Article

Decomposition Least-Squares-Based Iterative Identification Algorithms for Multivariable Equation-Error Autoregressive Moving Average Systems

Lijuan Wan ¹, Ximei Liu ¹, Feng Ding ^{2,3,*} and Chunping Chen ⁴

¹ College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China

² School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China

³ Department of Mathematics, King Abdulaziz University, Jeddah 21589, Arabia

⁴ Editorial Office of Journal of Qingdao University of Science and Technology (Natural Science Edition), Qingdao University of Science and Technology, Qingdao 266061, China

* Correspondence: fding@qust.edu.cn or fding@jiangnan.edu.cn

Received: 29 May 2019; Accepted: 28 June 2019; Published: 9 July 2019

Abstract: This paper is concerned with the identification problem for multivariable equation-error systems whose disturbance is an autoregressive moving average process. By means of the hierarchical identification principle and the iterative search, a hierarchical least-squares-based iterative (HLSI) identification algorithm is derived and a least-squares-based iterative (LSI) identification algorithm is given for comparison. Furthermore, a hierarchical multi-innovation least-squares-based iterative (HMILSI) identification algorithm is proposed using the multi-innovation theory. Compared with the LSI algorithm, the HLSI algorithm has smaller computational burden and can give more accurate parameter estimates and the HMILSI algorithm can track time-varying parameters. Finally, a simulation example is provided to verify the effectiveness of the proposed algorithms.

Keywords: least-squares; iterative identification; hierarchical; parameter estimation; multivariable system

1. Introduction

With the development of modern industry, multivariable systems have provided rich possibilities to system modeling and process control [1,2]. Compared with scalar systems, multivariable systems have more complex structures, complicated relationships between variables, high dimensions and stochastic disturbances. For decades, multivariable system identification has attracted increasing attention and many identification approaches have been reported [3]. Among them, how to improve the identification accuracy and to increase the identification efficiency of multivariable systems has become the core problems for researchers. Liu et al. proposed a partially-coupled gradient identification algorithm to obtain the more accurate estimates by filtering the input and output data and by transform the system into two subsystems [4].

Many methods have been used to deal with system identification problems [5–8]. Minimizing different criterion leads to different identification methods, such as neural network methods, fuzzy logic system identification methods, wavelet network system identification methods and so on. According to certain criterion function adopting different search strategies leads to different identification approaches, such as Newton identification methods, least-squares identification methods [9] and gradient identification methods [10,11]. The recursive and iterative identification

techniques can handle the parameter estimates of the system models. The recursive identification algorithms can avoid the matrix inversion and can be operated on-line [12,13] and the iterative algorithms can make use of all the given input–output data sufficiently and can improve the parameter estimation accuracy [14–17]. A filtering based multi-innovation gradient estimation algorithm has been proposed for nonlinear dynamical systems [18].

On the other hand, some identification principles are widely applied to parameter estimation for linear and nonlinear systems and many methods have been obtained for identifying multi-input multi-output systems. The multi-innovation identification theory is beneficial for deriving more accurate estimation algorithms by expanding the innovation from a scalar to a vector and from the vector to the matrix. The hierarchical identification principle is a decomposition-based identification technique, whose key is to decompose a system into several subsystems which can be identified easier by using the gradient-based or the least-squares-based identification algorithms. Hierarchical identification methods are suitable for large-scale and complex systems since they can reduce the dimension of the systems to be identified and reduce the computation load. The filtering technique is potential for identifying systems with colored noises [19,20]. An adaptive filtering based multi-innovation stochastic gradient algorithm was derived for bilinear systems with colored noise and can give small parameter estimation errors as the innovation length increases [21]. A multi-innovation gradient algorithm was developed based on the Kalman filtering to solve the joint state and parameter estimation problem for a nonlinear state space system with time-delay [22]. Many identification methods can be found in [23–32] and they can be applied to many areas [33–37].

Recently, using the hierarchical identification principle and the multi-innovation identification theory, some algorithms have been proposed to track the parameters of nonlinear systems and multivariable systems [38,39]. Considering that the least-squares algorithms have high estimation accuracy and rapid convergence [40–42], this paper focuses on the parameter identification problem of multivariable equation-error systems with autoregressive moving average noise process and presents iterative identification algorithms using the hierarchical identification principle and the multi-innovation method. The key is to decompose a system into two subsystems and to iteratively estimate the matrix of unknown parameters of each subsystem separately. The main contribution of this work are as follows.

- A decomposition least-squares-based iterative identification algorithm is derived for multivariable equation-error autoregressive moving average systems by using the hierarchical identification principle.
- Compared with the least-squares-based iterative algorithm, the proposed algorithm can improve the estimation accuracy and decrease the computation burden.
- A hierarchical multi-innovation least-squares-based iterative identification algorithm is proposed by using the multi-innovation theory, which can track time-varying parameters.

The remainder of the paper is organized as follows. Section 2 derives the identification models of multivariable equation-error systems with different forms of colored noise processes. A Least-squares-based iterative identification algorithm is introduced in Section 3. In Section 4, a hierarchical least-squares-based iterative algorithm is derived. Section 5 presents a hierarchical multi-innovation least-squares-based iterative identification algorithm. The numerical example results are presented to illustrate the effectiveness of the proposed algorithms in Section 6. Finally, Section 7 concludes the work.

2. System Description and Identification Model

Symbols Meaning

- 0: The zero matrix of appropriate sizes.
- $\mathbf{1}_{m \times n}$: An $m \times n$ matrix whose entries are all 1.
- $\mathbf{1}_n$: An n dimensional column vector whose entries are all 1.

I or I_n : The identity matrix of appropriate sizes or $n \times n$.

$\text{tr}[\mathbf{X}]$: The trace of the square matrix $\mathbf{X}$.

$\mathbf{X}^T$: The transpose of the vector or matrix $\mathbf{X}$.

$\|\mathbf{X}\|$: $\|\mathbf{X}\|^2 := \text{tr}[\mathbf{X}\mathbf{X}^T]$.

$A =: X$: X is defined by A .

$X := A$: X is defined by A .

s : The time variable.

$\hat{\theta}$: The estimate of the parameter matrix θ .

$\hat{\theta}(s)$: The estimate of θ at time s .

p_0 : A large positive constant, e.g., $p_0 = 10^6$.

Consider the following multivariable equation-error system with the colored noise,

$$\mathbf{A}(r)\mathbf{y}(s) = \mathbf{B}(r)\mathbf{u}(s) + \mathbf{w}(s), \quad (1)$$

where $\mathbf{u}(s) := [u_1(s), u_2(s), \dots, u_r(s)]^T \in \mathbb{R}^r$ is the input vector, $\mathbf{y}(s) := [y_1(s), y_2(s), \dots, y_m(s)]^T \in \mathbb{R}^m$ is the output vector, $\mathbf{A}(r)$ and $\mathbf{B}(r)$ are the matrix-coefficient polynomials in the unit backward shift operator r^{-1} (i.e., $r^{-1}\mathbf{x}(s) = \mathbf{x}(s-1)$) with degrees n_a and n_b , and they are defined as

$$\begin{aligned} \mathbf{A}(r) &:= \mathbf{I}_m + \mathbf{A}_1 r^{-1} + \mathbf{A}_2 r^{-2} + \dots + \mathbf{A}_{n_a} r^{-n_a}, \mathbf{A}_i \in \mathbb{R}^{m \times m}, \\ \mathbf{B}(r) &:= \mathbf{B}_1 r^{-1} + \mathbf{B}_2 r^{-2} + \dots + \mathbf{B}_{n_b} r^{-n_b}, \mathbf{B}_i \in \mathbb{R}^{m \times r}, \end{aligned}$$

where $\mathbf{w}(s) \in \mathbb{R}^m$ is a stochastic noise vector with zero mean, which may be a moving average (MA) process, an autoregressive (AR) process or an autoregressive moving average process (ARMA).

Considering the ARMA process, there are four forms for the description of the noise term as follows.

1. The first form is $\mathbf{w}(s) = \frac{d(r)}{c(r)}\mathbf{v}(s) \in \mathbb{R}^m$, where $\mathbf{v}(s) \in \mathbb{R}^m$ is a white noise vector, and $c(r)$ and $d(r)$ are scalar polynomials in r^{-1} with degrees n_c and n_d , which are expressed as

$$\begin{aligned} c(r) &:= 1 + c_1 r^{-1} + c_2 r^{-2} + \dots + c_{n_c} r^{-n_c}, c_i \in \mathbb{R}, \\ d(r) &:= 1 + d_1 r^{-1} + d_2 r^{-2} + \dots + d_{n_d} r^{-n_d}, d_i \in \mathbb{R}. \end{aligned}$$

2. The second form is $\mathbf{w}(s) = d(r)\mathbf{C}^{-1}(r)\mathbf{v}(s) \in \mathbb{R}^m$, where $d(r)$ is a scalar polynomial and $\mathbf{C}(r)$ is a matrix polynomial in r^{-1} , which is expressed as

$$\mathbf{C}(r) := \mathbf{I}_m + \mathbf{C}_1 r^{-1} + \mathbf{C}_2 r^{-2} + \dots + \mathbf{C}_{n_c} r^{-n_c}, \mathbf{C}_i \in \mathbb{R}^{m \times m}.$$

3. The third form is $\mathbf{w}(s) = \frac{D(r)}{c(r)}\mathbf{v}(s) \in \mathbb{R}^m$, where $c(r)$ is a scalar polynomial and $\mathbf{D}(r)$ is a matrix polynomial in r^{-1} , which is expressed as

$$\mathbf{D}(r) := \mathbf{I}_m + \mathbf{D}_1 r^{-1} + \mathbf{D}_2 r^{-2} + \dots + \mathbf{D}_{n_d} r^{-n_d}, \mathbf{D}_i \in \mathbb{R}^{m \times m}.$$

4. The last form is $\mathbf{w}(s) = \mathbf{C}^{-1}(r)\mathbf{D}(r)\mathbf{v}(s) \in \mathbb{R}^m$, where $\mathbf{C}(r)$ and $\mathbf{D}(r)$ are matrix polynomials in r^{-1} with degrees n_c and n_d .

Without loss of generality, we consider the multivariable systems with ARMA noise process, i.e.,

$$\mathbf{A}(r)\mathbf{y}(s) = \mathbf{B}(r)\mathbf{u}(s) + \mathbf{C}^{-1}(r)\mathbf{D}(r)\mathbf{v}(s), \quad (2)$$

where $\mathbf{A}_i$, $\mathbf{B}_i$, $\mathbf{C}_i$ and $\mathbf{D}_i$ are the coefficient matrices to be estimated. Assume that the orders n_a , n_b , n_c and n_d are known and $n := mn_a + rn_b + mn_c + mn_d$, $n_1 := mn_a + rn_b$, $n_2 := mn_c + mn_d$, $\mathbf{u}(s) = \mathbf{0}$, $\mathbf{y}(s) = \mathbf{0}$ and $\mathbf{v}(s) = \mathbf{0}$ for $s \leq 0$.

Define the parameter matrices θ , α and β as

$$\begin{aligned}\theta^T &:= [\alpha^T, \beta^T] \in \mathbb{R}^{m \times n}, \\ \alpha^T &:= [A_1, A_2, \dots, A_{n_a}, B_1, B_2, \dots, B_{n_b}] \in \mathbb{R}^{m \times n_1}, \\ \beta^T &:= [C_1, C_2, \dots, C_{n_c}, D_1, D_2, \dots, D_{n_d}] \in \mathbb{R}^{m \times n_2},\end{aligned}$$

and the information vectors $\varphi(s)$, $\phi_\alpha(s)$ and $\psi(s)$ as

$$\begin{aligned}\varphi(s) &:= [\phi_\alpha^T(s), \psi^T(s)]^T \in \mathbb{R}^n, \\ \phi_\alpha(s) &:= [-y^T(s-1), -y^T(s-2), \dots, -y^T(s-n_a), u^T(s-1), u^T(s-2), \dots, u^T(s-n_b)]^T, \quad (3) \\ \psi(s) &:= [-w^T(s-1), -w^T(s-2), \dots, -w^T(s-n_c), v^T(s-1), v^T(s-2), \dots, v^T(s-n_d)]^T. \quad (4)\end{aligned}$$

It follows that,

$$\begin{aligned}w(s) &= [I_m - C(r)]w(s) + D(r)v(s) \\ &= [I_m - C(r)]w(s) + [D(r) - I_m]v(s) + v(s) \\ &= \beta^T \psi(s) + v(s).\end{aligned} \quad (5)$$

Substituting Equation (5) into Equation (1) gives

$$\begin{aligned}y(s) &= [I_m - A(r)]y(s) + B(r)u(s) + w(s) \\ &= \alpha^T \phi_\alpha(s) + \beta^T \psi(s) + v(s) \quad (6) \\ &= \theta^T \varphi(s) + v(s).\end{aligned} \quad (7)$$

In this identification model, θ is the parameter matrix to be identified, which consists of two parameter matrices: the parameter matrix α of the system model and the parameter matrix β of the noise model.

