

Article

A Filtering-Based Stochastic Gradient Estimation Method for Multivariate Pseudo-Linear Systems Using the Partial Coupling Concept

Ping Ma ¹, Yuan Liu ¹ and Yiyang Chen ^{2,*}

- ¹ Jiangsu Key Laboratory of Media Design and Software Technology, School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China; pingma@jiangnan.edu.cn (P.M.); lyuan1800@sina.com (Y.L.)
- ² School of Mechanical and Electrical Engineering, Soochow University, Suzhou 215137, China
- * Correspondence: yychen90@suda.edu.cn

Abstract: Solutions for enhancing parameter identification effects for multivariate equation-error systems in random interference and parameter coupling conditions are considered in this paper. For the purpose of avoiding the impact of colored noises on parameter identification precision, an appropriate filter is utilized to process the autoregressive moving average noise. Then, the filtered system is transformed into a number of sub-identification models based on system output dimensions. Founded on negative gradient search, a new multivariate filtering algorithm employing a partial coupling approach is proposed, and a conventional gradient algorithm is derived for comparison. Parameter identification for multivariate equation-error systems has a high estimation accuracy and an efficient calculation speed with the application of the partial coupling approach and the data filtering method. Two simulations are performed to reveal the proposed method's effectiveness.

Keywords: gradient search; multivariate system; partial coupling; data filtering; parameter identification



Citation: Ma, P.; Liu, Y.; Chen, Y. A Filtering-Based Stochastic Gradient Estimation Method for Multivariate Pseudo-Linear Systems Using the Partial Coupling Concept. *Processes* **2023**, *11*, 2700. <https://doi.org/10.3390/pr11092700>

Academic Editors: Weicun Zhang, Hassan el Fadil and Quanmin Zhu

Received: 25 August 2023
Revised: 6 September 2023
Accepted: 7 September 2023
Published: 9 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The foundation of industrial automatic production and intelligent control is a precise model of the production processes [1]. With the expansion of production scale, multivariate systems have been widely used in production processes [2–4]. Parameter estimation for multivariate systems has taken a considerable role in system identification and has attracted much attention from researchers in recent decades. Multivariate systems are more difficult to identify than scalar systems because they have more unknown parameters, are accompanied by more complex random interferences, and the parameters of some channels are coupled [5–7]. Good identification results are often not achieved if the identification approaches for scalar systems are applied to multivariate systems without modification. Some improved methods for the identification of multivariate systems have been researched recently [8–10]. For instance, Mari et al. combined the Schur restabilization technique and a covariance fitting algorithm to propose a parameter estimation method for finite dimensional multivariate linear stochastic systems [11]. Luo and Manikas proposed an iterative method and a nonlinear optimization algorithm for suppressing the mutual target interference in the multitarget parameter estimation [12]. Zhang et al. identified the parameters of multivariate uncertain regression model with a maximum likelihood identification algorithm [13]. Oigard et al. researched an expectation maximization algorithm for heavy-tailed processes with a multivariate normal inverse Gaussian distribution, which has fast and accurate parameter identification effectiveness [14].

Although the identification methods for multivariate systems are gradually enriched, researchers have been devoted to finding methods with faster identification efficiency and higher identification accuracy for multivariate systems. In terms of improving estimation

efficiency, in addition to the decomposition identification method, the coupling identification approach also can effectively reduce the amount of identification computation [15–17]. The basis of the coupling identification approach is the transformation of multivariate systems into some identification subsystems, then coupling the identification results of each subsystem to make their results correlated [18–20]. There are some studies on coupling identification methods for multivariate systems. Ding researched parameter identification issues for non-uniformly sampled systems and proposed a partially coupled algorithm based upon the stochastic gradient method. A simulation in the paper revealed that the new algorithm requires less calculation than the standard stochastic gradient algorithm [21]. Zhou designed a nonlinear partially coupled parameter identification algorithm for multivariate radial basis function-based hybrid models inspired by the coupling concept which reduces the amount of calculation by dealing with the associated items brought by model decomposition [22]. Huang et al. provided a coupled probability representation regarding model coupling in feature-based image source identification which improved the identification accuracy significantly [23]. Wang solved parameter identification problems of nonlinear multivariate systems and developed a coupled gradient method by introducing the coupling idea, which can realize subsystem-coupled computation [24].

The data filtering approach in parameter estimation is the improvement of the parameter identification precision by modifying the structure of system interference noises through an appropriate filter without changing the system's input–output relationship [25–27]. The data filtering approach has been applied to scalar system identification in some studies. Ji and Jiang utilized a data filter to process collected data to deal with the disturbance of colored noise on identification precision for generalized time-varying systems [28]. Imani studied a maximum-likelihood parameter identification method for partially observed Boolean dynamical systems by using a Boolean Kalman filter [29]. Zhang developed a filtering hierarchical maximum likelihood iterative algorithm for nonlinear systems by applying the data filtering approach and multi-innovation identification method, which obtains highly precise parameter estimates and tracks time-varying parameters well [30]. Chen et al. proposed a multi-step-length gradient iterative algorithm for ARX models with the application of a modified Kalman filter. The Kalman filter was designed to enhance unmeasurable output estimates, which improved parameter identification accuracy [31]. Li and Liu addressed parameter identification problems in bilinear systems and presented iterative methods with high estimation accuracies by utilizing the particle filtering approach [32].

The least squares estimation algorithm, gradient estimation algorithm, least mean square estimation algorithm, and stochastic approximation estimation algorithm are all classical identification methods in the field of system identification. The least squares method is a basic parameter estimation method which can be used for dynamic system identification as well as for static system parameter fitting [33,34]. The gradient identification method is a search for parameter estimates along the direction of the negative gradient of the criterion functions [35–37]. Compared with the least squares method, the gradient identification algorithm has less computational complexity because it does not involve the covariance matrix. By extending the gradient identification algorithm and combining it with other methods, estimation algorithms with high identification performances can be obtained. Zhang and Ding proposed an optimal adaptive filtering algorithm for filter design by combining the data filtering approach with the gradient method [38]. Roman et al. derived a gradient descent method for identifying parameters of a linear wave equation from experimental boundary data [39]. Chen et al. identified parameters of time-delay rational state–space systems and presented two improved gradient descent algorithms by utilizing an intelligent search method and a momentum method, which had faster convergence speeds and higher computational efficiencies [40]. Kulikova researched adaptive filtering methods based on the gradient algorithm for identifying unknown parameters of pairwise linear Gaussian systems [41].

The data filtering approach can improve the parameter identification precision for multivariate systems by transforming colored noises with complex structures into white

noises with simple structures [42–44]. At the same time, the coupling identification method can effectively speed up identification and the gradient identification approach can quickly search for the optimal estimates [45–47]. Therefore, motivated by the significant advantages of these three methods, this paper combines the data filtering approach and the coupling identification method based upon gradient search to recognize parameters of multivariate equation-error systems. The introduction of the data filtering approach overcomes the influence of colored noise on identification precision. The use of the coupling identification method reduces the computation of the identification algorithm. The main highlights of this paper are summarized as follows.

- (1) A filter is used to transform the autoregressive moving average noise of multivariate pseudo-linear systems into white noise by applying the data filtering approach. The filtered system is converted into a number of subsystem identification models based upon system the output dimensions according to the coupling identification method.
- (2) A filtering-based multivariate gradient algorithm employing the partial coupling concept for multivariate pseudo-linear systems is proposed. Additionally, a conventional multivariate gradient algorithm is derived for comparison. The proposed algorithm has higher identification precision and faster computational efficiency than the conventional algorithm.

The structure of this paper is as follows. The multivariate pseudo-linear system is presented and the system identification obstacles are analyzed in Section 2. A new gradient algorithm based upon the coupling identification approach and the data filtering method is proposed in Section 3. Section 4 derives a conventional multivariate gradient algorithm. Convergence of the proposed method is discussed in Section 5. In Section 6, two simulations are performed to reveal the effectiveness of the proposed methods. Finally, Section 7 provides some conclusions of this paper.

2. Problem Description

At the beginning, we provide some notation to make the paper concise and clear. I_m is an identity matrix of size $m \times m$. $\mathbf{1}_{m \times n}$ denotes a matrix of size $m \times n$ whose elements are 1. The norm of a matrix A is defined by $\|A\|^2 := \text{tr}[AA^T]$, and the superscript T represents the matrix/vector transpose. The symbol \otimes stands for the Kronecker product, such as $X := [x_{ij}] \in \mathbb{R}^{m \times n}$, $Y := [y_{ij}] \in \mathbb{R}^{p \times q}$, $X \otimes Y = [x_{ij}y_{kl}] \in \mathbb{R}^{(mp) \times (nq)}$; generally, $X \otimes Y \neq Y \otimes X$. $\text{col}[B]$ denotes a vector consisting of all columns of matrix B arranged in order, that is, $B := [b_1, b_2, \dots, b_n] \in \mathbb{R}^{m \times n}$, $b_i \in \mathbb{R}^m$ ($i = 1, 2, \dots, n$), $\text{col}[B] := [b_1^T, b_2^T, \dots, b_n^T]^T \in \mathbb{R}^{mn}$. $\hat{H}(k)$ is the estimate of H at time k .

