machines

MDPI

Article

Auxiliary Model-Based Multi-Innovation Fractional Stochastic Gradient Algorithm for Hammerstein Output-Error Systems

Chen Xu [1] and Yawen Mao [2,*]

[1] Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China; chenxu@jiangnan.edu.cn
[2] School of Science, Jiangnan University, Wuxi 214122, China
* Correspondence: myw0530@163.com

Abstract: This paper focuses on the nonlinear system identification problem, which is a basic premise of control and fault diagnosis. For Hammerstein output-error nonlinear systems, we propose an auxiliary model-based multi-innovation fractional stochastic gradient method. The scalar innovation is extended to the innovation vector for increasing the data use based on the multi-innovation identification theory. By establishing appropriate auxiliary models, the unknown variables are estimated and the improvement in the performance of parameter estimation is achieved owing to the fractional-order calculus theory. Compared with the conventional multi-innovation stochastic gradient algorithm, the proposed method is validated to obtain better estimation accuracy by the simulation results.

Keywords: hammerstein output-error systems; auxiliary model; multi-innovation identification theory; fractional-order calculus theory

Citation: Xu, C.; Mao, Y. Auxiliary Model-Based Multi-Innovation Fractional Stochastic Gradient Algorithm for Hammerstein Output-Error Systems. *Machines* **2021**, *9*, 247. https://doi.org/10.3390/machines9110247

Academic Editors: Hongtian Chen, Kai Zhong, Guangtao Ran and Chao Cheng

Received: 29 August 2021
Accepted: 21 October 2021
Published: 23 October 2021

1. Introduction

The accuracy of a system model affects the performance and safety of industrial control systems [1–5], and system identification is a theory and method for constructing mathematical model of systems and has been widely implemented in practice [6–9]. The behavior of most modern industrial control systems and synthetic systems are nonlinear by nature. Presently, an important research field in modern signal processing is the research of parameter identification for nonlinear systems, in which the block-structure systems, such as the Hammerstein model, are among the most current nonlinear systems due to their efficiency and accuracy to model complex nonlinear systems [10–12]. The representative feature of a Hammerstein model is that its architecture consists of two blocks: a static nonlinear model followed by a linear dynamic model. The simplicity in structure makes it provide a good compromise between the accuracy of nonlinear systems and the tractability of linear systems, and thus promoting its use in different nonlinear applications such as automatic control [13–15], fault detection and diagnosis [16–18], and so on.

Recently, several new system identification methods and theories have been developed for nonlinear models in the literature, including the least squares methods [19], the gradient-based methods [20], the iterative methods [21],the subspace identification methods [22], the hierarchical identification theory [23], the auxiliary model and the multi-innovation (MI) identification theories [24]. One well-known algorithm is the stochastic gradient (SG) algorithm, which has lower computational cost and complexity than the recursive least squares algorithm, whereas slow-convergence phenomena are often observed. Therefore, different modifications of the SG algorithm were developed to enhance its performance [25–30]. In particular, by extending scalar innovation into innovation vectors, the MI identification theory was proposed to improve the convergence speed and estimation accuracy in [31], and the fractional-order calculus method was introduced to show that it can achieve more satisfactory performance in [32,33].

To the best of our knowledge, different fractional-order gradient methods have been produced [34–36]. For example, in [37], a fractional-order SG algorithm was designed to identify the Hammerstein nonlinear ARMAX systems by an improved fractional-order gradient method. Based on the MI theory and the fractional-order calculus, an MI fractional least mean squares identification algorithm was presented for the Hammerstein controlled autoregressive systems, where the update mechanism was composed of the first-order gradient and the fractional gradient [38]. However, the above-discussed papers only consider the Hammerstein equation-error systems, and the cross-products between the parameters in the linear block and nonlinear block can lead to many redundant parameters. When the dimensions of parameter vectors are large, it will cause high computational complexity and deteriorate the identification accuracy.

In this work, we study the identification problem of the Hammerstein output-error moving average (OEMA) systems, which have been less studied due to the difficulty in identification [39,40]. To avoid estimating the redundant parameters, the Hammerstein model is parameterized using the key-term separation principle [41]. Furthermore, based on the identification model, the fractional-order SG algorithm is extended to the identification of Hammerstein OEMA systems and an auxiliary model-based multi-innovation fractional stochastic gradient (AM-MIFSG) algorithm is presented by the auxiliary model identification idea. The proposed algorithm can generate higher estimation accuracy than the common multi-innovation stochastic gradient (MISG) algorithm, with fewer parameters required to be estimated.

The paper is structured as follows. Section 2 gives a description for Hammerstein OEMA systems. Section 3 introduces the multi-innovation identification theory and drives an auxiliary model-based multi-innovation stochastic gradient (AM-MISG) identification algorithm for a comparison purpose. Section 4 presents the AM-MIFSG identification algorithm for the Hammerstein OEMA systems. Section 5 gives the convergence analysis of the proposed AM-MIFSG algorithm. Section 6 verifies the results in this paper using a simulation example. Finally, concluding remarks are given in Section 7.

2. The System Description

Consider the Hammerstein OEMA systems shown in Figure 1,

$$y_k = \frac{B(z)}{A(z)}\bar{u}_k + D(z)v_k, \tag{1}$$

$$\bar{u}_k = c_1 f_1(u_k) + c_2 f_2(u_k) + \cdots + c_m f_m(u_k), \tag{2}$$

where $\{u_k\}$ and $\{y_k\}$ are the input and output sequences of the system, $\{\bar{u}_k\}$ is the output sequence of the nonlinear block, and it can be represented as a linear combination of a known basis $\boldsymbol{f}(u_k) := [f_1(u_k), f_2(u_k), \cdots, f_m(u_k)]$ with unknown coefficients c_i $(i = 1, 2, \cdots, m)$, $\{v_k\}$ is a stochastic white noise sequence with zero mean and variance σ^2, $A(z)$, $B(z)$ and $D(z)$ are the polynomials in the unit backward shift operator z^{-1} $[z^{-1}y_k = y_{k-1}]$, and defined as

$$\begin{aligned} A(z) &:= 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a}, \\ B(z) &:= 1 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b}, \\ D(z) &:= 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots + d_{n_d} z^{-n_d}. \end{aligned}$$

Assume that the orders of these polynomials n_a, n_b and n_d are known and $u_k = 0$, $y_k = 0$ and $v_k = 0$ for $k \leqslant 0$.

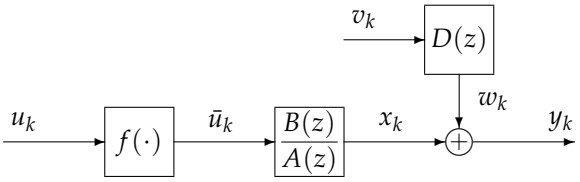


Figure 1. The Hammerstein OEMA systems.

Define the intermediate variables x_k and w_k as follows:

$$\begin{aligned} x_k &:= \frac{B(z)}{A(z)}\bar{u}_k \\ &= [1 - A(z)]x_k + B(z)\bar{u}_k \\ &= \bar{u}_k - \sum_{i=1}^{n_a} a_i x_{k-i} + \sum_{i=1}^{n_b} b_i \bar{u}_{k-i}, \end{aligned} \tag{3}$$

$$\begin{aligned} w_k &:= D(z)v_k \\ &= \sum_{i=1}^{n_d} d_i v_{k-i} + v_k. \end{aligned} \tag{4}$$

Take the first variable $\bar{u}_k$ on the right-hand side of (3) as a separated key-term. Based on the principle of key-term separation [42,43], substituting $\bar{u}_k$ in (2) into (3) gives

$$x_k = \sum_{i=1}^{m} c_i f_i(u_k) - \sum_{i=1}^{n_a} a_i x_{k-i} + \sum_{i=1}^{n_b} b_i \bar{u}_{k-i}. \tag{5}$$

Define the following related parameter vectors:

$$\begin{aligned} \boldsymbol{\theta} &:= \begin{bmatrix} \boldsymbol{\theta}_{\mathrm{s}} \\ \boldsymbol{d} \end{bmatrix} \in \mathbb{R}^n,\ n := n_a + n_b + n_d + m, \\ \boldsymbol{\theta}_{\mathrm{s}} &:= [\boldsymbol{a}^{\mathrm{T}}, \boldsymbol{b}^{\mathrm{T}}, \boldsymbol{c}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n_a+n_b+m}, \\ \boldsymbol{a} &:= [a_1, a_2, \cdots, a_{n_a}]^{\mathrm{T}} \in \mathbb{R}^{n_a}, \quad \boldsymbol{b} := [b_1, b_2, \cdots, b_{n_b}]^{\mathrm{T}} \in \mathbb{R}^{n_b}, \\ \boldsymbol{c} &:= [c_1, c_2, \cdots, c_m]^{\mathrm{T}} \in \mathbb{R}^{m}, \quad \boldsymbol{d} := [d_1, d_2, \cdots, d_{n_d}]^{\mathrm{T}} \in \mathbb{R}^{n_d}, \end{aligned}$$

and the information vectors:

$$\begin{aligned} \boldsymbol{\varphi}_k &:= \begin{bmatrix} \boldsymbol{\varphi}_{\mathrm{s},k} \\ \boldsymbol{\varphi}_{\mathrm{n},k} \end{bmatrix} \in \mathbb{R}^n, \\ \boldsymbol{\varphi}_{\mathrm{s},k} &:= [-x_{k-1}, -x_{k-2}, \cdots, -x_{k-n_a}, \bar{u}_{k-1}, \bar{u}_{k-2}, \cdots, \bar{u}_{k-n_b}, \boldsymbol{f}(u_k)]^{\mathrm{T}} \in \mathbb{R}^{n_a+n_b+m}, \\ \boldsymbol{\varphi}_{\mathrm{n},k} &:= [v_{k-1}, v_{k-2}, \cdots, v_{k-n_d}]^{\mathrm{T}} \in \mathbb{R}^{n_d}. \end{aligned}$$