Note that the unknown parameter matrices A_i , B_i , C_i and D_i are included in θ , the identification problem has many parameters to be estimated and it is significant to enhance the computational efficiency of the multivariable system identification algorithms. Aiming at this goal, this work studies the decomposition least-squares-based iterative estimation algorithms for the multivariable CARARMA systems by using the hierarchical identification principle from the measured input–output data $\{u(s), y(s): s = 1, 2, \dots, L\}$ (L denotes the data length).

3. The Least-Squares-Based Iterative Algorithm

Let $k = 1, 2, 3, \dots$ be an iterative variable and $\hat{\theta}_k := \begin{bmatrix} \hat{\alpha}_k \\ \hat{\beta}_k \end{bmatrix}$ be the estimate of $\theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ at iteration k . Assume that L is the data length ($L \gg n$). According to the data $\{u(s), y(s): s = 1, 2, \dots, L\}$ and based on Equation (7), define the stacked output matrix $Y(L)$ as

$$Y(L) = [y(1), y(2), \dots, y(L)] \in \mathbb{R}^{m \times L},$$

and the stacked information matrix $\Omega(L)$ is defined as

$$\Omega(L) = [\varphi(1), \varphi(2), \dots, \varphi(L)] \in \mathbb{R}^{n \times L}.$$

Define a quadratic criterion function,

$$J(\theta) := \|Y(L) - \theta^T \Omega(L)\|^2$$

$$= \sum_{s=1}^L [\mathbf{y}(s) - \boldsymbol{\theta}^T \boldsymbol{\varphi}(s)]^T [\mathbf{y}(s) - \boldsymbol{\theta}^T \boldsymbol{\varphi}(s)]. \quad (8)$$

To minimize $J(\boldsymbol{\theta})$, letting its partial derivatives with $\boldsymbol{\theta}$ at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ be zero gives

$$\left. \frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -2\boldsymbol{\Omega}(L)[\mathbf{Y}(L) - \hat{\boldsymbol{\theta}}^T \boldsymbol{\Omega}(L)]^T = \mathbf{0}$$

or

$$\left. \frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -2 \sum_{s=1}^L \boldsymbol{\varphi}(s) [\mathbf{y}(s) - \hat{\boldsymbol{\theta}}^T \boldsymbol{\varphi}(s)]^T = \mathbf{0}.$$

Then, the estimates of the parameter matrix $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = [\boldsymbol{\Omega}(L)\boldsymbol{\Omega}^T(L)]^{-1}\boldsymbol{\Omega}(L)\mathbf{Y}^T(L) = \left[\sum_{s=1}^L \boldsymbol{\varphi}(s)\boldsymbol{\varphi}^T(s) \right]^{-1} \sum_{s=1}^L \boldsymbol{\varphi}(s)\mathbf{y}^T(s). \quad (9)$$

It is observed that the information vector $\boldsymbol{\varphi}(s)$ in $\boldsymbol{\Omega}(L)$ involves the unknown noise terms $\mathbf{w}(s-i)$ ($i = 1, 2, \dots, n_c$) and $\mathbf{v}(s-j)$ ($j = 1, 2, \dots, n_d$), therefore the estimate $\hat{\boldsymbol{\theta}}$ in Equation (9) cannot be computed. To resolve this problem, we replace the unknown noise terms $\mathbf{w}(s-i)$ and $\mathbf{v}(s-j)$ in $\boldsymbol{\varphi}(s)$ with their corresponding estimates $\hat{\mathbf{w}}_{k-1}(s-i)$ and $\hat{\mathbf{v}}_{k-1}(s-j)$ at iteration $k-1$. Then, we can obtain the least-squares-based iterative (LSI) algorithm as follows:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_k &= [\hat{\boldsymbol{\Omega}}_k(L)\hat{\boldsymbol{\Omega}}_k^T(L)]^{-1}\hat{\boldsymbol{\Omega}}_k(L)\mathbf{Y}^T(L) \\ &= \left[\sum_{s=1}^L \hat{\boldsymbol{\varphi}}_k(s)\hat{\boldsymbol{\varphi}}_k^T(s) \right]^{-1} \sum_{s=1}^L \hat{\boldsymbol{\varphi}}_k(s)\mathbf{y}^T(s), \quad k = 1, 2, 3, \dots \end{aligned} \quad (10)$$

$$\mathbf{Y}(L) = [\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(L)], \quad (11)$$

$$\hat{\boldsymbol{\Omega}}_k(L) = [\hat{\boldsymbol{\varphi}}_k(1), \hat{\boldsymbol{\varphi}}_k(2), \dots, \hat{\boldsymbol{\varphi}}_k(L)], \quad (12)$$

$$\begin{aligned} \hat{\boldsymbol{\varphi}}_k(s) &= [\hat{\boldsymbol{\phi}}_\alpha^T(s), -\hat{\mathbf{w}}_{k-1}^T(s-1), -\hat{\mathbf{w}}_{k-1}^T(s-2), \dots, -\hat{\mathbf{w}}_{k-1}^T(s-n_c), \\ &\quad \hat{\mathbf{v}}_{k-1}^T(s-1), \hat{\mathbf{v}}_{k-1}^T(s-2), \dots, \hat{\mathbf{v}}_{k-1}^T(s-n_d)]^T, \end{aligned} \quad (13)$$

$$\begin{aligned} \hat{\boldsymbol{\phi}}_\alpha(s) &= [-\mathbf{y}^T(s-1), -\mathbf{y}^T(s-2), \dots, -\mathbf{y}^T(s-n_a), \\ &\quad \mathbf{u}^T(s-1), \mathbf{u}^T(s-2), \dots, \mathbf{u}^T(s-n_b)]^T, \end{aligned} \quad (14)$$

$$\hat{\mathbf{w}}_k(s) = \mathbf{y}(s) - \hat{\mathbf{a}}_k^T \hat{\boldsymbol{\phi}}_\alpha(s), \quad s = 1, 2, \dots, L, \quad (15)$$

$$\hat{\mathbf{v}}_k(s) = \mathbf{y}(s) - \hat{\boldsymbol{\theta}}_k^T \hat{\boldsymbol{\varphi}}(s), \quad (16)$$

$$\hat{\boldsymbol{\theta}}_k^T = [\hat{\mathbf{a}}_k^T, \hat{\boldsymbol{\beta}}_k^T], \quad (17)$$

$$\hat{\mathbf{a}}_k^T = [\hat{A}_{1,k}, \hat{A}_{2,k}, \dots, \hat{A}_{n_a,k}, \hat{B}_{1,k}, \hat{B}_{2,k}, \dots, \hat{B}_{n_b,k}], \quad (18)$$

$$\hat{\boldsymbol{\beta}}_k^T = [\hat{C}_{1,k}, \hat{C}_{2,k}, \dots, \hat{C}_{n_c,k}, \hat{D}_{1,k}, \hat{D}_{2,k}, \dots, \hat{D}_{n_d,k}]. \quad (19)$$

The computational efficiency of the LSI algorithm for the multivariable CARARMA systems is given in Table 1.

Table 1. The computational efficiency of the least-squares-based iterative identification algorithm.

Expressions	Number of Multiplications	Number of Additions
$\hat{\theta}_k = S'_k \beta_k \in \mathbb{R}^{n \times m}$	mn^2k	$mn(n-1)k$
$S'_k := S_k^{-1} \in \mathbb{R}^{n \times n}$	$(n^3 + n^2 - n + 1)k$	$(n^3 + n^2 - 2n)k$
$S_k := \hat{\Omega}_k(L) \hat{\Omega}_k^T(L) \in \mathbb{R}^{n \times n}$	n^2Lk	$n^2(L-1)k$
$\beta_k := \hat{\Omega}_k(L) Y^T(L) \in \mathbb{R}^{n \times m}$	$mnLk$	$mn(L-1)k$
$\hat{w}_k(s) = y(s) - \hat{\alpha}_k^T \phi_\alpha(s) \in \mathbb{R}^m$	mn_1Lk	mn_1Lk
$\hat{v}_k(s) = \hat{w}_k - \hat{\beta}_k^T \psi_k(s) \in \mathbb{R}^m$	mn_2Lk	mn_2Lk
Sum (Number of multiplications)	$[n^3 + (m+1+L)n^2 + (2mL-1)n + 1]k$	
Sum (Number of additions)	$[n^3 + (m+L)n^2 + 2(mL-m-1)n]k$	
Total flops	$N_1 := [2n^3 + (2m+2L+1)n^2 + (4mL-2m-3)n + 1]k$	

4. The Hierarchical Least-Squares-Based Iterative Algorithm

The multivariable system in Equation (6) is decomposed into two subsystems, one containing α and the other containing β , and each subsystem is identified separately by using the hierarchical identification principle and the iterative technique. Define two intermediate variables,

$$h(s) := y(s) - \beta^T \psi(s), \quad (20)$$

$$w(s) := y(s) - \alpha^T \phi_\alpha(s). \quad (21)$$

The system in Equation (1) can be decomposed into the following two fictitious subsystems:

$$h(s) = \alpha^T \phi_\alpha(s) + v(s), \quad (22)$$

$$w(s) = \beta^T \psi(s) + v(s). \quad (23)$$

Let L be the data length ($L \gg n$) and define the stacked output matrices $Y(L)$, $H(L)$, $W(L)$ and the stacked information matrices $\Phi(L)$ and $\Psi(L)$ as

$$Y(L) := [y(1), y(2), \dots, y(L)] \in \mathbb{R}^{m \times L}, \quad (24)$$

$$H(L) := [h(1), h(2), \dots, h(L)] \in \mathbb{R}^{m \times L},$$

$$W(L) := [w(1), w(2), \dots, w(L)] \in \mathbb{R}^{m \times L},$$

$$\Phi(L) := [\phi_\alpha(1), \phi_\alpha(2), \dots, \phi_\alpha(L)] \in \mathbb{R}^{n_1 \times L}, \quad (25)$$

$$\Psi(L) := [\psi(1), \psi(2), \dots, \psi(L)] \in \mathbb{R}^{n_2 \times L}.$$

From Equations (20) and (21), we have

$$H(L) = Y(L) - \beta^T \Psi(L), \quad (26)$$

$$W(L) = Y(L) - \alpha^T \Phi(L). \quad (27)$$

Define quadratic criterion functions:

$$J_1(\alpha) := \|H(L) - \alpha^T \Phi(L)\|^2 = \sum_{s=1}^L [h(s) - \alpha^T \phi_\alpha(s)]^T [h(s) - \alpha^T \phi_\alpha(s)],$$

$$J_2(\beta) := \|W(L) - \beta^T \Psi(L)\|^2 = \sum_{s=1}^L [w(s) - \beta^T \psi(s)]^T [w(s) - \beta^T \psi(s)].$$

