INTERNATIONAL JOURNAL OF ROBUST AND NONLINEAR CONTROL
Int. J. Robust Nonlinear Control 2023; **00**:1–15
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/acs

Hierarchical recursive gradient identification of Hammerstein nonlinear systems based on the key term separation

Lei Lv[1], Wei Sun[1]*and Jian Pan[2]

[1]*School of Aerospace Science and Technology, Xidian University, Xi'an 710126, China*
[2]*School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China*

SUMMARY

This article explores recursive algorithms for parameter identification issues of Hammerstein output-error systems. The proposed approach includes the key term separation auxiliary model recursive gradient algorithm, which utilizes the gradient search and the key term separation. To enhance computational efficiency, the system is decomposed into two or three subsystems through the hierarchical identification principle. Based on this, a key term separation auxiliary model two-stage recursive gradient algorithm and a key term separation auxiliary model three-stage recursive gradient algorithm are presented. The simulation results verify the validity of the obtained algorithms. Copyright © 2023 John Wiley & Sons, Ltd.

Received ...

KEY WORDS: Nonlinear system, parameter estimation, hierarchical identification, key term separation, gradient search

1. INTRODUCTION

In the realm of control systems, the establishment of accurate mathematical models that describe the essential features of the system is essential for its analysis and optimization [1, 2, 3]. Therefore, the foremost challenge is to overcome the hurdle of system modeling. System identification techniques that rely on measured data provide an efficient approach to establishing a mathematical model of the system [4, 5, 6, 7]. The estimation of system parameters is a crucial prerequisite for effective system identification and thus has been the subject of extensive scrutiny in the field of control engineering. Over the years, numerous parameter estimation methods have been proposed by researchers, contributing to the evolution and comprehensiveness of the theories of system identification [8, 9, 10].

Nonlinear systems are different from linear systems in that their output variables are not linearly related to their input variables [11, 12, 13, 14]. Block-oriented nonlinear systems are a typical class of nonlinear systems, characterized by splitting the dynamic nonlinear system into a dynamic linear subsystem and a static nonlinear subsystem [15, 16, 17, 18]. Hammerstein systems, also known as input nonlinear systems, are composed of a static nonlinear link in series with a linear dynamic subsystem and are widely used to describe dynamic systems with nonlinear input properties [19, 20, 21, 22]. In the literature of Hammerstein system identification, Kang et al. applied the key term separation technique for Hammerstein nonlinear autoregressive output-error systems

*Correspondence to: School of Aerospace Science and Technology, Xidian University, Xi'an 710126, China.
†Email: wsun@xidian.edu.cn

utilizing the hierarchical least squares method [23]; Wang et al. investigated a novel expectation maximization estimation method for Hammerstein systems with data loss [24].

Recursive methods are widely used in system identification and are capable of capturing real-time information about the systems by collecting dynamic data [25, 26, 27, 28]. The hierarchical identification principle achieves the purpose of reducing the dimensionality of the parameter vector by decomposing the system model into several virtual subsystems and identifying their parameters separately [29, 30, 31, 32]. Recently, Wang et al. [33] reconstructed the bilinear Hammerstein system into two virtual subsystems utilizing the hierarchical identification principle and investigated the estimation method of the system. The key term separation technique, which is important for studying the identification of nonlinear systems [34, 35], usually selects the output of the nonlinear block as the key term and separates the parameters of the nonlinear part, thereby realizing the simultaneous identification of the parameters of the linear part and the nonlinear part.

This article considers the parameter estimation problem of Hammerstein output-error (H-OE) systems. By using the key term separation technique [36, 37], a key term separation auxiliary model recursive gradient (KT-AM-RG) algorithm is proposed. In order to improve the model performance, we employ the hierarchical identification principle to decompose the key term separation identification model into two and three sub-models, and develop a key term separation auxiliary model two-stage recursive gradient (KT-AM-2S-RG) algorithm and a key term separation auxiliary model three-stage recursive gradient (KT-AM-3S-RG) algorithm.

The paper is organized as follows. Section 2 introduces some definitions and establishes the key term separation identification model of the H-OE system. Section 3 proposes a KT-AM-RG algorithm, a KT-AM-2S-RG algorithm, and a KT-AM-3S-RG algorithm. How to analyze and compare the computational efficiency of the proposed algorithms can be found in Section 4. Section 5 offers the simulation results to corroborate the obtained algorithms, and Section 6 summarize the research content of this paper.

Notations: Let "$A =: X$" or "$X := A$" represent "A is defined as X"; $\boldsymbol{I}_n$ denotes an identity matrix with proper size; $\mathbf{1}_n$ stands for an n-dimensional column vector whose elements are 1. Take z as the unit forward shift operator $[zy(t) = y(t+1),\ z^{-1}y(t) = y(t-1)]$; $\lambda_{\max}$ denotes the maximum eigenvalue of matrix. The superscript T represents the transpose of a matrix/vector; the norm of a matrix $\boldsymbol{X}$ is defined by $\|\boldsymbol{X}\|^2 := \mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$, in which $\boldsymbol{X} \in \mathbb{R}^{n\times n}$.

2. THE SYSTEM DESCRIPTION AND IDENTIFICATION MODEL

As shown in Figure 1, the H-OE system is expressed as

$$y(t) = \frac{B(z)}{A(z)}\bar{u}(t) + v(t). \tag{1}$$

The variable $y(t) \in \mathbb{R}$ and $u(t) \in \mathbb{R}$ are the output and input respectively. $v(t) \in \mathbb{R}$ is stochastic white noise with zero mean, $A(z)$ and $B(z)$ are the polynomials of known orders n_a and n_b, which can be defined as

$$A(z) := 1 + a_1z^{-1} + a_2z^{-2} + \cdots + a_{n_a}z^{-n_a}, \quad a_i \in \mathbb{R},$$
$$B(z) := b_0 + b_1z^{-1} + b_2z^{-2} + \cdots + b_{n_b}z^{-n_b}, \quad b_i \in \mathbb{R},$$

$\bar{u}(t) \in \mathbb{R}$ is the output of the nonlinear part and is a linear combination of a set of known basis functions $f_j(u(t))$ with parameters c_j's, that is

$$\bar{u}(t) = f(u(t)) = c_1f_1(u(t)) + c_2f_2(u(t)) + \cdots + c_mf_m(u(t)), \tag{2}$$

Suppose that the order m is known and $y(t) = 0$, $u(t) = 0$, and $v(t) = 0$ for $t \leqslant 0$.

Observing the H-OE systems, it can be seen that $\bar{u}(t)$ is a latent variable and any pair of $(\alpha f(*))$, $B(z)/\alpha)$ and $(f(*), B(z))$ with nonzero constant α will generate identical input and output relationship for the model in (1). To ensure identifiability, it is required to normalize the parameter of $\bar{u}(t)$ or $B(z)$. Here, we let $b_0 = 1$.

DOI: 10.1002/acs

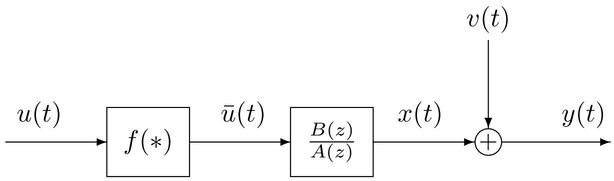

Figure 1. The H-OE system.

The noise-free output of the system in Figure 1 is given by

$$x(t) := \frac{B(z)}{A(z)}\bar{u}(t) \in \mathbb{R}. \tag{3}$$

Choosing $\bar{u}(t)$ as the key term to parameterize the noise-free output,

$$\begin{aligned} x(t) &= -\sum_{i=1}^{n_a} a_i x(t-i) + \sum_{i=1}^{n_b} b_i \bar{u}(t-i) + b_0\bar{u}(t) \\ &= -\sum_{i=1}^{n_a} a_i x(t-i) + \sum_{i=1}^{n_b} b_i \bar{u}(t-i) + \sum_{j=1}^{m} c_j f_j(u(t)). \end{aligned} \tag{4}$$

Define the sub-parameter vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$,

$$\begin{aligned} \boldsymbol{a} &:= [a_1, a_2, a_3, \cdots, a_{n_a}]^{\mathrm{T}} \in \mathbb{R}^{n_a}, \\ \boldsymbol{b} &:= [b_1, b_2, b_3, \cdots, b_{n_b}]^{\mathrm{T}} \in \mathbb{R}^{n_b}, \\ \boldsymbol{c} &:= [c_1, c_2, c_3, \cdots, c_{n_c}]^{\mathrm{T}} \in \mathbb{R}^{m}, \end{aligned}$$

and the sub-information vectors $\boldsymbol{\varphi}_a(t)$, $\boldsymbol{\varphi}_b(t)$ and $\boldsymbol{f}(t)$:

$$\begin{aligned} \boldsymbol{\varphi}_a(t) &:= [-x(t-1), -x(t-2), -x(t-3), \cdots, -x(t-n_a)]^{\mathrm{T}} \in \mathbb{R}^{n_a}, \\ \boldsymbol{\varphi}_b(t) &:= [\bar{u}(t-1), \bar{u}(t-2), \bar{u}(t-3), \cdots, \bar{u}(t-n_b)]^{\mathrm{T}} \in \mathbb{R}^{n_b}, \\ \boldsymbol{f}(t) &:= [f_1(u(t)), f_2(u(t)), f_3(u(t)), \cdots, f_m(u(t))]^{\mathrm{T}} \in \mathbb{R}^{m}. \end{aligned}$$

