International Journal of Control, Automation, and Systems 21(6) (2023) 1780-1792
http://dx.doi.org/10.1007/s12555-021-1028-6
ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

Parameter Estimation for Nonlinear Functions Related to System Responses

Ling Xu

Abstract: This paper considers the parameter estimation problem of nonlinear models, which are related to the impulse or step response functions of linear time-invariant (LTI) dynamical systems, based on the response data. In terms of the nonlinear characteristic of the models, the nonlinear dynamical optimization scheme is adopted for obtaining the system parameter estimates. By constructing a gradient criterion function, a gradient recursion algorithm is derived. In order to overcome the difficulty of determining the step-size in the gradient recursion algorithm, a trying method and a numerical approach are proposed to achieve the step-size. On this basis, a stochastic gradient estimation method is presented by using a recursive step-size. Furthermore, a multi-innovation stochastic gradient method is deduced for enhancing the estimation accuracy by using the dynamical window data. Finally, a dynamical length stochastic gradient estimation technique is offered to obtain more accurate parameter estimates by using dynamical length measured data from the step response. The examples are provided to examine the algorithm performance and the simulation results indicate that the presented approaches are effective.

Keywords: Gradient search, multi-innovation theory, nonlinear model, parameter estimation, recursive identification.

1. INTRODUCTION

The use of mathematical models is to accomplish different objectives in the design of control systems. Parameter estimation is an important step during the process of developing a mathematical model [1-5]. The aim of parameter estimation is to tune the model's parameters after the mathematical model structure is determined on the basis of some physical principles. In general, the parameter estimates are obtained by comparing the model outputs with system measurement outputs [6-9]. Parameter estimation methods based on statistical data have attracted much attention [10-13]. Recently, a parameter estimation approach was devised by using available observations on the slow component in terms of a slow-fast stochastic dynamical system [14], a gradient estimation algorithm was presented for estimating the parameters of the system with scarce measurement data [15].

System response is the system output under some excitation. The system response contains the dynamical information of systems. For instance, the controller design is to control the system response within the expectation range [16]. The transfer functions of linear time-invariant systems can be computed using the step response [17,18], the impulse response [19] and the frequency response experiments [20]. The system responses of dynamical systems are generally highly nonlinear functions with respect to the system parameters and thus the system parameters can be estimated by means of the system response measurement data. In order to obtain the parameter estimates, the nonlinear optimization techniques are adopted to deduce the identification algorithms for nonlinear models in this paper.

By constructing and optimizing a cost function, the parameter estimation method can be derived [21-24]. Many optimization strategies are used for deriving parameter estimation methods [25-30]. Carvalho *et al*. discussed the numerical difficulties of the nonlinear optimization formulation of parameter estimation and suggested an unconstrained derivative-free optimization method [31]. Lin *et al*. considered the parameter estimation of a nonlinear time-delay system using the dynamical optimization and the gradient search [32]. Isaksson *et al*. established an identification method using the moving horizon estimation and the nonlinear optimization for joint parameter estimation and state estimation [33]. On the basis of the hierarchical identification principle and the nonlinear optimization, this paper presents new parameter estimation

Manuscript received December 5, 2021; revised April 29, 2022; accepted June 29, 2022. Recommended by Associate Editor Joseph Kwon under the direction of Senior Editor Jay H. Lee. This work was supported by the National Natural Science Foundation of China (No. 61873111).

Ling Xu is with the School of Internet of Things and Artificial Intelligence, Wuxi Vocational Institute of Commerce, Wuxi 214153, China, the School of Microelectronics and Control Engineering, Changzhou University, Changzhou 213159, China, the School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China, and the School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China (e-mail: lingxu0848@163.com).

approaches for dynamical systems by means of the system response observation data. The contributions of this work are summarized as follows:

1) This paper presents some identification methods based on the dynamical responses of systems by applying the excitation signals and using the dynamical observations of the systems to be identified.
2) In order to test the accuracy of different number of observations for the recursive estimation, three schemes of using the dynamical data are designed, i.e., the single datum, the batch data and the dynamical length data, which are dynamically sampled and used in the recursive computation.
3) By collecting and employing the dynamical observations of the system responses, a recursive gradient algorithm, a multi-innovation recursive gradient algorithm and a dynamical length recursive gradient algorithm are presented based on different schemes of using the system observations.

