INTERNATIONAL JOURNAL OF ROBUST AND NONLINEAR CONTROL
Int. J. Robust Nonlinear Control 2022; **00**:1–21
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/acs

Recursive parameter estimation algorithms and convergence analysis for feedback nonlinear controlled autoregressive systems

Chun Wei[1], Xiao Zhang[1], Ling Xu[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 G1 1XJ, Scotland, UK*

SUMMARY

This paper deals with the problems of the parameter estimation for feedback nonlinear controlled autoregressive systems (i.e., feedback nonlinear equation-error systems). The bilinear-in-parameter identification model is formulated to describe the feedback nonlinear system. An overall recursive least squares algorithm is developed to handle the difficulty of the bilinear-in-parameter. For the purpose of avoiding the heavy computational burden, an overall stochastic gradient algorithm is deduced and the forgetting factor is introduced to improve the convergence rate. Furthermore, the convergence analysis of the proposed algorithms are established by means of the stochastic process theory. The effectiveness of the proposed algorithms are illustrated by the simulation example. Copyright © 2022 John Wiley & Sons, Ltd.

Received ...

KEY WORDS: Feedback nonlinear system, Bilinear-in-parameter model, Least squares, Gradient search, Convergence analysis

1. INTRODUCTION

System identification is an important tool to construct the mathematical models from the observed data [1–5]. The accurate mathematical models are the basis of the implementation of the control strategies [6–9]. In the area of system identification, much attention has been concentrated on linear systems, nonlinear systems, bilinear systems and so on. Due to the simple system structure, the parameter estimation methods have been matured for linear systems, such as the recursive identification [10], the iterative estimation [11, 12], the subspace identification [13] and the maximum likelihood identification [14–16]. However, nonlinear systems have the features of diverse structures, random interferences and great uncertainties. These factors make their identification problems more complex and significant. Recently, Zhang et al. investigated the iterative identification approach for the nonlinear time-delay systems by utilizing the observation data [17]. Liu et al. solved the parameter estimation problem by the expectation maximization algorithm for the nonlinear systems from the noisy output data [18].

Nonlinearity and feedback are the important characteristics of controlled systems. As a fundamental class of nonlinear systems, the feedback nonlinear systems consist of two parts, including the dynamic linear model in the forward channel and the static nonlinear model in the feedback channel. The dynamic linear models generally contain the finite impulse response

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

models, the equation-error models [19] and the output-error models [20]. Then, the nonlinearities in the feedback channel can be expressed as the function of the known nonlinear basis and the piecewise function with the saturation nonlinearities, the dead-zones nonlinearities and the hysteresis nonlinearities. On the identification of the feedback nonlinear output-error systems, an auxiliary model-based (decomposition) least squares identification approach was presented under the framework of the recursive least squares [21]. Taking advantage of the least squares principle and the harmonic balance method, the parameter estimation for the nonlinear systems with delayed feedback control was studied [22]. With the aim of improving the identification efficiency, the filtering-based hierarchical multi-innovation stochastic gradient algorithm was proposed for the feedback nonlinear systems with colored noises [23].

Over the years, the recursive identification has shown the great effectiveness in identifying various systems. The key idea of the recursive identification is to update the parameter estimates at the current time by adding the correction to the parameter estimates at the previous time. This method has been extensively applied in the modern industrial production because it can recursively compute the parameter estimates real-timely. In recent years, a number of recursive identification methods, particularly gradient-based [24–27] and least squares-based algorithms have been proposed for all kinds of systems, including linear systems and nonlinear systems. For example, Dong et al. developed the recursive least squares algorithm for the Wiener nonlinear systems based on the auxiliary model idea [28]. Du et al. investigated the online recursive identification approach to identifying the continuous-time switched nonlinear state-space models and proposed the subspace-based method to estimate the switching signal [29]. Zhang et al. derived the recursive projection algorithm to identify the finite impulse response systems with binary-valued observations [30]. The modified multi-innovation stochastic gradient algorithm was described to estimate the parameters of the Wiener-Hammerstein systems with unknown orders and backlash [31]. Moreover, some other parameter estimation technologies, such as the hierarchical identification principle [32, 33] and the data filtering technique [34], can also be combined to study new recursive parameter estimation algorithms. In practical applications, the recursive identification algorithms have also been researched for various systems, such as the wind tunnel systems [35], the alternating current motor drive systems and the air vehicle systems.

In this paper, we aim to develop the novel recursive algorithms for estimating the unknown parameters of the feedback nonlinear systems. The basic idea is to introduce the overall identification idea to handle the bilinear-in-parameter model and estimate the parameters. In addition, the convergence analysis is taken into consideration by means of the stochastic process theory. The main contributions are as follows.

- Inspired by the difficulty of the bilinear-in-parameter term, an overall recursive least squares (O-RLS) algorithm is proposed to identify the feedback nonlinear controlled autoregressive (FN-CAR) systems.
- Considering the computational efficiency, an overall stochastic gradient (O-SG) algorithm is deduced. Moreover, in order to speed up the convergence rate of the proposed algorithm, an overall forgetting factor stochastic gradient (O-FFSG) algorithm is derived through introducing the forgetting factor.
- The convergence analysis of the proposed O-RLS and O-SG algorithms are presented based on the stochastic process theory.

In brief, the outlines of this paper are organized as follows. Section 2 describes the FN-CAR model and its identification problem. Section 3 derives the O-RLS algorithm for the FN-CAR system by means of the least squares principle. Section 4 develops the O-SG algorithm for the FN-CAR system for the purpose of reducing the computational burden. The convergence analysis of the O-RLS and O-SG algorithms are presented in Section 5. The simulation example is provided to illustrate the effectiveness of the proposed algorithms in Section 6. Finally, Section 7 gives some concluding remarks.

Let us introduce some symbols for convenience. "$X := A$" or "$A := X$" stands for "A is defined as X"; the norm of the matrix is defined by $\|\boldsymbol{X}\|^2 := \mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$; the superscript T represents the matrix transpose; the determinant of the matrix is defined by $\det[\boldsymbol{X}] := |\boldsymbol{X}|$; $\boldsymbol{I}_n$ symbolizes an

DOI: 10.1002/acs

identity matrix of size $n \times n$ whose diagonal elements are 1 and the rest elements are 0; $\mathbf{1}_n$ denotes a column vector of dimension n whose entries are 1.

2. SYSTEM DESCRIPTION

Consider the feedback nonlinear controlled autoregressive system (i.e., feedback nonlinear equation-error system) whose forward channel of the system is a controlled autoregressive (CAR) model:

$$y(k)+\sum_{i=1}^{n_a} a_i y(k-i)=\sum_{i=1}^{n_b} b_i h(k-i)+v(k), \tag{1}$$

where $y(k)\in\mathbb{R}$ stands for the output of the system, $v(k)\in\mathbb{R}$ represents the random white noise with zero mean and variance σ^2, and $h(k)\in\mathbb{R}$ is the input of the CAR model. Assume that the orders are known, and $y(k)=0$ and $v(k)=0$ for $k\leqslant 0$.

The feedback channel of the system is the memoryless nonlinear block, of which the output $\bar{y}(k)$ can be regarded as a linear function of a known nonlinear basis $\boldsymbol{g}=(g_1,g_2,\cdots,g_m)$:

$$\begin{aligned}\bar{y}(k)&=g(y(k))=c_1g_1(y(k))+c_2g_2(y(k))+\cdots+c_mg_m(y(k))\\&=\sum_{j=1}^{m}c_jg_j(y(k))\\&=\boldsymbol{g}(y(k))\boldsymbol{c},\end{aligned} \tag{2}$$

where $\boldsymbol{c}:=[c_1,c_2,\cdots,c_m]^{\mathrm{T}}\in\mathbb{R}^m$ is the parameter vector of the nonlinear block, $\boldsymbol{g}(y(k)):=[g_1(y(k)),g_2(y(k)),\cdots,g_m(y(k))]\in\mathbb{R}^{1\times m}$ is the row vector of the basis function. From the above definitions, it is obvious that the input $h(k)$ of the CAR model and the output $\bar{y}(k)$ of the nonlinear block have the following relation:

$$h(k)=u(k)-\bar{y}(k), \tag{3}$$

where $u(k)\in\mathbb{R}$ is the input of the system and $u(k)=0$ for $k\leqslant 0$.

Recently, a recursive least squares and a hierarchical least squares identification algorithms have been proposed for feedback nonlinear equation-error systems in (1)–(3) [36]. This paper studies the convergence of the recursive least squares parameter estimation algorithm for feedback nonlinear equation-error systems.

Define the parameter vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ of the CAR model and $\boldsymbol{c}$ of the nonlinear block as

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

Define the output information vector $\boldsymbol{\varphi}_a(k)$, the input information vector $\boldsymbol{\varphi}_b(k)$ and the information matrix $\boldsymbol{G}(k)$ as

$$\boldsymbol{\varphi}_a(k):=[-y(k-1),-y(k-2),\cdots,-y(k-n_a)]^{\mathrm{T}}\in\mathbb{R}^{n_a}, \tag{4}$$

$$\boldsymbol{\varphi}_b(k):=[u(k-1),u(k-2),\cdots,u(k-n_b)]^{\mathrm{T}}\in\mathbb{R}^{n_b}, \tag{5}$$

$$\begin{aligned}\boldsymbol{G}(k)&:=\begin{bmatrix}-g_1(y(k-1)) & -g_2(y(k-1)) & \cdots & -g_m(y(k-1))\\ -g_1(y(k-2)) & -g_2(y(k-2)) & \cdots & -g_m(y(k-2))\\ \vdots & \vdots & & \vdots\\ -g_1(y(k-n_b)) & -g_2(y(k-n_b)) & \cdots & -g_m(y(k-n_b))\end{bmatrix}\\&=[-\boldsymbol{g}^{\mathrm{T}}(y(k-1)),-\boldsymbol{g}^{\mathrm{T}}(y(k-2)),\cdots,-\boldsymbol{g}^{\mathrm{T}}(y(k-n_b))]^{\mathrm{T}}\in\mathbb{R}^{n_b\times m}.\end{aligned} \tag{6}$$

DOI: 10.1002/acs

Substituting (2)–(6) into (1) gives

$$\begin{aligned}
y(k) &= -\sum_{i=1}^{n_a} a_i y(k-i) + \sum_{i=1}^{n_b} b_i[u(k-i) - \bar{y}(k-i)] + v(k) \\
&= -\sum_{i=1}^{n_a} a_i y(k-i) + \sum_{i=1}^{n_b} b_i u(k-i) - \sum_{i=1}^{n_b} b_i \boldsymbol{g}(y(k-i))\boldsymbol{c} + v(k) \\
&= -\sum_{i=1}^{n_a} a_i y(k-i) + \sum_{i=1}^{n_b} b_i u(k-i) - \sum_{i=1}^{n_b} b_i \sum_{j=1}^{m} c_j g_j(y(k-i)) + v(k) \\
&= \boldsymbol{\varphi}_a^{\mathrm{T}}(k)\boldsymbol{a} + \boldsymbol{\varphi}_b^{\mathrm{T}}(k)\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{G}(k)\boldsymbol{c} + v(k). \qquad (7)
\end{aligned}$$

The proposed parameter estimation algorithms in this paper are based on this identification model. Many identification methods are derived based on the identification models of the systems [37–40] and these methods can be used to estimate the parameters of other linear systems and nonlinear systems [41–44] and can be applied to other fields [45–50] such as chemical process control systems. There exists the product of the parameter vector $\boldsymbol{b}$ of the CAR model in the forward channel and $\boldsymbol{c}$ of the nonlinear block in the feedback channel in (7) such that the identification model is a typical bilinear-in-parameter model. It is worth pointing out that the bilinear-in-parameter models are the extension to the linear-in-parameter models and the existence of the product term increases the difficulty of identification. This motivates us to investigate new estimation methods of the bilinear-in-parameter systems on the basis of linear systems.