From (1)–(5), we have

$$\begin{aligned} y_k &= x_k + w_k \\ &= \boldsymbol{\varphi}_{\mathrm{s},k}^{\mathrm{T}}\boldsymbol{\theta}_{\mathrm{s}} + \boldsymbol{\varphi}_{\mathrm{n},k}^{\mathrm{T}}\boldsymbol{d} + v_k \\ &= \boldsymbol{\varphi}_k^{\mathrm{T}}\boldsymbol{\theta} + v_k. \end{aligned} \tag{6}$$

Equation (6) is the identification model of the Hammerstein OEMA system. Please note that the parameter vector $\boldsymbol{\theta}$ contains all the parameters of the system in (1)–(2), and the parameters in the linear and nonlinear blocks are separated. This means there is no need to identify redundant parameters. This paper aims to present an AM-MIFSG algorithm for Hammerstein OEMA systems to improve the parameter estimation accuracy.

3. The AM-MISG Algorithm

In this section, we introduce the auxiliary model and multi-innovation identification theories briefly, and derive the AM-MISG algorithm for the Hammerstein OEMA system.

Let $\hat{\boldsymbol{\theta}}_k$ denote the estimate of $\boldsymbol{\theta}$. Based on the search principle of negative gradient, defining and minimizing the cost function

$$J(\boldsymbol{\theta}) := \frac{1}{2}\sum_{j=1}^{k}[y_j - \boldsymbol{\varphi}_j^{\mathrm{T}}\boldsymbol{\theta}]^2,$$

the following SG algorithm can be obtained for estimating the parameter vector $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} - \mu_1\frac{\partial J(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{k-1} + \frac{\boldsymbol{\varphi}_k}{s_k}e_k, \tag{7}$$

$$e_k = y_k - \boldsymbol{\varphi}_k^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{k-1}, \tag{8}$$

$$s_k = s_{k-1} + \|\boldsymbol{\varphi}_k\|^2. \tag{9}$$

where μ_1 is the step size for the SG algorithm, which is taken as $\mu_1 = \frac{1}{s_k}$, and $s_0 = 1$.

However, it is worth noting that the variables x_{k-i}, $\bar{u}_{k-i}$ and v_{k-i} in $\boldsymbol{\varphi}_k$ are unknown, and thus the algorithms in (7)–(9) cannot be implemented directly. The solution is to use the idea of the auxiliary model to build the following auxiliary models based on the parameter estimate $\hat{\boldsymbol{\theta}}_k$:

$$\begin{aligned}\hat{x}_k &= \hat{\boldsymbol{\varphi}}_{\mathrm{s},k}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{\mathrm{s},k},\\ \hat{\bar{u}}_k &= \hat{c}_{1,k}f_1(u_k) + \hat{c}_{2,k}f_2(u_k) + \cdots + \hat{c}_{m,k}f_m(u_k),\\ \hat{v}_k &= y_k - \hat{\boldsymbol{\varphi}}_k^{\mathrm{T}}\hat{\boldsymbol{\theta}}_k,\end{aligned}$$

and use the outputs $\hat{x}_{k-i}$, $\hat{\bar{u}}_{k-i}$ and $\hat{v}_{k-i}$ of the auxiliary models instead of the unknown variables x_{k-i}, $\bar{u}_{k-i}$ and v_{k-i} to construct the estimates of the information vectors:

$$\begin{aligned}\hat{\boldsymbol{\varphi}}_k &= \begin{bmatrix}\hat{\boldsymbol{\varphi}}_{\mathrm{s},k}\\ \hat{\boldsymbol{\varphi}}_{\mathrm{n},k}\end{bmatrix},\\ \hat{\boldsymbol{\varphi}}_{\mathrm{s},k} &= [-\hat{x}_{k-1}, -\hat{x}_{k-2}, \cdots, -\hat{x}_{k-n_a}, \hat{\bar{u}}_{k-1}, \hat{\bar{u}}_{k-2}, \cdots, \hat{\bar{u}}_{k-n_b}, \boldsymbol{f}(u_k)]^{\mathrm{T}},\\ \hat{\boldsymbol{\varphi}}_{\mathrm{n},k} &= [\hat{v}_{k-1}, \hat{v}_{k-2}, \cdots, \hat{v}_{k-n_d}]^{\mathrm{T}}.\end{aligned}$$

The SG algorithm update the parameter estimate using the current data information, thus its computational complexity is low, but estimation accuracy needs to be improved. Based on the multi-innovation identification theory [44,45], a slide window of length p (i.e., innovation length) is built to improve the estimation performance of the SG algorithm, which contains the data information from the current time k to $k-p+1$, i.e.,

$$\boldsymbol{E}_{p,k} = \left[y_k - \hat{\boldsymbol{\varphi}}_k^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{k-1}, y_{k-1} - \hat{\boldsymbol{\varphi}}_{k-1}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{k-2}, \cdots, y_{k-p+1} - \hat{\boldsymbol{\varphi}}_{k-p+1}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{k-p}\right]^{\mathrm{T}}. \tag{10}$$

Define the stacked output vector $\boldsymbol{Y}_{p,k}$ and information matrix $\hat{\boldsymbol{\Phi}}_{p,k}$ as

$$\begin{aligned}\boldsymbol{Y}_{p,k} &:= [y_k, y_{k-1}, \cdots, y_{k-p+1}]^{\mathrm{T}} \in \mathbb{R}^p,\\ \hat{\boldsymbol{\Phi}}_{p,k} &:= [\hat{\boldsymbol{\varphi}}_k, \hat{\boldsymbol{\varphi}}_{k-1}, \cdots, \hat{\boldsymbol{\varphi}}_{k-p+1}] \in \mathbb{R}^{n\times p}.\end{aligned}$$

In principle, the estimate $\hat{\boldsymbol{\theta}}_{t-1}$ is closer to the optimal value $\boldsymbol{\theta}$ than $\hat{\boldsymbol{\theta}}_{t-i}$ for $i = 2, \cdots, p$, then Equation (10) can be approximated by

$$\boldsymbol{E}_{p,k} = \boldsymbol{Y}_{p,k} - \hat{\boldsymbol{\Phi}}_{p,k}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{k-1}.$$

In summary, we can obtain the AM-MISG algorithm as follows:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_k &= \hat{\boldsymbol{\theta}}_{k-1} + \frac{\hat{\boldsymbol{\Phi}}_{p,k}}{s_k}\boldsymbol{E}_{p,k}, && (11)\\
\boldsymbol{E}_{p,k} &= \boldsymbol{Y}_{p,k} - \hat{\boldsymbol{\Phi}}_{p,k}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{k-1}, && (12)\\
s_k &= s_{k-1} + \|\hat{\boldsymbol{\varphi}}_k\|^2, \quad s_0 = 1, && (13)\\
\boldsymbol{Y}_{p,k} &= [y_k, y_{k-1}, \cdots, y_{k-p+1}]^{\mathrm{T}}, && (14)\\
\hat{\boldsymbol{\Phi}}_{p,k} &= [\hat{\boldsymbol{\varphi}}_k, \hat{\boldsymbol{\varphi}}_{k-1}, \cdots, \hat{\boldsymbol{\varphi}}_{k-p+1}], && (15)\\
\hat{\bar{u}}_k &= \boldsymbol{f}(u_k)\hat{\boldsymbol{c}}_k, && (16)\\
\hat{x}_k &= \hat{\boldsymbol{\varphi}}_{\mathrm{s},k}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{\mathrm{s},k}, && (17)\\
\hat{v}_k &= y_k - \hat{\boldsymbol{\varphi}}_k^{\mathrm{T}}\hat{\boldsymbol{\theta}}_k, && (18)\\
\boldsymbol{f}(u_k) &= [f_1(u_k), f_2(u_k), \cdots, f_m(u_k)], && (19)\\
\hat{\boldsymbol{\varphi}}_k &= \begin{bmatrix} \hat{\boldsymbol{\varphi}}_{\mathrm{s},k} \\ \hat{\boldsymbol{\varphi}}_{\mathrm{n},k} \end{bmatrix}, && (20)\\
\hat{\boldsymbol{\varphi}}_{\mathrm{s},k} &= [-\hat{x}_{k-1}, -\hat{x}_{k-2}, \cdots, -\hat{x}_{k-n_a}, \hat{\bar{u}}_{k-1}, \hat{\bar{u}}_{k-2}, \cdots, \hat{\bar{u}}_{k-n_b}, \boldsymbol{f}(u_k)]^{\mathrm{T}}, && (21)\\
\hat{\boldsymbol{\varphi}}_{\mathrm{n},k} &= [\hat{v}_{k-1}, \hat{v}_{k-2}, \cdots, \hat{v}_{k-n_d}]^{\mathrm{T}}, && (22)\\
\hat{\boldsymbol{\theta}}_k &= \begin{bmatrix} \hat{\boldsymbol{\theta}}_{\mathrm{s},k} \\ \hat{\boldsymbol{d}}_k \end{bmatrix}, && (23)\\
\hat{\boldsymbol{\theta}}_{\mathrm{s},k} &= [\hat{\boldsymbol{a}}_k^{\mathrm{T}}, \hat{\boldsymbol{b}}_k^{\mathrm{T}}, \hat{\boldsymbol{c}}_k^{\mathrm{T}}]^{\mathrm{T}}. && (24)
\end{aligned}$$

Please note that the AM-MISG algorithm will reduce to the auxiliary model-based stochastic gradient (AM-SG) algorithm when $p = 1$.