Then, equation (1)–(3) can be reformulated as:

$$\bar{u}(t) = \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{c}, \tag{5}$$

$$x(t) = \boldsymbol{\varphi}_a^{\mathrm{T}}(t)\boldsymbol{a} + \boldsymbol{\varphi}_b^{\mathrm{T}}(t)\boldsymbol{b} + \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{c}, \tag{6}$$

$$\begin{aligned} y(t) &= x(t) + v(t) \\ &= \boldsymbol{\varphi}_a^{\mathrm{T}}(t)\boldsymbol{a} + \boldsymbol{\varphi}_b^{\mathrm{T}}(t)\boldsymbol{b} + \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{c} + v(t). \end{aligned} \tag{7}$$

Let $n := n_a + n_b + m$, define the parameter vector $\boldsymbol{\vartheta}$ and the information vector $\boldsymbol{\varphi}(t)$ as

$$\begin{aligned} \boldsymbol{\vartheta} &:= [\boldsymbol{a}^{\mathrm{T}}, \boldsymbol{b}^{\mathrm{T}}, \boldsymbol{c}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n}. \\ \boldsymbol{\varphi}(t) &:= [\boldsymbol{\varphi}_a^{\mathrm{T}}(t), \boldsymbol{\varphi}_b^{\mathrm{T}}(t), \boldsymbol{f}^{\mathrm{T}}(t)]^{\mathrm{T}} \in \mathbb{R}^{n}. \end{aligned}$$

System (1) can be rewritten as

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\vartheta} + v(t). \tag{8}$$

Hence, one can obtain the key term separation identification model in (8) of the H-OE system in (1)–(2), where parameter vector $\boldsymbol{\vartheta}$ to be identified contains all parameters of the original system.

DOI: 10.1002/acs

3. THE KEY TERM SEPARATION AUXILIARY MODEL RECURSIVE GRADIENT ALGORITHMS

3.1. KT-AM-RG algorithm

For the identification model in (8), defining the criterion function:

$$J_1(\boldsymbol{\vartheta}) := \frac{1}{2}\sum_{j=1}^{t}[y(j) - \boldsymbol{\varphi}^{\mathrm{T}}(j)\boldsymbol{\vartheta}]^2.$$

Define the stacked output vector $\boldsymbol{Y}(t)$, and the stacked information matrix $\boldsymbol{\Phi}(t)$ as

$$\begin{aligned}
&\boldsymbol{Y}(t) := [y(1), y(2), \cdots, y(t)]^{\mathrm{T}} \in \mathbb{R}^{t},\\
&\boldsymbol{\Phi}(t) := [\boldsymbol{\varphi}(1), \boldsymbol{\varphi}(2), \cdots, \boldsymbol{\varphi}(t)]^{\mathrm{T}} \in \mathbb{R}^{t\times n}.
\end{aligned}$$

Then the criterion function $J_1(\boldsymbol{\vartheta})$ can be expressed as

$$J_1(\boldsymbol{\vartheta}) = \frac{1}{2}\|\boldsymbol{Y}(t) - \boldsymbol{\Phi}(t)\boldsymbol{\vartheta}\|^2.$$

To better illustrate the identification problem, some useful recursive relations are defined:

$$\boldsymbol{\xi}(t) := \boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{Y}(t) = \boldsymbol{\xi}(t-1) + \boldsymbol{\varphi}(t)y(t) \in \mathbb{R}^{n}, \tag{9}$$

$$\boldsymbol{R}(t) := \boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{\Phi}(t) = \boldsymbol{R}(t-1) + \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) \in \mathbb{R}^{n\times n}. \tag{10}$$

The gradient vector of the criterion function $J_1(\boldsymbol{\vartheta})$ can be obtained by taking the first-order partial derivative of $J_1(\boldsymbol{\vartheta})$ with respect to the parameter vector $\boldsymbol{\vartheta}$

$$\mathrm{grad}[J_1(\boldsymbol{\vartheta})] := \frac{\partial J_1(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = -\boldsymbol{\Phi}^{\mathrm{T}}(t)[\boldsymbol{Y}(t) - \boldsymbol{\Phi}(t)\boldsymbol{\vartheta}] \in \mathbb{R}^{n}. \tag{11}$$

Let $\hat{\boldsymbol{\vartheta}}(t)$ be the estimate of $\boldsymbol{\vartheta}$ at time t. The gradient vector at $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}(t-1)$ is

$$\begin{aligned}
\mathrm{grad}[J_1(\hat{\boldsymbol{\vartheta}}(t-1))] &= \mathrm{grad}[J_1(\boldsymbol{\vartheta})]\big|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}(t-1)}\\
&= -\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{Y}(t) + \boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{\Phi}(t)\hat{\boldsymbol{\vartheta}}(t-1)\\
&= -[\boldsymbol{\xi}(t) - \boldsymbol{R}(t)\hat{\boldsymbol{\vartheta}}(t-1)].
\end{aligned} \tag{12}$$

Minimizing $J_1(\boldsymbol{\vartheta})$ by using the negative gradient search,

$$\begin{aligned}
\hat{\boldsymbol{\vartheta}}(t) &= \hat{\boldsymbol{\vartheta}}(t-1) - \mu(t)\mathrm{grad}[J_1(\hat{\boldsymbol{\vartheta}}(t-1))]\\
&= \hat{\boldsymbol{\vartheta}}(t-1) + \mu(t)[\boldsymbol{\xi}(t) - \boldsymbol{R}(t)\hat{\boldsymbol{\vartheta}}(t-1)]\\
&= [\boldsymbol{I}_n - \mu(t)\boldsymbol{R}(t)]\hat{\boldsymbol{\vartheta}}(t-1) + \mu(t)\boldsymbol{\xi}(t),
\end{aligned} \tag{13}$$

where $\mu(t) \leqslant 0$ is a step-size. To ensure the convergence of $\hat{\boldsymbol{\vartheta}}(t)$, all the eigenvalues of matrix $[\boldsymbol{I}_n - \mu(t)\boldsymbol{R}(t)]$ must be in the unit circle, this is to say $\mu(t)$ should satisfy $-\boldsymbol{I}_n \leqslant \boldsymbol{I}_n - \mu(t)\boldsymbol{R}(t) \leqslant \boldsymbol{I}_n$, so a reasonable choice of $\mu(t)$ is to satisfy

$$\mu(t) \leqslant \frac{2}{\lambda_{\max}[\boldsymbol{R}(t)]} = 2\lambda_{\max}^{-1}[\boldsymbol{R}(t)]. \tag{14}$$

To reduce the complexity of the calculation, an alternative way for $\mu(t)$ is to take

$$\mu(t) \leqslant \frac{2}{\mathrm{tr}[\boldsymbol{R}(t)]} = 2\{\mathrm{tr}[\boldsymbol{R}(t)]\}^{-1}. \tag{15}$$

Let $\mu(t) := \frac{1}{r(t)}$, $r(t) := \mathrm{tr}[\boldsymbol{R}(t)] = r(t-1) + \|\boldsymbol{\varphi}(t)\|^2$, $r(0) = 1$. Combining (9)–(15) constructs the following recursive relationships:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{1}{r(t)}[\boldsymbol{\xi}(t) - \boldsymbol{R}(t)\hat{\boldsymbol{\vartheta}}(t-1)]. \tag{16}$$

DOI: 10.1002/acs

Equation (16) is unable to determine the parameter estimate $\hat{\boldsymbol{\vartheta}}(t)$ because the information vector $\boldsymbol{\varphi}(t)$ includes the unknown entries such as $x(t-i)$ and $\bar{u}(t-i)$. To overcome this problem, an auxiliary model is constructed to improve the parameter estimation algorithm and address the issue of unmeasured intermediate variables. Specifically, based on equation (6), the following auxiliary model is constructed:

$$\begin{aligned} x_a(t) &:= \hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t) + \hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t) + \boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t) \\ &= \hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t), \end{aligned} \tag{17}$$

$$\bar{u}_a(t) := \boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t). \tag{18}$$

where $\hat{\boldsymbol{a}}(t)$, $\hat{\boldsymbol{b}}(t)$ and $\hat{\boldsymbol{c}}(t)$ are the parameter estimates of $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ at time t, respectively. $\hat{\boldsymbol{\varphi}}_a(t) := [-x_a(t-1), -x_a(t-2), \cdots, -x_a(t-n_a)]^{\mathrm{T}} \in \mathbb{R}^{n_a}$, $\hat{\boldsymbol{\varphi}}_b(t) := [\bar{u}_a(t-1), \bar{u}_a(t-2), \cdots, \bar{u}_a(t-n_b)]^{\mathrm{T}} \in \mathbb{R}^{n_b}$, and $\hat{\boldsymbol{\varphi}}(t) := [\hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t), \hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t), \boldsymbol{f}^{\mathrm{T}}(t)]^{\mathrm{T}} \in \mathbb{R}^{n}$.

