


## Article

# Partially Coupled Stochastic Gradient Estimation for Multivariate Equation-Error Systems

Ping Ma <sup>1</sup> and Lei Wang <sup>2,\*</sup><sup>1</sup> Jiangsu Key Laboratory of Media Design and Software Technology, School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China<sup>2</sup> School of Automation, Wuxi University, Wuxi 214105, China

\* Correspondence: leiwang@cwuxu.edu.cn

**Abstract:** This paper researches the identification problem for the unknown parameters of the multivariate equation-error autoregressive systems. Firstly, the original identification model is decomposed into several sub-identification models according to the number of system outputs. Then, based on the characteristic that the information vector and the parameter vector are common among the sub-identification models, the coupling identification concept is used to propose a partially coupled generalized stochastic gradient algorithm. Furthermore, by expanding the scalar innovation of each subsystem model to the innovation vector, a partially coupled multi-innovation generalized stochastic gradient algorithm is proposed. Finally, the numerical simulations indicate that the proposed algorithms are effective and have good parameter estimation performances.

**Keywords:** parameter estimation; coupling identification; stochastic gradient; multi-innovation theory; multivariate system

MSC: 93C35



**Citation:** Ma, P.; Wang, L. Partially Coupled Stochastic Gradient Estimation for Multivariate Equation-Error Systems. *Mathematics* **2022**, *10*, 2955. <https://doi.org/10.3390/math10162955>

Academic Editors: Leonid Piterbarg and Simeon Reich

Received: 5 July 2022

Accepted: 12 August 2022

Published: 16 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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

Parameter estimation is an important part in the field of system identification, which usually identifies the unknown parameters in the models according to the input and output data of the system when the model structure of a system is known [1–3]. Parameter estimation has been used in many fields in recent years, including chemistry, mechanics, engineering and so on [4–6]. For example, in chemical engineering field, Khalik et al. applied the parameter estimation approach by using current/voltage data in achieving physically meaningful parameters of the Doyle–Fuller–Newman model for Lithium-ion batteries [7]. In mechanics field, Shamrao et al. estimated the terramechanics parameters by applying dynamic Bayesian estimation techniques on the measurements from simple single-wheel tests [8]. In engineering, Calasan et al. proposed two algorithms for the transformer parameter estimation to improve the estimation process and prevent inaccuracies and parameter mismatch with the real parameters of the transformer [9].

In many identification objects of parameter estimation, multivariate systems is a very common class [10–12]. The majority of industrial processes are the multivariate systems [13,14]. It is not easy to estimate the unknown parameters of multivariate systems, because such systems are quite complex; for example, there are many variables, there is coupling between variables, or there is time delay [15–17]. In recent years, the identification of the multivariate systems has attracted the attention of many scholars. Shafin et al. studied the angle and delay estimation for 3D massive MIMO systems under a parametric channel modeling, which was crucial for such systems to realize the predicted capacity gains [18]. Kawaria et al. designed a Levy shuffled frog leaping algorithm with high parameter estimation efficiency to estimate the parameters of multiple-input–multiple-output

bilinear systems [19]. Roy et al. developed an online plant-parameter identification method for the multi-input–multi-output linear time-invariant systems [20].

The coupling identification concept is a useful identification method that has been developed in recent decades. It is suitable for parameter estimation problems for the multivariate systems with the same parameters among subsystems [21,22]. The identification algorithm based on the coupling concept has the advantage of lower computation amounts compared with the traditional identification algorithm for multivariate systems. Cui et al. combined the Kalman filtering principle and the coupling identification concept to derive a Kalman filtering-based partially coupled recursive least squares algorithm for jointly estimating the parameters and the states of the multivariable state–space system; their algorithm had high computational efficiency [23]. The multi-innovation identification theory is also an effective method that has been used to improve parameter estimation accuracy in recent years [24–26]. This method not only utilizes the data information of the current time, but also makes full use of the data information of the previous time [27–29]. Chaudhary et al. presented a multi-innovation fractional least mean square adaptive algorithm for the input-nonlinear systems by expanding the scalar innovation into a vector innovation by using the multi-innovation identification theory [30]. This method can also be applied to the parameter estimation of the multivariate systems.

We have studied the parameter identification problems of multivariate systems in the past. In [31], the colored noise of the original system was filtered into white noise by using the data filtering technique. In [32], the original identification system was decomposed into two sub-identification systems, where one contains the system model parameter vector and the other contains the noise model parameter matrix by applying the decomposition method. However, the coupling identification concept used in this paper is different from the previous two identification methods, as it decomposes the original system according to the number of system outputs to obtain several subsystems with partial parameter vectors and information vectors coupled. It has significant advantages in reducing the computational cost of the algorithm because it makes the original complex system simplified. Therefore, this paper presents new identification methods to estimate the parameters of the multivariate equation-error autoregressive system, which have high computational efficiency and high parameter estimation accuracy. The main contributions of this paper lie in the following.

1. This paper decomposes the multivariate equation-error autoregressive system into several sub-identification models according to the number of the system outputs.
2. A multivariate partially coupled generalized stochastic gradient (M-PC-GSG) algorithm is proposed for the multivariate equation-error system by utilizing the coupling identification concept, which can reduce the computation amounts compared with the traditional stochastic gradient algorithm.
3. A multivariate partially coupled multi-innovation generalized stochastic gradient (M-PC-MI-GSG) algorithm is proposed by using the multi-innovation identification theory, which has higher parameter estimation accuracy than the M-PC-GSG algorithm.

The rest of this paper is organized as follows. Section 2 presents a multivariate equation-error autoregressive system and describes its identification difficulties. Section 3 proposes a partially coupled generalized stochastic gradient algorithm and gives its schematic diagram. Section 4 proposes a partially coupled multi-innovation generalized stochastic gradient algorithm. Section 5 presents two numerical examples to indicate that the proposed algorithms are effective. Finally, we offer some concluding remarks in Section 6.

## 2. System Description and Identification Model

First of all, we give some notation in this paper.  $I_m$  denotes an identity matrix of size  $m \times m$ ;  $\mathbf{1}_n$  stands for an  $n$ -dimensional column vector whose elements are 1, that is  $\mathbf{1}_n = [1, 1, \dots, 1]^T \in \mathbb{R}^n$ ;  $\mathbf{1}_{m \times n}$  represents a matrix of size  $m \times n$  whose elements are 1; the norm of a matrix  $\mathbf{X}$  is defined by  $\|\mathbf{X}\|^2 := \text{tr}[\mathbf{X}\mathbf{X}^T]$ , the superscript T stands for the matrix/vector transpose, the symbol  $\otimes$  represents Kronecker product, for ex-

ample,  $A := [a_{ij}] \in \mathbb{R}^{m \times n}$ ,  $B := [b_{ij}] \in \mathbb{R}^{p \times q}$ ,  $A \otimes B = [a_{ij}b_{kl}] \in \mathbb{R}^{(mp) \times (nq)}$ , in general,  $A \otimes B \neq B \otimes A$ ;  $\text{col}[X]$  is defined as a vector consisting of all columns of matrix  $X$  arranged in order, for example,  $X := [x_1, x_2, \dots, x_n] \in \mathbb{R}^{m \times n}$ ,  $x_i \in \mathbb{R}^m$  ( $i = 1, 2, \dots, n$ ),  $\text{col}[X] := [x_1^T, x_2^T, \dots, x_n^T]^T \in \mathbb{R}^{mn}$ .

Consider the following multivariate equation-error autoregressive system,

$$y(t) = \Phi_s(t)\theta + C^{-1}(z)v(t), \quad (1)$$

where  $y(t) := [y_1(t), y_2(t), \dots, y_m(t)]^T \in \mathbb{R}^m$  is the system output vector,  $\Phi_s(t) \in \mathbb{R}^{m \times n}$  is the system information matrix consisting of the input-output data,  $\theta \in \mathbb{R}^n$  is the system parameter vector to be identified,  $v(t) := [v_1(t), v_2(t), \dots, v_m(t)]^T \in \mathbb{R}^m$  is a white noise process with zero mean,  $C(z) \in \mathbb{R}^{m \times m}$  is a polynomial matrix in the unit backward shift operator  $[z^{-1}y(t) = y(t-1)]$ :

$$C(z) := I_m + C_1z^{-1} + C_2z^{-2} + \dots + C_{n_c}z^{-n_c}, \quad C_i \in \mathbb{R}^{m \times m}.$$

Define the noise model,

$$w(t) := C^{-1}(z)v(t) \in \mathbb{R}^m. \quad (2)$$

Assume that the orders  $m$ ,  $n$  and  $n_c$  are known, and  $y(t) = 0$ ,  $\Phi_s(t) = 0$ ,  $w(t) = 0$  and  $v(t) = 0$  for  $t \leq 0$ .