3. THE OVERALL RECURSIVE LEAST SQUARES ALGORITHM

Considering separating the linear and nonlinear parameters in the identification model in (7), this section develops the overall recursive least squares (O-RLS) algorithm for the feedback nonlinear controlled autoregressive (FN-CAR) system. The details are as follows.

Define the overall information vector $\boldsymbol{\varphi}(\boldsymbol{b}, k)$ and the overall parameter vector $\boldsymbol{\vartheta}$ as

$$\boldsymbol{\varphi}(\boldsymbol{b}, k) := \begin{bmatrix} \boldsymbol{\varphi}_a(k) \\ \boldsymbol{\varphi}_b(k) \\ \boldsymbol{G}^{\mathrm{T}}(k)\boldsymbol{b} \end{bmatrix} \in \mathbb{R}^{n_a+n_b+m}, \quad \boldsymbol{\vartheta} := \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \\ \boldsymbol{c} \end{bmatrix} \in \mathbb{R}^{n_a+n_b+m}.$$

Then, the identification model for the FN-CAR system in (7) can be written as

$$\begin{aligned}
y(k) &= \boldsymbol{\varphi}_a^{\mathrm{T}}(k)\boldsymbol{a} + \boldsymbol{\varphi}_b^{\mathrm{T}}(k)\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{G}(k)\boldsymbol{c} + v(k) \\
&= [\boldsymbol{\varphi}_a^{\mathrm{T}}(k), \boldsymbol{\varphi}_b^{\mathrm{T}}(k), \boldsymbol{b}^{\mathrm{T}}\boldsymbol{G}(k)]\boldsymbol{\vartheta} + v(k) \\
&= \boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{b}, k)\boldsymbol{\vartheta} + v(k). \qquad (8)
\end{aligned}$$

Equation (8) can be viewed as a class of the pseudo-linear regression models, which contains $n := n_a + n_b + m$ parameters. In view of the fact that the overall information vector $\boldsymbol{\varphi}(\boldsymbol{b}, k)$ is unmeasurable due to the unknown parameter vector $\boldsymbol{b}$ in it, the least squares algorithm cannot directly work. The solution is to replace the unknown parameter vector $\boldsymbol{b}$ with its estimate so as to construct the estimate of the overall information vector $\boldsymbol{\varphi}(\boldsymbol{b}, k)$. Based on this idea, we define the criterion function:

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

Assume that $\hat{\boldsymbol{\vartheta}}(k) \in \mathbb{R}^n$ denotes the estimate of $\boldsymbol{\vartheta}$ at time k. Minimizing the criterion function $J_1(\boldsymbol{\vartheta})$, the least squares method is given by

$$\hat{\boldsymbol{\vartheta}}(k) = \left[\sum_{j=1}^{k}\boldsymbol{\varphi}(\boldsymbol{b}, j)\boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{b}, j)\right]^{-1}\left[\sum_{j=1}^{k}\boldsymbol{\varphi}(\boldsymbol{b}, j)y(j)\right]. \qquad (9)$$

DOI: 10.1002/acs

It is worth noting that the offline estimation method (9) may cause the heavy computational burden. In order to handle the problem, the covariance matrix $\boldsymbol{P}(k) \in \mathbb{R}^{n\times n}$ is taken into consideration,

$$\boldsymbol{P}^{-1}(k) := \sum_{j=1}^{k} \boldsymbol{\varphi}(\boldsymbol{b}, j)\boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{b}, j).$$

Therefore, we can derive the recursive relation for online identification from (9):

$$\hat{\boldsymbol{\vartheta}}(k) = \hat{\boldsymbol{\vartheta}}(k-1) + \boldsymbol{P}(k)\boldsymbol{\varphi}(\boldsymbol{b}, k)[y(k) - \boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{b}, k)\hat{\boldsymbol{\vartheta}}(k-1)], \tag{10}$$

$$\boldsymbol{P}^{-1}(k) = \boldsymbol{P}^{-1}(k-1) + \boldsymbol{\varphi}(\boldsymbol{b}, k)\boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{b}, k). \tag{11}$$

Obviously, the parameter vector $\boldsymbol{b}$ in the information vector $\boldsymbol{\varphi}(\boldsymbol{b}, k)$ in (10)–(11) is unknown. Faced with the situation, we utilize the estimate $\hat{\boldsymbol{b}}(k-1)$ of $\boldsymbol{b}$ to take the place of $\boldsymbol{b}$ in $\boldsymbol{\varphi}(\boldsymbol{b}, k)$. Then, the estimate $\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1), k)$ of $\boldsymbol{\varphi}(\boldsymbol{b}, k)$ is constructed as

$$\hat{\boldsymbol{\varphi}}_1(k) := \boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1), k) = [\boldsymbol{\varphi}_a^{\mathrm{T}}(k), \boldsymbol{\varphi}_b^{\mathrm{T}}(k), \hat{\boldsymbol{b}}^{\mathrm{T}}(k-1)\boldsymbol{G}(k)]^{\mathrm{T}}. \tag{12}$$

Replacing $\boldsymbol{\varphi}(\boldsymbol{b}, k)$ in (10)–(11) with its estimate $\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1), k)$ gives

$$\hat{\boldsymbol{\vartheta}}(k) = \hat{\boldsymbol{\vartheta}}(k-1) + \boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1), k)[y(k) - \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1), k)\hat{\boldsymbol{\vartheta}}(k-1)], \tag{13}$$

$$\boldsymbol{P}^{-1}(k) = \boldsymbol{P}^{-1}(k-1) + \boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1), k)\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1), k). \tag{14}$$

Using (4)–(6) and (12)–(14), we obtain the O-RLS algorithm, which is often used for the convergence analysis in this form. For the sake of avoiding the inverse of the covariance matrix and improving the computational efficiency, we introduce the gain vector $\boldsymbol{L}(k) := \boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1), k)$ and apply the matrix inversion lemma

$$(\boldsymbol{A} + \boldsymbol{BC})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{I} + \boldsymbol{CA}^{-1}\boldsymbol{B})^{-1}\boldsymbol{CA}^{-1}$$

to (14). Furthermore, combining (4)–(6), we have the overall recursive least squares (O-RLS) algorithm for estimating the parameter vector $\boldsymbol{\vartheta}$ of the FN-CAR system [36]:

$$\hat{\boldsymbol{\vartheta}}(k) = \hat{\boldsymbol{\vartheta}}(k-1) + \boldsymbol{L}(k)[y(k) - \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1), k)\hat{\boldsymbol{\vartheta}}(k-1)], \tag{15}$$

$$\boldsymbol{L}(k) = \boldsymbol{P}(k-1)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1), k)[1 + \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1), k)\boldsymbol{P}(k-1)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1), k)]^{-1}, \tag{16}$$

$$\boldsymbol{P}(k) = [\boldsymbol{I}_n - \boldsymbol{L}(k)\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1), k)]\boldsymbol{P}(k-1), \tag{17}$$

$$\hat{\boldsymbol{\varphi}}_1(k) = \boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1), k) = [\boldsymbol{\varphi}_a^{\mathrm{T}}(k), \boldsymbol{\varphi}_b^{\mathrm{T}}(k), \hat{\boldsymbol{b}}^{\mathrm{T}}(k-1)\boldsymbol{G}(k)]^{\mathrm{T}}, \tag{18}$$

$$\boldsymbol{\varphi}_a(k) = [-y(k-1), -y(k-2), \cdots, -y(k-n_a)]^{\mathrm{T}}, \tag{19}$$

$$\boldsymbol{\varphi}_b(k) = [u(k-1), u(k-2), \cdots, u(k-n_b)]^{\mathrm{T}}, \tag{20}$$

$$\boldsymbol{G}(k) = [-\boldsymbol{g}^{\mathrm{T}}(y(k-1)), -\boldsymbol{g}^{\mathrm{T}}(y(k-2)), \cdots, -\boldsymbol{g}^{\mathrm{T}}(y(k-n_b))]^{\mathrm{T}}, \tag{21}$$

$$\boldsymbol{g}(y(k)) = [g_1(y(k)), g_2(y(k)), \cdots, g_m(y(k))], \tag{22}$$

$$\hat{\boldsymbol{\vartheta}}(k) = [\hat{\boldsymbol{a}}^{\mathrm{T}}(k), \hat{\boldsymbol{b}}^{\mathrm{T}}(k), \hat{\boldsymbol{c}}^{\mathrm{T}}(k)]^{\mathrm{T}}. \tag{23}$$

Equations (15)–(23) form the O-RLS algorithm, which is generally used for online identification or real-time estimation. Taking the overall identification into consideration, the bilinear-in-parameter model is limited to the pseudo-linear regression model. The O-RLS algorithm is derived based on the linear least squares optimization. The O-RLS algorithm has a good performance in the parameter estimation accuracy because it requires to use new observed data at each recursion instead of the batch data. The steps of implementing the O-RLS algorithm in (15)–(23) are summed up as follows.

1. Initialization: set $k = 1$, and let the initial values $\hat{\boldsymbol{\vartheta}}(0) = [\hat{\boldsymbol{a}}^{\mathrm{T}}(0), \hat{\boldsymbol{b}}^{\mathrm{T}}(0), \hat{\boldsymbol{c}}^{\mathrm{T}}(0)]^{\mathrm{T}} = \mathbf{1}_n/p_0$, $\boldsymbol{P}(0) = p_0\boldsymbol{I}_n$, $y(k-j) = 0$, $u(k-j) = 0$, $j = 1, 2, \cdots, \max[n_a, n_b, m]$, and $p_0 = 10^6$. Give the basis function $g_i(\cdot)$ $(i = 1, 2, \cdots, m)$ and the data length $L_{\max}$.
2. Collect the measurement input and output data $u(k)$ and $y(k)$. Construct the output information vector $\boldsymbol{\varphi}_a(k)$, the input information vector $\boldsymbol{\varphi}_b(k)$ and the information matrix $\boldsymbol{G}(k)$ by (19)–(22).

DOI: 10.1002/acs

3. Form the information vector $\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)$ by (18). Compute the gain vector $\boldsymbol{L}(k)$ and the covariance matrix $\boldsymbol{P}(k)$ by (16)–(17).
4. Update the parameter estimation vector $\hat{\boldsymbol{\vartheta}}(k)$ using (15). Read out the parameter estimates $\hat{\boldsymbol{a}}(k)$, $\hat{\boldsymbol{b}}(k)$ and $\hat{\boldsymbol{c}}(k)$ from $\hat{\boldsymbol{\vartheta}}(k)$ in (23).
5. If $k < L_{\max}$, increase k by 1 and go to Step 2; otherwise, output the parameter estimates $\hat{\boldsymbol{a}}(k)$, $\hat{\boldsymbol{b}}(k)$ and $\hat{\boldsymbol{c}}(k)$, and terminate the recursive process.

4. THE OVERALL STOCHASTIC GRADIENT ALGORITHM

Involving the covariance matrix, the proposed O-RLS algorithm does not perform efficiently in the calculating quantities. Motivated by the negative gradient search, the overall stochastic gradient (O-SG) algorithm for the FN-CAR system is proposed in this section to improve the computational efficiency. The details are as follows.