The rest of this paper is organized as follows: Section 2 describes the problem of parameter estimation based on the response data. Section 3 derives the gradient based recursive method and the stochastic gradient algorithm. Section 4 deduces the multi-innovation stochastic gradient algorithm based on the moving data window. Section 5 devises the stochastic gradient algorithm based on the dynamical length data. Section 6 provides some simulation examples to show the performance of the proposed methods. Finally, Section 7 gives the concluding remarks.

2. PROBLEM DESCRIPTION

The dynamical characteristic of linear systems can be described by transfer functions. A general system transfer function is described as

$$G(s)=\frac{\beta_m s^m+\beta_{m-1}s^{m-1}+\cdots+\beta_0}{\alpha_n s^n+\alpha_{n-1}s^{n-1}+\cdots+\alpha_1 s+1}, \tag{1}$$

which can equivalently be expressed as the partial fractions, where n is the system order, α_1, α_2, $\cdots$, α_n, β_0, β_1, $\cdots$ and β_m are the system parameters ($\alpha_n \neq 0$, $m \leqslant n$).

For a practical system, it is easy to obtain its impulse (step) response $y(t)$ through imposing an excitation signal on the input terminal.

Remark 1: Although the system described by the transfer function $G(s)$ in (1) is linear, its outputs $y(t)$ (e.g., impulse response, step response, etc) are highly nonlinear about the system parameters a_i's and b_i's. For example, for the transfer function $G(s)=\frac{b_1}{s+a_1}+\frac{b_2}{s+a_2}$, its impulse response $g(t)=b_1\mathrm{e}^{-a_1 t}+b_2\mathrm{e}^{-a_2 t}$ is linear with respect to the parameters b_1 and b_2 but is highly nonlinear with respect to the parameters a_1 and a_2.

In general, the input signals in identification use step signals, impulse signals, sine signals and pseudo-random binary codes (PRBS) signals. Once the input signal is applied to the system as the excitation, one can collect the output data, which are the discrete sampled-data of the system response. For a dynamical system with the transfer function $G(s)$, its response is the function of the parameters a_1, a_2, $\cdots$, a_n, b_0, b_1, $\cdots$, b_m and time t. Therefore, the response function can be represented as $y(t)=f(\boldsymbol{\theta},t)$, where $\boldsymbol{\theta}:=[a_1,a_2,\cdots,a_n,b_0,b_1,\cdots,b_m]^{\mathrm{T}}$ is the unknown parameter vector.

Suppose that the discrete response data are $\{t_k,y(t_k),\ k=1,2,\cdots,L\}$, where L is the data length. The task of system identification is to study new methods for estimating the parameter vector $\boldsymbol{\theta}$ from the data $\{t_k,y(t_k)\}$. In order to obtain parameter estimates, we can construct equations with respect to the parameter vector $\boldsymbol{\theta}$ and solve these equations. Using the obtained data, the equations for estimating these parameters are given by $y(t_k)=f(\boldsymbol{\theta},t_k)$, $k=1,2,\cdots,L$.

The numerical solutions of the parameters a_i's and b_i's can be obtained by solving these equations $y(t_k)=f(\boldsymbol{\theta},t_k)$, $k=1,\ 2,\ \cdots,\ L$. However, there have no common ways for solving nonlinear equations. In general, the numerical solution of the nonlinear equations can be obtained by using some optimization method. Therefore, the numerical solutions are the approximate solutions by the successive approximation. This paper explores some novel numerical methods for solving such nonlinear equations for generating the parameter estimates of the systems.

3. THE GRADIENT BASED RECURSIVE ALGORITHM AND STOCHASTIC GRADIENT ALGORITHM

The function $y(t)=f(\boldsymbol{\theta},t)$ describes a mapping relationship over the definition domain of the parameter vector $\boldsymbol{\theta}$ and the time variable t. Since $y(t)=f(\boldsymbol{\theta},t)$ is nonlinear, it is almost impossible to directly solve these nonlinear equations $y(t_k)=f(\boldsymbol{\theta},t_k)$ with respect to the parameter vector $\boldsymbol{\theta}$ in order to obtain the parameter estimates. Thus, this paper derives several new recursive parameter estimation methods using the nonlinear optimization techniques.

