Data Filtering Based Stochastic Gradient Algorithms for Multivariable CARAR-Like Systems*

Dongqing Wang[a], Tong Shan[b] and Rui Ding[c]

[a] College of Automation Engineering, Qingdao University
 266071 Qingdao, China
[b] School of Computer Science, Fudan University
 201203 Shanghai, China
[c] School of Internet of Things Engineering, Jiangnan University
 214122 Wuxi, China
 E-mail(corresp.): dqwang64@163.com
 E-mail: 10300240057@fudan.edu.cn
 E-mail: rding2003@126.com

Abstract. This paper considers identification problems for a multivariable controlled autoregressive system with autoregressive noises. A hierarchical generalized stochastic gradient algorithm and a filtering based hierarchical stochastic gradient algorithm are presented to estimate the parameter vectors and parameter matrix of such multivariable colored noise systems, by using the hierarchical identification principle. The simulation results show that the proposed hierarchical gradient estimation algorithms are effective.

Keywords: parameter identification, mathematical model, error estimates.

AMS Subject Classification: 93B30; 62G05.

1 Introduction

Multivariable system modelling has received much attention in various practical systems, including magnetic compressors and magnetic fluids [9, 17], piston engines [8], distillation columns [13, 14], fault detection systems [12, 18] and

travelling waves [11], etc. As a consequence of this wide variety of applications, different identification algorithms for multivariable systems have been vastly reported in the literature, e.g., the gradient based iterative algorithm and the least squares based iterative algorithm for multivariable CARARMA systems [7], the hierarchical gradient-based iterative identification algorithms for multivariable CARAR-like systems [21], the stochastic gradient estimation algorithm for multivariable equation error systems [15], the auxiliary model-based multi-innovation stochastic gradient algorithm for multiple-input single-output systems [16], the bias compensation based identification algorithms for multivariable systems [22, 23], and the coupled-least-squares identification for multivariable systems [1].

In the previous work, Wang and Ding studied a filtering based least squares algorithm for CARARMA single variable systems [20]. Wang discussed filtering based recursive and iterative least squares algorithms for OEMA single variable systems [19]. Ding et al. proposed the hierarchical least squares and stochastic gradient algorithms for Hammerstein nonlinear systems [2, 6]. Ding and Chen presented the hierarchical least squares and stochastic gradient algorithms for multivariable systems [4, 5].

On the basis of the above research, this paper studies the identification problems for multivariable controlled autoregressive systems with autoregressive colored noises (multivariable CARAR-like systems) by using the data filtering technique [19, 20] and the hierarchical identification principle [4, 5]. First, by directly transforming the CARAR-like system, we get an identification model contains a parameter matrix and a parameter vector. By using the hierarchical identification principle [4, 5], a hierarchical generalized stochastic gradient (HGSG) algorithm is used to estimate the parameter matrix and the parameter vector in an alternating manner. Further, through filtering the input-output data of the system, we get an identification model contains a parameter matrix and two parameter vectors. Then a filtering based hierarchical stochastic gradient (F-HSG) algorithm is presented to estimate the parameters of the multivariable CARAR-like system by combining the filtering technique [19, 20] and the hierarchical identification principle [4, 5].

The basic idea of the F-HSG algorithm is to use a rational polynomial to filter the input-output data of the system, to transform a multivariable CARAR-like system into two identification models: a multivariable ARX-like system model which is parametrised by a parameter matrix and a parameter vector, and a multivariable AR noise model which is parametrised by only a parameter vector, and then to estimate the system parameter matrix and the two parameter vectors in an alternating manner by using the hierarchical identification principle.

The rest of the paper is organized as follows. Section 2 gives description of the HGSG algorithm for the multivariable CARAR-like system. Section 3 employs the filtering technique and presents the F-HSG identification algorithm. Section 4 gives an illustrative example to show the effectiveness of the proposed algorithms. Finally, the concluding remarks are given in Section 5.

2 Hierarchical Generalized Stochastic Gradient Algorithm

Let us introduce some notations first. The symbol $\mathbf{1}_n$ (or $\mathbf{1}_{m \times n}$) represents an n-dimensional column vector (or an $m \times n$-dimensional matrix) whose entries are 1; the superscript T denotes the matrix transpose; the norm of a matrix $\boldsymbol{X}$ is defined by $\|\boldsymbol{X}\|^2 := \mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$.
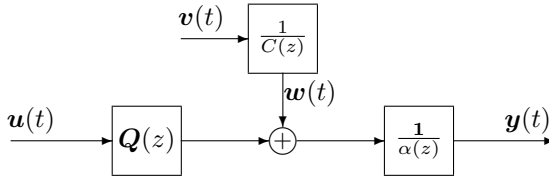

Figure 1. The considered multivariable CARAR-like model.

This section extends the Ding and Chen's hierarchical stochastic gradient method for multivariable systems [4] to study identification problems of multivariable controlled autoregressive systems with autoregressive colored noises (multivariable CARAR-like systems) described by the following equation [3]

$$\alpha(z)\boldsymbol{y}(t) = \boldsymbol{Q}(z)\boldsymbol{u}(t) + \frac{1}{C(z)}\boldsymbol{v}(t), \tag{2.1}$$

where $\boldsymbol{y}(t) \in \mathbb{R}^m$ is the system output vector, $\boldsymbol{u}(t) \in \mathbb{R}^r$ is the system input vector, $\boldsymbol{v}(t) \in \mathbb{R}^m$ is a white noise vector with zero mean, $\alpha(z)$ is the system characteristic polynomial in the unit backward shift operator z^{-1}: $z^{-1}\boldsymbol{y}(t) = \boldsymbol{y}(t-1)$, $\boldsymbol{Q}(z)$ is a matrix polynomial in z^{-1}, and $C(z)$ is a polynomial in z^{-1}, which are defined as

$$\alpha(z) := 1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \cdots + \alpha_{n_\alpha} z^{-n_\alpha}, \quad \alpha_i \in \mathbb{R},$$
$$\boldsymbol{Q}(z) := \boldsymbol{Q}_1 z^{-1} + \boldsymbol{Q}_2 z^{-2} + \cdots + \boldsymbol{Q}_{n_q} z^{-n_q}, \quad \boldsymbol{Q}_i \in \mathbb{R}^{m \times r},$$
$$C(z) := 1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}, \quad c_i \in \mathbb{R}.$$