According to the identification model in (8), define the gradient criterion function:

$$J_2(\boldsymbol{\vartheta}) := \frac{1}{2}[y(k)-\boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{b},k)\boldsymbol{\vartheta}]^2.$$

Let $\hat{\boldsymbol{\vartheta}}(k)\in\mathbb{R}^n$ be the estimate of $\boldsymbol{\vartheta}$ at time k and $\mu(k)$ be the convergence factor. By using the negative gradient search and minimizing the gradient criterion function $J_2(\boldsymbol{\vartheta})$, we can obtain the overall stochastic gradient (O-SG) algorithm of the FN-CAR system for estimating the parameter vector $\boldsymbol{\vartheta}$ [19]:

$$\hat{\boldsymbol{\vartheta}}(k)=\hat{\boldsymbol{\vartheta}}(k-1)+\frac{\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)}{r(k)}[y(k)-\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\hat{\boldsymbol{\vartheta}}(k-1)], \tag{24}$$

$$r(k)=r(k-1)+\|\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)\|^2, \tag{25}$$

$$\hat{\boldsymbol{\varphi}}_2(k)=\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)=[\boldsymbol{\varphi}_a^{\mathrm{T}}(k),\boldsymbol{\varphi}_b^{\mathrm{T}}(k),\hat{\boldsymbol{b}}^{\mathrm{T}}(k-1)\boldsymbol{G}(k)]^{\mathrm{T}}, \tag{26}$$

$$\boldsymbol{\varphi}_a(k)=[-y(k-1),-y(k-2),\cdots,-y(k-n_a)]^{\mathrm{T}}, \tag{27}$$

$$\boldsymbol{\varphi}_b(k)=[u(k-1),u(k-2),\cdots,u(k-n_b)]^{\mathrm{T}}, \tag{28}$$

$$\boldsymbol{G}(k)=[-\boldsymbol{g}^{\mathrm{T}}(y(k-1)),-\boldsymbol{g}^{\mathrm{T}}(y(k-2)),\cdots,-\boldsymbol{g}^{\mathrm{T}}(y(k-n_b))]^{\mathrm{T}}, \tag{29}$$

$$\boldsymbol{g}(y(k))=[g_1(y(k)),g_2(y(k)),\cdots,g_m(y(k))], \tag{30}$$

$$\hat{\boldsymbol{\vartheta}}(k)=[\hat{\boldsymbol{a}}^{\mathrm{T}}(k),\hat{\boldsymbol{b}}^{\mathrm{T}}(k),\hat{\boldsymbol{c}}^{\mathrm{T}}(k)]^{\mathrm{T}}. \tag{31}$$

In general, the parameter estimation errors of the stochastic gradient algorithm can converge to zero at a slow convergence rate. In order to improve the convergence rate of the O-SG algorithm, the forgetting factor λ is introduced to (25) and the relation is given by

$$r(k)=\lambda r(k-1)+\|\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)\|^2,\ 0<\lambda\leqslant 1. \tag{32}$$

Equations (24) and (26)–(32) form the overall forgetting factor stochastic gradient (O-FFSG) algorithm. When $\lambda=1$ in (32), the O-FFSG algorithm reduces to the O-SG algorithm. The O-SG and O-FFSG algorithms are developed through the negative gradient search and dynamical observations in order to reduce the computational burden. It is stressed that the introduction of forgetting factor can improve the transient performance of the proposed algorithm and obtain the high parameter estimation accuracy. The procedure of implementing the O-SG algorithm in (24)–(31) are listed as follows.

1. Initialization: set $k=1$, and let the initial values $\hat{\boldsymbol{\vartheta}}(0)=[\hat{\boldsymbol{a}}^{\mathrm{T}}(0),\hat{\boldsymbol{b}}^{\mathrm{T}}(0),\hat{\boldsymbol{c}}^{\mathrm{T}}(0)]^{\mathrm{T}}=\mathbf{1}_n/p_0$, $p_0=10^6$, $r(0)=1$, $y(k-i)=0$, and $u(k-i)=0$, $i=1,2,\cdots,\max[n_a,n_b,m]$. Give the basis function $g_i(\cdot)$ $(i=1,2,\cdots,m)$ and the data length $L_{\max}$.
2. Collect the measurement input and output data $u(k)$ and $y(k)$. Form the output information vector $\boldsymbol{\varphi}_a(k)$ by (27), the input information vector $\boldsymbol{\varphi}_b(k)$ by (28) and the information matrix $\boldsymbol{G}(k)$ by (29)–(30).

DOI: 10.1002/acs

3. Construct the information vector $\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)$ by (26) and compute $r(k)$ by (25).
4. Update the parameter estimation vector $\hat{\boldsymbol{\vartheta}}(k)$ according to (24). Read out the parameter estimates $\hat{\boldsymbol{a}}(k)$, $\hat{\boldsymbol{b}}(k)$ and $\hat{\boldsymbol{c}}(k)$ from $\hat{\boldsymbol{\vartheta}}(k)$ in (31).
5. If $k < L_{\max}$, increase k by 1 and go to Step 2; otherwise, obtain the parameter estimates $\hat{\boldsymbol{a}}(k)$, $\hat{\boldsymbol{b}}(k)$ and $\hat{\boldsymbol{c}}(k)$, and terminate the recursive process.

Remark 1 Moreover, the proposed algorithms in this article can combine other parameter estimation technologies [51–54] to investigate the identification of other bilinear systems and nonlinear systems [55–58] with white noises or colored noises and can be applied to other fields [59–70] such as signal processing, engineering application and so on.

Remark 2 Tables I and II demonstrate the computational efficiency of the O-RLS and O-SG algorithm. The numbers of multiplications and additions at each recursion and the total flops are listed to evaluate the computational costs. Comparing the total flops, it is clear that the O-RLS algorithm requires more computational cost than the O-SG algorithm, which verifies the high computational efficiency of the O-SG algorithm.

Table I. The computational efficiency of the O-RLS algorithm

Expressions	Multiplications	Additions
$\hat{\boldsymbol{\vartheta}}(k)=\hat{\boldsymbol{\vartheta}}(k-1)+\boldsymbol{L}(k)e(k)$	n	n
$e(k):=y(k)-\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\hat{\boldsymbol{\vartheta}}(k-1)$	n	n
$\boldsymbol{\zeta}(k):=\boldsymbol{P}(k-1)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)$	n^2	n^2-n
$\boldsymbol{L}(k)=\boldsymbol{\zeta}(k)/[1+\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{\zeta}(k)]$	$2n$	n
$\boldsymbol{P}(k)=\boldsymbol{P}(k-1)-\boldsymbol{L}(k)\boldsymbol{\zeta}^{\mathrm{T}}(k)$	n^2	n^2
$\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)=[\boldsymbol{\varphi}_a^{\mathrm{T}}(k),\boldsymbol{\varphi}_b^{\mathrm{T}}(k),\hat{\boldsymbol{b}}^{\mathrm{T}}(k-1)\boldsymbol{G}(k)]^{\mathrm{T}}$	mn_b	$m(n_b-1)$
Sum	$2n^2+4n+mn_b$	$2n^2+2n+m(n_b-1)$
Total flops	$4n^2+6n+m(2n_b-1)$	

Table II. The computational efficiency of the O-SG algorithm

Expressions	Multiplications	Additions
$\hat{\boldsymbol{\vartheta}}(k)=\hat{\boldsymbol{\vartheta}}(k-1)+\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)[e(k)/r(k)]$	$n+1$	n
$e(k):=y(k)-\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\hat{\boldsymbol{\vartheta}}(k-1)$	n	n
$r(k)=r(k-1)+\Vert\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)\Vert^2$	n	n
$\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)=[\boldsymbol{\varphi}_a^{\mathrm{T}}(k),\boldsymbol{\varphi}_b^{\mathrm{T}}(k),\hat{\boldsymbol{b}}^{\mathrm{T}}(k-1)\boldsymbol{G}(k)]^{\mathrm{T}}$	mn_b	$m(n_b-1)$
Sum	$3n+1+mn_b$	$3n+m(n_b-1)$
Total flops	$6n+1+m(2n_b-1)$	

5. THE CONVERGENCE ANALYSIS FOR THE O-RLS AND O-SG ALGORITHMS

In this section, combined with the stochastic process theory and the persistent excitation conditions, the convergence analysis of the O-RLS and O-SG algorithms for the FN-CAR system are proved.

Lemma 1 Assume that the non-negative random sequences $\{T(k)\}$, $\{\beta(k)\}$ and $\{\gamma(k)\}$ satisfy the inequality:

$$T(k+1)\leqslant T(k)-\gamma(k)+\beta(k),$$

and $\sum_{k=1}^{\infty}\beta(k)<\infty$, then we have $\sum_{k=1}^{\infty}\gamma(k)<\infty$, and $T(k)$ converges to a finite random variable T_0.

DOI: 10.1002/acs

Lemma 2 For the FN-CAR system in (8) and the O-RLS algorithm in (15)–(23), the following covariance matrix inequality holds

$$\sum_{k=1}^{\infty}\frac{\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)}{[\ln|\boldsymbol{P}^{-1}(k)|]^{c}}<\infty,\quad c>1.$$

Proof According to the definition of $\boldsymbol{P}^{-1}(k)$ in (14), we have

$$\begin{aligned}\boldsymbol{P}^{-1}(k-1)&=\boldsymbol{P}^{-1}(k)-\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\\&=\boldsymbol{P}^{-1}(k)[\boldsymbol{I}_n-\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)].\end{aligned}$$

Taking the determinant of both sides and using the relation $\det[\boldsymbol{I}_n+\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}]=1+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{a}$ give

$$|\boldsymbol{P}^{-1}(k-1)|=|\boldsymbol{P}^{-1}(k)|[1-\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)].$$

Thus, we have

$$\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)=\frac{|\boldsymbol{P}^{-1}(k)|-|\boldsymbol{P}^{-1}(k-1)|}{|\boldsymbol{P}^{-1}(k)|}.$$

Here, suppose that $\ln|\boldsymbol{P}^{-1}(k)|>0$. Dividing both sides by $[\ln|\boldsymbol{P}^{-1}(k)|]^{c}$ for any $c>1$ and summing for k from 1 to ∞ obtain

$$\begin{aligned}\sum_{k=1}^{\infty}\frac{\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)}{[\ln|\boldsymbol{P}^{-1}(k)|]^{c}}&=\sum_{k=1}^{\infty}\frac{|\boldsymbol{P}^{-1}(k)|-|\boldsymbol{P}^{-1}(k-1)|}{|\boldsymbol{P}^{-1}(k)|[\ln|\boldsymbol{P}^{-1}(k)|]^{c}}\\&\leqslant\sum_{k=1}^{\infty}\int_{|\boldsymbol{P}^{-1}(k-1)|}^{|\boldsymbol{P}^{-1}(k)|}\frac{\mathrm{d}x}{|\boldsymbol{P}^{-1}(k)|[\ln|\boldsymbol{P}^{-1}(k)|]^{c}}\\&\leqslant\int_{|\boldsymbol{P}^{-1}(0)|}^{|\boldsymbol{P}^{-1}(\infty)|}\frac{\mathrm{d}x}{x[\ln x]^{c}}\\&=\frac{1}{c-1}\left\{\frac{1}{[\ln|\boldsymbol{P}^{-1}(0)|]^{c-1}}-\frac{1}{[\ln|\boldsymbol{P}^{-1}(\infty)|]^{c-1}}\right\}<\infty.\end{aligned}$$

Lemma 2 is proved.

Theorem 1 For the FN-CAR system in (8) and the O-RLS algorithm in (15)–(23), suppose that $\{v(k)\}$ is the white noise sequence with zero mean and variance σ^2, which satisfies

(C1) $\mathrm{E}[v(k)]=0$,
(C2) $\mathrm{E}[v^2(k)]=\sigma^2<\infty$,
(C3) $\mathrm{E}[v(k)v(j)]=0,\ \ j\neq k$.

There exist the positive constants α_1 and α_2 and an integer k_0 such that the following excitation condition holds when $k\geqslant k_0$:

$$\text{(C4)}\ \ \alpha_1\boldsymbol{I}_n\leqslant\frac{1}{k}\sum_{j=1}^{k}\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(j-1),j)\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(j-1),j)\leqslant\alpha_2\boldsymbol{I}_n,\ \text{a.s.}.$$

Then, the parameter estimation error for the O-RLS algorithm converges to zero.