According to the type of colored noise, the multivariate pseudo-linear system can be divided into different types. In this paper, we consider systems where the noise is of the autoregressive moving average type. The system is widely present in industrial processes and its structure is described as

$$\mathbf{y}(k) = \Phi(k)\theta + \frac{D(z)}{c(z)}\mathbf{v}(k), \quad (1)$$

where $\mathbf{y}(k) := [y_1(k), y_2(k), \dots, y_m(k)]^T \in \mathbb{R}^m$ is the system output vector which can be measured, $\Phi(k) \in \mathbb{R}^{m \times n}$ is the system information matrix which is formed from system input–output data, $\theta \in \mathbb{R}^n$ is the system parameter vector which is unknown and to be identified, $\mathbf{v}(k) := [v_1(k), v_2(k), \dots, v_m(k)]^T \in \mathbb{R}^m$ is a white noise process with zero mean, $c(z) \in \mathbb{R}$ is a polynomial in the unit backward shift operator $[z^{-1}\mathbf{y}(k) = \mathbf{y}(k-1)]$, and $D(z) \in \mathbb{R}^{m \times m}$ is a polynomial matrices:

$$\begin{aligned} c(z) &:= 1 + c_1z^{-1} + c_2z^{-2} + \dots + c_{n_c}z^{-n_c}, c_i \in \mathbb{R}, \\ D(z) &:= I_m + D_1z^{-1} + D_2z^{-2} + \dots + D_{n_d}z^{-n_d}, D_i \in \mathbb{R}^{m \times m}. \end{aligned}$$

Define the noise model as,

$$w(k) := \frac{D(z)}{c(z)}v(k) \in \mathbb{R}^m. \tag{2}$$

In general, suppose that orders $m, n, n_c,$ and n_d of the system are known, and $y(k) = \mathbf{0}, \Phi(k) = \mathbf{0}$ and $v(k) = \mathbf{0}$ for $k \leq 0$.

Define the parameter vector η , the parameter matrix κ , the information vector $\psi(k)$, and the information matrix $\Psi(k)$ as

$$\eta := [\theta^T, c_1, c_2, \dots, c_{n_c}]^T \in \mathbb{R}^{n+n_c}, \tag{3}$$

$$\kappa^T := [D_1, D_2, \dots, D_{n_d}] \in \mathbb{R}^{m \times mn_d}, \tag{4}$$

$$\psi(k) := [v^T(k-1), v^T(k-2), \dots, v^T(k-n_d)]^T \in \mathbb{R}^{mn_d}, \tag{5}$$

$$\Psi(k) := [\Phi(k), -w(k-1), -w(k-2), \dots, -w(k-n_c)] \in \mathbb{R}^{m \times (n+n_c)}. \tag{6}$$

Based on Equation (2), we obtain

$$\begin{aligned} w(k) &= [1 - c(z)]w(k) + [D(z) - I_m]v(k) + v(k) \\ &= (-c_1z^{-1} - c_2z^{-2} - \dots - c_{n_c}z^{-n_c})w(k) \\ &\quad + (D_1z^{-1} + D_2z^{-2} + \dots + D_{n_d}z^{-n_d})v(k) + v(k) \\ &= -\sum_{i=1}^{n_c} c_i w(k-i) + \sum_{j=1}^{n_d} D_j v(k-j) + v(k). \end{aligned} \tag{7}$$

The identification model of the system in (1) is represented as

$$y(k) = \Phi(k)\theta + w(k) \tag{8}$$

$$\begin{aligned} &= \Phi(k)\theta - \sum_{i=1}^{n_c} c_i w(k-i) + \sum_{j=1}^{n_d} D_j v(k-j) + v(k) \\ &= \Psi(k)\eta + \kappa^T \psi(k) + v(k). \end{aligned} \tag{9}$$

Uniting the information matrix $\Psi(k)$ with the information vector $\psi(k)$, and the parameter vector η with the parameter matrix κ , the new information matrix $\Omega(k)$ and new parameter vector ϑ are

$$\Omega(k) := [\Psi(k), \psi^T(k) \otimes I_m] \in \mathbb{R}^{m \times n_0}, \quad n_0 := n + n_c + m^2 n_d, \tag{10}$$

$$\vartheta := \begin{bmatrix} \eta \\ \text{col}[\kappa] \end{bmatrix} \in \mathbb{R}^{n_0}. \tag{11}$$

Equation (9) is changed into

$$y(k) = \Omega(k)\vartheta + v(k). \tag{12}$$

The goal is to find effective identification methods to estimate unmeasurable parameters $\theta, c_i,$ and D_i which are in ϑ . The analysis shows that if the identification model in (12) is performed directly, superfluous calculations will be generated in the unknown parameter estimation processes because the Kronecker product calculation produces substantial zero elements to the information matrix $\Omega(k)$. With a view to enhance the identification performance for the system in (1), it is necessary to explore another efficient identification method.

3. The Filtering-Based Multivariate Partially Coupled Gradient Algorithm

By analyzing System (1), the existence of the noise reduces the parameter identification precision. For overcoming the adverse effects of disturbances, the data filtering method is adopted to convert colored noise into white noise. Setting $c(z)$ as the filter for System (1)

is an appropriate solution of the problem. First of all, multiply both sides of Equation (1) by $c(z)$:

$$c(z)\mathbf{y}(k) = c(z)\mathbf{\Phi}(k)\boldsymbol{\theta} + \mathbf{D}(z)\mathbf{v}(k). \quad (13)$$

Define the filtered output vector $\mathbf{y}_f(k)$ and the filtered information matrix $\mathbf{\Phi}_f(k)$ as

$$\begin{aligned} \mathbf{y}_f(k) &:= c(z)\mathbf{y}(k) \in \mathbb{R}^m, \\ \mathbf{\Phi}_f(k) &:= c(z)\mathbf{\Phi}(k) \in \mathbb{R}^{m \times n}. \end{aligned}$$

Then, Equation (13) can be rewritten as

$$\begin{aligned} \mathbf{y}_f(k) &= \mathbf{\Phi}_f(k)\boldsymbol{\theta} + \mathbf{D}(z)\mathbf{v}(k) \\ &= \mathbf{\Phi}_f(k)\boldsymbol{\theta} + \mathbf{D}_1\mathbf{v}(k-1) + \mathbf{D}_2\mathbf{v}(k-2) + \dots + \mathbf{D}_{n_d}\mathbf{v}(k-n_d) + \mathbf{v}(k) \\ &= \mathbf{\Phi}_f(k)\boldsymbol{\theta} + \boldsymbol{\kappa}^\top\boldsymbol{\psi}(k) + \mathbf{v}(k). \end{aligned} \quad (14)$$

Let $\boldsymbol{\kappa}_i^\top \in \mathbb{R}^{1 \times mn_d}$ be the i th row of the parameter matrix $\boldsymbol{\kappa}^\top$, $y_{fi}(k) \in \mathbb{R}$ be the i th row of the filtered output vector $\mathbf{y}_f(k)$, and $\boldsymbol{\Phi}_{fi}(k) \in \mathbb{R}^{1 \times n}$ be the i th row of the filtered information matrix $\mathbf{\Phi}_f(k)$, that is

$$\begin{aligned} \boldsymbol{\kappa}_i^\top &:= [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_{n_d}] := [\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2, \dots, \boldsymbol{\kappa}_m]^\top, \\ \mathbf{y}_f(k) &:= [y_{f1}(k), y_{f2}(k), \dots, y_{fm}(k)]^\top, \\ \boldsymbol{\Phi}_f(k) &:= [\boldsymbol{\Phi}_{f1}^\top(k), \boldsymbol{\Phi}_{f2}^\top(k), \dots, \boldsymbol{\Phi}_{fm}^\top(k)]^\top. \end{aligned}$$

Transform Equation (14) into m sub-identification models:

$$\begin{bmatrix} y_{f1}(k) \\ y_{f2}(k) \\ \vdots \\ y_{fm}(k) \end{bmatrix} := \begin{bmatrix} \boldsymbol{\Phi}_{f1}(k) \\ \boldsymbol{\Phi}_{f2}(k) \\ \vdots \\ \boldsymbol{\Phi}_{fm}(k) \end{bmatrix} \boldsymbol{\theta} + \begin{bmatrix} \boldsymbol{\kappa}_1^\top \\ \boldsymbol{\kappa}_2^\top \\ \vdots \\ \boldsymbol{\kappa}_m^\top \end{bmatrix} \boldsymbol{\psi}(k) + \begin{bmatrix} v_1(k) \\ v_2(k) \\ \vdots \\ v_m(k) \end{bmatrix}, \quad i = 1, 2, \dots, m. \quad (15)$$

Equation (14) is described by

$$y_{fi}(k) = \boldsymbol{\Phi}_{fi}(k)\boldsymbol{\theta} + \boldsymbol{\kappa}_i^\top\boldsymbol{\psi}(k) + v_i(k), \quad i = 1, 2, \dots, m. \quad (16)$$

In Equation (16), vectors $\boldsymbol{\theta}$ and $\boldsymbol{\psi}(k)$ are common in each subsystem, which is in line with the characteristics of the partially coupled-type identification model. Next, parameter estimation algorithm employing partial coupling concept for the identification model in (16) is derived in detail.

Define a gradient criterion function for the new identification model in (16) as

$$J_1(\boldsymbol{\theta}, \boldsymbol{\kappa}_i) := \|y_{fi}(k) - \boldsymbol{\Phi}_{fi}(k)\boldsymbol{\theta} - \boldsymbol{\kappa}_i^\top\boldsymbol{\psi}(k)\|^2, \quad i = 1, 2, \dots, m.$$

Minimizing $J_1(\boldsymbol{\theta}, \boldsymbol{\kappa}_i)$ based upon the gradient search, the gradient relationships are

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \frac{\boldsymbol{\Phi}_{fi}^\top(k)}{r_{\theta,i}(k)} [y_{fi}(k) - \boldsymbol{\Phi}_{fi}(k)\hat{\boldsymbol{\theta}}(k-1) - \hat{\boldsymbol{\kappa}}_i^\top(k-1)\boldsymbol{\psi}(k)], \quad (17)$$

$$r_{\theta,i}(k) = r_{\theta,i}(k-1) + \|\boldsymbol{\Phi}_{fi}(k)\|^2, \quad r_{\theta,i}(0) = 1, \quad (18)$$

$$\hat{\boldsymbol{\kappa}}_i(k) = \hat{\boldsymbol{\kappa}}_i(k-1) + \frac{\boldsymbol{\psi}(k)}{r_{\kappa,i}(k)} [y_{fi}(k) - \boldsymbol{\Phi}_{fi}(k)\hat{\boldsymbol{\theta}}(k-1) - \hat{\boldsymbol{\kappa}}_i^\top(k-1)\boldsymbol{\psi}(k)], \quad (19)$$

$$r_{\kappa,i}(k) = r_{\kappa,i}(k-1) + \|\boldsymbol{\psi}(k)\|^2, \quad r_{\kappa,i}(0) = 1. \quad (20)$$