Let $k = 1, 2, 3, \dots$ be an iterative variable, and $\hat{\alpha}_k$ and $\hat{\beta}_k$ be the estimates of parameter matrices α and β at iteration k , respectively. Minimizing $J_1(\alpha)$ and $J_2(\beta)$ and letting their partial derivatives with respect to α and β be zero, we can obtain the least-squares-based iterative relations, respectively:

$$\left. \frac{\partial J_1(\alpha)}{\partial \alpha} \right|_{\alpha=\hat{\alpha}} = -2\Phi(L)[H(L) - \hat{\alpha}^T \Phi(L)]^T = 0$$

or

$$\left. \frac{\partial J_1(\alpha)}{\partial \alpha} \right|_{\alpha=\hat{\alpha}} = -2 \sum_{s=1}^L \phi_\alpha(s)[h(s) - \hat{\alpha}^T \phi_\alpha(s)]^T = 0,$$

$$\left. \frac{\partial J_2(\beta)}{\partial \beta} \right|_{\beta=\hat{\beta}} = -2\Psi(L)[W(L) - \hat{\beta}^T \Psi(L)]^T = 0$$

or

$$\left. \frac{\partial J_2(\beta)}{\partial \beta} \right|_{\beta=\hat{\beta}} = -2 \sum_{s=1}^L \psi(s)[w(s) - \hat{\beta}^T \psi(s)]^T = 0.$$

Then, the estimates of the parameter matrices α and β are given by

$$\hat{\alpha} = [\Phi(L)\Phi^T(L)]^{-1}\Phi(L)[Y(L) - \hat{\beta}^T \Psi(L)]^T, \quad (28)$$

$$\hat{\beta} = [\Psi(L)\Psi^T(L)]^{-1}\Psi(L)[Y(L) - \hat{\alpha}^T \Phi(L)]^T. \quad (29)$$

Notice that the information vector $\psi(s)$ in $\Psi(L)$ contains the unknown noise items $w(s-i)$ and $v(s-j)$, thus the algorithm in Equations (28) and (29) cannot be implemented. The solution is to use the estimates $\hat{w}_{k-1}(s-i)$ and $\hat{v}_{k-1}(s-j)$ of the unknown noise vectors $w(s-i)$ and $v(s-j)$ at iteration $k-1$ to define the estimate $\hat{\psi}_k(s)$ of $\psi(s)$ as

$$\hat{\psi}_k(s) := [-\hat{w}_{k-1}^T(s-1), -\hat{w}_{k-1}^T(s-2), \dots, -\hat{w}_{k-1}^T(s-n_c), \hat{v}_{k-1}^T(s-1), \hat{v}_{k-1}^T(s-2), \dots, \hat{v}_{k-1}^T(s-n_d)]^T. \quad (30)$$

Replacing the unknown parameter matrix α in Equation (21) with its estimate $\hat{\alpha}_k$, we have the estimate of $w(s)$ at iteration k as

$$\hat{w}_k(s) := y(s) - \hat{\alpha}_k^T \phi_\alpha(s). \quad (31)$$

According to Equations (5) and (23), replacing $w(s)$ and α with their estimates $\hat{w}_k(s)$ and $\hat{\alpha}_k$, we have the estimate of $v(s)$ at iteration k as

$$\begin{aligned} \hat{v}_k(s) &:= \hat{w}_k(s) - \hat{\beta}_k^T \hat{\psi}_k(s) \\ &= y(s) - \hat{\alpha}_k^T \phi_\alpha(s) - \hat{\beta}_k^T \hat{\psi}_k(s). \end{aligned} \quad (32)$$

Use the estimate $\hat{\psi}_k(s)$ of $\psi(s)$ to define the estimate $\hat{\Psi}_k(L)$ of $\Psi(L)$ as

$$\hat{\Psi}_k(L) := [\hat{\psi}_k(1), \hat{\psi}_k(2), \dots, \hat{\psi}_k(L)] \in \mathbb{R}^{n_2 \times L}. \quad (33)$$

Replacing β , α and $\Psi(L)$ in Equations (28) and (29) with their estimates $\hat{\beta}_{k-1}$, $\hat{\alpha}_{k-1}$ and $\hat{\Psi}_k(L)$, and combining Equations (24), (25), (33), (3), and (30)–(32), we have the hierarchical least-squares-based iterative (HLSI) identification algorithm as follows:

$$\hat{\alpha}_k = [\Phi(L)\Phi^T(L)]^{-1}\Phi(L)[Y(L) - \hat{\beta}_{k-1}^T \hat{\Psi}_k(L)]^T \quad (34)$$

$$= \left[\sum_{s=1}^L \phi_\alpha(s)\phi_\alpha^T(s) \right]^{-1} \sum_{s=1}^L \phi_\alpha(s)[y(s) - \hat{\beta}_{k-1}^T \hat{\psi}_k(s)]^T, \quad k = 1, 2, 3, \dots$$

$$\hat{\beta}_k = [\hat{\Psi}_k(L)\hat{\Psi}_k^T(L)]^{-1}\hat{\Psi}_k(L)[Y(L) - \hat{\alpha}_{k-1}^T \Phi(L)]^T \quad (35)$$

$$= \left[\sum_{s=1}^L \hat{\psi}_k(s) \hat{\psi}_k^T(s) \right]^{-1} \sum_{s=1}^L \hat{\psi}_k(s) [\mathbf{y}(s) - \hat{\mathbf{a}}_{k-1}^T \boldsymbol{\phi}_\alpha(s)]^T,$$

$$\mathbf{Y}(L) = [\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(L)], \quad (36)$$

$$\boldsymbol{\Phi}(L) = [\boldsymbol{\phi}_\alpha(1), \boldsymbol{\phi}_\alpha(2), \dots, \boldsymbol{\phi}_\alpha(L)], \quad (37)$$

$$\hat{\Psi}_k(L) = [\hat{\psi}_k(1), \hat{\psi}_k(2), \dots, \hat{\psi}_k(L)], \quad (38)$$

$$\boldsymbol{\phi}_\alpha(s) = [-\mathbf{y}^T(s-1), -\mathbf{y}^T(s-2), \dots, -\mathbf{y}^T(s-n_a), \mathbf{u}^T(s-1), \mathbf{u}^T(s-2), \dots, \mathbf{u}^T(s-n_b)]^T, \quad (39)$$

$$\hat{\psi}_k(s) = [-\hat{\mathbf{w}}_{k-1}^T(s-1), -\hat{\mathbf{w}}_{k-1}^T(s-2), \dots, -\hat{\mathbf{w}}_{k-1}^T(s-n_c), \hat{\mathbf{v}}_{k-1}^T(s-1), \hat{\mathbf{v}}_{k-1}^T(s-2), \dots, \hat{\mathbf{v}}_{k-1}^T(s-n_d)]^T, \quad (40)$$

$$\hat{\mathbf{w}}_k(s) = \mathbf{y}(s) - \hat{\mathbf{a}}_k^T \boldsymbol{\phi}_\alpha(s), \quad s = 1, 2, 3, \dots, L, \quad (41)$$

$$\hat{\mathbf{v}}_k(s) = \mathbf{y}(s) - \hat{\mathbf{a}}_k^T \boldsymbol{\phi}_\alpha(s) - \hat{\boldsymbol{\beta}}_k^T \hat{\psi}_k(s), \quad (42)$$

$$\hat{\boldsymbol{\theta}}_k^T = [\hat{\mathbf{a}}_k^T, \hat{\boldsymbol{\beta}}_k^T], \quad (43)$$

$$\hat{\mathbf{a}}_k^T = [\hat{A}_{1,k}, \hat{A}_{2,k}, \dots, \hat{A}_{n_a,k}, \hat{B}_{1,k}, \hat{B}_{2,k}, \dots, \hat{B}_{n_b,k}], \quad (44)$$

$$\hat{\boldsymbol{\beta}}_k^T = [\hat{C}_{1,k}, \hat{C}_{2,k}, \dots, \hat{C}_{n_c,k}, \hat{D}_{1,k}, \hat{D}_{2,k}, \dots, \hat{D}_{n_d,k}]. \quad (45)$$

The steps of the HLSI algorithm to compute the parameter estimates $\hat{\mathbf{a}}_k$ and $\hat{\boldsymbol{\beta}}_k$ are listed as follows.

1. To initialize, let $k = 1$, give the data length L ($L \gg n$) and some small positive ε , and set the initial values: $\hat{\mathbf{a}}_0 = \mathbf{1}_{n_1 \times m} / p_0$, $\hat{\boldsymbol{\beta}}_0 = \mathbf{1}_{n_2 \times m} / p_0$, $p_0 = 10^6$.
2. Collect the input-output data $\{\mathbf{u}(s), \mathbf{y}(s): s = 1, 2, \dots, L\}$. Let $\hat{\mathbf{w}}_0(s)$ and $\hat{\mathbf{v}}_0(s)$ be random variables, and form the information vector $\boldsymbol{\phi}_\alpha(s)$ by Equation (39), the stacked output matrix $\mathbf{Y}(L)$ by Equation (36) and the stacked information matrix $\boldsymbol{\Phi}(L)$ by Equation (37).
3. Form the information vector $\hat{\psi}_k(s)$ by Equation (40), $s = 1, 2, \dots, L$, and form the stacked information matrix $\hat{\Psi}_k(L)$ by Equation (38).
4. Update the parameter estimates $\hat{\mathbf{a}}_k$ by Equation (34) and $\hat{\boldsymbol{\beta}}_k$ by Equation (35).
5. Compute $\hat{\mathbf{w}}_k(s)$ and $\hat{\mathbf{v}}_k(s)$ by Equations (41) and (42).
6. If $\|\hat{\mathbf{a}}_k - \hat{\mathbf{a}}_{k-1}\| + \|\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_{k-1}\| > \varepsilon$, increase k by 1 and go to Step 3; otherwise, obtain the iterative time k and the estimates $\hat{\mathbf{a}}_k$ and $\hat{\boldsymbol{\beta}}_k$, and terminate this procedure.

Remark 1. In the above algorithms, the unknown variables are replaced with their corresponding previous iteration estimates. Under some conditions and ensuring that the information matrices are persistently exciting, the identification algorithms based on the replacement of the unknown variables with their estimates can give satisfactory parameter estimates.

Remark 2. The computation is very important to evaluate the performance of the algorithm and it can be measured by the numbers of multiplication and addition operations. In particular, one division operation can be counted as one multiplication operation and one subtraction operation can be counted as one addition operation. One multiplication or one addition operation is called one floating point operation (flop for short). Here, we give the computational efficiency by measuring the flop numbers involved in the identification algorithms.

The multiplication and addition numbers of the HLSI algorithm are shown in Table 2.

Table 2. The computational efficiency of the hierarchical least-squares-based iterative identification algorithm.