Proof Define the parameter estimation error vector $\tilde{\boldsymbol{\vartheta}}(k)$ and the innovation $e(k)$ as

$$\begin{aligned}&\tilde{\boldsymbol{\vartheta}}(k):=\hat{\boldsymbol{\vartheta}}(k)-\boldsymbol{\vartheta},\\&e(k):=y(k)-\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\hat{\boldsymbol{\vartheta}}(k-1).\end{aligned}\tag{33}$$

Substituting (8) into (33) gives

$$e(k)=\boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{b},k)\boldsymbol{\vartheta}+v(k)-\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\hat{\boldsymbol{\vartheta}}(k-1)$$

DOI: 10.1002/acs

$$
\begin{aligned}
&= \boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{b},k)\boldsymbol{\vartheta} - \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)[\hat{\boldsymbol{\vartheta}}(k-1)-\boldsymbol{\vartheta}] - \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{\vartheta} + v(k)\\
&= -\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\tilde{\boldsymbol{\vartheta}}(k-1) + [\boldsymbol{\varphi}(\boldsymbol{b},k) - \boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)]^{\mathrm{T}}\boldsymbol{\vartheta} + v(k)\\
&=: -\tilde{y}(k) + \xi(k) + v(k),
\end{aligned} \tag{34}
$$

where

$$
\begin{aligned}
&\tilde{y}(k) := \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\tilde{\boldsymbol{\vartheta}}(k-1) \in \mathbb{R},\\
&\xi(k) := [\boldsymbol{\varphi}(\boldsymbol{b},k) - \boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)]^{\mathrm{T}}\boldsymbol{\vartheta} \in \mathbb{R}.
\end{aligned} \tag{35}
$$

Substituting (33) into (13) and subtracting $\boldsymbol{\vartheta}$ from both sides give

$$
\begin{aligned}
\tilde{\boldsymbol{\vartheta}}(k) &= \hat{\boldsymbol{\vartheta}}(k-1) + \boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)e(k) - \boldsymbol{\vartheta}\\
&= \tilde{\boldsymbol{\vartheta}}(k-1) + \boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)e(k).
\end{aligned} \tag{36}
$$

Define a non-negative function:

$$
T(k) := \tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}(k)\boldsymbol{P}^{-1}(k)\tilde{\boldsymbol{\vartheta}}(k).
$$

Using (36) and the relation $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}$, it follows that

$$
\begin{aligned}
T(k) &= [\tilde{\boldsymbol{\vartheta}}(k-1) + \boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)e(k)]^{\mathrm{T}}\boldsymbol{P}^{-1}(k)[\tilde{\boldsymbol{\vartheta}}(k-1) + \boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)e(k)]\\
&= \tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}(k-1)\boldsymbol{P}^{-1}(k)\tilde{\boldsymbol{\vartheta}}(k-1) + 2\tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}(k-1)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)e(k)\\
&\quad + \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)e^2(k).
\end{aligned} \tag{37}
$$

Substituting (14) and (34)–(35) into (37), we have

$$
\begin{aligned}
T(k) &= \tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}(k-1)[\boldsymbol{P}^{-1}(k-1) + \boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)]\tilde{\boldsymbol{\vartheta}}(k-1)\\
&\quad + 2\tilde{y}(k)e(k) + \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)e^2(k)\\
&= \tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}(k-1)\boldsymbol{P}^{-1}(k-1)\tilde{\boldsymbol{\vartheta}}(k-1) + \|\tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}(k-1)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)\|^2\\
&\quad + 2\tilde{y}(k)[-\tilde{y}(k) + \xi(k) + v(k)] + \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)\\
&\quad \times[-\tilde{y}(k) + \xi(k) + v(k)]^2\\
&= T(k-1) + \tilde{y}^2(k) - 2\tilde{y}^2(k) + 2\tilde{y}(k)[\xi(k) + v(k)] + \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)\\
&\quad \times\{\tilde{y}^2(k) + [\xi(k) + v(k)]^2 - 2\tilde{y}(k)[\xi(k) + v(k)]\}\\
&= T(k-1) - [1 - \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)]\tilde{y}^2(k)\\
&\quad + 2[1 - \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)]\tilde{y}(k)[\xi(k) + v(k)]\\
&\quad + \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)[\xi^2(k) + 2\xi(k)v(k) + v^2(k)].
\end{aligned} \tag{38}
$$

Note that

$$
\begin{aligned}
&1 - \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)\\
&= 1 - \frac{\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k-1)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)}{1 + \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k-1)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)}\\
&= \frac{1}{1 + \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k-1)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)} \geqslant 0.
\end{aligned}
$$

From (38), it is easy to get

$$
\begin{aligned}
T(k) &\leqslant T(k-1) + 2[1 - \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)]\tilde{y}(k)[\xi(k) + v(k)]\\
&\quad + \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)[\xi^2(k) + 2\xi(k)v(k) + v^2(k)].
\end{aligned} \tag{39}
$$

The noise sequence $\{v(k)\}$ is the white noise with zero mean and uncorrelated with the input signal $\{u(k)\}$. For given $\Delta_1 > 0$, if $\xi^2(k) \leqslant \Delta_1$, since $\tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}(k-1)\boldsymbol{P}^{-1}(k-1)\tilde{\boldsymbol{\vartheta}}(k-1)$, $\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)$, $\tilde{y}(k)$ and $\xi(k)$ are independent of $v(k)$, taking the expectation of both

DOI: 10.1002/acs

sides of (39) and using (C1)–(C3) give

$$\mathrm{E}[T(k)] \leqslant \mathrm{E}[T(k-1)] + \mathrm{E}[\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)](\sigma^2+\Delta_1). \tag{40}$$

Otherwise, let $\hat{\boldsymbol{\vartheta}}(k) := \hat{\boldsymbol{\vartheta}}(k-1)$ and $T(k) := T(k-1)$. Then, we introduce a new variable:

$$W(k) := \frac{T(k)}{[\ln|\boldsymbol{P}^{-1}(k)|]^c}.$$

Assume that $\ln|\boldsymbol{P}^{-1}(k)| > 0$ for $k \geqslant 1$. From (40), we have

$$\begin{aligned}\mathrm{E}[W(k)] &\leqslant \mathrm{E}\left[\frac{T(k-1)}{[\ln|\boldsymbol{P}^{-1}(k)|]^c}\right] + \mathrm{E}\left[\frac{\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)}{[\ln|\boldsymbol{P}^{-1}(k)|]^c}\right](\sigma^2+\Delta_1)\\ &\leqslant \mathrm{E}[W(k-1)] + \mathrm{E}\left[\frac{\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)}{[\ln|\boldsymbol{P}^{-1}(k)|]^c}\right](\sigma^2+\Delta_1). \end{aligned}\tag{41}$$

According to Lemma 2, the sum of the last term on the right-hand side of (41) for k from 1 to ∞ is bounded,

$$\sum_{k=1}^{\infty}\frac{\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{P}(k)\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)}{[\ln|\boldsymbol{P}^{-1}(k)|]^c}(\sigma^2+\Delta_1) < \infty.$$

Applying Lemma 1 to (41) gives that $W(k)$ converges to a finite random variable W_0 and

$$\begin{aligned} &W(k) \to W_0 < \infty,\\ &T(k) = O\left([\ln|\boldsymbol{P}^{-1}(k)|]^c\right). \end{aligned}$$

Since $\lambda_{\min}[\boldsymbol{X}]\|\boldsymbol{\Theta}\|^2 \leqslant \boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{\Theta} \leqslant \lambda_{\max}[\boldsymbol{X}]\|\boldsymbol{\Theta}\|^2$, where $\lambda_{\min}[\boldsymbol{X}]$ and $\lambda_{\max}[\boldsymbol{X}]$ represent the minimum eigenvalue and the maximum eigenvalue of a symmetric matrix, respectively, it follows that

$$\|\tilde{\boldsymbol{\vartheta}}(k)\|^2 \leqslant \frac{T(k)}{\lambda_{\min}[\boldsymbol{P}^{-1}(k)]} = O\left(\frac{[\ln|\boldsymbol{P}^{-1}(k)|]^c}{\lambda_{\min}[\boldsymbol{P}^{-1}(k)]}\right). \tag{42}$$

Using (14) and the persistent excitation condition (C4), we can obtain

$$\begin{aligned}\boldsymbol{P}^{-1}(k) &= \boldsymbol{P}^{-1}(k-1) + \boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k-1),k)\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k-1),k)\\ &= \boldsymbol{P}^{-1}(0) + \sum_{j=1}^{k}\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(j-1),j)\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(j-1),j)\\ &\leqslant \boldsymbol{I}_n/p_0 + \alpha_2 k\boldsymbol{I}_n = (\alpha_2 k + 1/p_0)\boldsymbol{I}_n. \end{aligned}$$

Taking the determinant and the logarithm of both sides gives $\ln|\boldsymbol{P}^{-1}(k)| \leqslant n\ln(\alpha_2 k + 1/p_0)$. Since $\lambda_{\min}[\boldsymbol{P}^{-1}(k)] \geqslant \alpha_1 k$, according to (42), it follows that

$$\begin{aligned}\|\tilde{\boldsymbol{\vartheta}}(k)\|^2 &= O\left(\frac{n^c[\ln(\alpha_2 k + 1/p_0)]^c}{\alpha_1 k}\right)\\ &= O\left(\frac{[\ln k]^c}{k}\right) = 0. \end{aligned}$$

The proof is completed.

Remark 3 Conditions (C1)–(C3) show that the noise has a zero mean value and bounded variance. Condition (C4) means that the system input is a persistent excitation signal. Theorem 1 indicates that under the assumption that Conditions (C1)–(C4) hold and the variable $\xi(k)$ are bounded, the parameter estimation error of the O-RLS algorithm converges to zero.

Theorem 2 For the FN-CAR system in (8) and the O-SG algorithm in (24)–(31), suppose that Conditions (C1)–(C3) are satisfied, and there exist an integer N and a positive constant α

DOI: 10.1002/acs

independent of k such that the following persistent excitation condition holds:

$$\text{(C5)} \quad \sum_{i=0}^{N-1} \frac{\boldsymbol{\varphi}(\hat{\boldsymbol{b}}(k+i-1),k+i)\boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(k+i-1),k+i)}{r(k+i)} \geqslant \alpha \boldsymbol{I}_n, \text{ a.s., } k>0.$$

Then, the parameter estimation error given by the O-SG algorithm converges to zero. The proof can be referred to the work in [19] and that is omitted here. The proposed approaches in the paper can combine other mathematical tools [71–74] and statistical strategies [75–80] to investigate the identification methods of nonlinear systems with different structures and disturbance noises and can be applied to literatures [81–90] such as engineering application systems.

Remark 4 Condition (C5) is a strong persistent excitation condition, which implies that the observed input and output data are rich enough. From Theorem 2, the convergence of the O-SG algorithm is guaranteed when Condition (C5) holds and the noise variance is bounded.

6. SIMULATION EXAMPLE

In this section, we test the effectiveness of the proposed algorithms by the simulation example. Consider the following FN-CAR system,

$$\begin{aligned}
&y(k)+a_1y(k-1)+a_2y(k-2)=b_1h(k-1)+b_2h(k-2)+v(k),\\
&h(k)=u(k)-g(y(k)),\\
&g(y(k))=c_1\sin(y^2(k))+c_2\sin(y^3(k)).
\end{aligned}$$

The parameter vector to be estimated is

$$\boldsymbol{\vartheta}=[a_1,a_2,b_1,b_2,c_1,c_2]^{\mathrm{T}}=[0.28,-0.12,0.72,-0.34,-0.08,0.18]^{\mathrm{T}}.$$

In simulation, the input $\{u(k)\}$ is taken as an uncorrelated Gaussian signal sequence with zero mean and unit variance. The noise $\{v(k)\}$ is taken as a white noise sequence with zero mean and variance σ^2. Using the simulation model parameters, the input signal $\{u(k)\}$ and the noise sequence $\{v(k)\}$ generates the output signal $\{y(k)\}$. The input and output signals of the system are shown in Figure 1. Set $L_e=3000$ as the data length for parameter estimation.

