

Article

Parameter Estimation Algorithms for Hammerstein Finite Impulse Response Moving Average Systems Using the Data Filtering Theory

Yan Ji ^{1,2} and Jinde Cao ^{1,3,*} ¹ School of Mathematics, Southeast University, Nanjing 210096, China; yjchinaqd@qust.edu.cn² College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China³ Yonsei Frontier Lab, Yonsei University, Seoul 03722, Korea

* Correspondence: jdcao@seu.edu.cn

Abstract: This paper considers the parameter estimation problems of Hammerstein finite impulse response moving average (FIR–MA) systems. Based on the matrix transformation and the hierarchical identification principle, the Hammerstein FIR–MA system is recast into two models, and a decomposition-based recursive least-squares algorithm is deduced for estimating the parameters of these two models. In order to further improve the accuracy of the parameter estimation, a multi-innovation hierarchical least-squares algorithm based on the data filtering theory proposed. Finally, a simulation example demonstrates the effectiveness of the proposed scheme.

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



Citation: Ji, Y.; Cao, J. Parameter Estimation Algorithms for Hammerstein Finite Impulse Response Moving Average Systems Using the Data Filtering Theory. *Mathematics* **2022**, *10*, 438. <https://doi.org/10.3390/math10030438>

Academic Editor: Ripon Kumar Chakraborty

Received: 1 December 2021

Accepted: 27 January 2022

Published: 29 January 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

The identification methods have been very mature for multivariable linear systems [1,2]. However, many systems in practical applications and industrial control are nonlinear multivariable systems [3,4]. Nonlinear multivariable systems are different from scalar systems or linear systems, and each of their outputs is usually controlled and affected by several inputs at the same time, so the parameter estimation of multivariable systems is more difficult [5–7]. In recent years, many scholars have studied the parameter estimation problem of nonlinear multivariable systems and proposed many identification methods. You et al. proposed an iterative algorithm for the identification of multiple-input single-output output-error systems with unknown time-delays, based on the basis pursuit denoising criterion and the auxiliary model identification idea [8]. Wang and Ding presented a novel recursive least-squares algorithm for multiple-input multiple-output (MIMO) systems with autoregressive moving average noise, employing the auxiliary model and the data filtering technique [9]. Lenka studied the asymptotic stability of equilibrium points of multivariable fractional order systems based on the fractional comparison principle and the Laplace transform [10]. Liu et al. investigated parameter estimation problems for multivariable controlled autoregressive autoregressive moving average systems, and derived a partially coupled generalized extended stochastic gradient algorithm by using the auxiliary model [11].

Least-squares and stochastic gradient algorithms are two methods for parameter estimation [12–15]. Compared with the stochastic gradient algorithm, the recursive least-squares algorithm has fast convergence rate and high computational effort [16]. The recursive least-squares algorithm needs to calculate the inverse of the covariance matrix in the identification process; when the dimension of the covariance matrix is high, the computational complexity is very large. In order to reduce the complexity of the calculation,

the decomposition technique can effectively solve the problem of a large computational burden in large-scale system identification [17–19]. The decomposition technique is used to decompose the original system into multiple subsystems and to identify the parameters in each subsystem separately [20,21]. For example, Wang et al. presented a hierarchical least-squares iterative algorithm to solve the difficulty that the identification model contains the unmeasurable variables and noise terms in the information matrix [22]. Ji et al. derived a hierarchical least-squares algorithm for two-input Hammerstein finite impulse response systems based on the hierarchical identification principle and the multi-innovation theory [23]. Wang et al. transformed a MIMO Hammerstein system with different types of coefficients into an over-parametrisation regression identification model, and then proposed a hierarchical extended stochastic gradient algorithm [24].

In order to reduce the error of parameter estimation, the data filtering technique can be used to get rid of the outliers and weaken the influence of noises to better extract useful information in the data [25,26]. Chen et al. combined the maximum likelihood principle, the decomposition technique and the data filtering technique, to present a maximum likelihood generalized extended gradient algorithm, and a data filtering-based maximum likelihood extended gradient algorithm [27]. Mao et al. presented a data filtering multi-innovation stochastic gradient identification algorithm for Hammerstein output-error autoregressive systems by means of the multi-innovation identification theory [28]. By eliminating the state variables in the systems, Li et al. proposed a filtering-based least-squares iterative algorithm for estimating the parameters of bilinear systems with colored noises [29]. The filtering technique can also be applied to signal processing and communication [30–32], and neural networks [33,34].

The parameter estimation of the system models is important for control system analysis and design. The parameters of the models can be estimated by using some identification methods [35–37] such as the hierarchical algorithms [38,39]. This paper studies the parameter estimation problems of two-input two-output Hammerstein finite impulse response moving average (FIR–MA) systems. The Hammerstein nonlinear system includes the nonlinear block and the linear block and its identification face more difficulties such as high computational complexity and parameter identifiability. In this paper, the decomposition technique and data filtering technique are employed to solve these problems. The main contributions of this paper lie in the following.

- Based on the decomposition technique, we decomposed the Hammerstein system into two models, each of which is expressed as a regression form in the parameters of the nonlinear part or in the parameters of the linear part, and we propose a hierarchical least-squares algorithm.
- By applying the data filtering technique, the input–output data are filtered, and a filtering-based hierarchical least-squares algorithm is presented for Hammerstein finite impulse response moving average systems to improve the accuracy of parameter estimation.

Briefly, the structure of this paper is organized as follows. Section 2 describes a two-input two-output Hammerstein FIR–MA system. Section 3 derives a hierarchical least-squares algorithm for the TITO Hammerstein system. Section 4 derives a filtering-based hierarchical least-squares algorithm for the TITO Hammerstein system. Section 5 provides an illustrative example to show the effectiveness of the proposed algorithms. Finally, we offer some concluding remarks in Section 6.

2. System Description and Identification Model

Let us define some symbols. “ $A =: B$ ” or “ $B := A$ ” stands for “ A is defined as B ”. The superscript T denotes the matrix/vector transpose; the norm of the matrix X is defined by $\|X\|^2 := \text{tr}[XX^T]$. The symbol I_n stands for an identity matrix of size $n \times n$. $\mathbf{1}_n$ represents a n -dimensional column vector whose elements are 1.

Consider the following Hammerstein FIR–MA system, depicted in Figure 1,

$$\mathbf{y}(t) = \boldsymbol{\theta}(z)\bar{\mathbf{u}}(t) + \alpha(z)v(t), \tag{1}$$

where $\mathbf{u}(t) := [\bar{u}_1(t), \bar{u}_2(t)]^T \in \mathbb{R}^2$ is the system nonlinear input, $\mathbf{y}(t) \in \mathbb{R}^2$ is the system output, $v(t) \in \mathbb{R}^2$ is the stochastic white noise with zero mean, $\boldsymbol{\theta}(z) := [\boldsymbol{\theta}_1(z), \boldsymbol{\theta}_2(z)]$ and $\alpha(z)$ are two matrix polynomials in the unit backward shift operator z^{-1} [$z^{-1}\mathbf{y}(t) = \mathbf{y}(t - 1)$], and defined by

$$\boldsymbol{\theta}_i(z) := \boldsymbol{\theta}_{i,1}z^{-1} + \boldsymbol{\theta}_{i,2}z^{-2} + \dots + \boldsymbol{\theta}_{i,n}z^{-n} \in \mathbb{R}^2, \quad i = 1, 2, \tag{2}$$

$$\alpha(z) := 1 + \alpha_1z^{-1} + \alpha_2z^{-2} + \dots + \alpha_nz^{-n} \in \mathbb{R}. \tag{3}$$

The unknown nonlinear input $\bar{u}_i(t)$ is assumed to be a static nonlinear function of the known basis $d_{i,k}$ with $\gamma_{i,k}$ as its parameters:

$$\bar{u}_i(t) = d_i[\mathbf{u}_i(t)] = \sum_{k=1}^s \gamma_{i,k}d_{i,k}[\mathbf{u}_i(t)]. \tag{4}$$

Assume that the orders n and s are known and $\bar{\mathbf{u}}(t) = \mathbf{0}$, $\mathbf{y}(t) = \mathbf{0}$ and $v(t) = \mathbf{0}$ for $t \leq 0$.

The purpose of this paper is to estimate three parameters $\boldsymbol{\theta}_{i,n}$, $\gamma_{i,k}$ and α_i .

