INTERNATIONAL JOURNAL OF ROBUST AND NONLINEAR CONTROL
Int. J. Robust Nonlinear Control 2021; **00**:1–17
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/acs

Identification of the nonlinear systems based on the kernel functions

Jimei Li[1], Feng Ding*[1†], Erfu Yang[2]

[1]*Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, PR China*
[2]*Department of Design, Manufacturing and Engineering Management, University of Strathclyde, Glasgow, Scotland, UK*

SUMMARY

Constructing an appropriate membership function is significant in fuzzy logic control. Based on the multi-model control theory, this paper constructs a novel kernel function which can implement the fuzzification and defuzzification processes and reflect the dynamic quality of the nonlinear systems accurately. Then we focus on the identification problems of the nonlinear systems based on the kernel functions. Applying the hierarchical identification principle, we present the hierarchical stochastic gradient (H-SG) algorithm for the nonlinear systems. Meanwhile, the one-dimensional search methods are proposed to solve the problem of determining the optimal step sizes. In order to improve the parameter estimation accuracy, we propose the hierarchical multi-innovation forgetting factor stochastic gradient (H-MIFG) algorithm by introducing the forgetting factor and using the multi-innovation identification theory. The simulation example is provided to test the proposed algorithms from the aspects of parameter estimation accuracy and prediction performance.

Received ...

KEY WORDS: Gaussian membership function, Nonlinear system, Hierarchical identification, Parameter estimation, Multi-innovation identification, Gradient search

1. INTRODUCTION

System identification is theory of establishing the mathematical models of dynamical systems by measuring the system inputs and outputs [1–3]. System modeling and parameter estimation are the basis of all the control problems [4–9], so it is significant to build an appropriate model for system prediction and control [10–15]. Many identification methods have been developed for linear systems, but most practical systems have nonlinear characteristics in nature. Although nonlinear system identification is challenging because of the difficulty to decide a general model structure which can represent data from a nonlinear system, the identification of nonlinear systems plays an important role in the design and analysis of control systems [16]. For decades, the control of nonlinear systems has drawn ever-increasing research interests. Meanwhile, in the field of system identification, a great deal of works have devoted to the identification of nonlinear systems [17–19]. For the Hammerstein nonlinear ARMAX systems, Cheng et al. proposed a multi-innovation fractional-order stochastic gradient algorithm [20].

The application of fuzzy logic theory provides an effective way to deal with the approximate and inexact nature of the real world [21,22]. Fuzzy logic control has been used to design control strategy in industrial process control, biomedical research, pattern recognition and other fields [23–25].

*Correspondence to: Feng Ding
†Email: fding@jiangnan.edu.cn

For the sake of completing a general fuzzy logic control, each of the components, that is, a fuzzification part, an inference engine and a defuzzification part have to be implemented [26]. Various membership functions are developed for these parts in order to get a good performance in the fuzzy logic control [27]. Among them, the Gaussian membership functions are widely used in many areas [28–30]. However, for the complicated unknown nonlinear systems, the simple fuzzification and defuzzification processes decrease the precision and dynamic quality of fuzzy logic control [31, 32]. To improve this situation, a novel kernel function is constructed in this paper, which has a good performance in fitting the nonlinear systems. This method can further apply other nonlinear functions to establish effective mathematical models which can also be used in the fuzzy control processes or other areas.

Many methods have been developed to estimate the Gaussian functions. In the literature, Guo derived a weighted least-squares method for estimating the parameters of a Gaussian function [33]. An algorithm was presented for fitting a Gaussian signal riding on a polynomial background by transforming the nonlinear least-squares fitting into a standard linear least-squares fitting [34]. However, it is worth noting that these methods are restricted to estimating the parameters of the Gaussian shaped functions. For decades, the hierarchical identification method has been used to solve the identification problems of nonlinear systems [35–38]. Simultaneously, the multi-innovation identification has shown the effectiveness in nonlinear system identification [39–41]. For example, a multi-innovation stochastic gradient algorithm and a filtering based multi-innovation stochastic gradient algorithm were proposed for a class of linear-in-parameters systems by expanding a scalar innovation into a multi-dimensional vector [42]. Combining the hierarchical identification technique with the multi-innovation theory, the novel methods are developed for identifying the nonlinear systems.

Fuzzy logic has become an important technique for artificial intelligence since the fuzzy logic allows introducing the human uncertain behaviour to the computer definite performance [43]. In order to ensure the effective application of the fuzzy theory to practical problems, the kernel functions are constructed based on the Gaussian membership functions. For the nonlinear systems based on the kernel functions, we apply the hierarchical identification principle and decompose a nonlinear system into two sub-systems, one of which contains the unknown parameter vector of the linear sub-system, and the other contains the unknown parameter of the nonlinear part. These parameter vectors are estimated interactively. In brief, we list the following contributions provided in this paper.

- Based on the Gaussian membership functions, a novel kernel function is constructed for the fuzzification and defuzzification processes, which has a good performance in fitting the nonlinear systems.
- Combining the hierarchical identification principle with the gradient search, a hierarchical stochastic gradient (H-SG) algorithm is proposed for the nonlinear systems. Then the one-dimensional search methods are used to derive the optimal step sizes.
- For improving the estimation accuracy of the H-SG algorithm, a hierarchical multi-innovation stochastic gradient (H-MISG) algorithm are presented for the nonlinear systems by making full use of the innovations.

In summary, the rest of this paper is organized as follows. Section 2 describes the construction process of the kernel functions. Based on the hierarchical identification principle, the H-SG algorithm is proposed for the nonlinear systems in Section 3. Section 4 derives the H-MISG algorithm and its variant by using the multi-innovation identification theory. Section 5 presents a numerical example for demonstrating the effectiveness of the proposed algorithms. Finally, some concluding remarks are given in Section 6.

2. PROBLEM DESCRIPTION

As one of the common membership functions in the fuzzy logic control, the Gaussian membership functions are usually utilized to accomplish the fuzzification and defuzzification processes due to

DOI: 10.1002/acs

its unique smoothness, symmetry and clear physical meaning [44], the expression is given as

$$\mu(x) = \exp\left[\frac{-(x-a)^2}{b^2}\right], \quad -\infty < x < \infty.$$

However, the simple fuzzy processing reduces the accuracy of the fuzzy models for the complex nonlinear systems. As a result, the fuzzy models cannot accurately reflect the dynamic performance of the complex nonlinear systems, and the development of new approaches becomes an unavoidable challenge for engineers and researchers.

Therefore, we construct a kernel function for the fuzzification and defuzzification processes, which has a good performance in fitting the nonlinear systems. Based on the Gaussian membership functions, the proposed kernel functions are given as

$$F(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, k) = \sum_{j=1}^{n} \alpha_j f_j(\boldsymbol{\beta}, \boldsymbol{\lambda}, k),$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_n]^{\mathrm{T}}$, $\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_n]^{\mathrm{T}}$, $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \cdots, \lambda_n]^{\mathrm{T}}$, $f_j(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)$ is the Gaussian membership function, i.e., $f_j(\boldsymbol{\beta}, \boldsymbol{\lambda}, k) = \exp[-\beta_j(x_k - \lambda_j)^2]$ and $F(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, k)$ denotes the membership of the element x_k corresponding to the fuzzy set.

According to the given input x_k and the observation data s_k ($k = 1, 2, \cdots, L$), the parameters α_j, β_j and λ_j ($j = 1, 2, \cdots, n$) can be determined. Suppose that the order n is known, the data s_k is measurable and the initial values are set to be $x_k = 0$ and $s_k = 0$ for $k \leq 0$. The optimal parameter vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ can be obtained by solving the optimization problem

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}} [s_k - \sum_{j=1}^{n} \alpha_j f_j(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)]^2. \tag{1}$$

The above equation indicate that the identification problem becomes a complex nonlinear optimization problem. The proposed parameter estimation algorithms in this paper are based on this criterion function. Many identification methods are derived based on the criterion functions of the systems [45–49] and can be used to estimate the parameters of other linear systems and nonlinear systems [50–54] and can be applied to other fields [55–60] such as chemical process control systems and information systems. The hierarchical identification methods have been used to solve these identification problems of nonlinear systems, which can reduce the computational burden by decomposing the complex identification models into several sub-models with smaller dimensions and fewer parameters. Applying the hierarchical identification principle, the following work focuses on the new recursive identification methods for the nonlinear systems based on the kernel functions.

3. THE HIERARCHICAL STOCHASTIC GRADIENT ALGORITHM

In this section, the nonlinear system based on the kernel functions is decomposed into two identification sub-systems, one contains the parameter vector $\boldsymbol{\alpha}$ and the other contains the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. In addition, the negative gradient search is widely adopted to deal with some optimization problems [61]. Based on the gradient search, an hierarchical stochastic gradient (H-SG) algorithm is proposed for the nonlinear systems.

3.1. Estimating the parameters of the linear sub-system

For any fixed $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$, the negative gradient search is used to compute $\boldsymbol{\alpha}$ in the linear sub-system. Define the information vector

$$\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k) := [\exp[-\beta_1(x_k - \lambda_1)^2], \exp[-\beta_2(x_k - \lambda_2)^2], \cdots, \exp[-\beta_n(x_k - \lambda_n)^2]]^{\mathrm{T}} \in \mathbb{R}^n.$$

DOI: 10.1002/acs

Define the error between the observation output s_k and the model output $F(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, k)$ as

$$e(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, k) := s_k - F(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, k) = s_k - \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)\boldsymbol{\alpha} \in \mathbb{R}.$$