From the above derivation, we can obtain the key term separation auxiliary model recursive gradient (KT-AM-RG) algorithm:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{1}{r(t)}[\boldsymbol{\xi}(t) - \boldsymbol{R}(t)\hat{\boldsymbol{\vartheta}}(t-1)], \tag{19}$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2, \tag{20}$$

$$\boldsymbol{\xi}(t) = \boldsymbol{\xi}(t-1) + \hat{\boldsymbol{\varphi}}(t)y(t) \in \mathbb{R}^n, \tag{21}$$

$$\boldsymbol{R}(t) = \boldsymbol{R}(t-1) + \hat{\boldsymbol{\varphi}}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t) \in \mathbb{R}^{n\times n}, \tag{22}$$

$$\hat{\boldsymbol{\varphi}}(t) = [\hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t), \hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t), \boldsymbol{f}^{\mathrm{T}}(t)]^{\mathrm{T}} \in \mathbb{R}^n, \quad n := n_a + n_b + m, \tag{23}$$

$$\hat{\boldsymbol{\varphi}}_a(t) = [-x_a(t-1), -x_a(t-2), \cdots, -x_a(t-n_a)]^{\mathrm{T}} \in \mathbb{R}^{n_a}, \tag{24}$$

$$\hat{\boldsymbol{\varphi}}_b(t) = [\bar{u}_a(t-1), \bar{u}_a(t-2), \cdots, \bar{u}_a(t-n_b)]^{\mathrm{T}} \in \mathbb{R}^{n_b}, \tag{25}$$

$$\boldsymbol{f}(t) = [f_1(u(t)), f_2(u(t)), \cdots, f_m(u(t))]^{\mathrm{T}} \in \mathbb{R}^m, \tag{26}$$

$$\bar{u}_a(t) = \boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t), \tag{27}$$

$$x_a(t) = \hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t) + \hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t) + \bar{u}_a(t), \tag{28}$$

$$\hat{\boldsymbol{\vartheta}}(t) = [\hat{\boldsymbol{a}}^{\mathrm{T}}(t), \hat{\boldsymbol{b}}^{\mathrm{T}}(t), \hat{\boldsymbol{c}}^{\mathrm{T}}(t)]^{\mathrm{T}}. \tag{29}$$

The KT-AM-RG algorithm involves the following steps:

1. Set all variables be zero when $t \leqslant 0$, and set the data length L and the basis function $f_j(*)$. Let $t=1$ and set the initial values: $\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_n/p_0$, $x_a(0) = 1/p_0$, $\bar{u}_a(0) = 1/p_0$, $\boldsymbol{\xi}(0) = \mathbf{1}_n/p_0$ and $\boldsymbol{R}(0) = \boldsymbol{I}_n/p_0$, $p_0 = 10^6$.
2. Collect the observation data $u(t)$ and $y(t)$, and construct the sub-information vector $\hat{\boldsymbol{\varphi}}_a(t)$, $\hat{\boldsymbol{\varphi}}_b(t)$ and $\boldsymbol{f}(t)$ using (24)–(26). From $\hat{\boldsymbol{\varphi}}(t)$ using (23).
3. Compute the step-size $r(t)$ by (20), the vector $\boldsymbol{\xi}(t)$ by (21), the matrix $\boldsymbol{R}(t)$ by (22).
4. Update the estimation vector $\hat{\boldsymbol{\vartheta}}(t)$ using (19) and extract $\hat{\boldsymbol{a}}(t)$, $\hat{\boldsymbol{b}}(t)$, and $\hat{\boldsymbol{c}}(t)$ from (29).
5. Compute the outputs $\bar{u}_a(t)$ and $x_a(t)$ of the auxiliary models by using (27)–(28).
6. Compare t with L: if $t \leqslant L$, increase t by 1 and go to Step 2; otherwise terminate this procedure, and output the estimation vector $\hat{\boldsymbol{\vartheta}}(L)$.

3.2. *KT-AM-2S-RG algorithm*

Next, to further reduce the computational burden, the model (8) is decomposed into two or three sub-models for identification using hierarchical identification principle.

According to Equations (7), one combines the linear partial parameter vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. Then model (8) can be equivalently expressed as

$$y(t) = \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\theta} + \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{c} + v(t), \tag{30}$$

$$\boldsymbol{\theta} := \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix} \in \mathbb{R}^{n_1}, \quad \boldsymbol{\psi}(t) := \begin{bmatrix} \boldsymbol{\varphi}_a(t) \\ \boldsymbol{\varphi}_b(t) \end{bmatrix} \in \mathbb{R}^{n_1}, \quad n_1 := n_a + n_b. \tag{31}$$

DOI: 10.1002/acs

For the above identification model (30), define two intermediate variables:

$$y_1(t) := y(t) - \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{c} \in \mathbb{R}, \tag{32}$$
$$y_2(t) := y(t) - \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\theta} \in \mathbb{R}. \tag{33}$$

Equation (30) can be decomposed into the following two fictitious subsystems:

$$y_1(t) = \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\theta} + v(t), \tag{34}$$
$$y_2(t) = \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{c} + v(t). \tag{35}$$

Thus, the key term separation two-stage identification model in (34)–(35) can be obtained. Define two criterion functions:

$$J_2(\boldsymbol{\theta}) := \frac{1}{2}\sum_{j=1}^{t}[y_1(j) - \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\theta}]^2,$$
$$J_3(\boldsymbol{c}) := \frac{1}{2}\sum_{j=1}^{t}[y_2(j) - \boldsymbol{f}^{\mathrm{T}}(j)\boldsymbol{c}]^2.$$

Define the stacked output vectors and the stacked information matrices as

$$\begin{aligned}
\boldsymbol{Y}_1(t) &:= [y_1(1), y_1(2), y_1(3), \cdots, y_1(t)]^{\mathrm{T}} \in \mathbb{R}^t,\\
\boldsymbol{Y}_2(t) &:= [y_2(1), y_2(2), y_2(3), \cdots, y_2(t)]^{\mathrm{T}} \in \mathbb{R}^t,\\
\boldsymbol{\Psi}(t) &:= [\boldsymbol{\psi}(1), \boldsymbol{\psi}(2), \boldsymbol{\psi}(3), \cdots, \boldsymbol{\psi}(t)]^{\mathrm{T}} \in \mathbb{R}^{t\times(n_1)},\\
\boldsymbol{F}(t) &:= [\boldsymbol{f}(1), \boldsymbol{f}(2), \boldsymbol{f}(3), \cdots, \boldsymbol{f}(t)]^{\mathrm{T}} \in \mathbb{R}^{t\times m}.
\end{aligned}$$

Hence, we have

$$J_2(\boldsymbol{\theta}) = \frac{1}{2}\|\boldsymbol{Y}_1(t) - \boldsymbol{\Psi}(t)\boldsymbol{\theta}\|^2,$$
$$J_3(\boldsymbol{c}) = \frac{1}{2}\|\boldsymbol{Y}_2(t) - \boldsymbol{F}(t)\boldsymbol{c}\|^2.$$

Define the following recursive relationship:

$$\boldsymbol{\xi}_1(t) := \boldsymbol{\Psi}^{\mathrm{T}}(t)\boldsymbol{Y}_1(t) = \boldsymbol{\xi}_1(t-1) + \boldsymbol{\psi}(t)y_1(t) \in \mathbb{R}^{n_1}, \tag{36}$$
$$\boldsymbol{R}_1(t) := \boldsymbol{\Psi}^{\mathrm{T}}(t)\boldsymbol{\Psi}(t) = \boldsymbol{R}_1(t-1) + \boldsymbol{\psi}(t)\boldsymbol{\psi}^{\mathrm{T}}(t) \in \mathbb{R}^{n_1\times n_1}, \tag{37}$$
$$\boldsymbol{\xi}_2(t) := \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{Y}_2(t) = \boldsymbol{\xi}_2(t-1) + \boldsymbol{f}(t)y_2(t) \in \mathbb{R}^{m}, \tag{38}$$
$$\boldsymbol{R}_2(t) := \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{F}(t) = \boldsymbol{R}_2(t-1) + \boldsymbol{f}(t)\boldsymbol{f}^{\mathrm{T}}(t) \in \mathbb{R}^{m\times m}. \tag{39}$$

Let $\hat{\boldsymbol{\theta}}(t) \in \mathbb{R}^{n_a+n_b}$ and $\hat{\boldsymbol{c}}(t) \in \mathbb{R}^m$ be the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{c}$ at time t. Minimizing $J_2(\boldsymbol{\theta})$ and $J_3(\boldsymbol{c})$ lead to the following recursive relations:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) - \frac{1}{r_1(t)}\mathrm{grad}[J_2(\hat{\boldsymbol{\theta}}(t-1))]\\
&= \hat{\boldsymbol{\theta}}(t-1) + \frac{1}{r_1(t)}[\boldsymbol{\xi}_1(t) - \boldsymbol{R}_1(t)\hat{\boldsymbol{\theta}}(t-1)],
\end{aligned} \tag{40}$$
$$r_1(t) := r_1(t-1) + \|\boldsymbol{\psi}(t)\|^2, \quad r_1(0) = 1, \tag{41}$$
$$\begin{aligned}
\hat{\boldsymbol{c}}(t) &= \hat{\boldsymbol{c}}(t-1) - \frac{1}{r_2(t)}\mathrm{grad}[J_3(\hat{\boldsymbol{c}}(t-1))]\\
&= \hat{\boldsymbol{c}}(t-1) + \frac{1}{r_2(t)}[\boldsymbol{\xi}_2(t) - \boldsymbol{R}_2(t)\hat{\boldsymbol{c}}(t-1)],
\end{aligned} \tag{42}$$
$$r_2(t) := r_2(t-1) + \|\boldsymbol{f}(t)\|^2. \quad r_2(0) = 1. \tag{43}$$

Unluckily, equations (40)–(43) cannot calculate the estimates $\hat{\boldsymbol{\theta}}(t)$ and $\hat{\boldsymbol{c}}(t)$, as these equations involve the unmeasured terms $x(t-i)$ and $\bar{u}(t-i)$. We adopt the auxiliary model identification idea

DOI: 10.1002/acs

to solve this problem. Replacing these unknown terms with their corresponding estimates $x_a(t-i)$ and $\bar{u}_a(t-i)$ based on the auxiliary model. From (17), we can compute the output $x_a(t)$ of the auxiliary model by $x_a(t)=\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t)+\boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t)$, $\hat{\boldsymbol{\psi}}(t):=[\hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t),\hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t)]^{\mathrm{T}}\in\mathbb{R}^{n_a+n_b}$.