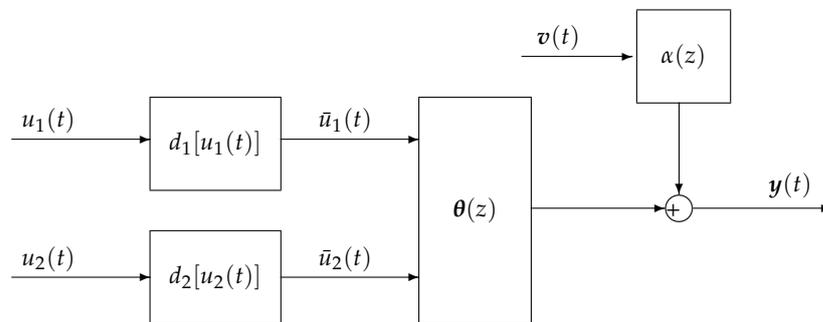


Figure 1. The two-input two-output Hammerstein FIR–MA system.

3. The Hierarchical Least-Squares Algorithm

From Equation (1), we have

$$\mathbf{y}(t) = \mathbf{A}_1(z)\bar{u}_1(t) + \mathbf{A}_2(z)\bar{u}_2(t) + \alpha(z)v(t). \tag{5}$$

The first two terms of the right hand side can be expressed by

$$\begin{aligned}
 \mathbf{A}_i(z)\bar{u}_i(t) &= [\mathbf{A}_{i,1}z^{-1} + \mathbf{A}_{i,2}z^{-2} + \dots + \mathbf{A}_{i,n}z^{-n}]\bar{u}_i(t) \\
 &= \mathbf{A}_{i,1}\bar{u}_i(t-1) + \mathbf{A}_{i,2}\bar{u}_i(t-2) + \dots + \mathbf{A}_{i,n}\bar{u}_i(t-n) \\
 &= [\mathbf{A}_{i,1}, \mathbf{A}_{i,2}, \dots, \mathbf{A}_{i,n}] \begin{bmatrix} \bar{u}_i(t-1) \\ \bar{u}_i(t-2) \\ \vdots \\ \bar{u}_i(t-n) \end{bmatrix} \\
 &= [\mathbf{A}_{i,1}, \mathbf{A}_{i,2}, \dots, \mathbf{A}_{i,n}] \begin{bmatrix} \sum_{k=1}^s \gamma_{i,k}d_{i,k}[u_i(t-1)] \\ \sum_{k=1}^s \gamma_{i,k}d_{i,k}[u_i(t-2)] \\ \vdots \\ \sum_{k=1}^s \gamma_{i,k}d_{i,k}[u_i(t-n)] \end{bmatrix}. \tag{6}
 \end{aligned}$$

Define

$$\gamma_i := \begin{bmatrix} \gamma_{i,1} \\ \gamma_{i,2} \\ \vdots \\ \gamma_{i,s} \end{bmatrix} \in \mathbb{R}^s, \quad \mathbf{d}_i(t) := \begin{bmatrix} d_{i,1}[u_i(t)] \\ d_{i,2}[u_i(t)] \\ \vdots \\ d_{i,s}[u_i(t)] \end{bmatrix}^T \in \mathbb{R}^{1 \times s}, \quad i = 1, 2.$$

Then Equation (5) can be rewritten as

$$\begin{aligned}
 \mathbf{y}(t) &= [\boldsymbol{\theta}_{1,1}, \boldsymbol{\theta}_{1,2}, \dots, \boldsymbol{\theta}_{1,n}] \begin{bmatrix} \mathbf{d}_1[u_1(t-1)]\gamma_1 \\ \mathbf{d}_1[u_1(t-2)]\gamma_1 \\ \vdots \\ \mathbf{d}_1[u_1(t-n)]\gamma_1 \end{bmatrix} \\
 &+ [\boldsymbol{\theta}_{2,1}, \boldsymbol{\theta}_{2,2}, \dots, \boldsymbol{\theta}_{2,n}] \begin{bmatrix} \mathbf{d}_2[u_2(t-1)]\gamma_2 \\ \mathbf{d}_2[u_2(t-2)]\gamma_2 \\ \vdots \\ \mathbf{d}_2[u_2(t-n)]\gamma_2 \end{bmatrix} \\
 &+ \alpha(z)\mathbf{v}(t). \tag{7}
 \end{aligned}$$

Define

$$\begin{aligned}
 \boldsymbol{\theta}_i &:= [\boldsymbol{\theta}_{i,1}, \boldsymbol{\theta}_{i,2}, \dots, \boldsymbol{\theta}_{i,n}] \in \mathbb{R}^{2 \times n}, \\
 \mathbf{D}_i(t) &:= \begin{bmatrix} \mathbf{d}_i[u_i(t-1)] \\ \mathbf{d}_i[u_i(t-2)] \\ \vdots \\ \mathbf{d}_i[u_i(t-n)] \end{bmatrix} \\
 &= \begin{bmatrix} d_{i,1}[u_i(t-1)] & d_{i,2}[u_i(t-1)] & \dots & d_{i,s}[u_i(t-1)] \\ d_{i,1}[u_i(t-2)] & d_{i,2}[u_i(t-2)] & \dots & d_{i,s}[u_i(t-2)] \\ \vdots & \vdots & \dots & \vdots \\ d_{i,1}[u_i(t-n)] & d_{i,2}[u_i(t-n)] & \dots & d_{i,s}[u_i(t-n)] \end{bmatrix} \in \mathbb{R}^{n \times s}.
 \end{aligned}$$

Then, Equation (7) can be rewritten in the following matrix form:

$$\begin{aligned}
 \mathbf{y}(t) &= [\boldsymbol{\theta}_{1,1}, \boldsymbol{\theta}_{1,2}, \dots, \boldsymbol{\theta}_{1,n}]\mathbf{D}_1(t)\gamma_1 + [\boldsymbol{\theta}_{2,1}, \boldsymbol{\theta}_{2,2}, \dots, \boldsymbol{\theta}_{2,n}]\mathbf{D}_2(t)\gamma_2 + \alpha(z)\mathbf{v}(t) \\
 &= \boldsymbol{\theta}_1\mathbf{D}_1(t)\gamma_1 + \boldsymbol{\theta}_2\mathbf{D}_2(t)\gamma_2 + \alpha(z)\mathbf{v}(t) \\
 &= \sum_{i=1}^2 \boldsymbol{\theta}_i\mathbf{D}_i(t)\gamma_i + \alpha(z)\mathbf{v}(t). \tag{8}
 \end{aligned}$$

Notice that in the identification model (8), it is difficult to simultaneously estimate the matrix θ_i and the vector γ_i . Here we use the hierarchical identification principle to transform the Hammerstein identification model into two different forms.

Define

$$\begin{aligned}
 D_\gamma(t) &:= \begin{bmatrix} D_1(t)\gamma_1 \\ D_2(t)\gamma_2 \end{bmatrix} \in \mathbb{R}^{2n}, \quad \theta := [\theta_1, \theta_2] \in \mathbb{R}^{2 \times 2n}, \\
 D_\theta(t) &:= [\theta_1 D_1(t), \theta_2 D_2(t)] \in \mathbb{R}^{2 \times 2s}, \quad \gamma := \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \in \mathbb{R}^{2s}, \\
 \alpha &:= [\alpha_1, \alpha_2, \dots, \alpha_n]^T \in \mathbb{R}^n, \quad \phi_s(t) := [v(t-1), v(t-2), \dots, v(t-n)]^T \in \mathbb{R}^{n \times 2}.
 \end{aligned}$$

The identification model in (8) can be transformed into the following two forms:

$$S_1: \quad y(t) = \theta D_\gamma(t) + \phi_s^T(t)\alpha + v(t), \tag{9}$$

$$S_2: \quad y(t) = D_\theta(t)\gamma + \phi_s^T(t)\alpha + v(t). \tag{10}$$

The parameter matrices θ , α and γ to be identified are included in these two new models. Then, define two cost functions about the parameter matrices θ , α and γ :

$$\begin{aligned}
 J_1(\theta, \alpha) &:= \sum_{l=1}^t [y(l) - \theta D_\gamma(l) - \phi_s^T(l)\alpha]^2, \\
 J_2(\gamma) &:= \sum_{l=1}^t [y(l) - D_\theta(l)\gamma - \phi_s^T(l)\alpha]^2.
 \end{aligned}$$

There are several difficulties in minimizing the two cost functions to obtain the parameter estimates. As the information matrix $D_\theta(t)$ contains the unknown parameter matrix θ_i , the information vector $D_\gamma(t)$ contains the unknown parameter vector γ_i , the information matrix $\phi_s(t)$ contains the unmeasurable noise term $v(t-i)$. The problem is solved by replacing the unknown parameter matrix θ_i with its estimate $\hat{\theta}_i(t)$, replacing the unknown parameter vector γ_i with its estimate $\hat{\gamma}_i(t)$, replacing the unmeasurable $v(t-i)$ with its estimate $\hat{v}(t-i)$.