However, estimates $\hat{\theta}(k)$ and $\hat{\kappa}_i(k)$ in (17)–(20) cannot be computed because $y_{fi}(k)$, $\Phi_{fi}(k)$, and $\psi(k)$ involve the unmeasurable terms c_i and $v(k)$. If we define,

$$\varphi_y(k) := [\mathbf{y}(k-1), \mathbf{y}(k-2), \dots, \mathbf{y}(k-n_c)] \in \mathbb{R}^{m \times n_c}, \quad (21)$$

$$\boldsymbol{\tau} := [c_1, c_2, \dots, c_{n_c}]^T \in \mathbb{R}^{n_c}. \quad (22)$$

Then, $\mathbf{y}_f(k)$ and $\Phi_f(k)$ are represented by

$$\begin{aligned} \mathbf{y}_f(k) &= c(z)\mathbf{y}(k) \\ &= \mathbf{y}(k) + c_1\mathbf{y}(k-1) + c_2\mathbf{y}(k-2) + \dots + c_{n_c}\mathbf{y}(k-n_c) \\ &= \mathbf{y}(k) + \varphi_y(k)\boldsymbol{\tau}, \end{aligned} \quad (23)$$

$$\begin{aligned} \Phi_f(k) &= c(z)\Phi(k) \\ &= \Phi(k) + c_1\Phi(k-1) + c_2\Phi(k-2) + \dots + c_{n_c}\Phi(k-n_c). \end{aligned} \quad (24)$$

However, $\mathbf{y}_f(k)$ and $\Phi_f(k)$ still cannot be calculated because c_i is unknown. The auxiliary model identification method [48] is a classical method which can solve some system identification problems with unmeasured variables. The essential thought behind this is to replace the unknown variable with the output of an auxiliary model. The problem is that parameter estimation cannot be calculated in the algorithm, which can be solved by replacing the unknown variable with its estimate when its value cannot be obtained. Here, utilizing the auxiliary model identification method, according to Equation (22), replacing c_i with estimates $\hat{c}_i(k)$, the estimate $\hat{\boldsymbol{\tau}}(k)$ is formed by

$$\hat{\boldsymbol{\tau}}(k) = [\hat{c}_1(k), \hat{c}_2(k), \dots, \hat{c}_{n_c}(k)]^T.$$

After that, replacing unknown parameters c_i and $\boldsymbol{\tau}$ with estimates $\hat{c}_i(k)$ and $\hat{\boldsymbol{\tau}}(k)$ in (23) and (24), estimates $\hat{\mathbf{y}}_f(k)$ and $\hat{\Phi}_f(k)$ are calculated by

$$\begin{aligned} \hat{\mathbf{y}}_f(k) &= \mathbf{y}(k) + \varphi_y(k)\hat{\boldsymbol{\tau}}(k) \\ &= [\hat{y}_{f1}(k), \hat{y}_{f2}(k), \dots, \hat{y}_{fm}(k)]^T, \end{aligned} \quad (25)$$

$$\begin{aligned} \hat{\Phi}_f(k) &= \Phi(k) + \hat{c}_1(k)\Phi(k-1) + \hat{c}_2(k)\Phi(k-2) + \dots + \hat{c}_{n_c}(k)\Phi(k-n_c) \\ &= [\hat{\Phi}_{f1}^T(k), \hat{\Phi}_{f2}^T(k), \dots, \hat{\Phi}_{fm}^T(k)]^T. \end{aligned} \quad (26)$$

Meanwhile, according to Equation (5), the estimate $\hat{\boldsymbol{\psi}}(k)$ is formed by

$$\hat{\boldsymbol{\psi}}(k) = [\hat{\boldsymbol{\psi}}^T(k-1), \hat{\boldsymbol{\psi}}^T(k-2), \dots, \hat{\boldsymbol{\psi}}^T(k-n_d)]^T. \quad (27)$$

Define the noise information matrix:

$$\boldsymbol{\chi}(k) := [\mathbf{w}(k-1), \mathbf{w}(k-2), \dots, \mathbf{w}(k-n_c)] \in \mathbb{R}^{m \times n_c}. \quad (28)$$

Equation (7) can be written as

$$\begin{aligned} \mathbf{w}(k) &= -\sum_{i=1}^{n_c} c_i \mathbf{w}(k-i) + \sum_{j=1}^{n_d} D_j \mathbf{v}(k-j) + \mathbf{v}(k) \\ &= \boldsymbol{\chi}(k)\boldsymbol{\tau} + \boldsymbol{\kappa}^T \boldsymbol{\psi}(k) + \mathbf{v}(k). \end{aligned}$$

Define an intermediate vector $\mathbf{w}_n(k) := \mathbf{w}(k) - \boldsymbol{\kappa}^T \boldsymbol{\psi}(k) \in \mathbb{R}^m$. Thus, the noise model is rewritten as

$$\mathbf{w}_n(k) := \boldsymbol{\chi}(k)\boldsymbol{\tau} + \mathbf{v}(k). \quad (29)$$

In order to obtain the value of estimate $\hat{\tau}(k)$, define another gradient criterion function for the noise model in Equation (29) as

$$J_2(\tau) := \|\mathbf{w}_n(k) - \chi(k)\tau\|^2.$$

Minimizing $J_2(\tau)$ based upon the gradient search, the gradient relationship is

$$\hat{\tau}(k) = \hat{\tau}(k-1) + \frac{\chi^T(k)}{r_\tau(k)} [\mathbf{w}_n(k) - \chi(k)\hat{\tau}(k-1)], \quad (30)$$

$$r_\tau(k) = r_\tau(k-1) + \|\chi(k)\|^2, \quad r_\tau(0) = 1. \quad (31)$$

Obviously, the parameter vector estimate $\hat{\tau}(k)$ cannot be calculated because $\mathbf{w}_n(k)$ and $\chi(k)$ are unknown. Replacing them with estimates $\hat{\mathbf{w}}_n(k)$ and $\hat{\chi}(k)$ can solve this problem. We have

$$\hat{\chi}(k) = [\hat{\mathbf{w}}(k-1), \hat{\mathbf{w}}(k-2), \dots, \hat{\mathbf{w}}(k-n_c)], \quad (32)$$

$$\hat{\mathbf{w}}_n(k) = \hat{\mathbf{w}}(k-1) - \hat{\kappa}^T(k-1)\hat{\psi}(k). \quad (33)$$

According to Equation (8), substitute the estimate $\hat{\theta}(k)$ for the unmeasurable term θ . Then, the estimate $\hat{\mathbf{w}}(k)$ is calculated by

$$\hat{\mathbf{w}}(k) = \mathbf{y}(k) - \Phi(k)\hat{\theta}(k). \quad (34)$$

Similarly, according to Equation (14), replacing unmeasurable terms $\Phi_f(k)$, θ , and κ with their estimates $\hat{\Phi}_f(k)$, $\hat{\theta}(k)$, and $\hat{\kappa}(k)$, the estimate $\hat{v}(k)$ is calculated by

$$\hat{v}(k) = \hat{y}_f(k) - \hat{\Phi}_f(k)\hat{\theta}(k) - \hat{\kappa}^T(k)\hat{\psi}(k). \quad (35)$$

There are superfluous estimates in algorithm (17)–(20) because θ is repeatedly computed m times. To reduce the excess computation, $\hat{\theta}_i$ is used instead of $\hat{\theta}$ in (17) and (19). Meanwhile, substitute estimates $\hat{\Phi}_{fi}^T(k)$, $\hat{y}_{fi}(k)$ and $\hat{\psi}(k)$ for unknown terms $\Phi_{fi}^T(k)$, $y_{fi}(k)$, and $\psi(k)$; then, the new algorithm is

$$\hat{\theta}_i(k) = \hat{\theta}_i(k-1) + \frac{\hat{\Phi}_{fi}^T(k)}{r_{\theta,i}(k)} [\hat{y}_{fi}(k) - \hat{\Phi}_{fi}(k)\hat{\theta}_i(k-1) - \hat{\kappa}_i^T(k-1)\hat{\psi}(k)], \quad (36)$$

$$r_{\theta,i}(k) = r_{\theta,i}(k-1) + \|\hat{\Phi}_{fi}(k)\|^2, \quad r_{\theta,i}(0) = 1, \quad (37)$$

$$\hat{\kappa}_i(k) = \hat{\kappa}_i(k-1) + \frac{\hat{\psi}(k)}{r_{\kappa,i}(k)} [\hat{y}_{fi}(k) - \hat{\Phi}_{fi}(k)\hat{\theta}_i(k-1) - \hat{\kappa}_i^T(k-1)\hat{\psi}(k)], \quad (38)$$

$$r_{\kappa,i}(k) = r_{\kappa,i}(k-1) + \|\hat{\psi}(k)\|^2, \quad r_{\kappa,i}(0) = 1. \quad (39)$$

In recursive algorithms, the estimated values of parameters tend to true values infinitely with the data length increasing. As is well known, the estimate $\hat{\theta}_{i-1}(k)$ of the $(i-1)$ th subsystem at time k is closer to the true value θ than the estimate $\hat{\theta}_i(k-1)$ of the i th subsystem at time $k-1$. Accordingly, in Algorithm (36)–(39), substitute $\hat{\theta}_{i-1}(k)$ for $\hat{\theta}_i(k-1)$ on the right-hand side of Equation (36), and substitute $\hat{\theta}_m(k-1)$ for $\hat{\theta}_1(k-1)$ in Equation (36) when $i=1$. To conclude, the filtering-based multivariate partially coupled generalized extended stochastic gradient (F-M-PC-GESG) algorithm is as follows.