Let $\hat{\boldsymbol{\alpha}}_k := [\hat{\alpha}_{1,k}, \hat{\alpha}_{2,k}, \cdots, \hat{\alpha}_{n,k}]^{\mathrm{T}}$ denote the estimate of $\boldsymbol{\alpha}$ at time k. Applying the negative gradient search to solve the optimization problem in (1), we derive the stochastic gradient algorithm for computing $\boldsymbol{\alpha}$:

$$\hat{\boldsymbol{\alpha}}_k = \hat{\boldsymbol{\alpha}}_{k-1} + \nu_{1,k}\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)[s_k - \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)\hat{\boldsymbol{\alpha}}_{k-1}], \tag{2}$$

$$\nu_{1,k} = \underset{\nu_{1,k} \geqslant 0}{\operatorname{argmin}} \frac{1}{2} e^2(\hat{\boldsymbol{\alpha}}_k, \boldsymbol{\beta}, \boldsymbol{\lambda}, k), \tag{3}$$

where $\nu_{1,k}$ is the step size. The choice of the step size is critical in determining the overall accuracy of the algorithm. Therefore, the one-dimensional search methods are derived to solve the difficulty in determining the step size. Define $h[\nu_{1,k}] := e^2(\hat{\boldsymbol{\alpha}}_k, \boldsymbol{\beta}, \boldsymbol{\lambda}, k)$ for computing $\nu_{1,k}$. The key idea is to determine the negative gradient direction and to compute the step size, which makes $h[\nu_{1,k}]$ minimal, by the one-dimensional search of the negative gradient direction. The step size is solved as follows

$$\begin{aligned}
h[\nu_{1,k}] &= [s_k - \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)\hat{\boldsymbol{\alpha}}_k]^2 \\
&= \{s_k - \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)[\hat{\boldsymbol{\alpha}}_{k-1} + \nu_{1,k}\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)e(\hat{\boldsymbol{\alpha}}_{k-1}, \boldsymbol{\beta}, \boldsymbol{\lambda}, k)]\}^2 \\
&= [s_k - \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)\hat{\boldsymbol{\alpha}}_{k-1} - \nu_{1,k}\boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)e(\hat{\boldsymbol{\alpha}}_{k-1}, \boldsymbol{\beta}, \boldsymbol{\lambda}, k)]^2 \\
&= [e(\hat{\boldsymbol{\alpha}}_{k-1}, \boldsymbol{\beta}, \boldsymbol{\lambda}, k) - \nu_{1,k}\|\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)\|^2 e(\hat{\boldsymbol{\alpha}}_{k-1}, \boldsymbol{\beta}, \boldsymbol{\lambda}, k)]^2 \\
&= e^2(\hat{\boldsymbol{\alpha}}_{k-1}, \boldsymbol{\beta}, \boldsymbol{\lambda}, k)[1 - \nu_{1,k}\|\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)\|^2]^2.
\end{aligned}$$

The optimal step-size $\nu_{1,k}$ can be obtained by minimizing $h_[\nu_{1,k}]$, and it is taken as

$$\nu_{1,k} = \frac{1}{\|\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)\|^2}.$$

To avoid the denominator being zero and cut down the sensitivity of the algorithm to noise in (2), we should ensure that the step-size $\nu_{1,k}$ tends to zero with k increasing. Therefore, $\nu_{1,k}$ is taken as

$$\nu_{1,k} := 1/r_{1,k}, \tag{4}$$

$$r_{1,k} := r_{1,k-1} + \|\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)\|^2, \ \ r_{1,0} = 1. \tag{5}$$

Equations (2) and (4)–(5) form the stochastic gradient (SG) algorithm for estimating $\boldsymbol{\alpha}$:

$$\hat{\boldsymbol{\alpha}}_k = \hat{\boldsymbol{\alpha}}_{k-1} + \frac{1}{r_{1,k}}\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)e(\hat{\boldsymbol{\alpha}}_{k-1}, \boldsymbol{\beta}, \boldsymbol{\lambda}, k), \tag{6}$$

$$r_{1,k} = r_{1,k-1} + \|\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)\|^2, \ \ r_{1,0} = 1. \tag{7}$$

For given $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$, the SG sub-algorithm in (6)–(7) can compute the parameter vector $\boldsymbol{\alpha}$ in a recursive way. Meanwhile, the gradient-based methods can be developed in the parameter identification problem of the nonlinear model.

3.2. Estimating the parameters of the nonlinear sub-system

For any fixed $\boldsymbol{\alpha}$, the negative gradient search is applied to derive the stochastic gradient algorithm for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ in the nonlinear sub-system. Define the derivations of the vector $\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)$ to the vectors $\boldsymbol{\beta}^{\mathrm{T}}$ and $\boldsymbol{\lambda}^{\mathrm{T}}$ respectively as

$$\begin{aligned}
\boldsymbol{\chi}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k) &:= \frac{\partial[\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{\lambda}, k)]}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \\
&= \operatorname{diag}[-(x_k - \lambda_1)^2 \exp[-\beta_1(x_k - \lambda_1)^2], -(x_k - \lambda_2)^2 \exp[-\beta_2(x_k - \lambda_2)^2], \cdots, \\
&\qquad -(x_k - \lambda_n)^2 \exp[-\beta_n(x_k - \lambda_n)^2]] \in \mathbb{R}^{n \times n},
\end{aligned}$$

DOI: 10.1002/acs

$$\begin{aligned}\boldsymbol{\Phi}(\boldsymbol{\beta},\boldsymbol{\lambda},k) &:= \frac{\partial[\boldsymbol{\eta}(\boldsymbol{\beta},\boldsymbol{\lambda},k)]}{\partial\boldsymbol{\lambda}^{\mathrm{T}}}\\ &= \mathrm{diag}[2\beta_1(x_k-\lambda_1)\exp[-\beta_1(x_k-\lambda_1)^2], 2\beta_2(x_k-\lambda_2)\exp[-\beta_2(x_k-\lambda_2)^2],\cdots,\\ &\qquad 2\beta_n(x_k-\lambda_n)\exp[-\beta_n(x_k-\lambda_n)^2]] \in \mathbb{R}^{n\times n}.\end{aligned}$$

Let $\hat{\boldsymbol{\beta}}_k := [\hat{\beta}_{1,k}, \hat{\beta}_{2,k}, \cdots, \hat{\beta}_{n,k}]^{\mathrm{T}}$ and $\hat{\boldsymbol{\lambda}}_k := [\hat{\lambda}_{1,k}, \hat{\lambda}_{2,k}, \cdots, \hat{\lambda}_{n,k}]^{\mathrm{T}}$ denote the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ at time k. Applying the negative gradient search to solve the optimization problem in (1), we derive the stochastic gradient algorithm for computing $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$:

$$\hat{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\beta}}_{k-1} + \nu_{2,k}\boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\alpha}[s_k - \boldsymbol{\eta}^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\alpha}], \tag{8}$$

$$\nu_{2,k} = \underset{\nu_{2,k}\geqslant 0}{\mathrm{argmin}} e^2(\boldsymbol{\alpha},\hat{\boldsymbol{\beta}}_k,\boldsymbol{\lambda},k), \tag{9}$$

$$\hat{\boldsymbol{\lambda}}_k = \hat{\boldsymbol{\lambda}}_{k-1} + \nu_{3,k}\boldsymbol{\Phi}(\boldsymbol{\beta},\hat{\boldsymbol{\lambda}}_{k-1},k)\boldsymbol{\alpha}[s_k - \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\beta},\hat{\boldsymbol{\lambda}}_{k-1},k)\boldsymbol{\alpha}], \tag{10}$$

$$\nu_{3,k} = \underset{\nu_{3,k}\geqslant 0}{\mathrm{argmin}} e^2(\boldsymbol{\alpha},\boldsymbol{\beta},\hat{\boldsymbol{\lambda}}_k,k), \tag{11}$$

where $\nu_{2,k}$ and $\nu_{3,k}$ are the step sizes. Similarly, the one-dimensional search methods are applied to derive the optimal step-sizes $\nu_{2,k}$ and $\nu_{3,k}$. Define $h[\nu_{2,k}] := e^2(\boldsymbol{\alpha},\hat{\boldsymbol{\beta}}_k,\boldsymbol{\lambda},k)$ for computing $\nu_{2,k}$. Then substituting the first-order Taylor expansion of $\boldsymbol{\eta}(\boldsymbol{\beta},\boldsymbol{\lambda},k)$ at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_{k-1}$ into the expression of $h[\nu_{2,k}]$ gives

$$\begin{aligned}h[\nu_{2,k}] &= [s_k - \boldsymbol{\eta}^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_k,\boldsymbol{\lambda},k)\boldsymbol{\alpha}]^2\\ &= \{s_k - [\boldsymbol{\eta}^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k) + (\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_{k-1})^{\mathrm{T}}\boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k) + o(\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_{k-1})]\boldsymbol{\alpha}\}^2\\ &= \{s_k - [\boldsymbol{\eta}^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k) + [\nu_{2,k}e(\boldsymbol{\alpha},\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\alpha}]^{\mathrm{T}}\boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\\ &\quad + o(\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_{k-1})]\boldsymbol{\alpha}\}^2\\ &= \{s_k - \boldsymbol{\eta}^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\alpha} - \nu_{2,k}e(\boldsymbol{\alpha},\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)[\boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\alpha}]^{\mathrm{T}}\boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\alpha}\\ &\quad + o(\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_{k-1})\}^2\\ &= [e(\boldsymbol{\alpha},\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k) - \nu_{2,k}e(\boldsymbol{\alpha},\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\|\boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\alpha}\|^2]^2 + o(\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_{k-1})^2\\ &= e^2(\boldsymbol{\alpha},\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)[1 - \nu_{2,k}\|\boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\alpha}\|^2]^2 + o(\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_{k-1})^2.\end{aligned}$$

The optimal step-size $\nu_{2,k}$ can be obtained by minimizing $h_[\nu_{2,k}]$, and it is taken as

$$\nu_{2,k} = \frac{1}{\|\boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\alpha}\|^2}.$$

Considering the stability of the identification algorithm, the above equation can be modified to

$$\nu_{2,k} := 1/r_{2,k},\ \ r_{2,k} := r_{2,k-1} + \|\boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\alpha}\|^2,\ \ r_{2,0} = 1. \tag{12}$$

Similarly, the step-size $\nu_{3,k}$ is taken as

$$\nu_{3,k} := 1/r_{3,k},\ \ r_{3,k} := r_{3,k-1} + \|\boldsymbol{\Phi}(\boldsymbol{\beta},\hat{\boldsymbol{\lambda}}_{k-1},k)\boldsymbol{\alpha}\|^2,\ \ r_{3,0} = 1. \tag{13}$$

Equations (8), (10) and (12)–(13) form the SG algorithm for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$:

$$\hat{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\beta}}_{k-1} + \frac{1}{r_{2,k}}e(\boldsymbol{\alpha},\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\alpha}, \tag{14}$$

$$r_{2,k} = r_{2,k-1} + \|\boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1},\boldsymbol{\lambda},k)\boldsymbol{\alpha}\|^2,\ \ r_{2,0} = 1, \tag{15}$$

$$\hat{\boldsymbol{\lambda}}_k = \hat{\boldsymbol{\lambda}}_{k-1} + \frac{1}{r_{3,k}}e(\boldsymbol{\alpha},\boldsymbol{\beta},\hat{\boldsymbol{\lambda}}_{k-1},k)\boldsymbol{\Phi}(\boldsymbol{\beta},\hat{\boldsymbol{\lambda}}_{k-1},k)\boldsymbol{\alpha}, \tag{16}$$

$$r_{3,k} = r_{3,k-1} + \|\boldsymbol{\Phi}(\boldsymbol{\beta},\hat{\boldsymbol{\lambda}}_{k-1},k)\boldsymbol{\alpha}\|^2,\ \ r_{3,0} = 1. \tag{17}$$

There are the common problems among the sub-algorithms in (2)–(3) and (14)–(17) for estimating the parameter vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$, i.e., the right-hand sides of (2)–(3) and (14)–(17) contain

DOI: 10.1002/acs

the unknown parameters in each sub-algorithm, the identification problem becomes a complex nonlinear optimization problem, the SG algorithms cannot be used directly for parameter estimation. To overcome this problem, the interactive estimation methods are presented with the hierarchical identification principle. The main idea is to use the previous estimates to substitute the unknown parameter vectors between the two sub-algorithms.