According to the definition of identification, construct a dynamical criterion function about the output $y(t_k)$ and the model output $f(\boldsymbol{\theta},t_k)$, e.g., the gradient criterion function $J(\boldsymbol{\theta}):=\frac{1}{2}[y(t_k)-f(\boldsymbol{\theta},t_k)]^2$, which varies dynamically as k increases and is a rolling optimization criterion function. Various estimation approaches are proposed according to the identification criterion functions of the systems [34-39] and these methods can be employed to identify other linear and nonlinear systems [40-45] and can be used in other areas [46-51] such as process engineering systems. The gradient search direction is determined by calculating the first derivative of the criterion function. Taking the first-order derivation of $J(\boldsymbol{\theta})$ with respect to

$\boldsymbol{\theta}$ gives

$$\mathrm{grad}[J(\boldsymbol{\theta})] := \frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{\partial f(\boldsymbol{\theta},t_k)}{\partial \boldsymbol{\theta}}[y(t_k) - f(\boldsymbol{\theta},t_k)].$$

Let $\hat{\boldsymbol{\theta}}(t_k)$ be the estimate of $\boldsymbol{\theta}$ at time $t = t_k$ and $\mu(t_k)$ be the step-size, and $n_0 := n+m+1$. Define the information vector

$$\boldsymbol{\varphi}(t_k) := \left.\frac{\partial f(\boldsymbol{\theta},t_k)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(t_{k-1})} = \frac{\partial f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k)}{\partial \boldsymbol{\theta}} \in \mathbb{R}^{n_0}.$$

Define the identification innovation $e(t_k) := y(t_k) - f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k) \in \mathbb{R}$. Minimizing the criterion function $J(\boldsymbol{\theta})$, we can obtain the recursive relation

$$\begin{aligned}\hat{\boldsymbol{\theta}}(t_k) &= \hat{\boldsymbol{\theta}}(t_{k-1}) + \mu(t_k)\boldsymbol{\varphi}(t_k)[y(t_k) - f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k)] \\ &= \hat{\boldsymbol{\theta}}(t_{k-1}) + \mu(t_k)\boldsymbol{\varphi}(t_k)e(t_k). \qquad (2)\end{aligned}$$

In the gradient recursion (GR) algorithm in (2), the step-size $\mu(t_k)$ is an important factor because it reflects the convergence rate and the estimation accuracy. Here, the step-size $\mu(t_k)$ can be determined by the following method. In order to determine the step-size $\mu(t_k)$ quickly, we can preset a small positive number $\mu(t_k)$ and substitute it into the criterion function at each recursion k. If $J(\hat{\boldsymbol{\theta}}(t_{k-1}) - \mu(t_k)\mathrm{grad}[J(\hat{\boldsymbol{\theta}}(t_{k-1})]) < J(\hat{\boldsymbol{\theta}}(t_{k-1}))$, the step-size $\mu(t_k)$ can make the criterion function decrease, then we use $\mu(t_k)$ as the step-size at the next recursion; otherwise, choose a new positive number and repeat the trying process.

After the gradient search direction is determined, the step-size can be obtained by the one-dimensional search method. An optimal step-size can be computed from

$$\mu(t_k) = \underset{\mu \geqslant 0}{\mathrm{argmin}} J[\hat{\boldsymbol{\theta}}(t_{k-1}) + \mu\boldsymbol{\varphi}(t_k)e(t_k)].$$

However, because the response function is nonlinear, it is difficult to determine the step-size $\mu(t_k)$ from the above equation. Here, we consider adopting the analytical method and the numerical solution method for finding $\mu(t_k)$. Let the quantity $\mu(t_k)$ be the step-size which makes the criterion function reach the minimum. Substituting $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(t_k)$ into the criterion function $J(\boldsymbol{\theta})$ gives

$$J(\boldsymbol{\theta}) = J(\hat{\boldsymbol{\theta}}(t_k)) = J(\hat{\boldsymbol{\theta}}(t_{k-1}) + \mu\boldsymbol{\varphi}(t_k)e(t_k)).$$