Substituting the vector $\boldsymbol{\psi}(t)$ in (40)–(43) with its estimate $\hat{\boldsymbol{\psi}}(t)$, we can derive a key term separation auxiliary model two-stage recursive gradient (KT-AM-2S-RG) algorithm:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{1}{r_1(t)}[\boldsymbol{\xi}_1(t)-\boldsymbol{R}_1(t)\hat{\boldsymbol{\theta}}(t-1)], \tag{44}$$

$$r_1(t)=r_1(t-1)+\|\hat{\boldsymbol{\psi}}(t)\|^2, \tag{45}$$

$$\begin{aligned}\boldsymbol{\xi}_1(t)&=\boldsymbol{\xi}_1(t-1)+\hat{\boldsymbol{\psi}}(t)y_1(t)\\&=\boldsymbol{\xi}_1(t-1)+\hat{\boldsymbol{\psi}}(t)[y(t)-\boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t-1)],\end{aligned} \tag{46}$$

$$\boldsymbol{R}_1(t)=\boldsymbol{R}_1(t-1)+\hat{\boldsymbol{\psi}}(t)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)\in\mathbb{R}^{n_1\times n_1}, \tag{47}$$

$$\hat{\boldsymbol{c}}(t)=\hat{\boldsymbol{c}}(t-1)+\frac{1}{r_2(t)}[\boldsymbol{\xi}_2(t)-\boldsymbol{R}_2(t)\hat{\boldsymbol{c}}(t-1)], \tag{48}$$

$$r_2(t)=r_2(t-1)+\|\boldsymbol{f}(t)\|^2, \tag{49}$$

$$\begin{aligned}\boldsymbol{\xi}_2(t)&=\boldsymbol{\xi}_2(t-1)+\boldsymbol{f}(t)y_2(t)\\&=\boldsymbol{\xi}_2(t-1)+\boldsymbol{f}(t)[y(t)-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)]\in\mathbb{R}^m,\end{aligned} \tag{50}$$

$$\boldsymbol{R}_2(t)=\boldsymbol{R}_2(t-1)+\boldsymbol{f}(t)\boldsymbol{f}^{\mathrm{T}}(t)\in\mathbb{R}^{m\times m}, \tag{51}$$

$$\hat{\boldsymbol{\psi}}(t)=[-x_a(t-1),\cdots,-x_a(t-n_a),\bar{u}_a(t-1),\cdots,\bar{u}_a(t-n_b)]^{\mathrm{T}}\in\mathbb{R}^{n_1}, \tag{52}$$

$$\boldsymbol{f}(t)=[f_1(u(t)),f_2(u(t)),\cdots,f_m(u(t))]^{\mathrm{T}},\in\mathbb{R}^m, \tag{53}$$

$$\bar{u}_a(t)=\boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t), \tag{54}$$

$$x_a(t)=\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t)+\bar{u}_a(t). \tag{55}$$

The KT-AM-2S-RG algorithm involves the following steps:

1. Set all variables be zero when $t\leqslant 0$, and set the data length L and the basis function $f_j(*)$. Let $t=1$ and set the initial values: $\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_{n_1}/p_0$, $x_a(0)=1/p_0$, $\bar{u}_a(0)=1/p_0$, $\boldsymbol{\xi}_1(0)=\mathbf{1}_{n_1}/p_0$, $\boldsymbol{R}_1(0)=\boldsymbol{I}_{n_1}/p_0$, $\boldsymbol{\xi}_2(0)=\mathbf{1}_m/p_0$, and $\boldsymbol{R}_2(0)=\boldsymbol{I}_m/p_0$, $p_0=10^6$.
2. Collect the observation data $u(t)$ and $y(t)$, and construct the sub-information vector $\hat{\boldsymbol{\varphi}}_1(t)$, and $\boldsymbol{f}(t)$ using (52)–(53).
3. Compute $\boldsymbol{\xi}_1(t)$ and $\boldsymbol{R}_1(t)$ by (46)–(47). Compute $\boldsymbol{\xi}_2(t)$ and $\boldsymbol{R}_2(t)$ by (50)–(51).
4. Compute the step-size $r_1(t)$ and $r_2(t)$ by (45) and (49). Update the estimation vectors $\hat{\boldsymbol{\theta}}(t)$ and $\hat{\boldsymbol{c}}(t)$ using (44) and (48).
5. Compute the outputs $\bar{u}_a(t)$ and $x_a(t)$ of the auxiliary models by (54)–(55).
6. compare t with L: if $t\leqslant L$, increase t by 1 and go to Step 2; otherwise terminate this procedure, and obtain the $\hat{\boldsymbol{\theta}}(L)$ and $\hat{\boldsymbol{c}}(L)$.

3.3. *KT-AM-3S-RG algorithm*

According to the identification model in (5)–(7), introduce three intermediate variables:

$$y_a(t):=y(t)-\boldsymbol{\varphi}_b^{\mathrm{T}}(t)\boldsymbol{b}-\boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{c}\in\mathbb{R}, \tag{56}$$

$$y_b(t):=y(t)-\boldsymbol{\varphi}_a^{\mathrm{T}}(t)\boldsymbol{a}-\boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{c}\in\mathbb{R}, \tag{57}$$

$$y_c(t):=y(t)-\boldsymbol{\varphi}_a^{\mathrm{T}}(t)\boldsymbol{a}-\boldsymbol{\varphi}_b^{\mathrm{T}}(t)\boldsymbol{b}\in\mathbb{R}. \tag{58}$$

Equation (7) can be decomposed into the following three fictitious subsystems:

$$y_a(t)=\boldsymbol{\varphi}_a^{\mathrm{T}}(t)\boldsymbol{a}+v(t), \tag{59}$$

$$y_b(t)=\boldsymbol{\varphi}_b^{\mathrm{T}}(t)\boldsymbol{b}+v(t), \tag{60}$$

$$y_c(t)=\boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{c}+v(t). \tag{61}$$

DOI: 10.1002/acs

This is a three-stage identification model based on the key terms separation. Define the stacked output vectors and the stacked information matrices,

$$\begin{aligned}
\boldsymbol{Y}_a(t) &:= [y_a(1), y_a(2), y_a(3), \cdots, y_a(t)]^{\mathrm{T}} \in \mathbb{R}^t, \\
\boldsymbol{Y}_b(t) &:= [y_b(1), y_b(2), y_b(3), \cdots, y_b(t)]^{\mathrm{T}} \in \mathbb{R}^t, \\
\boldsymbol{Y}_c(t) &:= [y_c(1), y_c(2), y_c(3), \cdots, y_c(t)]^{\mathrm{T}} \in \mathbb{R}^t, \\
\boldsymbol{\Phi}_a(t) &:= [\boldsymbol{\varphi}_a(1), \boldsymbol{\varphi}_a(2), \boldsymbol{\varphi}_a(3), \cdots, \boldsymbol{\varphi}_a(t)]^{\mathrm{T}} \in \mathbb{R}^{t\times n_a}, \\
\boldsymbol{\Phi}_b(t) &:= [\boldsymbol{\varphi}_b(1), \boldsymbol{\varphi}_b(2), \boldsymbol{\varphi}_b(3), \cdots, \boldsymbol{\varphi}_b(t)]^{\mathrm{T}} \in \mathbb{R}^{t\times n_b}, \\
\boldsymbol{F}(t) &:= [\boldsymbol{f}(1), \boldsymbol{f}(2), \boldsymbol{f}(3), \cdots, \boldsymbol{f}(t)]^{\mathrm{T}} \in \mathbb{R}^{t\times m}.
\end{aligned}$$

For the subsystems in (59)–(61), one defines three criterion functions:

$$\begin{aligned}
J_4(\boldsymbol{a}) &:= \frac{1}{2}\|\boldsymbol{Y}_a(t) - \boldsymbol{\Phi}_a(t)\boldsymbol{a}\|^2, \\
J_5(\boldsymbol{b}) &:= \frac{1}{2}\|\boldsymbol{Y}_b(t) - \boldsymbol{\Phi}_b(t)\boldsymbol{b}\|^2, \\
J_6(\boldsymbol{c}) &:= \frac{1}{2}\|\boldsymbol{Y}_c(t) - \boldsymbol{F}(t)\boldsymbol{c}\|^2.
\end{aligned}$$