Define the parameter matrix  $\gamma$  and the information vector  $\phi(t)$  as

$$\begin{aligned} \gamma^T &:= [C_1, C_2, \dots, C_{n_c}] \in \mathbb{R}^{m \times (mn_c)}, \\ \phi(t) &:= [-w^T(t-1), -w^T(t-2), \dots, -w^T(t-n_c)]^T \in \mathbb{R}^{mn_c}. \end{aligned}$$

From Equation (2), we have

$$\begin{aligned} w(t) &= [I_m - C(z)]w(t) + v(t), \\ &= (-C_1z^{-1} - C_2z^{-2} - \dots - C_{n_c}z^{-n_c})w(t) + v(t) \\ &= -C_1w(t-1) - C_2w(t-2) - \dots - C_{n_c}w(t-n_c) + v(t). \\ &= \gamma^T\phi(t) + v(t). \end{aligned} \quad (3)$$

Then, the multivariate equation-error autoregressive system in (1) can be transformed into the following identification model,

$$y(t) = \Phi_s(t)\theta + w(t) \quad (4)$$

$$= \Phi_s(t)\theta + \gamma^T\phi(t) + v(t). \quad (5)$$

For the identification model in (5), the objective of this paper is to identify the unknown parameters  $\theta$  and  $\gamma$  by researching some identification methods. Currently, the observable data are  $y(t)$  and  $\Phi_s(t)$ . Certainly, the most direct identification method is to integrate the information vector  $\theta$  and the information matrix  $\gamma$  into a new information matrix  $\vartheta$ , and the parameter matrix  $\Phi_s(t)$  and the parameter vector  $\phi(t)$  into a new parameter matrix  $\Phi(t)$ , then we can obtain the following identification model:

$$y(t) = \Phi(t)\vartheta + v(t), \quad (6)$$

$$\Phi(t) := [\Phi_s(t), I_m \otimes \phi^T(t)] \in \mathbb{R}^{m \times n_0}, \quad n_0 := n + m^2n_c, \quad (7)$$

$$\vartheta := \begin{bmatrix} \theta \\ \text{col}[\gamma] \end{bmatrix} \in \mathbb{R}^{n_0}. \quad (8)$$

However, the information matrix  $\Phi(t)$  in model (6) contains a large number of zero elements because it is calculated from the Kronecker product, which results in redundant computation amounts in the identification processes. Therefore, it is necessary to find

an alternative method with less computation amounts to estimate the identification model in (5). In this paper, two efficient algorithms with good performances are proposed to solve this problem by applying the coupling identification concept and the multi-innovation identification theory.

Suppose that  $\hat{X}(t)$  is the estimate of  $X$  at time  $t$ . That is to say  $\hat{c}_i(t)$ ,  $\hat{\gamma}(t)$  and  $\hat{w}(t)$  are estimates of  $c_i$ ,  $\gamma$  and  $w$  at time  $t$ , respectively.

### 3. The Partially Coupled Stochastic Gradient Algorithm

First of all, according to the number of the system outputs, decomposing the identification model in (5) into  $m$  sub-identification models:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{bmatrix} = \begin{bmatrix} \psi_1^T(t) \\ \psi_2^T(t) \\ \vdots \\ \psi_m^T(t) \end{bmatrix} \theta + \begin{bmatrix} \gamma_1^T \\ \gamma_2^T \\ \vdots \\ \gamma_m^T \end{bmatrix} \phi(t) + \begin{bmatrix} v_1(t) \\ v_2(t) \\ \vdots \\ v_m(t) \end{bmatrix}, \quad i = 1, 2, \dots, m, \quad (9)$$

where  $\gamma_i^T \in \mathbb{R}^{1 \times mn_c}$  are the  $i$ th row of the parameter matrix  $\gamma^T$ , and  $\psi_i^T(t) \in \mathbb{R}^{1 \times n}$  are the  $i$ th row of the information matrix  $\Phi_s(t)$ :

$$\begin{aligned} \gamma^T &:= [\gamma_1, \gamma_2, \dots, \gamma_m]^T \in \mathbb{R}^{m \times (mn_c)}, \\ \Phi_s(t) &:= [\psi_1(t), \psi_2(t), \dots, \psi_m(t)]^T \in \mathbb{R}^{m \times n}. \end{aligned}$$

For the  $m$  sub-identification models in (9), it can be seen that the information vector  $\theta$  and the parameter vector  $\phi(t)$  are common among the  $m$  subsystems. Thus, Equation (5) is a coupling identification model of partial parameter vectors and partial information vectors. Equation (9) can be represented as

$$\begin{aligned} y_i(t) &= \psi_i^T(t)\theta + \gamma_i^T\phi(t) + v_i(t) \\ &= \psi_i^T(t)\theta + \phi^T(t)\gamma_i + v_i(t), \quad i = 1, 2, \dots, m. \end{aligned} \quad (10)$$

For the  $m$  sub-identification models in (10), define a gradient criterion function,

$$J_1(\theta, \gamma_i) := [y_i(t) - \psi_i^T(t)\theta - \phi^T(t)\gamma_i]^2, \quad i = 1, 2, \dots, m.$$

Assuming that  $1/r_{\theta,i}(t)$  and  $1/r_{\gamma,i}(t)$  ( $i = 1, 2, \dots, m$ ) are the step-size, using the negative gradient search [33] and minimizing  $J_1(\theta, \gamma_i)$ , we have the gradient relationships:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\psi_i(t)}{r_{\theta,i}(t)} [y_i(t) - \psi_i^T(t)\hat{\theta}(t-1) - \phi^T(t)\hat{\gamma}_i(t-1)], \quad (11)$$

$$r_{\theta,i}(t) = r_{\theta,i}(t-1) + \|\psi_i(t)\|^2, \quad r_{\theta,i}(0) = 1, \quad (12)$$

$$\hat{\gamma}_i(t) = \hat{\gamma}_i(t-1) + \frac{\phi(t)}{r_{\gamma,i}(t)} [y_i(t) - \psi_i^T(t)\hat{\theta}(t-1) - \phi^T(t)\hat{\gamma}_i(t-1)], \quad (13)$$

$$r_{\gamma,i}(t) = r_{\gamma,i}(t-1) + \|\phi(t)\|^2, \quad r_{\gamma,i}(0) = 1. \quad (14)$$

The problem of identification in (11)–(14) is that the estimates  $\hat{\theta}(t)$  and  $\hat{\gamma}_i(t)$  can not be computed because  $\phi(t)$  contains the unmeasurable noise terms  $w(t-i)$ . The method to solve this problem is to replace the unmeasurable variables  $w(t-i)$  with their corresponding estimates  $\hat{w}(t-i)$ . Thus, the estimate of  $\phi(t)$  can be computed by  $\hat{w}(t-i)$ :

$$\hat{\phi}(t) := [-\hat{w}^T(t-1), -\hat{w}^T(t-2), \dots, -\hat{w}^T(t-n_c)]^T \in \mathbb{R}^{mn_c}. \quad (15)$$

It can be seen that the parameter vector  $\theta$  is estimated for  $m$  times in (11) and (12), which will lead to many redundant estimates. Using the estimate  $\hat{\theta}_i$  to replace the estimate  $\hat{\theta}$  in (11) can reduce the redundant computation. At the same time, replacing the

unknown information vector  $\phi(t)$  with its estimate  $\hat{\phi}(t)$ , we can obtain the new gradient relationships:

$$\hat{\theta}_i(t) = \hat{\theta}_i(t-1) + \frac{\psi_i(t)}{r_{\theta,i}(t)} [y_i(t) - \psi_i^T(t)\hat{\theta}_i(t-1) - \hat{\phi}^T(t)\hat{\gamma}_i(t-1)], \quad (16)$$

$$r_{\theta,i}(t) = r_{\theta,i}(t-1) + \|\psi_i(t)\|^2, \quad r_{\theta,i}(0) = 1, \quad (17)$$

$$\hat{\gamma}_i(t) = \hat{\gamma}_i(t-1) + \frac{\hat{\phi}(t)}{r_{\gamma,i}(t)} [y_i(t) - \psi_i^T(t)\hat{\theta}_i(t-1) - \hat{\phi}^T(t)\hat{\gamma}_i(t-1)], \quad (18)$$

$$r_{\gamma,i}(t) = r_{\gamma,i}(t-1) + \|\hat{\phi}(t)\|^2, \quad r_{\gamma,i}(0) = 1. \quad (19)$$

