Hierarchical gradient- and least squares-based iterative algorithms for input nonlinear output-error systems using the key term separation

Feng Ding[*,a], Hao Ma[a], Jian Pan[b], Erfu Yang[c]

[a]*School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, PR China*
[b]*School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, PR China*
[c]*Department of Design, Manufacturing and Engineering Management, University of Strathclyde, Glasgow G1 1XJ, Scotland, UK*

Abstract

This paper considers the parameter identification problems of the input nonlinear output-error (IN-OE) systems, that is the Hammerstein output-error systems. In order to overcome the excessive calculation amount of the over-parameterization method of the IN-OE systems. Through applying the hierarchial identification principle and decomposing the IN-OE system into three subsystems with a smaller number of parameters, we present the key term separation auxiliary model hierarchical gradient-based iterative algorithm and the key term separation auxiliary model hierarchical least squares-based iterative algorithm, which are called the key term separation auxiliary model three-stage gradient-based iterative algorithm and the key term separation auxiliary model three-stage least squares-based iterative algorithm. The comparison of the calculation amount and the simulation analysis indicate that the proposed algorithms are effective.

Key words: Parameter estimation, Iterative identification, Hierarchical identification, Key term separation, Nonlinear system

1. Introduction

Mathematical models are the basic requirement of studying natural sciences [1] and the solutions of numerous problems arising in the field of engineering and applied physics are based on the mathematical models [2, 3]. Parameter identification plays a key role in system identification [4, 5], and the parameter identification of linear and nonlinear systems is always a hot topic in the filed of system identification [6, 7, 8]. Nonlinear systems are simply categorized into the input nonlinear systems (i.e., Hammerstein nonlinear systems), the output nonlinear systems (i.e., Wiener nonlinear systems), the feedback nonlinear systems, and the input-output nonlinear systems (i.e., Hammertein–Wiener nonlinear systems). Recently, many methods have been proposed to solve the parameter identification problems of nonlinear systems [9, 10]. For instance, Zhang et al. proposed an adaptive-noise-correction integrated parameter identification method for time-delayed nonlinear systems [11]. By dividing the variables into the linear and nonlinear parts to simplify a class of nonlinear least squares problems, Gan et al. studied several separated algorithms for such problems and compared the performance of these algorithms by making use of the Monte Carlo experiments [12, 13, 14, 15, 16].

The input nonlinear systems consist of a static nonlinear block and a linear dynamical subsystem, and can be categorized into the input nonlinear equation-error systems and the input nonlinear output-error systems based on the features of the linear parts [17]. Many applications about the input nonlinear systems have been studied in the literatures. Various parameter identification methods have been proposed for input nonlinear systems, such as the auxiliary model identification idea [18], the multi-innovation identification theory [19]. For example, Albu and Nishikawa presented an iterative kernel algorithm for nonlinear acoustic echo cancellation [20]. Equation-error models and output-error models are two basic types of stochastic systems, and have received considerable attention in the field of system identification [21]. The output-error model is often used in practice, and the parameter estimation for output-error systems is essential [23, 24, 25, 26, 27]. Recently, Ding et al. studied the parameter identification of linear and nonlinear output-error models, and proposed a particle filtering based recursive least squares algorithm and a particle filtering based multi-innovation recursive least squares algorithm to identify the system parameters [28]. The recursive identification and the iterative identification are basic for parameter identification [29, 30, 31, 32]. Differently from the recursive algorithms, the iterative algorithms update the parameter estimates by making use of a batch of data. The gradient-based iterative and

☆This work was supported by the National Natural Science Foundation of China (No. 61873111) and the 111 Project (B12018).
*Corresponding author
Email address: fding@jiangnan.edu.cn (Feng Ding)

least squares-based iterative algorithms are the commonly used iterative algorithms for carrying out parameter estimation and for solving some matrix equations [33, 34, 35, 36].

This paper considers the parameter estimation problems of input nonlinear output-error systems and develops a key term separation auxiliary model three-stage gradient-based iterative (KT-AM-3S-GI) algorithm and a key term separation auxiliary model three-stage least squares-based iterative (KT-AM-3S-LSI) algorithm based on the key term separation technique and the hierarchical identification principle. The salient features of this work are summarized as follows.

- Based on the hierarchical identification principle, the key term separation identification model of the IN-OE model is decomposed into three sub-models to improve the computational efficiency.
- Based on the gradient search and least squares search, a KT-AM-3S-GI algorithm and a KT-AM-3S-LSI algorithm are proposed to estimate the parameters of the IN-OE systems.

This paper is organized as follows. Section 2 establishes the over-parameterization identification model. Section 3 establishes the key term separation identification model and the key term separation three-stage identification model of the input nonlinear output-error systems. A KT-AM-3S-GI algorithm is presented in Section 4. Section 5 proposes a KT-AM-3S-LSI algorithm. Section 6 analyzes the the calculation amount of the proposed algorithms and the existing algorithms. Section 7 offers an example to illustrate the effectiveness of the proposed algorithms. Finally, Section 8 gives some concluding remarks.

2. Over-parameterziation identification model and the least squares-based iterative algorithm

Regression analysis is a method of predictive modeling technology that studies the relationship between the dependent variable (target) and the independent variable (predictor). This technique is used in forecasting, time series modeling and finding causal relationships between variables. The regression analysis is an important tool for data modeling and analysis. The regression models are used to analyze the relationship between variables and to reveal the degree of influence of multiple independent variables on a dependent variable.

Consider the input nonlinear output-error system described by an input nonlinear output-error (IN-OE) model in Figure 1 [37], where the linear dynamical subsystem of the IN-OE model is given by

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

The variable $y(t) \in \mathbb{R}$ is the output of the system, $v(t) \in \mathbb{R}$ is a white noise with zero mean, $A(z)$ and $B(z)$ are the polynomials of the unit shift operator z^{-1} $[z^{-1}y(t) = y(t-1)]$ with following mathematical descriptions:

$$A(z) := 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a}, \quad a_i \in \mathbb{R},$$
$$B(z) := b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b}, \quad b_i \in \mathbb{R},$$

$\bar{u}(t) \in \mathbb{R}$ 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 γ_j's, that is

$$\bar{u}(t) = f(u(t)) = \sum_{j=1}^{m} \gamma_j f_j(u(t)) = \gamma_1 f_1(u(t)) + \gamma_2 f_2(u(t)) + \cdots + \gamma_m f_m(u(t)), \tag{2}$$

$u(t) \in \mathbb{R}$ is the input of the system. Assume that the degrees m, n_a and n_b are known, and $y(t) = 0$, $u(t) = 0$ and $v(t) = 0$ for $t \leqslant 0$.

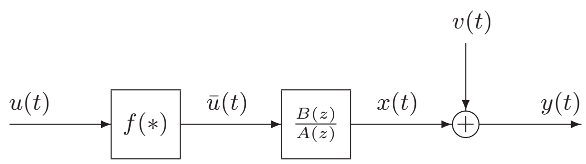

Figure 1: The input nonlinear output-error (IN-OE) system

From the IN-OE systems, we observe that $\bar{u}(t)$ is a hidden variable and is the output of the nonlinear part and the input of the linear dynamical subsystem. Since $\bar{u}(t)$ is an hidden variable located in the system, it cannot be measured directly. Meanwhile, the output $y(t)$ is the noisy measurement of $x(t)$. For the nonlinear part, there are many sets of basis functions such as polynomials, trigonometric functions, exponential functions, piecewise linear functions, etc., which can be chosen as the nonlinear basis functions $f_j(*)$.

Because the nonlinear function $f(*)$ is in series with the transfer function $\frac{B(z)}{A(z)}$, for any nonzero constant α, $(\alpha f(*), B(z)/\alpha)$ and $(f(*), B(z))$ result in the identified input and output relationship for Equations (2)–(1). In order to ensure that the system parameters are identifiable, a necessary work is to normalize $\bar{u}(t)$ or $B(z)$. Some different normalization assumptions in terms of different normalization methods are given as follows.

Assumption 1: Let b_1 be 1 or γ_1 be 1; another choice is to normalize any nonzero coefficient b_i or γ_i to 1.

Assumption 2: Let $B(1)$ be 1 or $B(1)/A(1)$ be 1, where $B(1)/A(1)$ is the gain of the linear dynamic subsystem.

Assumption 3: Let $b_1^2 + b_2^2 + \cdots + b_{n_b}^2$ be 1, $b_1 > 0$, or let $\gamma_1^2 + \gamma_2^2 + \cdots + \gamma_m^2$ be 1, $\gamma_1 > 0$.

The assumption scheme of the normalization in this paper is the first one, that is, $b_0 = 1$. Other choices of the normalization assumptions will lead to some different identification algorithms, for instance, $\gamma_1 = 1$.

In this section, an over-parameterization identification model is established for comparison. Let us discuss the over-parameterization identification model via the following analysis procedure. The noise-free output of the system in Figure 1 is given by

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

Uniting Equations (1) and (3) in accordance with the property of shift operator z^{-1} yields

$$\begin{aligned} x(t) &= [1 - A(z)]x(t) + B(z)\bar{u}(t) \\ &= -\sum_{i=1}^{n_a} a_i x(t-i) + \sum_{i=0}^{n_b} b_i \bar{u}(t-i) \\ &= -\sum_{i=1}^{n_a} a_i x(t-i) + \sum_{i=0}^{n_b} b_i \sum_{j=1}^{m} \gamma_j f_j(u(t-i)). \end{aligned} \tag{4}$$

Define the parameter vector $\boldsymbol{\vartheta}$ and the information vector $\boldsymbol{\varphi}_1(t)$ as

$$\boldsymbol{\vartheta} := [\boldsymbol{\vartheta}_0^{\mathrm{T}}, \boldsymbol{\theta}_0^{\mathrm{T}}, \boldsymbol{\theta}_1^{\mathrm{T}}, \cdots, \boldsymbol{\theta}_{n_b}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n_1}, \quad n_1 := n_a + m(n_b + 1), \tag{5}$$

$$\boldsymbol{\varphi}_1(t) := [\boldsymbol{\phi}^{\mathrm{T}}(t), \boldsymbol{\phi}_0^{\mathrm{T}}(t), \boldsymbol{\phi}_1^{\mathrm{T}}(t), \boldsymbol{\phi}_2^{\mathrm{T}}(t), \cdots, \boldsymbol{\phi}_{n_b}^{\mathrm{T}}(t)]^{\mathrm{T}} \in \mathbb{R}^{n_1}, \tag{6}$$

where the parameter vectors $\boldsymbol{\vartheta}_0$ and $\boldsymbol{\vartheta}_i$, and the information vectors $\boldsymbol{\phi}(t)$ and $\boldsymbol{\phi}_i(t)$ are defined as

$$\boldsymbol{\vartheta}_0 := [a_1, a_2, \cdots, a_{n_a}]^{\mathrm{T}} \in \mathbb{R}^{n_a}, \tag{7}$$

$$\boldsymbol{\theta}_i := [b_i\gamma_1, b_i\gamma_2, \cdots, b_i\gamma_m]^{\mathrm{T}} \in \mathbb{R}^m, \quad i = 0, 1, 2, \ldots, n_b, \tag{8}$$

$$\boldsymbol{\phi}(t) := [-x(t-1), -x(t-2), \cdots, -x(t-n_a)]^{\mathrm{T}} \in \mathbb{R}^{n_a}, \tag{9}$$

$$\boldsymbol{\phi}_i(t) := [f_1(u(t-i)), f_2(u(t-i)), \cdots, f_m(u(t-i))]^{\mathrm{T}} \in \mathbb{R}^m. \tag{10}$$

According to the aforementioned definitions, Equations (4) and (1) take the concise forms:

$$x(t) = \boldsymbol{\varphi}_1^{\mathrm{T}}(t)\boldsymbol{\vartheta}, \tag{11}$$

$$y(t) = x(t) + v(t) = \boldsymbol{\varphi}_1^{\mathrm{T}}(t)\boldsymbol{\vartheta} + v(t). \tag{12}$$

As a result, the over-parameterization identification model in (11)–(12) is established for the IN-OE system in (1)–(2). The proposed algorithms in this paper are based on this identification model in (11)–(12). Many identification methods are derived based on the identification models of the systems [38, 39, 40, 41] and can be used to estimate the parameters of other linear systems and nonlinear systems [42, 43, 44, 45, 46, 47] and can be applied to fields [48, 49, 50, 51, 52, 53] such as chemical process control systems.

Remark 1: The parameter vector $\boldsymbol{\vartheta}$ of the over-parameterization identification model in (12) contains all parameters of the whole system. However, there exists the product terms composed of the linear dynamical subsystem parameters b_i's and the static nonlinear part parameter γ_j's, which leads to a result that the number $n_1 := n_a + mn_b + m$ of the parameter vector $\boldsymbol{\vartheta}$ is larger than the number $n_a + n_b + 1 + m$ $(m, n_b \geqslant 2)$ of the all parameters of the original system, and generates many redundant parameter estimates. Hence, it is necessary to study new identification methods with smaller calculation amount.