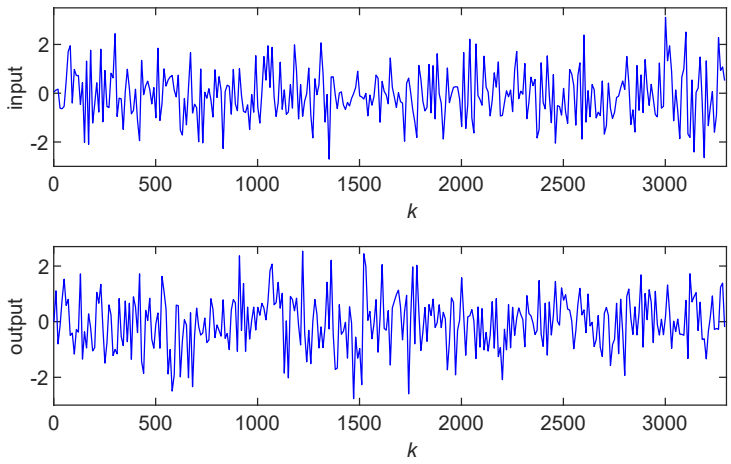

Figure 1. The input $y(k)$ and output $u(k)$

In order to illustrate the performance of the O-RLS and O-SG algorithms, we apply the proposed algorithms with different noise variances and forgetting factors to identify the parameters of the

DOI: 10.1002/acs

feedback nonlinear system. The parameter estimation errors $\delta := \|\hat{\boldsymbol{\vartheta}}(k) - \boldsymbol{\vartheta}\|/\|\boldsymbol{\vartheta}\|$ versus k are given to measure the parameter estimation accuracy. Tables III and IV show the parameter estimates and errors of the O-RLS and O-SG algorithms with the noise variances $\sigma^2 = 1.00^2$, $\sigma^2 = 0.50^2$ and $\sigma^2 = 0.10^2$, respectively. The parameter estimation errors δ versus k of the O-RLS and O-SG algorithms under different noise variances are shown in Figure 2. When the variance $\sigma^2 = 0.10^2$, Table V shows the parameter estimates and their errors of the O-FFSG algorithm with the forgetting factors $\lambda = 1$, $\lambda = 0.99$ and $\lambda = 0.98$, respectively. The parameter estimation errors δ versus k of the O-FFSG algorithm with different forgetting factors and the O-RLS algorithm for comparison are shown in Figure 3. The O-RLS and O-FFSG estimates of the parameters a_1, a_2, b_1 b_2, c_1 and c_2 versus k are plotted in Figures 4 to 7.

Table III. The O-RLS estimates and errors under different variances

σ^2	k	a_1	a_2	b_1	b_2	c_1	c_2	$\delta(\%)$
1.00^2	100	0.14318	-0.23090	0.57487	-0.36066	-0.29307	0.23975	36.40707
	200	0.19738	-0.22118	0.60837	-0.35377	-0.16745	0.18474	22.09703
	500	0.26460	-0.13944	0.61815	-0.34257	-0.07565	0.31131	19.21108
	1000	0.27069	-0.16004	0.67299	-0.34533	-0.13220	0.23957	11.54321
	2000	0.26184	-0.14630	0.69116	-0.34973	-0.09411	0.21473	6.61748
	3000	0.25835	-0.14961	0.71196	-0.34591	-0.05291	0.21949	6.98766
0.50^2	100	0.15244	-0.23057	0.65139	-0.39475	-0.13667	0.14471	23.04522
	200	0.18916	-0.19686	0.66265	-0.38136	-0.05025	0.18177	16.18178
	500	0.26302	-0.13399	0.66875	-0.34869	-0.09280	0.19679	6.88773
	1000	0.26182	-0.14674	0.69585	-0.35113	-0.07675	0.20398	5.52643
	2000	0.25769	-0.13935	0.70585	-0.35529	-0.07033	0.18209	4.28234
	3000	0.25728	-0.14257	0.71594	-0.35222	-0.07428	0.18593	4.05518
0.10^2	100	0.20272	-0.17251	0.70672	-0.38683	-0.09246	0.18244	12.12781
	200	0.23689	-0.14657	0.70852	-0.36514	-0.09470	0.18648	6.84390
	500	0.26064	-0.13265	0.70981	-0.35345	-0.09749	0.17884	3.83598
	1000	0.26144	-0.13475	0.71505	-0.35377	-0.08082	0.17869	3.18930
	2000	0.26153	-0.12988	0.71722	-0.35239	-0.08021	0.18851	2.96344
	3000	0.26723	-0.12719	0.71921	-0.34749	-0.08296	0.18925	2.18659
True values		0.28000	-0.12000	0.72000	-0.34000	-0.08000	0.18000	

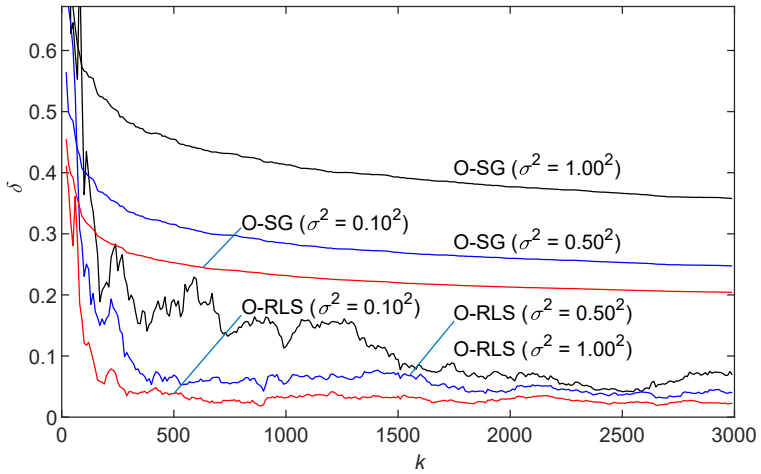

Figure 2. The estimation errors δ versus k with different variances

DOI: 10.1002/acs

Table IV. The O-SG estimates and errors under different variances

σ^2	k	a_1	a_2	b_1	b_2	c_1	c_2	$\delta(\%)$
1.00^2	100	0.05719	-0.21621	0.33052	-0.27428	-0.01992	0.01526	56.64618
	200	0.09226	-0.22543	0.36343	-0.29262	-0.01807	0.01622	51.92348
	500	0.14452	-0.22064	0.40178	-0.32105	-0.01774	0.02394	45.47697
	1000	0.17376	-0.22770	0.43608	-0.33351	-0.01960	0.02741	41.28132
	2000	0.18732	-0.21874	0.46610	-0.34294	-0.02079	0.03039	37.68042
	3000	0.19383	-0.21570	0.48340	-0.34562	-0.02050	0.03257	35.75415
0.50^2	100	0.15409	-0.18589	0.44737	-0.28178	-0.02632	0.02976	40.14129
	200	0.16968	-0.18566	0.47898	-0.30086	-0.02358	0.03299	36.30342
	500	0.19981	-0.18913	0.51396	-0.32693	-0.02543	0.04054	31.56069
	1000	0.21653	-0.19492	0.54288	-0.33882	-0.02551	0.04646	28.41534
	2000	0.22304	-0.19154	0.56730	-0.34714	-0.02686	0.04924	25.97461
	3000	0.22636	-0.19122	0.58062	-0.34966	-0.02760	0.05116	24.72799
0.10^2	100	0.26410	-0.15025	0.51022	-0.23243	-0.02031	0.03892	32.36853
	200	0.27511	-0.14606	0.54034	-0.25221	-0.02088	0.04141	28.77401
	500	0.29282	-0.14882	0.57130	-0.27342	-0.02230	0.04559	25.28142
	1000	0.30089	-0.14878	0.59262	-0.28426	-0.02286	0.04872	23.16865
	2000	0.30566	-0.14656	0.61216	-0.29258	-0.02394	0.05238	21.28098
	3000	0.30757	-0.14525	0.62191	-0.29613	-0.02451	0.05415	20.39997
True values		0.28000	-0.12000	0.72000	-0.34000	-0.08000	0.18000	

Table V. The O-FFSG estimates and errors with different forgetting factors ($\sigma^2 = 0.10^2$)

λ	k	a_1	a_2	b_1	b_2	c_1	c_2	$\delta(\%)$
1	100	0.26410	-0.15025	0.51022	-0.23243	-0.02031	0.03892	32.36853
	200	0.27511	-0.14606	0.54034	-0.25221	-0.02088	0.04141	28.77401
	500	0.29282	-0.14882	0.57130	-0.27342	-0.02230	0.04559	25.28142
	1000	0.30089	-0.14878	0.59262	-0.28426	-0.02286	0.04872	23.16865
	2000	0.30566	-0.14656	0.61216	-0.29258	-0.02394	0.05238	21.28098
	3000	0.30757	-0.14525	0.62191	-0.29613	-0.02451	0.05415	20.39997
0.99	100	0.27694	-0.15041	0.53834	-0.25039	-0.02243	0.04221	28.97259
	200	0.28994	-0.14137	0.58330	-0.27757	-0.02392	0.04648	24.00376
	500	0.31425	-0.13797	0.64387	-0.31197	-0.03055	0.05773	18.23994
	1000	0.31738	-0.12916	0.69735	-0.32391	-0.03508	0.07315	14.31450
	2000	0.30733	-0.11224	0.71733	-0.32482	-0.04671	0.10432	10.14597
	3000	0.31237	-0.11605	0.72354	-0.32724	-0.05704	0.12850	7.59645
0.98	100	0.28805	-0.14989	0.56631	-0.26668	-0.02465	0.04569	25.82450
	200	0.29998	-0.13553	0.62198	-0.29740	-0.02755	0.05218	20.17144
	500	0.31840	-0.12771	0.68065	-0.32639	-0.04109	0.07212	14.64496
	1000	0.31453	-0.12176	0.71840	-0.32672	-0.04687	0.09754	11.00434
	2000	0.30268	-0.10808	0.72266	-0.32504	-0.06197	0.14029	6.03643
	3000	0.30868	-0.11308	0.72325	-0.32959	-0.07289	0.16539	4.04581
True values		0.28000	-0.12000	0.72000	-0.34000	-0.08000	0.18000	

For the model validation, we use the rest $L_r = 300$ testing data from $k = L_e + 1 = 3001$ to $k = L_e + L_r = 3300$. For the variance $\sigma^2 = 0.10^2$, the predicted models are obtained by the O-RLS, O-SG and O-FFSG algorithms with the parameter estimates of the penultimate row in Tables III to V. The O-RLS predicted model output $\hat{y}_1(k)$, the O-SG predicted model output $\hat{y}_2(k)$ and the O-FFSG predicted model output $\hat{y}_3(k)$ are calculated as

$$\hat{y}_1(k) = \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(L_e - 1), k)\hat{\boldsymbol{\vartheta}}(L_e) = \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(2999), k)\hat{\boldsymbol{\vartheta}}(3000)$$