Additionally, according to Equations (4), we have

$$\hat{w}(t) = y(t) - \Phi_s(t)\hat{\theta}_m(t). \quad (20)$$

It is generally believed that the parameter estimate  $\hat{\theta}_{i-1}(t)$  of the  $(i-1)$ th subsystem at time  $t$  is closer to the true value  $\theta$  than the parameter estimate  $\hat{\theta}_i(t-1)$  of the  $i$ th subsystem at time  $(t-1)$ . Replacing  $\hat{\theta}_i(t-1)$  on the right-hand of Equation (16) with  $\hat{\theta}_{i-1}(t)$ , and replacing  $\hat{\theta}_1(t-1)$  of Equation (16) when  $i = 1$  with  $\hat{\theta}_m(t-1)$ . Combining Equations (15) and (20), we can obtain the multivariate partially coupled generalized stochastic gradient (M-PC-GSG) algorithm as

$$\hat{\theta}_1(t) = \hat{\theta}_m(t-1) + \frac{\psi_1(t)}{r_{\theta,1}(t)} [y_1(t) - \psi_1^T(t)\hat{\theta}_m(t-1) - \hat{\phi}^T(t)\hat{\gamma}_1(t-1)], \quad (21)$$

$$r_{\theta,1}(t) = r_{\theta,m}(t-1) + \|\psi_1(t)\|^2, \quad (22)$$

$$\hat{\gamma}_1(t) = \hat{\gamma}_1(t-1) + \frac{\hat{\phi}(t)}{r_{\gamma,1}(t)} [y_1(t) - \psi_1^T(t)\hat{\theta}_m(t-1) - \hat{\phi}^T(t)\hat{\gamma}_1(t-1)], \quad (23)$$

$$r_{\gamma,1}(t) = r_{\gamma,m}(t-1) + \|\hat{\phi}(t)\|^2, \quad (24)$$

$$\hat{\theta}_i(t) = \hat{\theta}_{i-1}(t) + \frac{\psi_i(t)}{r_{\theta,i}(t)} [y_i(t) - \psi_i^T(t)\hat{\theta}_{i-1}(t) - \hat{\phi}^T(t)\hat{\gamma}_i(t-1)], \quad (25)$$

$$r_{\theta,i}(t) = r_{\theta,i-1}(t) + \|\psi_i(t)\|^2, \quad (26)$$

$$\hat{\gamma}_i(t) = \hat{\gamma}_i(t-1) + \frac{\hat{\phi}(t)}{r_{\gamma,i}(t)} [y_i(t) - \psi_i^T(t)\hat{\theta}_{i-1}(t) - \hat{\phi}^T(t)\hat{\gamma}_i(t-1)], \quad (27)$$

$$r_{\gamma,i}(t) = r_{\gamma,i-1}(t) + \|\hat{\phi}(t)\|^2, \quad (28)$$

$$\Phi_s(t) = [\psi_1(t), \psi_2(t), \dots, \psi_m(t)]^T, \quad (29)$$

$$\hat{\phi}(t) = [-\hat{w}^T(t-1), -\hat{w}^T(t-2), \dots, -\hat{w}^T(t-n_c)]^T, \quad (30)$$

$$\hat{w}(t) = y(t) - \Phi_s(t)\hat{\theta}_m(t), \quad (31)$$

$$\hat{\gamma}(t) = [\hat{\gamma}_1(t), \hat{\gamma}_2(t), \dots, \hat{\gamma}_m(t)]. \quad (32)$$

The schematic diagram of the M-PC-GSG algorithm in (21)–(32) is shown in Figure 1, and the computation procedures are listed as follows.

1. Let  $t = 1$ , set the initial values  $\hat{\theta}_m(0) = \mathbf{1}_n/p_0$ ,  $r_{\theta,i}(0) = r_{\gamma,i}(0) = 1$ ,  $\hat{\gamma}_i(0) = \mathbf{1}_{mn_c}/p_0$ ,  $i = 1, 2, \dots, m$ ,  $\hat{w}(t-j) = \mathbf{0}$ ,  $j = 0, 1, \dots, n_c$ ,  $p_0 = 10^6$ , and set the data length  $K$ .
2. Collect the observation data  $\Phi_s(t)$  and  $y(t)$ , read  $\psi_i(t)$  from  $\Phi_s(t)$  using (29).
3. Construct  $\hat{\phi}(t)$  using (30).
4. Compute the step-size  $r_{\theta,1}(t)$  and  $r_{\gamma,1}(t)$  using (22) and (24).
5. Update the parameter estimates  $\hat{\theta}_1(t)$  and  $\hat{\gamma}_1(t)$  using (21) and (23).
6. When  $i = 2, 3, \dots, m$ , compute  $r_{\theta,i}(t)$  and  $r_{\gamma,i}(t)$  using (26) and (28), and update the parameter estimates  $\hat{\theta}_i(t)$  and  $\hat{\gamma}_i(t)$  using (25) and (27).
7. Compute  $\hat{w}(t)$  using (31) and construct  $\hat{\gamma}(t)$  using (32).
8. Increase  $t$  by 1 if  $t < K$ , and then go to Step 2. Otherwise, obtain parameter estimates  $\hat{\theta}(t)$  and  $\hat{\gamma}(t)$  and stop computing.

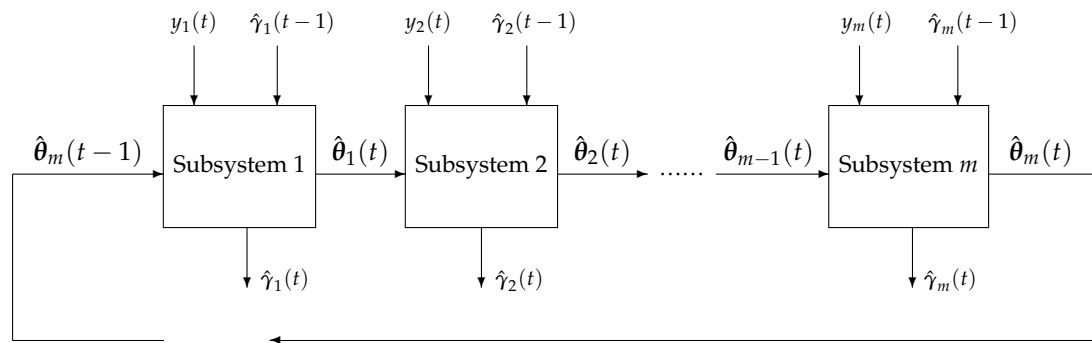


Figure 1. The schematic diagram of the M-PC-GSG algorithm.

#### 4. The Partially Coupled Multi-Innovation Stochastic Gradient Algorithm

In this section, we apply the multi-innovation identification theory based on the M-PC-GSG algorithm to further improve the parameter estimation accuracy. We introduce an innovation length  $p \geq 1$  to expand the innovation scalars of each subsystem models to the innovation vectors. According to the M-PC-GSG algorithm in (21)–(32), define the subsystem stacked information matrices  $\Lambda_1(p, t)$ ,  $\Lambda_i(p, t)$  ( $i = 2, 3, \dots, m$ ),  $\Xi(p, t)$  and the subsystem stacked output vectors  $Y_1(p, t)$  and  $Y_i(p, t)$  ( $i = 2, 3, \dots, m$ ) as

$$\begin{aligned}\Lambda_1(p, t) &:= [\psi_1(t), \psi_1(t-1), \dots, \psi_1(t-p+1)] \in \mathbb{R}^{n \times p}, \\ \Lambda_i(p, t) &:= [\psi_i(t), \psi_i(t-1), \dots, \psi_i(t-p+1)] \in \mathbb{R}^{n \times p}, i = 2, 3, \dots, m, \\ \Xi(p, t) &:= [\hat{\phi}(t), \hat{\phi}(t-1), \dots, \hat{\phi}(t-p+1)] \in \mathbb{R}^{mn_c \times p}, \\ Y_1(p, t) &:= [y_1(t), y_1(t-1), \dots, y_1(t-p+1)]^T \in \mathbb{R}^p, \\ Y_i(p, t) &:= [y_i(t), y_i(t-1), \dots, y_i(t-p+1)]^T \in \mathbb{R}^p, i = 2, 3, \dots, m.\end{aligned}$$

In general, define the subsystem innovation scalars  $e_1(t)$  and  $e_i(t)$  ( $i = 2, 3, \dots, m$ ) in the M-PC-GSG algorithm in (21)–(32) as

$$\begin{aligned}e_1(t) &:= y_1(t) - \psi_1^T(t) \hat{\theta}_m(t-1) - \hat{\phi}^T(t) \hat{\gamma}_1(t-1) \in \mathbb{R}, \\ e_i(t) &:= y_i(t) - \psi_i^T(t) \hat{\theta}_{i-1}(t) - \hat{\phi}^T(t) \hat{\gamma}_i(t-1) \in \mathbb{R}, i = 2, 3, \dots, m.\end{aligned}$$

According to the multi-innovation identification theory, expand the subsystem innovation scalars  $e_1(t)$  and  $e_i(t)$  ( $i = 2, 3, \dots, m$ ) into the subsystem innovation vectors  $E_1(p, t)$  and  $E_i(p, t)$  ( $i = 2, 3, \dots, m$ ):

$$\begin{aligned}E_1(p, t) &:= \begin{bmatrix} e_1(t) \\ e_1(t-1) \\ \vdots \\ e_1(t-p+1) \end{bmatrix} \\ &= \begin{bmatrix} y_1(t) - \psi_1^T(t) \hat{\theta}_m(t-1) - \hat{\phi}^T(t) \hat{\gamma}_1(t-1) \\ y_1(t-1) - \psi_1^T(t-1) \hat{\theta}_m(t-2) - \hat{\phi}^T(t-1) \hat{\gamma}_1(t-2) \\ \vdots \\ y_1(t-p+1) - \psi_1^T(t-p+1) \hat{\theta}_m(t-p) - \hat{\phi}^T(t-p+1) \hat{\gamma}_1(t-p) \end{bmatrix} \in \mathbb{R}^p, \quad (33)\end{aligned}$$

$$\begin{aligned}E_i(p, t) &:= \begin{bmatrix} e_i(t) \\ e_i(t-1) \\ \vdots \\ e_i(t-p+1) \end{bmatrix} \quad (i = 2, 3, \dots, m) \\ &= \begin{bmatrix} y_i(t) - \psi_i^T(t) \hat{\theta}_{i-1}(t) - \hat{\phi}^T(t) \hat{\gamma}_i(t-1) \\ y_i(t-1) - \psi_i^T(t-1) \hat{\theta}_{i-1}(t-1) - \hat{\phi}^T(t-1) \hat{\gamma}_i(t-2) \\ \vdots \\ y_i(t-p+1) - \psi_i^T(t-p+1) \hat{\theta}_{i-1}(t-p+1) - \hat{\phi}^T(t-p+1) \hat{\gamma}_i(t-p) \end{bmatrix} \in \mathbb{R}^p. \quad (34)\end{aligned}$$