Define the following recursive relationship:

$$\boldsymbol{\xi}_a(t) := \boldsymbol{\Phi}_a^{\mathrm{T}}(t)\boldsymbol{Y}_a(t) = \boldsymbol{\xi}_a(t-1) + \boldsymbol{\varphi}_a(t)y_a(t) \in \mathbb{R}^{n_a}, \tag{62}$$
$$\boldsymbol{R}_a(t) := \boldsymbol{\Phi}_a^{\mathrm{T}}(t)\boldsymbol{\Phi}_a(t) = \boldsymbol{R}_a(t-1) + \boldsymbol{\varphi}_a(t)\boldsymbol{\varphi}_a^{\mathrm{T}}(t) \in \mathbb{R}^{n_a\times n_a}, \tag{63}$$
$$\boldsymbol{\xi}_b(t) := \boldsymbol{\Phi}_b^{\mathrm{T}}(t)\boldsymbol{Y}_b(t) = \boldsymbol{\xi}_b(t-1) + \boldsymbol{\varphi}_b(t)y_b(t) \in \mathbb{R}^{n_b}, \tag{64}$$
$$\boldsymbol{R}_b(t) := \boldsymbol{\Phi}_b^{\mathrm{T}}(t)\boldsymbol{\Phi}_b(t) = \boldsymbol{R}_b(t-1) + \boldsymbol{\varphi}_b(t)\boldsymbol{\varphi}_b^{\mathrm{T}}(t) \in \mathbb{R}^{n_b\times n_b}, \tag{65}$$
$$\boldsymbol{\xi}_c(t) := \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{Y}_c(t) = \boldsymbol{\xi}_c(t-1) + \boldsymbol{f}(t)y_c(t) \in \mathbb{R}^{m}, \tag{66}$$
$$\boldsymbol{R}_c(t) := \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{F}(t) = \boldsymbol{R}_c(t-1) + \boldsymbol{f}(t)\boldsymbol{f}^{\mathrm{T}}(t) \in \mathbb{R}^{m\times m}. \tag{67}$$

Let $\hat{\boldsymbol{a}}(t)$, $\hat{\boldsymbol{b}}(t)$ and $\hat{\boldsymbol{c}}(t)$ be the estimates of the parameter vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ at time t. Minimizing $J_4(\boldsymbol{a})$, $J_5(\boldsymbol{b})$ and $J_6(\boldsymbol{c})$ by using the negative gradient search, we have

$$\begin{aligned}
\hat{\boldsymbol{a}}(t) &= \hat{\boldsymbol{a}}(t-1) - \frac{1}{r_a(t)}\mathrm{grad}[J_4(\hat{\boldsymbol{a}}(t-1))] \\
&= \hat{\boldsymbol{a}}(t-1) + \frac{1}{r_a(t)}[\boldsymbol{\xi}_a(t) - \boldsymbol{R}_a(t)\hat{\boldsymbol{a}}(t-1)],
\end{aligned} \tag{68}$$

$$\begin{aligned}
\hat{\boldsymbol{b}}(t) &= \hat{\boldsymbol{b}}(t-1) - \frac{1}{r_b(t)}\mathrm{grad}[J_5(\hat{\boldsymbol{b}}(t-1))] \\
&= \hat{\boldsymbol{b}}(t-1) + \frac{1}{r_b(t)}[\boldsymbol{\xi}_b(t) - \boldsymbol{R}_b(t)\hat{\boldsymbol{b}}(t-1)],
\end{aligned} \tag{69}$$

$$\begin{aligned}
\hat{\boldsymbol{c}}(t) &= \hat{\boldsymbol{c}}(t-1) - \frac{1}{r_c(t)}\mathrm{grad}[J_6(\hat{\boldsymbol{c}}(t-1))] \\
&= \hat{\boldsymbol{c}}(t-1) + \frac{1}{r_c(t)}[\boldsymbol{\xi}_c(t) - \boldsymbol{R}_c(t)\hat{\boldsymbol{c}}(t-1)].
\end{aligned} \tag{70}$$

where $1/r_a(t)$, $1/r_b(t)$ and $1/r_c(t)$ are the step-size and satisfy

$$r_a(t) := r_a(t-1) + \|\boldsymbol{\varphi}_a(t)\|^2, \quad , r_a(0) = 1, \tag{71}$$
$$r_b(t) := r_b(t-1) + \|\boldsymbol{\varphi}_b(t)\|^2, \quad , r_b(0) = 1, \tag{72}$$
$$r_c(t) := r_c(t-1) + \|\boldsymbol{f}(t)\|^2, \quad , r_c(0) = 1, . \tag{73}$$

However, similar problems arise. The vectors $\boldsymbol{\varphi}_a(t)$ and $\boldsymbol{\varphi}_b(t)$ contain the unmeasured terms $x(t-i)$ and $\bar{u}(t-i)$. Equations (68)–(73) cannot be executed directly. Using the similar approach as in KT-AM-RG algorithm, replace the unknown $x(t-i)$ and $\bar{u}(t-i)$ with their estimates $x_a(t-i)$ and $\bar{u}_a(t-i)$ based on the auxiliary model identification idea.

DOI: 10.1002/acs

Replacing $\boldsymbol{\varphi}_a(t)$ and $\boldsymbol{\varphi}_b(t)$ in (68)–(73) with their estimate $\hat{\boldsymbol{\varphi}}_a(t)$ and $\hat{\boldsymbol{\varphi}}_b(t)$, we can derive the key term separation auxiliary model three-stage recursive gradient (KT-AM-3S-RG) algorithm:

$$\hat{\boldsymbol{a}}(t) = \hat{\boldsymbol{a}}(t-1) + \frac{1}{r_a(t)}[\boldsymbol{\xi}_a(t) - \boldsymbol{R}_a(t)\hat{\boldsymbol{a}}(t-1)], \tag{74}$$
$$r_a(t) = r_a(t-1) + \|\hat{\boldsymbol{\varphi}}_a(t)\|^2, \tag{75}$$
$$\boldsymbol{\xi}_a(t) = \boldsymbol{\xi}_a(t-1) + \hat{\boldsymbol{\varphi}}_a(t)[y(t) - \hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1) - \boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t-1)] \in \mathbb{R}^{n_a}, \tag{76}$$
$$\boldsymbol{R}_a(t) = \boldsymbol{R}_a(t-1) + \hat{\boldsymbol{\varphi}}_a(t)\hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t) \in \mathbb{R}^{n_a \times n_a}, \tag{77}$$
$$\hat{\boldsymbol{b}}(t) = \hat{\boldsymbol{b}}(t-1) + \frac{1}{r_b(t)}[\boldsymbol{\xi}_b(t) - \boldsymbol{R}_b(t)\hat{\boldsymbol{b}}(t-1)], \tag{78}$$
$$r_b(t) = r_b(t-1) + \|\hat{\boldsymbol{\varphi}}_b(t)\|^2, \tag{79}$$
$$\boldsymbol{\xi}_b(t) = \boldsymbol{\xi}_b(t-1) + \hat{\boldsymbol{\varphi}}_b(t)[y(t) - \hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1) - \boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t-1)] \in \mathbb{R}^{n_b}, \tag{80}$$
$$\boldsymbol{R}_b(t) = \boldsymbol{R}_b(t-1) + \hat{\boldsymbol{\varphi}}_b(t)\hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t) \in \mathbb{R}^{n_b \times n_b}, \tag{81}$$
$$\hat{\boldsymbol{c}}(t) = \hat{\boldsymbol{c}}(t-1) + \frac{1}{r_c(t)}[\boldsymbol{\xi}_c(t) - \boldsymbol{R}_c(t)\hat{\boldsymbol{c}}(t-1)], \tag{82}$$
$$r_c(t) = r_c(t-1) + \|\boldsymbol{f}(t)\|^2, \tag{83}$$
$$\boldsymbol{\xi}_c(t) = \boldsymbol{\xi}_c(t-1) + \boldsymbol{f}(t)[y(t) - \hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1) - \hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1)] \in \mathbb{R}^{m}, \tag{84}$$
$$\boldsymbol{R}_c(t) = \boldsymbol{R}_c(t-1) + \boldsymbol{f}(t)\boldsymbol{f}^{\mathrm{T}}(t) \in \mathbb{R}^{m \times m}, \tag{85}$$
$$\hat{\boldsymbol{\varphi}}_a(t) = [-x_a(t-1), -x_a(t-2), \cdots, -x_a(t-n_a)]^{\mathrm{T}} \in \mathbb{R}^{n_a}, \tag{86}$$
$$\hat{\boldsymbol{\varphi}}_b(t) = [\bar{u}_a(t-1), \bar{u}_a(t-2), \cdots, \bar{u}_a(t-n_b)]^{\mathrm{T}} \in \mathbb{R}^{n_b}, \tag{87}$$
$$\boldsymbol{f}(t) = [f_1(u(t)), f_2(u(t)), \cdots, f_m(u(t))]^{\mathrm{T}} \in \mathbb{R}^{m}, \tag{88}$$
$$\bar{u}_a(t) = \boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t), \tag{89}$$
$$x_a(t) = \hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t) + \hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t) + \bar{u}_a(t). \tag{90}$$

The KT-AM-3S-RG algorithm involves the following steps:

1. Set all variables be zero when $t \leqslant 0$, and set the data length L and the basis function $f_j(*)$. Let $t = 1$ and set the initial values: $\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_n/p_0$, $x_a(0) = 1/p_0$, $\bar{u}_a(0) = 1/p_0$, $\boldsymbol{\xi}_a(0) = \mathbf{1}_{n_a}/p_0$, $\boldsymbol{R}_a(0) = \boldsymbol{I}_{n_a}/p_0$, $\boldsymbol{\xi}_b(0) = \mathbf{1}_{n_b}/p_0$, $\boldsymbol{R}_b(0) = \boldsymbol{I}_{n_b}/p_0$, $\boldsymbol{\xi}_c(0) = \mathbf{1}_m/p_0$, and $\boldsymbol{R}_c(0) = \boldsymbol{I}_m/p_0$, $p_0 = 10^6$.
2. Collect the observation data $u(t)$ and $y(t)$, and construct the sub-information vector $\hat{\boldsymbol{\varphi}}_a(t)$, $\hat{\boldsymbol{\varphi}}_b(t)$ and $\boldsymbol{f}(t)$ by (86)–(88).
3. Compute $\boldsymbol{\xi}_a(t)$ and $\boldsymbol{R}_a(t)$ by (76)–(77). Compute $\boldsymbol{\xi}_b(t)$ and $\boldsymbol{R}_b(t)$ by (80)–(81). Compute $\boldsymbol{\xi}_c(t)$ and $\boldsymbol{R}_c(t)$ by (84)–(85).
4. Compute the step-size $r_a(t)$, $r_b(t)$ and $r_c(t)$ by (75), (79) and (83). Update the parameter estimation vectors $\hat{\boldsymbol{a}}(t)$, $\hat{\boldsymbol{b}}(t)$ and $\hat{\boldsymbol{c}}(t)$ using (74), (78) and (82).
5. Compute the outputs $\bar{u}_a(t)$ and $x_a(t)$ of the auxiliary models by using (89)–(90).
6. Compare t with L: if $t \leqslant L$, increase t by 1 and go to Step 2; otherwise terminate this procedure, and obtain the estimates $\hat{\boldsymbol{a}}(L)$, $\hat{\boldsymbol{b}}(L)$ and $\hat{\boldsymbol{c}}(L)$.

4. CALCULATION ANALYSIS

The efficiency of the algorithm depends on its complexity, which can be measured by the computational burden. The computational burden of an algorithm can be evaluated by counting the number of floating-point operations, which is the sum of the number of multiplication and addition operations (note: division is considered as multiplication, and subtraction is considered as addition).

The computational burden of the KT-AM-RG, KT-AM-2S-RG, and KT-AM-3S-RG algorithms at each recursive calculation is shown in Tables I-III.

DOI: 10.1002/acs

Table I. Computational burden of the KT-AM-RG algorithm

Expressions	Multiplications	Additions
$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+[\boldsymbol{\xi}(t)-\boldsymbol{R}(t)\hat{\boldsymbol{\vartheta}}(t-1)]/r(t)$	n^2+n	n^2+n
$r(t)=r(t-1)+\Vert\hat{\boldsymbol{\varphi}}(t)\Vert^2$	n	n
$\boldsymbol{\xi}(t)=\boldsymbol{\xi}(t-1)+\hat{\boldsymbol{\varphi}}(t)y(t)$	n	n
$\boldsymbol{R}(t)=\boldsymbol{R}(t-1)+\hat{\boldsymbol{\varphi}}(t)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(t)$	n^2	n^2
$\bar{u}_a(t)=\boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t)$	m	$m-1$
$x_a(t)=\hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)+\hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t)+\bar{u}_a(t)$	n_a+n_b	n_a+n_b
Total flops	$N_1:=4n^2+8n-1$	

Table II. Computational burden of the KT-AM-2S-RG algorithm

Expressions	Multiplications	Additions
$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+[\boldsymbol{\xi}_1(t)-\boldsymbol{R}_1(t)\hat{\boldsymbol{\theta}}(t-1)]/r_1(t)$	$n_1^2+n_1$	$n_1^2+n_1$
$r_1(t)=r_1(t-1)+\Vert\hat{\boldsymbol{\psi}}(t)\Vert^2$	n_1	n_1
$\boldsymbol{\xi}_1(t)=\boldsymbol{\xi}_1(t-1)+\hat{\boldsymbol{\psi}}(t)[y(t)-\boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t-1)]$	n_1+m	n_1+m
$\boldsymbol{R}_1(t)=\boldsymbol{R}_1(t-1)+\hat{\boldsymbol{\psi}}(t)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)$	n_1^2	n_1^2
$\hat{\boldsymbol{c}}(t)=\hat{\boldsymbol{c}}(t-1)+[\boldsymbol{\xi}_2(t)-\boldsymbol{R}_2(t)\hat{\boldsymbol{c}}(t-1)]/r_2(t)$	m^2+m	m^2+m
$r_2(t)=r_2(t-1)+\Vert\boldsymbol{f}(t)\Vert^2$	m	m
$\boldsymbol{\xi}_2(t)=\boldsymbol{\xi}_2(t-1)+\boldsymbol{f}(t)[y(t)-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)]$	n_1+m	n_1+m
$\boldsymbol{R}_2(t)=\boldsymbol{R}_2(t-1)+\boldsymbol{f}(t)\boldsymbol{f}^{\mathrm{T}}(t)$	m^2	m^2
$\bar{u}_a(t)=\boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t)$	m	$m-1$
$x_a(t)=\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t)+\bar{u}_a(t)$	n_1	n_1
Total flops	$N_2:=4n_1^2+4m^2+10n_1+10m-1$	

Table III. Computational burden of the KT-AM-3S-RG algorithm

Expressions	Multiplications	Additions
$\hat{\boldsymbol{a}}(t)=\hat{\boldsymbol{a}}(t-1)+[\boldsymbol{\xi}_a(t)-\boldsymbol{R}_a(t)\hat{\boldsymbol{a}}(t-1)]/r_a(t)$	$n_a^2+n_a$	$n_a^2+n_a$
$r_a(t)=r_a(t-1)+\Vert\hat{\boldsymbol{\varphi}}_a(t)\Vert^2$	n_a	n_a
$\boldsymbol{\xi}_a(t)=\boldsymbol{\xi}_a(t-1)+\hat{\boldsymbol{\varphi}}_a(t)[y(t)-\hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1)-\boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t-1)]$	n_b+m	n_b+m
$\boldsymbol{R}_a(t)=\boldsymbol{R}_a(t-1)+\hat{\boldsymbol{\varphi}}_a(t)\hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t)$	n_a^2	n_a^2
$\hat{\boldsymbol{b}}(t)=\hat{\boldsymbol{b}}(t-1)+[\boldsymbol{\xi}_b(t)-\boldsymbol{R}_b(t)\hat{\boldsymbol{b}}(t-1)]/r_b(t)$	$n_b^2+n_b$	$n_b^2+n_b$
$r_b(t)=r_b(t-1)+\Vert\hat{\boldsymbol{\varphi}}_b(t)\Vert^2$	n_b	n_b
$\boldsymbol{\xi}_b(t)=\boldsymbol{\xi}_b(t-1)+\hat{\boldsymbol{\varphi}}_b(t)[y(t)-\hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)-\boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t-1)]$	n_a+m	n_a+m
$\boldsymbol{R}_b(t)=\boldsymbol{R}_b(t-1)+\hat{\boldsymbol{\varphi}}_b(t)\hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t)$	n_b^2	n_b^2
$\hat{\boldsymbol{c}}(t)=\hat{\boldsymbol{c}}(t-1)+[\boldsymbol{\xi}_c(t)-\boldsymbol{R}_c(t)\hat{\boldsymbol{c}}(t-1)]/r_c(t)$	m^2+m	m^2+m
$r_c(t)=r_c(t-1)+\Vert\boldsymbol{f}(t)\Vert^2$	m	m
$\boldsymbol{\xi}_c(t)=\boldsymbol{\xi}_c(t-1)+\boldsymbol{f}(t)[y(t)-\hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)-\hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1)]$	n_a+n_b	n_a+n_b
$\boldsymbol{R}_c(t)=\boldsymbol{R}_c(t-1)+\boldsymbol{f}(t)\boldsymbol{f}^{\mathrm{T}}(t)$	m^2	m^2
$\bar{u}_a(t)=\boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{c}}(t)$	m	$m-1$
$x_a(t)=\hat{\boldsymbol{\varphi}}_a^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t)+\hat{\boldsymbol{\varphi}}_b^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t)+\bar{u}_a(t)$	n_a+n_b	n_a+n_b
Total flops	$N_3:=4n_a^2+4n_b^2+4m^2+10n-1$	

From Tables I to III, the computational efficiencies of the KT-AM-RG, KT-AM-2S-RG, and KT-AM-3S-RG algorithms are

$$N_1:=4n^2+8n-1,$$

DOI: 10.1002/acs

$$\begin{aligned} N_2 &:= 4n_1^2 + 4m^2 + 10n_1 + 10m - 1 \\ &= 4n_1^2 + 4m^2 + 10n - 1, \\ N_3 &:= 4n_a^2 + 4n_b^2 + 4m^2 + 10(n_a + n_b + m) - 1 \\ &= 4n_a^2 + 4n_b^2 + 4m^2 + 10n - 1. \end{aligned}$$

When n_a, n_b, $m \geqslant 1$, the difference in computational burden between the three algorithms is,

$$\begin{aligned} N_1 - N_2 &= 4n^2 + 8n - 1 - (4n_1^2 + 4m^2 + 10n - 1) \\ &= 4n^2 - 4n_1^2 - 4m^2 - 2n \\ &= 8n_1 m - 2(n_1 + m) > 0, \\ N_2 - N_3 &= 4n_1^2 + 4m^2 + 10n - 1 - (4n_a^2 + 4n_b^2 + 4m^2 + 10n - 1) \\ &= 4n_1^2 - 4n_a^2 - 4n_b^2 \\ &= 8n_a n_b > 0. \end{aligned}$$

Obviously, $N_1 > N_2 > N_3$. To highlight the differences between each algorithm more clearly, we present a specific numerical comparison below. Assuming $n_a = 10$, $n_b = 10$, $m = 10$, we obtain $N_1 = 3839$, $N_2 = 2299$, and $N_3 = 1499$. As the system dimension increases, this gap becomes more pronounced, indicating that the computational efficiency of the KT-AM-3S-RG algorithm is superior to that of the KT-AM-RG and KT-AM-2S-RG algorithms.