Expressions	Number of Multiplications	Number of Additions
$\hat{\mathbf{a}}_k = \mathbf{S}'_1 \beta_{1,k} \in \mathbb{R}^{n_1 \times m}$	$mn_1^2 k$	$mn_1(n_1 - 1)k$
$\mathbf{S}'_1 := \mathbf{S}_1^{-1} \in \mathbb{R}^{n_1 \times n_1}$	$n_1^3 + n_1^2 - n_1 + 1$	$n_1^3 + n_1^2 - 2n_1$
$\mathbf{S}_1 := \Phi_\alpha(L) \Phi_\alpha^T(L) \in \mathbb{R}^{n_1 \times n_1}$	$n_1^2 L$	$n_1^2(L - 1)$
$\beta_{1,k} := \Phi(L) E_k^T(L) \in \mathbb{R}^{n_1 \times m}$	$mn_1 L k$	$mn_1(L - 1)k$
$\mathbf{E}_k(L) := \mathbf{Y}(L) - \hat{\mathbf{a}}_{k-1}^T \Phi(L) - \hat{\beta}_{k-1}^T \Psi_k(L)$	$mnLk$	$mnLk$
$\hat{\beta}_k = \mathbf{S}'_{2,k} \beta_{2,k} \in \mathbb{R}^{n_2 \times m}$	$mn_2^2 k$	$mn_2(n_2 - 1)k$
$\mathbf{S}'_{2,k} := \mathbf{S}_{2,k}^{-1} \in \mathbb{R}^{n_2 \times n_2}$	$[n_2^3 + n_2^2 - n_2 + 1]k$	$[n_2^3 + n_2^2 - 2n_2]k$
$\mathbf{S}_{2,k} := \Psi_k(L) \Psi_k^T(L) \in \mathbb{R}^{n_2 \times n_2}$	$n_2^2 L k$	$n_2^2(L - 1)k$
$\beta_{2,k} := \Psi_k(L) E_k^T(L) \in \mathbb{R}^{n_2 \times m}$	$mn_2 L k$	$mn_2(L - 1)k$
$\hat{\mathbf{w}}_k(s) = \mathbf{y}(s) - \hat{\mathbf{a}}_k^T \Phi_\alpha(s) \in \mathbb{R}^m$	$mn_1 L k$	$mn_1 L k$
$\hat{\mathbf{v}}_k(s) = \hat{\mathbf{w}}_k - \hat{\beta}_k^T \Psi_k(s) \in \mathbb{R}^m$	$mn_2 L k$	$mn_2 L k$
Sum (Number of multiplications)	$[m(n_1^2 + n_2^2) + 3mnL + n_2^3 + n_2^2(L + 1) - n_2 + 1]k$ $+ n_1^3 + n_1^2(L + 1) - n_1 + 1$	
Sum (Number of additions)	$[m(n_1^2 + n_2^2) + mn(3L - 2) + n_2^3 + n_2^2 L - 2n_2]k$ $+ n_1^3 + n_1^2 L - 2n_1$	
Total flops	$N_2 := [2m(n_1^2 + n_2^2) + 2mn(3L - 2) + 2n_2^3 + n_2^2(2L + 1) - 3n_2 + 1]k$ $+ 2n_1^3 + n_1^2(2L + 1) - 3n_1 + 1$	

The amount of calculations of the LSI and HLSI identification algorithms are compared as follows

$$\begin{aligned}
 N_1 - N_2 &= [2n^3 + (2m + 2L + 1)n^2 + (4mL - 2m - 3)n + 1]k \\
 &\quad - [2m(n_1^2 + n_2^2) + 2mn(3L - 2) + 2n_2^3 + n_2^2(2L + 1) - 3n_2 + 1] \\
 &\quad - [2n_1^3 + n_1^2(2L + 1) - 3n_1 + 1] \\
 &= [(4m^2 + 2mL - 4m - 3)n + 3n_1^3 + (2L + 3n_2 + 1)n_1^2 + 3n_1n_2^2 + 3n_2 - 1]k \\
 &\quad - [2n_1^3 + n_1^2(2L + 1) - 3n_1 + 1] > 0.
 \end{aligned}$$

It is obvious that the HLSI identification algorithm requires less computation than the LSI identification algorithm.

Compared with the LSI algorithm, the HLSI algorithm can obtain more accurate parameter estimates and higher computational efficiency for multivariable systems with the colored noise. Furthermore, the proposed approaches in the paper can combine other mathematical optimization approaches [43–47] and statistical strategies [48–54] to study the parameter estimation problems of linear and nonlinear systems with different disturbances [55–60], and can be applied to other fields [61–64] such as signal processing [65–70].

5. The Hierarchical Multi-Innovation Least-Squares-Based Iterative Identification Algorithm

To possess higher identification accuracy of the HLSI algorithm, we propose a hierarchical multi-innovation least-squares-based iterative algorithm by using the multi-innovation identification theory in this section.

Let p be the innovation length. Consider the data from time $l = s - p + 1$ to $l = s$ and define the stacked information matrices:

$$\mathbf{Y}(p, s) := [\mathbf{y}(s), \mathbf{y}(s - 1), \dots, \mathbf{y}(s - p + 1)] \in \mathbb{R}^{m \times p}, \quad (46)$$

$$\Phi(p, s) := [\Phi_\alpha(s), \Phi_\alpha(s - 1), \dots, \Phi_\alpha(s - p + 1)] \in \mathbb{R}^{n_1 \times p}, \quad (47)$$

$$\Psi(p, s) := [\Psi(s), \Psi(s - 1), \dots, \Psi(s - p + 1)] \in \mathbb{R}^{n_2 \times p}, \quad (48)$$

$$\mathbf{H}(p, s) := [\mathbf{h}(s), \mathbf{h}(s - 1), \dots, \mathbf{h}(s - p + 1)] = \mathbf{Y}(p, s) - \beta^T \Psi(p, s) \in \mathbb{R}^{m \times p},$$

$$\mathbf{W}(p, s) := [\mathbf{w}(s), \mathbf{w}(s - 1), \dots, \mathbf{w}(s - p + 1)] = \mathbf{Y}(p, s) - \mathbf{a}^T \Phi(p, s) \in \mathbb{R}^{m \times p}.$$

From the identification model in Equation (6), we define two cost functions:

$$\begin{aligned} J_3(\alpha) &:= \|\mathbf{H}(p, s) - \alpha^T \Phi(p, s)\|^2 \\ &= \sum_{l=s-p+1}^s [\mathbf{h}(l) - \alpha^T \phi_\alpha(l)]^T [\mathbf{h}(l) - \alpha^T \phi_\alpha(l)], \\ J_4(\beta) &:= \|\mathbf{W}(p, s) - \beta^T \Psi(p, s)\|^2 \\ &= \sum_{l=s-p+1}^s [\mathbf{w}(l) - \beta^T \psi(l)]^T [\mathbf{w}(l) - \beta^T \psi(l)]. \end{aligned}$$

Let $\hat{\alpha}_k(s)$ and $\hat{\beta}_k(s)$ be the k th iterative estimates of α and β at the current time s . Minimizing $J_3(\alpha)$ and $J_4(\beta)$, we have the iterative relations as follows:

$$\hat{\alpha}_k(s) = [\Phi(p, s) \Phi^T(p, s)]^{-1} \Phi(p, s) [Y(p, s) - \hat{\beta}_{k-1}^T(s) \Psi(p, s)]^T, \quad (49)$$

$$\hat{\beta}_k(s) = [\Psi(p, s) \Psi^T(p, s)]^{-1} \Psi(p, s) [Y(p, s) - \hat{\alpha}_{k-1}^T(s) \Phi(p, s)]^T. \quad (50)$$

Notice that the information vector $\psi(s)$ in $\Psi(p, s)$ contains the unmeasured variables $w(s-i)$ and $v(s-j)$, thus the algorithm in Equations (49) and (50) cannot be implemented. The solution is to use the estimates $\hat{w}_{k-1}(s-i)$ and $\hat{v}_{k-1}(s-j)$ of the unknown noise vectors $w(s-i)$ and $v(s-j)$ to define the estimate $\hat{\psi}_k(s)$ of $\psi(p, s)$ as

$$\hat{\psi}_k(s) := [-\hat{w}_{k-1}^T(s-1), -\hat{w}_{k-1}^T(s-2), \dots, -\hat{w}_{k-1}^T(s-n_c), \hat{v}_{k-1}^T(s-1), \hat{v}_{k-1}^T(s-2), \dots, \hat{v}_{k-1}^T(s-n_d)]^T. \quad (51)$$

Replacing the unknown parameter matrix α in Equation (22) with its estimate $\hat{\alpha}_k(s)$, we have the estimate of $w(s)$ at iteration k as

$$\hat{w}_k(s) := y(s) - \hat{\alpha}_k^T(s) \phi_\alpha(s). \quad (52)$$

Equation (23) gives the estimate of $v(s)$ at iteration k as

$$\hat{v}_k(s) := y(s) - \hat{\alpha}_k^T(s) \phi_\alpha(s) - \hat{\beta}_k^T(s) \hat{\psi}_k(s). \quad (53)$$

Define the estimate $\hat{\Psi}_k(p, s)$ of $\Psi(p, s)$ with the estimate $\hat{\psi}_k(s)$ of $\psi(s)$ as

$$\hat{\Psi}_k(p, s) := [\hat{\psi}_k(s), \hat{\psi}_k(s-1), \dots, \hat{\psi}_k(s-p+1)] \in \mathbb{R}^{n_2 \times p}. \quad (54)$$

Replacing β , α and $\Psi(p, t)$ in Equations (49) and (50) with their estimates $\hat{\beta}_{k-1}(s)$, $\hat{\alpha}_{k-1}(s)$ and $\hat{\Psi}_k(p, t)$, and combining Equations (46), (47), (54), (3), and (51)–(53), we have the hierarchical multi-innovation least-squares-based iterative identification algorithm for the multivariable CARARMA systems as follows:

$$\hat{\alpha}_k(s) = [\Phi(p, s) \Phi^T(p, s)]^{-1} \Phi(p, s) [Y(p, s) - \hat{\beta}_{k-1}^T(s) \hat{\Psi}_k(p, s)]^T \quad (55)$$

$$= \left[\sum_{l=s-p+1}^s \phi_\alpha(l) \phi_\alpha^T(l) \right]^{-1} \sum_{l=s-p+1}^s \phi_\alpha(l) [y(l) - \hat{\beta}_{k-1}^T(s) \hat{\psi}_k(l)]^T,$$

$$\hat{\beta}_k(s) = [\hat{\Psi}_k(p, s) \hat{\Psi}_k^T(p, s)]^{-1} \hat{\Psi}_k(p, s) [Y(p, s) - \hat{\alpha}_{k-1}^T(s) \Phi(p, s)]^T \quad (56)$$

$$= \left[\sum_{l=s-p+1}^s \hat{\psi}_k(l) \hat{\psi}_k^T(l) \right]^{-1} \sum_{l=s-p+1}^s \hat{\psi}_k(l) [y(l) - \hat{\alpha}_{k-1}^T(s) \phi_\alpha(l)]^T,$$

$$Y(p, s) = [y(s), y(s-1), \dots, y(s-p+1)], \quad (57)$$

$$\Phi(p, s) = [\phi_\alpha(s), \phi_\alpha(s-1), \dots, \phi_\alpha(s-p+1)], \quad (58)$$

$$\hat{\Psi}_k(p, s) = [\hat{\psi}_k(s), \hat{\psi}_k(s-1), \dots, \hat{\psi}_k(s-p+1)], \quad (59)$$

$$\boldsymbol{\phi}_\alpha(s) = [-\mathbf{y}^T(s-1), -\mathbf{y}^T(s-2), \dots, -\mathbf{y}^T(s-n_a), \mathbf{u}^T(s-1), \mathbf{u}^T(s-2), \dots, \mathbf{u}^T(s-n_b)]^T, \quad (60)$$

$$\hat{\boldsymbol{\psi}}_k(s) = [-\hat{\mathbf{w}}_{k-1}^T(s-1), -\hat{\mathbf{w}}_{k-1}^T(s-2), \dots, -\hat{\mathbf{w}}_{k-1}^T(s-n_c), \hat{\boldsymbol{\theta}}_{k-1}^T(s-1), \hat{\boldsymbol{\theta}}_{k-1}^T(s-2), \dots, \hat{\boldsymbol{\theta}}_{k-1}^T(s-n_d)]^T, \quad (61)$$

$$\hat{\mathbf{w}}_k(s) = \mathbf{y}(s) - \hat{\boldsymbol{\alpha}}_k^T(s) \boldsymbol{\phi}_\alpha(s), \quad s = 1, 2, \dots, s, \quad (62)$$

$$\hat{\boldsymbol{\theta}}_k(s) = \mathbf{y}(s) - \hat{\boldsymbol{\alpha}}_k^T(s) \boldsymbol{\phi}_\alpha(s) - \hat{\boldsymbol{\beta}}_k^T(s) \hat{\boldsymbol{\psi}}_k(s), \quad (63)$$

$$\hat{\boldsymbol{\theta}}_k^T(s) = [\hat{\boldsymbol{\alpha}}_k^T(s), \hat{\boldsymbol{\beta}}_k^T(s)], \quad (64)$$

$$\hat{\boldsymbol{\alpha}}_k^T(s) = [\hat{A}_{1,k}(s), \hat{A}_{2,k}(s), \dots, \hat{A}_{n_a,k}(s), \hat{B}_{1,k}(s), \hat{B}_{2,k}(s), \dots, \hat{B}_{n_b,k}(s)], \quad (65)$$

$$\hat{\boldsymbol{\beta}}_k^T(s) = [\hat{C}_{1,k}(s), \hat{C}_{2,k}(s), \dots, \hat{C}_{n_c,k}(s), \hat{D}_{1,k}(s), \hat{D}_{2,k}(s), \dots, \hat{D}_{n_d,k}(s)]. \quad (66)$$

The steps of the HMIGI algorithm for computing $\hat{\boldsymbol{\alpha}}_k(s)$ and $\hat{\boldsymbol{\beta}}_k(s)$ are listed as follows.