The system model is depicted in Figure 1. Referring to Figure 1, the colored noise has the following form

$$\boldsymbol{w}(t) := \frac{1}{C(z)}\boldsymbol{v}(t) \in \mathbb{R}^m, \tag{2.2}$$

or

$$\boldsymbol{w}(t) = \big[1 - C(z)\big]\boldsymbol{w}(t) + \boldsymbol{v}(t). \tag{2.3}$$

Then, (2.1) may be written as

$$\alpha(z)\boldsymbol{y}(t) = \boldsymbol{Q}(z)\boldsymbol{u}(t) + \boldsymbol{w}(t). \tag{2.4}$$

Note that the colored noise $\boldsymbol{w}(t) := \frac{1}{C(z)}\boldsymbol{v}(t)$ is an autoregressive (AR) process. Define the parameter vector $\boldsymbol{\vartheta}$, the parameter matrix $\boldsymbol{\theta}$, and the input information vector $\boldsymbol{\varphi}(t)$ and the information matrix $\boldsymbol{\psi}(t)$ as

$$\boldsymbol{\vartheta} := \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{c} \end{bmatrix} \in \mathbb{R}^{n_\alpha + n_c},$$

$$\boldsymbol{\alpha} := [\alpha_1, \alpha_2, \ldots, \alpha_{n_\alpha}]^\mathrm{T} \in \mathbb{R}^{n_\alpha}, \qquad \boldsymbol{c} := [c_1, c_2, \ldots, c_{n_c}]^\mathrm{T} \in \mathbb{R}^{n_c},$$

$$\boldsymbol{\theta}^\mathrm{T} := [\boldsymbol{Q}_1, \boldsymbol{Q}_2, \ldots, \boldsymbol{Q}_{n_q}] \in \mathbb{R}^{m \times (n_q r)},$$

$$\boldsymbol{\varphi}(t) := [\boldsymbol{u}^\mathrm{T}(t-1), \boldsymbol{u}^\mathrm{T}(t-2), \ldots, \boldsymbol{u}^\mathrm{T}(t-n_q)]^\mathrm{T} \in \mathbb{R}^{n_q r},$$

$$\boldsymbol{\psi}(t) := [\boldsymbol{\psi}_\mathrm{s}(t), \boldsymbol{\psi}_\mathrm{n}(t)] \in \mathbb{R}^{m \times (n_\alpha + n_c)},$$

$$\boldsymbol{\psi}_\mathrm{s}(t) := [\boldsymbol{y}(t-1), \boldsymbol{y}(t-2), \ldots, \boldsymbol{y}(t-n_\alpha)] \in \mathbb{R}^{m \times n_\alpha},$$

$$\boldsymbol{\psi}_\mathrm{n}(t) := [\boldsymbol{w}(t-1), \boldsymbol{w}(t-2), \ldots, \boldsymbol{w}(t-n_c)] \in \mathbb{R}^{m \times n_c}.$$

Equation (2.4) can be written as

$$\boldsymbol{y}(t) = -\boldsymbol{\psi}_\mathrm{s}(t)\boldsymbol{\alpha} + \boldsymbol{\theta}^\mathrm{T}\boldsymbol{\varphi}(t) + \boldsymbol{w}(t) \tag{2.5}$$

$$= -\boldsymbol{\psi}_\mathrm{s}(t)\boldsymbol{\alpha} - \boldsymbol{\psi}_\mathrm{n}(t)\boldsymbol{c} + \boldsymbol{\theta}^\mathrm{T}\boldsymbol{\varphi}(t) + \boldsymbol{v}(t)$$

$$= -\boldsymbol{\psi}(t)\boldsymbol{\vartheta} + \boldsymbol{\theta}^\mathrm{T}\boldsymbol{\varphi}(t) + \boldsymbol{v}(t). \tag{2.6}$$

Equation (2.6) is the regression model suitable for identification of the multivariable CARAR-like system in (2.1), and contains one parameter vector $\boldsymbol{\vartheta}$ which consists of the coefficients of two polynomials, and one parameter matrix $\boldsymbol{\theta}$ which consists of the coefficients of the matrix polynomial. Define a quadratic cost function [4]:

$$J_1(\boldsymbol{\vartheta}, \boldsymbol{\theta}) := \left\| \boldsymbol{y}(t) + \boldsymbol{\psi}(t)\boldsymbol{\vartheta} - \boldsymbol{\theta}^\mathrm{T}\boldsymbol{\varphi}(t) \right\|^2.$$

Let $\mu_1(t)$ and $\mu_2(t)$ be two convergence factors, and $\hat{\boldsymbol{\vartheta}}(t) := \begin{bmatrix} \hat{\boldsymbol{\alpha}}(t) \\ \hat{\boldsymbol{c}}(t) \end{bmatrix}$ and $\hat{\boldsymbol{\theta}}(t)$ represent the parameter estimates of $\boldsymbol{\vartheta} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{c} \end{bmatrix}$ and $\boldsymbol{\theta}$ at time t. Based on the hierarchical identification principle [4,5] and the negative gradient search [10], minimizing $J_1(\boldsymbol{\vartheta}, \boldsymbol{\theta})$ yields the generalized stochastic gradient algorithm:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) - \frac{\mu_1(t)}{2}\mathrm{grad}_\vartheta\left[J_1\left(\hat{\boldsymbol{\vartheta}}(t-1), \hat{\boldsymbol{\theta}}(t-1)\right)\right],$$

$$= \hat{\boldsymbol{\vartheta}}(t-1) - \mu_1(t)\boldsymbol{\psi}^\mathrm{T}(t)\left[\boldsymbol{y}(t) + \boldsymbol{\psi}(t)\hat{\boldsymbol{\vartheta}}(t-1) - \hat{\boldsymbol{\theta}}^\mathrm{T}(t-1)\boldsymbol{\varphi}(t)\right], \tag{2.7}$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\mu_2(t)}{2}\mathrm{grad}_\theta\left[J_1\left(\hat{\boldsymbol{\vartheta}}(t), \hat{\boldsymbol{\theta}}(t-1)\right)\right],$$

$$= \hat{\boldsymbol{\theta}}(t-1) + \mu_2(t)\boldsymbol{\varphi}(t)\left[\boldsymbol{y}(t) + \boldsymbol{\psi}(t)\hat{\boldsymbol{\vartheta}}(t) - \hat{\boldsymbol{\theta}}^\mathrm{T}(t-1)\boldsymbol{\varphi}(t)\right]^\mathrm{T}. \tag{2.8}$$