5. SIMULATION RESULTS

The following H-OE system is used for the simulation,

$$\begin{aligned} y(t) &= \frac{B(z)}{A(z)}\bar{u}(t) + v(t) \\ A(z) &= 1 + a_1 z^{-1} + a_2 z^{-2} = 1 + 0.40z^{-1} - 0.56z^{-2}, \\ B(z) &= b_0 + b_1 z^{-1} + b_2 z^{-2} = 1 - 0.32z^{-1} - 0.26z^{-2}, \\ \bar{u}(t) &= f(u(t)) = c_1 u(t) + c_2 u^2(t) + c_3 u^3(t) = -3.35u(t) - 1.86u^2(t) + 2.36u^3(t), \\ \boldsymbol{\vartheta} &= [a_1, a_2, b_1, b_2, c_1, c_2, c_3]^{\mathrm{T}} = [0.40, -0.56, -0.32, -0.26, -3.35, -1.86, 2.36]^{\mathrm{T}}, \end{aligned}$$

where $\{u(t)\}$ is taken as an independent persistent excitation signal sequence with zero mean and unit variance, and $\{v(t)\}$ is a white noise with zero mean and variance $\sigma_1^2 = 0.10^2$ and $\sigma_2^2 = 0.60^2$, respectively. Set the data length $L_e = 3000$ and apply the proposed algorithms to estimate the parameters of this example system. The parameter estimates and errors $\delta := \|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}\|/\|\boldsymbol{\vartheta}\|$ is shown in Tables IV–VI, the parameter estimation errors versus t are plotted in Figures 2–3.

From this simulation, some conclusions can be obtained as follow.

- The parameter estimates of all three methods are getting closer to the true values with t increasing. Hence, the proposed recursive algorithms are effective for this model.
- The KT-AM-3S-RG algorithm has a higher estimation accuracy compared to the other two algorithms under the same simulation conditions.
- It can be observed that the estimation accuracies of KT-AM-2S-RG and KT-AM-3S-RG algorithms are similar when the data length is large. The accuracy of the parameter estimates from the proposed algorithms improves with decreasing noise levels.

6. CONCLUSION

This paper investigates the parameter estimation of the Hammerstein output-error systems. To achieve this, we introduce the KT-AM-RG algorithm, which is derived from the key term separation technique. The presented KT-AM-2S-RG and KT-AM-3S-RG algorithms utilize the hierarchical

DOI: 10.1002/acs

Table IV. The KT-AM-RG estimates and errors under different σ^2

σ^2	t	a_1	a_2	b_1	b_2	c_1	c_2	c_3	δ (%)
0.10^2	100	0.48257	-0.62631	-0.35853	-0.22114	-0.25581	-1.61125	0.82878	75.76551
	200	0.48015	-0.50028	-0.53214	-0.09344	-0.75329	-2.00215	1.13486	63.20271
	500	0.38954	-0.52346	-0.39593	-0.25366	-1.45364	-2.07173	1.45585	46.22993
	1000	0.45127	-0.50168	-0.29253	-0.26345	-2.18688	-1.95730	1.81566	28.23173
	2000	0.50813	-0.46041	-0.23473	-0.25066	-2.79771	-1.91106	2.10294	13.88215
	3000	0.50577	-0.45718	-0.23232	-0.24105	-3.07824	-1.88478	2.23070	7.60945
0.60^2	100	0.36319	-0.57269	-0.41838	-0.16995	-0.15748	-1.66081	0.72776	78.61923
	200	0.33698	-0.53620	-0.53421	-0.19697	-0.72051	-2.05058	0.99673	65.12947
	500	0.34722	-0.55723	-0.43802	-0.26496	-1.38651	-2.17678	1.38932	48.49767
	1000	0.44141	-0.51119	-0.29352	-0.26912	-2.12777	-1.99574	1.77370	29.84273
	2000	0.51372	-0.45581	-0.22587	-0.25261	-2.73874	-1.92828	2.06969	15.39589
	3000	0.51393	-0.44967	-0.22507	-0.23896	-3.03344	-1.89765	2.20183	8.78473
True values		0.40000	-0.56000	-0.32000	-0.26000	-3.35000	-1.86000	2.23600	

Table V. The KT-AM-2S-RG estimates and errors under different σ^2

σ^2	t	a_1	a_2	b_1	b_2	c_1	c_2	c_3	δ (%)
0.10^2	100	0.27749	-0.74607	-0.64188	-0.10531	-2.75337	-1.80724	2.03452	17.52649
	200	0.20296	-0.73220	-0.59708	-0.17639	-3.45427	-1.83547	2.40786	8.91333
	500	0.20733	-0.72826	-0.52240	-0.26039	-3.41853	-1.90508	2.38778	7.38306
	1000	0.27536	-0.67744	-0.44163	-0.27680	-3.37987	-1.90557	2.37073	4.76719
	2000	0.34716	-0.61073	-0.36937	-0.27282	-3.36507	-1.89964	2.36451	2.16415
	3000	0.37626	-0.58262	-0.34187	-0.26799	-3.35871	-1.89461	2.36119	1.17656
0.60^2	100	0.24326	-0.82164	-0.63344	-0.19643	-2.84984	-1.93404	2.04801	16.19865
	200	0.14341	-0.80328	-0.62993	-0.22796	-3.46706	-2.02361	2.43019	11.31459
	500	0.15107	-0.78176	-0.54956	-0.29944	-3.51319	-2.00566	2.44582	10.27505
	1000	0.25540	-0.69918	-0.43912	-0.30646	-3.38958	-1.97173	2.38335	5.83822
	2000	0.33209	-0.62736	-0.37112	-0.29402	-3.34867	-1.94556	2.36222	3.11278
	3000	0.35722	-0.60251	-0.35270	-0.28343	-3.34085	-1.93205	2.35231	2.25115
True values		0.40000	-0.56000	-0.32000	-0.26000	-3.35000	-1.86000	2.23600	

identification principle to reduce the computational burden. In terms of computation complexity, both the KT-AM-2S-RG and KT-AM-3S-RG algorithms have a lower computationally intensive than the KT-AM-RG algorithm. The effectiveness of the presented methods have been verified by the simulation, and the accuracy of parameter estimation can be improved by using hierarchical identification principle.

ACKNOWLEDGEMENT

This work was supported by National Grants No. 61671356, Science and Technology Nature Science Foundation of China (NSFC) under Program of Shaanxi Province under G2020GY-136 and 2019ZDLGY14-02-03, Key industrial innovation chain project of Shaanxi Province key research and development plan No.2022ZDLGY03-01.

Table VI. The KT-AM-3S-RG estimates and errors under different σ^2

σ^2	t	a_1	a_2	b_1	b_2	c_1	c_2	c_3	δ (%)
0.10^2	100	0.49325	-0.48924	-0.49124	0.05491	-2.59528	-1.63895	2.05047	20.24565
	200	0.44988	-0.51738	-0.36682	-0.12323	-3.30638	-1.72877	2.34487	4.61767
	500	0.41684	-0.51917	-0.30659	-0.21204	-3.40239	-1.75340	2.38358	3.02299
	1000	0.42725	-0.52922	-0.30440	-0.22352	-3.39186	-1.77508	2.37886	2.45401
	2000	0.42930	-0.52851	-0.30083	-0.23466	-3.38060	-1.80017	2.37807	1.91998
	3000	0.43000	-0.52950	-0.29875	-0.23904	-3.36043	-1.81104	2.36684	1.58866
0.60^2	100	0.41111	-0.58119	-0.54909	0.03876	-2.15320	-1.57373	1.93905	29.62455
	200	0.37582	-0.59711	-0.45592	-0.10956	-2.87220	-1.72822	2.22225	12.13515
	500	0.33784	-0.58609	-0.37828	-0.23092	-3.22949	-1.74034	2.34183	4.26210
	1000	0.35098	-0.60039	-0.37416	-0.23753	-3.29102	-1.76734	2.34992	3.06570
	2000	0.35514	-0.59968	-0.37227	-0.24876	-3.31166	-1.80397	2.35952	2.30006
	3000	0.35660	-0.60017	-0.37148	-0.25088	-3.30563	-1.81860	2.34625	2.19873
True values		0.40000	-0.56000	-0.32000	-0.26000	-3.35000	-1.86000	2.23600	

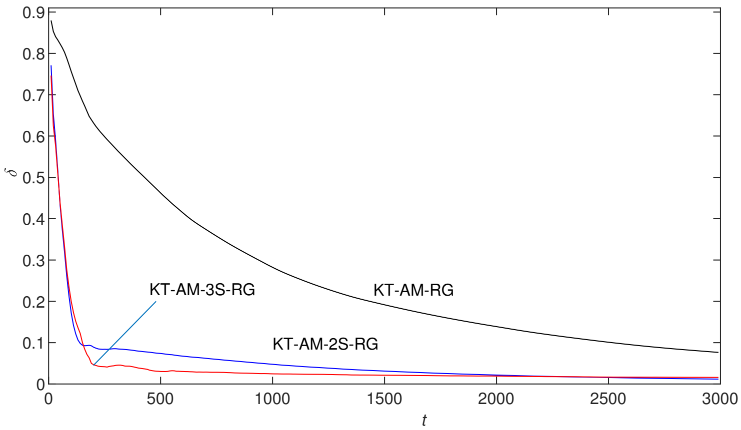

Figure 2. The estimation errors δ versus t (σ^2=0.10^2).