DOI: 10.1002/acs

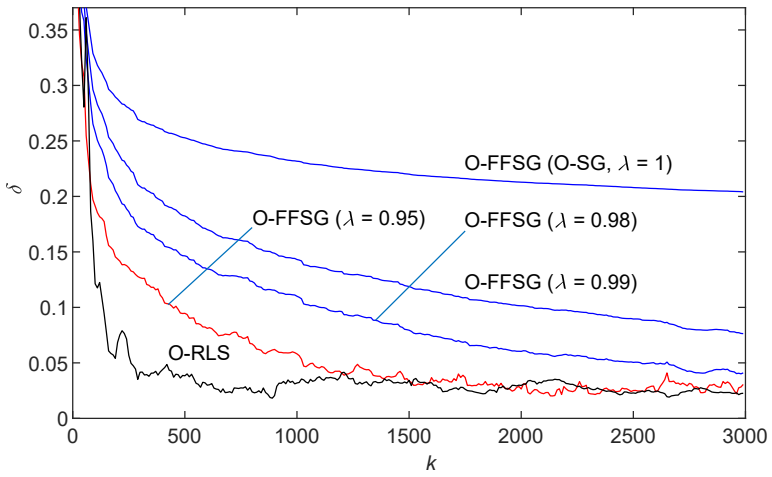

Figure 3. The estimation errors δ versus k ($\sigma^2 = 0.10^2$)

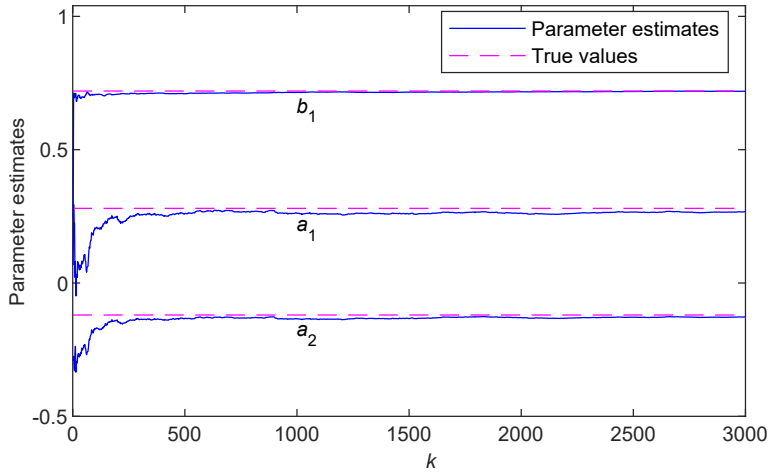

Figure 4. The O-RLS parameter estimates of a_1, a_2 and b_1 versus k ($\sigma^2 = 0.10^2$)

$$
\begin{aligned}
&= -0.26723y(k-1) + 0.12719y(k-2) + 0.71921u(k-1) - 0.34749u(k-2) \\
&\quad +0.71910 \times [-0.08296\sin(y^2(k-1)) + 0.18925\sin(y^3(k-1))] \\
&\quad -0.34739 \times [-0.08296\sin(y^2(k-2)) + 0.18925\sin(y^3(k-2))], \qquad (43) \\
\hat{y}_2(k) &= \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(L_e-1),k)\hat{\boldsymbol{\vartheta}}(L_e) = \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(2999),k)\hat{\boldsymbol{\vartheta}}(3000) \\
&= -0.30757y(k-1) + 0.14525y(k-2) + 0.62191u(k-1) - 0.29613u(k-2) \\
&\quad +0.62185 \times [-0.02451\sin(y^2(k-1)) + 0.05415\sin(y^3(k-1))] \\
&\quad -0.29615 \times [-0.02451\sin(y^2(k-2)) + 0.05415\sin(y^3(k-2))], \qquad (44) \\
\hat{y}_3(k) &= \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(L_e-1),k)\hat{\boldsymbol{\vartheta}}(L_e) = \boldsymbol{\varphi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(2999),k)\hat{\boldsymbol{\vartheta}}(3000) \\
&= -0.30868y(k-1) + 0.11308y(k-2) + 0.72325u(k-1) - 0.32959u(k-2) \\
&\quad +0.72182 \times [-0.07289\sin(y^2(k-1)) + 0.16539\sin(y^3(k-1))] \\
&\quad -0.32990 \times [-0.07289\sin(y^2(k-2)) + 0.16539\sin(y^3(k-2))]. \qquad (45)
\end{aligned}
$$

DOI: 10.1002/acs

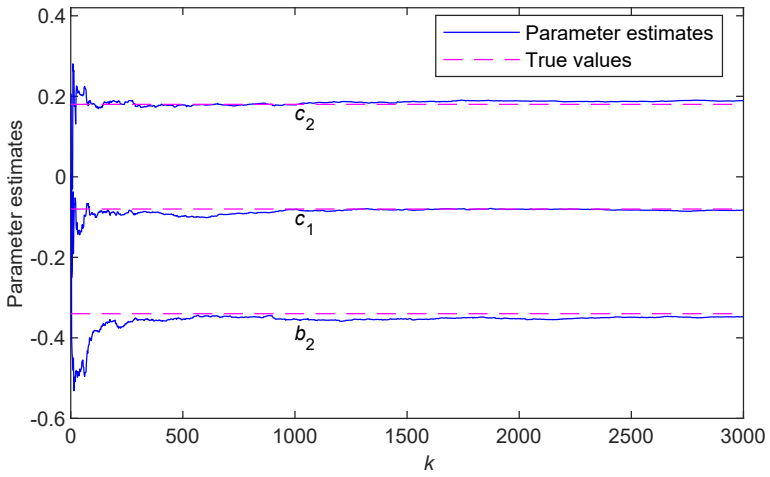

Figure 5. The O-RLS parameter estimates of b_2, c_1 and c_2 versus k ($\sigma^2 = 0.10^2$)

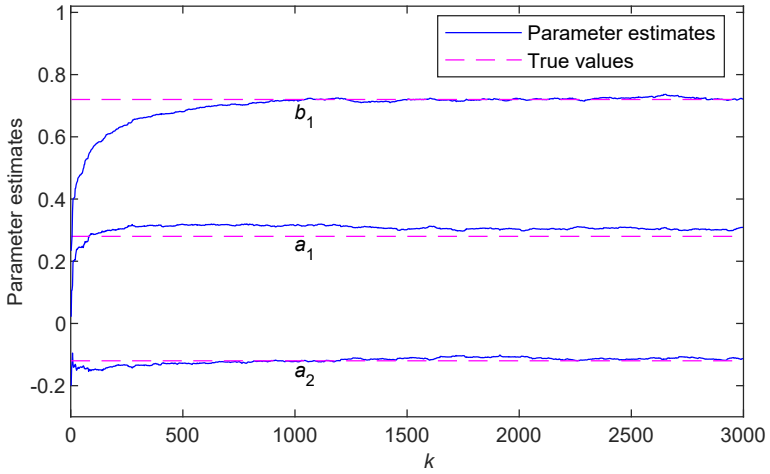

Figure 6. The O-FFSG parameter estimates of a_1, a_2 and b_1 versus k ($\lambda = 0.98, \sigma^2 = 0.10^2$)

The measurement output $y(k)$, the estimated outputs $\hat{y}_1(k)$, $\hat{y}_2(k)$ and $\hat{y}_3(k)$ in (43)–(45), and their estimation errors $\hat{y}_1(k) - y(k)$, $\hat{y}_2(k) - y(k)$ and $\hat{y}_3(k) - y(k)$ are plotted in Figures 8 to 10. We compute the root mean square errors to evaluate the model performance:

$$\text{Error}(\hat{y}_1(k)) = \sqrt{\frac{1}{L_r}\sum_{k=L_e+1}^{L_e+L_r}[\hat{y}_1(k) - y(k)]^2} = 0.10384,$$

$$\text{Error}(\hat{y}_2(k)) = \sqrt{\frac{1}{L_r}\sum_{k=L_e+1}^{L_e+L_r}[\hat{y}_2(k) - y(k)]^2} = 0.15415,$$

$$\text{Error}(\hat{y}_3(k)) = \sqrt{\frac{1}{L_r}\sum_{k=L_e+1}^{L_e+L_r}[\hat{y}_3(k) - y(k)]^2} = 0.10605.$$

From Tables III to V and Figures 1 to 10, we can draw the following conclusions:

DOI: 10.1002/acs

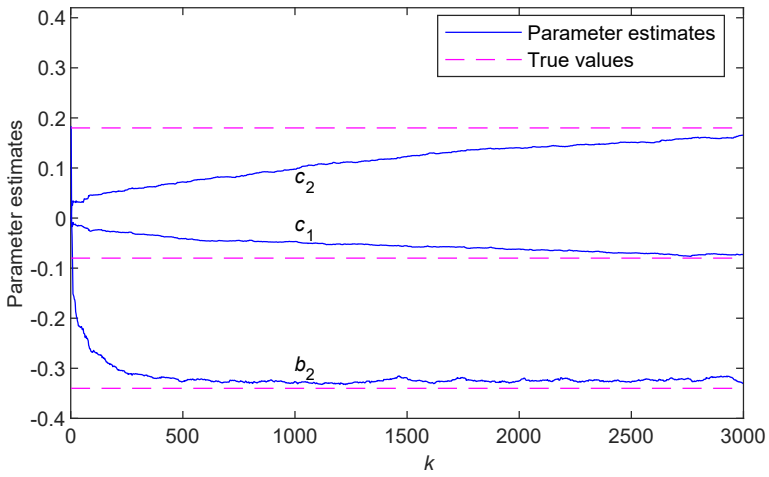

Figure 7. The O-FFSG parameter estimates of b_2, c_1 and c_2 versus k ($\lambda = 0.98$, $\sigma^2 = 0.10^2$)

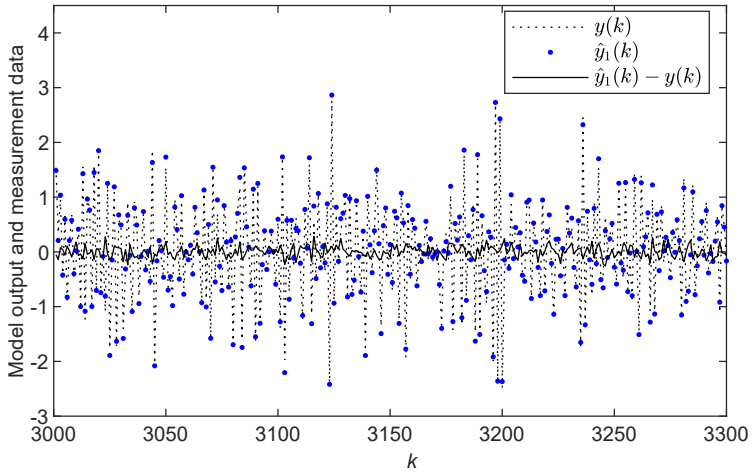

Figure 8. The model output $\hat{y}_1(k)$ based on the O-RLS estimated model and the measurement data $y(k)$ versus k ($\sigma^2 = 0.10^2$)

1. The parameter estimation errors given by the O-RLS and O-SG algorithms keep decreasing with the increasing of the data length – see Tables III to V.
2. For the same noise variance, the parameter estimation accuracy of the O-RLS algorithm becomes higher than that of the O-SG algorithm. With the noise levels decreasing, it is obvious that the parameter estimation errors of the O-RLS and O-SG algorithms become smaller, but the O-RLS algorithm still gives more accurate parameter estimation than the O-SG algorithm – see Figure 2.
3. Introducing the forgetting factor λ, the convergence rate of the O-FFSG algorithm becomes faster compared with the O-SG algorithm, and the O-FFSG algorithm can faster track the parameters. The parameter estimation errors of the O-FFSG algorithm become smaller as the forgetting factor λ decreases – see Figure 3.
4. For an appropriate forgetting factor $\lambda = 0.95$, the parameter estimates of the O-FFSG algorithm are very close to those of the O-RLS algorithm. The O-RLS and O-FFSG parameter estimation errors approach zero for large k – see Figure 3.