3.3. Estimating the parameters of the nonlinear systems

Considering the parameter coupling between these sub-algorithms, the previous estimates $\hat{\boldsymbol{\alpha}}_{k-1}$, $\hat{\boldsymbol{\beta}}_{k-1}$ and $\hat{\boldsymbol{\lambda}}_{k-1}$ are used to substitute the unknown parameter vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ in (2)–(3) and (14)–(17). Define the parameter vector

$$\boldsymbol{\theta} := [\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\lambda}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{3n}.$$

Let $\hat{\boldsymbol{\theta}}_k := [\hat{\boldsymbol{\alpha}}_k^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_k^{\mathrm{T}}, \hat{\boldsymbol{\lambda}}_k^{\mathrm{T}}]^{\mathrm{T}}$ denote the estimate of $\boldsymbol{\theta}$ at time k. Applying the hierarchical identification methods, we derive the H-SG algorithm for the nonlinear systems:

$$\hat{\boldsymbol{\alpha}}_k = \hat{\boldsymbol{\alpha}}_{k-1} + \frac{1}{r_{1,k}}\hat{\boldsymbol{\eta}}_k \hat{e}_k, \tag{18}$$

$$\hat{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\beta}}_{k-1} + \frac{1}{r_{2,k}}\hat{\boldsymbol{\chi}}_k \hat{\boldsymbol{\alpha}}_{k-1} \hat{e}_k, \tag{19}$$

$$\hat{\boldsymbol{\lambda}}_k = \hat{\boldsymbol{\lambda}}_{k-1} + \frac{1}{r_{3,k}}\hat{\boldsymbol{\Phi}}_k \hat{\boldsymbol{\alpha}}_{k-1} \hat{e}_k, \tag{20}$$

$$\begin{aligned} \hat{e}_k &= e(\hat{\boldsymbol{\alpha}}_{k-1}, \hat{\boldsymbol{\beta}}_{k-1}, \hat{\boldsymbol{\lambda}}_{k-1}, k) \\ &= s_k - \hat{\boldsymbol{\eta}}_k^{\mathrm{T}} \hat{\boldsymbol{\alpha}}_{k-1}, \end{aligned} \tag{21}$$

$$\begin{aligned} \hat{\boldsymbol{\eta}}_k &= \boldsymbol{\eta}(\hat{\boldsymbol{\beta}}_{k-1}, \hat{\boldsymbol{\lambda}}_{k-1}, k) \\ &= [\exp[-\hat{\beta}_{1,k-1}(x_k - \hat{\lambda}_{1,k-1})^2], \exp[-\hat{\beta}_{2,k-1}(x_k - \hat{\lambda}_{2,k-1})^2], \cdots, \\ &\quad \exp[-\hat{\beta}_{n,k-1}(x_k - \hat{\lambda}_{n,k-1})^2]]^{\mathrm{T}}, \end{aligned} \tag{22}$$

$$\begin{aligned} \hat{\boldsymbol{\chi}}_k &= \boldsymbol{\chi}(\hat{\boldsymbol{\beta}}_{k-1}, \hat{\boldsymbol{\lambda}}_{k-1}, k) \\ &= \mathrm{diag}[-(x_k - \hat{\lambda}_{1,k-1})^2 \exp[-\hat{\beta}_{1,k-1}(x_k - \hat{\lambda}_{1,k-1})^2], \\ &\qquad -(x_k - \hat{\lambda}_{2,k-1})^2 \exp[-\hat{\beta}_{2,k-1}(x_k - \hat{\lambda}_{2,k-1})^2], \cdots, \\ &\qquad -(x_k - \hat{\lambda}_{n,k-1})^2 \exp[-\hat{\beta}_{n,k-1}(x_k - \hat{\lambda}_{n,k-1})^2]], \end{aligned} \tag{23}$$

$$\begin{aligned} \hat{\boldsymbol{\Phi}}_k &= \boldsymbol{\Phi}(\hat{\boldsymbol{\beta}}_{k-1}, \hat{\boldsymbol{\lambda}}_{k-1}, k) \\ &= \mathrm{diag}[2\hat{\beta}_{1,k-1}(x_k - \hat{\lambda}_{1,k-1}) \exp[-\hat{\beta}_{1,k-1}(x_k - \hat{\lambda}_{1,k-1})^2], \\ &\qquad 2\hat{\beta}_{2,k-1}(x_k - \hat{\lambda}_{2,k-1}) \exp[-\hat{\beta}_{2,k-1}(x_k - \hat{\lambda}_{2,k-1})^2], \cdots, \\ &\qquad 2\hat{\beta}_{n,k-1}(x_k - \hat{\lambda}_{n,k-1}) \exp[-\hat{\beta}_{n,k-1}(x_k - \hat{\lambda}_{n,k-1})^2]], \end{aligned} \tag{24}$$

$$r_{1,k} = r_{1,k-1} + \|\hat{\boldsymbol{\eta}}_k\|^2, \tag{25}$$

$$r_{2,k} = r_{2,k-1} + \|\hat{\boldsymbol{\chi}}_k \hat{\boldsymbol{\alpha}}_{k-1}\|^2, \tag{26}$$

$$r_{3,k} = r_{3,k-1} + \|\hat{\boldsymbol{\Phi}}_k \hat{\boldsymbol{\alpha}}_{k-1}\|^2, \tag{27}$$

$$\hat{\boldsymbol{\theta}}_k = [\hat{\boldsymbol{\alpha}}_k^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_k^{\mathrm{T}}, \hat{\boldsymbol{\lambda}}_k^{\mathrm{T}}]^{\mathrm{T}}, \tag{28}$$

$$\hat{\boldsymbol{\alpha}}_k = [\hat{\alpha}_{1,k}, \hat{\alpha}_{2,k}, \cdots, \hat{\alpha}_{n,k}]^{\mathrm{T}}, \tag{29}$$

$$\hat{\boldsymbol{\beta}}_k = [\hat{\beta}_{1,k}, \hat{\beta}_{2,k}, \cdots, \hat{\beta}_{n,k}]^{\mathrm{T}}, \tag{30}$$

$$\hat{\boldsymbol{\lambda}}_k = [\hat{\lambda}_{1,k}, \hat{\lambda}_{2,k}, \cdots, \hat{\lambda}_{n,k}]^{\mathrm{T}}. \tag{31}$$

The computation procedure of the H-SG algorithm in (18)–(31) is summarized as follows.

1. To initialize, let $k = 1$, $\hat{\boldsymbol{\alpha}}_0 = [\hat{\alpha}_{1,0}, \hat{\alpha}_{2,0}, \cdots, \hat{\alpha}_{n,0}]^{\mathrm{T}} = \mathbf{1}_n/p_0$, $\hat{\boldsymbol{\beta}}_0 = [\hat{\beta}_{1,0}, \hat{\beta}_{2,0}, \cdots, \hat{\beta}_{n,0}]^{\mathrm{T}} = \mathbf{1}_n/p_0$, $\hat{\boldsymbol{\lambda}}_0 = [\hat{\lambda}_{1,0}, \hat{\lambda}_{2,0}, \cdots, \hat{\lambda}_{n,0}]^{\mathrm{T}} = \mathbf{1}_n/p_0$, $p_0 = 10^6$, $r_{1,0} = r_{2,0} = r_{3,0} = 1$, give an error tolerance $\varepsilon > 0$.

DOI: 10.1002/acs

2. Collect the observation data x_k and s_k, compute and form the information vector $\hat{\boldsymbol{\eta}}_k$ using (22).
3. Compute and form the information matrixes $\hat{\boldsymbol{\chi}}_k$ and $\hat{\boldsymbol{\Phi}}_k$ using (23)–(24).
4. Compute the innovation $\hat{e}_k$ using (21), compute the reciprocals of the step-sizes $r_{1,k}$, $r_{2,k}$ and $r_{3,k}$ using (25)–(27).
5. Update the parameter estimation vectors $\hat{\boldsymbol{\alpha}}_k$, $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\lambda}}_k$ using (18)–(20), form the estimate $\hat{\boldsymbol{\theta}}_k$ using (28).
6. Read out the estimates $\hat{\alpha}_{j,k}$, $\hat{\beta}_{j,k}$ and $\hat{\lambda}_{j,k}$, $j = 1, 2, \cdots, n$, from $\hat{\boldsymbol{\alpha}}_k$, $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\lambda}}_k$ in (29)–(31).
7. Compare $\hat{\boldsymbol{\theta}}_k$ with $\hat{\boldsymbol{\theta}}_{k-1}$: if $\|\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{k-1}\| > \varepsilon$, increase k by 1, and go to Step 2; otherwise, obtain the parameter estimates and terminate this procedure.

The H-SG algorithm in (18)–(31) estimates the parameter vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ in an interactive way. The innovation $\hat{e}_k$ in (21) is a scalar. In order to make full use of the observation data, we apply the multi-innovation theory and derive an interactive multi-innovation parameter estimation method in the next section.

