimality of the threshold dividend strategy for the compound Poisson model, Statistics & Probability Letters 81 (12) (2011) 1841-1846.

[50] C.C. Yin, Y.Z. Wen, Exit problems for jump processes with applications to dividend problems, Journal of Computational and Applied Mathematics 245 (2013) 30-52.

[51] C.C. Yin, Y.Z. Wen, Optimal dividend problem with a terminal value for spectrally positive Levy processes, Insurance Mathematics & Economics 53 (3) (2013) 769-773.

[52] P.C Gong, W.Q. Wang, F.C. Li, H. Cheung, Sparsity-aware transmit beamspace design for FDA-MIMO radar, Signal Processing 144 (2018) 99-103.

[53] Z.H. Rao, C.Y. Zeng, 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 12 (1) (2018) 413-435.

[54] N. Zhao, R. Liu, Y. Chen, M. Wu, Y. Jiang, W. Xiong, C. Liu. Contract design for relay incentive mechanism under dual asymmetric information in cooperative networks, Wireless Networks (2018). doi: 10.1007/s11276-017-1518-x