Normally, we reach an agreement that the estimates  $\hat{\theta}_m(t-1)$  and  $\hat{\gamma}_i(t-1)$  ( $i = 1, 2, \dots, m$ ) at time  $(t-1)$  are closer to the true values  $\theta_m$  and  $\gamma_i$  than the estimates  $\hat{\theta}_m(t-j)$  and  $\hat{\gamma}_i(t-j)$  at time  $(t-j)$  ( $j \geq 2$ ). Similarly, the estimate  $\hat{\theta}_{i-1}(t)$  at time  $t$  is closer to the true value  $\theta_{i-1}$  than the estimate  $\hat{\theta}_{i-1}(t-j)$  at time  $(t-j)$  ( $j \geq 2$ ). Therefore, replacing the terms  $\hat{\theta}_m(t-j)$ ,  $\hat{\gamma}_i(t-j)$  and  $\hat{\theta}_{i-1}(t-j)$  ( $j \geq 2$ ) with  $\hat{\theta}_m(t-1)$ ,  $\hat{\gamma}_i(t-1)$  and  $\hat{\theta}_{i-1}(t)$  in (33) and (34), and the subsystem innovation vectors  $E_1(p, t)$  and  $E_i(p, t)$  ( $i = 2, 3, \dots, m$ ) are modified into

$$E_1(p, t) := \begin{bmatrix} y_1(t) - \psi_1^T(t)\hat{\theta}_m(t-1) - \hat{\phi}^T(t)\hat{\gamma}_1(t-1) \\ y_1(t-1) - \psi_1^T(t-1)\hat{\theta}_m(t-1) - \hat{\phi}^T(t-1)\hat{\gamma}_1(t-1) \\ \vdots \\ y_1(t-p+1) - \psi_1^T(t-p+1)\hat{\theta}_m(t-1) - \hat{\phi}^T(t-p+1)\hat{\gamma}_1(t-1) \end{bmatrix} \\ = Y_1(p, t) - \Lambda_1^T(p, t)\hat{\theta}_m(t-1) - \Xi^T(p, t)\hat{\gamma}_1(t-1) \in \mathbb{R}^p, \quad (35)$$

$$E_i(p, t) := \begin{bmatrix} y_i(t) - \psi_i^T(t)\hat{\theta}_{i-1}(t) - \hat{\phi}^T(t)\hat{\gamma}_i(t-1) \\ y_i(t-1) - \psi_i^T(t-1)\hat{\theta}_{i-1}(t) - \hat{\phi}^T(t-1)\hat{\gamma}_i(t-1) \\ \vdots \\ y_i(t-p+1) - \psi_i^T(t-p+1)\hat{\theta}_{i-1}(t) - \hat{\phi}^T(t-p+1)\hat{\gamma}_i(t-1) \end{bmatrix}, \\ = Y_i(p, t) - \Lambda_i^T(p, t)\hat{\theta}_{i-1}(t) - \Xi^T(p, t)\hat{\gamma}_i(t-1) \in \mathbb{R}^p, \quad i = 2, 3, \dots, m. \quad (36)$$

Thus, based on the the M-PC-GSG algorithm in (21)–(32), we can obtain the following multivariate partially coupled multi-innovation generalized stochastic gradient (M-PC-MI-GSG) algorithm with an innovation length  $p$ :

$$\hat{\theta}_1(t) = \hat{\theta}_m(t-1) + \frac{\Lambda_1(p, t)}{r_{\theta,1}(t)} E_1(p, t), \quad (37)$$

$$Y_1(p, t) = [y_1(t), y_1(t-1), \dots, y_1(t-p+1)]^T, \quad (38)$$

$$\Lambda_1(p, t) = [\psi_1(t), \psi_1(t-1), \dots, \psi_1(t-p+1)], \quad (39)$$

$$\Xi(p, t) = [\hat{\phi}(t), \hat{\phi}(t-1), \dots, \hat{\phi}(t-p+1)], \quad (40)$$

$$E_1(p, t) = Y_1(p, t) - \Lambda_1^T(p, t)\hat{\theta}_m(t-1) - \Xi^T(p, t)\hat{\gamma}_1(t-1), \quad (41)$$

$$r_{\theta,1}(t) = r_{\theta,m}(t-1) + \|\psi_1(t)\|^2, \quad (42)$$

$$\hat{\gamma}_1(t) = \hat{\gamma}_1(t-1) + \frac{\Xi(p, t)}{r_{\gamma,1}(t)} E_1(p, t), \quad (43)$$

$$r_{\gamma,1}(t) = r_{\gamma,m}(t-1) + \|\hat{\phi}(t)\|^2, \quad (44)$$

$$\hat{\theta}_i(t) = \hat{\theta}_{i-1}(t) + \frac{\Lambda_i(p, t)}{r_{\theta,i}(t)} E_i(p, t), \quad (45)$$

$$Y_i(p, t) = [y_i(t), y_i(t-1), \dots, y_i(t-p+1)]^T, \quad (46)$$

$$\Lambda_i(p, t) = [\psi_i(t), \psi_i(t-1), \dots, \psi_i(t-p+1)], \quad (47)$$

$$E_i(p, t) = Y_i(p, t) - \Lambda_i^T(p, t)\hat{\theta}_{i-1}(t) - \Xi^T(p, t)\hat{\gamma}_i(t-1), \quad (48)$$

$$r_{\theta,i}(t) = r_{\theta,i-1}(t) + \|\psi_i(t)\|^2, \quad (49)$$

$$\hat{\gamma}_i(t) = \hat{\gamma}_i(t-1) + \frac{\Xi(p, t)}{r_{\gamma,i}(t)} E_i(p, t), \quad (50)$$

$$r_{\gamma,i}(t) = r_{\gamma,i-1}(t) + \|\hat{\phi}(t)\|^2, \quad (51)$$

$$\Phi_s(t) = [\psi_1(t), \psi_2(t), \dots, \psi_m(t)]^T, \quad (52)$$

$$\hat{\phi}(t) = [-\hat{w}^T(t-1), -\hat{w}^T(t-2), \dots, -\hat{w}^T(t-n_c)]^T, \quad (53)$$

$$\hat{w}(t) = y(t) - \Phi_s(t)\hat{\theta}_m(t), \quad (54)$$

$$\hat{\gamma}(t) = [\hat{\gamma}_1(t), \hat{\gamma}_2(t), \dots, \hat{\gamma}_m(t)]. \quad (55)$$

The computation procedures of the M-PC-MI-GSG algorithm in (37)–(55) are listed as follows.



1. Let  $t = 1$ , choose an innovation length  $p$ , set the initial values  $\hat{\theta}_m(0) = \mathbf{1}_n/p_0$ ,  $r_{\theta,i}(0) = r_{\gamma,i}(0) = 1$ ,  $\hat{\gamma}_i(0) = \mathbf{1}_{mn_c}/p_0$ ,  $i = 1, 2, \dots, m$ ,  $\hat{w}(t-j) = \mathbf{0}$ ,  $j = 0, 1, \dots, n_c$ ,  $p_0 = 10^6$ , and set the data length  $K$ .
2. Collect the observation data  $\Phi_s(t)$  and  $y(t)$ , read  $\psi_i(t)$  from  $\Phi_s(t)$  using (52).
3. Construct  $\hat{\Phi}(t)$  using (53).
4. Construct  $Y_1(p, t)$ ,  $\Lambda_1(p, t)$  and  $\Xi(p, t)$  using (38)–(40).
5. Compute  $E_1(p, t)$  using (41).
6. Compute  $r_{\theta,1}(t)$  and  $r_{\gamma,1}(t)$  using (42) and (44).
7. Update the parameter estimates  $\hat{\theta}_1(t)$  and  $\hat{\gamma}_1(t)$  using (37) and (43).
8. When  $i = 2, 3, \dots, m$ , construct  $Y_i(p, t)$  and  $\Lambda_i(p, t)$  using (46) and (47).
9. Compute  $E_i(p, t)$  using (48).
10. Compute  $r_{\theta,i}(t)$  and  $r_{\gamma,i}(t)$  using (49) and (51).
11. Update the parameter estimates  $\hat{\theta}_i(t)$  and  $\hat{\gamma}_i(t)$  using (45) and (50).
12. Compute  $\hat{w}(t)$  using (54) and construct  $\hat{\gamma}(t)$  using (55).
13. Increase  $t$  by 1 if  $t < K$ , and then go to Step 2. Otherwise, obtain parameter estimates  $\hat{\theta}(t)$  and  $\hat{\gamma}(t)$  and stop computing.