$$\hat{\theta}_1(k) = \hat{\theta}_m(k-1) + \frac{\hat{\Phi}_{f1}^T(k)}{r_{\theta,1}(k)} [\hat{y}_{f1}(k) - \hat{\Phi}_{f1}(k)\hat{\theta}_m(k-1) - \hat{\kappa}_1^T(k-1)\hat{\psi}(k)], \quad (40)$$

$$r_{\theta,1}(k) = r_{\theta,1}(k-1) + \|\hat{\Phi}_{f1}(k)\|^2, \quad r_{\theta,1}(0) = 1, \quad (41)$$

$$\hat{\kappa}_1(k) = \hat{\kappa}_1(k-1) + \frac{\hat{\psi}(k)}{r_{\kappa,1}(k)} [\hat{y}_{f1}(k) - \hat{\Phi}_{f1}(k)\hat{\theta}_m(k-1) - \hat{\kappa}_1^T(k-1)\hat{\psi}(k)], \quad (42)$$

$$r_{\kappa,1}(k) = r_{\kappa,1}(k-1) + \|\hat{\psi}(k)\|^2, \quad r_{\kappa,1}(0) = 1, \quad (43)$$

$$\hat{\theta}_i(k) = \hat{\theta}_{i-1}(k) + \frac{\hat{\Phi}_{fi}^T(k)}{r_{\theta,i}(k)} [\hat{y}_{fi}(k) - \hat{\Phi}_{fi}(k)\hat{\theta}_{i-1}(k) - \hat{\kappa}_i^T(k-1)\hat{\psi}(k)], \quad (44)$$

$$r_{\theta,i}(k) = r_{\theta,i}(k-1) + \|\hat{\Phi}_{fi}(k)\|^2, \quad r_{\theta,i}(0) = 1, \quad (45)$$

$$\hat{\kappa}_i(k) = \hat{\kappa}_i(k-1) + \frac{\hat{\psi}(k)}{r_{\kappa,i}(k)} [\hat{y}_{fi}(k) - \hat{\Phi}_{fi}(k)\hat{\theta}_{i-1}(k) - \hat{\kappa}_i^T(k-1)\hat{\psi}(k)], \quad (46)$$

$$r_{\kappa,i}(k) = r_{\kappa,i}(k-1) + \|\hat{\psi}(k)\|^2, \quad r_{\kappa,i}(0) = 1, \quad (47)$$

$$\hat{y}_f(k) = \mathbf{y}(k) + \boldsymbol{\varphi}_y(k)\hat{\tau}(k) = [\hat{y}_{f1}(k), \hat{y}_{f2}(k), \dots, \hat{y}_{fm}(k)]^T, \quad (48)$$

$$\begin{aligned} \hat{\Phi}_f(k) &= \Phi(k) + \hat{c}_1(k)\Phi(k-1) + \hat{c}_2(k)\Phi(k-2) + \dots + \hat{c}_{n_c}(k)\Phi(k-n_c) \\ &= [\hat{\Phi}_{f1}^T(k), \hat{\Phi}_{f2}^T(k), \dots, \hat{\Phi}_{fm}^T(k)]^T, \end{aligned} \quad (49)$$

$$\hat{\psi}(k) = [\hat{v}^T(k-1), \hat{v}^T(k-2), \dots, \hat{v}^T(k-n_d)]^T, \quad (50)$$

$$\boldsymbol{\varphi}_y(k) = [\mathbf{y}(k-1), \mathbf{y}(k-2), \dots, \mathbf{y}(k-n_c)], \quad (51)$$

$$\begin{aligned} \hat{\tau}(k) &= \hat{\tau}(k-1) + \frac{\hat{\chi}^T(k)}{r_\tau(k)} [\hat{w}_n(k) - \hat{\chi}(k)\hat{\tau}(k-1)] \\ &= [\hat{c}_1(k), \hat{c}_2(k), \dots, \hat{c}_{n_c}(k)]^T, \end{aligned} \quad (52)$$

$$r_\tau(k) = r_\tau(k-1) + \|\hat{\chi}(k)\|^2, \quad r_\tau(0) = 1, \quad (53)$$

$$\hat{\chi}(k) = [\hat{w}(k-1), \hat{w}(k-2), \dots, \hat{w}(k-n_c)], \quad (54)$$

$$\hat{w}_n(k) = \hat{w}(k-1) - \hat{\kappa}^T(k-1)\hat{\psi}(k), \quad (55)$$

$$\hat{w}(k) = \mathbf{y}(k) - \Phi(k)\hat{\theta}_m(k), \quad (56)$$

$$\hat{v}(k) = \hat{y}_f(k) - \hat{\Phi}_f(k)\hat{\theta}(k) - \hat{\kappa}^T(k)\hat{\psi}(k), \quad (57)$$

$$\hat{\kappa}(k) = [\hat{\kappa}_1(k), \hat{\kappa}_2(k), \dots, \hat{\kappa}_m(k)]. \quad (58)$$

The calculation steps of the F-M-PC-GESG algorithm in (40)–(58) are presented as follows.

1. Let $k = 1$, set the initial values $\hat{\theta}_m(0) = \mathbf{1}_n/p_0$, $\hat{\kappa}(0) = \mathbf{1}_{mn_d \times m}/p_0$, $\hat{\tau}(0) = \mathbf{1}_{n_c}/p_0$, $r_{\theta,i}(0) = r_{\kappa,i}(0) = r_\tau(0) = 1, i = 1, 2, \dots, m$, $\hat{w}(k-j) = \mathbf{0}, \hat{v}(k-j) = \mathbf{0}, j = 0, 1, \dots, \max[n_c, n_d]$, $p_0 = 10^6$, and set the data length K .
2. Acquire the observation data $\Phi(k)$ and $\mathbf{y}(k)$. Configure $\hat{\psi}(k)$, $\boldsymbol{\varphi}_y(k)$ and $\hat{\chi}(k)$ using (50), (51) and (54).
3. Calculate $r_\tau(k)$ and $\hat{w}_n(k)$ utilizing (53) and (55). Update $\hat{\tau}(k)$ using (52) and read $\hat{c}_i(k)$ from $\hat{\tau}(k)$.
4. Compute $\hat{y}_f(k)$ and $\hat{\Phi}_f(k)$ using (48) and (49). Read $\hat{y}_{fi}(k)$ from $\hat{y}_f(k)$ and $\hat{\Phi}_{fi}^T(k)$ form $\hat{\Phi}_f(k)$.
5. Compute the step size $r_{\theta,1}(k)$ and $r_{\kappa,1}(k)$ utilizing (41) and (43).
6. Refresh the parameter estimates $\hat{\theta}_1(k)$ and $\hat{\kappa}_1(k)$ utilizing (40) and (42).
7. Calculate $r_{\theta,i}(k)$ and $r_{\kappa,i}(k)$ utilizing (45) and (47) when $i = 2, 3, \dots, m$. Refresh the parameter estimates $\hat{\theta}_i(k)$ and $\hat{\kappa}_i(k)$ utilizing (44) and (46).
8. Construct $\hat{\kappa}(k)$ using (58). Calculate $\hat{v}(k)$ and $\hat{w}(k)$ using (57) and (56).
9. Gain k by 1 if $k < K$, and then skip to Step 2. If not, obtain the parameter estimates $\hat{\theta}(k)$, $\hat{\kappa}(k)$, and $\hat{\tau}(k)$ and quit.

By analyzing the whole calculation steps, the F-M-PC-GESG algorithm uses the method of interactive estimation. That is, the value of estimate $\hat{\tau}(k)$ can be obtained first after setting the initial values, and then using the value of estimate $\hat{\tau}(k)$ to calculate

the values of estimates $\hat{\theta}_i(k)$ and $\hat{\kappa}_i(k)$. The loop continues until the estimates $\hat{\tau}(k)$, $\hat{\theta}_i(k)$ and $\hat{\kappa}_i(k)$ are stable.

The schematic diagram of the F-M-PC-GESG algorithm in (40)–(58) is given in Figure 1. It indicates that $\hat{\theta}_i(k)$ in each subsystem are common whereas $\hat{\kappa}_i(k)$ is separate. Figure 1 clearly shows how the partially coupled identification approach of the F-M-PC-GESG algorithm operates.

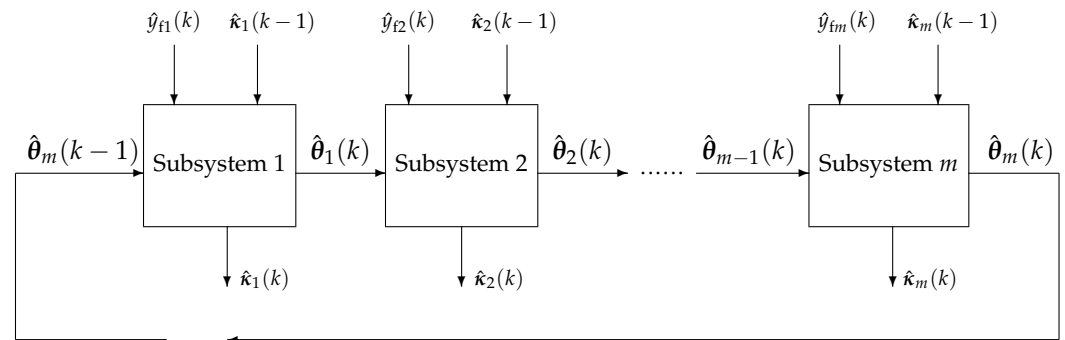


Figure 1. The schematic diagram of the F-M-PC-GESG algorithm.