4. The AM-MIFSG Algorithm

This section deduces an AM-MIFSG algorithm to improve the parameter estimation performance of above AM-MISG identification algorithm.

In (7), the first-order gradient is used to update the parameter vector. In contrast to the integer order, for the quadratic objective function, the derivative of a fractional-order near a point is uncertain, so its essential property is nonlocal. This excellent property enables the fractional-order gradient method to jump out of local optimum and reach global minimum point more quicker. Here, we propose to add the fractional-order gradient in addition to the first-order gradient, and the final update relation is written as:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} - \mu_1\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \mu_\alpha\frac{\partial^\alpha J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \tag{25}$$

where μ_α is the step size for the factional order derivative ∂^α. According to the Caputo and Riemann–Liouville definition [46,47], the fractional derivation of a power function $f(t) = t^n$ $(n > -1)$is defined as:

$$D_t^\alpha t^n = \frac{\Gamma(n+1)}{\Gamma(n+1-\alpha)}t^{n-\alpha}, \tag{26}$$

where D_t^α is the fractional derivative operator of order α and Γ is the gamma function which defined as $\Gamma(n) = (n-1)!$.

According to (26), the fractional-order gradient in Equation (25) can be written as follows:

$$\frac{\partial^\alpha J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\boldsymbol{\varphi}_k\left(\frac{\partial^\alpha \boldsymbol{\theta}}{\partial \boldsymbol{\theta}}\right) = -\boldsymbol{\varphi}_k\left(\frac{\Gamma(2)}{\Gamma(2-\alpha)}\boldsymbol{\theta}^{1-\alpha}\right), \tag{27}$$

where $\Gamma(2) = 1$. Then Equation (25) can be approximated as follows:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} + \frac{\boldsymbol{\varphi}_k}{s_k} e_k + \frac{\boldsymbol{\psi}_k}{s_{\alpha,k}} e_k, \quad 0 < \alpha < 1, \tag{28}$$

$$s_{\alpha,k} = s_{\alpha,k-1} + \|\boldsymbol{\psi}_k\|^2, \quad s_{\alpha,0} = 1, \tag{29}$$

$$\boldsymbol{\psi}_k = \frac{\operatorname{diag}(\boldsymbol{\varphi}_k)(|\boldsymbol{\theta}|_{k-1}^{1-\alpha})}{\Gamma(2-\alpha)}. \tag{30}$$

Please note that the absolute value of $\boldsymbol{\theta}$ is used to avoid complex values, this is a common way of dealing with fractional-order gradient [38]. The introduction of fractional-order parameter α provides additional degrees of freedom and increases the flexibility of the parameter estimation.

Similar to the AM-MISG algorithm in Section 3, expanding the information vector $\boldsymbol{\psi}_k$ to the information matrix

$$\boldsymbol{\Psi}_{p,k} = [\boldsymbol{\psi}_k, \boldsymbol{\psi}_{k-1}, \cdots, \boldsymbol{\psi}_{k-p+1}],$$

and applying the auxiliary model identification idea, we can obtain the following AM-MIFSG algorithm:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} + \left(\frac{\hat{\boldsymbol{\Phi}}_{p,k}}{s_k} + \frac{\hat{\boldsymbol{\Psi}}_{p,k}}{s_{\alpha,k}} \right) \boldsymbol{E}_{p,k}, \tag{31}$$

$$\boldsymbol{E}_{p,k} = \boldsymbol{Y}_{p,k} - \hat{\boldsymbol{\Phi}}_{p,k}^{\mathrm{T}} \hat{\boldsymbol{\theta}}_{k-1}, \tag{32}$$

$$s_k = s_{k-1} + \|\hat{\boldsymbol{\varphi}}_k\|^2, \quad s_0 = 1, \tag{33}$$

$$s_{\alpha,k} = s_{\alpha,k-1} + \|\hat{\boldsymbol{\psi}}_k\|^2, \quad s_{\alpha,0} = 1, \tag{34}$$

$$\boldsymbol{Y}_{p,k} = [y_k, y_{k-1}, \cdots, y_{k-p+1}]^{\mathrm{T}}, \tag{35}$$

$$\hat{\boldsymbol{\Phi}}_{p,k} = [\hat{\boldsymbol{\varphi}}_k, \hat{\boldsymbol{\varphi}}_{k-1}, \cdots, \hat{\boldsymbol{\varphi}}_{k-p+1}], \tag{36}$$

$$\hat{\boldsymbol{\Psi}}_{p,k} = [\hat{\boldsymbol{\psi}}_k, \hat{\boldsymbol{\psi}}_{k-1}, \cdots, \hat{\boldsymbol{\psi}}_{k-p+1}], \tag{37}$$

$$\hat{\boldsymbol{\psi}}_j = \frac{\operatorname{diag}(\hat{\boldsymbol{\varphi}}_j)(|\hat{\boldsymbol{\theta}}|_{k-1}^{1-\alpha})}{\Gamma(2-\alpha)}, \quad j = k, k-1, \cdots, k-p+1, \tag{38}$$

$$\hat{\tilde{u}}_k = \boldsymbol{f}(u_k)\hat{\boldsymbol{c}}_k, \tag{39}$$

$$\hat{x}_k = \hat{\boldsymbol{\varphi}}_{\mathrm{s},k}^{\mathrm{T}} \hat{\boldsymbol{\theta}}_{\mathrm{s},k}, \tag{40}$$

$$\hat{v}_k = y_k - \hat{\boldsymbol{\varphi}}_k^{\mathrm{T}} \hat{\boldsymbol{\theta}}_k, \tag{41}$$

$$\boldsymbol{f}(u_k) = [f_1(u_k), f_2(u_k), \cdots, f_m(u_k)], \tag{42}$$

$$\hat{\boldsymbol{\varphi}}_k = \begin{bmatrix} \hat{\boldsymbol{\varphi}}_{\mathrm{s},k} \\ \hat{\boldsymbol{\varphi}}_{\mathrm{n},k} \end{bmatrix}, \tag{43}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{s},k} = [-\hat{x}_{k-1}, -\hat{x}_{k-2}, \cdots, -\hat{x}_{k-n_a}, \hat{\tilde{u}}_{k-1}, \hat{\tilde{u}}_{k-2}, \cdots, \hat{\tilde{u}}_{k-n_b}, \boldsymbol{f}(u_k)]^{\mathrm{T}}, \tag{44}$$

$$\hat{\boldsymbol{\varphi}}_{\mathrm{n},k} = [\hat{v}_{k-1}, \hat{v}_{k-2}, \cdots, \hat{v}_{k-n_d}]^{\mathrm{T}}, \tag{45}$$

$$\hat{\boldsymbol{\theta}}_k = \begin{bmatrix} \hat{\boldsymbol{\theta}}_{\mathrm{s},k} \\ \hat{\boldsymbol{d}}_k \end{bmatrix}, \tag{46}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{s},k} = [\hat{\boldsymbol{a}}_k^{\mathrm{T}}, \hat{\boldsymbol{b}}_k^{\mathrm{T}}, \hat{\boldsymbol{c}}_k^{\mathrm{T}}]^{\mathrm{T}}. \tag{47}$$

Here, the above AM-MIFSG algorithm reduces to the auxiliary model-based fractional stochastic gradient (AM-FSG) algorithm when $p = 1$.

Remark 1. *In general, as the innovation length p increases, the collected data are being used more fully, and therefore the estimation accuracy is gradually improved. However, the computational amount increases at the same time. How to choose optimal innovation p is an open problem to be solved. In practice, we often choose* $p < n$.

Remark 2. *The differential order α is chose in the range of (0,1). The orders may show different characteristics for different systems, and can be adjusted during the procedure as needed.*

The implementation of the AM-MIFSG algorithm is listed as follows.