It can be easily seen that the M-PC-MI-GSG algorithm with the innovation length  $p = 1$  is equal to the M-PC-GSG algorithm. Since the past data information of the system are utilized sufficiently, the M-PC-MI-GSG algorithm has higher parameter estimation accuracy than the M-PC-GSG algorithm.

## 5. The Simulation Examples

In this section, two numerical simulations are given to show the good performances of the newly proposed algorithms.

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

$$\begin{aligned}
 y(t) &= \Phi_s(t)\theta + C^{-1}(z)v(t), \\
 \Phi_s(t) &= \begin{bmatrix} \sin(u_1(t-2))y_1(t-1) & \cos(u_1(t-1))y_1(t-2) \\ \cos(u_1(t-2))y_2(t-1) & \sin(u_1(t-2))y_2(t-2) \\ \sin(u_2(t-2))y_2(t-1) & \cos(u_2(t-1))y_2(t-2) \\ \cos(u_2(t-2))y_1(t-1) & \sin(u_2(t-1))y_1(t-2) \end{bmatrix}, \\
 \theta &= [\theta_1, \theta_2, \theta_3, \theta_4]^T = [0.42, 0.93, 0.56, 0.31]^T, \\
 C(z) &= I_2 + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} z^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} -0.25 & 0.68 \\ -0.33 & 0.44 \end{bmatrix} z^{-1}.
 \end{aligned}$$

The parameter vector to be estimated is

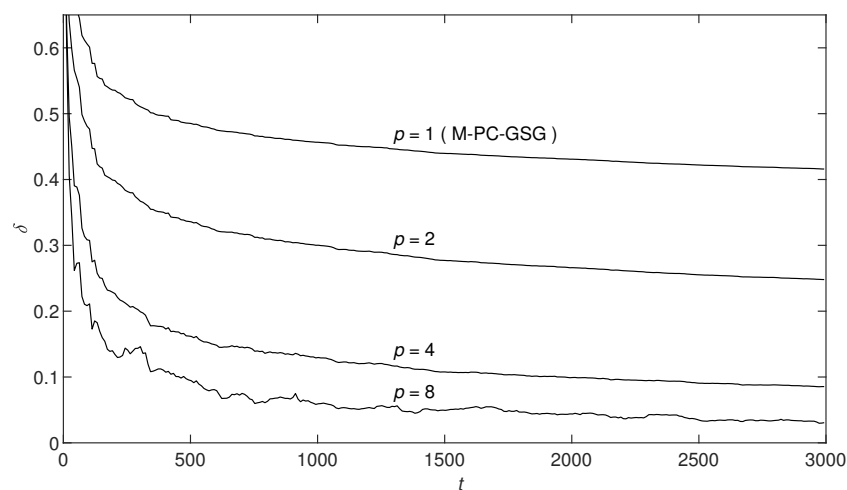
$$\begin{aligned}
 \vartheta &= [\theta_1, \theta_2, \theta_3, \theta_4, c_{11}, c_{12}, c_{21}, c_{22}]^T \\
 &= [0.42, 0.93, 0.56, 0.31, -0.25, 0.68, -0.33, 0.44]^T.
 \end{aligned}$$

In this simulation,  $u(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} \in \mathbb{R}^2$  is the input vector, which is a random sequence with zero mean and variance one;  $y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} \in \mathbb{R}^2$  is the output vector;  $v(t) = \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} \in \mathbb{R}^2$  is a white noise vector with zero mean;  $\sigma_1^2$  and  $\sigma_2^2$  are the variances of  $v_1(t)$  and  $v_2(t)$ . Taking the noise variances  $\sigma_1^2 = 0.40^2$  and  $\sigma_2^2 = 0.30^2$ , using the M-PC-GSG algorithm (i.e., the M-PC-MI-GSG algorithm with  $p = 1$ ) and the M-PC-MI-GSG algorithm with  $p = 2$ ,  $p = 4$  and  $p = 8$  to estimate the parameters of this example system. We obtain the parameter estimates and their errors  $\delta := \|\hat{\vartheta}(t) - \vartheta\| / \|\vartheta\|$  shown in Table 1. The parameter estimation errors versus  $t$  are shown in Figure 2. The parameter estimates  $\hat{\theta}_1(t)$ ,  $\hat{\theta}_2(t)$ ,  $\hat{\theta}_3(t)$ ,  $\hat{\theta}_4(t)$  and  $\hat{c}_{11}(t)$ ,  $\hat{c}_{12}(t)$ ,  $\hat{c}_{21}(t)$ ,  $\hat{c}_{22}(t)$  versus  $t$  of the M-PC-MI-GSG algorithm with  $p = 8$  are shown in Figures 3 and 4.



**Table 1.** Parameter estimates and errors ( $\sigma_1^2 = 0.40^2, \sigma_2^2 = 0.30^2$ ).

Algorithms	$t$	100	200	500	1000	2000	3000	True Values
M-PC-GSG	$\theta_1$	0.19022	0.28281	0.33788	0.36637	0.36787	0.36879	0.42000
	$\theta_2$	0.77045	0.84529	0.88400	0.91050	0.90956	0.90832	0.93000
	$\theta_3$	0.42624	0.49137	0.54550	0.57360	0.57695	0.57314	0.56000
	$\theta_4$	0.25045	0.25999	0.29032	0.31802	0.32379	0.32917	0.31000
	$c_{11}$	−0.16279	−0.13896	−0.12959	−0.13205	−0.13608	−0.13540	−0.25000
	$c_{12}$	−0.06264	0.01024	0.07393	0.10879	0.14383	0.16385	0.68000
	$c_{21}$	−0.18783	−0.21131	−0.22760	−0.22982	−0.23502	−0.23697	−0.33000
	$c_{22}$	0.06255	0.05905	0.07305	0.09172	0.10850	0.12021	0.44000
	$\delta$ (%)	60.04604	53.54881	48.52141	45.66110	43.08714	41.59002	
M-PC-MI-GSG ( $p = 2$ )	$\theta_1$	0.24413	0.35242	0.38045	0.39205	0.38805	0.38809	0.42000
	$\theta_2$	0.80411	0.89433	0.91071	0.92629	0.92058	0.91797	0.93000
	$\theta_3$	0.49437	0.54833	0.57802	0.58115	0.58224	0.57102	0.56000
	$\theta_4$	0.30721	0.28679	0.29391	0.32301	0.32347	0.32964	0.31000
	$c_{11}$	−0.05715	−0.06016	−0.08608	−0.10747	−0.13040	−0.13492	−0.25000
	$c_{12}$	0.12099	0.22052	0.30492	0.34234	0.38260	0.40361	0.68000
	$c_{21}$	−0.23138	−0.26019	−0.26756	−0.26447	−0.26951	−0.27084	−0.33000
	$c_{22}$	0.11762	0.11798	0.15107	0.18454	0.20984	0.22701	0.44000
	$\delta$ (%)	47.51695	39.87369	33.61530	30.01753	26.59647	24.80351	
M-PC-MI-GSG ( $p = 4$ )	$\theta_1$	0.26824	0.39655	0.39206	0.40527	0.40117	0.40344	0.42000
	$\theta_2$	0.81764	0.91822	0.91009	0.93027	0.92223	0.92506	0.93000
	$\theta_3$	0.53011	0.56224	0.57291	0.58006	0.58163	0.56221	0.56000
	$\theta_4$	0.33781	0.28934	0.27962	0.32670	0.31900	0.33021	0.31000
	$c_{11}$	−0.02859	−0.07525	−0.14398	−0.17985	−0.21158	−0.20979	−0.25000
	$c_{12}$	0.40063	0.49444	0.57042	0.57794	0.60699	0.61726	0.68000
	$c_{21}$	−0.27689	−0.30473	−0.29946	−0.29170	−0.29725	−0.29909	−0.33000
	$c_{22}$	0.22503	0.21310	0.25734	0.29682	0.32439	0.34271	0.44000
	$\delta$ (%)	30.60589	22.77807	16.20218	12.94875	9.87303	8.55799	
M-PC-MI-GSG ( $p = 8$ )	$\theta_1$	0.22875	0.41123	0.39312	0.41067	0.40847	0.41377	0.42000
	$\theta_2$	0.82538	0.91675	0.89962	0.92660	0.91916	0.93461	0.93000
	$\theta_3$	0.50062	0.56129	0.56967	0.58480	0.57905	0.55554	0.56000
	$\theta_4$	0.34784	0.29455	0.27015	0.32493	0.31639	0.33324	0.31000
	$c_{11}$	−0.08032	−0.14448	−0.22262	−0.25289	−0.27162	−0.25101	−0.25000
	$c_{12}$	0.60874	0.66577	0.71758	0.67480	0.69550	0.69273	0.68000
	$c_{21}$	−0.31148	−0.33276	−0.30958	−0.29949	−0.30552	−0.31357	−0.33000
	$c_{22}$	0.32292	0.26569	0.31977	0.36227	0.39067	0.40749	0.44000
	$\delta$ (%)	20.99172	13.61264	9.45336	5.90521	4.39162	3.04341	