The symbol grad_x denotes the gradient with respect to x. Like in [10], we take two convergence factors to be

$$\mu_1(t) := 1/r_1(t), \qquad r_1(t) = r_1(t-1) + \left\| \boldsymbol{\psi}(t) \right\|^2,$$

$$\mu_2(t) := 1/r_2(t), \qquad r_2(t) = r_2(t-1) + \left\| \boldsymbol{\varphi}(t) \right\|^2.$$

Thus, we have

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) - \frac{\boldsymbol{\psi}^{\mathrm{T}}(t)}{r_1(t)} \big[\boldsymbol{y}(t) + \boldsymbol{\psi}(t)\hat{\boldsymbol{\vartheta}}(t-1) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t) \big], \qquad (2.9)$$

$$r_1(t) = r_1(t-1) + \big\| \boldsymbol{\psi}(t) \big\|^2, \qquad r_1(0) = 1, \qquad (2.10)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\varphi}(t)}{r_2(t)} \big[\boldsymbol{y}(t) + \boldsymbol{\psi}(t)\hat{\boldsymbol{\vartheta}}(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t) \big]^{\mathrm{T}}, \qquad (2.11)$$

$$r_2(t) = r_2(t-1) + \big\| \boldsymbol{\varphi}(t) \big\|^2, \qquad r_2(0) = 1. \qquad (2.12)$$

Because the information matrix $\boldsymbol{\psi}(t)$ contains the unmeasurable noise term $\boldsymbol{w}(t-i)$, the algorithms in (2.9)–(2.12) cannot be implemented. The problem is solved by replacing the unmeasurable $\boldsymbol{w}(t-i)$ with its estimate $\hat{\boldsymbol{w}}(t-i)$.

Thus we have the hierarchical generalized stochastic gradient (HGSG) algorithm for estimating $\boldsymbol{\vartheta}$ and $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) - \frac{\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)}{r_1(t)} \big[\boldsymbol{y}(t) + \hat{\boldsymbol{\psi}}(t)\hat{\boldsymbol{\vartheta}}(t-1) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t) \big], \qquad (2.13)$$

$$r_1(t) = r_1(t-1) + \big\| \hat{\boldsymbol{\psi}}(t) \big\|^2, \qquad r_1(0) = 1, \qquad (2.14)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\varphi}(t)}{r_2(t)} \big[\boldsymbol{y}(t) + \hat{\boldsymbol{\psi}}(t)\hat{\boldsymbol{\vartheta}}(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t) \big]^{\mathrm{T}}, \qquad (2.15)$$

$$r_2(t) = r_2(t-1) + \big\| \boldsymbol{\varphi}(t) \big\|^2, \qquad r_2(0) = 1, \qquad (2.16)$$

$$\boldsymbol{\varphi}(t) = \big[\boldsymbol{u}^{\mathrm{T}}(t-1), \boldsymbol{u}^{\mathrm{T}}(t-2), \ldots, \boldsymbol{u}^{\mathrm{T}}(t-n_q) \big]^{\mathrm{T}}, \qquad (2.17)$$

$$\boldsymbol{\psi}_{\mathrm{s}}(t) = \big[\boldsymbol{y}(t-1), \boldsymbol{y}(t-2), \ldots, \boldsymbol{y}(t-n_\alpha) \big], \qquad (2.18)$$

$$\hat{\boldsymbol{\psi}}(t) = \big[\boldsymbol{\psi}_{\mathrm{s}}(t), \hat{\boldsymbol{w}}(t-1), \hat{\boldsymbol{w}}(t-2), \ldots, \hat{\boldsymbol{w}}(t-n_c) \big], \qquad (2.19)$$

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) + \boldsymbol{\psi}_{\mathrm{s}}(t)\hat{\boldsymbol{\alpha}}(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t). \qquad (2.20)$$

The procedures of computing $\hat{\boldsymbol{\vartheta}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ in the HGSG algorithm are listed as follows

1) Let $t = 1$, $\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_{n_\alpha + n_c}/p_0$, $\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_{(n_q r) \times m}/p_0$, $\hat{\boldsymbol{w}}(i) = \mathbf{1}_m/p_0$, $\boldsymbol{u}(i) = \mathbf{0}$, $\boldsymbol{y}(i) = \mathbf{0}$ as $i \leq 0$, $p_0 = 10^6$.

2) Collect the input–output data $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$, form $\boldsymbol{\varphi}(t)$, $\boldsymbol{\psi}_{\mathrm{s}}(t)$ and $\hat{\boldsymbol{\psi}}(t)$ by (2.17)–(2.19).

3) Compute $r_1(t)$ and $r_2(t)$ by (2.14) and (2.16), respectively.

4) Update the parameter estimates $\hat{\boldsymbol{\vartheta}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ by (2.13) and (2.15), respectively.

5) Compute $\hat{\boldsymbol{w}}(t)$ by (2.20).

6) Increase t by 1 and go to Step 2.

3 Filtering Based Hierarchical Stochastic Gradient Algorithm

For the CARAR-like model (2.1), if the input-output data are filtered through the polynomial $C(z)$ (a linear filter), model (2.1) can be transformed into two identification models: a multivariable ARX-like model and a multivariable AR noise model, then the hierarchical stochastic gradient algorithm can be applied. The identification method based on this idea is called the filtering based hierarchical stochastic gradient algorithm (F-HSG). Because $C(z)$ is unknown, its estimate $\hat{C}(t, z)$ is used to filter the input-output data [20].