1. To initialize, choose an innovation length p , let $s = p$, give some small positive ε and set the maximum iterative number $k_{\max}$. Set the initial values: $\hat{\boldsymbol{\theta}}_0 = \begin{bmatrix} \hat{\boldsymbol{\alpha}}_0(s) \\ \hat{\boldsymbol{\beta}}_0(s) \end{bmatrix} = \mathbf{1}_{n \times m} / p_0$, $\hat{\mathbf{w}}_0(s)$ and $\hat{\boldsymbol{\theta}}_0(s)$ to be random vectors.
2. Let $k = 1$ and collect the input-output data $\{\mathbf{u}(s), \mathbf{y}(s)\}$. Form the stacked output matrix $\mathbf{Y}(p, s)$ by Equation (57), form the information vector $\boldsymbol{\phi}_\alpha(s)$ by Equation (60) and the stacked information matrix $\boldsymbol{\Phi}(p, s)$ by Equation (58).
3. Form the information vector $\hat{\boldsymbol{\psi}}_k(s)$ by Equation (61) and the stacked information matrix $\hat{\boldsymbol{\Psi}}_k(p, s)$ by Equation (59).
4. Update the parameter estimates $\hat{\boldsymbol{\alpha}}_k(s)$ by Equation (55) and $\hat{\boldsymbol{\beta}}_k(s)$ by Equation (56).
5. Compute $\hat{\mathbf{w}}_k(s)$ and $\hat{\boldsymbol{\theta}}_k(s)$ by Equations (62) and (63).
6. If $k < k_{\max}$, increase k by 1 and go to Step 3; otherwise, proceed to the next step.
7. Compare $\hat{\boldsymbol{\theta}}_k(s) := \begin{bmatrix} \hat{\boldsymbol{\alpha}}_k(s) \\ \hat{\boldsymbol{\beta}}_k(s) \end{bmatrix}$ and $\hat{\boldsymbol{\theta}}_{k-1}(s)$, for the given small positive ε , if $\|\hat{\boldsymbol{\alpha}}_k(s) - \hat{\boldsymbol{\alpha}}_{k-1}(s)\| + \|\hat{\boldsymbol{\beta}}_k(s) - \hat{\boldsymbol{\beta}}_{k-1}(s)\| > \varepsilon$, let $\hat{\boldsymbol{\theta}}_0(s+1) := \hat{\boldsymbol{\theta}}_k(s)$ and increase s by 1 and go to Step 2; otherwise, obtain the iterative estimate $\hat{\boldsymbol{\theta}}_k(s)$ and terminate this procedure.

6. Example

Consider the following simulation system

$$\begin{aligned} \mathbf{A}(r)\mathbf{y}(s) &= \mathbf{B}(r)\mathbf{u}(s) + \mathbf{C}^{-1}(r)\mathbf{D}(r)\mathbf{v}(s), \\ \mathbf{A}(r) &= \mathbf{I}_2 + \mathbf{A}_1 r^{-1} = \begin{bmatrix} 1 + 0.50r^{-1} & 0.40r^{-1} \\ -0.70r^{-1} & 1 + 0.80r^{-1} \end{bmatrix}, \\ \mathbf{B}(r) &= \mathbf{B}_1 r^{-1} = \begin{bmatrix} 1.00r^{-1} & -0.50r^{-1} \\ -0.40r^{-1} & 1.20r^{-1} \end{bmatrix}, \\ \mathbf{C}(r) &= \mathbf{I}_2 + \mathbf{C}_1 r^{-1} = \begin{bmatrix} 1 - 0.55r^{-1} & 0.20r^{-1} \\ -1.90r^{-1} & 1 + 0.03r^{-1} \end{bmatrix}, \\ \mathbf{D}(r) &= \mathbf{I}_2 + \mathbf{D}_1 r^{-1} = \begin{bmatrix} 1 + 0.05r^{-1} & -0.05r^{-1} \\ -0.05r^{-1} & 1 + 0.05r^{-1} \end{bmatrix}, \\ \boldsymbol{\theta}^T &= [\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1, \mathbf{D}_1]. \end{aligned}$$

In simulation, $\{u_1(s)\}$ and $\{u_2(s)\}$ are taken as two independent persistent excitation signal sequences with zero mean and unit variances, $\{v_1(s)\}$ and $\{v_2(s)\}$ as two independent white noise sequences with zero mean and variances σ_1^2 for $v_1(s)$ and σ_2^2 for $v_2(s)$. Taking the data length $L = 3000$, $\sigma_1^2 = \sigma_2^2 = 0.05^2$, $\sigma_1^2 = \sigma_2^2 = 1.00^2$ and $\sigma_1^2 = \sigma_2^2 = 2.00^2$, respectively. Applying the LSI and

HLSI algorithms to estimate the parameters of this example system, the parameter estimates and the estimation errors $\delta := \|\hat{\theta}_k - \theta\| / \|\theta\|$ of the LSI algorithm are shown in Tables 3–5, and the parameter estimates and the estimation errors $\delta := \sqrt{\frac{\|\hat{\alpha}_k - \alpha\|^2 + \|\hat{\beta}_k - \beta\|^2}{\|\alpha\|^2 + \|\beta\|^2}}$ of the HLSI algorithm are shown in Tables 6–8.

Table 3. The LSI estimates and errors ($\sigma_1^2 = \sigma_2^2 = 0.50^2$).

k	a_{11}	a_{12}	b_{11}	b_{12}	c_{11}	c_{12}	d_{11}	d_{12}
1	0.45255	0.40643	0.98869	−0.50900	−0.00688	0.00264	−0.00796	−0.00359
2	0.48634	0.38699	0.99227	−0.50971	−0.54477	0.22674	−0.00537	−0.00023
3	0.50958	0.39596	0.99203	−0.50976	−0.56916	0.21497	−0.00523	0.00100
4	0.50615	0.40461	0.99211	−0.50968	−0.56544	0.20941	−0.00531	0.00003
5	0.49758	0.40336	0.99219	−0.50966	−0.55529	0.21183	−0.00537	−0.00049
10	0.49925	0.40139	0.99217	−0.50968	−0.55793	0.21329	−0.00534	−0.00021
True values	0.50000	0.40000	1.00000	−0.50000	−0.55000	0.20000	0.05000	−0.05000

k	a_{21}	a_{22}	b_{21}	b_{22}	c_{21}	c_{22}	d_{21}	d_{22}	δ (%)
1	−0.79945	0.73056	−0.38776	1.18668	−0.02309	−0.02288	−0.02250	0.00397	68.20082
2	−0.65715	0.75762	−0.39320	1.19655	−1.91411	0.05582	−0.01435	0.01394	4.08239
3	−0.66034	0.80889	−0.39353	1.19729	−1.91203	−0.00101	−0.01470	0.00996	3.81620
4	−0.70285	0.80968	−0.39302	1.19743	−1.86645	0.00885	−0.01503	0.00676	3.61680
5	−0.71357	0.79623	−0.39296	1.19737	−1.85580	0.02099	−0.01502	0.00710	3.66328
10	−0.70093	0.79640	−0.39309	1.19732	−1.87117	0.01880	−0.01492	0.00802	3.45265
True values	−0.70000	0.80000	−0.40000	1.20000	−1.90000	0.03000	−0.05000	0.05000	

Table 4. The LSI estimates and errors ($\sigma_1^2 = \sigma_2^2 = 1.00^2$).

k	a_{11}	a_{12}	b_{11}	b_{12}	c_{11}	c_{12}	d_{11}	d_{12}
1	0.37911	0.41821	0.97582	−0.51920	−0.02077	0.00610	−0.01571	−0.00216
2	0.45862	0.36770	0.98462	−0.51944	−0.51850	0.24583	−0.01069	−0.00056
3	0.52330	0.37841	0.98443	−0.51962	−0.58539	0.22828	−0.01071	0.00344
4	0.52177	0.41221	0.98429	−0.51938	−0.58362	0.20060	−0.01068	0.00061
5	0.49205	0.40950	0.98436	−0.51932	−0.54986	0.20628	−0.01068	−0.00126
10	0.49662	0.40205	0.98442	−0.51939	−0.55669	0.21270	−0.01067	−0.00036
True values	0.50000	0.40000	1.00000	−0.50000	−0.55000	0.20000	0.05000	−0.05000

k	a_{21}	a_{22}	b_{21}	b_{22}	c_{21}	c_{22}	d_{21}	d_{22}	δ (%)
1	−0.93419	0.64539	−0.39021	1.17886	−0.05758	−0.04446	−0.04030	0.01929	67.62883
2	−0.59263	0.66521	−0.38479	1.19285	−1.97468	0.14646	−0.02931	0.03218	8.64671
3	−0.57894	0.81485	−0.38713	1.19469	−1.99036	−0.01829	−0.02970	0.02306	6.64949
4	−0.69557	0.83966	−0.38683	1.19518	−1.86685	−0.02063	−0.02958	0.01271	4.35994
5	−0.74887	0.79252	−0.38647	1.19502	−1.81223	0.02568	−0.02959	0.01277	4.77498
10	−0.70220	0.78945	−0.38650	1.19486	−1.86665	0.02395	−0.02957	0.01625	3.46273
True values	−0.70000	0.80000	−0.40000	1.20000	−1.90000	0.03000	−0.05000	0.05000	

Table 5. The LSI estimates and errors ($\sigma_1^2 = \sigma_2^2 = 2.00^2$).

k	a_{11}	a_{12}	b_{11}	b_{12}	c_{11}	c_{12}	d_{11}	d_{12}
1	0.30598	0.43273	0.95321	−0.54078	−0.05132	0.00896	−0.03407	0.00703
2	0.42316	0.35154	0.96908	−0.53879	−0.48423	0.26160	−0.02134	−0.00104
3	0.54052	0.33373	0.97007	−0.53902	−0.60385	0.27146	−0.02198	0.00733
4	0.55436	0.42567	0.96900	−0.53866	−0.61803	0.18606	−0.02160	0.00172
5	0.48269	0.42471	0.96851	−0.53864	−0.54117	0.19036	−0.02131	−0.00235
10	0.49093	0.40387	0.96891	−0.53872	−0.55204	0.21020	−0.02138	−0.00053
True values	0.50000	0.40000	1.00000	−0.50000	−0.55000	0.20000	0.05000	−0.05000

k	a_{21}	a_{22}	b_{21}	b_{22}	c_{21}	c_{22}	d_{21}	d_{22}	δ (%)
1	−1.05293	0.57820	−0.40261	1.16284	−0.12697	−0.08390	−0.08149	0.05918	66.29107
2	−0.52615	0.53991	−0.36897	1.18798	−2.03624	0.27031	−0.05907	0.05664	15.54031
3	−0.44989	0.78904	−0.37305	1.18991	−2.11502	0.00375	−0.06016	0.04746	12.73434
4	−0.67116	0.92651	−0.37609	1.19057	−1.88402	−0.10881	−0.05836	0.02468	8.21885
5	−0.82963	0.79822	−0.37539	1.19022	−1.72334	0.01847	−0.05820	0.02430	8.51037
10	−0.71231	0.78084	−0.37425	1.19008	−1.85093	0.03082	−0.05865	0.03211	4.13713
True values	−0.70000	0.80000	−0.40000	1.20000	−1.90000	0.03000	−0.05000	0.05000	

Table 6. The HLSI estimates and errors ($\sigma_1^2 = \sigma_2^2 = 0.50^2$).