The estimate of $\phi_s(t)$, $D_\gamma(t)$ and $D_\theta(t)$ can be written as

$$\begin{aligned}
 \hat{\phi}_s(t) &:= [\hat{v}(t-1), \hat{v}(t-2), \dots, \hat{v}(t-n)]^T. \\
 \hat{D}_\gamma(t) &:= \begin{bmatrix} D_1(t)\hat{\gamma}_1(t-1) \\ D_2(t)\hat{\gamma}_2(t-1) \end{bmatrix}, \\
 \hat{D}_\theta(t) &:= [\hat{\theta}_1(t)D_1(t), \hat{\theta}_2(t)D_2(t)].
 \end{aligned}$$

Then, the estimate $\hat{v}(t)$ can be computed by

$$\hat{v}(t) = y(t) - \hat{\theta}(t)\hat{D}_\gamma(t) - \hat{\phi}_s^T(t)\hat{\alpha}(t).$$

Then, we obtain the hierarchical least-squares (HLS) parameter estimation algorithm for the TITO Hammerstein FIR-MA models as follows:

$$\hat{\theta}^T(t) = \hat{\theta}^T(t-1) + L_\theta(t)[y(t) - \hat{\theta}(t-1)\hat{D}_\gamma(t) - \hat{\phi}_s^T(t)\hat{\alpha}(t-1)]^T, \tag{11}$$

$$L_\theta(t) = P_\theta(t-1)\hat{D}_\gamma(t)[1 + \hat{D}_\gamma^T(t)P_\theta(t-1)\hat{D}_\gamma(t)]^{-1}, \tag{12}$$

$$P_\theta(t) = [I_n - L_\theta(t)\hat{D}_\gamma^T(t)]P_\theta(t-1), \quad P_\theta(0) = p_0I_n, \tag{13}$$

$$\hat{\alpha}(t) = \hat{\alpha}(t-1) + L_\alpha(t)[y(t) - \hat{\theta}(t)\hat{D}_\gamma(t) - \hat{\phi}_s^T(t)\hat{\alpha}(t-1)], \tag{14}$$

$$L_\alpha(t) = P_\alpha(t-1)\hat{\phi}_s^T(t)[I_n + \hat{\phi}_s(t)P_\alpha(t-1)\hat{\phi}_s^T(t)]^{-1} \tag{15}$$

$$P_\alpha(t) = [I_n - L_\alpha(t)\hat{\phi}_s^T(t)]P_\alpha(t-1), \quad P_\alpha(0) = p_0I_n, \tag{16}$$

$$\hat{\gamma}(t) = \hat{\gamma}(t-1) + L_\gamma(t)[y(t) - \hat{D}_\theta(t)\hat{\gamma}(t-1) - \hat{\phi}_s^T(t)\hat{\alpha}(t-1)], \tag{17}$$

$$L_\gamma(t) = P_\gamma(t-1)\hat{D}_\theta^T(t)[I_s + \hat{D}_\theta(t)P_\gamma(t-1)\hat{D}_\theta^T(t)]^{-1}, \tag{18}$$

$$P_\gamma(t) = [I_s - L_\gamma(t)\hat{D}_\theta^T(t)]P_\gamma(t-1), \quad P_\gamma(0) = p_0I_s, \tag{19}$$

$$d_i(t) = [d_{i,1}[u_i(t)], d_{i,2}[u_i(t)], \dots, d_{i,s}[u_i(t)]] \in \mathbb{R}^{1 \times s}, \quad i = 1, 2, \tag{20}$$

$$D_i(t) = \begin{bmatrix} d_{i,1}[u_i(t-1)] & d_{i,2}[u_i(t-1)] & \dots & d_{i,s}[u_i(t-1)] \\ d_{i,1}[u_i(t-2)] & d_{i,2}[u_i(t-2)] & \dots & d_{i,s}[u_i(t-2)] \\ \vdots & \vdots & \ddots & \vdots \\ d_{i,1}[u_i(t-n)] & d_{i,2}[u_i(t-n)] & \dots & d_{i,s}[u_i(t-n)] \end{bmatrix}, \tag{21}$$

$$\hat{D}_\gamma(t) = \begin{bmatrix} D_1(t)\hat{\gamma}_1(t-1) \\ D_2(t)\hat{\gamma}_2(t-1) \end{bmatrix}, \tag{22}$$

$$\hat{D}_\theta(t) = [\hat{\theta}_1(t)D_1(t), \hat{\theta}_2(t)D_2(t)], \tag{23}$$

$$\hat{\phi}_s(t) = [\hat{v}(t-1), \hat{v}(t-2), \dots, \hat{v}(t-n)]^T, \tag{24}$$

$$\hat{v}(t) = y(t) - \hat{\theta}(t)\hat{D}_\gamma(t) - \hat{\phi}_s^T(t)\hat{\alpha}(t), \tag{25}$$

$$\hat{\theta}_i(t) = [\hat{\theta}_{1i}(t), \hat{\theta}_{2i}(t), \dots, \hat{\theta}_{ni}(t)], \tag{26}$$

$$\hat{\gamma}_i(t) = [\hat{\gamma}_{i,1}(t), \hat{\gamma}_{i,2}(t), \dots, \hat{\gamma}_{i,s}(t)]^T, \tag{27}$$

$$\hat{\theta}(t) = [\hat{\theta}_1(t), \hat{\theta}_2(t)], \tag{28}$$

$$\hat{\gamma}(t) = [\hat{\gamma}_1^T(t), \hat{\gamma}_2^T(t)]^T. \tag{29}$$

The steps of computing the parameter estimation matrices $\hat{\theta}(t)$, $\hat{\alpha}(t)$ and $\hat{\gamma}(t)$ by the HLS algorithm in (11)–(29) are summarised as follows.

1. Set the initial values: let $t = 1$, $P_\theta(0) = p_0I_n$, $P_\alpha(0) = p_0I_n$, $P_\gamma(0) = p_0I_s$, $\hat{\theta}(0) = \mathbf{1}_{2 \times 2n}/p_0$, $\hat{\alpha}(0) = \mathbf{1}_n/p_0$, $\hat{\gamma}(0) = \mathbf{1}_{2s}/p_0$, $p_0 = 10^6$; $u(t) = \mathbf{0}$ and $y(t) = \mathbf{0}$ for $t \leq 0$ and set a small positive number ϵ .
2. Collect the input and output data $u(t)$ and $y(t)$, and form $D_i(t)$ by (21), $\hat{\phi}_s(t)$ by (24).
3. Compute $\hat{D}_\gamma(t)$, $L_\theta(t)$ and $P_\theta(t)$ by (22), (12) and (13), and update the parameter estimate $\hat{\theta}(t)$ by (11).
4. Compute $\hat{D}_\theta(t)$, $L_\gamma(t)$ and $P_\gamma(t)$ by (23), (18) and (19), and update the parameter estimate $\hat{\gamma}(t)$ by (17).
5. Compute $L_\alpha(t)$ and $P_\alpha(t)$ by (15) and (16), and update the parameter estimate $\hat{\alpha}(t)$ by (14).
6. Compute the noise term $\hat{v}(t)$ using (25).
7. Compare $\hat{\theta}(t)$ with $\hat{\theta}(t-1)$ and compare $\hat{\gamma}(t)$ with $\hat{\gamma}(t-1)$: if $\|\hat{\theta}(t) - \hat{\theta}(t-1)\| < \epsilon$ and $\|\hat{\gamma}(t) - \hat{\gamma}(t-1)\| < \epsilon$, terminate recursive calculation procedure and obtain $\hat{\theta}(t)$, $\hat{\alpha}(t)$ and $\hat{\gamma}(t)$; otherwise, increase t by 1 and go to step 2.

The flowchart of computing the parameter estimates $\hat{\theta}(t)$, $\hat{\alpha}(t)$ and $\hat{\gamma}(t)$ in the HLS algorithm is shown in Figure 2.

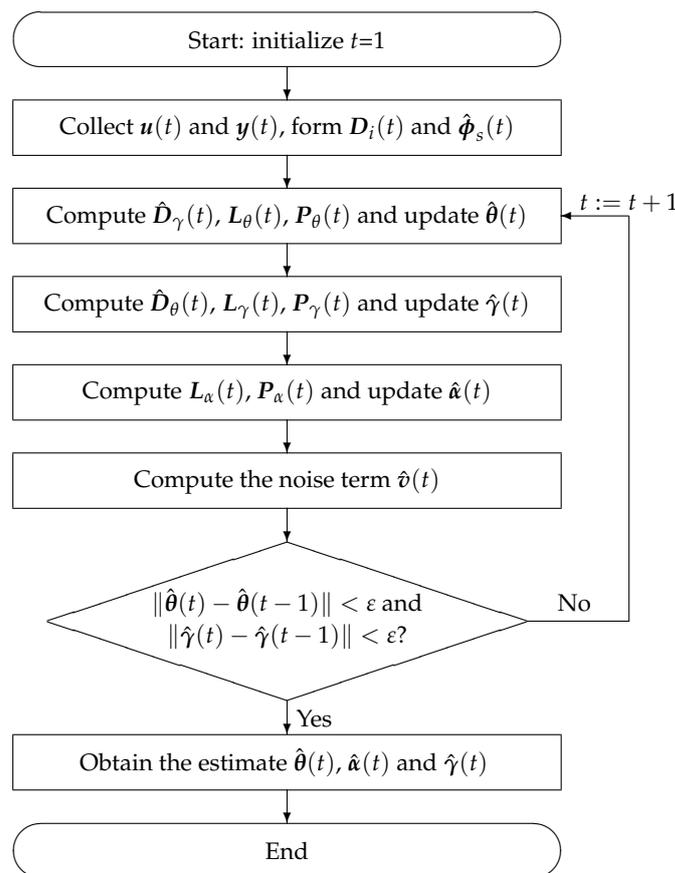


Figure 2. The flowchart of the HLS algorithm.