3. The key term separation three-stage identification model

In this section, a key term separation identification model is established based on the key term separation technique. Furthermore, in order to improve the computational efficiency, the key term separation identification model is decomposed into three submodels with fewer variables and a key term separation three-stage identification model is established.

3.1. The key term separation identification model

Choosing $\bar{u}(t)$ as the key term to parameterize the noise-free output, and Equation (3) can be expressed as

$$\begin{aligned} x(t) &= -\sum_{i=1}^{n_a} a_i x(t-i) + \sum_{i=1}^{n_b} b_i \bar{u}(t-i) + b_0 \bar{u}(t) \\ &= -\sum_{i=1}^{n_a} a_i x(t-i) + \sum_{i=1}^{n_b} b_i \bar{u}(t-i) + \sum_{j=1}^{m} \gamma_j f_j(u(t)). \end{aligned} \tag{13}$$

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

$$\begin{aligned} \boldsymbol{a} &:= [a_1, a_2, a_3, \cdots, a_{n_a}]^{\mathrm{T}} \in \mathbb{R}^{n_a}, \\ \boldsymbol{b} &:= [b_1, b_2, b_3, \cdots, b_{n_b}]^{\mathrm{T}} \in \mathbb{R}^{n_b}, \\ \boldsymbol{\gamma} &:= [\gamma_1, \gamma_2, \gamma_3, \cdots, \gamma_m]^{\mathrm{T}} \in \mathbb{R}^{m}, \\ \boldsymbol{\varphi}_a(t) &:= [-x(t-1), -x(t-2), -x(t-3), \cdots, -x(t-n_a)]^{\mathrm{T}} \in \mathbb{R}^{n_a}, \\ \boldsymbol{\varphi}_b(t) &:= [\bar{u}(t-1), \bar{u}(t-2), \bar{u}(t-3), \cdots, \bar{u}(t-n_b)]^{\mathrm{T}} \in \mathbb{R}^{n_b}, \\ \boldsymbol{f}(t) &:= [f_1(u(t)), f_2(u(t)), f_3(u(t)), \cdots, f_m(u(t))]^{\mathrm{T}} \in \mathbb{R}^{m}. \end{aligned} \tag{14}$$

Using these definitions, Equations (2)–(3) take the concise forms:

$$\bar{u}(t) = \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{\gamma}, \tag{15}$$

$$x(t) = \boldsymbol{\varphi}_a^{\mathrm{T}}(t)\boldsymbol{a} + \boldsymbol{\varphi}_b^{\mathrm{T}}(t)\boldsymbol{b} + \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{\gamma}, \tag{16}$$

$$\begin{aligned} y(t) &= x(t) + v(t) \\ &= \boldsymbol{\varphi}_a^{\mathrm{T}}(t)\boldsymbol{a} + \boldsymbol{\varphi}_b^{\mathrm{T}}(t)\boldsymbol{b} + \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{\gamma} + v(t). \end{aligned} \tag{17}$$

Let $n := n_a + n_b + m$, define the parameter vector $\boldsymbol{\vartheta}$ and the information vector $\boldsymbol{\varphi}_2(t)$ as

$$\boldsymbol{\vartheta} := [\boldsymbol{a}^{\mathrm{T}}, \boldsymbol{b}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n}, \tag{18}$$

$$\boldsymbol{\varphi}_2(t) := [\boldsymbol{\varphi}_a^{\mathrm{T}}(t), \boldsymbol{\varphi}_b^{\mathrm{T}}(t), \boldsymbol{f}^{\mathrm{T}}(t)]^{\mathrm{T}} \in \mathbb{R}^{n}. \tag{19}$$

Then, Equation (17) can be rewritten as

$$y(t) = \boldsymbol{\varphi}_2^{\mathrm{T}}(t)\boldsymbol{\vartheta} + v(t). \tag{20}$$

Therefore, we can get the key term separation identification model in (20) of the IN-OE system in (2)–(1), where $\boldsymbol{\vartheta}$ is the parameter vector to be identified and contains all parameters of the original system.

Remark 2: Obviously, the dimension of the parameter vector $\boldsymbol{\vartheta}$ of the key term separation identification model in (20) is smaller than that of the over-parameterization identification model in (12). Hence, the identification algorithms based on the key term separation identification model have less computation than the identification algorithms based on the over-parameterization identification model.

3.2. The key term separation three-stage identification model

According to the key term separation identification model in (15)–(17), define three intermediate variables:

$$\begin{aligned} y_a(t) &:= y(t) - \boldsymbol{\varphi}_b^{\mathrm{T}}(t)\boldsymbol{b} - \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{\gamma}, \\ y_b(t) &:= y(t) - \boldsymbol{\varphi}_a^{\mathrm{T}}(t)\boldsymbol{a} - \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{\gamma}, \\ y_\gamma(t) &:= y(t) - \boldsymbol{\varphi}_a^{\mathrm{T}}(t)\boldsymbol{a} - \boldsymbol{\varphi}_b^{\mathrm{T}}(t)\boldsymbol{b}. \end{aligned}$$

Equation (17) can be decomposed into the following three fictitious subsystems:

$$y_a(t) = \boldsymbol{\varphi}_a^{\mathrm{T}}(t)\boldsymbol{a} + v(t), \tag{21}$$

$$y_b(t) = \boldsymbol{\varphi}_b^{\mathrm{T}}(t)\boldsymbol{b} + v(t), \tag{22}$$

$$y_\gamma(t) = \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{\gamma} + v(t). \tag{23}$$

The original parameter vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{\gamma}$ to be identified are included in these three fictitious subsystems, respectively.

Remark 3: On handling the key term separation three-stage identification models in (21)–(23), a difficulty arises because the sub-information vectors $\boldsymbol{\varphi}_a(t)$ and $\boldsymbol{\varphi}_b(t)$ are constructed by the unknown terms $x(t-i)$ and $\bar{u}(t-i)$. The approach taken here is to establish the auxiliary models, whose outputs are regarded as the estimates of these unknown terms, respectively, and to construct the sub-information vectors with their estimates $\hat{x}_{k-1}(t-i)$ and $\hat{\bar{u}}_{k-1}(t-i)$ at iteration $(k-1)$.

4. The key term separation auxiliary model three-stage gradient-based iterative algorithm

Consider the data from $t=1$ to $t=L$, and define the stacked output vectors $\boldsymbol{Y}(L)$, $\boldsymbol{Y}_a(L)$, $\boldsymbol{Y}_b(L)$ and $\boldsymbol{Y}_\gamma(L)$, and the stacked information matrices $\boldsymbol{\Phi}_a(L)$, $\boldsymbol{\Phi}_b(L)$ and $\boldsymbol{F}(L)$ as

$$\boldsymbol{Y}(L) := [y(1), y(2), y(3), \cdots, y(L)]^{\mathrm{T}} \in \mathbb{R}^L, \tag{24}$$

$$\begin{aligned}
\boldsymbol{Y}_a(L) &:= [y_a(1), y_a(2), y_a(3), \cdots, y_a(L)]^{\mathrm{T}} = \boldsymbol{Y}(L) - \boldsymbol{\Phi}_b(L)\boldsymbol{b} - \boldsymbol{F}(L)\boldsymbol{\gamma} \in \mathbb{R}^L,\\
\boldsymbol{Y}_b(L) &:= [y_b(1), y_b(2), y_b(3), \cdots, y_b(L)]^{\mathrm{T}} = \boldsymbol{Y}(L) - \boldsymbol{\Phi}_a(L)\boldsymbol{a} - \boldsymbol{F}(L)\boldsymbol{\gamma} \in \mathbb{R}^L,\\
\boldsymbol{Y}_\gamma(L) &:= [y_\gamma(1), y_\gamma(2), y_\gamma(3), \cdots, y_\gamma(L)]^{\mathrm{T}} = \boldsymbol{Y}(L) - \boldsymbol{\Phi}_a(L)\boldsymbol{a} - \boldsymbol{\Phi}_b(L)\boldsymbol{b} \in \mathbb{R}^L,\\
\boldsymbol{\Phi}_a(L) &:= [\boldsymbol{\varphi}_a(1), \boldsymbol{\varphi}_a(2), \boldsymbol{\varphi}_a(3), \cdots, \boldsymbol{\varphi}_a(L)]^{\mathrm{T}} \in \mathbb{R}^{L\times n_a},\\
\boldsymbol{\Phi}_b(L) &:= [\boldsymbol{\varphi}_b(1), \boldsymbol{\varphi}_b(2), \boldsymbol{\varphi}_b(3), \cdots, \boldsymbol{\varphi}_b(L)]^{\mathrm{T}} \in \mathbb{R}^{L\times n_b},\\
\boldsymbol{F}(L) &:= [\boldsymbol{f}(1), \boldsymbol{f}(2), \boldsymbol{f}(3), \cdots, \boldsymbol{f}(L)]^{\mathrm{T}} \in \mathbb{R}^{L\times m}.
\end{aligned} \tag{25}$$

Let the norm of a matrix (or a column vector) $\boldsymbol{X}$ is denoted by $\|bfX\|$ where $\|\boldsymbol{X}\|^2 := \mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$. According to Equations (21)–(23), define three criterion functions:

$$\begin{aligned}
J_a(\boldsymbol{a}) &:= \frac{1}{2}\|\boldsymbol{Y}_a(L) - \boldsymbol{\Phi}_a(L)\boldsymbol{a}\|^2,\\
J_b(\boldsymbol{b}) &:= \frac{1}{2}\|\boldsymbol{Y}_b(L) - \boldsymbol{\Phi}_b(L)\boldsymbol{b}\|^2,\\
J_\gamma(\boldsymbol{\gamma}) &:= \frac{1}{2}\|\boldsymbol{Y}_\gamma(L) - \boldsymbol{F}(L)\boldsymbol{\gamma}\|^2.
\end{aligned}$$

Let $k=1,2,3,\cdots$ be an iterative variable, $\hat{\boldsymbol{a}}_k \in \mathbb{R}^{n_a}$, $\hat{\boldsymbol{b}}_k \in \mathbb{R}^{n_b}$ and $\hat{\gamma}_k \in \mathbb{R}^m$ be the estimates of the parameter vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{\gamma}$ at iteration k, and $\mu_{1,k}$, $\mu_{2,k}$ and μ_3 be three convergence factors. Using the negative gradient search and minimizing $J_a(\boldsymbol{a})$, $J_b(\boldsymbol{b})$ and $J_\gamma(\boldsymbol{\gamma})$ lead to the following gradient-based iterative relations for computing $\hat{\boldsymbol{a}}_k$, $\hat{\boldsymbol{b}}_k$ and $\hat{\gamma}_k$:

$$\begin{aligned}
\hat{\boldsymbol{a}}_k &= \hat{\boldsymbol{a}}_{k-1} - \mu_{1,k}\mathrm{grad}[J_a(\hat{\boldsymbol{a}}_{k-1})]\\
&= \hat{\boldsymbol{a}}_{k-1} + \mu_{1,k}\boldsymbol{\Phi}_a^{\mathrm{T}}(L)[\boldsymbol{Y}_a(L) - \boldsymbol{\Phi}_a(L)\hat{\boldsymbol{a}}_{k-1}]\\
&= \hat{\boldsymbol{a}}_{k-1} + \mu_{1,k}\boldsymbol{\Phi}_a^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \boldsymbol{\Phi}_b(L)\hat{\boldsymbol{b}}_{k-1} - \boldsymbol{F}(L)\hat{\gamma}_{k-1} - \boldsymbol{\Phi}_a(L)\hat{\boldsymbol{a}}_{k-1}]\\
&= \hat{\boldsymbol{a}}_{k-1} + \mu_{1,k}\boldsymbol{\Phi}_a^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \boldsymbol{\Phi}_a(L)\hat{\boldsymbol{a}}_{k-1} - \boldsymbol{\Phi}_b(L)\hat{\boldsymbol{b}}_{k-1} - \boldsymbol{F}(L)\hat{\gamma}_{k-1}],
\end{aligned} \tag{26}$$

$$\begin{aligned}
\hat{\boldsymbol{b}}_k &= \hat{\boldsymbol{b}}_{k-1} - \mu_{2,k}\mathrm{grad}[J_b(\hat{\boldsymbol{b}}_{k-1})]\\
&= \hat{\boldsymbol{b}}_{k-1} + \mu_{2,k}\boldsymbol{\Phi}_b^{\mathrm{T}}(L)[\boldsymbol{Y}_b(L) - \boldsymbol{\Phi}_b(L)\hat{\boldsymbol{b}}_{k-1}]\\
&= \hat{\boldsymbol{b}}_{k-1} + \mu_{2,k}\boldsymbol{\Phi}_b^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \boldsymbol{\Phi}_a(L)\hat{\boldsymbol{a}}_{k-1} - \boldsymbol{F}(L)\hat{\gamma}_{k-1} - \boldsymbol{\Phi}_b(L)\hat{\boldsymbol{b}}_{k-1}]\\
&= \hat{\boldsymbol{b}}_{k-1} + \mu_{2,k}\boldsymbol{\Phi}_b^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \boldsymbol{\Phi}_a(L)\hat{\boldsymbol{a}}_{k-1} - \boldsymbol{\Phi}_b(L)\hat{\boldsymbol{b}}_{k-1} - \boldsymbol{F}(L)\hat{\gamma}_{k-1}],
\end{aligned} \tag{27}$$

$$\begin{aligned}
\hat{\gamma}_k &= \hat{\gamma}_{k-1} - \mu_3\mathrm{grad}[J_\gamma(\hat{\gamma}_{k-1})]\\
&= \hat{\gamma}_{k-1} + \mu_3\boldsymbol{F}^{\mathrm{T}}(L)[\boldsymbol{Y}_\gamma(L) - \boldsymbol{F}(L)\hat{\gamma}_{k-1}]\\
&= \hat{\gamma}_{k-1} + \mu_3\boldsymbol{F}^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \boldsymbol{\Phi}_a(L)\hat{\boldsymbol{a}}_{k-1} - \boldsymbol{\Phi}_b(L)\hat{\boldsymbol{b}}_{k-1} - \boldsymbol{F}(L)\hat{\gamma}_{k-1}].
\end{aligned} \tag{28}$$