1. Choose p, α and initialize: let $k = 1$, $\hat{\boldsymbol{\theta}}_0 = \begin{bmatrix} \hat{\boldsymbol{\theta}}_{s,0} \\ \hat{\boldsymbol{d}}_0 \end{bmatrix} = 1_{n/p_0}$, $s_0 = 1$, $s_{\alpha,0} = 1$, and set $\hat{x}_i = 1/p_0$, $\hat{u}_i = 1/p_0$ and $\hat{v}_i = 1/p_0$ for $i \leqslant 0$, $p_0 = 10^6$, and give the base function $f_i(\cdot)$.
2. Collect the input-output data u_k and y_k, form the basis function vector $\boldsymbol{f}(u_k)$ by (42), and the information vectors $\hat{\boldsymbol{\varphi}}_k$ by (43), $\hat{\boldsymbol{\varphi}}_{s,k}$ by (44) and $\hat{\boldsymbol{\varphi}}_{n,k}$ by (45).
3. Compute $\hat{\boldsymbol{\psi}}_j$ by (38). Form the stacked output vector $\boldsymbol{Y}_{p,k}$ by (35), the information matrices $\hat{\boldsymbol{\Phi}}_{p,k}$ and $\hat{\boldsymbol{\Psi}}_{p,k}$ by (36) and (37).
4. Compute the innovation vector $\boldsymbol{E}_{p,k}$ by (32), s_k by (33) and $s_{\alpha,k}$ by (34).
5. Update the parameter estimate $\hat{\boldsymbol{\theta}}_k$ by (31), compute the estimates $\hat{u}_k$ by (39), $\hat{x}_k$ by (40), $\hat{v}_k$ by (41).
6. Increase k by 1, go to step 2.

The algorithm obtained above combined with the method in [48–53] can cope with linear and nonlinear systems with different disturbances. Furthermore, prediction models or soft sensor models can be obtained with the assistance of other parameter estimation algorithms [54–59] and can be applied to process control and other fields [60–65].

5. Convergence Analysis

Theorem 1. *For the system in (1)–(2) and the AM-MIFSG algorithm in (31)–(47), assume that the noise sequence $\{v_k\}$ satisfies*

$$(\text{A1})\ \mathrm{E}[v_k|\mathcal{F}_t] = 0,\ \text{a.s.},\ \mathrm{E}[v_k^2|\mathcal{F}_t] \leqslant \sigma^2 < \infty,\ \text{a.s.},$$

and there exist an integer N_k and a positive constant ϱ independent of k such that the following persistent excitation condition holds,

$$(\text{A2})\ \sum_{i=0}^{N_k} \frac{\hat{\boldsymbol{\Phi}}_{\alpha,p,k+i}^{\mathrm{T}} \hat{\boldsymbol{\Phi}}_{\alpha,p,k+i}}{s_{k+i}} \geqslant \varrho \boldsymbol{I},\ \text{a.s.}, \tag{48}$$

where $\hat{\boldsymbol{\Phi}}_{\alpha,p,k} = [\hat{\boldsymbol{\varphi}}_k \odot \boldsymbol{\theta}_\alpha, \hat{\boldsymbol{\varphi}}_{k-1} \odot \boldsymbol{\theta}_\alpha, \cdots, \hat{\boldsymbol{\varphi}}_{k-p+1} \odot \boldsymbol{\theta}_\alpha]$, $\boldsymbol{\theta}_\alpha := 1_n + \hat{\boldsymbol{\theta}}_{k-1}^{1-\alpha}$, $\odot$ denotes an element-by-element multiplication of vectors. Then the parameter estimation error given by the AM-MIFSG algorithm satisfies $\lim_{k\to\infty} \mathrm{E}[\|\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}\|^2] \to 0$.

Proof. Define the parameter estimation error $\bar{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta} \in \mathbb{R}^n$. To simplify the proof, assuming $s_{\alpha,k} = s_k/\Gamma(2-\alpha)$. Inserting (32) into (31) and rearranging, we have

$$\begin{aligned}
\bar{\boldsymbol{\theta}}_k &= \bar{\boldsymbol{\theta}}_{k-1} + \frac{\hat{\boldsymbol{\Phi}}_{p,k}}{s_k}\left[\boldsymbol{Y}_{p,k} - \hat{\boldsymbol{\Phi}}_{p,k}^{\mathrm{T}} \hat{\boldsymbol{\theta}}_{k-1}\right] \odot \boldsymbol{\theta}_\alpha \\
&= \bar{\boldsymbol{\theta}}_{k-1} + \frac{\hat{\boldsymbol{\Phi}}_{p,k}}{s_k}\left[\boldsymbol{\Phi}_{p,k}^{\mathrm{T}} \boldsymbol{\theta}_{k-1} + \boldsymbol{V}_{p,k} - \hat{\boldsymbol{\Phi}}_{p,k}^{\mathrm{T}} \hat{\boldsymbol{\theta}}_{k-1}\right] \odot \boldsymbol{\theta}_\alpha \\
&=: \bar{\boldsymbol{\theta}}_{k-1} + \frac{\hat{\boldsymbol{\Phi}}_{p,k}}{s_k}\left[\boldsymbol{\mu}_{p,k} - \boldsymbol{\varsigma}_{p,k} + \boldsymbol{V}_{p,k}\right] \odot \boldsymbol{\theta}_\alpha,
\end{aligned} \tag{49}$$

where

$$\begin{aligned}
\boldsymbol{\mu}_{q,t} &:= [\boldsymbol{\Phi}_{p,k} - \hat{\boldsymbol{\Phi}}_{p,k}]^{\mathrm{T}} \boldsymbol{\theta} \in \mathbb{R}^p, \quad \boldsymbol{\varsigma}_{q,t} := \hat{\boldsymbol{\Phi}}_{p,k}^{\mathrm{T}} \bar{\boldsymbol{\theta}}_{k-1} \in \mathbb{R}^p, \\
\boldsymbol{V}_{p,k} &:= [v_k, v_{k-1}, \cdots, v_{k-p+1}] \in \mathbb{R}^p.
\end{aligned}$$

Pre-multiplying (49) by $\bar{\boldsymbol{\theta}}_k^{\mathrm{T}}$ gives

$$\begin{aligned}\bar{\boldsymbol{\theta}}_k^{\mathrm{T}}\bar{\boldsymbol{\theta}}_k &= \bar{\boldsymbol{\theta}}_{k-1}^{\mathrm{T}}\bar{\boldsymbol{\theta}}_{k-1} + \frac{2}{s_k}\bar{\boldsymbol{\theta}}_{k-1}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_{\alpha,p,k}[\boldsymbol{\mu}_{p,k} - \boldsymbol{\varsigma}_{p,k} + \boldsymbol{V}_{p,k}] \\ &\quad + \frac{1}{r_k^2}[\boldsymbol{\mu}_{p,k} - \boldsymbol{\varsigma}_{p,k} + \boldsymbol{V}_{p,k}]^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_{\alpha,p,k}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_{\alpha,p,k}[\boldsymbol{\mu}_{p,k} - \boldsymbol{\varsigma}_{p,k} + \boldsymbol{V}_{p,k}].\end{aligned}$$

□

The rest can be proved in a similar to the way in [66].

6. Examples

Consider the following Hammerstein OEMA system:

$$\begin{aligned} y_k &= \frac{B(z)}{A(z)}\bar{u}_k + D(z)v_k, \\ A(z) &= 1 + a_1z^{-1} + a_2z^{-2} = 1 + 0.45z^{-1} + 0.56z^{-2}, \\ B(z) &= 1 + b_1z^{-1} + b_2z^{-2} = 1 + 0.25z^{-1} - 0.35z^{-2}, \\ D(z) &= 1 + d_1z^{-1} = 1 - 0.54z^{-1}, \\ \bar{u}_k &= c_1u_k + c_2u_k^2 + c_3u_k^3 = 0.52u_k + 0.54u_k^2 + 0.82u_k^3, \\ \boldsymbol{\theta} &= [a_1, a_2, b_1, b_2, c_1, c_2, c_3, d_1]^{\mathrm{T}} = [0.45, 0.56, 0.25, -0.35, 0.52, 0.54, 0.82, -0.54]^{\mathrm{T}}. \end{aligned}$$

In this example, the input $\{u_k\}$ is a persistently excited signal sequence and $\{v_k\}$ is a white noise sequence with zero mean and variances $\sigma^2 = 0.80^2$. The data length is taken as $L = 4000$, where the first 3500 samples are assigned for system identification and the remaining 500 samples are assigned for prediction and validation. The details are as follows.

1. Firstly, applying the AM-MISG algorithm and the AM-MIFSG algorithm with $\alpha = 0.94$ to estimate the parameters of considered system. Tables 1 and 2 show the parameter estimates and their errors with $p = 1, 2, 4$ and 6. Figures 2 and 3 indicate the parameter estimation errors $\delta := \|\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}\|/\|\boldsymbol{\theta}\|$ versus k.