DOI: 10.1002/acs

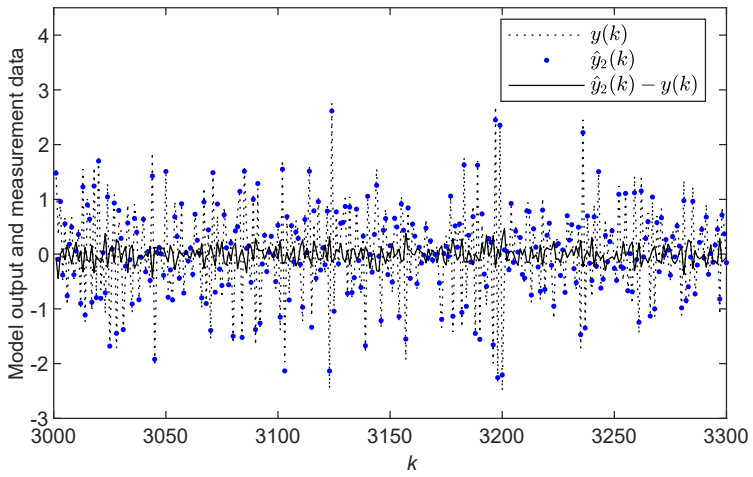

Figure 9. The model output $\hat{y}_2(k)$ based on the O-SG estimated model and the measurement data $y(k)$ versus k ($\sigma^2 = 0.10^2$)

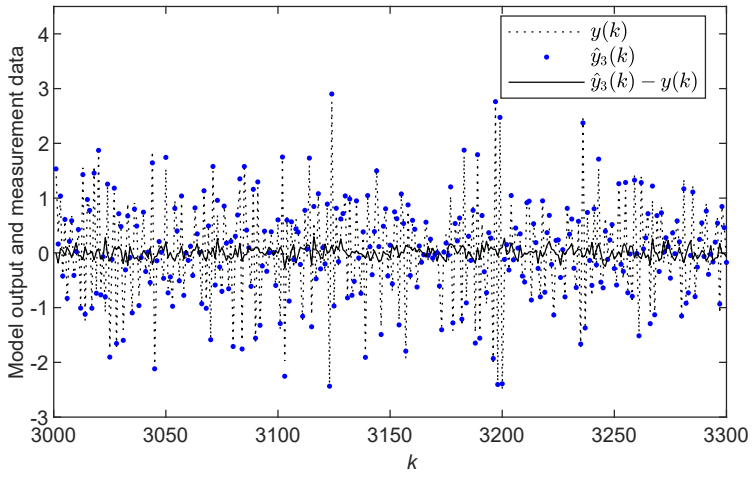

Figure 10. The model output $\hat{y}_3(k)$ based on the O-FFSG estimated model and the measurement data $y(k)$ versus k ($\lambda = 0.98$, $\sigma^2 = 0.10^2$)

5. At the initial period of the algorithm, the O-RLS algorithm can quickly converge around the true values while the O-FFSG algorithm has the slower convergence rate. With the increasing of k gradually, the parameter estimates of the O-RLS and O-FFSG algorithms gradually approach the true values and the parameter estimation accuracy of the O-RLS and O-FFSG algorithms are improved – see Figures 4 to 7.
6. The predicted model outputs obtained by the O-RLS, O-SG and O-FFSG algorithms are close to the measurement output, which indicates that the estimated model can capture the system dynamics. Compared with the root mean square errors among the three algorithms, it can be seen that the O-RLS algorithm has the smallest root mean square errors and it is the closest to the noise standard deviation σ. Therefore, the O-RLS algorithm performs better than the O-SG and O-FFSG algorithms – see Figures 8 to 10.

DOI: 10.1002/acs

7. CONCLUSIONS

This paper focuses on the parameter estimation problems for the feedback nonlinear system. Based on the least squares and the gradient search, the O-RLS, O-SG and O-FFSG algorithms are presented for the FN-CAR system. In terms of the O-RLS and O-SG algorithms, the convergence analysis proves that their parameter estimates can converge to their true values under the persistent excitation conditions. Moreover, the O-RLS algorithm can provide more accurate parameter estimates than the O-SG algorithm in the simulation example. For the O-FFSG algorithm, introducing the forgetting factor can enhance the parameter estimation accuracy. The proposed parameter identification algorithm for fedback nonlinear stochastic systems in this paper can be extended to other systems [91–95] and can be applied to other control and information processing systems [96–103] and so on.

ACKNOWLEDGEMENT

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

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

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. Kang Z, Ji Y, Liu XM. Hierarchical recursive least squares algorithms for Hammerstein nonlinear autoregressive output-error systems. *Int J Adapt Control Signal Process*. 2021;35(11):2276-2295.
2. Pan J, Jiang X, Wan XK, Ding W. A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems. *Int J Control Autom Syst*. 2017;15(3):1189-1197.
3. Ding F, Ma H, Pan J, Yang EF. Hierarchical gradient- and least squares-based iterative algorithms for input nonlinear output-error systems using the key term separation. *J Frankl Inst*. 2021;358(9):5113-5135.
4. Xu L. Separable multi-innovation Newton iterative modeling algorithm for multi-frequency signals based on the sliding measurement window. *Circuits Syst Signal Process*. 2022;41(2):805-830.
5. Xu L. Separable Newton recursive estimation method through system responses based on dynamically discrete measurements with increasing data length. *Int J Control Autom Syst*. 2022;20(2)432-443.
6. Pan J, Li W, Zhang HP. Control algorithms of magnetic suspension systems based on the improved double exponential reaching law of sliding mode control. *Int J Control Autom Syst*. 2018;16(6):2878-2887.
7. Xiong JX, Pan J, Chen GY. Sliding mode dual-channel disturbance rejection attitude control for a quadrotor. *IEEE Trans Ind Electron*. 2022. https://doi.org/10.1109/TIE.2021.3137600.
8. Wang YJ, Tang SH, Gu XB. Parameter estimation for nonlinear Volterra systems by using the multi-innovation identification theory and tensor decomposition. *J Frankl Inst*. 2022;359(2):1782-1802.
9. Wang YJ, Yang L. An efficient recursive identification algorithm for multilinear systems based on tensor decomposition. *Int J Robust Nonlinear Control*. 2021;31(11):7920-7936.
10. Pan J, Ma H, Zhang X, et al. Recursive coupled projection algorithms for multivariable output-error-like systems with coloured noises. *IET Signal Process*. 2020;14(7):455-466.
11. Oyerinde OO. Reweighted regularised variable step size normalised least mean square-based iterative channel estimation for multicarrier-interleave division multiple access systems. *IET Signal Process*. 2016;10(8):947-954.
12. Oliveira AM, Costa OLV. An iterative approach for the discrete-time dynamic control of Markov jump linear systems with partial information. *Int J Robust Nonlinear Control*. 2020;30(2):495-511.

13. Hou J, Chen FW, Li PH, Zhu ZQ. Gray-box parsimonious subspace identification of Hammerstein-type systems. *IEEE Trans Ind Electron*. 2021;68(10):9941-9951.
14. Li MH, Liu XM. Maximum likelihood hierarchical least squares-based iterative identification for dual-rate stochastic systems. *Int J Adapt Control Signal Process*. 2021;35(2):240-261.
15. Li MH, Liu XM. Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique. *Int J Control Autom Syst*. 2020;18(6):1581-1592.
16. Candan C. Making linear prediction perform like maximum likelihood in Gaussian autoregressive model parameter estimation. *Signal Process*. 2020;166:107256.
17. Zhang T, Lu ZR, Liu JK, et al. Parameter identification of nonlinear systems with time-delay from time-domain data. *Nonlinear Dyn*. 2021;104(4):4045-4061.
18. Liu X, Yang XQ, Yin S. Nonlinear system identification with robust multiple model approach. *IEEE Trans Control Syst Technol*. 2020;28(6):2728-2735.
19. Wei C, Xu L, Ding F. Stochastic gradient identification algorithm and its convergence for feedback nonlinear systems. *Control Theory Appl*. 2021;38:1-10.
20. Wang Z, An HR, Luo XL. Adaptive filtering-based recursive identification for time-varying Wiener output-error systems with unknown noise statistics. *J Frankl Inst*. 2020;357(2):1280-1298.
21. Hu PP, Ding F, Sheng J. Auxiliary model based least squares parameter estimation algorithm for feedback nonlinear systems using the hierarchical identification principle. *J Frankl Inst*. 2013;350(10):3248-3259.
22. Zhang XX, Xu J. Identification of time delay in nonlinear systems with delayed feedback control. *J Frankl Inst*. 2015;352(8):2987-2998.
23. Shen BB, Ding F, Xu L, et al. Data filtering based multi-innovation gradient identification methods for feedback nonlinear systems. *Int J Control Autom Syst*. 2018;16(5):2225-2234.
24. 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. *Int J Robust Nonlinear Control*. 2020;30(9):3607-3623.
25. Zhou YH. Partially-coupled nonlinear parameter optimization algorithm for a class of multivariate hybrid models. *Appl Math Comput*. 2022;414:126663.
26. Zhou YH. Modeling nonlinear processes using the radial basis function-based state-dependent autoregressive models. *IEEE Signal Process Lett*. 2020;27:1600-1604.
27. Zhou YH. Hierarchical estimation approach for RBF-AR models with regression weights based on the increasing data length. *IEEE Trans Circuits Syst II Express Briefs*. 2021;68(12):3597-3601.
28. Dong SJ, Yu L, Zhang WA, et al. Recursive identification for Wiener non-linear systems with non-stationary disturbances. *IET Control Theory Appl*. 2019;13(16):2648-2657.
29. Du YW, Liu FZ, Qiu JB, et al. A novel recursive approach for online identification of continuous-time switched nonlinear systems. *Int J Robust Nonlinear Control*. 2021;31(15):7546-7565.
30. Zhang H, Wang T, Zhao YL. Asymptotically efficient recursive identification of FIR systems with binary-valued observations. *IEEE Trans Syst Man Cybern Syst*. 2021;51(5):2687-2700.
31. Li LW, Ren XM, Guo FM. Modified multi-innovation stochastic gradient algorithm for Wiener-Hammerstein systems with backlash. *J Frankl Inst*. 2018;355(9):4050-4075.
32. Ji Y, Kang Z. Three-stage forgetting factor stochastic gradient parameter estimation methods for a class of nonlinear systems. *Int J Robust Nonlinear Control*. 2021;31(3):871-987.
33. Ji Y, Jiang XK, Wan LJ. Hierarchical least squares parameter estimation algorithm for two-input Hammerstein finite impulse response systems. *J Frankl Inst*. 2020;357(8):5019-5032.
34. Ji Y, Kang Z, Zhang C. Two-stage gradient-based recursive estimation for nonlinear models by using the data filtering. *Int J Control Autom Syst*. 2021;19(8):2706-2715.
35. Du N, Zhang L, Long XH, et al. Recursive identification for choke finger system in wind tunnel. *ISA Trans*. 2020;107:173-180.
36. Ding F, Liu XM, Hayat T. Hierarchical least squares identification for feedback nonlinear equation-error systems. *J Frankl Inst*. 2020;357(5):2958-2977.
37. Ding J, Chen LJ, Cao ZX, et al. Convergence analysis of the modified adaptive extended Kalman filter for the parameter estimation of a brushless DC motor. *Int J Robust Nonlinear Control*. 2021;31(16):7606-7620.
38. Li ZJ, Ding J, Lin JX. Discrete fractional order PID controller design for nonlinear systems. *Int J Syst Sci*. 2021;52(15):3206-3213.
39. Ding J, Cao ZX, Chen JZ, et al. Weighted parameter estimation for Hammerstein nonlinear ARX systems. *Circ Syst Signal Process*. 2020;39(4):2178-2192.
40. Feng L, Ding J, Han YY. Improved sliding mode based EKF for SOC estimation of lithium-ion batteries. *Ionics*. 2020;26(6):2875-2882.
41. Xu L, Yang EF. Auxiliary model multiinnovation stochastic gradient parameter estimation methods for nonlinear sandwich systems. *Int J Robust Nonlinear Control*. 2021;31(1):148-165.
42. Xu L, Hayat T. Hierarchical recursive signal modeling for multi-frequency signals based on discrete measured data. *Int J Adapt Control Signal Process*. 2021;35(5):676-693.
43. Xu L, Zhu QM. Decomposition strategy-based hierarchical least mean square algorithm for control systems from the impulse responses. *Int J Syst Sci*. 2021;52(9):1806-1821.
44. Wan LJ. Decomposition- and gradient-based iterative identification algorithms for multivariable systems using the multi-innovation theory. *Circ Syst Signal Process*. 2019;38(7):2971-2991.
45. Ding JL, Zhang WH. Finite-time adaptive control for nonlinear systems with uncertain parameters based on the command filters. *Int J Adapt Control Signal Process*. 2021;35(9):1754-1767.
46. Ma P, Wang L. Filtering-based recursive least squares estimation approaches for multivariate equation-error systems by using the multiinnovation theory. *Int J Adapt Control Signal Process*. 2021;35(9):1898-1915.
47. Mao YW, Liu S, Liu JF. Robust economic model predictive control of nonlinear networked control systems with communication delays. *Int J Adapt Control Signal Process*. 2020;34(5):614-637.