**Figure 2.** The M-PC-MI-GSG estimation errors versus  $t$ .

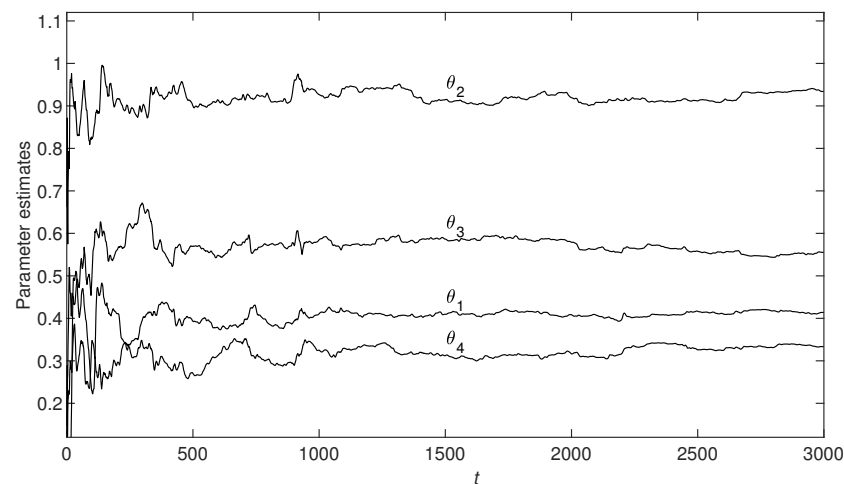


Figure 3. Parameter estimates  $\hat{\theta}_1(t)$ ,  $\hat{\theta}_2(t)$ ,  $\hat{\theta}_3(t)$ ,  $\hat{\theta}_4(t)$  versus  $t$ .

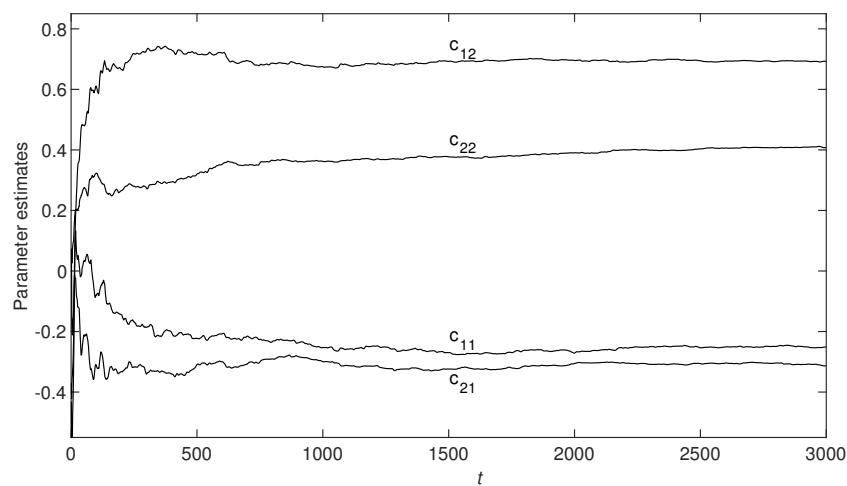


Figure 4. Parameter estimates  $\hat{c}_{11}(t)$ ,  $\hat{c}_{12}(t)$ ,  $\hat{c}_{21}(t)$ ,  $\hat{c}_{22}(t)$  versus  $t$ .

**Example 2.** Consider another multivariate equation-error autoregressive systems,

$$\begin{aligned}
 y(t) &= \Phi_s(t)\theta + C^{-1}(z)v(t), \\
 \Phi_s(t) &= \begin{bmatrix} u_1(t-1)y_1(t-2) & u_1(t-2)u_2(t-2) & u_2(t-1)\cos(t) \\ u_1(t-2)y_1(t-1) & y_1(t-2)y_2(t-2) & y_2(t-1)\sin(t) \\ \cos^2(t-1) & u_2(t-2)\sin(t-1) & \cos(u_1(t-1)) \\ \sin^2(t-1) & y_2(t-2)\cos(t-1) & \sin(y_1(t-1)) \end{bmatrix}, \\
 \theta &= [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6]^T = [-0.36, 0.22, 0.34, 0.45, 0.25, 0.11]^T, \\
 C(z) &= I_2 + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} z^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} -0.41 & -0.48 \\ 0.35 & -0.31 \end{bmatrix} z^{-1}.
 \end{aligned}$$

The parameter vector to be estimated is

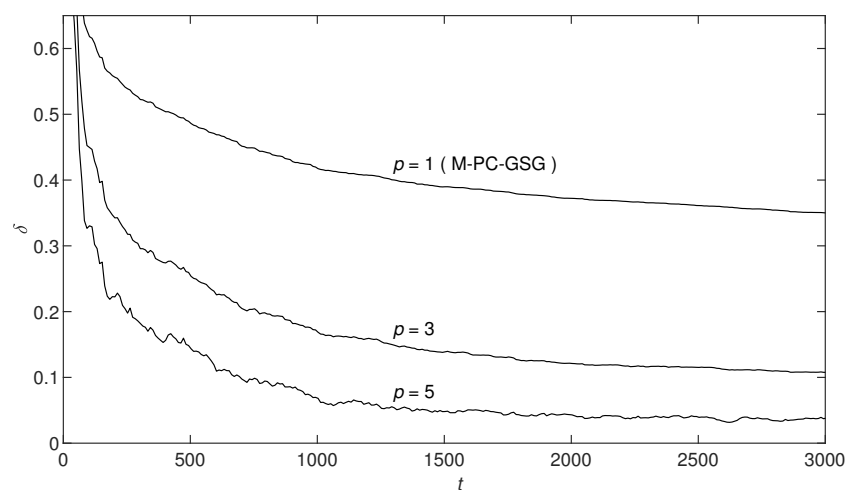
$$\begin{aligned}
 \vartheta &= [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, c_{11}, c_{12}, c_{21}, c_{22}]^T \\
 &= [-0.36, 0.22, 0.34, 0.45, 0.25, 0.11, -0.41, -0.48, 0.35, -0.31]^T.
 \end{aligned}$$

Here, the simulation conditions are similar to Example 1. Taking the noise variances  $\sigma_1^2 = \sigma_2^2 = 0.50^2$ , using the M-PC-GSG algorithm (i.e., the M-PC-MI-GSG algorithm with  $p = 1$ ) and the M-PC-MI-GSG algorithm with  $p = 3$  and  $p = 5$  to estimate the parameters of this example system. The parameter estimates and their errors are shown in Table 2. The parameter estimation errors versus  $t$  are shown in Figure 5. For model validation, we use the estimated model obtained

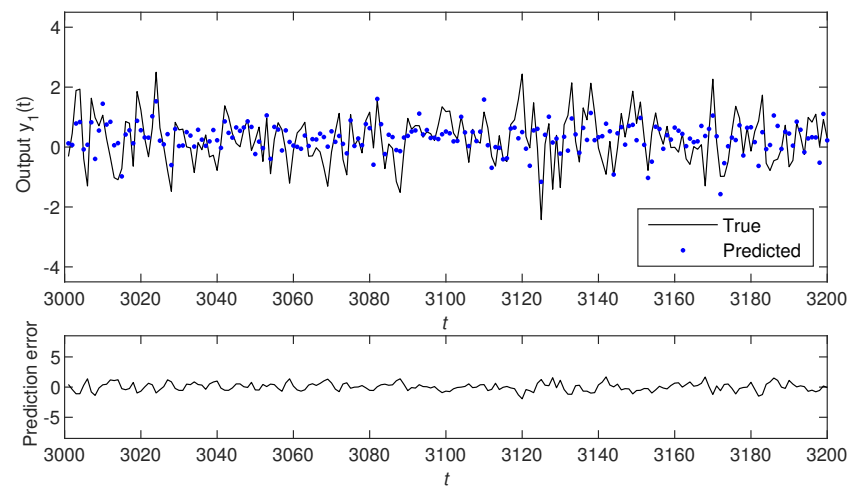
by the M-PC-MI-GSG algorithm with  $p = 5$  to predict the system outputs with 200 samples from  $t = 3001$  to  $t = 3200$ . The true output  $y_1(t)$  and the predicted output  $\hat{y}_1(t)$  as well as their errors are shown in Figure 6. The true output  $y_2(t)$  and the predicted output  $\hat{y}_2(t)$  as well as their errors are shown in Figure 7.