Taking the second-order Taylor approximation of $J(\hat{\boldsymbol{\theta}}(t_{k-1}) + \mu\boldsymbol{\varphi}(t_k)e(t_k))$ at $\hat{\boldsymbol{\theta}}(t_{k-1})$ gives

$$\begin{aligned}J(\hat{\boldsymbol{\theta}}(t_k)) &= J(\hat{\boldsymbol{\theta}}(t_{k-1}) + \mu\boldsymbol{\varphi}(t_k)e(t_k)) \approx J(\hat{\boldsymbol{\theta}}(t_{k-1})) \\ &\quad + \mu\boldsymbol{\varphi}^{\mathrm{T}}(t_k)\frac{\partial J(\hat{\boldsymbol{\theta}}(t_{k-1}))}{\partial \boldsymbol{\theta}}e(t_k) \\ &\quad + \frac{1}{2}\mu^2 e^2(t_k)\boldsymbol{\varphi}^{\mathrm{T}}(t_k)\frac{\partial^2 J(\hat{\boldsymbol{\theta}}(t_{k-1}))}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}}\boldsymbol{\varphi}(t_k), \qquad (3)\end{aligned}$$

$$\begin{aligned}\frac{\partial J(\hat{\boldsymbol{\theta}}(t_{k-1}))}{\partial \boldsymbol{\theta}} &:= -\mathrm{grad}[f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k)] \\ &\quad \times [y(t_k) - f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k)] \\ &= -\boldsymbol{\varphi}(t_k)e(t_k), \qquad (4)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 J(\hat{\boldsymbol{\theta}}(t_{k-1}))}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}} &:= \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\frac{\partial J(\hat{\boldsymbol{\theta}}(t_{k-1}))}{\partial \boldsymbol{\theta}} \\ &= -\frac{\partial \boldsymbol{\varphi}(t_k)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}e(t_k) - \boldsymbol{\varphi}(t_k)\frac{\partial e(t_k)}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \\ &= -\frac{\partial \boldsymbol{\varphi}(t_k)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}e(t_k) + \boldsymbol{\varphi}(t_k)\boldsymbol{\varphi}^{\mathrm{T}}(t_k). \qquad (5)\end{aligned}$$

Define $\rho(t_k) := \boldsymbol{\varphi}^{\mathrm{T}}(t_k)\boldsymbol{H}(t_k)\boldsymbol{\varphi}(t_k)$ and

$$\boldsymbol{H}(t_k) := \frac{\partial \boldsymbol{\varphi}(t_k)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}e(t_k) + \boldsymbol{\varphi}(t_k)\boldsymbol{\varphi}^{\mathrm{T}}(t_k) \in \mathbb{R}^{n_0 \times n_0}. \qquad (6)$$

Substituting (4) and (6) into (3) gives

$$\begin{aligned}J(\hat{\boldsymbol{\theta}}(t_k)) &= J(\hat{\boldsymbol{\theta}}(t_{k-1}) + \mu\boldsymbol{\varphi}(t_k)e(t_k)) \\ &\approx J(\hat{\boldsymbol{\theta}}(t_{k-1})) - \mu\boldsymbol{\varphi}^{\mathrm{T}}(t_k)\boldsymbol{\varphi}(t_k)e^2(t_k) \\ &\quad + \frac{1}{2}\mu^2 e^2(t_k)\boldsymbol{\varphi}^{\mathrm{T}}(t_k)\boldsymbol{H}(t_k)\boldsymbol{\varphi}(t_k) \\ &= J(\hat{\boldsymbol{\theta}}(t_{k-1})) - \mu\|\boldsymbol{\varphi}(t_k)\|^2 e^2(t_k) \\ &\quad + \frac{1}{2}\mu^2 e^2(t_k)\boldsymbol{\varphi}^{\mathrm{T}}(t_k)\boldsymbol{H}(t_k)\boldsymbol{\varphi}(t_k). \qquad (7)\end{aligned}$$

Note that $J(\hat{\boldsymbol{\theta}}(t_{k-1})) = \frac{1}{2}e^2(t_k)$. Let $g(\mu) := 2J(\hat{\boldsymbol{\theta}}(t_{k-1}) + \mu\boldsymbol{\varphi}(t_k)e(t_k))$. Then, (7) becomes

$$\begin{aligned}g(\mu) &= e^2(t_k) - 2\mu\|\boldsymbol{\varphi}(t_k)\|^2 e^2(t_k) + \mu^2 e^2(t_k)\rho(t_k) \\ &= [1 - 2\mu\|\boldsymbol{\varphi}(t_k)\|^2 + \mu^2\rho(t_k)]e^2(t_k).\end{aligned}$$