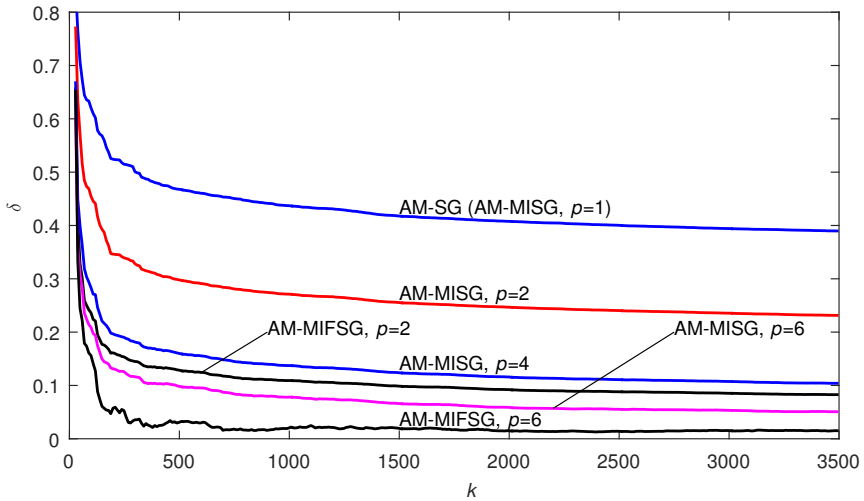


Figure 2. The AM-MISG estimation error δ versus k with $p = 1, 2, 4$ and 6 and the AM-MIFSG estimation error δ versus k with $p = 2$ and 6.

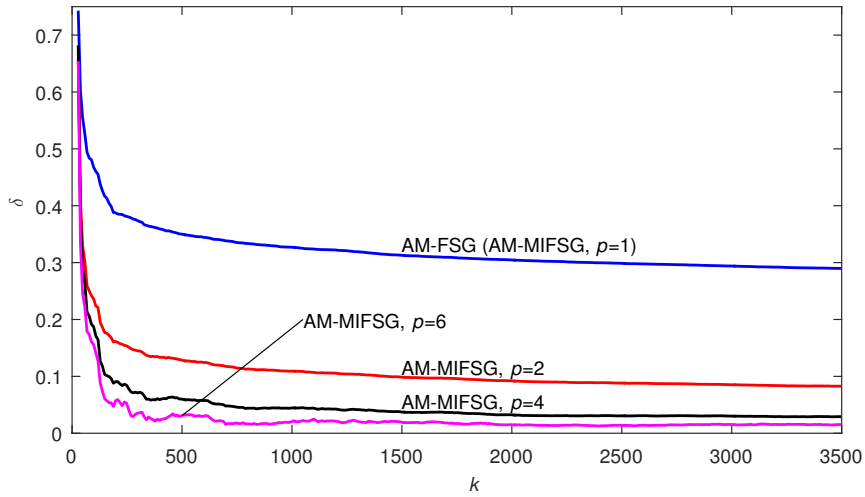


Figure 3. The AM-MIFSG estimation error δ versus k with $p = 1, 2, 4$ and 6.

Table 1. The AM-MISG estimates and errors $p = 1, 2, 4$ and 6.

p	k	a_1	a_2	b_1	b_2	c_1	c_2	c_3	d_1	$\delta(\%)$
1	100	0.03695	0.23553	−0.03327	−0.20875	0.27338	0.03195	0.49278	−0.31825	61.82507
	200	0.07113	0.34613	−0.03653	−0.29482	0.32261	0.07100	0.59507	−0.37276	52.41770
	500	0.10951	0.42150	−0.04098	−0.34040	0.35083	0.10812	0.64790	−0.39006	46.76773
	1000	0.13632	0.46020	−0.03759	−0.36082	0.36386	0.12852	0.67688	−0.39765	43.71611
	2000	0.16601	0.49396	−0.03164	−0.38052	0.37629	0.14894	0.70178	−0.40765	40.77833
	3000	0.18127	0.50703	−0.02857	−0.38630	0.38212	0.15952	0.71245	−0.41028	39.42179
2	100	0.08718	0.45055	−0.02857	−0.34926	0.38250	0.12367	0.64113	−0.44165	45.20563
	200	0.21252	0.53700	−0.00743	−0.40687	0.43140	0.19230	0.74170	−0.46339	34.63613
	500	0.27602	0.53381	0.01178	−0.39497	0.44676	0.23408	0.76740	−0.46163	29.78555
	1000	0.29539	0.53658	0.03287	−0.38875	0.45350	0.26006	0.78181	−0.45922	27.12588
	2000	0.31601	0.54589	0.05191	−0.39039	0.45935	0.28327	0.79322	−0.46058	24.67852
	3000	0.32385	0.54810	0.06104	−0.38751	0.46189	0.29484	0.79696	−0.46005	23.55543
4	100	0.27486	0.60968	0.03978	−0.39097	0.50049	0.23079	0.79164	−0.50729	28.18166
	200	0.37933	0.57741	0.09918	−0.37644	0.52086	0.30714	0.83586	−0.48517	19.67676
	500	0.38912	0.54069	0.13209	−0.34877	0.51837	0.35074	0.82239	−0.48274	16.01070
	1000	0.38933	0.54894	0.15511	−0.34580	0.51921	0.37894	0.82425	−0.48325	13.73325
	2000	0.40081	0.55745	0.17376	−0.34949	0.51943	0.40309	0.82417	−0.48612	11.58755
	3000	0.40202	0.55950	0.18102	−0.34717	0.51845	0.41433	0.82086	−0.48675	10.74226
6	100	0.35169	0.61755	0.08824	−0.35474	0.53581	0.30223	0.83886	−0.52654	20.81443
	200	0.40866	0.59002	0.16182	−0.34923	0.53654	0.37953	0.84693	−0.50426	13.13328
	500	0.42035	0.54952	0.18298	−0.32811	0.52821	0.42093	0.81930	−0.50377	9.82957
	1000	0.41823	0.56425	0.20204	−0.33346	0.53010	0.44667	0.82367	−0.50609	7.80786
	2000	0.42981	0.56613	0.21766	−0.34005	0.52962	0.46847	0.82213	−0.50962	5.89007
	3000	0.42678	0.56730	0.22265	−0.33911	0.52764	0.47733	0.81677	−0.51053	5.32836
True values		0.45000	0.56000	0.25000	−0.35000	0.52000	0.54000	0.82000	−0.54000	

Table 2. The AM-MIFSG estimates and errors with $p = 1$, 2, 4 and 6.

p	k	a_1	a_2	b_1	b_2	c_1	c_2	c_3	d_1	$\delta(\%)$
1	100	0.15708	0.59506	−0.10818	−0.36395	0.38627	0.08191	0.73420	−0.36577	46.47154
	200	0.27641	0.58549	−0.07338	−0.37468	0.42517	0.13842	0.81526	−0.37310	38.73219
	500	0.29432	0.55031	−0.03625	−0.35604	0.43184	0.17816	0.82178	−0.37342	34.99604
	1000	0.29755	0.54549	−0.01005	−0.35097	0.43470	0.20499	0.82736	−0.37378	32.70875
	2000	0.30878	0.55280	0.01095	−0.35501	0.43788	0.22916	0.83306	−0.37827	30.47638
	3000	0.31387	0.55453	0.02109	−0.35366	0.43918	0.24158	0.83382	−0.37952	29.40027
2	100	0.35829	0.62744	0.15273	−0.37663	0.55109	0.23357	0.78715	−0.59971	23.46638
	200	0.42746	0.58770	0.19005	−0.36598	0.56024	0.31456	0.81037	−0.55565	16.12428
	500	0.43121	0.55375	0.20459	−0.34652	0.55703	0.35970	0.79703	−0.54665	12.87485
	1000	0.42642	0.56004	0.21858	−0.34550	0.55880	0.38789	0.80138	−0.54408	10.92266
	2000	0.43305	0.56476	0.22928	−0.34967	0.55890	0.41208	0.80204	−0.54336	9.22414
	3000	0.43194	0.56535	0.23216	−0.34777	0.55770	0.42315	0.79911	−0.54236	8.52773
4	100	0.37614	0.62528	0.15435	−0.35937	0.56108	0.31419	0.83373	−0.62653	18.87062
	200	0.43032	0.60035	0.22302	−0.36272	0.55406	0.42250	0.82836	−0.56139	9.08948
	500	0.44845	0.55152	0.22616	−0.33881	0.54491	0.46286	0.79775	−0.55411	6.00591
	1000	0.44218	0.56900	0.23948	−0.34615	0.55133	0.48607	0.81308	−0.55361	4.43982
	2000	0.44975	0.56313	0.24931	−0.35057	0.55048	0.50632	0.81111	−0.55405	3.24848
	3000	0.44202	0.56473	0.25018	−0.34923	0.54786	0.51266	0.80477	−0.55341	3.01322
6	100	0.37154	0.63684	0.14344	−0.33325	0.54734	0.38404	0.84946	−0.62029	15.86854
	200	0.43128	0.61835	0.23888	−0.36493	0.53142	0.48850	0.83328	−0.55888	5.77068
	500	0.45536	0.55467	0.23160	−0.34012	0.52315	0.50947	0.79826	−0.55434	3.08071
	1000	0.44922	0.57123	0.24453	−0.35088	0.53625	0.52484	0.82763	−0.55500	2.04837
	2000	0.45296	0.55918	0.25378	−0.35400	0.53417	0.53985	0.82120	−0.55601	1.49579
	3000	0.44123	0.56369	0.25237	−0.35328	0.53073	0.54151	0.81199	−0.55532	1.53252
True values		0.45000	0.56000	0.25000	−0.35000	0.52000	0.54000	0.82000	−0.54000	

2. Secondly, to validate the influence of the fraction order α, in the AM-MIFSG algorithm, we take $p = 5$ and 6, and α =0.80, 0.90 and 0.92, respectively, the simulation results are shown in Tables 3 and 4, and Figures 4 and 5.