4. THE HIERARCHICAL MULTI-INNOVATION STOCHASTIC GRADIENT ALGORITHM

The multi-innovation identification is the innovation expansion based identification. Applying the multi-innovation identification theory into basic recursive identification algorithms such as the least squares algorithm and the stochastic gradient algorithm can improve the parameter estimation accuracy [62, 63]. In the following, based on the multi-innovation identification theory, taking advantage of the innovations generated by the newest l data $\{s_k, s_{k-1}, \cdots, s_{k-l+1}\}$ to expand the innovation $\hat{e}_k$ in (18)–(20) into the innovation vector $\boldsymbol{E}(l)$ gives

$$\boldsymbol{E}(l) := \boldsymbol{S}(l) - \hat{\boldsymbol{\Gamma}}^{\mathrm{T}}(l)\hat{\boldsymbol{\alpha}}_{k-1} \in \mathbb{R}^l, \tag{32}$$

where l denotes the innovation length, $\boldsymbol{S}(l)$ is the stacked output vector of s_k, $\hat{\boldsymbol{\Gamma}}(l)$ is the stacked information matrix of $\hat{\boldsymbol{\eta}}_k$, i.e.,

$$\boldsymbol{S}(l) := [s_k, s_{k-1}, \cdots, s_{k-l+1}]^{\mathrm{T}} \in \mathbb{R}^l, \tag{33}$$

$$\hat{\boldsymbol{\Gamma}}(l) := [\hat{\boldsymbol{\eta}}_k, \hat{\boldsymbol{\eta}}_{k-1}, \cdots, \hat{\boldsymbol{\eta}}_{k-l+1}] \in \mathbb{R}^{n\times l}. \tag{34}$$

Since $\boldsymbol{E}(1) = \hat{e}_k$, $\boldsymbol{S}(1) = s_k$ and $\hat{\boldsymbol{\Gamma}}(1) = \hat{\boldsymbol{\eta}}_k$ for $l = 1$, based on the stochastic gradient algorithm in (18)–(20) and (25)–(27), the H-MISG algorithm for estimating $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ can be written as

$$\hat{\boldsymbol{\alpha}}_k = \hat{\boldsymbol{\alpha}}_{k-1} + \frac{1}{r_{1,k}}\hat{\boldsymbol{\Gamma}}(l)\boldsymbol{E}(l), \tag{35}$$

$$\hat{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\beta}}_{k-1} + \frac{1}{r_{2,k}}\hat{\boldsymbol{\alpha}}_{k-1} \circ \hat{\boldsymbol{\Psi}}(l)\boldsymbol{E}(l), \tag{36}$$

$$\hat{\boldsymbol{\lambda}}_k = \hat{\boldsymbol{\lambda}}_{k-1} + \frac{1}{r_{3,k}}\hat{\boldsymbol{\alpha}}_{k-1} \circ \hat{\boldsymbol{\Xi}}(l)\boldsymbol{E}(l), \tag{37}$$

$$r_{1,k} = r_{1,k-1} + \|\hat{\boldsymbol{\eta}}_k\|^2, \ \ r_{1,0} = 1, \tag{38}$$

$$r_{2,k} = r_{2,k-1} + \|\hat{\boldsymbol{\alpha}}_{k-1} \circ \hat{\boldsymbol{\psi}}_k\|^2, \ \ r_{2,0} = 1, \tag{39}$$

$$r_{3,k} = r_{3,k-1} + \|\hat{\boldsymbol{\alpha}}_{k-1} \circ \hat{\boldsymbol{\zeta}}_k\|^2, \ \ r_{3,0} = 1, \tag{40}$$

where the symbol $\boldsymbol{B} \circ \boldsymbol{C}$ represents the Hadamard product, that is, $\boldsymbol{B} = [b_{ij}] \in \mathbb{R}^{p\times q}$, $\boldsymbol{C} = [c_{ij}] \in \mathbb{R}^{p\times q}$, $\boldsymbol{B} \circ \boldsymbol{C} = [b_{ij}c_{ij}] \in \mathbb{R}^{p\times q}$, the information vectors $\hat{\boldsymbol{\psi}}_k$ and $\hat{\boldsymbol{\zeta}}_k$ and the stacked information matrixes $\hat{\boldsymbol{\Psi}}(l)$ and $\hat{\boldsymbol{\Xi}}(l)$ are defined respectively as

$$\begin{aligned}\hat{\boldsymbol{\psi}}_k := [&-(x_k - \hat{\lambda}_{1,k-1})^2 \exp[-\hat{\beta}_{1,k-1}(x_k - \hat{\lambda}_{1,k-1})^2],\\ &-(x_k - \hat{\lambda}_{2,k-1})^2 \exp[-\hat{\beta}_{2,k-1}(x_k - \hat{\lambda}_{2,k-1})^2], \cdots,\end{aligned}$$

DOI: 10.1002/acs

$$-(x_k-\hat{\lambda}_{n,k-1})^2\exp[-\hat{\beta}_{n,k-1}(x_k-\hat{\lambda}_{n,k-1})^2]]^{\mathrm{T}}\in\mathbb{R}^n, \tag{41}$$

$$\begin{aligned}\hat{\boldsymbol{\zeta}}_k:=&[2\hat{\beta}_{1,k-1}(x_k-\hat{\lambda}_{1,k-1})\exp[-\hat{\beta}_{1,k-1}(x_k-\hat{\lambda}_{1,k-1})^2],\\ &2\hat{\beta}_{2,k-1}(x_k-\hat{\lambda}_{2,k-1})\exp[-\hat{\beta}_{2,k-1}(x_k-\hat{\lambda}_{2,k-1})^2],\cdots,\\ &2\hat{\beta}_{n,k-1}(x_k-\hat{\lambda}_{n,k-1})\exp[-\hat{\beta}_{n,k-1}(x_k-\hat{\lambda}_{n,k-1})^2]]^{\mathrm{T}}\in\mathbb{R}^n,\end{aligned} \tag{42}$$

$$\hat{\boldsymbol{\Psi}}(l):=[\hat{\boldsymbol{\psi}}_k,\hat{\boldsymbol{\psi}}_{k-1},\cdots,\hat{\boldsymbol{\psi}}_{k-l+1}]\in\mathbb{R}^{n\times l}, \tag{43}$$

$$\hat{\boldsymbol{\Xi}}(l):=[\hat{\boldsymbol{\zeta}}_k,\hat{\boldsymbol{\zeta}}_{k-1},\cdots,\hat{\boldsymbol{\zeta}}_{k-l+1}]\in\mathbb{R}^{n\times l}. \tag{44}$$

Equations (32)–(44) form the H-MISG algorithm for the nonlinear systems:

$$\hat{\boldsymbol{\alpha}}_k=\hat{\boldsymbol{\alpha}}_{k-1}+\frac{1}{r_{1,k}}\hat{\boldsymbol{\Gamma}}(l)\boldsymbol{E}(l), \tag{45}$$

$$\hat{\boldsymbol{\beta}}_k=\hat{\boldsymbol{\beta}}_{k-1}+\frac{1}{r_{2,k}}\hat{\boldsymbol{\alpha}}_{k-1}\circ\hat{\boldsymbol{\Psi}}(l)\boldsymbol{E}(l), \tag{46}$$

$$\hat{\boldsymbol{\lambda}}_k=\hat{\boldsymbol{\lambda}}_{k-1}+\frac{1}{r_{3,k}}\hat{\boldsymbol{\alpha}}_{k-1}\circ\hat{\boldsymbol{\Xi}}(l)\boldsymbol{E}(l), \tag{47}$$

$$\boldsymbol{E}(l)=\boldsymbol{S}(l)-\hat{\boldsymbol{\Gamma}}^{\mathrm{T}}(l)\hat{\boldsymbol{\alpha}}_{k-1}, \tag{48}$$

$$\boldsymbol{S}(l)=[s_k,s_{k-1},\cdots,s_{k-l+1}]^{\mathrm{T}}, \tag{49}$$

$$\hat{\boldsymbol{\Gamma}}(l)=[\hat{\boldsymbol{\eta}}_k,\hat{\boldsymbol{\eta}}_{k-1},\cdots,\hat{\boldsymbol{\eta}}_{k-l+1}], \tag{50}$$

$$\hat{\boldsymbol{\Psi}}(l)=[\hat{\boldsymbol{\psi}}_k,\hat{\boldsymbol{\psi}}_{k-1},\cdots,\hat{\boldsymbol{\psi}}_{k-l+1}], \tag{51}$$

$$\hat{\boldsymbol{\Xi}}(l)=[\hat{\boldsymbol{\zeta}}_k,\hat{\boldsymbol{\zeta}}_{k-1},\cdots,\hat{\boldsymbol{\zeta}}_{k-l+1}], \tag{52}$$

$$\begin{aligned}\hat{\boldsymbol{\eta}}_k=&[\exp[-\hat{\beta}_{1,k-1}(x_k-\hat{\lambda}_{1,k-1})^2],\exp[-\hat{\beta}_{2,k-1}(x_k-\hat{\lambda}_{2,k-1})^2],\cdots,\\ &\exp[-\hat{\beta}_{n,k-1}(x_k-\hat{\lambda}_{n,k-1})^2]]^{\mathrm{T}},\end{aligned} \tag{53}$$

$$\begin{aligned}\hat{\boldsymbol{\psi}}_k=&[-(x_k-\hat{\lambda}_{1,k-1})^2\exp[-\hat{\beta}_{1,k-1}(x_k-\hat{\lambda}_{1,k-1})^2],\\ &-(x_k-\hat{\lambda}_{2,k-1})^2\exp[-\hat{\beta}_{2,k-1}(x_k-\hat{\lambda}_{2,k-1})^2],\cdots,\\ &-(x_k-\hat{\lambda}_{n,k-1})^2\exp[-\hat{\beta}_{n,k-1}(x_k-\hat{\lambda}_{n,k-1})^2]]^{\mathrm{T}},\end{aligned} \tag{54}$$

$$\begin{aligned}\hat{\boldsymbol{\zeta}}_k=&[2\hat{\beta}_{1,k-1}(x_k-\hat{\lambda}_{1,k-1})\exp[-\hat{\beta}_{1,k-1}(x_k-\hat{\lambda}_{1,k-1})^2],\\ &2\hat{\beta}_{2,k-1}(x_k-\hat{\lambda}_{2,k-1})\exp[-\hat{\beta}_{2,k-1}(x_k-\hat{\lambda}_{2,k-1})^2],\cdots,\\ &2\hat{\beta}_{n,k-1}(x_k-\hat{\lambda}_{n,k-1})\exp[-\hat{\beta}_{n,k-1}(x_k-\hat{\lambda}_{n,k-1})^2]]^{\mathrm{T}},\end{aligned} \tag{55}$$

$$r_{1,k}=r_{1,k-1}+\|\hat{\boldsymbol{\eta}}_k\|^2, \tag{56}$$

$$r_{2,k}=r_{2,k-1}+\|\hat{\boldsymbol{\alpha}}_{k-1}\circ\hat{\boldsymbol{\psi}}_k\|^2, \tag{57}$$

$$r_{3,k}=r_{3,k-1}+\|\hat{\boldsymbol{\alpha}}_{k-1}\circ\hat{\boldsymbol{\zeta}}_k\|^2, \tag{58}$$

$$\hat{\boldsymbol{\theta}}_k=[\hat{\boldsymbol{\alpha}}_k^{\mathrm{T}},\hat{\boldsymbol{\beta}}_k^{\mathrm{T}},\hat{\boldsymbol{\lambda}}_k^{\mathrm{T}}]^{\mathrm{T}}, \tag{59}$$

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

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

$$\hat{\boldsymbol{\lambda}}_k=[\hat{\lambda}_{1,k},\hat{\lambda}_{2,k},\cdots,\hat{\lambda}_{n,k}]^{\mathrm{T}}. \tag{62}$$

The computation procedure of the H-MISG algorithm in (45)–(62) is summarized as follows.