4. The Multivariate Generalized Extended Stochastic Gradient Algorithm

The gradient algorithm is a classical identification method which does not generate covariance matrix in the calculation process and has a significant effect on improving the computational efficiency. In this section, the direct stochastic gradient method without improvement is utilized to identify parameters of the multivariate system. Define another quadratic criterion function for the identification model in (12):

$$J_3(\boldsymbol{\theta}) := \|\boldsymbol{\Omega}(k)\boldsymbol{\theta} - \mathbf{y}(k)\|^2.$$

Suppose that $\mu(k)$ is the step size. Minimizing $J_3(\boldsymbol{\theta})$ based upon the gradient search, the gradient relationship is

$$\begin{aligned} \hat{\boldsymbol{\theta}}(k) &= \hat{\boldsymbol{\theta}}(k-1) - \mu(k)\text{grad}[J_3(\hat{\boldsymbol{\theta}}(k-1))] \\ &= \hat{\boldsymbol{\theta}}(k-1) + \mu(k)\boldsymbol{\Omega}^T(k)[\mathbf{y}(k) - \boldsymbol{\Omega}(k)\hat{\boldsymbol{\theta}}(k-1)]. \end{aligned}$$

Let $\mu(k) := 1/r(k)$; then, the relationship is changed into

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \frac{\boldsymbol{\Omega}^T(k)}{r(k)}[\mathbf{y}(k) - \boldsymbol{\Omega}(k)\hat{\boldsymbol{\theta}}(k-1)], \tag{59}$$

$$r(k) = r(k-1) + \|\boldsymbol{\Omega}(k)\|^2. \tag{60}$$

The obstacle in identification is that $\hat{\boldsymbol{\theta}}(k)$ cannot be calculated because $\mathbf{v}(k)$ and $\mathbf{w}(k)$ in $\boldsymbol{\Omega}(k)$ are unmeasurable. Substitute estimates $\hat{\mathbf{v}}(k)$ and $\hat{\mathbf{w}}(k)$ for terms $\mathbf{v}(k)$ and $\mathbf{w}(k)$. From Equations (8) and (12), estimates $\hat{\mathbf{v}}(k)$ and $\hat{\mathbf{w}}(k)$ are calculated by

$$\hat{\mathbf{v}}(k) = \mathbf{y}(k) - \hat{\boldsymbol{\Omega}}(k)\hat{\boldsymbol{\theta}}(k), \tag{61}$$

$$\hat{\mathbf{w}}(k) = \mathbf{y}(k) - \boldsymbol{\Phi}(k)\hat{\boldsymbol{\theta}}(k). \tag{62}$$

Considering that $\boldsymbol{\Psi}(k)$ also involves the unmeasurable $\mathbf{w}(k)$, define $\hat{\boldsymbol{\Psi}}(k)$ by the estimate $\hat{\mathbf{w}}(k)$:

$$\hat{\boldsymbol{\Psi}}(k) = [\boldsymbol{\Phi}(k), -\hat{\mathbf{w}}^T(k-1), -\hat{\mathbf{w}}^T(k-2), \dots, -\hat{\mathbf{w}}^T(k-n_c)]^T. \tag{63}$$

Similarly, because $\boldsymbol{\psi}(k)$ involves the unmeasurable $\mathbf{v}(k-i)$, define $\hat{\boldsymbol{\psi}}(k)$ by the estimate $\hat{\mathbf{v}}(k-i)$ as

$$\hat{\boldsymbol{\psi}}(k) = [\hat{\boldsymbol{\theta}}^T(k-1), \hat{\boldsymbol{\theta}}^T(k-2), \dots, \hat{\boldsymbol{\theta}}^T(k-n_d)]^T. \tag{64}$$

Additionally, because $\Omega(k)$ involves the unmeasurable terms $\psi(k)$ and $\Psi(k)$, define $\hat{\Omega}(k)$ by estimates $\hat{\psi}(k)$ and $\hat{\Psi}(k)$:

$$\hat{\Omega}(k) = [\hat{\Psi}(k), \hat{\psi}^T(k) \otimes I_m].$$

Subsequently, the multivariate generalized extended stochastic gradient (M-GESG) algorithm is obtained as follows.

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{\hat{\Omega}^T(k)}{r(k)} [\mathbf{y}(k) - \hat{\Omega}(k)\hat{\theta}(k-1)], \quad (65)$$

$$r(k) = r(k-1) + \|\hat{\Omega}(k)\|^2, \quad r(0) = 1, \quad (66)$$

$$\hat{\Omega}(k) = [\hat{\Psi}(k), \hat{\psi}^T(k) \otimes I_m], \quad (67)$$

$$\hat{\Psi}(k) = [\Phi(k), -\hat{w}^T(k-1), -\hat{w}^T(k-2), \dots, -\hat{w}^T(k-n_c)]^T, \quad (68)$$

$$\hat{\psi}(k) = [\hat{v}^T(k-1), \hat{v}^T(k-2), \dots, \hat{v}^T(k-n_d)]^T, \quad (69)$$

$$\hat{w}(k) = \mathbf{y}(k) - \Phi(k)\hat{\theta}(k), \quad (70)$$

$$\hat{v}(k) = \mathbf{y}(k) - \hat{\Omega}(k)\hat{\theta}(k), \quad (71)$$

$$\hat{\theta}(k) = \begin{bmatrix} \hat{\eta}(k) \\ \text{col}[\hat{\mathbf{x}}(k)] \end{bmatrix}, \quad (72)$$

$$\hat{\eta}(k) = [\hat{\theta}^T(k), \hat{c}_1(k), \hat{c}_2(k), \dots, \hat{c}_{n_c}(k)]^T. \quad (73)$$

The computation procedures of the M-GESG algorithm in (65)–(73) are presented as follows.

1. Let $k = 1$, set the initial values $\hat{\theta}(0) = \mathbf{1}_{n_0}/p_0$, $r(0) = 1$, $\hat{w}(k-j) = \mathbf{0}$, $\hat{v}(k-j) = \mathbf{0}$, $j = 0, 1, \dots, \max[n_c, n_d]$, $p_0 = 10^6$, and set the data length K .
2. Acquire the observation data $\Phi(k)$ and $\mathbf{y}(k)$. Configure the matrix $\hat{\Psi}(k)$ and vector $\hat{\psi}(k)$ utilizing (68) and (69).
3. Calculate $\hat{\Omega}(k)$ using (67), and calculate $r(k)$ utilizing (66).
4. Refresh the parameter estimate vector $\hat{\theta}(k)$ utilizing (65). Read $\hat{\theta}(k)$ form $\hat{\theta}(k)$ in (72) and (73).
5. Calculate $\hat{v}(k)$ and $\hat{w}(k)$ utilizing (71) and (70).
6. Gain k by 1 if $k < K$, and then skip to Step 2. If not, acquire the parameter estimate $\hat{\theta}(k)$ and quit.

5. Convergence Analysis

The convergence of the F-M-PC-GESG algorithm is analyzed in this section.

Suppose that the σ algebra sequence $\mathcal{F}_k = \sigma(v(k), v(k-1), v(k-2), \dots)$ generated by $v(k)$, and $\{v(k), \mathcal{F}_k\}$ is a Martingale difference sequence on a probability space $\{\Omega, \mathcal{F}, P\}$ [49]. The sequence $\{v(k)\}$ satisfies

$$(Q1) \quad \mathbb{E}[v(k)|\mathcal{F}_{k-1}] = \mathbf{0}, \text{ a.s.},$$

$$(Q2) \quad \mathbb{E}[\|v(k)\|^2|\mathcal{F}_{k-1}] \leq \sigma^2 < \infty, \text{ a.s.}$$

Lemma 1. For the systems in (16) and (29) and the F-M-PC-GESG algorithm in (40)–(58), the following inequalities hold:

$$\sum_{k=1}^{\infty} \frac{\|\hat{\Phi}_{fi}(k)\|^2}{r_{\theta,i}^2(k)} < \infty, \quad i = 1, 2, \dots, m,$$

$$\sum_{k=1}^{\infty} \frac{\|\hat{\psi}(k)\|^2}{r_{\kappa,i}^2(k)} < \infty, \quad i = 1, 2, \dots, m,$$

$$\sum_{k=1}^{\infty} \frac{\|\hat{\chi}(k)\|^2}{r_{\tau}^2(k)} < \infty.$$

Theorem 1. For the systems in (16) and (29) and the F-M-PC-GESG algorithm in (40)–(58), suppose that (Q1) and (Q2) hold. There exists the positive constants λ_1 , λ_2 , and λ_3 independent of k , and an integer K such that the following persistent excitation condition holds:

$$(Q3) \quad \sum_{j=0}^K \frac{\hat{\Phi}_{fi}(k+j)\hat{\Phi}_{fi}^T(k+j)}{r_{\theta,i}(k+j)} \geq \lambda_1 \mathbf{I}, \quad i = 1, 2, \dots, m,$$

$$(Q4) \quad \sum_{j=0}^K \frac{\hat{\psi}(k+j)\hat{\psi}^T(k+j)}{r_{\kappa,i}(k+j)} \geq \lambda_2 \mathbf{I}, \quad i = 1, 2, \dots, m,$$

$$(Q5) \quad \sum_{j=0}^K \frac{\hat{\chi}(k+j)\hat{\chi}^T(k+j)}{r_{\tau}(k+j)} \geq \lambda_3 \mathbf{I}.$$

Then, the parameter estimation errors $\|\hat{\theta}_i(k) - \theta\|$, $\|\hat{\kappa}(k) - \kappa\|$, and $\|\hat{\tau}(k) - \tau\|$ given by the F-M-PC-GESG algorithm converge to zero in the mean square sense.

Through the above analysis, we can determine that the proposed algorithm can make parameter estimation errors of multivariate pseudo-linear systems converge to zero in the case of random disturbance. In other words, the proposed algorithm not only has the ability to estimate the unknown parameters accurately, but also has certain stability.

Theorem 1 can be proved in a similar to the way in [50] and is omitted here.

6. Simulations

This part is to demonstrate the superiority of F-M-PC-GESG algorithm in identification performances by conducting two simulations.