3. In the end, a different data set ($L_e = 500$ samples from $k = 3501$ to 4000) and the estimated model obtained by the AM-MIFSG algorithm with $p = 6$ and $\alpha = 0.92$ are used for model validation. The predicted output and true output are plotted in Figure 6 from $k = 3501$ to 3700 and Figure 7 from $k = 3501$ to 4000, where the average predicted output error is

$$\delta_e = \frac{1}{L_e}\left[\sum_{k=3501}^{4000}[\hat{y}_k - y_k]^2\right]^{1/2} = 0.0658,$$

and the dots line is the output $\hat{y}_k$ of the estimated model and the solid line is the true output y_k.

From Tables 1–4 and Figures 2–7, we can draw the following conclusions: (1) with the innovation length p increases, both the AM-MISG and the AM-MIFSG algorithm can give higher parameter estimation accuracy; (2) in general, the AM-MIFSG algorithm has a faster convergence rate than the AM-MISG algorithm in the same situation, and the introduction of the fractional-order can improve the parameter estimation accuracy; (3) the convergence rate of the AM-MIFSG increases as the fractional-order α increases, the α within the range of [0.90, 0.95] seems to be an appropriate choice which can give better estimation results for the Hammerstein output-error systems; (4) the estimated model obtained by the AM-MIFSG algorithm can well capture system dynamics.

Table 3. The AM-MIFSG estimates and errors with $\alpha = 0.80$, 0.90 and 0.92 ($p = 5$).

α	k	a_1	a_2	b_1	b_2	c_1	c_2	c_3	d_1	$\delta(\%)$
0.80	100	0.23581	0.59985	0.09673	−0.40126	0.35249	0.17777	0.82055	−0.48512	32.54342
	200	0.43614	0.59824	0.19561	−0.40423	0.39862	0.34638	0.91657	−0.45204	18.57503
	500	0.44156	0.53689	0.21507	−0.35179	0.39149	0.42275	0.86423	−0.46396	13.37312
	1000	0.43930	0.56210	0.23408	−0.35731	0.40462	0.46695	0.88603	−0.47323	11.19137
	2000	0.44866	0.55624	0.24854	−0.35742	0.40774	0.49743	0.88189	−0.48368	9.82294
	3000	0.43832	0.56284	0.25029	−0.35452	0.40630	0.50670	0.87138	−0.48769	9.37582
0.90	100	0.25965	0.58762	0.16306	−0.43464	0.41011	0.32565	0.75782	−0.62289	23.25864
	200	0.43757	0.58825	0.25570	−0.43466	0.45124	0.46639	0.85201	−0.55254	9.35233
	500	0.46161	0.53207	0.24689	−0.37626	0.45338	0.49568	0.83323	−0.55019	6.10290
	1000	0.45443	0.55626	0.25376	−0.37420	0.46646	0.51604	0.85868	−0.55042	5.04977
	2000	0.45629	0.55114	0.25927	−0.36918	0.46731	0.53246	0.85369	−0.55163	4.58180
	3000	0.44425	0.55646	0.25785	−0.36480	0.46517	0.53551	0.84496	−0.55137	4.29344
0.92	100	0.36347	0.63518	0.16075	−0.38255	0.49377	0.31352	0.84459	−0.60486	18.82927
	200	0.44017	0.60438	0.24641	−0.38609	0.49205	0.43973	0.85042	−0.53959	8.24557
	500	0.45824	0.54625	0.24047	−0.35090	0.48640	0.47823	0.81880	−0.53739	4.87871
	1000	0.45085	0.56619	0.25046	−0.35726	0.49783	0.50209	0.84203	−0.53918	3.35483
	2000	0.45508	0.55783	0.25789	−0.35824	0.49792	0.52152	0.83812	−0.54161	2.43473
	3000	0.44406	0.56191	0.25685	−0.35603	0.49533	0.52624	0.82969	−0.54185	2.13756
True values		0.45000	0.56000	0.25000	−0.35000	0.52000	0.54000	0.82000	−0.54000	

Table 4. The AM-MIFSG estimates and errors with $\alpha = 0.80$, 0.90 and 0.92 ($p = 6$).

α	k	a_1	a_2	b_1	b_2	c_1	c_2	c_3	d_1	$\delta(\%)$
0.80	100	0.26215	0.64436	0.06610	−0.35700	0.41270	0.26270	0.84456	−0.55007	27.25073
	200	0.42892	0.61731	0.19706	−0.36999	0.43708	0.42526	0.90034	−0.50010	12.53946
	500	0.44308	0.55154	0.21290	−0.33527	0.42696	0.47865	0.84239	−0.50720	8.39876
	1000	0.44126	0.56985	0.23379	−0.34728	0.44327	0.50887	0.87299	−0.51449	6.95024
	2000	0.44893	0.55908	0.24812	−0.35102	0.44420	0.53026	0.86479	−0.52163	6.06418
	3000	0.43731	0.56558	0.24875	−0.35051	0.44213	0.53393	0.85338	−0.52364	5.87107
0.90	100	0.32475	0.62263	0.16067	−0.40949	0.45832	0.35530	0.80222	−0.63261	18.70725
	200	0.44220	0.60427	0.26639	−0.41415	0.47360	0.48866	0.84739	−0.55477	7.38987
	500	0.46480	0.54124	0.24811	−0.36447	0.47263	0.50941	0.82077	−0.55289	4.30671
	1000	0.45622	0.56250	0.25474	−0.36674	0.48814	0.52662	0.85226	−0.55350	3.52333
	2000	0.45641	0.55356	0.25988	−0.36364	0.48753	0.54148	0.84452	−0.55464	3.17023
	3000	0.44318	0.55972	0.25723	−0.36073	0.48473	0.54283	0.83450	−0.55411	2.90188
0.92	100	0.38062	0.64556	0.16445	−0.36053	0.51634	0.33828	0.85954	−0.61558	17.41472
	200	0.44313	0.61605	0.25735	−0.38049	0.50254	0.46548	0.84460	−0.54615	6.92261
	500	0.46199	0.55073	0.24347	−0.34802	0.49615	0.49609	0.80967	−0.54393	3.60595
	1000	0.45377	0.56875	0.25282	−0.35650	0.51054	0.51672	0.83983	−0.54582	2.32078
	2000	0.45586	0.55746	0.25944	−0.35758	0.50944	0.53419	0.83334	−0.54795	1.60509
	3000	0.44334	0.56272	0.25702	−0.35604	0.50628	0.53692	0.82365	−0.54786	1.35745
True values		0.45000	0.56000	0.25000	−0.35000	0.52000	0.54000	0.82000	−0.54000	

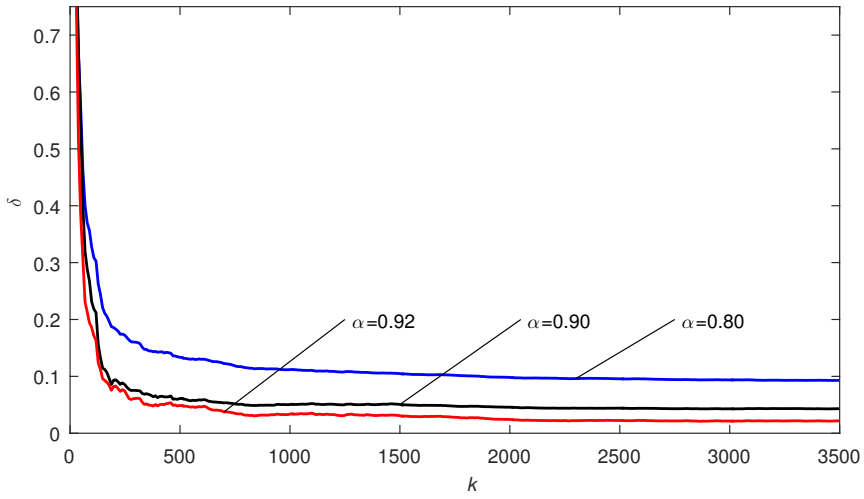


Figure 4. The AM-MIFSG estimation error δ versus k with $\alpha = 0.80$, 0.90 and 0.92 ($p = 5$).