DOI: 10.1002/acs

48. Chen J, Huang B, Gan M, Chen CLP. A novel reduced-order algorithm for rational models based on Arnoldi process and Krylov subspace. *Automatica*. 2021;129, Article Number: 109663.
49. Chen J, Zhu QM, Liu YJ. Modified Kalman filtering based multi-step-length gradient iterative algorithm for ARX models with random missing outputs. *Automatica*. 2020;118, Article Number: 109034.
50. Chen J, Shen QY, Ma JX, Liu YJ. Stochastic average gradient algorithm for multirate FIR models with varying time delays using self-organizing maps. *Int J Adapt Control Signal Process*. 2020;34(7):955-970.
51. Ma H, Pan J, Ding W. Partially-coupled least squares based iterative parameter estimation for multi-variable output-error-like autoregressive moving average systems. *IET Control Theory Appl*. 2019;13(18):3040-3051.
52. Wang JW, Ji Y, Zhang C. Iterative parameter and order identification for fractional-order nonlinear finite impulse response systems using the key term separation. *Int J Adapt Control Signal Process*. 2021;35(8):1562-1577.
53. Ji Y, Zhang C, Kang Z, et al. Parameter estimation for block-oriented nonlinear systems using the key term separation. *Int J Robust Nonlinear Control*. 2020;30(9):3727-3752.
54. Li MH, Liu XM. Iterative identification methods for a class of bilinear systems by using the particle filtering technique. *Int J Adapt Control Signal Process*. 2021;35(10):2056-2074.
55. Ji Y, Kang Z, Liu XM. The data filtering based multiple-stage Levenberg-Marquardt algorithm for Hammerstein nonlinear systems. *Int J Robust Nonlinear Control*. 2021;31(15):7007-7025.
56. Fan YM, Liu XM. Two-stage auxiliary model gradient-based iterative algorithm for the input nonlinear controlled autoregressive system with variable-gain nonlinearity. *Int J Robust Nonlinear Control*. 2020;30(14):5492-5509.
57. Liu XM, Fan YM. Maximum likelihood extended gradient-based estimation algorithms for the input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity. *Int J Robust Nonlinear Control*. 2021;31(9):4017-4036.
58. Fan YM, Liu XM. Auxiliary model-based multi-innovation recursive identification algorithms for an input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity. *Int J Adapt Control Signal Process*. 2022;36(3):690-707.
59. Cao Y, Sun YK, Xie G, et al. A sound-based fault diagnosis method for railway point machines based on two-stage feature selection strategy and ensemble classifier. *IEEE Trans Intell Transp Syst*. 2022. doi:10.1109/TITS.2021.3109632.
60. Cao Y, Wen JK, Hobiny A, et al. Parameter-varying artificial potential field control of virtual coupling system with nonlinear dynamics. *Fractals*. 2022. doi: 10.1142/S0218348X22400990
61. Cao Y, Wen JK, Ma LC. Tracking and collision avoidance of virtual coupling train control system. *Alex Eng J*. 2021;60(2):2115-2125.
62. Sun YK, Cao Y, Ma LC. A fault diagnosis method for train plug doors via sound signals. *IEEE Intell Transp Syst Mag*. 2021;13(3):107-117.
63. Sun Y, Cao Y, Xie G, Wen T. Sound based fault diagnosis for RPMs based on multi-scale fractional permutation entropy and two-scale algorithm. *IEEE Trans Veh Technol*. 2021;70(11):11184-11192.
64. Su S, Wang XK, Cao Y, et al. An energy-efficient train operation approach by integrating the metro timetabling and eco-driving. *IEEE Trans Intell Transp Syst*. 2020;21(10):4252-4268.
65. Cao Y, Wang Z. Liu F, et al. Bio-inspired speed curve optimization and sliding mode tracking control for subway trains. *IEEE Trans Veh Technol*. 2019;68(7):6331-6342.
66. 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 Trans Veh Technol*. 2019;68(8):7544-7551.
67. Cao Y, Ma LC, Xiao S, et al. Standard analysis for transfer delay in CTCS-3. *Chinese J Electron*. 2017;26(5):1057-1063.
68. Su S, She JF, Li KC, et al. A nonlinear safety equilibrium spacing based model predictive control for virtually coupled train set over gradient terrains. *IEEE Trans Transp Electr*. 2022. doi:10.1109/TTE.2021.3134669.
69. Su S, Han L, Li SK. Finite-time event-triggered consensus control for high-speed train with gradient resistance. *J Frankl Inst*. 2022;359. https://doi.org/10.1016/j.jfranklin.2021.11.012.
70. Su S, Tang T, Xun J, et al. Design of running grades for energy-efficient train regulation: A case study for beijing yizhuang line. *IEEE Intell Transp Syst Mag*. 2021;13(2):189-200.
71. Li XY, Wang HL, Wu BY. A stable and efficient technique for linear boundary value problems by applying kernel functions. *Appl Numer Math*. 2022;172:206-214.
72. Li XY, Wu BY. A kernel regression approach for identification of first order differential equations based on functional data. *Appl Math Lett*. 2022;127:107832.
73. Geng FZ, Wu XY. Reproducing kernel functions based univariate spline interpolation. *Appl Math Lett*. 2021;122:107525.
74. Li XY, Wu BY. Superconvergent kernel functions approaches for the second kind Fredholm integral equations. *Appl Numer Math*. 2021;167:202-210.
75. Dong H, Yin CC, Dai HS. Spectrally negative Levy risk model under Erlangized barrier strategy. *J Comput Appl Math*. 2019;351:101-116.
76. Sha XY, Xu ZS, Yin CC. Elliptical distribution-based weight-determining method for ordered weighted averaging operators. *Int J Intell Syst*. 2019;34(5):858-877.
77. Yin CC, Wen YZ. An extension of Paulsen-Gjessing's risk model with stochastic return on investments. *Insur Math Econom*. 2013;52(3):469-476.
78. Zhao YX, Chen P, Yang HL. Optimal periodic dividend and capital injection problem for spectrally positive Levy processes. *Insur Math Econom*. 2017;74:135-146.
79. Zhao XH, Dong H, Dai HS. On spectrally positive Levy risk processes with Parisian implementation delays in dividend payments. *Stat Prob Lett*. 2018;140:176-184.
80. Zhao YX, Yin CC. The expected discounted penalty function under a renewal risk model with stochastic income. *Appl Math Comput*. 2012;218(10):6144-6154.
81. Xiong W, Jia X, Yang D, Ai M, Li L, Wang S., DP-LinkNet: A convolutional network for historical document image binarization. *KSII Transactions on Internet and Information Systems*. 2021;15(5):1778-1797.

82. Xiong W, Zhou L, Yue L, Li L, Wang S. An enhanced binarization framework for degraded historical document images. *EURASIP Journal on Image and Video Processing*. 2021;2021, Article number: 13.
83. Chang C, Wu YT, Jiang JC, Jiang Y, Tian AN, Li TY, Gao Y. Prognostics of the state of health for lithium-ion battery packs in energy storage applications. *Energy*. 2022;239, Part B: 122189.
84. Zhao G, Gao TH, Wang YD, et al. Optimal sizing of isolated microgrid containing photovoltaic/photothermal/wind/diesel/battery. *Int J Photoenergy*. 2021;2021:5566597.
85. Wang XG, Zhao M, Zhou Y, Wan ZW, Xu W. Design and analysis for multi-disc coreless axial-flux permanent-magnet synchronous machine. *IEEE Transn Appl Superconductivity*. 2021;31(8):5203804.
86. Wang XG, Wan ZW, Tang L, Xu W, Zhao M. Electromagnetic performance analysis of an axial flux hybrid excitation motor for HEV drives. *IEEE Trans Appl Superconductivity*. 2021;31(8):5205605.
87. Li M, Xu G, Lai Q, Chen J. A chaotic strategy-based quadratic Opposition-Based Learning adaptive variable-speed whale optimization algorithm. *Mathematics and Computers in Simulation*. 2022;193:71-99.
88. Shu J, He JC, Li L. MSIS: Multispectral instance segmentation method for power equipment. *Comput Intell Neurosci*. 2022;2022, Article ID 2864717. https://doi.org/10.1155/2022/2864717
89. An Y, Zhang Y, Gao W, Tong Z, He ZQ. A lightweight and practical anonymous authentication protocol based on bit-self-test PUF. *Electronics*. 2022;11.
90. Wang H, Fan H, Pan J. Complex dynamics of a four-dimensional circuit system. *Int J Bifurcation Chaos*. 2021;31(14):2150208.
91. Li MH. 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.
92. Zhang X. Adaptive parameter estimation for a general dynamical system with unknown states. *Int J Robust Nonlinear Control*. 2020;30(4):1351-1372.
93. Zhang X. Recursive parameter estimation methods and convergence analysis for a special class of nonlinear systems. *Int J Robust Nonlinear Control*. 2020;30(4):1373-1393.
94. Zhao ZY, Zhou YQ, Wang XY, Wang ZY, Bai YT. Water quality evolution mechanism modeling and health risk assessment based on stochastic hybrid dynamic systems. *Expert Systems With Applications*. 2022;193:116404.
95. Chen Q, Zhao ZY, Wang XY, et al. Microbiological predictive modeling and risk analysis based on the one-step kinetic integrated Wiener process. *Innovat Food Sci Emerg Technol*. 2022;75:102912.
96. Jin X, Gong W, Kong J, Bai Y, Su T. PFVAE: a planar flow-based variational auto-encoder prediction model for time-series data. *Mathematics*. 2022;10(4):610.
97. Jin X, Zhang J, Kong J, Su T, Bai Y. A reversible automatic selection normalization (RASN) deep network for predicting in the smart agriculture system. *Agronomy*. 2022;12(3):591.
98. Jin X, Gong W, Kong J, Bai Y, Su T. A variational Bayesian deep network with data self-screening layer for massive time-series data forecasting. *Entropy*. 2022;24(3):335.
99. Zhao N, Wu A, Pei Y, et al. Spatial-temporal aggregation graph convolution network for efficient mobile cellular traffic prediction. IEEE Commun Lett. 2022;26 doi: 10.1109/LCOMM.2021.3138075.
100. Chen YF, Zhang C, Liu CY, et al. Atrial fibrillation detection using feedforward neural network. J Med Biolog Eng 2022.
101. Lu MY, Wu J, Zhan XS, Han T, Yan HC. Consensus of second-order heterogeneous multi-agent systems with and without input saturation. *ISA Trans*. 2022. https://doi.org/10.1016/j.isatra.2021.08.001
102. Lin J, Li Y, Yang GC. FPGAN: Face de-identification method with generative adversarial networks for social robots. *Neural Networks*. 2021;133:132-147.
103. 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):49.