Example 1. Consider the following multivariate equation-error autoregressive moving average systems,

$$\mathbf{y}(k) = \Phi(k)\theta + \frac{D(z)}{c(z)}\mathbf{v}(k),$$

$$\Phi(k) = \begin{bmatrix} u_1(k-2) & \sin(y_1(k-2)) & y_1(k-1) & \cos(u_2(k-2)) \\ y_2(k-1) & \cos(u_1(k-1)) & u_2(k-1) & \sin(y_2(k-2)) \end{bmatrix},$$

$$\theta = [\theta_1, \theta_2, \theta_3, \theta_4]^T = [0.24, -0.14, -0.60, 0.02]^T,$$

$$c(z) = 1 + c_1z^{-1} + c_2z^{-2} = 1 + 0.65z^{-1} - 0.31z^{-2},$$

$$D(z) = \mathbf{I}_2 + \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} z^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} -0.12 & 0.17 \\ -0.03 & -0.24 \end{bmatrix} z^{-1}.$$

The parameter vector to be identified is

$$\vartheta = [\theta_1, \theta_2, \theta_3, \theta_4, c_1, c_2, d_{11}, d_{12}, d_{21}, d_{22}]^T$$

$$= [0.24, -0.14, -0.60, 0.02, 0.65, -0.31, -0.12, 0.17, -0.03, -0.24]^T.$$

In this simulation, the input vector $\mathbf{u}(k) = \begin{bmatrix} u_1(k) \\ u_2(k) \end{bmatrix} \in \mathbb{R}^2$ is a random sequence with zero mean and variance one. The output vector is $\mathbf{y}(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} \in \mathbb{R}^2$. The white noise vector $\mathbf{v}(k) = \begin{bmatrix} v_1(k) \\ v_2(k) \end{bmatrix} \in \mathbb{R}^2$ is with zero mean. σ_1^2 and σ_2^2 are variances of $v_1(k)$ and $v_2(k)$. Taking the noise variances $\sigma_1^2 = \sigma_2^2 = 0.20^2$, utilizing the M-GESG algorithm and the F-M-PC-GESG algorithm to identify system parameters, parameter estimates and their errors $\delta := \|\hat{\boldsymbol{\theta}}(k) - \boldsymbol{\theta}\| / \|\boldsymbol{\theta}\|$ are given in Table 1. The parameter identification errors versus k are given in Figure 2. The parameter estimates $\hat{\theta}_1(k), \hat{\theta}_2(k), \hat{\theta}_3(k), \hat{\theta}_4(k), \hat{c}_1(k), \hat{c}_2(k), \hat{d}_{11}(k), \hat{d}_{12}(k), \hat{d}_{21}(k),$ and $\hat{d}_{22}(k)$ versus k of the F-M-PC-GESG algorithm are given in Figures 3 and 4.

Table 1. Parameter estimates and errors ($\sigma_1^2 = \sigma_2^2 = 0.20^2$).

Algorithms	k	100	200	500	1000	2000	3000	True Value
M-GESG	θ_1	0.24221	0.24151	0.24223	0.23292	0.22983	0.22863	0.24000
	θ_2	0.05203	0.02792	-0.00854	-0.02886	-0.04654	-0.05528	-0.14000
	θ_3	-0.61451	-0.61142	-0.59979	-0.61122	-0.60372	-0.60521	-0.60000
	θ_4	-0.00923	0.00479	0.00416	0.01756	0.01947	0.02117	0.02000
	c_1	0.42969	0.44032	0.44813	0.46623	0.47098	0.47768	0.65000
	c_2	-0.43009	-0.43931	-0.44511	-0.45792	-0.45798	-0.46094	-0.31000
	d_{11}	-0.05862	-0.06323	-0.06660	-0.06933	-0.07090	-0.07281	-0.12000
	d_{12}	-0.00128	-0.00411	-0.00552	-0.00673	-0.00676	-0.00691	0.17000
	d_{21}	-0.02181	-0.02119	-0.02004	-0.01749	-0.01566	-0.01452	-0.03000
	d_{22}	0.00430	0.00556	0.00304	-0.00250	-0.00599	-0.00827	-0.24000
	δ (%)	42.79473	41.60834	40.01046	38.71188	37.83113	37.30980	
F-M-PC-GESG	θ_1	0.24743	0.23989	0.24511	0.23901	0.23664	0.23713	0.24000
	θ_2	-0.10583	-0.11597	-0.12770	-0.13172	-0.13459	-0.13672	-0.14000
	θ_3	-0.59934	-0.58195	-0.59534	-0.59614	-0.59351	-0.59624	-0.60000
	θ_4	0.01463	0.02155	0.01422	0.02609	0.02376	0.02292	0.02000
	c_1	0.65312	0.66910	0.68801	0.70132	0.71257	0.71886	0.65000
	c_2	-0.34080	-0.32821	-0.31140	-0.29937	-0.28767	-0.28125	-0.31000
	d_{11}	-0.16261	-0.17089	-0.15802	-0.14864	-0.13486	-0.13573	-0.12000
	d_{12}	0.03206	0.04136	0.04804	0.06647	0.07554	0.07958	0.17000
	d_{21}	-0.02175	-0.01882	-0.01885	-0.02282	-0.02068	-0.02177	-0.03000
	d_{22}	-0.14040	-0.13357	-0.13861	-0.15208	-0.15712	-0.16094	-0.24000
	δ (%)	17.67220	17.44393	16.38489	14.48147	13.95661	13.86945	

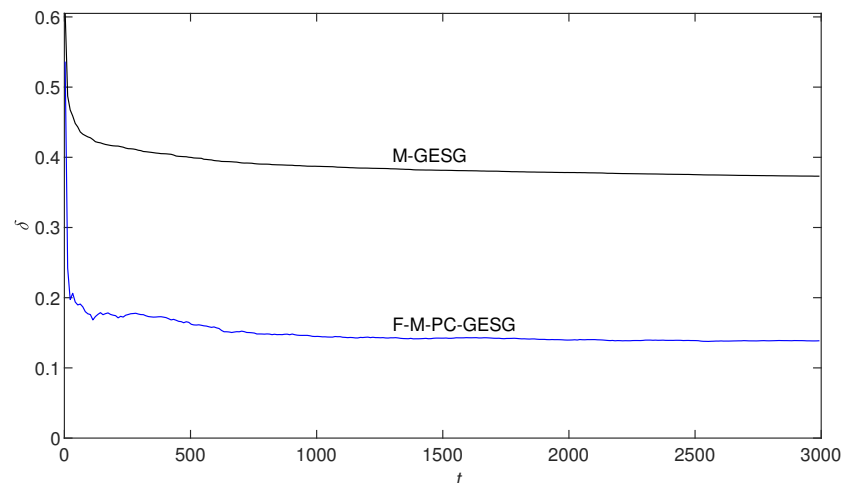


Figure 2. Identification errors versus k ($\sigma_1^2 = \sigma_2^2 = 0.20^2$).

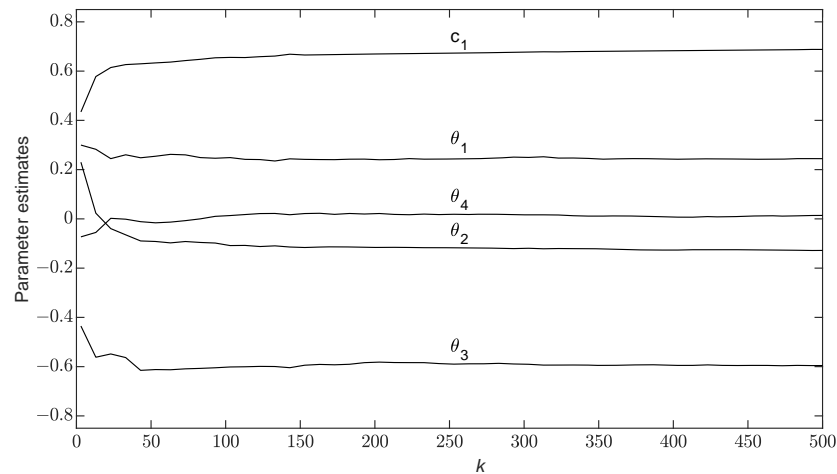


Figure 3. Parameter estimates $\hat{\theta}_1(k), \hat{\theta}_2(k), \hat{\theta}_3(k), \hat{\theta}_4(k), \hat{c}_1(k)$ versus k .

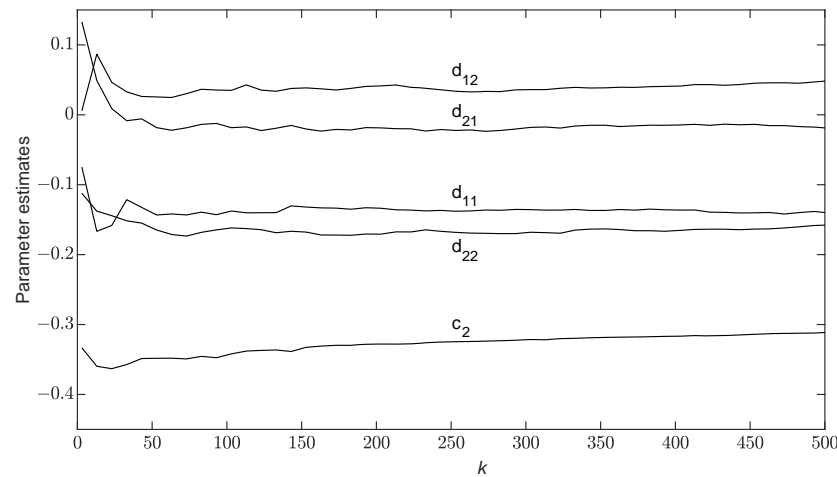


Figure 4. Parameter estimates $\hat{c}_2(k), \hat{d}_{11}(k), \hat{d}_{12}(k), \hat{d}_{21}(k), \hat{d}_{22}(k)$ versus k .