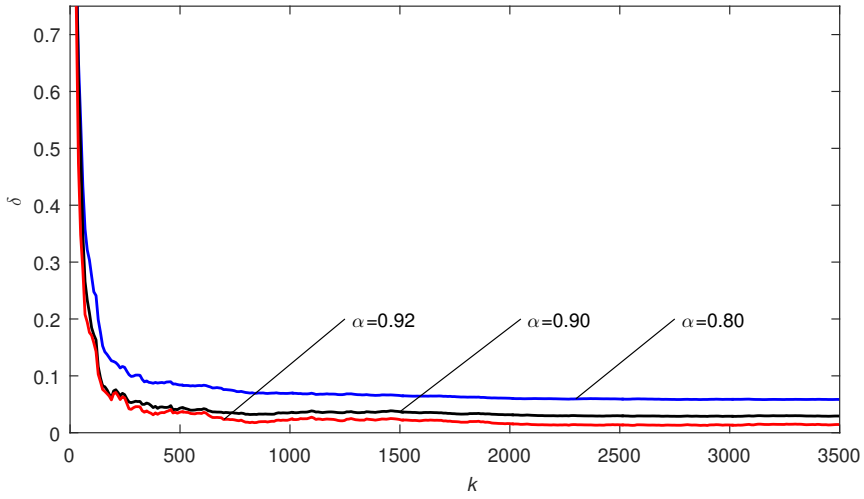
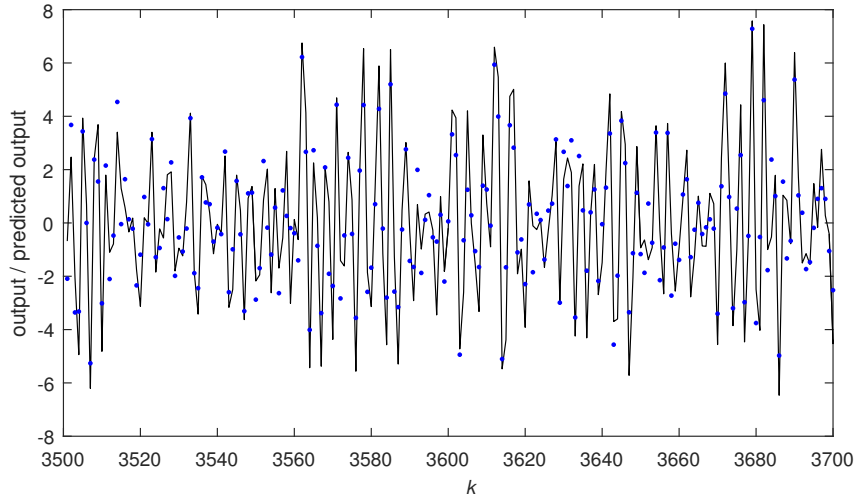
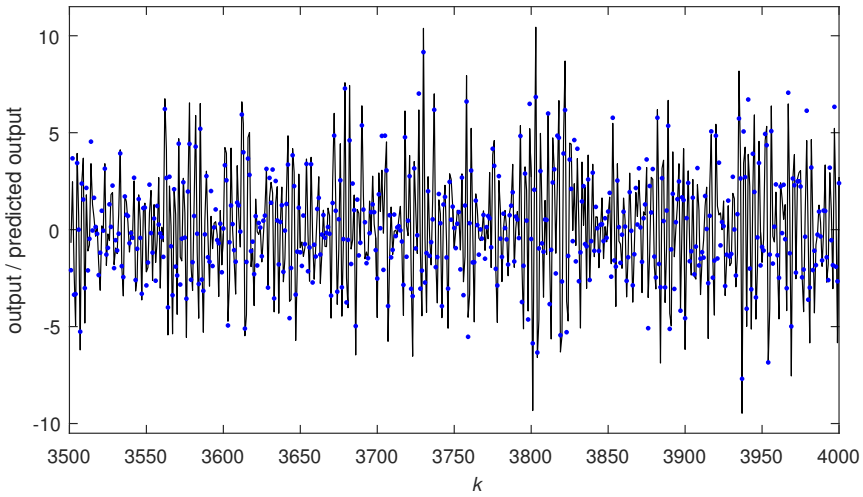


Figure 5. The AM-MIFSG estimation error δ versus k with $\alpha = 0.80$, 0.90 and 0.92 ($p = 6$).



Solid line: the true output y_k, dots: the predicted output $\hat{y}_k$.

Figure 6. The predicted output $\hat{y}_k$ and true output y_k from k = 3501 to 3700.



Solid line: the true output y_k, dots: the predicted output $\hat{y}_k$.

Figure 7. The predicted output $\hat{y}_k$ and true output y_k from k = 3501 to 4000.

7. Conclusions

This paper derives an AM-MIFSG estimation algorithm for Hammerstein output-error systems based on the key-term separation principle and auxiliary model identification idea. By means of the key-term separation principle, all the parameters in the linear and nonlinear blocks are separated, and the unknown variables in the identification model are replaced by the outputs of the auxiliary models. The analysis of the simulation results shows that the proposed algorithm obtains better parameter estimation performance than the AM-MISG algorithm. However, there also exist many topics that need to be further discussed. For example, is this algorithm still effective for systems with missing data? And is the performance of the algorithm can be improved by introducing a time-varying differential order α? These topics remain as open problems for future studies.

Author Contributions: Conceptualization, C.X.; methodology, Y.M.; software, C.X.; validation, C.X.; formal analysis, C.X.; investigation, C.X.; resources, Y.M.; data curation, Y.M.; writing-original draft preparation, Y.M.; writing—review and editing, C.X.; visualization, C.X.; supervision, Y.M.; project administration, Y.M.; funding acquisition, C.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (No. 62103167), and in part by the Natural Science Foundation of Jiangsu Province (No. BK20210451), and the research project of Jiangnan University (Nos. JUSRP12028 and JUSRP12040).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Wang, L.J.; Guo, J.; Xu, C.; Wu, T.Z.; Lin, H.P. Hybrid model predictive control strategy of supercapacitor energy storage system based on double active bridge. *Energies* **2019**, *12*, 2134. [CrossRef]
2. Zhang, Y.; Yan, Z.; Zhou, C.C.; Wu, T.Z.; Wang, Y.Y. Capacity allocation of HESS in micro-grid based on ABC algorithm. *Int. J. Low-Carbon Technol.* **2020**, *15*, 496-505. [CrossRef]
3. Chen, H.T.; Jiang, B.; Chen, W.; Yi, H. Data-driven detection and diagnosis of incipient faults in electrical drives of high-speed trains. *IEEE Trans. Ind. Electron.* **2019**, *66*, 4716-4725. [CrossRef]
4. Chen, H.T.; Jiang, B.; Ding, S.X.; Huang, B. Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives. *IEEE Trans. Intell. Transp. Syst.* **2020**. [CrossRef]

5. 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]
6. Ding, F., Liu, Y.J.; Bao, B. Gradient based and least squares based iterative estimation algorithms for multi-input multi-output systems. *Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng.* **2012**, *226*, 43–55. [CrossRef]
7. Xu, L.; Song, G.L. A recursive parameter estimation algorithm for modeling signals with multi-frequencies. *Circuits Syst. Signal Process.* **2020**, *39*, 4198–4224. [CrossRef]
8. 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]
9. Zhang, X.; Yang, E.F. Highly computationally efficient state filter based on the delta operator. *Int. J. Adapt. Control Signal Process.* **2019**, *33*, 875–889. [CrossRef]
10. Cuevas, E.; Díaz, P.; Avalos, O.; Zaldivar, D.; Pérez-Cisneros, M. Nonlinear system identification based on ANFIS-Hammerstein model using Gravitational search algorithm. *Appl. Intell.* **2018**, *48*, 182–203. [CrossRef]
11. Mukhopadhyay, S.; Mukherjee, A. ImdLMS: An imputation based LMS algorithm for linear system identification with missing input data. *IEEE Trans. Signal Process.* **2020**, *68*, 2370–2385. [CrossRef]
12. Sepulveda, N. E.; Sinha, J. Mathematical validation of experimentally optimised parameters used in a vibration-based machine-learning model for fault diagnosis in rotating machines. *Machines* **2021**, *9*, 155. [CrossRef]
13. Zhao, S.Y.; Yuriy, S.; Ahn, C.; Zhao, C.H. Probabilistic monitoring of correlated sensors for nonlinear processes in state space. *IEEE Trans. Ind. Electron.* **2020**, *67*, 2294–2303. [CrossRef]
14. Du, J.; Zhang, L.; Chen, J.; Li, J.; Jiang, X.; Zhu, C. Self-adjusted decomposition for multi-model predictive control of Hammerstein systems based on included angle. *ISA Trans.* **2020**, *103*, 19–27. [CrossRef] [PubMed]
15. Chen, H.T.; Jiang, B. A review of fault detection and diagnosis for the traction system in high-speed trains. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 450–465. [CrossRef]
16. Yang, M.; Wang, J.; Zhang, Y. Fault detection and diagnosis for plasticizing process of single-base gun propellant using mutual information weighted MPCA under limited batch samples modelling. *Machines* **2021**, *9*, 166. [CrossRef]
17. Chandra, S.; Hayashibe, M.; Thondiyath, A. Muscle fatigue induced hand tremor clustering in dynamic laparoscopic manipulation. *IEEE Trans. Syst. Man Cybern. -Syst.* **2020**, *50*, 5420–5431. [CrossRef]
18. Jalaleddini, K.; Kearney, R. E. Subspace identification of SISO Hammerstein systems: application to stretch reflex identification. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 2725–2734. [CrossRef]
19. Ding, F.; Chen, H.; Xu, L.; Dai, J.; Li, Q.; 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]
20. Ding, J.; Cao, Z.; Chen, J.; Jiang, G. Weighted parameter estimation for Hammerstein nonlinear ARX systems. *Circuits Syst. Signal Process.* **2020**, *39*, 2178–2192. [CrossRef]
21. Kazemi, M.; Arefi, M. M. A fast iterative recursive least squares algorithm for Wiener model identification of highly nonlinear systems. *ISA Trans.* **2017**, *67*, 382–388. [CrossRef]
22. Hou, J.; Liu, T.; Wang, Q. G. Subspace identification of Hammerstein-type nonlinear systems subject to unknown periodic disturbance. *Int. J. Control* **2021**, *94*, 849–859. [CrossRef]
23. Wang, L.; Ji, Y.; Wan, L.; Bu, N. Hierarchical recursive generalized extended least squares estimation algorithms for a class of nonlinear stochastic systems with colored noise. *J. Frankl. Inst.* **2019**, *356*, 10102–10122. [CrossRef]
24. Wan, L.; 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]
25. Loizou, N.; Richtárik, P. Momentum and stochastic momentum for stochastic gradient, newton, proximal point and subspace descent methods. *Comput. Optim. Appl.* **2020**, *77*, 653–710. [CrossRef]
26. Pan, J.; Jiang, X.; Wan, X.; 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]
27. Ma, H.; Pan, J.; Lv, L.. Recursive algorithms for multivariable output-error-like ARMA systems. *Mathematics* **2019**, *7*, 558. [CrossRef]
28. Pan, J.; Ma, H.; Zhang, X. Recursive coupled projection algorithms for multivariable output-error-like systems with coloured noises. *IET Signal Process.* **2020**, *14*, 455–466. [CrossRef]
29. Ma, H.; Zhang, X.; Liu, Q.Y. Hayat, T. Partially-coupled gradient-based iterative algorithms for multivariable output-error-like systems with autoregressive moving average noises. *IET Contr. Theory Appl.* **2020**, *14*, 2613–2627. [CrossRef]
30. Khan, S.; Ahmad, J.; Naseem, I.; Moinuddin, M. A novel fractional gradient-based learning algorithm for recurrent neural networks. *Circuits Syst. Signal Process.* **2018**, *37*, 593–612. [CrossRef]
31. Ding, F.; Chen, T. Performance analysis of multi-innovation gradient type identification methods. *Automatica* **2007**, *43*, 1–14. [CrossRef]
32. Chaudhary, N. I.; Raja, M. A. Z. Identification of Hammerstein nonlinear ARMAX systems using nonlinear adaptive algorithms. *Nonlinear Dyn.* **2015**, *79*, 1385–1397. [CrossRef]
33. Chaudhary, N. I.; Raja, M. A. Z. Design of fractional adaptive strategy for input nonlinear Box-Jenkins systems. *Signal Process.* **2015**, *116*, 141–151. [CrossRef]
34. Wang, Y.; Li, M.; Chen, Z. Experimental study of fractional-order models for lithium-ion battery and ultra-capacitor: Modeling, system identification, and validation. *Appl. Energy* **2020**, *278*, 115736. [CrossRef]