Define the filtered input vector $\boldsymbol{u}_f(t)$, the filtered output vector $\boldsymbol{y}_f(t)$, the filtered information vector $\boldsymbol{\varphi}_f(t)$ and the filtered information matrix $\boldsymbol{\psi}_f(t)$ as

$$\boldsymbol{u}_f(t) := C(z)\boldsymbol{u}(t), \qquad \boldsymbol{y}_f(t) := C(z)\boldsymbol{y}(t),$$

$$\boldsymbol{\varphi}_f(t) := \left[\boldsymbol{u}_f^T(t-1), \boldsymbol{u}_f^T(t-2), \ldots, \boldsymbol{u}_f^T(t-n_q)\right]^T \in \mathbb{R}^{n_q r},$$

$$\boldsymbol{\psi}_f(t) := \left[\boldsymbol{y}_f(t-1), \boldsymbol{y}_f(t-2), \ldots, \boldsymbol{y}_f(t-n_\alpha)\right] \in \mathbb{R}^{m \times n_\alpha}.$$

Thus multiplying both sides of (2.1) by $C(z)$ gives

$$\alpha(z)\boldsymbol{y}_f(t) = \boldsymbol{Q}(z)\boldsymbol{u}_f(t) + \boldsymbol{v}(t),$$

which can be rewritten as

$$\boldsymbol{y}_f(t) + \boldsymbol{\psi}_f(t)\boldsymbol{\alpha} = \boldsymbol{\theta}^T \boldsymbol{\varphi}_f(t) + \boldsymbol{v}(t). \tag{3.1}$$

Equation (2.3) can be written as

$$\boldsymbol{w}(t) = -\boldsymbol{\psi}_n(t)\boldsymbol{c} + \boldsymbol{v}(t), \tag{3.2}$$

For two identification models in (3.1) and (3.2) of the multivariable CARAR-like system, we define two cost functions,

$$J_2(\boldsymbol{\alpha}, \boldsymbol{\theta}) := \left\| \boldsymbol{y}_f(t) + \boldsymbol{\psi}_f(t)\boldsymbol{\alpha} - \boldsymbol{\theta}^T \boldsymbol{\varphi}_f(t) \right\|^2, \quad J_3(\boldsymbol{c}) := \left\| \boldsymbol{w}(t) + \boldsymbol{\psi}_n(t)\boldsymbol{c} \right\|^2.$$

Let $\mu_3(t)$, $\mu_4(t)$ and $\mu_5(t)$ be three convergence factors. Using the negative gradient search [10], and minimizing $J_2(\boldsymbol{\alpha}, \boldsymbol{\theta})$ and $J_3(\boldsymbol{c})$ yields the following filtering based stochastic gradient algorithm [4]:

$$\hat{\boldsymbol{\alpha}}(t) = \hat{\boldsymbol{\alpha}}(t-1) - \frac{\mu_3(t)}{2}\text{grad}_\alpha\left[J_2\big(\hat{\boldsymbol{\alpha}}(t-1), \hat{\boldsymbol{\theta}}(t-1)\big)\right],$$

$$= \hat{\boldsymbol{\alpha}}(t-1) - \mu_3(t)\boldsymbol{\psi}_f^T(t)\left[\boldsymbol{y}_f(t) + \boldsymbol{\psi}_f(t)\hat{\boldsymbol{\alpha}}(t-1) - \hat{\boldsymbol{\theta}}^T(t-1)\boldsymbol{\varphi}_f(t)\right], \quad (3.3)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\mu_4(t)}{2}\text{grad}_\theta\left[J_2\big(\hat{\boldsymbol{\alpha}}(t), \hat{\boldsymbol{\theta}}(t-1)\big)\right],$$

$$= \hat{\boldsymbol{\theta}}(t-1) + \mu_4(t)\boldsymbol{\varphi}_f(t)\left[\boldsymbol{y}_f(t) + \boldsymbol{\psi}_f(t)\hat{\boldsymbol{\alpha}}(t) - \hat{\boldsymbol{\theta}}^T(t-1)\boldsymbol{\varphi}_f(t)\right]^T, \quad (3.4)$$

$$\hat{\boldsymbol{c}}(t) = \hat{\boldsymbol{c}}(t-1) - \frac{\mu_5(t)}{2}\text{grad}_c\left[J_3\big(\hat{\boldsymbol{c}}(t-1)\big)\right],$$

$$= \hat{\boldsymbol{c}}(t-1) - \mu_5(t)\boldsymbol{\psi}_n^T(t)\left[\boldsymbol{w}(t) + \boldsymbol{\psi}_n(t)\hat{\boldsymbol{c}}(t-1)\right]. \tag{3.5}$$

Setting

$$\mu_3(t) := \frac{1}{r_3(t)}, \qquad r_3(t) = r_3(t-1) + \left\| \boldsymbol{\psi}_{\mathrm{f}}(t) \right\|^2,$$

$$\mu_4(t) := \frac{1}{r_4(t)}, \qquad r_4(t) = r_4(t-1) + \left\| \boldsymbol{\varphi}_{\mathrm{f}}(t) \right\|^2,$$

$$\mu_5(t) := \frac{1}{r_5(t)}, \qquad r_5(t) = r_5(t-1) + \left\| \boldsymbol{\psi}_{\mathrm{n}}(t) \right\|^2,$$

we have

$$\hat{\boldsymbol{\alpha}}(t) = \hat{\boldsymbol{\alpha}}(t-1) - \frac{\boldsymbol{\psi}_{\mathrm{f}}^{\mathrm{T}}(t)}{r_3(t)} \left[\boldsymbol{y}_{\mathrm{f}}(t) + \boldsymbol{\psi}_{\mathrm{f}}(t)\hat{\boldsymbol{\alpha}}(t-1) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}_{\mathrm{f}}(t) \right], \qquad (3.6)$$

$$r_3(t) = r_3(t-1) + \left\| \boldsymbol{\psi}_{\mathrm{f}}(t) \right\|^2, \qquad r_3(0) = 1, \qquad (3.7)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\varphi}_{\mathrm{f}}(t)}{r_4(t)} \left[\boldsymbol{y}_{\mathrm{f}}(t) + \boldsymbol{\psi}_{\mathrm{f}}(t)\hat{\boldsymbol{\alpha}}(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}_{\mathrm{f}}(t) \right]^{\mathrm{T}}, \qquad (3.8)$$

$$r_4(t) = r_4(t-1) + \left\| \boldsymbol{\varphi}_{\mathrm{f}}(t) \right\|^2, \qquad r_4(0) = 1, \qquad (3.9)$$

$$\hat{\boldsymbol{c}}(t) = \hat{\boldsymbol{c}}(t-1) - \frac{\boldsymbol{\psi}_{\mathrm{n}}^{\mathrm{T}}(t)}{r_5(t)} \left[\boldsymbol{w}(t) + \boldsymbol{\psi}_{\mathrm{n}}(t)\hat{\boldsymbol{c}}(t-1) \right], \qquad (3.10)$$

$$r_5(t) = r_5(t-1) + \left\| \boldsymbol{\psi}_{\mathrm{n}}(t) \right\|^2, \qquad r_5(0) = 1. \qquad (3.11)$$