Equations (26)–(28) cannot figure out the parameter estimates $\hat{\boldsymbol{a}}_k$, $\hat{\boldsymbol{b}}_k$ and $\hat{\gamma}_k$ because the stacked information matrices $\boldsymbol{\Phi}_a(L)$ and $\boldsymbol{\Phi}_b(L)$ contain the unknown entries $x(t-i)$ and $\bar{u}(t-i)$. The solution is to replace these unknown entries in the stacked information matrices $\boldsymbol{\Phi}_a(L)$ and $\boldsymbol{\Phi}_b(L)$ with their corresponding estimates $\hat{x}_{k-1}(t-i)$ and $\hat{\bar{u}}_{k-1}(t-i)$ at iteration $(k-1)$. Define the estimates $\hat{\boldsymbol{\varphi}}_{a,k}(t)$, $\hat{\boldsymbol{\varphi}}_{b,k}(t)$, $\hat{\boldsymbol{\Phi}}_{a,k}(L)$ and $\hat{\boldsymbol{\Phi}}_{b,k}(L)$ of $\boldsymbol{\varphi}_a(t)$, $\boldsymbol{\varphi}_b(t)$, $\boldsymbol{\Phi}_a(L)$ and $\boldsymbol{\Phi}_b(L)$ as

$$\hat{\boldsymbol{\varphi}}_{a,k}(t) := [-\hat{x}_{k-1}(t-1), -\hat{x}_{k-1}(t-2), \cdots, -\hat{x}_{k-1}(t-n_a)]^{\mathrm{T}} \in \mathbb{R}^{n_a}, \tag{29}$$

$$\hat{\boldsymbol{\varphi}}_{b,k}(t) := [\hat{\bar{u}}_{k-1}(t-1), \hat{\bar{u}}_{k-1}(t-2), \cdots, \hat{\bar{u}}_{k-1}(t-n_b)]^{\mathrm{T}} \in \mathbb{R}^{n_b}, \tag{30}$$

$$\hat{\boldsymbol{\Phi}}_{a,k}(L) := [\hat{\boldsymbol{\varphi}}_{a,k}(1), \hat{\boldsymbol{\varphi}}_{a,k}(2), \hat{\boldsymbol{\varphi}}_{a,k}(3), \cdots, \hat{\boldsymbol{\varphi}}_{a,k}(L)]^{\mathrm{T}} \in \mathbb{R}^{L\times n_a}, \tag{31}$$

$$\hat{\boldsymbol{\Phi}}_{b,k}(L) := [\hat{\boldsymbol{\varphi}}_{b,k}(1), \hat{\boldsymbol{\varphi}}_{b,k}(2), \hat{\boldsymbol{\varphi}}_{b,k}(3), \cdots, \hat{\boldsymbol{\varphi}}_{b,k}(L)]^{\mathrm{T}} \in \mathbb{R}^{L\times n_b}. \tag{32}$$

Replacing $\boldsymbol{\Phi}_a(L)$ and $\boldsymbol{\Phi}_b(L)$ in (26)–(28) with their estimates $\hat{\boldsymbol{\Phi}}_{a,k}(L)$ and $\hat{\boldsymbol{\Phi}}_{b,k}(L)$ yields the following gradient-based iterative algorithm for estimating $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{\gamma}$:

$$\begin{aligned}
\hat{\boldsymbol{a}}_k &= \hat{\boldsymbol{a}}_{k-1} + \mu_{1,k}\hat{\boldsymbol{\Phi}}_{a,k}^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \hat{\boldsymbol{\Phi}}_{a,k}(L)\hat{\boldsymbol{a}}_{k-1} - \hat{\boldsymbol{\Phi}}_{b,k}(L)\hat{\boldsymbol{b}}_{k-1} - \boldsymbol{F}(L)\hat{\gamma}_{k-1}] \qquad (33)\\
&= [\boldsymbol{I}_L - \mu_{1,k}\hat{\boldsymbol{\Phi}}_{a,k}^{\mathrm{T}}(L)\hat{\boldsymbol{\Phi}}_{a,k}(L)]\hat{\boldsymbol{a}}_{k-1} + \mu_{1,k}\hat{\boldsymbol{\Phi}}_{a,k}^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \hat{\boldsymbol{\Phi}}_{b,k}(L)\hat{\boldsymbol{b}}_{k-1} - \boldsymbol{F}(L)\hat{\gamma}_{k-1}],
\end{aligned}$$

$$\hat{\boldsymbol{b}}_k = \hat{\boldsymbol{b}}_{k-1} + \mu_{2,k}\hat{\boldsymbol{\Phi}}_{b,k}^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \hat{\boldsymbol{\Phi}}_{a,k}(L)\hat{\boldsymbol{a}}_{k-1} - \hat{\boldsymbol{\Phi}}_{b,k}(L)\hat{\boldsymbol{b}}_{k-1} - \boldsymbol{F}(L)\hat{\gamma}_{k-1}] \tag{34}$$

$$= [\boldsymbol{I}_L - \mu_{2,k}\hat{\boldsymbol{\Phi}}^{\mathrm{T}}_{b,k}(L)\hat{\boldsymbol{\Phi}}_{b,k}(L)]\hat{\boldsymbol{b}}_{k-1} + \mu_{2,k}\hat{\boldsymbol{\Phi}}^{\mathrm{T}}_{b,k}(L)[\boldsymbol{Y}(L) - \hat{\boldsymbol{\Phi}}_{a,k}(L)\hat{\boldsymbol{a}}_{k-1} - \boldsymbol{F}(L)\hat{\boldsymbol{\gamma}}_{k-1}],$$

$$\hat{\boldsymbol{\gamma}}_k = \hat{\boldsymbol{\gamma}}_{k-1} + \mu_3\boldsymbol{F}^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \hat{\boldsymbol{\Phi}}_{a,k}(L)\hat{\boldsymbol{a}}_{k-1} - \hat{\boldsymbol{\Phi}}_{b,k}(L)\hat{\boldsymbol{b}}_{k-1} - \boldsymbol{F}(L)\hat{\boldsymbol{\gamma}}_{k-1}] \tag{35}$$

$$= [\boldsymbol{I}_L - \mu_3\boldsymbol{F}^{\mathrm{T}}(L)\boldsymbol{F}(L)]\hat{\boldsymbol{\gamma}}_{k-1} + \mu_3\boldsymbol{F}^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \hat{\boldsymbol{\Phi}}_{a,k}(L)\hat{\boldsymbol{a}}_{k-1} - \hat{\boldsymbol{\Phi}}_{b,k}(L)\hat{\boldsymbol{b}}_{k-1}],$$

The above equations can be viewed as three discrete-time systems. In order to guarantee the convergence of the parameter estimates $\hat{\boldsymbol{a}}_k$, $\hat{\boldsymbol{b}}_k$ and $\hat{\boldsymbol{\gamma}}_k$, all characteristic values of the matrices $[\boldsymbol{I}_L - \mu_{1,k}\hat{\boldsymbol{\Phi}}^{\mathrm{T}}_{a,k}(L)\hat{\boldsymbol{\Phi}}_{a,k}(L)]$, $[\boldsymbol{I}_L - \mu_{2,k}\hat{\boldsymbol{\Phi}}^{\mathrm{T}}_{b,k}(L)\hat{\boldsymbol{\Phi}}_{b,k}(L)]$ and $[\boldsymbol{I}_L - \mu_3\boldsymbol{F}^{\mathrm{T}}(L)\boldsymbol{F}(L)]$ must be inside the unit circle. The conservative choices of $\mu_{1,k}$, $\mu_{2,k}$ and μ_3 are to satisfy

$$\mu_{1,k} \leqslant \frac{2}{\lambda_{\max}[\hat{\boldsymbol{\Phi}}^{\mathrm{T}}_{a,k}(L)\hat{\boldsymbol{\Phi}}_{a,k}(L)]} = 2\lambda^{-1}_{\max}[\hat{\boldsymbol{\Phi}}^{\mathrm{T}}_{a,k}(L)\hat{\boldsymbol{\Phi}}_{a,k}(L)], \tag{36}$$

$$\mu_{2,k} \leqslant \frac{2}{\lambda_{\max}[\hat{\boldsymbol{\Phi}}^{\mathrm{T}}_{b,k}(L)\hat{\boldsymbol{\Phi}}_{b,k}(L)]} = 2\lambda^{-1}_{\max}[\hat{\boldsymbol{\Phi}}^{\mathrm{T}}_{b,k}(L)\hat{\boldsymbol{\Phi}}_{b,k}(L)], \tag{37}$$

$$\mu_3 \leqslant \frac{2}{\lambda_{\max}[\boldsymbol{F}^{\mathrm{T}}(L)\boldsymbol{F}(L)]} = 2\lambda^{-1}_{\max}[\boldsymbol{F}^{\mathrm{T}}(L)\boldsymbol{F}(L)], \tag{38}$$

where $\lambda_{\max}[\boldsymbol{X}]$ denotes the maximum eigenvalue of the square matrix $\boldsymbol{X}$. Calculating the convergence factors through Equations (36)–(38) is very complicated, so the convergence factors can be simply taken as

$$\mu_{1,k} \leqslant 2\|\hat{\boldsymbol{\Phi}}_{a,k}(L)\|^{-2}, \tag{39}$$

$$\mu_{2,k} \leqslant 2\|\hat{\boldsymbol{\Phi}}_{b,k}(L)\|^{-2}, \tag{40}$$

$$\mu_3 \leqslant 2\|\boldsymbol{F}(L)\|^{-2}, \tag{41}$$

Based on Equations (15)–(16), the estimates $\hat{\hat{u}}_k(t)$ and $\hat{x}_k(t)$ at iteration k can be calculated through two auxiliary models:

$$\hat{\hat{u}}_k(t) := \boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{\gamma}}_k, \tag{42}$$

$$\hat{x}_k(t) := \hat{\boldsymbol{\varphi}}^{\mathrm{T}}_{a,k}(t)\hat{\boldsymbol{a}}_k + \hat{\boldsymbol{\varphi}}^{\mathrm{T}}_{b,k}(t)\hat{\boldsymbol{b}}_k + \hat{\hat{u}}_k(t), \tag{43}$$