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

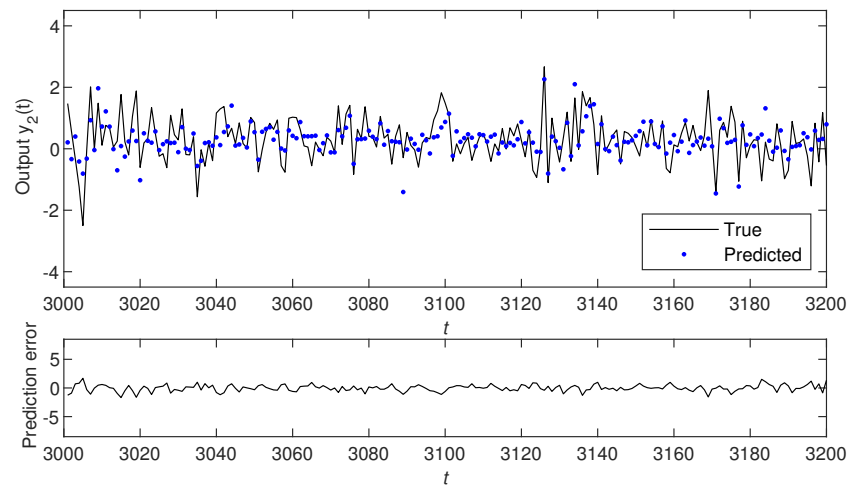
Algorithms	$t$	100	200	500	1000	2000	3000	True Values
M-PC-GSG	$\theta_1$	−0.38109	−0.34290	−0.35197	−0.35358	−0.35240	−0.35140	−0.36000
	$\theta_2$	0.17154	0.19202	0.22318	0.20571	0.20894	0.20290	0.22000
	$\theta_3$	0.52049	0.46240	0.42063	0.40418	0.38885	0.38130	0.34000
	$\theta_4$	0.36772	0.39041	0.40431	0.43059	0.44228	0.45084	0.45000
	$\theta_5$	0.32582	0.28799	0.27265	0.25708	0.26433	0.26468	0.25000
	$\theta_6$	−0.05956	−0.02720	−0.02478	−0.02008	−0.00934	−0.00876	0.11000
	$c_{11}$	−0.32562	−0.33413	−0.35093	−0.36678	−0.37198	−0.37339	−0.41000
	$c_{12}$	−0.42387	−0.42801	−0.43950	−0.44549	−0.45522	−0.45784	−0.48000
	$c_{21}$	−0.07449	−0.04447	0.00357	0.05541	0.08891	0.10599	0.35000
	$c_{22}$	0.13315	0.09440	0.04896	0.00007	−0.03206	−0.04982	−0.31000
	$\delta$ (%)	62.46630	55.62215	48.66825	41.81593	37.21977	34.99402	
M-PC-MI-GSG ( $p = 3$ )	$\theta_1$	−0.34351	−0.30662	−0.36709	−0.35952	−0.35025	−0.35053	−0.36000
	$\theta_2$	0.13616	0.21845	0.24620	0.20140	0.21882	0.20575	0.22000
	$\theta_3$	0.46447	0.34810	0.32286	0.33605	0.34131	0.33736	0.34000
	$\theta_4$	0.53467	0.51841	0.49198	0.50896	0.49490	0.49143	0.45000
	$\theta_5$	0.34052	0.24404	0.24150	0.23199	0.26443	0.26536	0.25000
	$\theta_6$	−0.08458	−0.00717	−0.00412	0.01422	0.05155	0.05032	0.11000
	$c_{11}$	−0.40747	−0.39750	−0.42113	−0.42870	−0.41996	−0.41446	−0.41000
	$c_{12}$	−0.44953	−0.45067	−0.47122	−0.47509	−0.49160	−0.49275	−0.48000
	$c_{21}$	−0.00019	0.05273	0.13849	0.22633	0.25865	0.27774	0.35000
	$c_{22}$	−0.10619	−0.14079	−0.18170	−0.23788	−0.25384	−0.26009	−0.31000
	$\delta$ (%)	45.05598	34.25524	25.50852	16.96953	12.12756	10.74341	
M-PC-MI-GSG ( $p = 5$ )	$\theta_1$	−0.35745	−0.30970	−0.38425	−0.36288	−0.34050	−0.34946	−0.36000
	$\theta_2$	0.12136	0.23250	0.25401	0.20572	0.22745	0.20850	0.22000
	$\theta_3$	0.40415	0.28647	0.30611	0.33618	0.34589	0.33781	0.34000
	$\theta_4$	0.51978	0.48285	0.46617	0.48824	0.46462	0.46024	0.45000
	$\theta_5$	0.38782	0.22572	0.22904	0.22473	0.27031	0.26832	0.25000
	$\theta_6$	−0.02008	0.06621	0.05334	0.06901	0.10837	0.09507	0.11000
	$c_{11}$	−0.41690	−0.39157	−0.43170	−0.43710	−0.42419	−0.41720	−0.41000
	$c_{12}$	−0.46440	−0.45497	−0.48430	−0.48122	−0.50181	−0.49994	−0.48000
	$c_{21}$	0.09086	0.14050	0.23086	0.31994	0.33401	0.34857	0.35000
	$c_{22}$	−0.24541	−0.23841	−0.25163	−0.30483	−0.29672	−0.29278	−0.31000
	$\delta$ (%)	32.60246	22.30989	14.47746	6.91170	4.31185	3.73785	



**Figure 5.** The M-PC-MI-GSG estimation errors versus  $t$ .



**Figure 6.** The true output, the predicted output of  $y_1(t)$  and the prediction error.



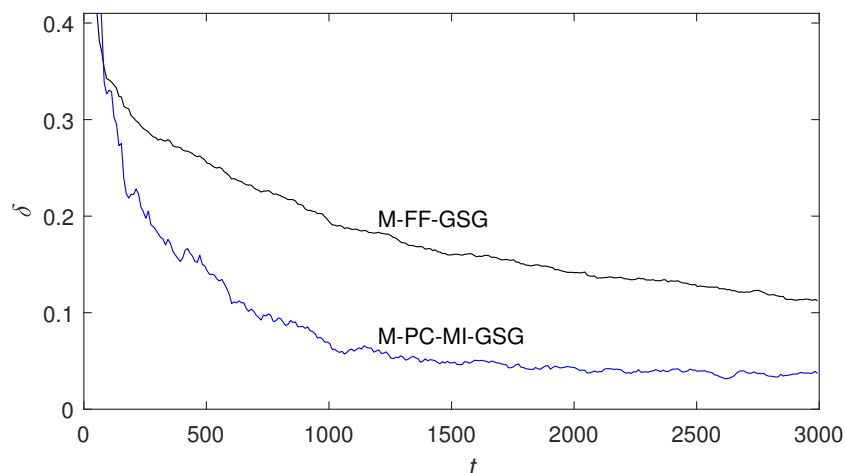
**Figure 7.** The true output, the predicted output of  $y_2(t)$  and the prediction error.

From Tables 1 and 2 and Figures 2–7, the following conclusions are obtained.

1. From Tables 1 and 2, Figures 2 and 5, it can be shown that the parameter estimation errors of the M-PC-GSG and the M-PC-MI-GSG algorithms become smaller as the data length  $t$  increases, which means that the proposed algorithms are effective in parameter estimation for the multivariate autoregressive system.
2. Figures 2 and 5 show that the M-PC-MI-GSG algorithm has higher parameter estimation accuracy than the M-PC-GSG algorithm under the same noise variances and same data length. Introducing the innovation length  $p$  can effectively improve the parameter estimation accuracy for the M-PC-GSG algorithm, and the parameter estimates can be more stationary as the innovation length  $p$  increases.
3. It can be seen from Figures 3 and 4 that the M-PC-MI-GSG algorithm can obtain accurate parameter estimates.
4. From Figures 6 and 7, we can see that the predicted outputs of the M-PC-MI-GSG algorithm are very close to the true outputs, which indicates that the estimated model can capture the dynamics of the system.

**Remark 1.** In order to show the advantages of the identification performances of the proposed algorithm in this paper, the forgetting factor stochastic gradient identification method proposed in [34] is compared with the proposed algorithm. The forgetting factor stochastic gradient identification method is applied to identify the multivariate equation-error autoregressive systems in this paper, and the multivariate forgetting factor generalized stochastic gradient (M-FF-GSG) algorithm is

obtained. The M-FF-GSG algorithm is compared with the M-PC-MI-GSG algorithm with  $p = 5$  through simulation, and the simulation experimental conditions are the same as those in Example 2. The comparison results are shown in Figure 8. It can be seen that the proposed algorithm in this paper has faster identification speed and higher estimation accuracy.



**Figure 8.** The comparison of algorithms M-FF-GSG with M-PC-MI-GSG.

## 6. Conclusions

Coupling identification concept is an emerging method in the field of system identification in recent decades, one which is usually used in the parameter estimation of the multivariate systems. Its main idea is to utilize the parameter-coupling characteristic, and to identify the parameter of each subsystem model separately and then connected, which can greatly reduce the calculation amounts in the estimation processes. This paper combines the coupling identification concept with the stochastic gradient identification method to propose a new identification algorithm for multivariate equation-error systems. The proposed algorithm has the advantage of a lower computation than the traditional stochastic gradient identification algorithm. Additionally, the multi-innovation identification theory is also a promising identification method, which makes full use of the data information collected in the past to identify the unknown parameters. Based on the partially coupled stochastic gradient algorithm, this paper then introduces the innovation length by applying the multi-innovation identification theory, and proposes the partially coupled multi-innovation stochastic gradient algorithm. The new algorithm also has the advantage of higher parameter estimation accuracy.

The proposed coupling and multi-innovation-based identification methods can be extended to study other multivariate systems with different structures and disturbance noises. Meanwhile, the idea of the algorithms can be utilized when the system identification model has the coupled terms. The future research opportunities are to apply the proposed algorithms to actual engineering production systems to improve the computational efficiency and accuracy of the system identification in production practice. Additionally, the proposed methods in the paper can also combine other mathematical tools and statistical strategies to research the performances of some parameter estimation algorithms for other linear or nonlinear systems with colored noises.