4. The Convergence Analysis of the Hierarchical Least-Squares Algorithm

The convergence analysis of the proposed hierarchical identification principle-based least-squares algorithm for the TITO Hammerstein system is illustrated as follows. Assume the estimated information vectors $\hat{D}_\gamma(t)$, $\hat{D}_\theta(t)$, and $\hat{\phi}_s(t)$ are persistently exciting, i.e., there exist constants $M_i > 0$ and an integer N such that for $t > N$, the following strong persistent excitation conditions hold:

$$\begin{aligned}
 (A1) \quad \mathbf{M}_1 \mathbf{I} &\leq \frac{1}{N} \sum_{j=0}^{N-1} \hat{D}_\gamma(t) \hat{D}_\gamma^T(t) \leq \mathbf{M}_2 \mathbf{I}, \quad a.s., \\
 (A2) \quad \mathbf{M}_3 \mathbf{I} &\leq \frac{1}{N} \sum_{j=0}^{N-1} \hat{D}_\theta(t) \hat{D}_\theta^T(t) \leq \mathbf{M}_4 \mathbf{I}, \quad a.s., \\
 (A3) \quad \mathbf{M}_5 \mathbf{I} &\leq \frac{1}{N} \sum_{j=0}^{N-1} \hat{\phi}_s(t) \hat{\phi}_s^T(t) \leq \mathbf{M}_6 \mathbf{I}, \quad a.s.
 \end{aligned}
 \tag{30}$$

Theorem 1. For the the hierarchical least-squares algorithm in (11)–(29), suppose that (A1)–(A3) hold. Additionally, assume that the σ algebra sequence $F_t = \sigma(v(t), v(t-1), v(t-2), \dots)$, which is generated by $v(t)$, and $\{v(t), F_t\}$ is a martingale difference sequence on a probability space (Ω, F, P) [24]. The sequence $v(t)$ satisfies

$$\begin{aligned}
 (B1) \quad &E[v(t)|F_{t-1}] = 0, a.s., \\
 (B2) \quad &E[|v(t)|^2|F_{t-1}] = \sigma^2(t) \leq \sigma^2 < \infty, a.s., \\
 (B3) \quad &\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t v^2(i) \leq \sigma^2 < \infty, a.s.
 \end{aligned}
 \tag{31}$$

Then, the parameter estimation vector $\hat{\theta}_i(t)$, $\hat{\gamma}(t)$, and $\hat{\theta}(t)$ consistently converges to the true parameter vector θ_i , γ and θ .

5. The Filtering Based Recursive Least-Squares Algorithm

This section proposes the filtering-based recursive least-squares algorithm for the two-input two-output system with MA noise by using the data filtering technique. The input–output data are filtered through a rational polynomial $\alpha(z)$. Multiplying both sides of Equation (8) by $\frac{1}{\alpha(z)}$ yields

$$\frac{1}{\alpha(z)}\mathbf{y}(t) = \frac{\theta_1 D_1(t)\gamma_1}{\alpha(z)} + \frac{\theta_2 D_2(t)\gamma_2}{\alpha(z)} + \mathbf{v}(t). \tag{32}$$

Define the filtered input and the filtered output as

$$\mathbf{y}_f(t) := \frac{1}{\alpha(z)}\mathbf{y}(t) = \mathbf{y}(t) + [1 - \alpha(z)]\mathbf{y}_f(t), \tag{33}$$

$$D_{f,i}(t) := \frac{1}{\alpha(z)}D_i(t) = D_i(t) + [1 - \alpha(z)]D_{f,i}(t). \tag{34}$$

Then, Equation (32) can be rewritten as

$$\mathbf{y}_f(t) = \theta_1 D_{f,1}(t)\gamma_1 + \theta_2 D_{f,2}(t)\gamma_2 + \mathbf{v}(t). \tag{35}$$

Substituted (33) into (35), we can have

$$\mathbf{y}(t) = \theta_1 D_{f,1}(t)\gamma_1 + \theta_2 D_{f,2}(t)\gamma_2 + [\alpha(z) - 1]\mathbf{y}_f(t) + \mathbf{v}(t). \tag{36}$$

Define

$$D_{f,\gamma}(t) := \begin{bmatrix} D_{f,1}(t)\gamma_1 \\ D_{f,2}(t)\gamma_2 \end{bmatrix} \in \mathbb{R}^{2n}, \quad D_{f,\theta}(t) := [\theta_1 D_{f,1}(t), \theta_2 D_{f,2}(t)] \in \mathbb{R}^{2 \times 2s},$$

$$\phi_f(t) := [\mathbf{y}_f(t-1), \mathbf{y}_f(t-2), \dots, \mathbf{y}_f(t-n)]^T \in \mathbb{R}^{n \times 2}.$$

Then, Equation (36) can be transformed into the following two forms:

$$S_{f,1} : \mathbf{y}(t) = \theta D_{f,\gamma}(t) + \phi_f^T(t)\alpha + \mathbf{v}(t), \tag{37}$$

$$S_{f,2} : \mathbf{y}(t) = D_{f,\theta}(t)\gamma + \phi_f^T(t)\alpha + \mathbf{v}(t). \tag{38}$$

The parameter matrices θ , α and γ to be identified are included in these two new models. Then, define two cost functions about the parameter matrices θ , α and γ :

$$J_3(\theta, \alpha) := \sum_{l=1}^t [\mathbf{y}(l) - \theta D_{f,\gamma}(l) - \phi_f^T(l)\alpha]^2,$$

$$J_4(\gamma) := \sum_{l=1}^t [\mathbf{y}(l) - D_{f,\theta}(l)\gamma - \phi_f^T(l)\alpha]^2.$$

There are several difficulties in minimizing the two cost functions to obtain the parameter estimates. Since the information matrix $D_{f,\gamma}(t)$ contains the unknown parameter matrix θ_i and $D_{f,i}(t)$, the information vector $D_{f,\theta}(t)$ contains the unknown parameter vector γ_i and $D_{f,i}(t)$, the information matrix $\phi_f(t)$ contains the unmeasurable noise term $\mathbf{y}_f(t-i)$. The problem is solved by replacing the unknown parameter matrix θ_i with its estimate $\hat{\theta}_i(t)$, replacing the unknown parameter vector γ_i with its estimate $\hat{\gamma}_i(t)$, replacing the unmeasurable $D_{f,i}(t)$ and $\mathbf{y}_f(t-i)$ with their estimate $\hat{D}_{f,i}(t)$ and $\hat{\mathbf{y}}_f(t-i)$.

The estimate of $D_{f,i}(t)$ can be computed by

$$\hat{D}_{f,i}(t) := D_i(t) - \sum_{i=1}^2 \alpha_i \hat{D}_{f,i}(t-i).$$

Then, the estimate of $\phi_f(t)$, $D_{f,\gamma}(t)$ and $D_{f,\theta}(t)$ can be written as

$$\begin{aligned} \hat{\phi}_f(t) &:= [\hat{y}_f(t-1), \hat{y}_f(t-2), \dots, \hat{y}_f(t-n)]^T, \\ \hat{D}_{f,\gamma}(t) &:= \begin{bmatrix} \hat{D}_{f,1}(t)\hat{\gamma}_1(t-1) \\ \hat{D}_{f,2}(t)\hat{\gamma}_2(t-1) \end{bmatrix}, \\ \hat{D}_{f,\theta}(t) &:= [\hat{\theta}_1(t)\hat{D}_{f,1}(t), \hat{\theta}_2(t)\hat{D}_{f,2}(t)]. \end{aligned}$$