1. To initialize, let $k=1$, $\hat{\boldsymbol{\alpha}}_0=[\hat{\alpha}_{1,0},\hat{\alpha}_{2,0},\cdots,\hat{\alpha}_{n,0}]^{\mathrm{T}}=\mathbf{1}_n/p_0$, $\hat{\boldsymbol{\beta}}_0=[\hat{\beta}_{1,0},\hat{\beta}_{2,0},\cdots,\hat{\beta}_{n,0}]^{\mathrm{T}}=\mathbf{1}_n/p_0$, $\hat{\boldsymbol{\lambda}}_0=[\hat{\lambda}_{1,0},\hat{\lambda}_{2,0},\cdots,\hat{\lambda}_{n,0}]^{\mathrm{T}}=\mathbf{1}_n/p_0$, $p_0=10^6$, $r_{1,0}=r_{2,0}=r_{3,0}=1$, give an error tolerance $\varepsilon>0$.
2. Collect the observation data x_k and s_k, form the stacked output vector $\boldsymbol{S}(l)$ using (49).
3. Compute and form the information vector $\hat{\boldsymbol{\eta}}_k$ using (53), form the stacked information matrix $\hat{\boldsymbol{\Gamma}}(l)$ using (50).
4. Compute and form the information vectors $\hat{\boldsymbol{\psi}}_k$ and $\hat{\boldsymbol{\zeta}}_k$ using (54)–(55), form the stacked information matrixes $\hat{\boldsymbol{\Psi}}(l)$ and $\hat{\boldsymbol{\Xi}}(l)$ using (51)–(52).

DOI: 10.1002/acs

5. Compute the innovation vector $\boldsymbol{E}(l)$ using (48), compute the reciprocals of the step-sizes $r_{1,k}$, $r_{2,k}$ and $r_{3,k}$ using (56)–(58).
6. Update the parameter estimation vectors $\hat{\boldsymbol{\alpha}}_k$, $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\lambda}}_k$ using (45)–(47), form the estimate $\hat{\boldsymbol{\theta}}_k$ using (59).
7. Read out the estimates $\hat{\alpha}_{j,k}$, $\hat{\beta}_{j,k}$ and $\hat{\lambda}_{j,k}$, $j = 1, 2, \cdots, n$, from $\hat{\boldsymbol{\alpha}}_k$, $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\lambda}}_k$ in (60)–(62).
8. Compare $\hat{\boldsymbol{\theta}}_k$ with $\hat{\boldsymbol{\theta}}_{k-1}$: if $\|\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{k-1}\| > \varepsilon$, increase k by 1, and go to Step 2; otherwise, obtain the parameter estimates and terminate this procedure.

At each recursion, the H-SG algorithm involves the current measurement data and innovation, the H-MISG algorithm applies all the current and the preceding $(l-1)$ measurement data and innovations, which makes the latter have a higher parameter estimation accuracy. The introduction of the innovation length increases the number of data used in the algorithm compared with the H-SG algorithm. Although the parameter estimation accuracy can be effectively improved, too long innovation length leads to a large amount of computation. However, the increased computation burden is not heavy, and it is tolerable.

In order to obtain more accurate parameter estimates without increasing the computational cost of the H-MISG algorithm, we introduce the forgetting factors γ_1, γ_2 and γ_3 into (56)–(58):

$$r_{1,k} = \gamma_1 r_{1,k-1} + \|\hat{\boldsymbol{\eta}}_k\|^2, \ \ 0 \leq \gamma_1 < 1, \tag{63}$$

$$r_{2,k} = \gamma_2 r_{2,k-1} + \|\hat{\boldsymbol{\alpha}}_{k-1} \circ \hat{\boldsymbol{\psi}}_k\|^2, \ \ 0 \leq \gamma_2 < 1, \tag{64}$$

$$r_{3,k} = \gamma_3 r_{3,k-1} + \|\hat{\boldsymbol{\alpha}}_{k-1} \circ \hat{\boldsymbol{\zeta}}_k\|^2, \ \ 0 \leq \gamma_3 < 1. \tag{65}$$

Replacing (56)–(58) in the H-MISG algorithm with (63)–(65), we obtain the variant of the H-MISG algorithm, that is, the H-MIFG algorithm for the nonlinear systems. The methods proposed in this paper can combine some statistical tools and optimal strategies [64–72] and idenfitication algorithms [73–80, 80, 81] to study the parameter estimation issues of linear and nonlinear systems with colored noises and can be applied to other fields [82–86] such as mechnical systems and power systems. Before using the proposed algorithms to identify the nonlinear systems, we should consider the order determination of the kernel functions. The orthogonalization procedure and correlation analysis can be adopted to determine the orders [87].

5. EXAMPLE

Consider the following nonlinear system:

$$\begin{aligned} s_k &= \alpha_1 \exp[-\beta_1 (x_k - \lambda_1)^2] + \alpha_2 \exp[-\beta_2 (x_k - \lambda_2)^2] + v_k \\ &= 0.9 \exp[-0.25 (x_k - 1.2)^2] + 0.45 \exp[-0.64 (x_k + 1.2)^2] + v_k. \end{aligned}$$

The parameters to be estimated are

$$\boldsymbol{\theta} = [\alpha_1, \alpha_2, \beta_1, \beta_2, \lambda_1, \lambda_2]^{\mathrm{T}} = [0.90, 0.45, 0.25, 0.64, 1.20, -1.20]^{\mathrm{T}}.$$

In simulation, the input $\{x_k\}$ is taken as a uniformly distributed random signal sequence between the intervals $[-5, 5]$ with zero mean, the disturbance $\{v_k\}$ is taken as a white noise sequence with zero mean and variance σ^2. The measurement data length is taken as $L_e = 2000$.

To illustrate the advantage of the proposed multi-innovation identification algorithm, fix the noise variance $\sigma^2 = 0.10^2$, and use the H-SG algorithm and the H-MISG algorithm with the innovation lengths $l = 2$ and $l = 3$ to identify the nonlinear system, respectively. The parameter estimates and their errors are shown in Tables I–II, and the parameter estimation errors $\delta := \|\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}\| / \|\boldsymbol{\theta}\|$ versus k are shown in Figure 1, from which the following conclusions can be drawn: (1) the parameter estimation errors decrease as the data length k increases for all the algorithms proposed in this paper, the H-MISG algorithm has higher parameter estimation accuracy than the H-SG algorithm; (2) the parameter estimation accuracy of the H-MISG algorithm increases with the innovation length l increasing.

DOI: 10.1002/acs

Table I. The H-SG estimates and errors with $\sigma^2 = 0.10^2$

k	α_1	α_2	β_1	β_2	λ_1	λ_2	$\delta(\%)$
100	1.24809	0.30464	0.42691	0.31561	1.08186	-1.21408	25.90844
200	1.17344	0.32579	0.39428	0.34327	1.07897	-1.23349	22.18602
500	1.10139	0.34515	0.35066	0.37079	1.09779	-1.26123	18.43658
1000	1.05922	0.34868	0.33046	0.38711	1.11689	-1.27472	16.47982
2000	1.02677	0.35935	0.30482	0.40012	1.13062	-1.28524	14.90152
True values	0.90000	0.45000	0.25000	0.64000	1.20000	-1.20000	

Table II. The H-MISG estimates and errors with $l = 2$, $l = 3$ and $\sigma^2 = 0.10^2$

l	k	α_1	α_2	β_1	β_2	λ_1	λ_2	$\delta(\%)$
2	100	1.06884	0.29424	0.28249	0.53340	0.95716	-1.16588	16.94499
	200	1.01139	0.30826	0.25719	0.53955	0.99778	-1.21046	13.84334
	500	0.95958	0.31519	0.23337	0.54549	1.06234	-1.25980	11.06718
	1000	0.93944	0.32196	0.22838	0.54854	1.07148	-1.27356	10.56413
	2000	0.92290	0.33117	0.22156	0.55174	1.07876	-1.28379	10.14977
3	100	0.96552	0.42678	0.25872	0.63463	1.06055	-1.21844	7.52629
	200	0.94461	0.42616	0.25475	0.63987	1.13168	-1.25819	4.93622
	500	0.92685	0.42054	0.24314	0.64324	1.19381	-1.26884	3.83622
	1000	0.91942	0.42250	0.24529	0.64488	1.17695	-1.25277	3.20929
	2000	0.91185	0.42721	0.24025	0.64688	1.17659	-1.24294	2.70537
True values		0.90000	0.45000	0.25000	0.64000	1.20000	-1.20000	

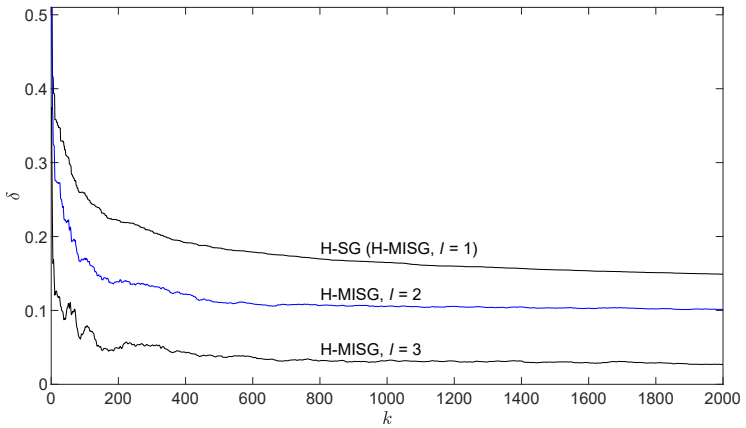

Figure 1. The H-SG and H-MISG estimation errors δ against k with $l = 2$, $l = 3$ and $\sigma^2 = 0.10^2$

To study the influence of different noises on parameter estimation, fix $l = 3$, and apply the H-MISG algorithm with the noise variances $\sigma^2 = 0.10^2$, $\sigma^2 = 0.15^2$ and $\sigma^2 = 0.20^2$ to identify the nonlinear system, respectively. The results are exhibited in Table III and Figure 2, from which we can see that the parameter estimation errors of the H-MISG algorithm become small as the noise variance decreases.