k	a_{11}	a_{12}	b_{11}	b_{12}	c_{11}	c_{12}	d_{11}	d_{12}
1	0.45263	0.40637	0.98875	−0.50909	−0.04541	0.02727	0.02581	0.01375
2	0.45491	0.40356	0.98880	−0.51009	−0.51110	0.20729	−0.00526	0.00002
3	0.48853	0.39232	0.99200	−0.50967	−0.51309	0.20921	−0.00489	0.00030
4	0.50095	0.39372	0.99204	−0.50988	−0.54442	0.21420	−0.00477	0.00107
5	0.50362	0.40108	0.99221	−0.50987	−0.55317	0.21647	−0.00489	0.00036
10	0.49955	0.40119	0.99217	−0.50969	−0.55804	0.21307	−0.00534	−0.00017
True values	0.50000	0.40000	1.00000	−0.50000	−0.55000	0.20000	0.05000	−0.05000

k	a_{21}	a_{22}	b_{21}	b_{22}	c_{21}	c_{22}	d_{21}	d_{22}	δ (%)
1	−0.79936	0.73063	−0.38834	1.18748	0.01185	−0.13866	0.00423	0.01831	69.14794
2	−0.79589	0.74816	−0.39168	1.18917	−1.77185	0.08294	−0.00952	0.01716	7.20851
3	−0.66539	0.75709	−0.39207	1.19513	−1.76251	0.05227	−0.01002	0.01671	6.24422
4	−0.67843	0.79006	−0.39212	1.19557	−1.88355	0.05344	−0.01313	0.01080	3.52902
5	−0.68501	0.80187	−0.39217	1.19649	−1.88049	0.02427	−0.01446	0.00823	3.41217
10	−0.70042	0.79617	−0.39310	1.19731	−1.87272	0.01863	−0.01494	0.00806	3.43715
True values	−0.70000	0.80000	−0.40000	1.20000	−1.90000	0.03000	−0.05000	0.05000	

Table 7. The HLSI estimates and errors ($\sigma_1^2 = \sigma_2^2 = 1.00^2$).

k	a_{11}	a_{12}	b_{11}	b_{12}	c_{11}	c_{12}	d_{11}	d_{12}
1	0.37942	0.41810	0.97612	−0.51907	−0.06555	0.02995	0.02371	0.03103
2	0.38912	0.41194	0.97660	−0.52021	−0.43935	0.19532	−0.01176	0.00257
3	0.45524	0.38505	0.98314	−0.51937	−0.45002	0.19959	−0.01084	0.00294
4	0.48057	0.37803	0.98320	−0.51970	−0.51466	0.21120	−0.01033	0.00417
5	0.50132	0.39142	0.98407	−0.51981	−0.52689	0.22128	−0.00933	0.00318
10	0.49770	0.39981	0.98448	−0.51946	−0.56257	0.21197	−0.01094	−0.00043
True values	0.50000	0.40000	1.00000	−0.50000	−0.55000	0.20000	0.05000	−0.05000

k	a_{21}	a_{22}	b_{21}	b_{22}	c_{21}	c_{22}	d_{21}	d_{22}	δ (%)
1	−0.93342	0.64534	−0.39098	1.18112	0.02877	−0.15072	−0.00904	−0.00210	70.25212
2	−0.93091	0.68495	−0.39579	1.18382	−1.63403	0.16609	−0.02072	0.04766	15.04949
3	−0.64662	0.66320	−0.38416	1.18926	−1.61660	0.09883	−0.02194	0.04712	12.33823
4	−0.65462	0.72677	−0.38511	1.18991	−1.87187	0.13390	−0.02323	0.02942	5.98792
5	−0.61429	0.75987	−0.38266	1.19109	−1.85595	0.06066	−0.02495	0.02355	5.08516
10	−0.70984	0.79018	−0.38689	1.19508	−1.88460	0.02909	−0.03028	0.01504	3.34293
True values	−0.70000	0.80000	−0.40000	1.20000	−1.90000	0.03000	−0.05000	0.05000	

Table 8. The HLSI estimates and errors ($\sigma_1^2 = \sigma_2^2 = 2.00^2$).

k	a_{11}	a_{12}	b_{11}	b_{12}	c_{11}	c_{12}	d_{11}	d_{12}
1	0.30659	0.43256	0.95390	−0.54000	−0.10583	0.03531	0.01951	0.06559
2	0.33248	0.42280	0.95566	−0.54103	−0.36770	0.18045	−0.02732	0.01179
3	0.40565	0.38654	0.96508	−0.53914	−0.39917	0.18838	−0.02489	0.01057
4	0.44228	0.36624	0.96510	−0.53908	−0.46950	0.19745	−0.02448	0.01207
5	0.47839	0.38036	0.96662	−0.53919	−0.48983	0.22303	−0.02097	0.01083
10	0.50459	0.38505	0.96907	−0.53937	−0.58299	0.21447	−0.02257	−0.00096
True values	0.50000	0.40000	1.00000	−0.50000	−0.55000	0.20000	0.05000	−0.05000

k	a_{21}	a_{22}	b_{21}	b_{22}	c_{21}	c_{22}	d_{21}	d_{22}	δ (%)
1	−1.05128	0.57789	−0.40369	1.16799	0.06260	−0.17483	−0.03557	−0.04291	72.22302
2	−1.06002	0.63997	−0.40964	1.17231	−1.51135	0.23182	−0.05330	0.10898	22.60909
3	−0.65570	0.56386	−0.37654	1.18036	−1.48940	0.13195	−0.05586	0.11046	18.48183
4	−0.64791	0.63257	−0.37934	1.18113	−1.85360	0.21998	−0.04887	0.06935	10.56573
5	−0.51920	0.65844	−0.37044	1.18255	−1.81902	0.13118	−0.04833	0.06187	10.21901
10	−0.72624	0.78837	−0.37432	1.19006	−1.96662	0.08720	−0.05894	0.01722	5.15136
True values	−0.70000	0.80000	−0.40000	1.20000	−1.90000	0.03000	−0.05000	0.05000	

From the simulation results in Tables 3–8, we can draw the following conclusions.

- The estimation errors given by the LSI and HLSI algorithms become smaller as the iterative variable k increases (see the estimation errors δ in Tables 3–8).

- The HLSI algorithm leads to smaller parameter estimation errors than the LSI algorithm for the same data length L and noise level (see Tables 4 and 7).
- With iterations increase, the parameter estimations given by the HLSI algorithm are close to the true parameters, which verify the effectiveness of the HLSI algorithm proposed in this paper (see Tables 6–8).

To validate the obtained model, we use the LSI estimates and the HLSI estimates to construct the estimated model, respectively. The predicted outputs are

$$\hat{\mathbf{y}}(s) = \mathbf{y}(s) - \mathbf{D}^{-1}(r)\mathbf{C}(r)[\mathbf{A}(r)\mathbf{y}(s) - \mathbf{B}(r)\mathbf{u}(s)]. \quad (67)$$

Using the LSI parameter estimates in Table 4 with the noise variance $\sigma_1^2 = \sigma_2^2 = 0.50^2$ and $k = 10$ to construct the LSI estimated model:

$$\begin{aligned} \hat{\mathbf{y}}_{LS}(s) &= \mathbf{y}(s) - \hat{\mathbf{D}}_{LS}^{-1}(r)\hat{\mathbf{C}}_{LS}(r)[\hat{\mathbf{A}}_{LS}(r)\mathbf{y}(s) - \hat{\mathbf{B}}_{LS}(r)\mathbf{u}(s)], \\ \hat{\mathbf{A}}_{LS}(r) &= \mathbf{I}_2 + \hat{\mathbf{A}}_{LS1}r^{-1} = \begin{bmatrix} 1 + 0.49925r^{-1} & 0.40139r^{-1} \\ -0.70093r^{-1} & 1 + 0.7640r^{-1} \end{bmatrix}, \\ \hat{\mathbf{B}}_{LS}(r) &= \hat{\mathbf{B}}_{LS1}r^{-1} = \begin{bmatrix} 0.992172r^{-1} & -0.50968r^{-1} \\ -0.39309r^{-1} & 1.19732r^{-1} \end{bmatrix}, \\ \hat{\mathbf{C}}_{LS}(r) &= \mathbf{I}_2 + \hat{\mathbf{C}}_{LS1}r^{-1} = \begin{bmatrix} 1 - 0.55793r^{-1} & 0.21329r^{-1} \\ -1.87117r^{-1} & 1 + 0.01880r^{-1} \end{bmatrix}, \\ \hat{\mathbf{D}}_{LS}(r) &= \mathbf{I}_2 + \hat{\mathbf{D}}_{LS1}r^{-1} = \begin{bmatrix} 1 - 0.00534r^{-1} & -0.00021r^{-1} \\ -0.01492r^{-1} & 1 + 0.00802r^{-1} \end{bmatrix}. \end{aligned}$$

Using the HLSI parameter estimates in Table 7 with the noise variance $\sigma_1^2 = \sigma_2^2 = 0.50^2$ and $k = 10$ to construct the HLSI estimated model as

$$\begin{aligned} \hat{\mathbf{y}}_H(s) &= \mathbf{y}(s) - \hat{\mathbf{D}}_H^{-1}(r)\hat{\mathbf{C}}_H(r)[\hat{\mathbf{A}}_H(r)\mathbf{y}(s) - \hat{\mathbf{B}}_H(r)\mathbf{u}(s)], \\ \hat{\mathbf{A}}_H(r) &= \mathbf{I}_2 + \hat{\mathbf{A}}_{H1}r^{-1} = \begin{bmatrix} 1 + 0.49955r^{-1} & 0.40119r^{-1} \\ -0.70042r^{-1} & 1 + 0.79617r^{-1} \end{bmatrix}, \\ \hat{\mathbf{B}}_H(r) &= \hat{\mathbf{B}}_{H1}r^{-1} = \begin{bmatrix} 0.99217r^{-1} & -0.50969r^{-1} \\ -0.39310r^{-1} & 1.19731r^{-1} \end{bmatrix}, \\ \hat{\mathbf{C}}_H(r) &= \mathbf{I}_2 + \hat{\mathbf{C}}_{H1}r^{-1} = \begin{bmatrix} 1 - 0.55804r^{-1} & 0.21307r^{-1} \\ -1.87272r^{-1} & 1 + 0.01863r^{-1} \end{bmatrix}, \\ \hat{\mathbf{D}}_H(r) &= \mathbf{I}_2 + \hat{\mathbf{D}}_{H1}r^{-1} = \begin{bmatrix} 1 - 0.00534r^{-1} & -0.00017r^{-1} \\ -0.01494r^{-1} & 1 + 0.00806r^{-1} \end{bmatrix}. \end{aligned}$$

For model validation, we use a different dataset with a length of $L_r = 100$, whose data are from $s = L + 1$ to $s = L + L_r$, to compute the predicted outputs $\hat{\mathbf{y}}(s) = [\hat{y}_1(s), \hat{y}_2(s)]^T \in \mathbb{R}^2$ of the system. The actual outputs $\mathbf{y}_i(s)$ ($i = 1, 2$), the predicted output $\hat{\mathbf{y}}_{LSI}(s)$ based on the LSI parameter estimates, and the predicted output $\hat{\mathbf{y}}_{HLSI}(s)$ based on the HLSI parameter estimates are plotted in Figures 1 and 2. Define the root-mean-square errors (RMSEs) as

$$Error(\hat{\mathbf{y}}) := \sqrt{\frac{1}{L_r} \sum_{s=L}^{L+L_r} [\hat{\mathbf{y}}(s) - \mathbf{y}(s)]^2}. \quad (68)$$

Using the estimated outputs to compute the root-mean-square errors for the LSI and HLSI estimated models as

$$\text{Error}(\hat{\mathbf{y}}_{LS}) := \sqrt{\frac{1}{L_r} \sum_{s=L}^{L+L_r} [\hat{\mathbf{y}}_{LS}(s) - \mathbf{y}(s)]^2} = [0.89203, 1.24336]^T \in \mathbb{R}^2,$$

$$\text{Error}(\hat{\mathbf{y}}_H) := \sqrt{\frac{1}{L_r} \sum_{s=L}^{L+L_r} [\hat{\mathbf{y}}_H(s) - \mathbf{y}(s)]^2} = [0.65907, 1.22223]^T \in \mathbb{R}^2.$$

Figure 1. The true output $y_1(s)$, the predicted outputs $\hat{y}_{LSI1}(s)$ and $\hat{y}_{HLSI1}(s)$ versus s .

Figure 2. The true output $y_2(s)$, the predicted outputs $\hat{y}_{LSI2}(s)$ and $\hat{y}_{HLSI2}(s)$ versus s .