Similarly, the polynomial $C(z)$ is unknown, so are $\boldsymbol{u}_{\mathrm{f}}(t)$ and $\boldsymbol{y}_{\mathrm{f}}(t)$, and the information vector $\boldsymbol{\varphi}_{\mathrm{f}}(t)$ and matrix $\boldsymbol{\psi}_{\mathrm{f}}(t)$ are unknown, the estimates $\hat{\boldsymbol{\alpha}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ in (3.6) and (3.8) cannot be implemented. The information matrix $\boldsymbol{\psi}_{\mathrm{n}}(t)$ contains the unavailable noise term $\boldsymbol{w}(t-i)$, so the parameter estimate $\hat{\boldsymbol{c}}(t)$ in (3.10) cannot be implemented.

The problem is solved by replacing the unknown filtered vectors $\boldsymbol{u}_{\mathrm{f}}(t)$ and $\boldsymbol{y}_{\mathrm{f}}(t)$ with their estimates $\hat{\boldsymbol{u}}_{\mathrm{f}}(t)$ and $\hat{\boldsymbol{y}}_{\mathrm{f}}(t)$ and replacing the unavailable noise term $\boldsymbol{w}(t-i)$ with its estimate $\hat{\boldsymbol{w}}(t-i)$. Equation (2.5) gives

$$\boldsymbol{w}(t) = \boldsymbol{y}(t) + \boldsymbol{\psi}_{\mathrm{s}}(t)\boldsymbol{\alpha} - \boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(t). \qquad (3.12)$$

Replacing the unknown $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ in (3.12) with the estimates $\hat{\boldsymbol{\alpha}}(t-1)$ and $\hat{\boldsymbol{\theta}}(t-1)$, the estimate $\hat{\boldsymbol{w}}(t)$ can be computed by

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) + \boldsymbol{\psi}_{\mathrm{s}}(t)\hat{\boldsymbol{\alpha}}(t-1) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t).$$

Use $\hat{\boldsymbol{w}}(t-i)$ to construct the estimate of $\boldsymbol{\psi}_{\mathrm{n}}(t)$:

$$\hat{\boldsymbol{\psi}}_{\mathrm{n}}(t) = \left[\hat{\boldsymbol{w}}(t-1), \hat{\boldsymbol{w}}(t-2), \dots, \hat{\boldsymbol{w}}(t-n_c) \right].$$

Use the parameter estimate $\hat{\boldsymbol{c}}(t) = [\hat{c}_1(t), \hat{c}_2(t), \dots, \hat{c}_{n_c}(t)]^{\mathrm{T}}$ to construct the estimates of $C(z)$:

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

By filtering u(t) and y(t) with $\hat{C}(t;z)$, one obtains the estimates of $\boldsymbol{u}_{\mathrm{f}}(t)$ and $\boldsymbol{y}_{\mathrm{f}}(t)$:

$$\hat{\boldsymbol{u}}_{\mathrm{f}}(t) = \hat{C}(t, z)\boldsymbol{u}(t), \qquad \hat{\boldsymbol{y}}_{\mathrm{f}}(t) = \hat{C}(t, z)\boldsymbol{y}(t).$$

Also, $\hat{\boldsymbol{u}}_{\mathrm{f}}(t)$ and $\hat{\boldsymbol{y}}_{\mathrm{f}}(t)$ can be recursively computed by

$$\hat{\boldsymbol{u}}_{\mathrm{f}}(t) = \boldsymbol{u}(t) + \hat{c}_1(t)\boldsymbol{u}(t-1) + \hat{c}_2(t)\boldsymbol{u}(t-2) + \cdots + \hat{c}_{n_c}(t)\boldsymbol{u}(t-n_c),$$
$$\hat{\boldsymbol{y}}_{\mathrm{f}}(t) = \boldsymbol{y}(t) + \hat{c}_1(t)\boldsymbol{y}(t-1) + \hat{c}_2(t)\boldsymbol{y}(t-2) + \cdots + \hat{c}_{n_c}(t)\boldsymbol{y}(t-n_c).$$

Construct the estimates of $\boldsymbol{\varphi}_{\mathrm{f}}(t)$ and $\boldsymbol{\psi}_{\mathrm{f}}(t)$ with $\hat{\boldsymbol{u}}_{\mathrm{f}}(t-i)$ and $\hat{\boldsymbol{y}}_{\mathrm{f}}(t-i)$ as follows:

$$\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t) = \left[\hat{\boldsymbol{u}}_{\mathrm{f}}^{\mathrm{T}}(t-1), \hat{\boldsymbol{u}}_{\mathrm{f}}^{\mathrm{T}}(t-2), \ldots, \hat{\boldsymbol{u}}_{\mathrm{f}}^{\mathrm{T}}(t-n_q)\right]^{\mathrm{T}},$$
$$\hat{\boldsymbol{\psi}}_{\mathrm{f}}(t) = \left[\hat{\boldsymbol{y}}_{\mathrm{f}}(t-1), \hat{\boldsymbol{y}}_{\mathrm{f}}(t-2), \ldots, \hat{\boldsymbol{y}}_{\mathrm{f}}(t-n_\alpha)\right].$$