Then, the estimate $\hat{y}_f(t)$ can be computed by

$$\hat{y}_f(t) = y(t) - \hat{\phi}_f^T(t)\hat{\alpha}(t).$$

Then, we obtain the filtering-based hierarchical least-squares (F-HLS) parameter estimation algorithm for the TITO Hammerstein FIR-MA models as follows:

$$\hat{\theta}^T(t) = \hat{\theta}^T(t-1) + L_\theta(t)[y(t) - \hat{\theta}(t-1)\hat{D}_{f,\gamma}(t) - \hat{\phi}_f^T(t)\hat{\alpha}(t-1)]^T, \tag{39}$$

$$L_\theta(t) = P_\theta(t-1)\hat{D}_{f,\gamma}(t)[1 + \hat{D}_{f,\gamma}^T(t)P_\theta(t-1)\hat{D}_{f,\gamma}(t)]^{-1}, \tag{40}$$

$$P_\theta(t) = [I_n - L_\theta(t)\hat{D}_{f,\gamma}^T(t)]P_\theta(t-1), \quad P_\theta(0) = p_0I_n, \tag{41}$$

$$\hat{\alpha}(t) = \hat{\alpha}(t-1) + L_\alpha(t)[y(t) - \hat{\theta}(t)\hat{D}_{f,\gamma}(t) - \hat{\phi}_f^T(t)\hat{\alpha}(t-1)], \tag{42}$$

$$L_\alpha(t) = P_\alpha(t-1)\hat{\phi}_s^T(t)[I_n + \hat{\phi}_s(t)P_\alpha(t-1)\hat{\phi}_s^T(t)]^{-1} \tag{43}$$

$$P_\alpha(t) = [I_n - L_\alpha(t)\hat{\phi}_s^T(t)]P_\alpha(t-1), \quad P_\alpha(0) = p_0I_n, \tag{44}$$

$$\hat{\gamma}(t) = \hat{\gamma}(t-1) + L_\gamma(t)[y(t) - \hat{D}_{f,\theta}(t)\hat{\gamma}(t-1) - \hat{\phi}_s^T(t)\hat{\alpha}(t-1)], \tag{45}$$

$$L_\gamma(t) = P_\gamma(t-1)\hat{D}_{f,\theta}^T(t)[I_s + \hat{D}_{f,\theta}(t)P_\gamma(t-1)\hat{D}_{f,\theta}^T(t)]^{-1}, \tag{46}$$

$$P_\gamma(t) = [I_s - L_\gamma(t)\hat{D}_{f,\theta}^T(t)]P_\gamma(t-1), \quad P_\gamma(0) = p_0I_s, \tag{47}$$

$$d_i(t) = [d_{i,1}[u_i(t)], d_{i,2}[u_i(t)], \dots, d_{i,s}[u_i(t)]] \in \mathbb{R}^{1 \times s}, \quad i = 1, 2, \tag{48}$$

$$D_i(t) = \begin{bmatrix} d_{i,1}[u_i(t-1)] & d_{i,2}[u_i(t-1)] & \dots & d_{i,s}[u_i(t-1)] \\ d_{i,1}[u_i(t-2)] & d_{i,2}[u_i(t-2)] & \dots & d_{i,s}[u_i(t-2)] \\ \vdots & \vdots & \dots & \vdots \\ d_{i,1}[u_i(t-n)] & d_{i,2}[u_i(t-n)] & \dots & d_{i,s}[u_i(t-n)] \end{bmatrix}, \tag{49}$$

$$\hat{D}_{f,i}(t) := D_i(t) - \sum_{i=1}^2 \alpha_i \hat{D}_{f,i}(t-i), \tag{50}$$

$$\hat{D}_{f,\gamma}(t) := \begin{bmatrix} \hat{D}_{f,1}(t)\hat{\gamma}_1(t-1) \\ \hat{D}_{f,2}(t)\hat{\gamma}_2(t-1) \end{bmatrix}, \tag{51}$$

$$\hat{D}_{f,\theta}(t) := [\hat{\theta}_1(t)\hat{D}_{f,1}(t), \hat{\theta}_2(t)\hat{D}_{f,2}(t)]. \tag{52}$$

$$\hat{\phi}_f(t) := [\hat{y}_f(t-1), \hat{y}_f(t-2), \dots, \hat{y}_f(t-n)]^T, \tag{53}$$

$$\hat{y}_f(t) = y(t) - \hat{\phi}_f^T(t)\hat{\alpha}(t), \tag{54}$$

$$\hat{\theta}_i(t) = [\hat{\theta}_{1i}(t), \hat{\theta}_{2i}(t), \dots, \hat{\theta}_{ni}(t)], \tag{55}$$

$$\hat{\gamma}_i(t) = [\hat{\gamma}_{i,1}(t), \hat{\gamma}_{i,2}(t), \dots, \hat{\gamma}_{i,s}(t)]^T, \tag{56}$$

$$\hat{\theta}(t) = [\hat{\theta}_1(t), \hat{\theta}_2(t)], \tag{57}$$

$$\hat{\gamma}(t) = [\hat{\gamma}_1^T(t), \hat{\gamma}_2^T(t)]^T. \tag{58}$$

The steps of computing the parameter estimation matrices $\hat{\theta}(t)$, $\hat{\alpha}(t)$ and $\hat{\gamma}(t)$ by the F-HLS algorithm in (39)–(58) are summarised as follows.

1. Set the initial values: let $t = 1$, $P_{\theta}(0) = p_0 I_n$, $P_{\alpha}(0) = p_0 I_n$, $P_{\gamma}(0) = p_0 I_s$, $\hat{\theta}(0) = \mathbf{1}_{2 \times 2n} / p_0$, $\hat{\alpha}(0) = \mathbf{1}_n / p_0$, $\hat{\gamma}(0) = \mathbf{1}_{2s} / p_0$, $\hat{y}_f(t - i) = \mathbf{1}_n / p_0$, $p_0 = 10^6$; $\mathbf{u}(t) = \mathbf{0}$ and $\mathbf{y}(t) = \mathbf{0}$ for $t \leq 0$ and set a small positive number ε .
2. Collect the input and output data $\mathbf{u}(t)$ and $\mathbf{y}(t)$, form $D_i(t)$ by (49), $\hat{\phi}_f(t)$ by (53), and compute $\hat{D}_{f,i}(t)$ by (50).
3. Compute $\hat{D}_{f,\gamma}(t)$, $L_{\theta}(t)$ and $P_{\theta}(t)$ by (51), (40) and (41), and update the parameter estimate $\hat{\theta}(t)$ by (39).
4. Compute $\hat{D}_{f,\beta}(t)$, $L_{\gamma}(t)$ and $P_{\gamma}(t)$ by (52), (46) and (47), and update the parameter estimate $\hat{\gamma}(t)$ by (45).
5. Compute $L_{\alpha}(t)$ and $P_{\alpha}(t)$ by (43) and (44), and update the parameter estimate $\hat{\alpha}(t)$ by (42).
6. Compute the noise term $\hat{v}(t)$ using (54).
7. Compare $\hat{\theta}(t)$ with $\hat{\theta}(t - 1)$ and compare $\hat{\gamma}(t)$ with $\hat{\gamma}(t - 1)$: if $\|\hat{\theta}(t) - \hat{\theta}(t - 1)\| < \varepsilon$ and $\|\hat{\gamma}(t) - \hat{\gamma}(t - 1)\| < \varepsilon$, terminate recursive calculation procedure and obtain $\hat{\theta}(t)$, $\hat{\alpha}(t)$ and $\hat{\gamma}(t)$; otherwise, increase t by 1 and go to step 2.

6. Examples

Consider the following TITO Hammerstein FIR-MA system:

$$\begin{aligned} \mathbf{y}(t) &= \boldsymbol{\theta}(z)\bar{\mathbf{u}}(t) + \alpha(z)v(t), \\ \bar{\mathbf{u}}(t) &= \begin{bmatrix} \bar{u}_1(t) \\ \bar{u}_2(t) \end{bmatrix} = \begin{bmatrix} -0.35u_1(t) - 0.40u_1^2(t) \\ 0.14u_2(t) - 0.34u_2^2(t) \end{bmatrix}, \\ \mathbf{y}(t) &= \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}, \quad \mathbf{v}(t) = \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix}, \\ \alpha(z) &= 1 + 1.00z^{-1} + 0.11z^{-2}, \\ \boldsymbol{\theta}_1(z) &= \begin{bmatrix} -0.28 \\ 0.06 \end{bmatrix} z^{-1} + \begin{bmatrix} -0.34 \\ -0.27 \end{bmatrix} z^{-2}, \\ \boldsymbol{\theta}_2(z) &= \begin{bmatrix} 1.10 \\ 1.40 \end{bmatrix} z^{-1} + \begin{bmatrix} 0.41 \\ 0.54 \end{bmatrix} z^{-2}, \\ \boldsymbol{\theta}_{1,1} &= \begin{bmatrix} -0.28 \\ 0.06 \end{bmatrix}, \quad \boldsymbol{\theta}_{1,2} = \begin{bmatrix} -0.34 \\ -0.27 \end{bmatrix}, \quad \boldsymbol{\theta}_{2,1} = \begin{bmatrix} 1.10 \\ 1.40 \end{bmatrix}, \quad \boldsymbol{\theta}_{2,2} = \begin{bmatrix} 0.41 \\ 0.54 \end{bmatrix}, \\ \boldsymbol{\gamma}_1 &= \begin{bmatrix} -0.35 \\ -0.40 \end{bmatrix}, \quad \boldsymbol{\gamma}_2 = \begin{bmatrix} 0.14 \\ -0.34 \end{bmatrix}, \quad \boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} 1.00 \\ 0.11 \end{bmatrix}. \end{aligned}$$

In the simulation, the input $\bar{\mathbf{u}}(t)$ was taken as an uncorrelated persistent excitation signal vector sequence with zero mean and unit variance, $\mathbf{v}(t)$ was taken as a white noise sequence with zero mean and variance $\sigma^2 = 1.00^2$ and we applied the HLS algorithm and the F-HLS algorithm to identify this example system. The parameter estimates and the errors are shown in Tables 1–4. The parameter estimation errors: $\delta := \sqrt{\frac{\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|^2 + \|\hat{\boldsymbol{\gamma}}(t) - \boldsymbol{\gamma}\|^2}{\|\hat{\boldsymbol{\theta}}\|^2 + \|\hat{\boldsymbol{\gamma}}\|^2}}$ versus t are shown in Figures 3 and 4.