Combining Equations (24)–(43), we can obtain the following key term separation auxiliary model three-stage gradient-based iterative (KT-AM-3S-GI) algorithm [37]:

$$\hat{\boldsymbol{a}}_k = \hat{\boldsymbol{a}}_{k-1} + \mu_{1,k}\hat{\boldsymbol{\Phi}}^{\mathrm{T}}_{a,k}(L)[\boldsymbol{Y}(L) - \hat{\boldsymbol{\Phi}}_{a,k}(L)\hat{\boldsymbol{a}}_{k-1} - \hat{\boldsymbol{\Phi}}_{b,k}(L)\hat{\boldsymbol{b}}_{k-1} - \boldsymbol{F}(L)\hat{\boldsymbol{\gamma}}_{k-1}], \tag{44}$$

$$\mu_{1,k} \leqslant 2\lambda^{-1}_{\max}[\hat{\boldsymbol{\Phi}}^{\mathrm{T}}_{a,k}(L)\hat{\boldsymbol{\Phi}}_{a,k}(L)], \quad \text{or} \quad \mu_{1,k} \leqslant 2\|\hat{\boldsymbol{\Phi}}_{a,k}(L)\|^{-2}, \tag{45}$$

$$\hat{\boldsymbol{b}}_k = \hat{\boldsymbol{b}}_{k-1} + \mu_{2,k}\hat{\boldsymbol{\Phi}}^{\mathrm{T}}_{b,k}(L)[\boldsymbol{Y}(L) - \hat{\boldsymbol{\Phi}}_{a,k}(L)\hat{\boldsymbol{a}}_{k-1} - \hat{\boldsymbol{\Phi}}_{b,k}(L)\hat{\boldsymbol{b}}_{k-1} - \boldsymbol{F}(L)\hat{\boldsymbol{\gamma}}_{k-1}], \tag{46}$$

$$\mu_{2,k} \leqslant 2\lambda^{-1}_{\max}[\hat{\boldsymbol{\Phi}}^{\mathrm{T}}_{b,k}(L)\hat{\boldsymbol{\Phi}}_{b,k}(L)], \quad \text{or} \quad \mu_{2,k} \leqslant 2\|\hat{\boldsymbol{\Phi}}_{b,k}(L)\|^{-2}, \tag{47}$$

$$\hat{\boldsymbol{\gamma}}_k = \hat{\boldsymbol{\gamma}}_{k-1} + \mu_3\boldsymbol{F}^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \hat{\boldsymbol{\Phi}}_{a,k}(L)\hat{\boldsymbol{a}}_{k-1} - \hat{\boldsymbol{\Phi}}_{b,k}(L)\hat{\boldsymbol{b}}_{k-1} - \boldsymbol{F}(L)\hat{\boldsymbol{\gamma}}_{k-1}], \tag{48}$$

$$\mu_3 \leqslant 2\lambda^{-1}_{\max}[\boldsymbol{F}^{\mathrm{T}}(L)\boldsymbol{F}(L)], \quad \text{or} \quad \mu_3 \leqslant 2\|\boldsymbol{F}(L)\|^{-2}. \tag{49}$$

$$\boldsymbol{Y}(L) = [y(1), y(2), y(3), \cdots, y(L)]^{\mathrm{T}}, \tag{50}$$

$$\hat{\boldsymbol{\Phi}}_{a,k}(L) = [\hat{\boldsymbol{\varphi}}_{a,k}(1), \hat{\boldsymbol{\varphi}}_{a,k}(2), \hat{\boldsymbol{\varphi}}_{a,k}(3), \cdots, \hat{\boldsymbol{\varphi}}_{a,k}(L)]^{\mathrm{T}}, \tag{51}$$

$$\hat{\boldsymbol{\Phi}}_{b,k}(L) = [\hat{\boldsymbol{\varphi}}_{b,k}(1), \hat{\boldsymbol{\varphi}}_{b,k}(2), \hat{\boldsymbol{\varphi}}_{b,k}(3), \cdots, \hat{\boldsymbol{\varphi}}_{b,k}(L)]^{\mathrm{T}}, \tag{52}$$

$$\boldsymbol{F}(L) = [\boldsymbol{f}(1), \boldsymbol{f}(2), \boldsymbol{f}(3), \cdots, \boldsymbol{f}(L)]^{\mathrm{T}}, \tag{53}$$

$$\hat{\boldsymbol{\varphi}}_{a,k}(t) = [-\hat{x}_{k-1}(t-1), -\hat{x}_{k-1}(t-2), \cdots, -\hat{x}_{k-1}(t-n_a)]^{\mathrm{T}}, \tag{54}$$

$$\hat{\boldsymbol{\varphi}}_{b,k}(t) = [\hat{\hat{u}}_{k-1}(t-1), \hat{\hat{u}}_{k-1}(t-2), \cdots, \hat{\hat{u}}_{k-1}(t-n_b)]^{\mathrm{T}}, \tag{55}$$

$$\boldsymbol{f}(t) = [f_1(u(t)), f_2(u(t)), \cdots, f_m(u(t))]^{\mathrm{T}}, \tag{56}$$

$$\hat{\hat{u}}_k(t) = \boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{\gamma}}_k, \tag{57}$$

$$\hat{x}_k(t) = \hat{\boldsymbol{\varphi}}^{\mathrm{T}}_{a,k}(t)\hat{\boldsymbol{a}}_k + \hat{\boldsymbol{\varphi}}^{\mathrm{T}}_{b,k}(t)\hat{\boldsymbol{b}}_k + \hat{\hat{u}}_k(t), \tag{58}$$

$$\hat{\boldsymbol{a}}_k = [\hat{a}_{1,k}, \hat{a}_{2,k}, \cdots, \hat{a}_{n_a,k}]^{\mathrm{T}}, \tag{59}$$

$$\hat{\boldsymbol{b}}_k = [\hat{b}_{1,k}, \hat{b}_{2,k}, \cdots, \hat{b}_{n_b,k}]^{\mathrm{T}}, \tag{60}$$

$$\hat{\boldsymbol{\gamma}}_k = [\hat{\gamma}_{1,k}, \hat{\gamma}_{2,k}, \cdots, \hat{\gamma}_{m,k}]^{\mathrm{T}}. \tag{61}$$

The steps involved in the KT-AM-3S-GI algorithm in (44)–(61) are listed as follows.

1. For $t \leqslant 0$, all variables are set to zero. Set the data length L ($L \gg n$), the parameter estimation accuracy ε and the basis function $f_j(*)$. Let $k = 1$, set $\hat{x}_0(t) = 1/p_0$ and $\hat{\hat{u}}_0(t) = 1/p_0$, $t = 1, 2, \cdots, L$. Preset $\hat{\boldsymbol{a}}_0 = \boldsymbol{1}_{n_a}/p_0$, $\hat{\boldsymbol{b}}_0 = \boldsymbol{1}_{n_b}/p_0$ and $\hat{\boldsymbol{\gamma}}_0 = \boldsymbol{1}_m/p_0$, $p_0 = 10^6$.

2. Collect the observation data $u(t)$ and $y(t)$, and construct the sub-information vector $\boldsymbol{f}(t)$ using (56), $t = 1, 2, \cdots, L$. Construct the stacked output vector $\boldsymbol{Y}(L)$ and the stacked input information matrix $\boldsymbol{F}(L)$ using (50) and (53). Select a large step-size μ_3 by (49).
3. Construct the sub-information vectors $\hat{\boldsymbol{\varphi}}_{a,k}(t)$ and $\hat{\boldsymbol{\varphi}}_{b,k}(t)$ using (54)–(55), $t = 1, 2, \cdots, L$. Construct the stacked information matrices $\hat{\boldsymbol{\Phi}}_{a,k}(L)$ and $\hat{\boldsymbol{\Phi}}_{b,k}(L)$ using (51) and (52).
4. Select a large step-sizes $\mu_{1,k}$ and $\mu_{2,k}$ by (45) and (47), and update the parameter estimation vectors $\hat{\boldsymbol{a}}_k$, $\hat{\boldsymbol{b}}_k$ and $\hat{\boldsymbol{\gamma}}_k$ using (44), (46) and (48).
5. Calculate the outputs $\hat{\bar{u}}_k(t)$ and $\hat{x}_k(t)$ of the auxiliary models by using (57)–(58).
6. Compare $\hat{\boldsymbol{a}}_k$ with $\hat{\boldsymbol{a}}_{k-1}$, $\hat{\boldsymbol{b}}_k$ with $\hat{\boldsymbol{b}}_{k-1}$, $\hat{\boldsymbol{\gamma}}_k$ with $\hat{\boldsymbol{\gamma}}_{k-1}$: if

$$Error := \|\hat{\boldsymbol{a}}_k - \hat{\boldsymbol{a}}_{k-1}\| + \|\hat{\boldsymbol{b}}_k - \hat{\boldsymbol{b}}_{k-1}\| + \|\hat{\boldsymbol{\gamma}}_k - \hat{\boldsymbol{\gamma}}_{k-1}\| > \varepsilon,$$

increase k by 1 and go to Step 3; otherwise obtain the parameter estimates $\hat{\boldsymbol{a}}_k$, $\hat{\boldsymbol{b}}_k$ and $\hat{\boldsymbol{\gamma}}_k$, and terminate this procedure.

5. The key term separation auxiliary model three-stage least squares-based iterative algorithm

In this section, a key term separation auxiliary model three-stage least squares-based iterative algorithm is derived for the parameter estimation of the IN-OE systems. Differently from the gradient-based iterative algorithm [54], the least squares-based iterative algorithm needs fewer iterations to achieve the same parameter estimation accuracy.

Minimizing the criterion functions $J_a(\boldsymbol{a})$, $J_b(\boldsymbol{b})$ and $J_\gamma(\boldsymbol{\gamma})$ and setting the derivatives to be zero give

$$\begin{aligned}
\frac{\partial J_a(\boldsymbol{a})}{\partial \boldsymbol{a}} &= -\boldsymbol{\Phi}_a^{\mathrm{T}}(L)[\boldsymbol{Y}_a(L) - \boldsymbol{\Phi}_a(L)\boldsymbol{a}] = \mathbf{0},\\
\frac{\partial J_b(\boldsymbol{b})}{\partial \boldsymbol{b}} &= -\boldsymbol{\Phi}_b^{\mathrm{T}}(L)[\boldsymbol{Y}_b(L) - \boldsymbol{\Phi}_b(L)\boldsymbol{b}] = \mathbf{0},\\
\frac{\partial J_\gamma(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} &= -\boldsymbol{F}^{\mathrm{T}}(L)[\boldsymbol{Y}_\gamma(L) - \boldsymbol{F}(L)\boldsymbol{\gamma}] = \mathbf{0},
\end{aligned}$$

or

$$[\boldsymbol{\Phi}_a^{\mathrm{T}}(L)\boldsymbol{\Phi}_a(L)]\boldsymbol{a} = \boldsymbol{\Phi}_a^{\mathrm{T}}(L)\boldsymbol{Y}_a(L), \tag{62}$$

$$[\boldsymbol{\Phi}_b^{\mathrm{T}}(L)\boldsymbol{\Phi}_b(L)]\,\boldsymbol{b} = \boldsymbol{\Phi}_b^{\mathrm{T}}(L)\boldsymbol{Y}_b(L), \tag{63}$$

$$[\boldsymbol{F}^{\mathrm{T}}(L)\boldsymbol{F}(L)]\,\boldsymbol{\gamma} = \boldsymbol{F}^{\mathrm{T}}(L)\boldsymbol{Y}_\gamma(L). \tag{64}$$

Assume that the data length L is much greater than the dimension of $\boldsymbol{\vartheta}$, i.e., $L \gg n$, when the information vectors $\boldsymbol{\varphi}_a(t)$, $\boldsymbol{\varphi}_b(t)$ and $\boldsymbol{f}(t)$ are persistently exciting and the matrices $[\boldsymbol{\Phi}_a^{\mathrm{T}}(L)\boldsymbol{\Phi}_a(L)]$, $[\boldsymbol{\Phi}_b^{\mathrm{T}}(L)\boldsymbol{\Phi}_b(L)]$ and $[\boldsymbol{F}^{\mathrm{T}}(L)\boldsymbol{F}(L)]$ are invertible, we can obtained the least squares-based iterative relations for computing $\hat{\boldsymbol{a}}_k$, $\hat{\boldsymbol{b}}_k$ and $\hat{\boldsymbol{\gamma}}_k$:

$$\begin{aligned}
\hat{\boldsymbol{a}}_k &= [\boldsymbol{\Phi}_a^{\mathrm{T}}(L)\boldsymbol{\Phi}_a(L)]^{-1}\boldsymbol{\Phi}_a^{\mathrm{T}}(L)\boldsymbol{Y}_a(L)\\
&= [\boldsymbol{\Phi}_a^{\mathrm{T}}(L)\boldsymbol{\Phi}_a(L)]^{-1}\boldsymbol{\Phi}_a^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \boldsymbol{\Phi}_b(L)\boldsymbol{b} - \boldsymbol{F}(L)\boldsymbol{\gamma}], && (65)\\
\hat{\boldsymbol{b}}_k &= [\boldsymbol{\Phi}_b^{\mathrm{T}}(L)\boldsymbol{\Phi}_b(L)]^{-1}\boldsymbol{\Phi}_b^{\mathrm{T}}(L)\boldsymbol{Y}_b(L)\\
&= [\boldsymbol{\Phi}_b^{\mathrm{T}}(L)\boldsymbol{\Phi}_b(L)]^{-1}\boldsymbol{\Phi}_b^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \boldsymbol{\Phi}_a(L)\boldsymbol{a} - \boldsymbol{F}(L)\boldsymbol{\gamma}], && (66)\\
\hat{\boldsymbol{\gamma}}_k &= [\boldsymbol{F}^{\mathrm{T}}(L)\boldsymbol{F}(L)]^{-1}\boldsymbol{F}^{\mathrm{T}}(L)\boldsymbol{Y}_\gamma(L)\\
&= [\boldsymbol{F}^{\mathrm{T}}(L)\boldsymbol{F}(L)]^{-1}\boldsymbol{F}^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \boldsymbol{\Phi}_a(L)\boldsymbol{a} - \boldsymbol{\Phi}_b(L)\boldsymbol{b}]. && (67)
\end{aligned}$$