Replacing the unknown information vector $\boldsymbol{\psi}_{\mathrm{f}}(t)$ in (3.6)–(3.8) with $\hat{\boldsymbol{\psi}}_{\mathrm{f}}(t)$, $\boldsymbol{\varphi}_{\mathrm{f}}(t)$ in (3.6), (3.8) and (3.9) with $\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t)$, and $\boldsymbol{\psi}_{\mathrm{n}}(t)$ in (3.10) and (3.11) with $\hat{\boldsymbol{\psi}}_{\mathrm{n}}(t)$, the unknown filtered output $\boldsymbol{y}_{\mathrm{f}}(t)$ in (3.6) and (3.8) with $\hat{\boldsymbol{y}}_{\mathrm{f}}(t)$, and the unknown vector $\boldsymbol{w}(t)$ in (3.10) with $\hat{\boldsymbol{w}}(t)$, we obtain the filtering based hierarchical stochastic gradient (F-HSG) algorithm for estimating $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$ and $\boldsymbol{c}$ for the CARAR-like systems:

$$\hat{\boldsymbol{\alpha}}(t) = \hat{\boldsymbol{\alpha}}(t-1) - \frac{\hat{\boldsymbol{\psi}}_{\mathrm{f}}^{\mathrm{T}}(t)}{r_3(t)}\left[\hat{\boldsymbol{y}}_{\mathrm{f}}(t) + \hat{\boldsymbol{\psi}}_{\mathrm{f}}(t)\hat{\boldsymbol{\alpha}}(t-1) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t)\right], \qquad (3.13)$$

$$r_3(t) = r_3(t-1) + \left\|\hat{\boldsymbol{\psi}}_{\mathrm{f}}(t)\right\|^2, \qquad r_3(0) = 1, \qquad (3.14)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t)}{r_4(t)}\left[\hat{\boldsymbol{y}}_{\mathrm{f}}(t) + \hat{\boldsymbol{\psi}}_{\mathrm{f}}(t)\hat{\boldsymbol{\alpha}}(t) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t)\right]^{\mathrm{T}}, \qquad (3.15)$$

$$r_4(t) = r_4(t-1) + \left\|\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t)\right\|^2, \qquad r_4(0) = 1, \qquad (3.16)$$

$$\hat{\boldsymbol{c}}(t) = \hat{\boldsymbol{c}}(t-1) - \frac{\hat{\boldsymbol{\psi}}_{\mathrm{n}}^{\mathrm{T}}(t)}{r_5(t)}\left[\hat{\boldsymbol{w}}(t) + \hat{\boldsymbol{\psi}}_{\mathrm{n}}(t)\hat{\boldsymbol{c}}(t-1)\right], \qquad (3.17)$$

$$r_5(t) = r_5(t-1) + \left\|\hat{\boldsymbol{\psi}}_{\mathrm{n}}(t)\right\|^2, \qquad r_5(0) = 1, \qquad (3.18)$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t) = \left[\hat{\boldsymbol{u}}_{\mathrm{f}}^{\mathrm{T}}(t-1), \hat{\boldsymbol{u}}_{\mathrm{f}}^{\mathrm{T}}(t-2), \ldots, \hat{\boldsymbol{u}}_{\mathrm{f}}^{\mathrm{T}}(t-n_q)\right]^{\mathrm{T}}, \qquad (3.19)$$

$$\hat{\boldsymbol{\psi}}_{\mathrm{f}}(t) = \left[\hat{\boldsymbol{y}}_{\mathrm{f}}(t-1), \hat{\boldsymbol{y}}_{\mathrm{f}}(t-2), \ldots, \hat{\boldsymbol{y}}_{\mathrm{f}}(t-n_\alpha)\right], \qquad (3.20)$$

$$\hat{\boldsymbol{\psi}}_{\mathrm{n}}(t) = \left[\hat{\boldsymbol{w}}(t-1), \hat{\boldsymbol{w}}(t-2), \ldots, \hat{\boldsymbol{w}}(t-n_c)\right], \qquad (3.21)$$

$$\boldsymbol{\varphi}(t) = \left[\boldsymbol{u}^{\mathrm{T}}(t-1), \boldsymbol{u}^{\mathrm{T}}(t-2), \ldots, \boldsymbol{u}^{\mathrm{T}}(t-n_q)\right]^{\mathrm{T}}, \qquad (3.22)$$

$$\boldsymbol{\psi}_{\mathrm{s}}(t) = \left[\boldsymbol{y}(t-1), \boldsymbol{y}(t-2), \ldots, \boldsymbol{y}(t-n_\alpha)\right], \qquad (3.23)$$

$$\hat{\boldsymbol{u}}_{\mathrm{f}}(t) = \boldsymbol{u}(t) + \hat{c}_1(t)\boldsymbol{u}(t-1) + \hat{c}_2(t)\boldsymbol{u}(t-2) + \cdots + \hat{c}_{n_c}(t)\boldsymbol{u}(t-n_c), \qquad (3.24)$$

$$\hat{\boldsymbol{y}}_{\mathrm{f}}(t) = \boldsymbol{y}(t) + \hat{c}_1(t)\boldsymbol{y}(t-1) + \hat{c}_2(t)\boldsymbol{y}(t-2) + \cdots + \hat{c}_{n_c}(t)\boldsymbol{y}(t-n_c), \qquad (3.25)$$

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) + \boldsymbol{\psi}_{\mathrm{s}}(t)\hat{\boldsymbol{\alpha}}(t-1) - \hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t), \qquad (3.26)$$

$$\hat{\boldsymbol{\alpha}}(t) = \left[\hat{\alpha}_1(t), \hat{\alpha}_2(t), \ldots, \hat{\alpha}_{n_\alpha}(t)\right]^{\mathrm{T}}, \qquad (3.27)$$

$$\hat{\boldsymbol{c}}(t) = \left[\hat{c}_1(t), \hat{c}_2(t), \ldots, \hat{c}_{n_c}(t)\right]^{\mathrm{T}}, \qquad (3.28)$$

$$\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t) = \left[\hat{\boldsymbol{Q}}_1(t), \hat{\boldsymbol{Q}}_2(t), \ldots, \hat{\boldsymbol{Q}}_{n_q}(t)\right]. \qquad (3.29)$$

The procedures of computing $\hat{\boldsymbol{\alpha}}(t)$, $\hat{\boldsymbol{\theta}}(t)$ and $\hat{\boldsymbol{c}}(t)$ in the F-HSG algorithm are listed as follows:

1) Let $t = 1$, $\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_{n_\alpha + n_c}/p_0$ (i.e. $\hat{\boldsymbol{\alpha}}(0) = \mathbf{1}_{n_\alpha}/p_0$ and $\hat{\boldsymbol{c}}(0) = \mathbf{1}_{n_c}/p_0$), $\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_{(n_q r) \times m}/p_0$, $\hat{\boldsymbol{u}}_{\mathrm{f}}(i) = \mathbf{1}_r/p_0$, $\hat{\boldsymbol{y}}_{\mathrm{f}}(i) = \mathbf{1}_m/p_0$, $\hat{\boldsymbol{w}}(i) = \mathbf{1}_m/p_0$, $\boldsymbol{u}(i) = 0$, $\boldsymbol{y}(i) = 0$ as $i \le 0$, $p_0 = 10^6$.