35. Aslam, M. S.; Chaudhary, N. I.; Raja, M. A. Z. A sliding-window approximation-based fractional adaptive strategy for Hammerstein nonlinear ARMAX systems. *Nonlinear Dyn.* **2017**, *87*, 519–533. [CrossRef]
36. Chaudhary, N. I.; Raja, M. A. Z.; Khan, A. U. R. Design of modified fractional adaptive strategies for Hammerstein nonlinear control autoregressive systems. *Nonlinear Dyn.* **2015**, *82*, 1811–1830. [CrossRef]
37. Cheng, S.; Wei, Y.; Sheng, D.; Chen, Y.; Wang, Y. Identification for Hammerstein nonlinear ARMAX systems based on multi-innovation fractional order stochastic gradient. *Signal Process.* **2018**, *142*, 1–10. [CrossRef]
38. Chaudhary, N. I.; Raja, M. A. Z.; He, Y.; Khan, Z. A.; Machado, J. T. Design of multi innovation fractional LMS algorithm for parameter estimation of input nonlinear control autoregressive systems. *Appl. Math. Model.* **2021**, *93*, 412–425. [CrossRef]
39. Ding, F.; Shi, Y.; Chen, T. Auxiliary model-based least-squares identification methods for Hammerstein output-error systems. *Syst. Control Lett.* **2007**, *56*, 373–380. [CrossRef]
40. Zhang, Q. Nonlinear system identification with output error model through stabilized simulation. *IFAC Proc. Vol.* **2004**, *37*, 501–506. [CrossRef]
41. Vörös, J. Parameter identification of discontinuous Hammerstein systems. *Automatica* **1997**, *33*, 1141–1146. [CrossRef]
42. Vörös, J. Recursive identification of Hammerstein systems with discontinuous nonlinearities containing dead-zones. *IEEE Trans. Autom. Control* **2003**, *48*, 2203–2206. [CrossRef]
43. Vörös, J. Identification of nonlinear cascade systems with output hysteresis based on the key term separation principle. *Appl. Math. Model.* **2015**, *39*, 5531–5539. [CrossRef]
44. Mao, Y.; Ding, F.; Yang, E. Adaptive filtering based multi-innovation gradient algorithm for input nonlinear systems with autoregressive noise. *Int. J. Adapt. Control Signal Process.* **2017**, *31*, 1388–1400. [CrossRef]
45. Liu, Y.; Yu, L.; Ding, F. Multi-innovation extended stochastic gradient algorithm and its performance analysis. *Circuits Syst. Signal Process.* **2010**, *29*, 649–667. [CrossRef]
46. Shah, S. M. Riemann-Liouville operator-based fractional normalised least mean square algorithm with application to decision feedback equalisation of multipath channels. *IET Signal Process.* **2016**, *10*, 575–582. [CrossRef]
47. Li, C.; Qian, D.; Chen, Y. On Riemann-Liouville and caputo derivatives. In *Discrete Dynamics in Nature and Society*; Hindawi Limited: London, UK, 2011.
48. Li, M.H.; Liu, X.M. Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique. *Int. J. Control Autom. Syst.* **2020**, *18*, 1581–1592. [CrossRef]
49. Ding, F.; Xu, L.; Meng, D.D. Gradient estimation algorithms for the parameter identification of bilinear systems using the auxiliary model. *J. Comput. Appl. Math.* **2020**, *369*, 112575. [CrossRef]
50. Li, M.H.; Liu, X.M. Maximum likelihood hierarchical least squares-based iterative identification for dual-rate stochastic systems. *Int. J. Adapt. Control Signal Process.* **2021**, *35*, 240–261. [CrossRef]
51. 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]
52. Ji, Y.; Kang, Z.; Zhang, C. Two-stage gradient-based recursive estimation for nonlinear models by using the data filtering. *Int. J. Control Autom. Syst.* **2021**, *19*, 2706–2715. [CrossRef]
53. Xu, L.; Yang,E.F. Auxiliary model multiinnovation stochastic gradient parameter estimation methods for nonlinear sandwich systems. *Int. J. Robust Nonlinear Control* **2021**, *31*, 148–165. [CrossRef]
54. Wang JW, Ji Y, Zhang C. Iterative parameter and order identification for fractional-order nonlinear finite impulse response systems using the key term separation. *Int. J. Adapt. Control Signal Process.* **2021**, *35*, 1562–1577. [CrossRef]
55. Zhang, X. Adaptive parameter estimation for a general dynamical system with unknown states. *Int. J. Robust Nonlinear Control* **2020**, *30*, 1351–1372. [CrossRef]
56. Zhang, X. Recursive parameter estimation methods and convergence analysis for a special class of nonlinear systems. *Int. J. Robust Nonlinear Control* **2020**, *30*, 1373–1393. [CrossRef]
57. Mao, Y.W.; Liu, S,; Liu, J.F. Robust economic model predictive control of nonlinear networked control systems with communication delays. *Int. J. Adapt. Control Signal Process.* **2020**, *34*, 614–637. [CrossRef]
58. Ding, F. State filtering and parameter estimation for state space systems with scarce measurements. *Signal Process.* **2014**, *104*, 369–380. [CrossRef]
59. Ding, F. Combined state and least squares parameter estimation algorithms for dynamic systems. *Appl. Math. Modell.* **2014**, *38*, 403–412. [CrossRef]
60. Li, M.H.; Liu, X.M. Iterative identification methods for a class of bilinear systems by using the particle filtering technique. *Int. J. Adapt. Control Signal Process.* **2021**, *35*, 2056–2074. [CrossRef]
61. Zhao, S.Y.; Huang, B.; Liu F. Linear optimal unbiased filter for time-variant systems without apriori information on initial conditions. *IEEE Trans. Autom. Control* **2017**, *62*, 882–887. [CrossRef]
62. Liu, Y.J.; Shi, Y. An efficient hierarchical identification method for general dual-rate sampled-data systems. *Automatica* **2014**, *50*, 962–970. [CrossRef]
63. Ding, F.; Qiu, L.; Chen, T. Reconstruction of continuous-time systems from their non-uniformly sampled discrete-time systems. *Automatica* **2009**, *45*, 324–332. [CrossRef]
64. Zhao, S.Y.; Yuriy, S.; Ahn, C.; Liu F. Adaptive-horizon iterative UFIR filtering algorithm with applications. *IEEE Trans. Ind. Electron.* **2018**, *65*, 6393–6402. [CrossRef]

65. Zhao, S.Y.; Huang, B. Trial-and-error or avoiding a guess? Initialization of the Kalman filter. *Automatica* **2020**, *121*, 109184. [CrossRef]
66. Ding, F.; Liu, G.; Liu, X.P. Parameter estimation with scarce measurements. *Automatica* **2011**, *47*, 1646–1655. [CrossRef]