However, Equations (65)–(67) cannot calculate the estimates $\hat{\boldsymbol{a}}_k$, $\hat{\boldsymbol{b}}_k$ and $\hat{\boldsymbol{\gamma}}_k$ because these equations contain the unknown parameter vectors $\boldsymbol{b}$, $\boldsymbol{\gamma}$ and $\boldsymbol{a}$ and the unknown information matrices $\boldsymbol{\Phi}_b(L)$ and $\boldsymbol{\Phi}_a(L)$. The approach taken here is to replace these unknown terms with their estimates at the previous iteration based on the auxiliary model identification idea. Then, we can obtain the following key term separation auxiliary model three-stage least squares-based iterative (KT-AM-3S-LSI) algorithm [37]:

$$\hat{\boldsymbol{a}}_k = [\hat{\boldsymbol{\Phi}}_{a,k}^{\mathrm{T}}(L)\hat{\boldsymbol{\Phi}}_{a,k}(L)]^{-1}\hat{\boldsymbol{\Phi}}_{a,k}^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \hat{\boldsymbol{\Phi}}_{b,k}(L)\hat{\boldsymbol{b}}_{k-1} - \boldsymbol{F}(L)\hat{\boldsymbol{\gamma}}_{k-1}], \tag{68}$$

$$\hat{\boldsymbol{b}}_k = [\hat{\boldsymbol{\Phi}}_{b,k}^{\mathrm{T}}(L)\hat{\boldsymbol{\Phi}}_{b,k}(L)]^{-1}\hat{\boldsymbol{\Phi}}_{b,k}^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \hat{\boldsymbol{\Phi}}_{a,k}(L)\hat{\boldsymbol{a}}_{k-1} - \boldsymbol{F}(L)\hat{\boldsymbol{\gamma}}_{k-1}], \tag{69}$$

$$\hat{\boldsymbol{\gamma}}_k = [\boldsymbol{F}^{\mathrm{T}}(L)\boldsymbol{F}(L)]^{-1}\boldsymbol{F}^{\mathrm{T}}(L)[\boldsymbol{Y}(L) - \hat{\boldsymbol{\Phi}}_{a,k}(L)\hat{\boldsymbol{a}}_{k-1} - \hat{\boldsymbol{\Phi}}_{b,k}(L)\hat{\boldsymbol{b}}_{k-1}], \tag{70}$$

$$\boldsymbol{Y}(L) = [y(1), y(2), y(3), \cdots, y(L)]^{\mathrm{T}}, \tag{71}$$

$$\hat{\boldsymbol{\Phi}}_{a,k}(L) = [\hat{\boldsymbol{\varphi}}_{a,k}(1), \hat{\boldsymbol{\varphi}}_{a,k}(2), \cdots, \hat{\boldsymbol{\varphi}}_{a,k}(L)]^{\mathrm{T}}, \tag{72}$$

$$\hat{\boldsymbol{\Phi}}_{b,k}(L) = [\hat{\boldsymbol{\varphi}}_{b,k}(1), \hat{\boldsymbol{\varphi}}_{b,k}(2), \cdots, \hat{\boldsymbol{\varphi}}_{b,k}(L)]^{\mathrm{T}}, \tag{73}$$

$$\boldsymbol{F}(L) = [\boldsymbol{f}(1), \boldsymbol{f}(2), \cdots, \boldsymbol{f}(L)]^{\mathrm{T}}, \tag{74}$$

$$\hat{\boldsymbol{\varphi}}_{a,k}(t) = [-\hat{x}_{k-1}(t-1), -\hat{x}_{k-1}(t-2), \cdots, -\hat{x}_{k-1}(t-n_a)]^{\mathrm{T}}, \tag{75}$$

$$\hat{\boldsymbol{\varphi}}_{b,k}(t) = [\hat{\bar{u}}_{k-1}(t-1), \hat{\bar{u}}_{k-1}(t-2), \cdots, \hat{\bar{u}}_{k-1}(t-n_b)]^{\mathrm{T}}, \tag{76}$$

$$\boldsymbol{f}(t) = [f_1(u(t)), f_2(u(t)), \cdots, f_m(u(t))]^{\mathrm{T}}, \tag{77}$$

$$\hat{\bar{u}}_k(t) = \boldsymbol{f}^{\mathrm{T}}(t)\hat{\boldsymbol{\gamma}}_k, \tag{78}$$

$$\hat{x}_k(t) = \hat{\boldsymbol{\varphi}}_{a,k}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}_k + \hat{\boldsymbol{\varphi}}_{b,k}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}_k + \hat{\bar{u}}_k(t), \tag{79}$$

$$\hat{\boldsymbol{a}}_k = [\hat{a}_{1,k}, \hat{a}_{2,k}, \cdots, \hat{a}_{n_a,k}]^{\mathrm{T}}, \tag{80}$$

$$\hat{\boldsymbol{b}}_k = [\hat{b}_{1,k}, \hat{b}_{2,k}, \cdots, \hat{b}_{n_b,k}]^{\mathrm{T}}, \tag{81}$$

$$\hat{\boldsymbol{\gamma}}_k = [\hat{\gamma}_{1,k}, \hat{\gamma}_{2,k}, \cdots, \hat{\gamma}_{m,k}]^{\mathrm{T}}. \tag{82}$$

The proposed approaches in the paper can combine some mathematical tools and statistical strategies [55, 56, 57, 58, 59, 60] and some identification algorithms [61-67] to study the performances of the parameter estimation algorithms of other linear stochastic systems and nonlinear stochastic systems with different structures and disturbance noises [68-75] and can be applied to literatures [76-82] such as paper-making systems. The steps involved in the KT-AM-3S-LSI algorithm in (68)–(82) are listed in the following.

1. For $t \leqslant 0$, all variables are set to zero. Set the data length L ($L \gg n$), the parameter estimation accuracy ε and the basis function $f_j(*)$. Let $k = 1$, $\hat{x}_0(t)$ and $\hat{\bar{u}}_0(t)$ be random numbers, $t = 1, 2, \cdots, L$. Preset $\hat{\boldsymbol{a}}_0 = \mathbf{1}_{n_a}/p_0$, $\hat{\boldsymbol{b}}_0 = \mathbf{1}_{n_b}/p_0$ and $\hat{\boldsymbol{\gamma}}_0 = \mathbf{1}_m/p_0$, $p_0 = 10^6$.
2. Collect the observation data $u(t)$ and $y(t)$, and construct the sub-information vector $\boldsymbol{f}(t)$ using (77), $t = 1, 2, \cdots, L$. Construct the stacked output vector $\boldsymbol{Y}(L)$ and the stacked input information matrix $\boldsymbol{F}(L)$ using (71) and (74).
3. Construct the sub-information vectors $\hat{\boldsymbol{\varphi}}_{a,k}(t)$ and $\hat{\boldsymbol{\varphi}}_{b,k}(t)$ using (75)–(76), $t = 1, 2, \cdots, L$. Construct the stacked information matrices $\hat{\boldsymbol{\Phi}}_{a,k}(L)$ and $\hat{\boldsymbol{\Phi}}_{b,k}(L)$ using (72) and (73).
4. Update the parameter estimation vectors $\hat{\boldsymbol{a}}_k$, $\hat{\boldsymbol{b}}_k$ and $\hat{\boldsymbol{\gamma}}_k$ using (68)–(70).
5. Calculate the outputs $\hat{\bar{u}}_k(t)$ and $\hat{x}_k(t)$ of the auxiliary models using (78)–(79).
6. Compare $\hat{\boldsymbol{a}}_k$ with $\hat{\boldsymbol{a}}_{k-1}$, $\hat{\boldsymbol{b}}_k$ with $\hat{\boldsymbol{b}}_{k-1}$, $\hat{\boldsymbol{\gamma}}_k$ with $\hat{\boldsymbol{\gamma}}_{k-1}$: if

$$Error := \|\hat{\boldsymbol{a}}_k - \hat{\boldsymbol{a}}_{k-1}\| + \|\hat{\boldsymbol{b}}_k - \hat{\boldsymbol{b}}_{k-1}\| + \|\hat{\boldsymbol{\gamma}}_k - \hat{\boldsymbol{\gamma}}_{k-1}\| > \varepsilon,$$

   increase k by 1 and go to Step 3; otherwise obtain the parameter estimates $\hat{\boldsymbol{a}}_k$, $\hat{\boldsymbol{b}}_k$ and $\hat{\boldsymbol{\gamma}}_k$, and terminate this procedure.

6. Calculation analysis

This part compares the calculation amount of the the KT-AM-3S-GI algorithm and the KT-AM-3S-LSI algorithms with the existing algorithms for estimating the parameters of the IN-OE systems, namely the O-AM-GI algorithm, the O-AM-LSI algorithm, the KT-AM-GI algorithm and the KT-AM-LSI algorithm.

Table 1: The computational efficiency of each algorithm

Algorithms	Total flops
O-AM-GI	$N_1 = 6n_1L + 3n_1 + 2mn_b - 1$
O-AM-LSI	$N_2 = \frac{8}{3}n_1^3 + \frac{1}{6}n_1^2 + 2n_1^2L - \frac{11}{6}n_1 + 2mn_b + 2n_1L + 2n_1 - 1$
KT-AM-GI	$N_3 = 6nL + 3n + 2m - 2$
KT-AM-LSI	$N_4 = \frac{8}{3}n^3 + (2L + \frac{1}{6})n^2 + \frac{1}{6}n + 2nL + 2m - 2$
KT-AM-3S-GI	$N_5 = 6L(n_a + n_b + m) + 3(n_a + n_b + m) - 2$
KT-AM-3S-LSI	$N_6 = \frac{8}{3}(n_a^3 + n_b^3 + m^3) + (2L + \frac{1}{6})(n_a^2 + n_b^2 + m^2)$ $+6L(n_a + n_b + m) + \frac{1}{6}(n_a + n_b + m) - 2$

The calculation amount of each algorithm at each step is given in the supplementary file, and the calculation amount of each algorithm is shown in Table 1, where flop represents the floating point operations, and the stacked information matrices $\hat{\boldsymbol{\Phi}}_{1,k}(L)$ and $\hat{\boldsymbol{\Phi}}_{2,k}(L)$ are defined as

$$\hat{\boldsymbol{\Phi}}_{1,k}(L) := [\boldsymbol{\varphi}_{1,k}(1), \boldsymbol{\varphi}_{1,k}(2), \boldsymbol{\varphi}_{1,k}(3), \cdots, \boldsymbol{\varphi}_{1,k}(L)]^{\mathrm{T}} \in \mathbb{R}^{L\times n_1}, \quad n_1 := n_a + m(n_b + 1),$$

$$\hat{\boldsymbol{\Phi}}_{2,k}(L) := [\boldsymbol{\varphi}_{2,k}(1), \boldsymbol{\varphi}_{2,k}(2), \boldsymbol{\varphi}_{2,k}(3), \cdots, \boldsymbol{\varphi}_{2,k}(L)]^{\mathrm{T}} \in \mathbb{R}^{L\times n}, \quad n := n_a + n_b + m.$$

The calculation amount of each algorithm comprehensively satisfies the following relationship when m, $n_b \geqslant 2$,

$$N_5 < N_3 < N_1, \tag{83}$$

$$N_6 < N_4 < N_2, \tag{84}$$

where N_i $(i = 1, 2, \cdots, 6)$ represents the total flops of the O-AM-GI, O-AM-LSI, KT-AM-GI, KT-AM-LSI, KT-AM-3S-GI and KT-AM-3S-LSI in turn. Hence, the proposed algorithms have a small amount of calculation. Furthermore, for the same identification model, the gradient-based iterative algorithms have a smaller amount of calculation than the least squares-based iterative algorithms. In order to show the gap between each algorithm more clearly, a specific numerical comparison is given below. Set $n_a = 10$, $n_b = 10$, $m = 10$ and $L = 1000$, we can obtain N_1, N_2, N_3, N_4, N_5 and N_6 in Table 2.

Obviously, as the data length increases and the system dimension increases, this gap becomes more and more obvious, that is, the computational efficiency of the algorithms proposed in this paper is superior to the existing O-AM-GI, O-AM-LSI, KT-AM-GI and KT-AM-LSI algorithms.

Table 2: The computational efficiencies of each algorithm

$N_i (i = 1, \cdots, 6)$	Total flops	$N_i (i = 1, \cdots, 6)$	Total flops
N_1	7.20559×10^5	N_2	3.3650619×10^7
N_3	1.80108×10^5	N_4	1.932173×10^6
N_5	1.80088×10^5	N_6	7.88053×10^5

7. Simulation Example

The O-AM-GI algorithm, the O-AM-LSI algorithm, the KT-AM-GI algorithm and the KT-AM-LSI algorithm are chosen as the comparators. All the simulations are implemented on the MatLab R2019b version and the a machine with an Intel Core i7-10700 2.9 GHz CPU and 8 GB RAM.