Remark 3. Figures 1 and 2 demonstrate that the predicted outputs are close to the measurement outputs, which means that the LSI and HLSI estimation algorithms are effective and they can give accurate estimates of the multivariable CARARMA systems.

Remark 4. In Figures 1 and 2, we can see that the predicted outputs of the LSI and HLSI algorithms are close to true values for $y_1(s)$. However, for $y_2(s)$, the predicted outputs of the LSI algorithm has a significant discrepancy while the HLSI estimates remain close to true values, which shows that the LSI method has a limitation for application to this example. On the other hand, the proposed HLSI algorithm achieves good estimation performance for identifying the multivariable CARARMA systems. Generally speaking, the combination of the

decomposition techniques and the iterative technique results in more accurate parameter estimates and better computational efficiency for multivariable systems.

7. Conclusions

This work proposes a hierarchical least-squares-based iterative identification algorithm and a hierarchical multi-innovation least-squares-based iterative identification algorithm for multivariable CARARMA systems based on the hierarchical identification principle and the iterative technique. Compared with the least-squares-based iterative identification algorithm, the proposed algorithms achieve highly accurate parameter estimates and improve the performance the algorithms. The proposed decomposition least-squares-based iterative identification algorithms for multivariable equation-error autoregressive moving average systems can combine other estimation methods [71–74] and the mathematical techniques [75–77] to explore the parameter identification methods of other scalar, multivariable linear, nonlinear systems with colored noises [78–80], and can be extended to other scientific fields [81–88] such as signal modeling and communication networked systems [89–92].

Author Contributions: Conceptualization and methodology, L.W. and F.D.; and validation and analysis, X.L. and C.C.

Funding: This work was supported by the National Natural Science Foundation of China (No. 61304093) and the Natural Science Foundation of Shandong Province (ZR201702170236).

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

References

1. Xu, L.; Chen, L.; Xiong, W.L. Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration. *Nonlinear Dyn.* **2015**, *79*, 2155–2163. [\[CrossRef\]](#)
2. Xu, L. The parameter estimation algorithms based on the dynamical response measurement data. *Adv. Mech. Eng.* **2017**, *9*, 1–12. [\[CrossRef\]](#)
3. Pan, J.; Jiang, X.; Wan, X.K.; Ding, W. A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems. *Int. J. Control Autom. Syst.* **2017**, *15*, 1189–1197. [\[CrossRef\]](#)
4. Liu, Q.Y.; Ding, F.; Xu, L.; Yang, E.F. Partially coupled gradient estimation algorithm for multivariable equation-error autoregressive moving average systems using the data filtering technique. *IET Control Theory Appl.* **2019**, *13*, 642–650. [\[CrossRef\]](#)
5. Xu, L. The damping iterative parameter identification method for dynamical systems based on the sine signal measurement. *Signal Process* **2016**, *120*, 660–667. [\[CrossRef\]](#)
6. Wang, D.Q.; Li, L.W.; Ji, Y.; Yan, Y.R. Model recovery for Hammerstein systems using the auxiliary model based orthogonal matching pursuit method. *Appl. Math. Model.* **2018**, *54*, 537–550. [\[CrossRef\]](#)
7. Wang, D.Q.; Yan, Y.R.; Liu, Y.J.; Ding, J.H. Model recovery for Hammerstein systems using the hierarchical orthogonal matching pursuit method. *J. Comput. Appl. Math.* **2019**, *345*, 135–145. [\[CrossRef\]](#)
8. Xu, L. A proportional differential control method for a time-delay system using the Taylor expansion approximation. *Appl. Math. Comput.* **2014**, *236*, 391–399. [\[CrossRef\]](#)
9. Liu, Q.Y.; Ding, F. Auxiliary model-based recursive generalized least squares algorithm for multivariate output-error autoregressive systems using the data filtering. *Circuits Syst. Signal Process.* **2019**, *38*, 590–610. [\[CrossRef\]](#)
10. Ge, Z.W.; Ding, F.; Xu, L.; Alsaedi, A.; Hayat, T. Gradient-based iterative identification method for multivariate equation-error autoregressive moving average systems using the decomposition technique. *J. Frankl. Inst.* **2019**, *356*, 658–1676. [\[CrossRef\]](#)
11. Pan, J.; Ma, H.; Jiang, X.; Ding, W. Adaptive gradient-based iterative algorithm for multivariate controlled autoregressive moving average systems using the data filtering technique. *Complexity* **2018**, *2018*, 9598307. [\[CrossRef\]](#)
12. Li, J.H.; Zheng, W.X.; Gu, J.P.; Hua, L. A recursive identification algorithm for Wiener nonlinear systems with linear state-space subsystem. *Circuits Syst. Signal Process.* **2018**, *37*, 2374–2393. [\[CrossRef\]](#)

13. Xu, H.; Ding, F.; Yang, E.F. Modeling a nonlinear process using the exponential autoregressive time series model. *Nonlinear Dyn.* **2019**, *95*, 2079–2092.
14. Ding, J.L. The hierarchical iterative identification algorithm for multi-input-output-error systems with autoregressive noise. *Complexity* **2017**, *2017*, 5292894. [[CrossRef](#)]
15. Ding, J.L. Recursive and iterative least squares parameter estimation algorithms for multiple-input-output-error systems with autoregressive noise. *Circuits Syst. Signal Process.* **2018**, *37*, 1884–1906. [[CrossRef](#)]
16. Xu, L. Application of the Newton iteration algorithm to the parameter estimation for dynamical systems. *J. Comput. Appl. Math.* **2015**, *288*, 33–43. [[CrossRef](#)]
17. Liu, S.Y.; Ding, F.; Xu, L.; Hayat, T. Hierarchical principle-based iterative parameter estimation algorithm for dual-frequency signals. *Circuits Syst. Signal Process.* **2019**, *38*, 3251–3268. [[CrossRef](#)]
18. Wang, Y.J.; Ding, F. A filtering based multi-innovation gradient estimation algorithm and performance analysis for nonlinear dynamical systems. *IMA J. Appl. Math.* **2017**, *82*, 1171–1191. [[CrossRef](#)]
19. Wang, Y.J.; Ding, F. Novel data filtering based parameter identification for multiple-input multiple-output systems using the auxiliary model. *Automatica* **2016**, *71*, 308–313. [[CrossRef](#)]
20. Ding, J.; Chen, J.Z.; Lin, J.X.; Wan, L.J. Particle filtering based parameter estimation for systems with output-error type model structures. *J. Frankl. Inst.* **2019**, *356*, 5521–5540. [[CrossRef](#)]
21. Zhang, X.; Xu, L.; Ding, F.; Hayat, T. Combined state and parameter estimation for a bilinear state space system with moving average noise. *J. Frankl. Inst.* **2018**, *355*, 3079–3103. [[CrossRef](#)]
22. Zhang, X.; Ding, F.; Alsaadi, F.E.; Hayat, T. Recursive parameter identification of the dynamical models for bilinear state space systems. *Nonlinear Dyn.* **2017**, *89*, 2415–2429. [[CrossRef](#)]
23. Ding, F.; Chen, H.B.; Xu, L.; Dai, J.Y.; Li, Q.S.; Hayat, T. A hierarchical least squares identification algorithm for Hammerstein nonlinear systems using the key term separation. *J. Frankl. Inst.* **2018**, *355*, 3737–3752. [[CrossRef](#)]
24. Xu, L.; Ding, F. Parameter estimation for control systems based on impulse responses. *Int. J. Control Autom. Syst.* **2017**, *15*, 2471–2479. [[CrossRef](#)]
25. Xu, L.; Ding, F. Parameter estimation algorithms for dynamical response signals based on the multi-innovation theory and the hierarchical principle. *IET Signal Process.* **2017**, *11*, 228–237. [[CrossRef](#)]
26. 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](#)]
27. Ding, F. Coupled-least-squares identification for multivariable systems. *IET Control Theory Appl.* **2013**, *7*, 68–79. [[CrossRef](#)]
28. Ding, F. Two-stage least squares based iterative estimation algorithm for CARARMA system modeling. *Appl. Math. Model.* **2013**, *37*, 4798–4808. [[CrossRef](#)]
29. Ding, F.; Wang, Y.J.; Dai, J.Y.; Li, Q.S.; Chen, Q.J. A recursive least squares parameter estimation algorithm for output nonlinear autoregressive systems using the input-output data filtering. *J. Frankl. Inst.* **2017**, *354*, 6938–6955. [[CrossRef](#)]
30. Xu, L.; Ding, F. Recursive least squares and multi-innovation stochastic gradient parameter estimation methods for signal modeling. *Circuits Syst. Signal Process.* **2017**, *36*, 1735–1753. [[CrossRef](#)]
31. Ding, F. Decomposition based fast least squares algorithm for output error systems. *Signal Process.* **2013**, *93*, 1235–1242. [[CrossRef](#)]
32. Zhang, X.; Ding, F.; Xu, L.; Yang, E.F. State filtering-based least squares parameter estimation for bilinear systems using the hierarchical identification principle. *IET Control Theory Appl.* **2018**, *12*, 1704–1713. [[CrossRef](#)]
33. Cao, Y.; Wang, Z.; Liu, F.; Li, P.; Xie, G. Bio-inspired speed curve optimization and sliding mode tracking control for subway trains. *IEEE Trans. Veh. Technol.* **2019**. [[CrossRef](#)]
34. Cao, Y.; Lu, H.; Wen, T. A safety computer system based on multi-sensor data processing. *Sensors* **2019**, *19*, 818. [[CrossRef](#)]
35. Cao, Y.; Zhang, Y.; Wen, T.; Li, P. Research on dynamic nonlinear input prediction of fault diagnosis based on fractional differential operator equation in high-speed train control system. *Chaos* **2019**, *29*, 013130. [[CrossRef](#)]
36. Cao, Y.; Li, P.; Zhang, Y. Parallel processing algorithm for railway signal fault diagnosis data based on cloud computing. *Future Gener. Comput. Syst.* **2018**, *88*, 279–283. [[CrossRef](#)]