From Tables 1–4 and Figures 3–8, we can draw the following conclusions.

- The estimation errors given by the HLS algorithm and the F-HLS algorithm become generally smaller and smaller as t increases.
- Compared with the HLS algorithm, the F-HLS algorithm has higher parameter estimation accuracy.
- The predictions of the residuals are close to the true residuals, and the estimated outputs are close to the true outputs.

Table 1. The HLS estimates and their errors with $\sigma^2 = 1.00^2$.

<i>t</i>	100	200	500	1000	2000	3000	True Values
$\theta_{1,11}$	-0.50914	-0.30750	-0.19701	-0.22835	-0.29013	-0.29596	-0.28000
$\theta_{1,12}$	0.97766	0.56617	0.51217	0.32163	0.14490	0.13167	0.06000
$\theta_{1,21}$	0.11384	0.21689	0.03764	-0.11184	-0.19503	-0.25967	-0.34000
$\theta_{1,22}$	-0.80709	-0.63473	-0.25479	-0.17204	-0.24346	-0.25086	-0.27000
$\theta_{2,11}$	0.74403	0.81074	0.74457	0.81258	0.92845	0.91909	1.10000
$\theta_{2,12}$	0.44263	1.10990	0.95009	1.12074	1.14401	1.21750	1.40000
$\theta_{2,21}$	-0.12767	0.00039	0.27178	0.21795	0.28146	0.33896	0.41000
$\theta_{2,22}$	0.28715	0.69936	0.51335	0.40011	0.43399	0.42026	0.54000
$\gamma_{1,1}$	-0.52470	-0.44508	-0.39763	-0.37525	-0.37048	-0.39093	-0.35000
$\gamma_{1,2}$	-0.20855	-0.21843	-0.24514	-0.25373	-0.27027	-0.27842	-0.40000
$\gamma_{2,1}$	0.28346	0.34036	0.27891	0.26296	0.23975	0.23523	0.14000
$\gamma_{2,2}$	-0.29712	-0.34020	-0.32389	-0.33012	-0.33797	-0.34912	-0.34000
α_1	0.14994	0.26180	0.45948	0.68975	0.90551	1.01653	1.00000
α_2	0.50161	0.40490	0.23834	0.11798	0.06374	0.06636	0.11000
δ (%)	83.88885	57.71077	44.53619	30.11516	18.88155	15.33941	

Table 2. The HLS estimates and their errors with $\sigma^2 = 0.50^2$.

<i>t</i>	100	200	500	1000	2000	3000	True Values
$\theta_{1,11}$	-0.44432	-0.40006	-0.38992	-0.31198	-0.26693	-0.26510	-0.28000
$\theta_{1,12}$	0.56573	-0.09028	0.00964	-0.06114	-0.06901	0.00239	0.06000
$\theta_{1,21}$	-0.12045	-0.04354	-0.31996	-0.30311	-0.33949	-0.33982	-0.34000
$\theta_{1,22}$	-0.22202	-0.11342	0.01576	-0.12291	-0.19827	-0.23634	-0.27000
$\theta_{2,11}$	1.04323	1.00861	1.21022	1.08627	1.01177	1.03408	1.10000
$\theta_{2,12}$	0.80616	1.08325	1.03559	1.20325	1.32250	1.35832	1.40000
$\theta_{2,21}$	0.41692	0.42105	0.36272	0.43777	0.45083	0.41805	0.41000
$\theta_{2,22}$	1.13348	0.87373	0.73620	0.53233	0.49740	0.50385	0.54000
$\gamma_{1,1}$	-0.37277	-0.35597	-0.40370	-0.38316	-0.40989	-0.42941	-0.35000
$\gamma_{1,2}$	-0.41897	-0.38966	-0.42279	-0.39275	-0.41758	-0.43019	-0.40000
$\gamma_{2,1}$	-0.05560	0.01713	0.03196	0.08860	0.11363	0.12266	0.14000
$\gamma_{2,2}$	-0.40705	-0.38088	-0.38310	-0.35594	-0.35451	-0.36378	-0.34000
α_1	0.72392	0.83949	1.00022	1.05937	1.10617	1.14589	1.00000
α_2	-0.20897	-0.13856	-0.02692	0.09362	0.15665	0.18479	0.11000
δ (%)	48.67846	29.82225	24.42703	12.72200	10.38521	9.39920	

Table 3. The F-HLS estimates and their errors with $\sigma^2 = 1.00^2$.

<i>t</i>	100	200	500	1000	2000	3000	True Values
$\theta_{1,11}$	-0.46534	-0.47768	-0.40413	-0.32596	-0.27511	-0.27812	-0.28000
$\theta_{1,12}$	0.54645	0.51758	0.32302	0.22073	0.12924	0.10861	0.06000
$\theta_{1,21}$	0.34879	0.18063	-0.05504	-0.22830	-0.36081	-0.33090	-0.34000
$\theta_{1,22}$	-1.13416	-1.18888	-0.91111	-0.53083	-0.36501	-0.39249	-0.27000
$\theta_{2,11}$	1.37312	1.07871	0.98702	1.02387	1.06733	1.06065	1.10000
$\theta_{2,12}$	1.20017	1.40717	1.39712	1.26345	1.32548	1.35968	1.40000
$\theta_{2,21}$	-0.50420	-0.17069	0.29524	0.40066	0.39513	0.47011	0.41000
$\theta_{2,22}$	0.46208	0.26937	0.49625	0.60009	0.72538	0.68512	0.54000
$\gamma_{1,1}$	-0.28981	-0.29722	-0.34828	-0.35355	-0.38810	-0.38610	-0.35000
$\gamma_{1,2}$	-0.46484	-0.45777	-0.42673	-0.41237	-0.41568	-0.42091	-0.40000
$\gamma_{2,1}$	0.19233	0.17823	0.17974	0.14425	0.14231	0.12582	0.14000
$\gamma_{2,2}$	-0.18200	-0.18516	-0.18579	-0.19230	-0.22195	-0.23310	-0.34000
α_1	0.47710	0.58835	0.74277	0.88623	0.97788	1.02303	1.00000
α_2	-0.17863	-0.14081	-0.14153	-0.08345	0.01088	0.05732	0.11000
δ (%)	73.00597	61.84658	37.76548	19.80014	12.36094	10.78310	

Table 4. The F-HLS estimates and their errors with $\sigma^2 = 0.50^2$.

t	100	200	500	1000	2000	3000	True Values
$\theta_{1,11}$	-0.37740	-0.41815	-0.37182	-0.31015	-0.25496	-0.26324	-0.28000
$\theta_{1,12}$	0.51706	0.52888	0.34646	0.24585	0.14247	0.12024	0.06000
$\theta_{1,21}$	0.21828	0.05468	-0.08284	-0.18580	-0.31081	-0.27801	-0.34000
$\theta_{1,22}$	-0.88659	-0.96682	-0.73059	-0.42252	-0.26874	-0.30345	-0.27000
$\theta_{2,11}$	1.50800	1.15388	1.02343	1.05184	1.08214	1.07516	1.10000
$\theta_{2,12}$	1.27160	1.47267	1.42432	1.25340	1.32645	1.36304	1.40000
$\theta_{2,21}$	-0.47274	-0.14330	0.25538	0.31426	0.32086	0.38596	0.41000
$\theta_{2,22}$	0.34738	0.16458	0.40585	0.52964	0.61534	0.57783	0.54000
$\gamma_{1,1}$	-0.34054	-0.33682	-0.37719	-0.37203	-0.40047	-0.39525	-0.35000
$\gamma_{1,2}$	-0.45282	-0.44870	-0.41742	-0.40349	-0.40820	-0.41487	-0.40000
$\gamma_{2,1}$	0.21347	0.20244	0.21104	0.17728	0.17622	0.15573	0.14000
$\gamma_{2,2}$	-0.22228	-0.22713	-0.22743	-0.22955	-0.26159	-0.27247	-0.34000
α_1	0.52496	0.64401	0.80149	0.91977	1.00864	1.05000	1.00000
α_2	-0.13503	-0.10317	-0.11367	-0.03595	0.05006	0.09420	0.11000
δ (%)	64.49938	53.52304	31.39408	17.18469	8.81068	6.52457	