Example 2. Consider another multivariate equation-error autoregressive moving average system,

$$\Phi(t) = \begin{bmatrix} u_1^2(k-1) & \cos^2(u_1(k-2)) & u_2(k-1)y_1(k-2) \\ u_2^2(k-1) & \sin^2(y_1(k-2)) & u_1(k-1)y_2(k-2) \\ y_1^3(k-1) & \sin^2(y_2(k-1)) \\ y_2^2(k-2) & \cos^2(y_2(k-1)) \end{bmatrix}'$$

$$\theta = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5]^T = [0.08, -0.09, 0.02, -0.01, -0.03]^T,$$

$$c(z) = 1 + c_1z^{-1} = 1 + 0.89z^{-1},$$

$$D(z) = I_2 + D_1z^{-1} + D_2z^{-2}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.22 & -0.29 \\ -0.03 & 0.10 \end{bmatrix} z^{-1} + \begin{bmatrix} 0.04 & -0.08 \\ -0.05 & 0.12 \end{bmatrix} z^{-2}.$$

The parameter vector to be identified is

$$\vartheta = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, c_1, d_{11}, d_{12}, d_{21}, d_{22}, d_{12}, d_{21}, d_{12}, d_{21}, d_{12}, d_{21}]^T$$

$$= [0.08, -0.09, 0.02, -0.01, -0.03, 0.89, 0.22, -0.29, 0.04, -0.08, -0.03, 0.10, -0.05, 0.12]^T.$$

The configuration of the simulation in this example is the same as in Example 1. Set noise variances $\sigma_1^2 = 0.50^2$ and $\sigma_2^2 = 0.40^2$, and utilize the M-GESG algorithm and the F-M-PC-GESG algorithm to identify the system parameters. The parameter estimates and

their errors are given in Table 2. The parameter identification errors versus t are given in Figure 5.

Table 2. Parameter estimates and errors ($\sigma_1^2 = 0.50^2, \sigma_2^2 = 0.40^2$).

Algorithms	k	100	200	500	1000	2000	3000	True Value
M-GESG	θ_1	0.09112	0.08857	0.06625	0.08761	0.08713	0.08118	0.08000
	θ_2	-0.10932	-0.10922	-0.11672	-0.11222	-0.11331	-0.11291	-0.09000
	θ_3	0.10330	0.06629	0.04765	0.03156	0.02940	0.02798	0.02000
	θ_4	-0.08128	-0.05999	-0.05284	-0.04692	-0.04326	-0.04202	-0.01000
	θ_5	0.05408	0.04776	0.03644	0.03423	0.02818	0.02637	-0.03000
	c_1	0.63454	0.64477	0.66918	0.69922	0.71524	0.73325	0.89000
	$d1_{11}$	0.16746	0.16823	0.16294	0.15883	0.15732	0.15446	0.22000
	$d1_{12}$	-0.11067	-0.10817	-0.10595	-0.10446	-0.10122	-0.10005	-0.29000
	$d2_{11}$	-0.07832	-0.08579	-0.09696	-0.10167	-0.10818	-0.11186	0.04000
	$d2_{12}$	-0.02125	-0.02080	-0.02172	-0.02349	-0.02480	-0.02649	-0.08000
	$d1_{21}$	0.19580	0.19361	0.19552	0.20044	0.19851	0.19854	-0.03000
	$d1_{22}$	-0.03530	-0.03103	-0.02681	-0.02341	-0.02217	-0.02248	0.10000
	$d2_{21}$	-0.19341	-0.19073	-0.18208	-0.17582	-0.17124	-0.16896	-0.05000
	$d2_{22}$	0.30383	0.29812	0.29649	0.29994	0.29897	0.29897	0.12000
	δ (%)		51.69208	50.09867	48.69494	47.52852	46.83530	46.27683
F-M-PC-GESG	θ_1	0.06055	0.08860	0.07901	0.08999	0.08715	0.07885	0.08000
	θ_2	-0.07339	-0.08860	-0.09644	-0.09032	-0.09163	-0.08886	-0.09000
	θ_3	0.07020	0.02402	0.01128	0.01168	0.01124	0.00870	0.02000
	θ_4	-0.00816	0.00063	0.00035	0.00275	0.00146	-0.00635	-0.01000
	θ_5	0.00114	-0.02136	-0.03064	-0.03338	-0.03768	-0.03383	-0.03000
	c_1	0.99456	0.98815	0.98077	0.98014	0.97418	0.97488	0.89000
	$d1_{11}$	0.16879	0.15751	0.18199	0.19942	0.21985	0.21906	0.22000
	$d1_{12}$	-0.27855	-0.29182	-0.30472	-0.29502	-0.29439	-0.29678	-0.29000
	$d2_{11}$	0.10959	0.06096	0.05244	0.04869	0.02627	0.02232	0.04000
	$d2_{12}$	-0.20205	-0.19511	-0.16709	-0.14945	-0.13756	-0.13392	-0.08000
	$d1_{21}$	0.05176	0.05175	0.03801	0.01758	0.01343	0.00674	-0.03000
	$d1_{22}$	0.02924	0.06773	0.08395	0.08658	0.09398	0.09718	0.10000
	$d2_{21}$	0.02688	0.04268	0.03090	0.02807	0.01438	0.00469	-0.05000
	$d2_{22}$	0.14854	0.13947	0.11760	0.11307	0.10907	0.10424	0.12000
	δ (%)		23.94783	21.27826	17.32330	15.14836	13.24001	12.48907

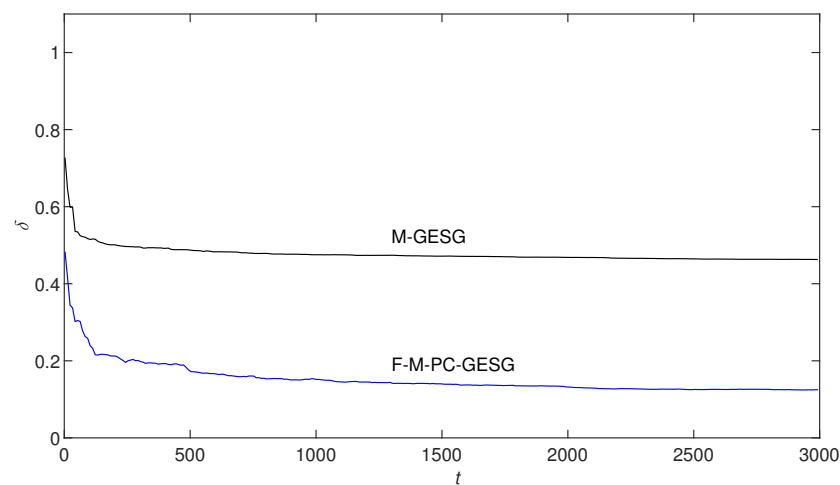


Figure 5. Identification errors versus k ($\sigma_1^2 = 0.50^2, \sigma_2^2 = 0.40^2$).

With Tables 1 and 2 and Figures 2–5, identification performances of proposed algorithms are analyzed as follows.

1. Tables 1 and 2, Figures 2 and 5 indicate that parameter identification errors of the M-GESG and the F-M-PC-GESG algorithms decrease with increasing data length k .

This reveals that the proposed algorithms are valid in parameter identification for the multivariate equation-error autoregressive moving average system.

2. Through Figures 2 and 5, we can see that the F-M-PC-GESG algorithm has superiority over the M-GESG algorithm in parameter identification precision under the same data length and noise variances.
3. Figures 3 and 4 show that the F-M-PC-GESG algorithm can rapidly obtain access to precise parameter estimates.

7. Conclusions

This paper presents methods of how to improve parameter identification effects for multivariate pseudo-linear systems under conditions of random interference and parameter coupling, which provides modular solutions for modeling and forecasting of real multivariate time series. Taking into account colored noises and high-dimensional unknown parameters, the original system is filtered by the designed filter and then be transformed into several subsystem identification models by utilizing the coupling identification approach. A new filtering-based multivariate gradient algorithm is proposed, which has higher parameter estimation precision and faster identification efficiency than the conventional multivariate gradient algorithm. Convergence analysis and simulation experiments confirms that the F-M-PC-GESG algorithm has performance that can obtain access to unknown parameter estimates precisely and rapidly. Future research directions include applying proposed methods to parameter identification problems for other linear or nonlinear models under random interference in various engineering systems.

Author Contributions: Methodology, P.M.; Writing—original draft, P.M.; Writing—review & editing, Y.C.; Supervision, Y.L.; Funding acquisition, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 62103293), by the Natural Science Foundation of Jiangsu Province (No. BK20210709), by the Suzhou Municipal Science and Technology Bureau (No. SYG202138), and by the ‘Taihu Light’ Basic Research Project on Scientific and Technological Breakthroughs of Wuxi (No. K20221006).

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, Y.Y.; Jiang, W.; Charalambous, T. Machine learning based iterative learning control for non-repetitive time-varying systems. *Int. J. Robust Nonlinear Control* **2022**, *33*, 4098–4116. [[CrossRef](#)]
2. Zhang, X.; Lei, Y.; Chen, H.; Zhang, L.; Zhou, Y. Multivariate time-series modeling for forecasting sintering temperature in rotary kilns using DCGNet. *IEEE Trans. Ind. Inform.* **2021**, *17*, 4635–4645. [[CrossRef](#)]
3. Terbuch, A.; O’Leary, P.; Khalili-Motlagh-Kasmaei, N.; Auer, P.; Zohrer, A.; Winter, V. Detecting anomalous multivariate time-series via hybrid machine learning. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 2503711. [[CrossRef](#)]
4. Niaki, S.T.A.; Davoodi, M. Designing a multivariate-multistage quality control system using artificial neural networks. *Int. J. Prod. Res.* **2009**, *47*, 251–271. [[CrossRef](#)]
5. Wen, C.; Xie, Y.; Qiao, Z.; Xu, L.; Qian, Y. A tensor generalized weighted linear predictor for FDA-MIMO radar parameter estimation. *IEEE Trans. Veh. Technol.* **2022**, *71*, 6059–6072. [[CrossRef](#)]
6. Olivier, L.E.; Craig, I.K. Model-plant mismatch detection and model update for a run-of-mine ore milling circuit under model predictive control. *J. Process Control* **2013**, *23*, 100–107. [[CrossRef](#)]
7. Zhang, Z.; Ren, J.; Bai, W. MIMO non-parametric modeling of ship maneuvering motion for marine simulator using adaptive moment estimation locally weighted learning. *Ocean. Eng.* **2022**, *261*, 112103. [[CrossRef](#)]
8. Qi, R.; Tao, G.; Jiang, B. Adaptive control of MIMO time-varying systems with indicator function based parametrization. *Automatica* **2014**, *50*, 1369–1380. [[CrossRef](#)]
9. Wang, Z.; Xie, L.; Wan, Q. Beam-space joint azimuth, elevation, and delay estimation for large-scale MIMO-OFDM system. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 9506412. [[CrossRef](#)]
10. Ibrir, S. Joint state and parameter estimation of non-linearly parameterized discrete-time nonlinear systems. *Automatica* **2018**, *97*, 226–233. [[CrossRef](#)]