2) Collect the input-output data $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$, form $\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t)$, $\hat{\boldsymbol{\psi}}_{\mathrm{f}}(t)$, $\hat{\boldsymbol{\psi}}_{\mathrm{n}}(t)$, $\boldsymbol{\varphi}(t)$ and $\boldsymbol{\psi}_{\mathrm{s}}(t)$ by (3.19)–(3.23).

3) Compute $\hat{\boldsymbol{w}}(t)$ by (3.26), and $r_5(t)$ by (3.18).

4) Update the parameter estimate $\hat{\boldsymbol{c}}(t)$ by (3.17).

5) Compute $\hat{\boldsymbol{u}}_{\mathrm{f}}(t)$ and $\hat{\boldsymbol{y}}_{\mathrm{f}}(t)$ by (3.24) and (3.25), respectively.

6) Compute $r_3(t)$ and $r_4(t)$ by (3.14) and (3.16), respectively.

7) Update the parameter estimates $\hat{\boldsymbol{\alpha}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ by (3.13) and (3.15), respectively.

8) Increase t by 1 and go to Step 2.

4   Example

Consider a two-input two-output controlled autoregressive systems with autoregressive noises,

$$\alpha(z)\boldsymbol{y}(t) = \boldsymbol{Q}(z)\boldsymbol{u}(t) + \frac{1}{C(z)}\boldsymbol{v}(t),$$

$$\boldsymbol{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}, \quad \boldsymbol{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}, \quad \boldsymbol{v}(t) = \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix},$$

$$\alpha(z) = 1 + 0.24z^{-1} + 0.62z^{-2}, \quad C(z) = 1 - 0.10z^{-1},$$

$$\boldsymbol{Q}(z) = \begin{bmatrix} 1.20 & -0.60 \\ 0.66 & -1.05 \end{bmatrix} z^{-1}.$$

The inputs $\{u_1(t)\}$ and $\{u_2(t)\}$ are taken as two random 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 $\sigma_1^2 = \sigma_2^2 = 0.50^2$. Using the HGSG algorithm and the F-HSG algorithm to estimate the parameters of this example system, the parameter estimates and the estimation errors

$$\delta := \sqrt{\frac{\|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}\|^2 + \|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2}{\|\boldsymbol{\vartheta}\|^2 + \|\boldsymbol{\theta}\|^2}} \times 100\%$$

$$= \sqrt{\frac{\|\hat{\boldsymbol{\alpha}}(t) - \boldsymbol{\alpha}\|^2 + \|\hat{\boldsymbol{c}}(t) - \boldsymbol{c}\|^2 + \|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2}{\|\boldsymbol{\alpha}\|^2 + \|\boldsymbol{c}\|^2 + \|\boldsymbol{\theta}\|^2}} \times 100\%$$

are shown in Table 1 and Figure 2.

From Table 1 and Figure 2, we can draw the following conclusions:

Table 1. The F-HSG and HGSG estimates and their errors ($\sigma^2 = 0.50^2$).

| Algorithms	t	$\hat{\alpha}_1(t)$	$\hat{\alpha}_2(t)$	$\hat{Q}_1(1,1)$	$\hat{Q}_1(1,2)$	$\hat{Q}_1(2,1)$	$\hat{Q}_1(2,2)$	$\hat{c}_1(t)$	δ (%)
F-HSG	100	0.30662	0.58555	1.21531	−0.61010	0.63106	−1.04450	0.05751	9.13669
	200	0.29706	0.59695	1.19650	−0.60943	0.62966	−1.05899	−0.00974	5.86469
	500	0.26722	0.61028	1.21024	−0.62429	0.65396	−1.05652	−0.04840	3.35702
	1000	0.23590	0.61142	1.20619	−0.61947	0.65490	−1.03340	−0.09153	1.52456
HGSG	100	0.25794	0.66565	1.11525	−0.55899	0.54916	−0.93779	0.05788	12.69383
	200	0.26122	0.65567	1.12647	−0.57267	0.57009	−0.98298	0.05310	10.76113
	500	0.25623	0.64438	1.16228	−0.60280	0.61349	−1.01063	0.04959	8.65297
	1000	0.25003	0.63348	1.17669	−0.60784	0.62715	−1.00527	0.04407	8.07667
True values		0.24000	0.62000	1.20000	−0.60000	0.66000	−1.05000	−0.10000	
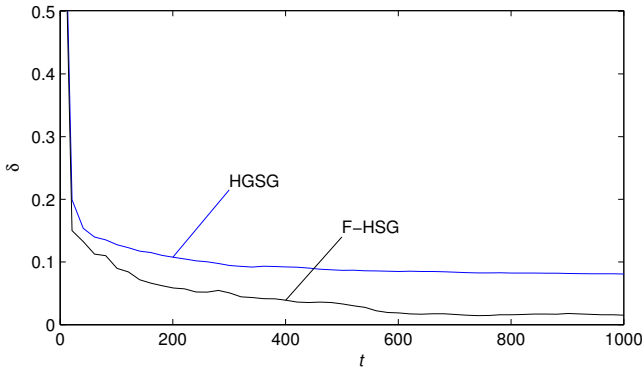
Figure 2. The estimation errors δ versus t ($\sigma^2 = 0.50^2$).

- The parameter estimation errors generally decrease with the data length t increasing, and parameter estimates are closer to their true values in the case of the F-HSG algorithm.

- The parameter estimates given by the F-HSG algorithm converge faster to their true values compared with the HGSG algorithm.