37. Cao, Y.; Ma, L.C.; Xiao, S.; Zhang, X.; Xu, W. Standard analysis for transfer delay in CTCS-3. *Chin. J. Electron.* **2017**, *26*, 1057–1063. [\[CrossRef\]](#)
38. Salhi, H.; Kamoun, S. A recursive parametric estimation algorithm of multivariable nonlinear systems described by Hammerstein mathematical models. *Appl. Math. Model.* **2015**, *39*, 4951–4962. [\[CrossRef\]](#)
39. Wan, L.J.; Ding, F. Decomposition- and gradient-based iterative identification algorithms for multivariable systems using the multi-innovation theory. *Circuits Syst. Signal Process.* **2019**, *38*, 2971–2991. [\[CrossRef\]](#)
40. Li, M.H.; Liu, X.M. The least squares based iterative algorithms for parameter estimation of a bilinear system with autoregressive noise using the data filtering technique. *Signal Process.* **2018**, *147*, 23–34. [\[CrossRef\]](#)
41. Xu, L.; Ding, F. Iterative parameter estimation for signal models based on measured data. *Circuits Syst. Signal Process.* **2018**, *37*, 3046–3069. [\[CrossRef\]](#)
42. Xu, L.; Xiong, W.L.; Alsaedi, A.; Hayat, T. Hierarchical parameter estimation for the frequency response based on the dynamical window data. *Int. J. Control Autom. Syst.* **2018**, *16*, 1756–1764. [\[CrossRef\]](#)
43. Zhang, X.; Ding, F.; Xu, L.; Alsaedi, A.; Hayat, T. A hierarchical approach for joint parameter and state estimation of a bilinear system with autoregressive noise. *Mathematics* **2019**, *7*, 356. [\[CrossRef\]](#)
44. Ding, F.; Pan, J.; Alsaedi, A.; Hayat, T. Gradient-based iterative parameter estimation algorithms for dynamical systems from observation data. *Mathematics* **2019**, *7*, 428. [\[CrossRef\]](#)
45. Ma, H.; Pan, J.; Lv, L.; Xu, G.H.; Ding, F.; Alsaedi, A.; Hayat, T. Recursive algorithms for multivariable output-error-like ARMA systems. *Mathematics* **2019**, *7*, 558. [\[CrossRef\]](#)
46. Wang, Y.J.; Ding, F.; Xu, L. Some new results of designing an IIR filter with colored noise for signal processing. *Digit. Signal Process.* **2018**, *72*, 44–58. [\[CrossRef\]](#)
47. Zhang, X.; Ding, F.; Xu, L.; Yang, E.F. Highly computationally efficient state filter based on the delta operator. *Int. J. Adapt. Control Signal Process.* **2019**, *33*, 875–889. [\[CrossRef\]](#)
48. Yin, C.C.; Wen, Y.Z.; Zhao, Y.X. On the optimal dividend problem for a spectrally positive levy process. *Astin Bull.* **2014**, *44*, 635–651. [\[CrossRef\]](#)
49. Yin, C.C.; Wen, Y.Z. Optimal dividend problem with a terminal value for spectrally positive Levy processes. *Insur. Math. Econ.* **2013**, *53*, 769–773. [\[CrossRef\]](#)
50. Yin, C.C.; Zhao, J.S. Nonexponential asymptotics for the solutions of renewal equations, with applications. *J. Appl. Probab.* **2006**, *43*, 815–824. [\[CrossRef\]](#)
51. Yin, C.C.; Wang, C.W. The perturbed compound Poisson risk process with investment and debit interest. *Methodol. Comput. Appl. Probab.* **2010**, *12*, 391–413. [\[CrossRef\]](#)
52. Yin, C.C.; Wen, Y.Z. Exit problems for jump processes with applications to dividend problems. *J. Comput. Appl. Math.* **2013**, *245*, 30–52. [\[CrossRef\]](#)
53. Wen, Y.Z.; Yin, C.C. Solution of Hamilton-Jacobi-Bellman equation in optimal reinsurance strategy under dynamic VaR constraint. *J. Funct. Spaces* **2019**, *2019*, 6750892. [\[CrossRef\]](#)
54. Sha, X.Y.; Xu, Z.S.; Yin, C.C. Elliptical distribution-based weight-determining method for ordered weighted averaging operators. *Int. J. Intell. Syst.* **2019**, *34*, 858–877. [\[CrossRef\]](#)
55. Pan, J.; Li, W.; Zhang, H.P. Control algorithms of magnetic suspension systems based on the improved double exponential reaching law of sliding mode control. *Int. J. Control Autom. Syst.* **2018**, *16*, 2878–2887. [\[CrossRef\]](#)
56. Li, M.H.; Liu, X.M. Auxiliary model based least squares iterative algorithms for parameter estimation of bilinear systems using interval-varying measurements. *IEEE Access* **2018**, *6*, 21518–21529. [\[CrossRef\]](#)
57. Sun, Z.Y.; Zhang, D.; Meng, Q.; Chen, C.C. Feedback stabilization of time-delay nonlinear systems with continuous time-varying output function. *Int. J. Syst. Sci.* **2019**, *50*, 244–255. [\[CrossRef\]](#)
58. Zhan, X.S.; Cheng, L.L.; Wu, J.; Yang, Q.S.; Han, T. Optimal modified performance of MIMO networked control systems with multi-parameter constraints. *ISA Trans.* **2019**, *84*, 111–117. [\[CrossRef\]](#)
59. Zhan, X.S.; Guan, Z.H.; Zhang, X.H.; Yuan, F.S. Optimal tracking performance and design of networked control systems with packet dropout. *J. Frankl. Inst.* **2013**, *350*, 3205–3216. [\[CrossRef\]](#)
60. Jiang, C.M.; Zhang, F.F.; Li, T.X. Synchronization and antisynchronization of N-coupled fractional-order complex chaotic systems with ring connection. *Math. Methods Appl. Sci.* **2018**, *41*, 2625–2638. [\[CrossRef\]](#)
61. Wang, T.; Liu, L.; Zhang, J.; Schaeffer, E.; Wang, Y. A M-EKF fault detection strategy of insulation system for marine current turbine. *Mech. Syst. Signal Process.* **2019**, *115*, 269–280. [\[CrossRef\]](#)
62. Gong, P.C.; Wang, W.Q.; Wan, X.R. Adaptive weight matrix design and parameter estimation via sparse modeling for MIMO radar. *Signal Process.* **2017**, *139*, 1–11. [\[CrossRef\]](#)

63. Gong, P.C.; Wang, W.Q.; Li, F.C.; Cheung, H. Sparsity-aware transmit beamspace design for FDA-MIMO radar. *Signal Process.* **2018**, *144*, 99–103. [[CrossRef](#)]
64. Wan, X.K.; Li, Y.; Xia, C.; Wu, M.H.; Liang, J.; Wang, N. A T-wave alternans assessment method based on least squares curve fitting technique. *Measurement* **2016**, *86*, 93–100. [[CrossRef](#)]
65. Zhao, N. Joint Optimization of cooperative spectrum sensing and resource allocation in multi-channel cognitive radio sensor networks. *Circuits Syst. Signal Process.* **2016**, *35*, 2563–2583. [[CrossRef](#)]
66. Zhao, N.; Chen, Y.; Liu, R.; Wu, M.H.; Xiong, W. Monitoring strategy for relay incentive mechanism in cooperative communication networks. *Comput. Electr. Eng.* **2017**, *60*, 14–29. [[CrossRef](#)]
67. Zhao, N.; Wu, M.H.; Chen, J.J. Android-based mobile educational platform for speech signal processing. *Int. J. Electr. Eng. Edu.* **2017**, *54*, 3–16. [[CrossRef](#)]
68. Zhao, N.; Liang, Y.; Pei, Y. Dynamic contract incentive mechanism for cooperative wireless networks. *IEEE Trans. Veh. Technol.* **2018**, *67*, 10970–10982. [[CrossRef](#)]
69. Zhao, X.L.; Liu, F.; Fu, B.; Na, F. Reliability analysis of hybrid multi-carrier energy systems based on entropy-based Markov model. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2016**, *230*, 561–569. [[CrossRef](#)]
70. Zhao, X.L.; Lin, Z.Y.; Fu, B.; He, L.; Na, F. Research on automatic generation control with wind power participation based on predictive optimal 2-degree-of-freedom PID strategy for multi-area interconnected power system. *Energies* **2018**, *11*, 3325. [[CrossRef](#)]
71. Xu, L.; Ding, F.; Gu, Y.; Alsaedi, A.; Hayat, T. A multi-innovation state and parameter estimation algorithm for a state space system with d-step state-delay. *Signal Process.* **2017**, *140*, 97–103. [[CrossRef](#)]
72. Gu, Y.; Liu, J.; Li, X.; Chou, Y.; Ji, Y. State space model identification of multirate processes with time-delay using the expectation maximization. *J. Frankl. Inst.* **2019**, *356*, 1623–1639. [[CrossRef](#)]
73. Gu, Y.; Chou, Y.; Liu, J.; Ji, Y. Moving horizon estimation for multirate systems with time-varying time-delays. *J. Frankl. Inst.* **2019**, *356*, 2325–2345. [[CrossRef](#)]
74. Ding, F.; Xu, L.; Alsaadi, F.E.; Hayat, T. Iterative parameter identification for pseudo-linear systems with ARMA noise using the filtering technique. *IET Control Theory Appl.* **2018**, *12*, 892–899. [[CrossRef](#)]
75. Liu, F.; Xue, Q.; Yabuta, K. Boundedness and continuity of maximal singular integrals and maximal functions on Triebel-Lizorkin spaces. *Sci. China Math.* **2019**. [[CrossRef](#)]
76. Liu, F. Boundedness and continuity of maximal operators associated to polynomial compound curves on Triebel-Lizorkin spaces. *Math. Inequal. Appl.* **2019**, *22*, 25–44. [[CrossRef](#)]
77. Liu, F.; Fu, Z.; Jhang, S. Boundedness and continuity of Marcinkiewicz integrals associated to homogeneous mappings on Triebel-Lizorkin spaces. *Front. Math. China* **2019**, *14*, 95–122. [[CrossRef](#)]
78. Zhang, W.H.; Xue, L.; Jiang, X. Global stabilization for a class of stochastic nonlinear systems with SISS-like conditions and time delay. *Int. J. Robust Nonlinear Control* **2018**, *28*, 3909–3926. [[CrossRef](#)]
79. Li, N.; Guo, S.; Wang, Y. Weighted preliminary-summation-based principal component analysis for non-Gaussian processes. *Control Eng. Pract.* **2019**, *87*, 122–132. [[CrossRef](#)]
80. Wang, Y.; Si, Y.; Huang, B.; Lou, Z. Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008–2017. *Can. J. Chem. Eng.* **2018**, *96*, 2073–2085. [[CrossRef](#)]
81. Wu, M.H.; Li, X.; Liu, C.; Liu, M.; Zhao, N.; Wang, J.; Wan, X.K.; Rao, Z.H.; Zhu, L. Robust global motion estimation for video security based on improved k-means clustering. *J. Ambient Intell. Humaniz. Comput.* **2019**, *10*, 439–448. [[CrossRef](#)]
82. Wan, X.K.; Wu, H.; Qiao, F.; Li, F.; Li, Y.; Yan, Y.; Wei, J. Electrocardiogram baseline wander suppression based on the combination of morphological and wavelet transformation based filtering. *Computat. Math. Methods Med.* **2019**, *2019*, 7196156. [[CrossRef](#)] [[PubMed](#)]
83. Feng, L.; Li, Q.X.; Li, Y.F. Imaging with 3-D aperture synthesis radiometers. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2395–2406. [[CrossRef](#)]
84. Shi, W.X.; Liu, N.; Zhou, Y.M.; Cao, X.A. Effects of postannealing on the characteristics and reliability of polyfluorene organic light-emitting diodes. *IEEE Trans. Electron Devices* **2019**, *66*, 1057–1062. [[CrossRef](#)]
85. Fu, B.; Ouyang, C.X.; Li, C.S.; Wang, J.W.; Gul, E. An improved mixed integer linear programming approach based on symmetry diminishing for unit commitment of hybrid power system. *Energies* **2019**, *12*, 833. [[CrossRef](#)]

86. Wu, T.Z.; Shi, X.; Liao, L.; Zhou, C.J.; Zhou, H.; Su, Y.H. A capacity configuration control strategy to alleviate power fluctuation of hybrid energy storage system based on improved particle swarm optimization. *Energies* **2019**, *12*, 642. [[CrossRef](#)]
87. Zhao, X.L.; Lin, Z.Y.; Fu, B.; He, L.; Li, C.S. Research on the predictive optimal PID plus second order derivative method for AGC of power system with high penetration of photovoltaic and wind power. *J. Electr. Eng. Technol.* **2019**, *14*, 1075–1086. [[CrossRef](#)]
88. Liu, N.; Mei, S.; Sun, D.; Shi, W.; Feng, J.; Zhou, Y.M.; Mei, F.; Xu, J.; Jiang, Y.; Cao, X.A. Effects of charge transport materials on blue fluorescent organic light-emitting diodes with a host-dopant system. *Micromachines* **2019**, *10*, 344. [[CrossRef](#)]
89. Tian, X.P.; Niu, H.M. A bi-objective model with sequential search algorithm for optimizing network-wide train timetables. *Comput. Ind. Eng.* **2019**, *127*, 1259–1272. [[CrossRef](#)]
90. Li, X.Y.; Li, H.X.; Wu, B.Y. Piecewise reproducing kernel method for linear impulsive delay differential equations with piecewise constant arguments. *Appl. Math. Comput.* **2019**, *349*, 304–313. [[CrossRef](#)]
91. Ma, F.Y.; Yin, Y.K.; Li, M. Start-up process modelling of sediment microbial fuel cells based on data driven. *Math. Probl. Eng.* **2019**, *2019*, 7403732. [[CrossRef](#)]
92. Yang, F.; Zhang, P.; Li, X.X. The truncation method for the Cauchy problem of the inhomogeneous Helmholtz equation. *Appl. Anal.* **2019**, *98*, 991–1004. [[CrossRef](#)]

© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).