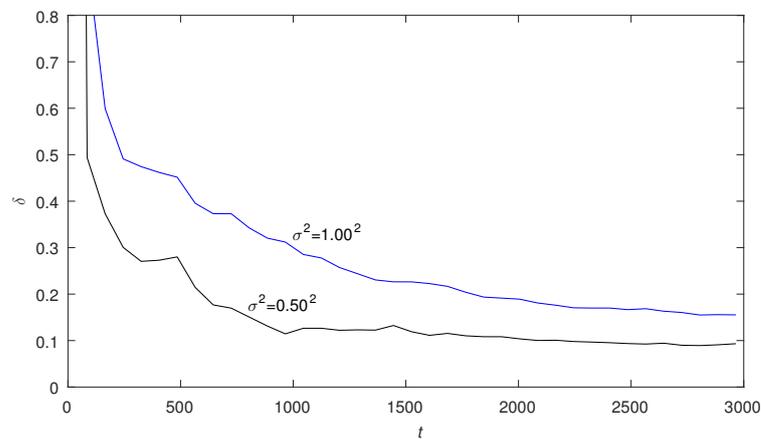


Figure 3. The HLS estimation errors δ versus t with $\sigma^2 = 0.50^2$ and $\sigma^2 = 1.00^2$.

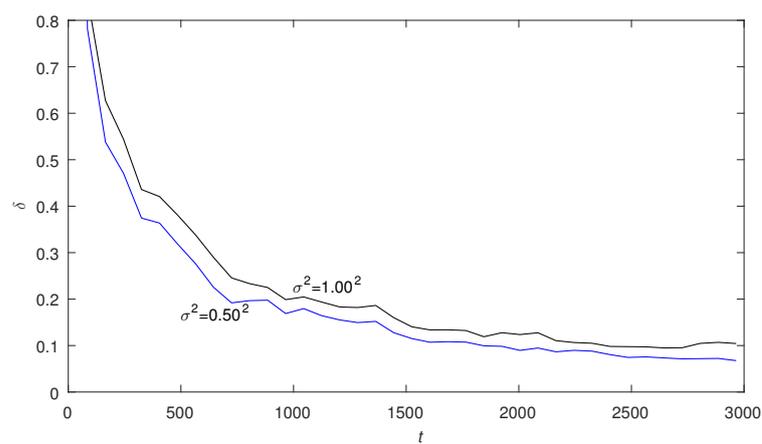


Figure 4. The F-HLS estimation errors δ versus t with $\sigma^2 = 0.50^2$ and $\sigma^2 = 1.00^2$.

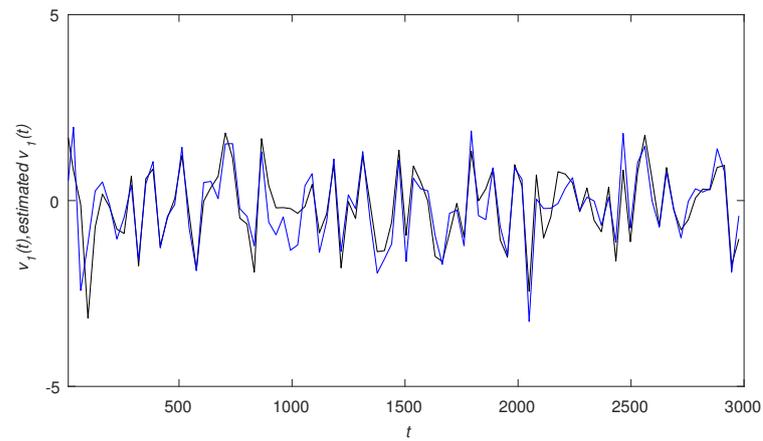


Figure 5. The F-HLS residual $v_1(t)$ (black line)/ $\hat{v}_1(t)$ (blue line) versus t under $\sigma^2 = 1.00^2$.

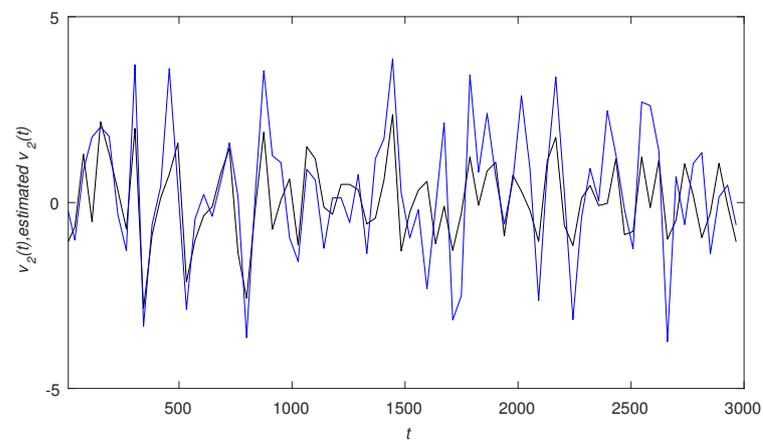


Figure 6. The F-HLS residual $v_2(t)$ (black line)/ $\hat{v}_2(t)$ (blue line) versus t under $\sigma^2 = 1.00^2$.

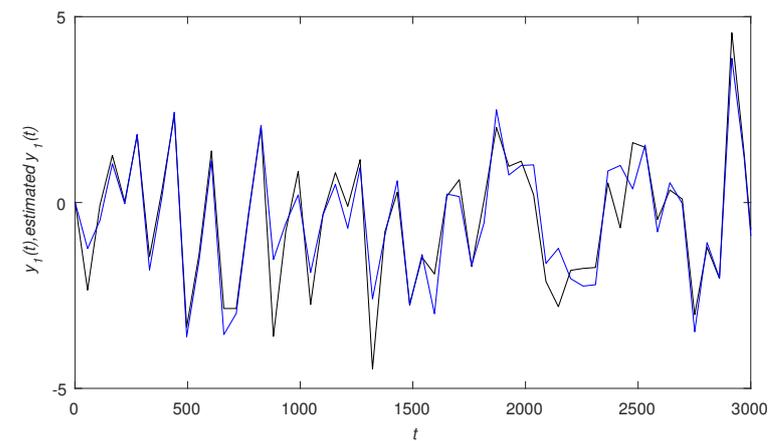


Figure 7. The F-HLS residual $y_1(t)$ (black line)/ $\hat{y}_1(t)$ (blue line) versus t under $\sigma^2 = 1.00^2$.

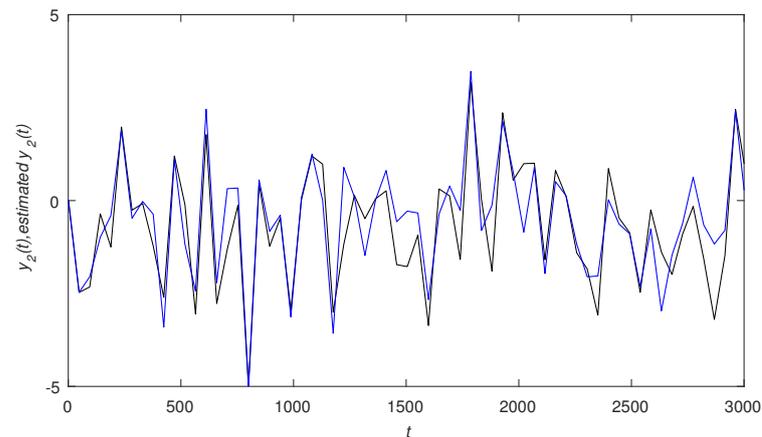


Figure 8. The F-HLS residual $y_2(t)$ (black line)/ $\hat{y}_2(t)$ (blue line) versus t under $\sigma^2 = 1.00^2$.

7. Conclusions

We derived a hierarchical least-squares algorithm and a filtering-based hierarchical least-squares algorithm for the two-input two-output Hammerstein finite impulse response moving average systems based on the hierarchical identification principle and the data filtering theory. The proposed F-HLS algorithm can effectively identify the parameters of the TITO Hammerstein FIR–MA systems and has a higher parameter estimation accuracy compared with the HLS algorithm. The proposed model parameter estimation methods in the paper can combine some mathematical strategies and other estimation algorithms to study the parameter identification problems of linear and nonlinear systems with different disturbances, and can be applied to other fields, such as signal processing and engineering application systems.