5 Conclusions

By filtering the input-output data of a multivariable controlled autoregressive system with autoregressive noise using the estimated noise model, this paper develops a filtering based hierarchical stochastic gradient algorithm to estimate the system parameters of multivariable CARAR-like systems by using the negative search and the hierarchical identification principle. The proposed algorithm can be extended to other type of linear multivariable systems or nonlinear multivariable systems.

References

[1] F. Ding. Coupled-least-squares identification for multivariable systems. *IET Control Theory and Applications*, **7**(1):68–79, 2013. http://dx.doi.org/10.1049/iet-cta.2012.0171.

[2] F. Ding. Hierarchical multi-innovation stochastic gradient algorithm for Hammerstein nonlinear system modeling. *Appl. Math. Model.*, **37**, 2013. http://dx.doi.org/10.1016/j.apm.2012.04.039.

[3] F. Ding. *System Identification – New Theory and Methods*. Science Press, Beijing, 2013.

[4] F. Ding and T. Chen. Hierarchical gradient-based identification of multivariable discrete-time systems. *Automatica*, **41**(2):315–325, 2005. http://dx.doi.org/10.1016/j.automatica.2004.10.010.

[5] F. Ding and T. Chen. Hierarchical least squares identification methods for multivariable systems. *IEEE Trans. Automat. Control*, **50**(3):397–402, 2005. http://dx.doi.org/10.1109/TAC.2005.843856.

[6] F. Ding, X.G. Liu and J. Chu. Gradient-based and least-squares-based iterative algorithms for hammerstein systems using the hierarchical identification principle. *IET Control Theory and Applications*, **7**, 2013. http://dx.doi.org/10.1049/iet-cta.2012.0313.

[7] F. Ding, Y.J. Liu and B. Bao. Gradient based and least squares based iterative estimation algorithms for multi-input multi-output systems. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, **226**(1):43–55, 2012. http://dx.doi.org/10.1177/0959651811409491.

[8] M. Emans, Z. Žunič, B. Basara and S. Frolov. A novel SIMPLE—Based pressure-enthalpy coupling scheme for engine flow problems. *Math. Model. Anal.*, **17**(1):1–20, 2012. http://dx.doi.org/10.3846/13926292.2012.643504.

[9] M. Galanin, A. Lototsky and A. Rodin. The mathematical modelling of liner movement in a magnetic compressor: elastic, liquid and plastic liner models comparison. *Math. Model. Anal.*, **17**(1):31–46, 2012. http://dx.doi.org/10.3846/13926292.2012.644406.

[10] G.C. Goodwin and K.S. Sin. *Adaptive filtering prediction and control*. Prentice–Hall, Englewood Cliffs, NJ, 1984.

[11] A. Krylovas and R. Kriauzienė. Asymptotical analysis of some coupled nonlinear wave equations. *Math. Model. Anal.*, **16**(1):97–108, 2011. http://dx.doi.org/10.3846/13926292.2011.560618.

[12] W. Li and X.L. Tian. Fault detection in discrete dynamic systems with uncertainty based on interval optimization. *Math. Model. Anal.*, **16**(4):549–557, 2011. http://dx.doi.org/10.3846/13926292.2011.628418.

[13] X.G. Liu, C.Y. Wang, L. Cong and F. Ding. Adaptive generalised predictive control of high purity internal thermally coupled distillation column. *Can. J. Chem. Eng.*, **90**(2):420–428, 2012. http://dx.doi.org/10.1002/cjce.20539.

[14] X.G. Liu, Y.X. Zhou, L. Cong and F. Ding. High-purity control of internal thermally coupled distillation columns based on nonlinear wave model. *Journal of Process Control*, **21**(6):920–926, 2011. http://dx.doi.org/10.1016/j.jprocont.2011.03.002.

[15] Y.J. Liu, J. Sheng and R.F. Ding. Convergence of stochastic estimation gradient algorithm for multivariable ARX-like systems. *Comput. Math. Appl.*, **59**(8):2615–2627, 2010. http://dx.doi.org/10.1016/j.camwa.2010.01.030.

[16] Y.J. Liu, Y.S. Xiao and X.L. Zhao. Multi-innovation stochastic gradient algorithm for multiple-input single-output systems using the auxiliary model. *Appl. Math. Comput.*, **215**(4):1477–1483, 2009. http://dx.doi.org/10.1016/j.amc.2009.07.012.

[17] V. Polevikov and L. Tobiska. ADI approach to the particle diffusion problem in magnetic fluids. *Math. Model. Anal.*, **16**(1):62–71, 2011. http://dx.doi.org/10.3846/13926292.2011.560195.

[18] M. Rodrigues, D. Theilliol, M. Adam-Medina and D. Sauter. A fault detection and isolation scheme for industrial systems based on multiple operating models. *Control Engineering Practice*, **16**(2):225–239, 2008. http://dx.doi.org/10.1016/j.conengprac.2006.02.020.

[19] D.Q. Wang. Least squares-based recursive and iterative estimation for output error moving average systems using data filtering. *IET Control Theory Appl.*, **5**(14):1648–1657, 2011. http://dx.doi.org/10.1049/iet-cta.2010.0416.

[20] D.Q. Wang and F. Ding. Input–output data filtering based recursive least squares identification for CARARMA systems. *Digital Signal Processing*, **20**(4):991–999, 2010. http://dx.doi.org/10.1016/j.dsp.2009.12.006.

[21] D.Q. Wang, R. Ding and X.Z. Dong. Iterative parameter estimation for a class of multivariable systems based on the hierarchical identification principle and the gradient search. *Circuits Systems Signal Process.*, **31**(6):2167–2177, 2012. http://dx.doi.org/10.1007/s00034-012-9425-y.

[22] Y. Zhang. Unbiased identification of a class of multi-input single-output systems with correlated disturbances using bias compensation methods. *Math. Comput. Model.*, **53**(9–10):1810–1819, 2011. http://dx.doi.org/10.1016/j.mcm.2010.12.059.

[23] Y. Zhang and G.M. Cui. Bias compensation methods for stochastic systems with colored noise. *Appl. Math. Model.*, **35**(4):1709–1716, 2011. http://dx.doi.org/10.1016/j.apm.2010.10.003.