To verify the performance of the H-MIFG algorithm, fix $\sigma^2 = 0.10^2$, and use the H-SG algorithm, the H-MISG algorithm with $l = 3$ and the H-MIFG algorithm with $l = 3$ and $\gamma_1 = 0.99$ to estimate the nonlinear system, respectively. The parameter estimates and their errors are listed in Tables II and IV, the parameter estimation errors δ versus k are shown in Figure 3. From which we can see that the parameter estimation errors decrease after introducing a forgetting factor, and the H-MIFG

DOI: 10.1002/acs

Table III. The H-MISG estimates and errors with $l = 3$, $\sigma^2 = 0.15^2$ and $\sigma^2 = 0.20^2$

σ^2	k	α_1	α_2	β_1	β_2	λ_1	λ_2	$\delta(\%)$
0.15^2	100	0.95769	0.45762	0.27223	0.51632	1.08362	-1.18255	8.69853
	200	0.94065	0.45135	0.27097	0.52903	1.18894	-1.22337	5.87715
	500	0.92138	0.44130	0.25482	0.53789	1.27437	-1.23411	6.36388
	1000	0.91171	0.44280	0.25819	0.54318	1.24537	-1.20525	5.18116
	2000	0.90110	0.44949	0.25005	0.54934	1.24401	-1.19100	4.84347
0.20^2	100	0.96172	0.48846	0.28504	0.48432	1.07916	-1.15785	10.39167
	200	0.94642	0.47710	0.28610	0.50268	1.22123	-1.20562	7.34268
	500	0.92571	0.46145	0.26465	0.51529	1.32658	-1.21690	8.67809
	1000	0.91322	0.46198	0.26843	0.52282	1.28640	-1.17965	7.14288
	2000	0.89989	0.46948	0.25720	0.53153	1.28444	-1.16259	6.89113
True values		0.90000	0.45000	0.25000	0.64000	1.20000	-1.20000	

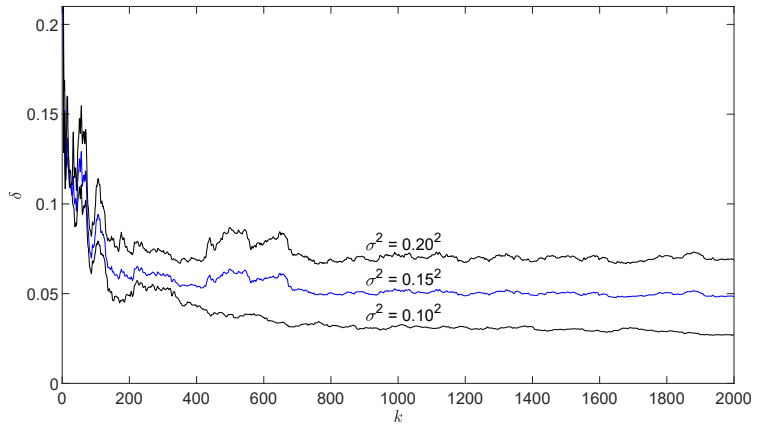

Figure 2. The H-MISG estimation errors δ against k with $l = 3$, $\sigma^2 = 0.10^2$, $\sigma^2 = 0.15^2$ and $\sigma^2 = 0.20^2$

algorithm with the appropriate innovation length and forgetting factor is effective to identify the nonlinear system.

Table IV. The H-MIFG estimates and errors with $l = 3$, $\gamma_1 = 0.99$ and $\sigma^2 = 0.10^2$

k	α_1	α_2	β_1	β_2	λ_1	λ_2	$\delta(\%)$
100	0.95854	0.44696	0.25986	0.63426	1.07941	-1.20499	6.44581
200	0.93072	0.44529	0.25736	0.63932	1.15212	-1.24413	3.47194
500	0.92741	0.42271	0.24329	0.64154	1.20988	-1.25015	3.08588
1000	0.91150	0.44905	0.24625	0.64276	1.18898	-1.23554	1.87778
2000	0.91004	0.46284	0.24447	0.64450	1.19583	-1.22542	1.49837
True values	0.90000	0.45000	0.25000	0.64000	1.20000	-1.20000	

Figure 4 shows the parameter search process (solid line: the estimates, dotted line: the true values) of the H-MIFG algorithm under $l = 3$, $\gamma_1 = 0.99$ and $\sigma^2 = 0.10^2$. As shown in Figure 4, all the estimated parameters gradually approach the true values as the data length k increases.

For the model validation, we use $L_r = 100$ observations from $k = L_e + 1$ to $k = L_e + L_r$ and the predicted model by the H-MIFG algorithm with $l = 3$, $\gamma_1 = 0.99$ and $\sigma^2 = 0.10^2$. In order to verify the fitting effect, the input $\{x_k\}$ between $k = L_e + 1$ and $k = L_e + L_r$ is taken as a uniformly distributed signal sequence between the intervals $[-5, 5]$. The predicted data $\hat{s}_k$ and the measurement

DOI: 10.1002/acs

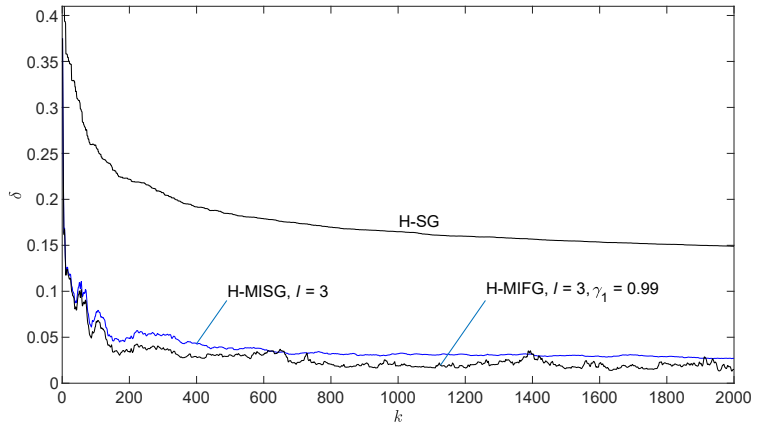

Figure 3. The H-SG, H-MISG and H-MIFG estimation errors δ against k with $\sigma^2 = 0.10^2$

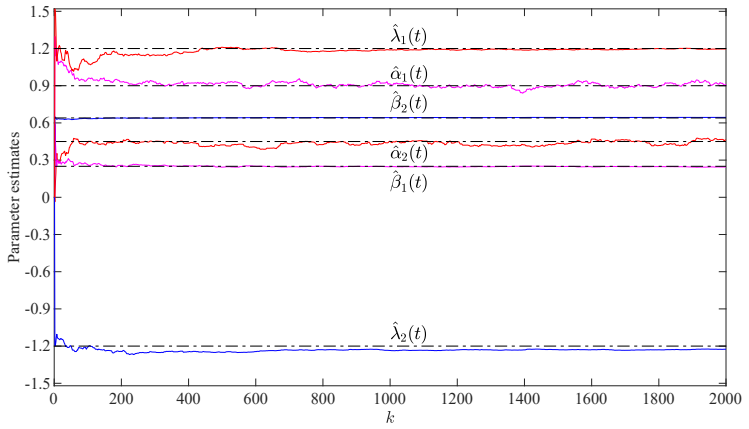

Figure 4. Parameter estimates against k for the H-MIFG algorithm with $l = 3$, $\gamma_1 = 0.99$ and $\sigma^2 = 0.10^2$

data s_k are plotted in Figure 5. To evaluate the prediction performance, we define and compute the mean square error (MSE) as follows

$$\text{MSE} := \left[\frac{1}{L_r}\sum_{k=L_e+1}^{L_e+L_r}[\hat{s}_k - s_k]^2\right]^{\frac{1}{2}} = 0.10875.$$

The MSE is close to the standard deviation of the noise. Combine with Figure 5, it is clear that the predicted data is close to the measurement data, which means the kernel functions have a good performance in fitting the nonlinear systems.

6. CONCLUSIONS

The novel kernel function is constructed for fitting the nonlinear systems in this paper, which can be applied to the processes of fuzzy control. Then we focus on the identification problems of the nonlinear systems based on the kernel functions. Applying the hierarchical principle, we propose the H-SG algorithm for the nonlinear systems. The optimal step sizes are deduced by using the one-dimensional search methods. To improve the estimation accuracy, we present the H-MISG algorithm

DOI: 10.1002/acs

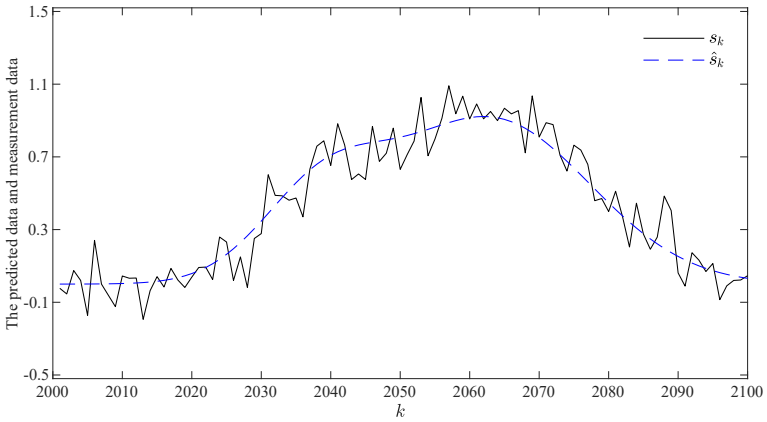

Figure 5. The predicted data $\hat{s}_k$ and the measurement data s_k for the H-MIFG algorithm

for the nonlinear systems by using the multi-innovation theory. Moreover, the forgetting factors are introduced to obtain the H-MIFG algorithm which has more accurate parameter estimates. The simulation results indicate that the H-MIFG algorithm with the appropriate innovation length and forgetting factor is effective to identify the nonlinear systems. In the future work, we will extend the methods to other nonlinear systems with different structures and disturbances, and study the estimation methods of model orders. The proposed approaches proposed in the paper can be extended to study the parameter identification problems of other nonlinear systems with colored noises and can be applied to other fields [88–93] such as signal analysis and engineering application systems [94–103], information processing, transportation communication systems [104–111] and so on.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (No. 61873111) and the 111 Project (B12018).