REFERENCES

1. Pan J, Li W, Zhang HP. Control algorithms of magnetic suspension systems based on the improved double exponential reaching law of sliding mode control. *International Journal of Control Automation and Systems*. 2018;16(6):2878-2887.
2. Wang YJ, Tang SH, Deng MQ. Modeling nonlinear systems using the tensor network B-spline and the multi-innovation identification theory. *Int J Robust Nonlinear Control*. 2022; 32(13):7304-7318.
3. Zhang KW, Shi Y. Adaptive model predictive control for a class of constrained linear systems with parametric uncertainties. *Automatica*. 2020;117;108974.
4. Xu L. Separable multi-innovation Newton iterative modeling algorithm for multi-frequency signals based on the sliding measurement window. *Circuits Syst Signal Process*. 2022; 41(2):805-830.
5. Li MH, Liu XM. Maximum likelihood hierarchical least squares-based iterative identification for dual-rate stochastic systems. *International Journal of Adaptive Control and Signal Processing*. 2021;35(2):240-261.
6. Noshadi A, Shi J, Lee WS, et al. System identification and robust control of multi-input multi-output active magnetic bearing systems. *IEEE Transactions on Control Systems Technology*. 2016;24(4):1227-1239.
7. Xu N, Ding F. Parameter estimation for a class of time-varying systems with the invariant matrix. *Int J Robust Nonlinear Control*. 2023; 33(3):2163-2181.
8. Xu L, Song GL. A recursive parameter estimation algorithm for modeling signals with multi-frequencies. *Circuits Systems and Signal Processing*. 2020;39(8):4198-4224.
9. Ma H, Pan J, Ding F, et al. Partially-coupled least squares based iterative parameter estimation for multi-variable output-error-like. *IET Control Theory & Applications*. 2019;13(18):3040-3051.

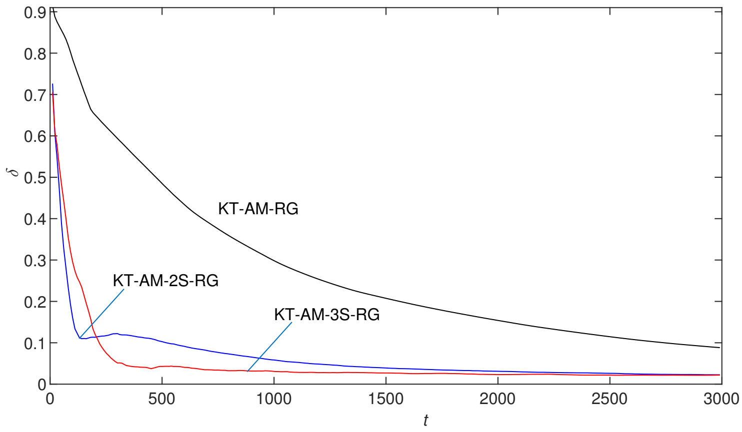

Figure 3. The estimation errors δ versus t (σ^2=0.60^2).

10. Kazemi M, Arefi MM. A fast iterative recursive least squares algorithm for Wiener model identification of highly nonlinear systems. *ISA Transactions*. 2017;67:382-388.
11. Chen J, Jiang B. Modified stochastic gradient parameter estimation algorithms for a nonlinear two-variable difference system. *International Journal of Control Automation and Systems*. 2016;14(6):1493-1500.
12. Li JM, Ding F, Hayat T. A novel nonlinear optimization method for fitting a noisy Gaussian activation function. *International Journal of Adaptive Control and Signal Processing*. 2022; 36 (3):690-707.
13. Brewick PT and Masri SF. An evaluation of data-driven identification strategies for complex nonlinear dynamic systems. *Nonlinear Dynamics*. 2016;85(2):1297-1318.
14. Xu L, Ding F, Yang EF. Auxiliary model multiinnovation stochastic gradient parameter estimation methods for nonlinear sandwich systems. *Int J Robust Nonlinear Control*. 2021; 31(1):148-165.
15. Ji Y, Zhang C, Kang Z, Yu T. Parameter estimation for block-oriented nonlinear systems using the key term separation. *International Journal of Robust and Nonlinear Control*. 2020;30(9):3727-3752.
16. Li F, Chen LY, Wo SL, et al. Modeling and parameter learning method for the Hammerstein-Wiener model with disturbance. *Measurement and Control*. 2020;53(5-6):971-982.
17. Ji Y, Kang Z. Three-stage forgetting factor stochastic gradient parameter estimation methods for a class of nonlinear systems. *Int J Robust Nonlinear Control*. 2021; 31(3):971-987.
18. Li JH, Zheng W, Gu JP, Hua L. A recursive identification algorithm for Wiener nonlinear systems with linear state-space subsystem. *Circuits Systems and Signal Processing*. 2018;37(6):2374-2393.
19. Li JH, Zhang JL. Maximum likelihood identification of dual-rate Hammerstein output-error moving average system. *IET Control Theory and Applications*. 2020;14(8):1089-1101.
20. Dong SJ, Yu L, Zhang WA, Chen B. Robust extended recursive least squares identification algorithm for Hammerstein systems with dynamic disturbances. *Digital Signal Processing*. 2020;101:102716.
21. Liu XM, Fan YM. Maximum likelihood extended gradient-based estimation algorithms for the input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity. *Int J Robust Nonlinear Control*. 2021; 31(9):4017-4036.
22. Ji Y, Kang Z, Liu XM. The data filtering based multiple-stage Levenberg-Marquardt algorithm for Hammerstein nonlinear systems. *Int J Robust Nonlinear Control*. 2021; 31(15):7007-7025.
23. Kang Z, Ji Y, and Liu X. Hierarchical recursive least squares algorithms for Hammerstein nonlinear autoregressive output-error systems. *International Journal of Adaptive Control and Signal Processing*. 2021;35(11):2276-2295.
24. Wang DQ, Zhang S, Gan M, Qiu JL. A novel EM identification method for Hammerstein systems With missing output data. *IEEE Transactions on Industrial Informatics*. 2020;16(4);2500-2508.
25. Xu L. Separable Newton recursive estimation method through system responses based on dynamically discrete measurements with increasing data length. *International Journal of Control Automation and Systems*. 2022;20(2):432-443.
26. Meng DD. Recursive least squares and multi-innovation gradient estimation algorithms for bilinear stochastic systems. *Circuits Systems and Signal Processing*. 2017;36(3):1052-1065.
27. Wang YJ, Yang L. An efficient recursive identification algorithm for multilinear systems based on tensor decomposition. *Int J Robust Nonlinear Control*. 2021; 31(11):7920-7936.
28. Ding F, Xu L, Zhang X, Zhou YH. Filtered auxiliary model recursive generalized extended parameter estimation methods for Box-Jenkins systems by means of the filtering identification idea. *Int J Robust Nonlinear Control*. 2023; 33. doi:10.1002/rnc.6657
29. Wang JW, Ji Y, Zhang X, Xu L, Two-stage gradient-based iterative algorithms for the fractional-order nonlinear systems by using the hierarchical identification principle. *International Journal of Adaptive Control and Signal Processing*. 2022; 36(7):1778-1796.
30. Ding F, Ma H, Pan J, Yang EF. Hierarchical gradient- and least squares-based iterative algorithms for input nonlinear output-error systems using the key term separation. *Journal of the Franklin Institute*. 2021;358(9):5113-5135.

31. Xu L, Xiong WL, Alsaedi A, Hayat T. Hierarchical parameter estimation for the frequency response based on the dynamical window data. *International Journal of Control Automation and Systems*. 2018;16(4):1756-1764.
32. Ji Y, Jiang XK, Wan LJ. Hierarchical least squares parameter estimation algorithm for two-input Hammerstein finite impulse response systems. *Journal of the Franklin Institute*. 2020;357(8):5019-5032.
33. Wang DQ, Yan YR, Liu YJ, and Ding JH. Model recovery for Hammerstein systems using the hierarchical orthogonal matching pursuit method. *Journal of Computational and Applied Mathematics*. 2019;345:135-145.
34. Vörös J. Identification of nonlinear dynamic systems with input saturation and output backlash using three-block cascade models. *Journal of the Franklin Institute*. 2014;351(12);5455-5466.
35. Vörös J. Identification of nonlinear cascade systems with output hysteresis based on the key term separation principle. *Applied Mathematical Modelling*. 2015;39(18):5531-5539.
36. Wang JW, Ji Y, Zhang C. Iterative parameter and order identification for fractional-order nonlinear finite impulse response systems using the key term separation. *International Journal of Adaptive Control and Signal Processing*. 2021;35(8):1562-1577.
37. Ding F, Chen HB, Xu L, et al. A hierarchical least squares identification algorithm for Hammerstein nonlinear systems using the key term separation. *Journal of the Franklin Institute*. 2018;355(8):3737-3752.