11. Mari, J.; Stoica, P.; McKelvey, T. Vector ARMA estimation: A reliable subspace approach. *IEEE Trans. Signal Process.* **2000**, *48*, 2092–2104. [[CrossRef](#)]
12. Luo, K.; Manikas, A. Superresolution multitarget parameter estimation in MIMO radar. *IEEE Trans. Geosci. Remote. Sens.* **2013**, *51*, 3683–3693. [[CrossRef](#)]
13. Zhang, G.D.; Sheng, Y.H.; Shi, Y.X. Uncertain hypothesis testing of multivariate uncertain regression model. *J. Intell. Fuzzy Syst.* **2022**, *43*, 7341–7350. [[CrossRef](#)]
14. Oigard, T.A.; Hanssen, A.; Hansen, R.E.; Godtliebsen, F. EM-estimation and modeling of heavy-tailed processes with the multivariate normal inverse Gaussian distribution. *Signal Process.* **2005**, *85*, 1655–1673. [[CrossRef](#)]
15. Ding, F.; Liu, X.M.; Chen, H.B.; Yao, G.Y. Hierarchical gradient based and hierarchical least squares based iterative parameter identification for CARARMA systems. *Signal Process.* **2014**, *97*, 31–39. [[CrossRef](#)]
16. Xu, L.; Ding, F. Separable synthesis estimation methods and convergence analysis for multivariable systems. *J. Comput. Appl. Math.* **2023**, *427*, 115104. [[CrossRef](#)]
17. Xu, L. Separable Newton recursive estimation method through system responses based on dynamically discrete measurements with increasing data length. *Int. J. Control Autom. Syst.* **2022**, *20*, 432–443. [[CrossRef](#)]
18. Cui, T.; Chen, F.; Ding, F.; Sheng, J. Combined estimation of the parameters and states for a multivariable state-space system in presence of colored noise. *Int. J. Adapt. Control Signal Process.* **2020**, *34*, 590–613. [[CrossRef](#)]
19. Zheng, M.; Li, L.; Peng, H.; Xiao, J.; Yang, Y.; Zhao, H. Parameters estimation and synchronization of uncertain coupling recurrent dynamical neural networks with time-varying delays based on adaptive control. *Neural Comput. Appl.* **2018**, *30*, 2217–2227. [[CrossRef](#)]
20. Geng, X.; Xie, L. Learning the LMP-load coupling from data: A support vector machine based approach. *IEEE Trans. Power Syst.* **2017**, *32*, 1127–1138.
21. Ding, F.; Liu, G.; Liu, X.P. Partially coupled stochastic gradient identification methods for non-uniformly sampled systems. *IEEE Trans. Autom. Control* **2010**, *55*, 1976–1981. [[CrossRef](#)]
22. Zhou, Y.; Zhang, X.; Ding, F. Partially-coupled nonlinear parameter optimization algorithm for a class of multivariate hybrid models. *Appl. Math. Comput.* **2022**, *414*, 126663. [[CrossRef](#)]
23. Huang, Y.G.; Cao, L.B.; Zhang, J.; Pan, L.; Liu, Y.Y. Exploring feature coupling and model coupling for image source identification. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 3108–3121. [[CrossRef](#)]
24. Wang, X.H.; Ding, F. Partially coupled extended stochastic gradient algorithm for nonlinear multivariable output error moving average systems. *Eng. Comput.* **2017**, *34*, 629–647. [[CrossRef](#)]
25. Ma, P.; Wang, L. Filtering-based recursive least squares estimation approaches for multivariate equation-error systems by using the multiinnovation theory. *Int. J. Adapt. Control Signal Process.* **2021**, *35*, 1898–1915. [[CrossRef](#)]
26. Ma, J.; Xiong, W.; Ding, F.; Alsaedi, A.; Hayat, T. Data filtering based forgetting factor stochastic gradient algorithm for Hammerstein systems with saturation and preload nonlinearities. *J. Frankl.-Inst.-Eng. Appl. Math.* **2016**, *353*, 4280–4299. [[CrossRef](#)]
27. Li, M.; Liu, X. The least squares based iterative algorithms for parameter estimation of a bilinear system with autoregressive noise using the data filtering technique. *Signal Process.* **2018**, *147*, 23–34. [[CrossRef](#)]
28. Ji, Y.; Jiang, A. Filtering-based accelerated estimation approach for generalized time-varying systems with disturbances and colored noises. *IEEE Trans. Circuits Syst. II Express Briefs* **2023**, *70*, 206–210. [[CrossRef](#)]
29. Imani, M.; Braga-Neto, U.M. Maximum-likelihood adaptive filter for partially observed boolean dynamical systems. *IEEE Trans. Signal Process.* **2017**, *65*, 359–371. [[CrossRef](#)]
30. Zhang, C.; Liu, H.; Ji, Y. Gradient parameter estimation of a class of nonlinear systems based on the maximum likelihood principle. *Int. J. Control Autom. Syst.* **2022**, *20*, 1393–1404. [[CrossRef](#)]
31. Chen, J.; Zhu, Q.M.; Liu, Y.J. Modified Kalman filtering based multi-step length gradient iterative algorithm for ARX models with random missing outputs. *Automatica* **2020**, *118*, 109034.
32. Li, M.H.; Liu, X.M. Iterative identification methods for a class of bilinear systems by using the particle filtering technique. *Int. J. Adapt. Control Signal Process.* **2021**, *35*, 2056–2074. [[CrossRef](#)]
33. Ding, F.; Liu, X.M.; Ma, X.Y. Kalman state filtering based least squares iterative parameter estimation for observer canonical state space systems using decomposition. *J. Comput. Appl. Math.* **2016**, *301*, 135–143. [[CrossRef](#)]
34. Li, M.H.; Liu, X.M. Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique. *Int. J. Control Autom. Syst.* **2020**, *18*, 1581–1592. [[CrossRef](#)]
35. Chen, Y.Y.; Zhou, Y.W. Machine learning based decision making for time varying systems: Parameter estimation and performance optimization. *Knowl.-Based Syst.* **2020**, *190*, 105479. [[CrossRef](#)]
36. Altbawi, S.M.A.; Khalid, S.B.A.; Bin Mokhtar, A.S.; Shareef, H.; Husain, N.; Yahya, A.; Haider, S.A.; Moin, L.; Alsisi, R.H. An improved gradient-based optimization algorithm for solving complex optimization problems. *Processes* **2023**, *11*, 498. [[CrossRef](#)]
37. Xu, L. Parameter estimation for nonlinear functions related to system responses. *Int. J. Control Autom. Syst.* **2023**, *21*, 1780–1792. [[CrossRef](#)]
38. Zhang, X.; Ding, F. Optimal adaptive filtering algorithm by using the fractional-order derivative. *IEEE Signal Process. Lett.* **2022**, *29*, 399–403. [[CrossRef](#)]

39. Roman, C.; Ferrante, F.; Prieur, C. Parameter identification of a linear wave equation from experimental boundary data. *IEEE Trans. Control Syst. Technol.* **2021**, *29*, 2166–2179. [[CrossRef](#)]
40. Chen, J.; Zhu, Q.M.; Hu, M.F.; Guo, L.X.; Narayan, P. Improved gradient descent algorithms for time-delay rational state-space systems: Intelligent search method and momentum method. *Nonlinear Dyn.* **2020**, *101*, 361–373. [[CrossRef](#)]
41. Kulikova, M.V. Gradient-based parameter estimation in pairwise linear Gaussian system. *IEEE Trans. Autom. Control* **2017**, *62*, 1511–1517. [[CrossRef](#)]
42. Imani, M.; Braga-Neto, U.M. Particle filters for partially-observed Boolean dynamical systems. *Automatica* **2018**, *87*, 238–250. [[CrossRef](#)]
43. Ding, F. Least squares parameter estimation and multi-innovation least squares methods for linear fitting problems from noisy data. *J. Comput. Appl. Math.* **2023**, *426*, 115107. [[CrossRef](#)]
44. Ding, F.; Xu, L.; Zhang, X.; Zhou, Y. Filtered auxiliary model recursive generalized extended parameter estimation methods for Box-Jenkins systems by means of the filtering identification idea. *Int. J. Robust Nonlinear Control* **2023**, *33*, 5510–5535. [[CrossRef](#)]
45. Ma, P.; Wang, L. Partially coupled stochastic gradient estimation for multivariate equation-error systems. *Mathematics* **2022**, *10*, 2955. [[CrossRef](#)]
46. Ding, F. Coupled-least-squares identification for multivariable systems. *IET Control Theory Appl.* **2013**, *7*, 68–79. [[CrossRef](#)]
47. Xu, L.; Ding, F.; Zhu, Q. Separable synchronous multi-innovation gradient-based iterative signal modeling from on-line measurements. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 6501313. [[CrossRef](#)]
48. Wang, Y.J. Novel data filtering based parameter identification for multiple-input multiple-output systems using the auxiliary model. *Automatica* **2016**, *71*, 308–313. [[CrossRef](#)]
49. Goodwin, G.C.; Sin, K.S. *Adaptive Filtering, Prediction and Control*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1984.
50. Ding, F. *System Identification-Performances Analysis for Identification Methods*; Science Press: Beijing, China, 2014.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.