DATA AVAILABILITY STATEMENT

All data generated or analyzed during this study are included in this article.

ORCID

Feng Ding https://orcid.org/0000-0002-2721-2025

REFERENCES

1. Li MH, Liu XM. Iterative parameter estimation methods for dual-rate sampled-data bilinear systems by means of the data filtering technique. *IET Control Theory and Applications*. 2021;15. doi:10.1049/cth2.12118
2. Ding J, Cao ZX, Chen JZ, Jiang GP. Weighted parameter estimation for Hammerstein nonlinear ARX systems. *Circuits Systems and Signal Processing*. 2020;39(4):2178-2192.
3. Chakravorti T, Satyanarayana P. Non linear system identification using kernel based exponentially extended random vector functional link network. *Applied Soft Computing*. 2020;89:106117.
4. Li MH, Liu XM. 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*. 2020;18(6):1581-1592.
5. Ding F, 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*. 2017;62:211-223.

6. Wang XH, et al. The modified extended Kalman filter based recursive estimation for Wiener nonlinear systems with process noise and measurement noise. *International Journal of Adaptive Control and Signal Processing*. 2020;34(10):1321-1340.
7. Ding F, Xu L, Zhu Q. Performance analysis of the generalised projection identification for time-varying systems. *IET Control Theory and Applications*. 2016;10(18):2506-2514.
8. Xu L, Chen L, Xiong WL. Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration. *Nonlinear Dynamics*. 2015;79(3):2155-2163.
9. Xu L. The damping iterative parameter identification method for dynamical systems based on the sine signal measurement. *Signal Processing*. 2016;120:660-667.
10. Li JH, Zong TC, Gu JP. Parameter estimation of Wiener systems based on the particle swarm iteration and gradient search principle. *Circuits Systems and Signal Processing*. 2020;39(7):3470-3495.
11. Wan LJ, et al. A new iterative least squares parameter estimation approach for equation-error autoregressive systems. *International Journal of Control Automation and Systems*. 2020;18(3):780-790.
12. 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.
13. 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.
14. Xu L. The parameter estimation algorithms based on the dynamical response measurement data. *Advances in Mechanical Engineering*. 2017;9(11). Article Numbeer: 1687814017730003.
15. Xu L, 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*. 2019;50(1):141-151.
16. Vaezi M, Izadian A. Piecewise affine system identification of a hydraulic wind power transfer system. *IEEE Transactions on Control Systems Technology*. 2015;23(6):2077-2086.
17. Ji Y, Zhang C, Kang Z, et al. Parameter estimation for block-oriented nonlinear systems using the key term separation. *International Journal of Robust and Nonlinear Control*. 2020;30(9):3727-3752.
18. Fan YM, Liu XM. 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*. 2020;30(14):5492-5509.
19. Gu Y, Zhu Q, Nouri H. Bias compensation-based parameter and state estimation for a class of time-delay nonlinear state-space models. *IET Control Theory and Applications*. 2020;14(15):2176-2185.
20. Cheng SS, Wei YH, Sheng D. Identification for Hammerstein nonlinear ARMAX systems based on multi-innovation fractional order stochastic gradient. *Signal Processing*. 2018;142:1-10.
21. Lin CM, Li HY. Adaptive dynamic sliding-mode fuzzy CMAC for voice coil motor using asymmetric Gaussian membership function. *IEEE Transactions on Industrial Electronics*. 2014;61(10):5662-5671.
22. Li JD. A data-driven improved fuzzy logic control optimization-simulation tool for reducing flooding volume at downstream urban drainage systems. *Science of the Total Environment*. 2020;732:138931.
23. Wang YX, Qin FF, Qu K, et al. Temperature control for a polymer electrolyte membrane fuel cell by using fuzzy rule. *IEEE Transactions on Energy Conversion*. 2016;31(2):674-682.
24. Wei ZX, Doctor F, Liu YX, et al. An optimized type-2 self-organizing fuzzy logic controller applied in anesthesia for propofol dosing to regulate BIS. *IEEE Transactions on Fuzzy Systems*. 2020;28(6):1062-1072.
25. Lindblad J, Sladoje N. Linear time distances between fuzzy sets with applications to pattern matching and classification. *IEEE Transactions on Image Processing*. 2014;23(1):126-136.
26. Bozorgmehr A, Jooq MKQ, Moaiyeri MH. A high-performance fully programmable membership function generator based on 10 nm gate-all-around CNTFETs. *International Journal of Electronics and Communications*. 2020;123:153293.
27. Jiang ZQ, Wu WJ, Qin H. Optimization of fuzzy membership function of runoff forecasting error based on the optimal closeness. *Journal of Hydrology*. 2019;570:51-61.
28. Sihag P, Tiwari NK, Ranjan S. Prediction of unsaturated hydraulic conductivity using adaptive neuro-fuzzy inference system (ANFIS). *ISH Journal of Hydraulic Engineering*. 2019;25(2):132-142.
29. Chattopadhyay S. Neurofuzzy models to automate the grading of old-age depression. *Expert Systems*. 2014;31(1):48-55.
30. Pradhan SK, Parhi DR, Panda AK. Fuzzy logic techniques for navigation of several mobile robots. *Applied Soft Computing*. 2009;9(1):290-304.
31. Huynh TT, Lin CM, Le TL. A new self-organizing fuzzy cerebellar model articulation controller for uncertain nonlinear systems using overlapped Gaussian membership functions. *IEEE Transactions on Industrial Electronics*. 2020;67(11):9671-9682.
32. Li CS, Zhou JZ, Chang L. T-S fuzzy model identification based on a novel hyperplane-shaped membership function. *IEEE Transactions on Fuzzy Systems*. 2017;25(5):1364-1370.
33. Guo H. A simple algorithm for fitting a Gaussian function. *IEEE Signal Processing Magazine*. 2011;28(5):134-137.
34. Roonizi EK. A new algorithm for fitting a Gaussian function riding on the polynomial background. *IEEE Signal Processing Letters*. 2013;20(11):1062-1065.
35. Wang LJ, Ji Y, Wan LJ, et al. Hierarchical recursive generalized extended least squares estimation algorithms for a class of nonlinear stochastic systems with colored noise. *Journal of the Franklin Institute*. 2019;356(16):10102-10122.
36. Zhou YH, et al. Modeling nonlinear processes using the radial basis function-based state-dependent autoregressive models. *IEEE Signal Processing Letters*. 2020;27:1600-1604.
37. Zhou YH, et al. Hierarchical estimation approach for RBF-AR models with regression weights based on the increasing data length. *IEEE Transactions on Circuits and Systems-II: Express Briefs*. 2021;68.
38. Li MH, Liu XM. The least squares based iterative algorithms for parameter estimation of a bilinear system with autoregressive noise using the data filtering technique. *Signal Processing*. 2018;147:23-34.