**Author Contributions:** Writing—original draft, P.M.; Writing—review and editing, L.W. 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. 61873111), the Fundamental Research Funds for the Central Universities (No. JUSRP121071), the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (No. 22KJB120009), and the Start-up Fund for Introducing Talent of Wuxi University (No. 2021r045).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data generated or analyzed during this study are included in this article.

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

## Abbreviations

The following abbreviations are used in this manuscript:

Abbreviations	Explanations
M-PC-GSG	the multivariate partially coupled generalized stochastic gradient algorithm
M-PC-MI-GSG	the multivariate partially coupled multi-innovation generalized stochastic gradient algorithm
M-FF-GSG	the multivariate forgetting factor generalized stochastic gradient algorithm

## References

- Na, J.; Yang, J.; Wu, X.; Guo, Y. Robust adaptive parameter estimation of sinusoidal signals. *Automatica* **2015**, *53*, 376–384. [\[CrossRef\]](#)
- Wang, J.; Efimov, D.; Bobtsov, A.A. On robust parameter estimation in finite-time without persistence of excitation. *IEEE Trans. Autom. Control* **2020**, *65*, 1731–1738. [\[CrossRef\]](#)
- Xu, L. Application of the Newton iteration algorithm to the parameter estimation for dynamical systems. *J. Comput. Appl. Math.* **2015**, *288*, 33–43. [\[CrossRef\]](#)
- Mo, D.; Duarte, M.F. Compressive parameter estimation via K-median. *Signal Process.* **2018**, *142*, 36–48. [\[CrossRef\]](#)
- Demirli, R.; Saniie, J. Asymmetric Gaussian chirplet model and parameter estimation for generalized echo representation. *J. Frankl. Inst.* **2014**, *351*, 907–921. [\[CrossRef\]](#)
- Upadhyay, R.K.; Paul, C.; Mondal, A.; Vishwakarma, G.K. Estimation of biophysical parameters in a neuron model under random fluctuations. *Appl. Math. Comput.* **2018**, *329*, 364–373. [\[CrossRef\]](#)
- Khalik, Z.; Donkers, M.C.F.; Sturm, J.; Bergveld, H.J. Parameter estimation of the Doyle-Fuller-Newman model for Lithium-ion batteries by parameter normalization, grouping, and sensitivity analysis. *J. Power Sources* **2021**, *499*, 229901. [\[CrossRef\]](#)
- Padmanabhan, C.; Gupta, S.; Mylswamy, A. Estimation of terramechanics parameters of wheel-soil interaction model using particle filtering. *J. Terramechanics* **2018**, *79*, 79–95.
- Calasan, M.P.; Jovanovic, A.; Rubezic, V.; Mujicic, D.; Deriszadeh, A. Notes on parameter estimation for single-phase transformer. *IEEE Trans. Ind. Appl.* **2020**, *56*, 3710–3718. [\[CrossRef\]](#)
- Zhang, X.; El Korso, M.N.; Pesavento, M. MIMO radar target localization and performance evaluation under SIRP clutter. *Signal Process.* **2017**, *130*, 217–232. [\[CrossRef\]](#)
- Kulikova, M.V.; Tsyganova, J.V.; Kulikov, G.Y. UD-based pairwise and MIMO kalman-like filtering for estimation of econometric model structures. *IEEE Trans. Autom. Control* **2020**, *65*, 4472–4479. [\[CrossRef\]](#)
- Weijtens, W.; Sitter, G.D.; Devriendt, C.; Guillaume, P. Operational modal parameter estimation of MIMO systems using transmissibility functions. *Automatica* **2014**, *50*, 559–564. [\[CrossRef\]](#)
- Wang, H.; Xu, L.W.; Yang, Z.Q.; Gulliver, T.A. Low-complexity MIMO-FBMC sparse channel parameter estimation for industrial big data communications. *IEEE Trans. Ind. Inform.* **2021**, *17*, 3422–3430. [\[CrossRef\]](#)
- Cerone, V.; Razza, V.; Regruto, D. Set-membership errors-in-variables identification of MIMO linear systems. *Automatica* **2018**, *90*, 25–37. [\[CrossRef\]](#)
- Liu, L.; Wang, Y.; Wang, C.; Ding, F.; Hayat, T. Maximum likelihood recursive least squares estimation for multivariate equation-error ARMA systems. *J. Frankl. Inst.* **2018**, *355*, 7609–7625. [\[CrossRef\]](#)
- Cecilio, I.M.; Ottewill, J.R.; Fretheim, H.; Thornhill, N.F. Multivariate detection of transient disturbances for uni- and multirate systems. *IEEE Trans. Control Syst. Technol.* **2015**, *23*, 1477–1493. [\[CrossRef\]](#)
- Ma, J.; Xiong, W.; Chen, J.; Feng, D. Hierarchical identification for multivariate Hammerstein systems by using the modified Kalman filter. *IET Control Theory Appl.* **2017**, *11*, 857–869. [\[CrossRef\]](#)
- Shafin, R.; Liu, L.J.; Li, Y.; Wang, A.D.; Zhang, J.Z. Angle and delay estimation for 3-D massive MIMO/FD-MIMO systems based on parametric channel modeling. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 5370–5383. [\[CrossRef\]](#)
- Kawaria, N.; Patidar, R.; George, N.V. Parameter estimation of MIMO bilinear systems using a Levy shuffled frog leaping algorithm. *Soft Comput.* **2017**, *21*, 3849–3858. [\[CrossRef\]](#)
- Roy, S.B.; Bhasin, S.; Kar, I.N. Combined MRAC for unknown MIMO LTI systems with parameter convergence. *IEEE Trans. Autom. Control* **2018**, *63*, 283–290. [\[CrossRef\]](#)
- Ma, H.; Zhang, X.; Liu, Q.; Ding, F.; Jin, X.B.; Alsaedi, A.; Hayat, T. Partially-coupled gradient-based iterative algorithms for multivariable output-error-like systems with autoregressive moving average noises. *IET Control Theory Appl.* **2020**, *14*, 2613–2627. [\[CrossRef\]](#)

22. Huang, W.; Ding, F.; Hayat, T.; Alsaedi, A. Coupled stochastic gradient identification algorithms for multivariate output-error systems using the auxiliary model. *Int. J. Control Autom. Syst.* **2017**, *15*, 1622–1631. [[CrossRef](#)]
23. Cui, T.; Chen, F.Y.; 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](#)]
24. Xu, L.; Yang, E.F. Auxiliary model multiinnovation stochastic gradient parameter estimation methods for nonlinear sandwich systems. *Int. J. Robust Nonlinear Control* **2021**, *31*, 148–165. [[CrossRef](#)]
25. Xu, L.; Sheng, J. Separable multi-innovation stochastic gradient estimation algorithm for the nonlinear dynamic responses of systems. *Int. J. Adapt. Control Signal Process.* **2020**, *34*, 937–954. [[CrossRef](#)]
26. Jin, Q.B.; Wang, Z.; Liu, X.P. Auxiliary model-based interval-varying multi-innovation least squares identification for multivariable OE-like systems with scarce measurements. *J. Process Control* **2015**, *35*, 154–168. [[CrossRef](#)]
27. Zhang, G.Q.; Zhang, X.K.; Pang, H.S. Multi-innovation auto-constructed least squares identification for 4 DOF ship manoeuvring modelling with full-scale trial data. *ISA Trans.* **2015**, *58*, 186–195. [[CrossRef](#)]
28. Wang, C.; Zhu, L. Parameter identification of a class of nonlinear systems based on the multi-innovation identification theory. *J. Frankl. Inst.* **2015**, *352*, 4624–4637. [[CrossRef](#)]
29. Pan, J.; Jiang, X.; Wan, X.K.; Ding, W.F. A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems. *Int. J. Control Autom. Syst.* **2017**, *15*, 1189–1197. [[CrossRef](#)]
30. Chaudhary, N.I.; Raja, M.A.Z.; He, Y.G.; Khan, Z.A.; Machado, J.A.T. Design of multi innovation fractional LMS algorithm for parameter estimation of input nonlinear control autoregressive systems. *Appl. Math. Model.* **2021**, *93*, 412–425. [[CrossRef](#)]
31. Ma, P.; Ding, F. New gradient based identification methods for multivariate pseudo-linear systems using the multi-innovation and the data filtering. *J. Frankl. Inst.* **2017**, *354*, 1568–1583. [[CrossRef](#)]
32. Ma, P.; Ding, F.; Alsaedi, A.; Hayat, T. Decomposition-based gradient estimation algorithms for multivariate equation-error autoregressive systems using the multi-innovation theory. *Circuits Syst. Signal Process.* **2018**, *37*, 1846–1862. [[CrossRef](#)]
33. Ding, F.; Xu, L.; Meng, D.D.; Jin, X.B.; Alsaedi, A.; Hayat, T. Gradient estimation algorithms for the parameter identification of bilinear systems using the auxiliary model. *J. Comput. Appl. Math.* **2020**, *369*, 112575. [[CrossRef](#)]
34. Ji, Y.; Kang, Z. Three-stage forgetting factor stochastic gradient parameter estimation methods for a class of nonlinear systems. *Int. J. Robust Nonlinear Control* **2021**, *31*, 971–987. [[CrossRef](#)]