Example 1: In this section, the system identification problems of the IN-OE model with seven elements in the parameter vector $\boldsymbol{\vartheta}$ is evaluated along with the derails as given below:

$$\begin{aligned}
y(t) &= \frac{B(z)}{A(z)}\bar{u}(t) + v(t),\\
A(z) &= 1 + a_1 z^{-1} + a_2 z^{-2} = 1 + 0.84z^{-1} + 0.31z^{-2},\\
B(z) &= b_0 + b_1 z^{-1} + b_2 z^{-2} = 1 - 0.57z^{-1} + 0.86z^{-2},\\
\bar{u}(t) &= f(u(t)) = \gamma_1 u(t) + \gamma_2 u^2(t) + \gamma_3 u^3(t)\\
&= -1.50u(t) - 2.60u^2(t) + 3.20u^3(t),\\
\boldsymbol{\vartheta} &= [a_1, a_2, b_1, b_2, \gamma_1, \gamma_2, \gamma_3]^{\mathrm{T}}\\
&= [0.84, 0.31, -0.57, 0.86, -1.50, -2.60, 3.20]^{\mathrm{T}}.
\end{aligned}$$

In this simulation, the input $\{u(t)\}$ is taken as an independent persistent excitation signal sequence with zero mean and unit variance, $\{v(t)\}$ is taken as a white noise with zero mean and variance $\sigma_1^2 = 0.20^2$ and $\sigma_2^2 = 1.00^2$, respectively. Applying the KT-AM-3S-GI algorithm and the KT-AM-3S-LSI algorithm with the data length $L = 1000$ to estimate the parameters of this system, the parameter estimation errors $\delta = \|\hat{\boldsymbol{\vartheta}}_k - \boldsymbol{\vartheta}\|/\|\boldsymbol{\vartheta}\|$ versus k are shown in Figures 2 and 3.

Adjust the input value, other simulation conditions are the same as above. Applying the KT-AM-3S-GI algorithm and the KT-AM-3S-LSI algorithm with the data length $L = 500$ and $L = 1000$ to estimate the parameters of this system, respectively, the estimation errors δ versus k are shown in Figures 4 and 5.

From Figures 2–5, we can draw the following conclusions.

- The parameter estimation errors given by the KT-AM-3S-GI and KT-AM-3S-LSI algorithms become smaller as the iteration k increases. Thus the proposed algorithms are effective for the IN-OE systems.

- To achieve the same parameter estimation accuracy, compared with the KT-AM-3S-GI algorithm, the KT-AM-3S-LSI algorithm requires fewer iterative steps.

- A lower noise variance leads to higher parameter estimation accuracy given by the KT-AM-3S-GI algorithm and the KT-AM-3S-LSI algorithm under the same data length.

- Under the same noise variance, the parameter estimation accuracy given by the the KT-AM-3S-GI algorithm and the KT-AM-3S-LSI algorithm becomes smaller as the data length L increases.

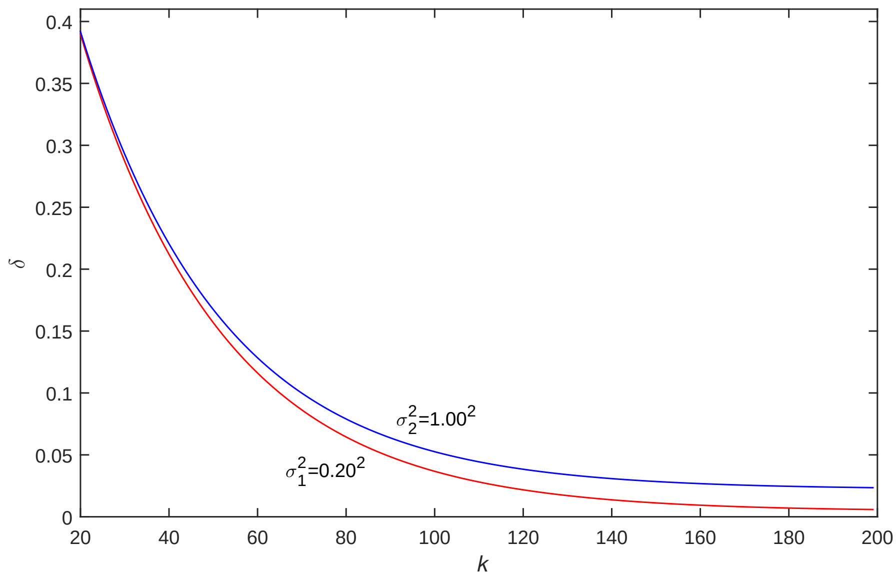

Figure 2: The KT-AM-3S-GI estimation errors δ versus k with different noise variance

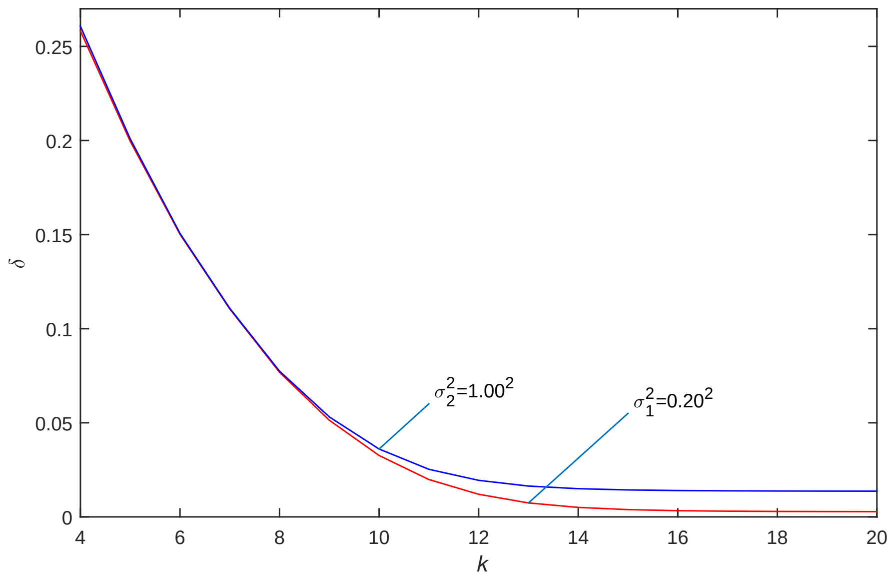

Figure 3: The KT-AM-3S-LSI estimation errors δ versus k with different noise variance

Example 2: In the above simulation analysis of the proposed algorithms, the influence of the noise variance and the data length on the parameter estimation accuracy is studied. Next, we will consider a model with more parameters, which is used as a basis to compare the parameter estimation accuracy of the presented algorithms and the existing algorithms. Consider the following IN-OE model:

$$\begin{aligned}
y(t) &= \frac{B(z)}{A(z)}\bar{u}(t) + v(t), \\
A(z) &= 1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} + a_4 z^{-4} + a_5 z^{-5} \\
&= 1 + 0.14z^{-1} + 0.21z^{-2} + 0.23z^{-3} + 0.31z^{-4} + 0.23z^{-5}, \\
B(z) &= b_0 + b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3} + b_4 z^{-4} + b_5 z^{-5} \\
&= 1 - 0.61z^{-1} + 0.87z^{-2} + 0.36z^{-3} + 0.23z^{-4} + 0.24z^{-5}, \\
\bar{u}(t) &= f(u(t)) = \gamma_1 \cos u(t) + \gamma_2 \sin u^2(t) + \gamma_3 u^3(t) \\
&= -1.50 \cos u(t) - 2.60 \sin u^2(t) + 3.20 u^3(t),
\end{aligned}$$

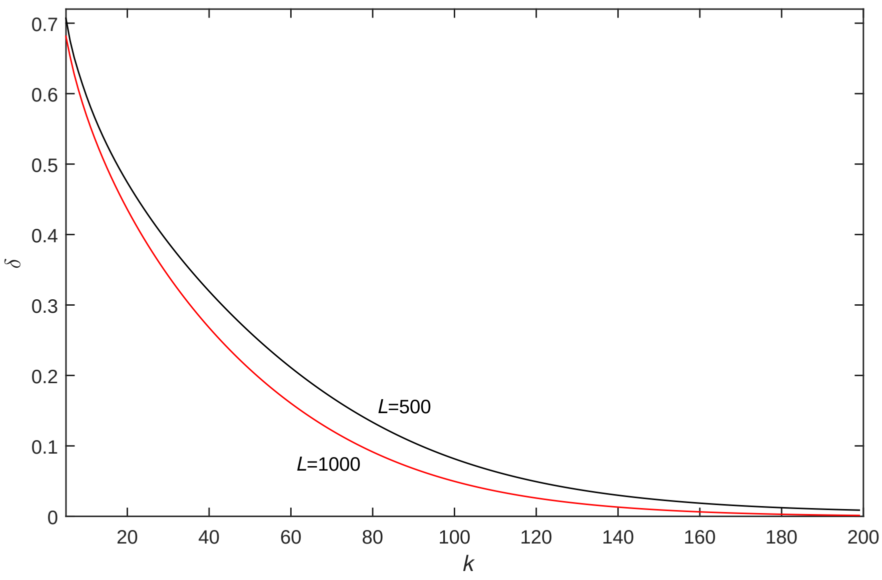

Figure 4: The KT-AM-3S-GI estimation errors δ versus k with different data length

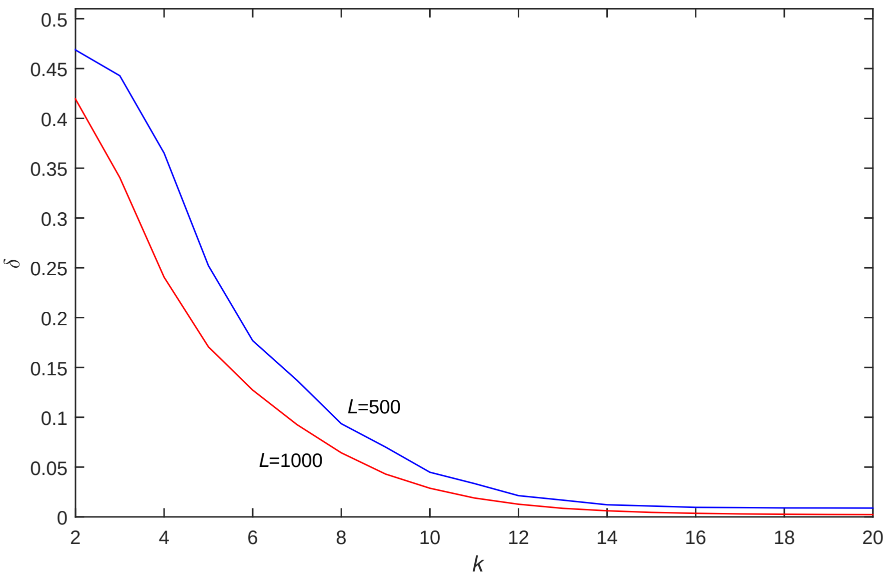

Figure 5: The KT-AM-3S-LSI estimation errors δ versus k with different data length

$$\begin{aligned}\boldsymbol{\vartheta} &= [a_1, a_2, a_3, a_4 a_5, b_1, b_2, b_3, b_4, b_5, \gamma_1, \gamma_2, \gamma_3]^{\mathrm{T}} \\ &= [0.14, 0.21, 0.23, 0.31, 0.23, -0.61, 0.87, 0.36, 0.23, 0.24, -1.50, -2.60, 3.20]^{\mathrm{T}}.\end{aligned}$$

In the simulation, the input $\{u(t)\}$ is taken as an independent persistent excitation signal sequence with zero mean and unit variance, $\{v(t)\}$ is taken as a white noise with zero mean and variance $\sigma^2 = 0.80^2$. Applying the O-AM-GI and O-AM-LSI algorithms, the KT-AM-GI and KT-AM-LSI algorithms, the KT-AM-3S-GI and KT-AM-3S-LSI algorithms with the data length $L = 500$ to estimate the parameters of the example system, the parameter estimation errors δ versus k are shown in Figures 6–7.

From Figures 6–7, we can draw the following conclusions.

- Under the same simulation conditions, the parameter estimation errors of the KT-AM-3S-GI algorithm are smaller than the parameter estimation accuracy of the O-AM-GI algorithm and the KT-AM-GI algorithm.
- Under the same simulation conditions, the parameter estimation accuracy of KT-AM-3S-LSI algorithm is higher than that of O-AM-LSI algorithm and KT-AM-LSI algorithm.