DOI: 10.1002/acs

39. Cui T, et al. Joint multi-innovation recursive extended least squares parameter and state estimation for a class of state-space systems. *International Journal of Control Automation and Systems*. 2020;18(6):1412-1424.
40. Li MH, 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*. 2019;33(7):1189-1211.
41. 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.
42. Wang C, Zhu L. Parameter identification of a class of nonlinear systems based on the multi-innovation identification theory. *Journal of the Franklin Institute*. 2015;352(10):4624-4637.
43. Miguel Adanez J, Mohammed Al-Hadithi B, Jimenez A. Multidimensional membership functions in T-S fuzzy models for modelling and identification of nonlinear multivariable systems using genetic algorithms. *Applied Soft Computing*. 2019;75:607-615.
44. Kim S, Lee M, Lee J. A study of fuzzy membership functions for dependence decision-making in security robot system. *Neural Computing & Applications*. 2017;28(1):155-164.
45. Zhang X, et al. Highly computationally efficient state filter based on the delta operator. *International Journal of Adaptive Control and Signal Processing*. 2019;33(6):875-889.
46. Zhang X, 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*. 2019;33(7):1157-1173.
47. Zhang X, et al. State filtering-based least squares parameter estimation for bilinear systems using the hierarchical identification principle. *IET Control Theory and Applications*. 2018;12(12):1704-1713.
48. Zhang X,et al. Combined state and parameter estimation for a bilinear state space system with moving average noise. *Journal of the Franklin Institute*. 2018;355(6):3079-3103.
49. Zhang X, et al. Recursive paramete
50. Zhan XS, Hu JW, Wu J, Yan HC. Analysis of optimal performance of MIMO NCS with encoding and packet dropout constraints. *IET Control Theory and Applications*. 2020;14(13):1762-1768.
51. Han T, Zheng WX. Bipartite output consensus for heterogeneous multi-agent systems via output regulation approach. *IEEE Transactions on Circuits and Systems II: Express Briefs*. 2021;68(1):281-285.
52. H. Ma, X. Zhang, Q.Y. Liu, F. Ding, X.B. Jin, A. Alsaedi, T. Hayat, 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.
53. Pan J, Jiang X, Wan XK, Ding W. A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems. *International Journal of Control Automation and Systems*. 2017;15(3):1189-1197.
54. 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.
55. Wu MH, Yue HH, Wang J, et al. Object detection based on RGC mask R-CNN, *IET Image Processing*. 2020;14(8):1502-1508.
56. Wan XK, Jin ZY, Wu HB et al, Heartbeat classification algorithm based on one-dimensional convolution neural network. *Journal of Mechanics in Medicine and Biology*. 2020;20(7):2050046.
57. Zhang Y, Yan Z, Zhou CC, et al. Capacity allocation of HESS in micro-grid based on ABC algorithm. *International Journal of Low-Carbon Technologies*. 2020;15(4):496-505.
58. Zhao N, Fan P, Cheng Y. Dynamic contract incentives mechanism for traffic offloading in multi-UAV networks. *Wireless Communications and Mobile Computing*. 2020;2020. Article ID 2361029.
59. Wang LJ, Guo J, Xu C, et al. Hybrid model predictive control strategy of supercapacitor energy storage system based on double active bridge. *Energies*. 2019;12(11):2134.
60. Liu H, Zou QX, Zhang ZP. Energy disaggregation of appliances consumptions using ham approach. *IEEE Access*. 2019;7:185977-185990.
61. Ji Y, Kang Z. Three-stage forgetting factor stochastic gradient parameter estimation methods for a class of nonlinear systems. *International Journal of Robust and Nonlinear Control*. 2021;31(3):971-987.
62. Wang LJ, Ji Y, Yang HL, 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*. 2020;30(9):3607-3623.
63. 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.
64. Lv LL, Chen JB, Zhang Z, et al. A numerical solution of a class of periodic coupled matrix equations. *Journal of the Franklin Institute*. 2021;358(3):2039-2059.
65. Zhang L, Tang SY, Lv LL. An finite iterative algorithm for sloving periodic Sylvester bimatrix equations. *Journal of the Franklin Institute*. 2020;357(15):10757-10772.
66. Zhang L, Xu CB, Gao YH, Han Y, Du XJ, Tian ZH. Improved Dota2 lineup recommendation model based on a bidirectional LSTM. *Tsinghua Science and Technology*. 2020l;25(6):712-720.
67. Dong H, Yin CC, Dai HS. Spectrally negative Levy risk model under Erlangized barrier strategy. *Journal of Computational and Applied Mathematics*. 2019;351:101-116.
68. Sha XY, Xu ZS, Yin CC. Elliptical distribution-based weight-determining method for ordered weighted averaging operators. *International Journal of Intelligent Systems*. 2019;34(5):858-877.
69. Yin CC, Wen YZ. An extension of Paulsen-Gjessing's risk model with stochastic return on investments. *Insurance Mathematics & Economics*. 2013;52(3):469-476.
70. Zhao YX, Chen P, Yang HL. Optimal periodic dividend and capital injection problem for spectrally positive Levy processes. *Insurance Mathematics & Economics*. 2017;74:135-146.
71. Zhao XH, Dong H, Dai HS. On spectrally positive Levy risk processes with Parisian implementation delays in dividend payments. *Statistics & Probability Letters*. 2018;140:176-184.

DOI: 10.1002/acs

72. Zhao YX, Yin CC. The expected discounted penalty function under a renewal risk model with stochastic income. *Applied Mathematics and Computation*. 2012;218(10):6144-6154.
73. Zhang X, et al. Adaptive parameter estimation for a general dynamical system with unknown states. *International Journal of Robust and Nonlinear Control*. 2020;30(4):1351-1372.
74. Zhang X, et al, Xu L. Recursive parameter estimation methods and convergence analysis for a special class of nonlinear systems. *International Journal of Robust and Nonlinear Control*. 2020;30(4):1373-1393.
75. Zhang X, et al. Recursive parameter estimation and its convergence for bilinear systems. *IET Control Theory and Applications*. 2020;14(5):677-688.
76. Zhang X, et al. Hierarchical parameter and state estimation for bilinear systems. *International Journal of Systems Science*. 2020;51(2):275-290.
77. Zhang X, et al. Recursive identification of bilinear time-delay systems through the redundant rule. *Journal of the Franklin Institute*. 2020;357(1):726-747.
78. Xu L, et al. Hierarchical multi-innovation generalised extended stochastic gradient methods for multivariable equation-error autoregressive moving average systems. *IET Control Theory and Applications*. 2020;14(10):1276-1286.
79. Xu L, et al. Separable multi-innovation stochastic gradient estimation algorithm for the nonlinear dynamic responses of systems. *International Journal of Adaptive Control and Signal Processing*. 2020;34(7):937-954.
80. Xu L, et al. Separable recursive gradient algorithm for dynamical systems based on the impulse response signals. *International Journal of Control Automation and Systems*. 2020;18(12):3167-3177.
81. Xu L, et al. Auxiliary model multiinnovation stochastic gradient parameter estimation methods for nonlinear sandwich systems. *International Journal of Robust and Nonlinear Control*. 2021;31(1):148-165.
82. Z.P. Zhou, Y.H. Tan, X.F. Liu, Zhou ZP, Tan YH, Liu XF. State estimation of dynamic systems with sandwich structure and hysteresis. *Mechanical Systems and Signal Processing*. 2019;126:82-97.
83. Zhou ZP, Liu XF. State and fault estimation of sandwich systems with hysteresis. *International Journal of Robust and Nonlinear Control*. 2018;28(13):3974-3986.
84. Zhou ZP, Tan YH, Xie YQ, Dong RL. State estimation of a compound non-smooth sandwich system with backlash and dead zone. Mechanical Systems and Signal Processing. 2017:83:439-449.
85. Zhou ZP, Tan YH, Shi, P. Fault detection of a sandwich system with dead-zone based on robust observer. *Systems & Control Letters*. 2016;96:132-140.
86. Zhou ZP, Tan YH, Xie YQ, Dong RL. Soft measurement of states of sandwich system with dead zone and its application. *Measurement*. 2016;78:219-234.
87. Zhang B, Billings SA. Identification of continuous-time nonlinear systems: The nonlinear difference equation with moving average noise (NDEMA) framework. *Mechanical Systems and Signal Processing*. 2015;60-61:810-835.
88. Chen DC, Zhang XX, Tang J, 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*. 2019;363:346-357.
89. Cui H, Chen DC, Zhang Y, et al. Dissolved gas analysis in transformer oil using Pd catalyst decorated MoSe2 monolayer: A first-principles theory. *Sustainable Materials and Technologies*. 2019;20:e00094.
90. Cui H, Liu T, Zhang Y, 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*. 2019;19(13):5249-5255.
91. Cui H, Zhang XX, Chen DC, et al. Adsorption mechanism of SF6 decomposed species on pyridine-like PtN3 embedded CNT: A DFT study. *Applied Surface Science*. 2018;447:594-598.
92. Cui H, Zhang XX, Li Y, et al. First-principles insight into Ni-doped InN monolayer as a noxious gases scavenger. *Applied Surface Science*. 2019;494:859-866.
93. Cui H, Zhang XX, Zhang GZ, et al. Pd-doped MoS2 monolayer: A promising candidate for DGA in transformer oil based on DFT method. *Applied Surface Science*. 2019;470:1035-1042.
94. Cui H, Zhang GZ, Zhang XX, et al. Rh-doped MoSe2 as a toxic gas scavenger: a first-principles study. *Nanoscale Advances*. 2019;1(2):772-780.
95. Zhang XX, Gui YG, Xiao HY, et al. Analysis of adsorption properties of typical partial discharge gases on Ni-SWCNTs using density functional theory. *Applied Surface Science*. 2016;379:47-54.
96. Zhang XX, Yu L, Gui YG, et al. First-principles study of SF6 decomposed gas adsorbed on Au-decorated graphene. *Applied Surface Science*. 2016;367:259-269.
97. Zhang XX, Yu L, Wu XQ, et al. Experimental sensing and density functional theory study of H2S and SOF2 adsorption on Au-modified graphene. *Advanced Science*. 2015;2(11), Article Number:1500101.
98. Zhang XX, Zhang JB, Jia YC, et al. TiO2 nanotube array sensor for detecting the SF6 decomposition product SO2. *Sensors*. 2012;12(3):3302-3313.
99. Zhao XL, Lin ZY, Fu B, Gong SL. Research on frequency control method for micro-grid with a hybrid approach of FFR-OPPT and pitch angle of wind turbine. *International Journal of Electrical Power & Energy Systems*. 2021;127, Article Number: 106670.
100. Tian SS, Zhang XX, Xiao S, et al. Application of C6F12O/CO2 mixture in 10 kV medium-voltage switchgear. *IET Science Measurement and Technology*. 2019;13(9):1225-1230.
101. Zhang XX, Zhang Y, Huang Y, Li Y, Cheng HT, Xiao S. Detection of decomposition products of C4F7N-CO2 gas mixture based on infrared spectroscopy. *Vibrational Spectroscopy*. 2020;110, Article Number: 103114.
102. Jin XB, Yu XH, Su TL, Yang DN, Bai YT, Kong JL, Wang L. Distributed deep fusion predictor for a multi-sensor system based on causality. *Entropy*. 2021;23(2):219.
103. Jin XB, RobertJeremiah RJ, Su TL, et al. The new trend of state estimation: from model-driven to hybrid-driven methods. *Sensors*. 2021;21(6):2085.
104. Cao Y, Wen JK, Ma LC. Tracking and collision avoidance of virtual coupling train control system. *Alexandria Engineering Journal*. 2021;60(2):2115-2125.
105. Cao Y, Wen JK, Ma LC. Tracking and collision avoidance of virtual coupling train control system. *Future Generation Computer Systems*. 2021;120:76-90.

106. Su S, Wang XK, Cao Y, Yin JT. An energy-efficient train operation approach by integrating the metro timetabling and eco-driving. *IEEE Transactions on Intelligent Transportation Systems*. 2020;21(10):4252-4268.
107. Su S, Tang T, Xun J, Cao F, Wang YH. 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.
108. Cao Y, Wang Z. Liu F, et al. Bio-inspired speed curve optimization and sliding mode tracking control for subway trains. *IEEE Transactions on Vehicular Technology*. 2019;68(7):6331-6342.
109. Cao Y, Sun YK, Xie G, et al. 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*. 2019;68(8):7544-7551.
110. Lin J, Li Y, Yang GC. FPGAN: Face de-identification method with generative adversarial networks for social robots. *Neural Networks*. 2021;133:132-147.
111. Yang GC, Chen ZJ, Li Y, Su ZD. Rapid relocation method for mobile robot based on improved ORB-SLAM2 algorithm.*Remote Sensing*. 2019;11(2), Article No. 149.

DOI: 10.1002/acs