Author Contributions: Conceptualization, methodology and software, Y.J. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Shandong Province (ZR2020MF160 and SDYKC20090).

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.

Acknowledgments: This work was supported by the Natural Science Foundation of Shandong Province (ZR2020MF160 and SDYKC20090).

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

References

- Pillonetto, G.; Chiuso, A.; Nicolao, G. Stable spline identification of linear systems under missing data. *Automatica* **2019**, *108*, 108493. [CrossRef]
- Li, H.Y.; Zhang, K. Accurate and fast parameter identification of conditionally Gaussian Markov jump linear system with input control. *Automatica* **2022**, *137*, 109928. [CrossRef]
- Ding, F.; Zhang, X.; Xu, L. The innovation algorithms for multivariable state-space models. *Int. J. Adapt. Control Signal Process.* **2019**, *33*, 1601–1608. [CrossRef]
- Li, X.D.; Li, P. Stability of time-delay systems with impulsive control involving stabilizing delays. *Automatica* **2021**, *124*, 109336. [CrossRef]
- Peng, H.X.; He, W.J.; Zhang, Y.L.; Li, X.; Ding, Y.; Menon, V.G.; Verma, S. Covert non-orthogonal multiple access communication assisted by multi-antenna jamming Author links open overlay. *Phys. Commun.* **2022**, *52*, 101598. [CrossRef]
- Ding, F.; Liu, G.; Liu, X.P. Parameter estimation with scarce measurements. *Automatica* **2011**, *47*, 1646–1655. [CrossRef]

7. Shu, J.; He, J.C.; Li, L. MSIS: Multispectral instance segmentation method for power equipment. *Comput. Intell. Neurosci.* **2022**, *2022*, 2864717. [[CrossRef](#)] [[PubMed](#)]
8. You, J.Y.; Liu, Y.J. Iterative identification for multivariable systems with time-delays based on basis pursuit de-noising and auxiliary model. *Algorithms* **2018**, *11*, 180. [[CrossRef](#)]
9. 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](#)]
10. Lenka, B.K. Fractional comparison method and asymptotic stability results for multivariable fractional order systems. *Commun. Nonlinear Sci. Numer. Simul.* **2019**, *69*, 398–415. [[CrossRef](#)]
11. Liu, Q.Y.; Ding, F.; Yang, E.F. Parameter estimation algorithm for multivariable controlled autoregressive autoregressive moving average systems. *Digit. Signal Process.* **2018**, *83*, 323–331. [[CrossRef](#)]
12. Pan, J.; Ma, H.; Liu, Q.Y. Recursive coupled projection algorithms for multivariable output-error-like systems with coloured noises. *IET Signal Process* **2020**, *14*, 455–466. [[CrossRef](#)]
13. Xu, L.; Zhu, Q.M. Decomposition strategy-based hierarchical least mean square algorithm for control systems from the impulse responses. *Int. J. Syst. Sci.* **2021**, *52*, 1806–1821. [[CrossRef](#)]
14. Ding, F.; Chen, T. Parameter estimation of dual-rate stochastic systems by using an output error method. *IEEE Trans. Autom. Control* **2005**, *50*, 1436–1441. [[CrossRef](#)]
15. Zhang, X.; Yang, E.F. State estimation for bilinear systems through minimizing the covariance matrix of the state estimation errors. *Int. J. Adapt. Control. Signal Process.* **2019**, *33*, 1157–1173. [[CrossRef](#)]
16. Zhang, X. Hierarchical parameter and state estimation for bilinear systems. *Int. J. Syst. Sci.* **2020**, *51*, 275–290. [[CrossRef](#)]
17. Ding, F.; Lv, L.; Pan, J.; Wan, X.; Jin, X.B. Two-stage gradient-based iterative estimation methods for controlled autoregressive systems using the measurement data. *Int. J. Control Autom. Syst.* **2020**, *18*, 886–896. [[CrossRef](#)]
18. Xu, L.; Chen, F.Y.; Hayat, T. Hierarchical recursive signal modeling for multi-frequency signals based on discrete measured data. *Int. J. Adapt. Control Signal Process.* **2021**, *35*, 676–693. [[CrossRef](#)]
19. Katahira, K. How hierarchical models improve point estimates of model parameters at the individual level. *J. Math. Psychol.* **2016**, *73*, 37–58. [[CrossRef](#)]
20. Atitallah, A.; Bedoui, S.; Abderrahim, K. Multistage for identification of Wiener time delay systems based on hierarchical gradient approach. *Math. Comput. Model. Dyn. Syst.* **2017**, *23*, 222–239. [[CrossRef](#)]
21. Wang, Y.J.; Wu, M.H. Recursive parameter estimation algorithm for multivariate output-error systems. *J. Frankl. Inst.* **2018**, *355*, 5163–5181. [[CrossRef](#)]
22. Ding, F.; Wang, Y.J. The filtering based iterative identification for multivariable systems. *IET Control Theory Appl.* **2016**, *10*, 894–902.
23. Ji, Y.; Jiang, X.K.; Wan, L.J. Hierarchical least squares parameter estimation algorithm for two-input Hammerstein finite impulse response systems. *J. Frankl. Inst.* **2020**, *357*, 5019–5032. [[CrossRef](#)]
24. Wang, D.Q.; Mao, L.; Ding, F. Recasted models-based hierarchical extended stochastic gradient method for MIMO nonlinear systems. *IET Control Theory Appl.* **2017**, *11*, 476–485. [[CrossRef](#)]
25. Li, X.D.; Li, P. Input-to-state stability of nonlinear systems: Event-triggered impulsive control. *IEEE Trans. Autom. Control* **2021**. [[CrossRef](#)]
26. Li, P.; Li, X.D.; Lu, J.Q. Input-to-state stability of impulsive delay systems with multiple impulses. *IEEE Trans. Autom. Control* **2021**, *66*, 362–368. [[CrossRef](#)]
27. Chen, F.Y.; Xu, L.; Hayat, T. Data filtering based maximum likelihood extended gradient method for multivariable systems with autoregressive moving average noise. *J. Frankl. Inst.* **2018**, *355*, 3381–3398. [[CrossRef](#)]
28. Mao, Y.W.; Ding, F. Parameter estimation for nonlinear systems by using the data filtering and the multi-innovation identification theory. *Int. J. Comput. Math.* **2016**, *93*, 1869–1885. [[CrossRef](#)]
29. 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](#)]
30. Zhao, Z.Y.; Zhou, Y.Q.; Wang, X.Y.; Wang, Z.; Bai, Y. Water quality evolution mechanism modeling and health risk assessment based on stochastic hybrid dynamic systems. *Expert Syst. Appl.* **2022**, *193*, 116404. [[CrossRef](#)]
31. Li, W.L.; Jia, Y.M.; Du, J.P. Event-triggered Kalman consensus filter over sensor networks. *IET Control Theory Appl.* **2015**, *10*, 103–110. [[CrossRef](#)]
32. Kruzick, S.; Moura, J.M.F. Optimal filter design for signal processing on random graphs: Accelerated consensus. *IEEE Trans. Signal Process.* **2018**, *66*, 1258–1272. [[CrossRef](#)]
33. Pan, X.; Zhao, J.; Xu, J. An object-based and heterogeneous segment filter convolutional neural network for high-resolution remote sensing image classification. *Int. J. Remote Sens.* **2019**, *40*, 5892–5916. [[CrossRef](#)]
34. Zhou, Y.H.; Zhang, X. Hierarchical estimation approach for RBF-AR models with regression weights based on the increasing data length. *IEEE Trans. Circuits Syst. Express Briefs* **2021**, *68*, 3597–3601. [[CrossRef](#)]
35. Ding, F.; Chen, T. Combined parameter and output estimation of dual-rate systems using an auxiliary model. *Automatica* **2004**, *40*, 1739–1748. [[CrossRef](#)]
36. Ding, F.; Shi, Y.; Chen, Y.T. Auxiliary model-based least-squares identification methods for Hammerstein output-error systems. *Syst. Control Lett.* **2007**, *56*, 373–380. [[CrossRef](#)]

-
37. Zhou, Y.H. Modeling nonlinear processes using the radial basis function-based state-dependent autoregressive models. *IEEE Signal Process. Lett.* **2020**, *27*, 1600–1604. [[CrossRef](#)]
 38. Liu, Y.J.; Shi, Y. An efficient hierarchical identification method for general dual-rate sampled-data systems. *Automatica* **2014**, *50*, 962–970. [[CrossRef](#)]
 39. Zhang, X. Optimal adaptive filtering algorithm by using the fractional-order derivative. *IEEE Signal Process. Lett.* **2022**, *29*, 399–403. [[CrossRef](#)]