The best step-size can be obtained by minimizing $g(\mu)$. Taking the first-order derivative of $g(\mu)$ with respect to μ and letting it be zero give

$$g'(\mu) = 2\mu\rho(t_k)e^2(t_k) - 2\|\boldsymbol{\varphi}(t_k)\|^2 e^2(t_k) = 0.$$

When $\rho(t_k) \neq 0$, solving the above equation gives

$$\mu(t_k) = \frac{\|\boldsymbol{\varphi}(t_k)\|^2}{\rho(t_k)} = \frac{\|\boldsymbol{\varphi}(t_k)\|^2}{\boldsymbol{\varphi}^{\mathrm{T}}(t_k)\boldsymbol{H}(t_k)\boldsymbol{\varphi}(t_k)}.$$

Next, we introduce the mathematical method for finding the optimal step-size. The basic idea is to obtain the optimal value through the iterative method. The optimization method can adopt the gradient search, the Newton search and so on. Substituting $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(t_k)$ into $J(\boldsymbol{\theta})$ gives

$$\begin{aligned}g(\mu) &:= 2J(\hat{\boldsymbol{\theta}}(t_k)) \\ &= [y(t_k) - f(\hat{\boldsymbol{\theta}}(t_{k-1}) - \mu\boldsymbol{\varphi}(t_k)e(t_k),t_k)]^2.\end{aligned}$$

The optimal step-size $\mu(t_k)$ can be determined by solving $g'(\mu) = 0$. Taking the first-order derivative of $g(\mu)$ with respect to μ gives

$$g'(\mu) = 2\frac{\partial J(\hat{\boldsymbol{\theta}}(t_k))}{\partial \mu}$$

$$= -2[y(t_k) - f(\hat{\boldsymbol{\theta}}(t_{k-1}) - \mu\boldsymbol{\varphi}(t_k)e(t_k),t_k)] \times f'_\mu(\hat{\boldsymbol{\theta}}(t_{k-1}) - \mu\boldsymbol{\varphi}(t_k)e(t_k),t_k).$$

Letting $g'(\mu) = 0$ gives

$$[y(t_k) - f(\hat{\boldsymbol{\theta}}(t_{k-1}) - \mu(t_k)\boldsymbol{\varphi}(t_k)e(t_k),t_k)] \times f'_\mu(\hat{\boldsymbol{\theta}}(t_{k-1}) - \mu\boldsymbol{\varphi}(t_k)e(t_k),t_k) = 0, \tag{8}$$

where $f'_\mu(\hat{\boldsymbol{\theta}}(t_{k-1}) - \mu\boldsymbol{\varphi}(t_k)e(t_k),t_k)$ is the first-order derivative with respect to μ. Solving (8) can obtain the optimal step-size $\mu(t_k)$. If $f(\boldsymbol{\theta},t_k)$ is linear, it is easy to solve by direct solving. If $f(\boldsymbol{\theta},t_k)$ is nonlinear, the iterative method can be used to get an approximation solution.

Remark 2: Equation (8) is a one-degree equation with respect to μ. It is described as $\lambda(\mu) = 0$. When $f(\boldsymbol{\theta},t)$ is linear on the parameter space, we can obtain the algebraic solution of $\lambda(\mu) = 0$. But when $f(\boldsymbol{\theta},t)$ is not linear, one can only use the iterative method to solve $\lambda(\mu) = 0$, such as the Newton iterative method, the gradient iterative method and so on. The iterative formula of the Newton iterative algorithm is given by $\mu_{l+1} = \mu_l - \lambda'(\mu_l)/\lambda''(\mu_l)$, $l = 0, 1, 2, \cdots$, where l is the iterative variable. After giving an initial value of the step-size, the satisfied step-size can be obtained by multiple iterations. The gradient iteration does not need computing the second-order derivative and has less computation load. The gradient iterative method can be used to obtain the numerical solution of the step-size. The formula of gradient iterative method is given by $\mu_{l+1} = \mu_l - \eta_l\lambda'(\mu_l)$, $l = 0, 1, 2, \cdots$. This process needs multiple iterations to achieve the optimal value μ_l.