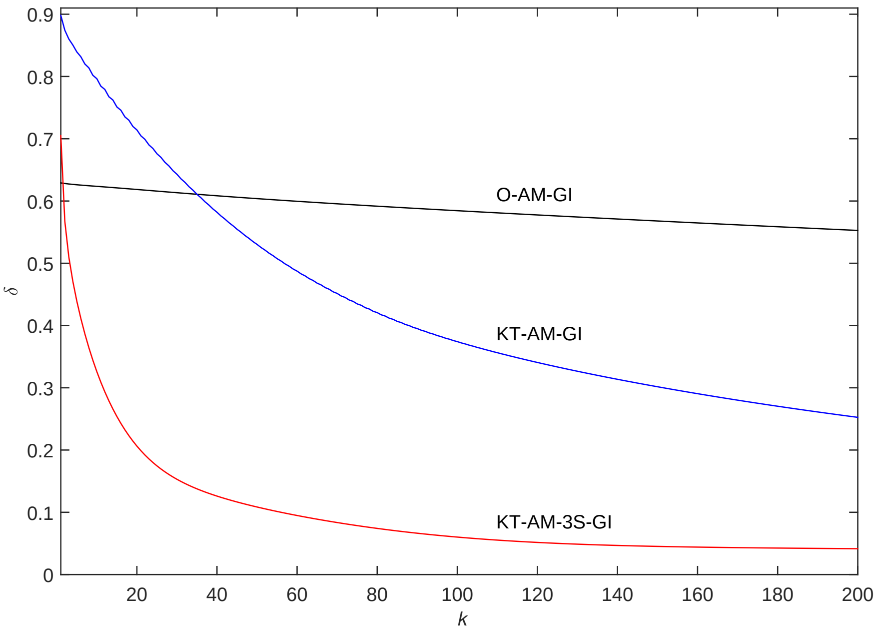

Figure 6: The estimation errors δ obtained from three gradient-based iterative algorithms versus k

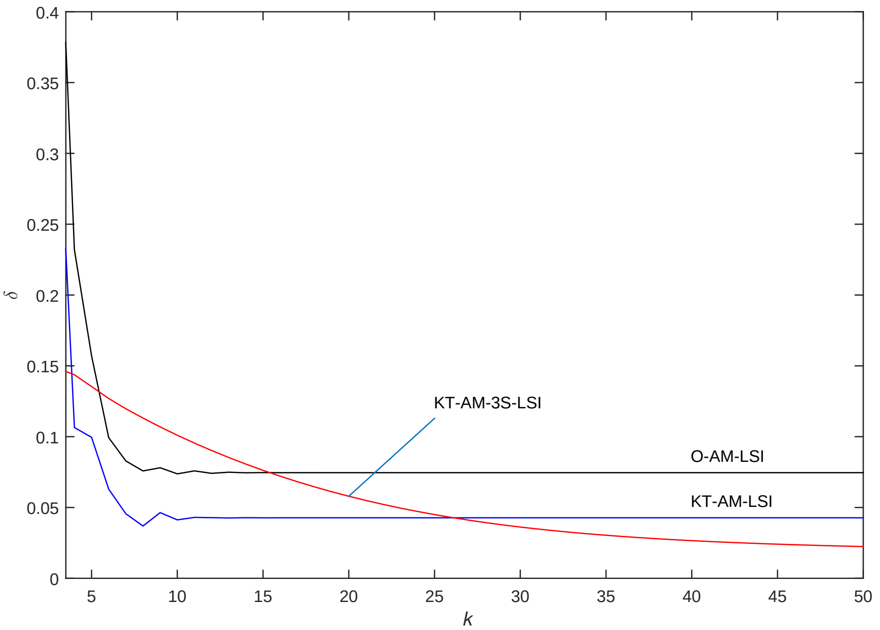

Figure 7: The estimation errors δ obtained from three least squares-based iterative algorithms versus k

8. Conclusions

In this paper, we have studied the parameter identification problems of the dynamical systems described by the input nonlinear output-error model. A key term separation auxiliary model three-stage gradient-based iterative algorithm and a key term separation auxiliary model three-stage least squares-based iterative algorithm have been presented based on the key term separation technique and the hierarchical identification principle. In order to fully demonstrate the performance of the proposed algorithms, we compares the difference in the calculation and the parameter estimation accuracy between the proposed algorithms and the existing algorithms in detail. Both the data and the simulation waveforms fully confirm that the presented algorithms have less calculation and higher parameter estimation accuracy. However, there are still two problems worth thinking about: when the system's input and output dimensions increase, that is the system becomes a multiple-input multiple-output system, and when the disturbances become complex colored noises, whether the methods studied in this paper still have good performance needs further research. The proposed methods in this paper can be applied to other fields [83, 84, 85, 86, 87] such as signal processing and engineering application systems [88, 89, 90, 91, 92, 93] the information processing and transportation communication systems [94, 95, 96, 97, 98, 99] and so on.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F. Ding, *System Identification - Auxiliary Model Identification Idea and Methods*, Science Press, Beijing, 2017.

[2] X.Q. Yang, W.L. Xiong, J.X. Ma, et al. Robust identification of Wiener time-delay system with expectation-maximization algorithm, Journal of the Franklin Institute 354 (13) (2017) 5678-5693

[3] M.H. Li, et al, The filtering-based maximum likelihood iterative estimation algorithms for a special class of nonlinear systems with autoregressive moving average noise using the hierarchical identification principle, International Journal of Adaptive Control and Signal Processing 33 (7) (2019) 1189-1211.

[4] M.H. Li, X.M. Liu, Maximum likelihood hierarchical least squares-based iterative identification for dual-rate stochastic systems, International Journal of Adaptive Control and Signal Processing 35 (2) (2021) 240-261.

[5] Y. Gu, J. Liu, X. Li, Y. Chou, Y. Ji, State space model identification of multirate processes with time-delay using the expectation maximization, Journal of the Franklin Institute 356 (3) (2019) 1623-1639.

[6] Y. Ji, C. Zhang, Z. Kang, T. Yu, Parameter estimation for block-oriented nonlinear systems using the key term separation, International Journal of Robust and Nonlinear Control 30 (9) (2020) 3727-3752.

[7] Y. Ji, X.K. Jiang, L.J. Wan, Hierarchical least squares parameter estimation algorithm for two-input Hammerstein finite impulse response systems, Journal of the Franklin Institute 357 (8) (2020) 5019-5032.

[8] G.Y. Chen, M. Gan, C.L.P. Chen, et al, A regularized variable projection algorithm for separable nonlinear least-squares problems, IEEE Transactions on Automatic Control 64 (2) (2019) 526-537. February 2019.

[9] L. Xu, et al, Auxiliary model multiinnovation stochastic gradient parameter estimation methods for nonlinear sandwich systems, International Journal of Robust and Nonlinear Control 31 (1) (2021) 148-165. January 2020.

[10] X. Zhang, et al, Recursive parameter estimation and its convergence for bilinear systems, IET Control Theory and Applications 14 (5) (2020) 677-688.

[11] X.X. Zhang, J.C. Ji, X. Jian, Parameter identification of time-delayed nonlinear systems: An integrated method with adaptive noise correction, Journal of the Franklin Institute 356 (11) (2019) 5858-5880.

[12] M. Gan, G.Y. Chen, L. Chen, C.L.P. Chen, Term selection for a class of separable nonlinear models, IEEE Transactions on Neural Networks and Learning Systems 31 (2) (2020) 445-451.

[13] G.Y. Chen, et al, Modified Gram-Schmidt method-based variable projection algorithm for separable nonlinear models, IEEE Transactions on Neural Networks and Learning Systems 30 (8) (2019) 2410-2418.

[14] M. Gan, C.L.P. Chen, G.Y. Chen, L. Chen, On some separated algorithms for separable nonlinear squares problems, IEEE Transactions on Cybernetics 48 (10) (2018) 2866-2874.

[15] M. Gan, Y. Guan, G.Y. Chen, C.L.P. Chen. Recursive variable projection algorithm for a class of separable nonlinear models. IEEE Transactions on on Neural Networks and Learning System (2021). doii:10.1109/TNNLS.2020.3026482

[16] G.Y. Chen, M. Gan, C.L.P. Chen, H.X. Li, Basis function matrix-based flexible coefficient autoregressive models: A framework for time series and nonlinear system modeling, IEEE Transactions on Cybernetics 51 (2) (2021) 614-623.

[17] J. Ding, Z.X. Cao, J.Z. Chen, et al, Weighted parameter estimation for Hammerstein nonlinear ARX systems, Circuits Systems and Signal Processing 39 (4) (2020) 2178-2192.

[18] J. Chen, Y.J. Liu, et al., Gradient-based particle filter algorithm for an ARX model with nonlinear communication output, IEEE Transactions on Systems Man and Cybernetics: Systems 50 (6) (2020) 2198-2207.

[19] H.F. Xia, Y. Ji, Y.J. Liu, et al., Maximum likelihood-based multi-innovation stochastic gradient method for multivariable systems, International Journal of Control Automation and Systems 17 (3) (2019) 565-574.

[20] F. Albu, K. Nishikawa, New iterative kernel algorithms for nonlinear acoustic echo cancellation, 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA) 16-19 December 2015, Hong Kong, China, pp. 734-739.

[21] J. Ding, J.Z. Chen, J.X. Lin, et al., Particle filtering-based recursive identification for controlled auto-regressive systems with quantised output, IET Control Theory and Applications 13 (14) (2019) 2181-2187.

[22] J. Ma, S. Xu, Y. Li, et al., Neural networks-based adaptive output feedback control for a class of uncertain nonlinear systems with input delay and disturbances, Journal of the Franklin Institute 355 (13) (2018) 5503-5519.

[23] M.H. Li and X.M. Liu, The least squares based iterative algorithms for parameter estimation of a bilinear system with autoregressive noise using the data filtering technique, Signal Processing 147 (2018) 23-34.

[24] D.Q. Wang, L.W. Li, Y. Ji, Y.R. Yan. Model recovery for Hammerstein systems using the auxiliary model based orthogonal matching pursuit method, Applied Mathematical Modelling 54 (2018) 537-550.

[25] D.Q. Wang, Y.R. Yan, Y.J. Liu, et al., Model recovery for Hammerstein systems using the hierarchical orthogonal matching pursuit method, Journal of Computational and Applied Mathematics 345 (2019) 135-145.

[26] D.Q. Wang, S. Zhang, M. Gan, and J.L. Qiu, A novel EM identification method for Hammerstein systems with missing output data, IEEE Transactions on Industrial Informatics 16 (4) (2020) 2500-2508.

[27] D.Q. Wang, Q.H. Fan, Y. Ma, An interactive maximum likelihood estimation method for multivariable Hammerstein systems, Journal of the Franklin Institute 357 (17) (2020) 12986-13005.

[28] J. Ding, J.Z. Chen, J.X. Lin, et al., Particle filtering based parameter estimation for systems with output-error type model structures, Journal of the Franklin Institute 356 (10) (2019) 5521-5540.

[29] H. Ma, X. Zhang, Q.Y. Liu, et al, Partiallly-coupled gradient-based iterative algorithms for multivariable output-error-like systems with autoregressive moving average noises, IET Control Theory and Applications 14 (17) (2020) 2613-2627.

[30] M.H. Li, X.M. Liu, Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique, International Journal of Control Automation and Systems 18 (6) (2020) 1581-1592.

[31] L.J. Wang, Y. Ji, H.L. Yang, et al, Decomposition-based multiinnovation gradient identification algorithms for a special bilinear system based on its input-output representation, International Journal of Robust and Nonlinear Control 30 (9) (2020) 3607-3623.

[32] L.J. Wang, Y. Ji, L.J. Wan, N. Bu, Hierarchical recursive generalized extended least squares estimation algorithms for a class of nonlinear stochastic systems with colored noise, Journal of the Franklin Institute 356 (16) (2019) 10102-10122.

[33] L. Zhang, S.Y. Tang, L.L. Lv, An finite iterative algorithm for sloving periodic Sylvester bimatrix equations, Journal of the Franklin Institute 357 (15) (2020) 10757-10772.

[34] L. Zhang, C.B. Xu, Y.H. Gao, Y. Han, X.J. Du, Z.H. Tian, Improved Dota2 lineup recommendation model based on a bidirectional LSTM, Tsinghua Science and Technology 25 (6) (2020) 712-720.

[35] F.Z. Geng, X.Y. Wu, Reproducing kernel function-based Filon and Levin methods for solving highly oscillatory integral, Applied Mathematics and Computation 397 (2021), Article Number: 125980.

[36] F.Z. Geng, Piecewise reproducing kernel-based symmetric collocation approach for linear stationary singularly perturbed problems, AIMS Mathematics 5 (6) (2020) 6020-6029.

[37] F. Ding, *System Identification-Iterative Search Principle and Identification Methods*, Science Press, Beijing, 2018.

[38] L. Xu, The parameter estimation algorithms based on the dynamical response measurement data, Advances in Mechanical Engineering 9 (11) (2017) Article Number: 1687814017730003.

[39] L. Xu, L. Chen, W.L. Xiong, Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration, Nonlinear Dynamics 79 (3) (2015) 2155-2163.

[40] L. Xu, The damping iterative parameter identification method for dynamical systems based on the sine signal measurement, Signal Processing 120 (2016) 660-667.

[41] J. Pan, X. Jiang, X.K. Wan, et al., A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems, International Journal of Control Automation and Systems 15 (3) (2017) 1189-1197.

[42] X. Zhang, et al, Recursive parameter estimation methods and convergence analysis for a special class of nonlinear systems, International Journal of Robust and Nonlinear Control 30 (4) (2020) 1373-1393.

[43] X. Zhang, et al, Recursive identification of bilinear time-delay systems through the redundant rule, Journal of the Franklin Institute 357 (1) (2020) 726-747.

[44] X. Zhang, et al, Hierarchical parameter and state estimation for bilinear systems, International Journal of Systems Science 51 (2) (2020) 275-290.

[45] H.Y. Wan, X.L. Luan, H.R. Karimi, F. Liu, Dynamic self-triggered controller codesign for Markov jump systems, IEEE Transactions on Automatic Control 66 (3) (2021) 1353-1360.

[46] X.S. Zhan, J.W. Hu, J. Wu, H.C. Yan, Analysis of optimal performance of MIMO NCS with encoding and packet dropout constraints, IET Control Theory and Applications 14 (13) (2020) 1762-1768.

[47] L.L. Hao, X.S. Zhan, J. Wu, T. Han, H.C. Yan, Fixed-time group consensus of nonlinear multi-agent systems via pinning control, International Journal of Control Automation and Systems 19 (1) (2021) 200-208.

[48] Y. Gu, Y. Chou, J. Liu, Y. Ji, Moving horizon estimation for multirate systems with time-varying time-delays, Journal of the Franklin Institute 356 (4) (2019) 2325-2345.

[49] Y. Gu, Q. Zhu, H. Nouri, Bias compensation-based parameter and state estimation for a class of time-delay nonlinear state-space models, IET Control Theory and Applications 14 (15) (2020) 2176-2185.

[50] Y. Ji, Z. Kang, Three-stage forgetting factor stochastic gradient parameter estimation methods for a class of nonlinear systems, International Journal of Robust and Nonlinear Control 31 (3) (2021) 871-987.

[51] Y.M. Fan, X.M. Liu, Two-stage auxiliary model gradient-based iterative algorithm for the input nonlinear controlled autoregressive system with variable-gain nonlinearity, International Journal of Robust and Nonlinear Control 30 (14) (2020) 5492-5509.

[52] X.M. Liu, Y.M. Fan, Maximum likelihood extended gradient-based estimation algorithms for the input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity, International Journal of Robust and Nonlinear Control 31 (2021) doi:10.1002/rnc.5450

[53] N. Bu, J.X. Pang, M. Deng, Robust fault tolerant tracking control for the multi-joint manipulator based on operator theory, Journal of the Franklin Institute 357 (5) (2020) 2696-2714.

[54] J. Chen, J. Li, Y.J. Liu, Gradient iterative algorithm for dual-rate nonlinear systems based on a novel particle filter, Journal of the Franklin Institute 354 (11) (2017) 4425-4437.

[55] H. Dong, C.C. Yin, H.S. Dai, Spectrally negative Levy risk model under Erlangized barrier strategy, Journal of Computational and Applied Mathematics 351 (2019) 101-116.

[56] X.Y. Sha, Z.S. Xu, C.C. Yin, Elliptical distribution-based weight-determining method for ordered weighted averaging operators, International Journal of Intelligent Systems 34 (5) (2019) 858-877.

[57] C.C. Yin, Y.Z. Wen, An extension of Paulsen-Gjessing's risk model with stochastic return on investments, Insurance Mathematics & Economics 52 (3) (2013) 469-476.

[58] Y.X. Zhao, P. Chen, H.L. Yang, Optimal periodic dividend and capital injection problem for spectrally positive Levy processes, Insurance Mathematics & Economics 74 (2017) 135-146.

[59] X.H. Zhao, H. Dong, H.S. Dai, On spectrally positive Levy risk processes with Parisian implementation delays in dividend payments, Statistics & Probability Letters 140 (2018) 176-184.

[60] Y.X. Zhao, C.C. Yin, The expected discounted penalty function under a renewal risk model with stochastic income, Applied Mathematics and Computation 218 (10) (2012) 6144-6154.

[61] X. Zhang, et al, Recursive parameter identification of the dynamical models for bilinear state space systems, Nonlinear Dynamics 89 (4) (2017) 2415-2429.

[62] L. Xu, et al, Hierarchical multi-innovation generalised extended stochastic gradient methods for multivariable equation-error autoregressive moving average systems, IET Control Theory and Applications 14 (10) (2020) 1276-1286.

[63] X. Zhang, et al, Combined state and parameter estimation for a bilinear state space system with moving average noise, Journal of the Franklin Institute 355 (6) (2018) 3079-3103.

[64] X. Zhang, et al, State filtering-based least squares parameter estimation for bilinear systems using the hierarchical identification principle, IET Control Theory and Applications 12 (12) (2018) 1704-1713.

[65] L. Xu, et al, Hierarchical Newton and least squares iterative estimation algorithm for dynamic systems by transfer functions based on the impulse responses, International Journal of Systems Science 50 (1) (2019) 141-151.

[66] X. Zhang, et al, Highly computationally efficient state filter based on the delta operator, International Journal of Adaptive Control and Signal Processing 33 (6) (2019) 875-889.

[67] F. Ding, X.H. Wang, L. Mao, et al, Joint state and multi-innovation parameter estimation for time-delay linear systems and its convergence based on the Kalman filtering, Digital Signal Processing 62 (2017) 211-223.

[68] X. Zhang, et al, State estimation for bilinear systems through minimizing the covariance matrix of the state estimation errors, International Journal of Adaptive Control and Signal Processing 33 (7) (2019) 1157-1173.

[69] F. Ding, et al, Performance analysis of the generalised projection identification for time-varying systems, IET Control Theory and Applications 10 (18) (2016) 2506-2514.

[70] F. Ding, et al, Parameter estimation for pseudo-linear systems using the auxiliary model and the decomposition technique, IET Control Theory and Applications 11 (3) (2017) 390-400.

[71] L. Xu, W.L. Xiong, A. Alsaedi, T. Hayat, Hierarchical parameter estimation for the frequency response based on the dynamical window data, International Journal of Control Automation and Systems 16 (4) (2018) 1756-1764.

[72] L. Xu, et al, Separable multi-innovation stochastic gradient estimation algorithm for the nonlinear dynamic responses of systems, International Journal of Adaptive Control and Signal Processing 34 (7) (2020) 937-954.

[73] J. Pan, W. Li, H.P. Zhang, 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 16 (6) (2018) 2878-2887.

[74] L. Xu, G.L. Song, A recursive parameter estimation algorithm for modeling signals with multi-frequencies, Circuits Systems and Signal Processing 39 (8) (2020) 4198-4224.

[75] L.Xu, et al, Separable recursive gradient algorithm for dynamical systems based on the impulse response signals, International Journal of Control Automation and Systems 18 (12) (2020) 3167-3177.

[76] X.M. Zhang, Z.Y. Zhao, Z.Y. Wang, X.Y. Wang, Fault detection and identification method for quadcopter based on airframe vibration signals, Sensors 21 (2) (2021) Article Number: 581.

[77] X.B. Jin, X.H. Yu, T.L. Su, D.N. Yang, Y.T. Bai, J.L. Kong, L. Wang, Distributed deep fusion predictor for a multi-sensor system based on causality, Entropy 23 (2) (2021) Article Number: 219.

[78] X.B. Jin, R.J. RobertJeremiah, T.L. Su, Y.T. Bai, J.L. Kong, The new trend of state estimation: from model-driven to hybrid-driven methods, Sensors 21 (6) (2021) Article Number: 2085.

[79] G.C. Yang, Z.J. Chen, Y. Li, Z.D. Su, Rapid relocation method for mobile robot based on improved ORB-SLAM2 algorithm, Remote Sensing 11 (2) (2019). Article Number: 149.

[80] J. Lin, Y. Li, G.C. Yang, FPGAN: Face de-identification method with generative adversarial networks for social robots, Neural Networks 133 (2021) 132-147.

[81] S.S. Tian, X.X. Zhang, S. Xiao, J. Zhang, Q. Chen, Y. Li, Application of C6F12O/CO2 mixture in 10 kV medium-voltage switchgear, IET Science Measurement & Technology 13 (9) (2019) 1225-1230.

[82] M.H. Wu, H.H. Yue, J. Wang, et al, Object detection based on RGC mask R-CNN, IET Image Processing 14 (8) (2020) 1502-1508.

[83] D.C. Chen, X.X. Zhang, J. Tang, et al., Pristine and Cu decorated hexagonal InN monolayer, a promising candidate to detect and scavenge SF6 decompositions based on first-principle study, Journal of Hazardous Materials 363 (2019) 346-357.

[84] H. Cui, D.C. Chen, Y. Zhang, et al., Dissolved gas analysis in transformer oil using Pd catalyst decorated MoSe2 monolayer: A first-principles theory, Sustainable Materials and Technologies 20 (2019) Article Number: e00094.

[85] H. Cui, T. Liu, Y. Zhang, et al., Ru-InN monolayer as a gas scavenger to guard the operation status of SF6 insulation devices: A first-principles theory, IEEE Sensors Journal 19 (13) (2019) 5249-5255.

[86] H. Cui, X.X. Zhang, D.C. Chen, et al., Adsorption mechanism of SF6 decomposed species on pyridine-like PtN3 embedded CNT: A DFT study, Applied surface science 447 (2018) 594-598.

[87] H. Cui, X.X. Zhang, Y. Li, et al., First-principles insight into Ni-doped InN monolayer as a noxious gases scavenger, Applied Surface Science 494 (2019) 859-866.

[88] H. Cui, X.X. Zhang, G.Z. Zhang, et al., Pd-doped MoS2 monolayer: A promising candidate for DGA in transformer oil based on DFT method, Applied Surface Science 470 (2019) 1035-1042.

[89] H. Cui, G.Z. Zhang, X.X. Zhang, et al., Rh-doped MoSe2 as a toxic gas scavenger: a first-principles study, Nanoscale Advances 1 (2) (2019) 772-780.

[90] X.X. Zhang, Y.G. Gui, H.Y. Xiao, et al., Analysis of adsorption properties of typical partial discharge gases on Ni-SWCNTs using density functional theory, Applied Surface Science 379 (2016) 47-54.

[91] X.X. Zhang, L. Yu, Y.G. Gui, et al., First-principles study of SF6 decomposed gas adsorbed on Au-decorated graphene, Applied Surface Science 367 (2016) 259-269.

[92] X.X. Zhang, L. Yu, X.Q. Wu, et al., Experimental sensing and density functional theory study of H2S and SOF2 adsorption on Au-modified graphene, Advanced Science 2 (11) (2015) Article Number: 1500101.

[93] X.X. Zhang, Y. Zhang, Y. Huang, Y. Li, H.T. Cheng, S. Xiao, Detection of decomposition products of C4F7N-CO2 gas mixture based on infrared spectroscopy, Vibrational Spectroscopy 110 (2020), Article Number: 103114.

[94] Y. Cao, J.K. Wen, L.C. Ma, Tracking and collision avoidance of virtual coupling train control system, Alexandria Engineering Journal 60 (2) (2021) 2115-2125.

[95] Y. Cao, J.K. Wen, L.C. Ma, Tracking and collision avoidance of virtual coupling train control system, Future Generation Computer Systems 120 (2021) 76-90.

[96] S. Su, X.K. Wang, Y. Cao, J.T. Yin, An energy-efficient train operation approach by integrating the metro timetabling and eco-driving, IEEE Transactions on Intelligent Transportation Systems 21 (10) (2020) 4252-4268.

[97] S. Su, T. Tang, J. Xun, F. Cao, Y.H. Wang, Design of running grades for energy-efficient train regulation: A case study for beijing yizhuang line, IEEE Intelligent Transportation Systems Magazine (2021). doi:10.1109/MITS.2019.2907681.

[98] Y. Cao, Z. Wang, F. Liu, P. Li, G. Xie, Bio-inspired speed curve optimization and sliding mode tracking control for subway trains, IEEE Transactions on Vehicular Technology 68 (7) (2019) 6331-6342.

[99] Y. Cao, Y.K. Sun, G. Xie, T. Wen, Fault diagnosis of train plug door based on a hybrid criterion for IMFs selection and fractional wavelet package energy entropy, IEEE Transactions on Vehicular Technology 68 (8) (2019) 7544-7551.