Finally, the GR algorithm for computing the parameter estimation vector $\hat{\boldsymbol{\theta}}(t_k)$ can be expressed as

$$\hat{\boldsymbol{\theta}}(t_k) = \hat{\boldsymbol{\theta}}(t_{k-1}) + \mu(t_k)\boldsymbol{\varphi}(t_k)e(t_k), \tag{9}$$

$$e(t_k) = y(t_k) - f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k), \tag{10}$$

$$\boldsymbol{\varphi}(t_k) = \mathrm{grad}[f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k)], \tag{11}$$

$$\mu(t_k) = \underset{\mu\geqslant 0}{\mathrm{argmin}} J[\hat{\boldsymbol{\theta}}(t_{k-1}) + \mu\boldsymbol{\varphi}(t_k)e(t_k))], \text{ or} \tag{12}$$

$$\mu(t_k) = \frac{\|\boldsymbol{\varphi}(t_k)\|^2}{\rho(t_k)} = \frac{\|\boldsymbol{\varphi}(t_k)\|^2}{\boldsymbol{\varphi}^{\mathrm{T}}(t_k)\boldsymbol{H}(t_k)\boldsymbol{\varphi}(t_k)}. \tag{13}$$

The steps of the GR algorithm are as follows: Firstly, let $k = 0$ and give $\hat{\boldsymbol{\theta}}(t_0)$. Pre-set a recursion length L. Secondly, gather the output data $y(t_k)$. Thirdly, compute the innovation $e(t_k)$ using (10) and the information vector $\boldsymbol{\varphi}(t_k)$ using (11). Fourthly, compute the optimal step-size $\mu(t_k)$ according to (12). Finally, update the parameter estimation vector $\hat{\boldsymbol{\theta}}(t_k)$, if $k = L$, terminate this recursive process; otherwise let $k := k+1$ and repeat this recursive process.

In order to provide a convenient method to determine the step-size $\mu(t_k)$, referring to [52], we have the following stochastic gradient (SG) parameter estimation algorithm:

$$\hat{\boldsymbol{\theta}}(t_k) = \hat{\boldsymbol{\theta}}(t_{k-1}) + \boldsymbol{\varphi}(t_k)e(t_k)/r(t_k), \tag{14}$$

$$\boldsymbol{\varphi}(t_k) = \frac{\partial f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k)}{\partial\boldsymbol{\theta}}, \tag{15}$$

$$e(t_k) = y(t_k) - f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k), \tag{16}$$

$$r(t_k) = r(t_{k-1}) + \|\boldsymbol{\varphi}(t_k)\|^2, \quad r(t_0) = 1. \tag{17}$$

Remark 3: Compared the SG algorithm in (14)-(17) with the GR algorithm in (9)-(13), their difference lies in the step-sizes. For the SG algorithm, the determination of the step-size is based on the information vector and does not need the one-dimensional optimization. So the SG method is more convenient than the GR method.

4. THE MULTI-INNOVATION GRADIENT ALGORITHM BASED ON THE MOVING DATA WINDOW

This section applies the multi-innovation identification theory to study the parameter estimation for nonlinear models. In linear cases, the multi-innovation algorithms have been developed for a linear output-error system [15] and a linear regression model [53]. The method here is for nonlinear models (e.g., the output of dynamical systems is nonlinear about the systems' parameters) and completely differs from the previous work and is the extension from linear cases in [15,53] to nonlinear cases.

For convenience, assume that $t = t_k$ is the current time. For the GR algorithm or the SG algorithm, the current parameter estimate $\hat{\boldsymbol{\theta}}(t_k)$ is obtained by the previous estimate $\hat{\boldsymbol{\theta}}(t_{k-1})$ plus a modification term $\frac{\boldsymbol{\varphi}(t_k)}{r(t_k)}e(t_k)$ or $\mu(t_k)\boldsymbol{\varphi}(t_k)e(t_k)$. The modification term is composed of the product of the innovation $e(t_k)$ and the gain vector $\frac{\boldsymbol{\varphi}(t_k)}{r(t_k)}$ or $\mu(t_k)\boldsymbol{\varphi}(t_k)$. The innovation is the useful information which can improve the estimation accuracy. In the SG algorithm and the GR algorithm, the innovation is a scalar and only the current data $y(t_k)$ is used to generate new parameter estimate. These two algorithms use only less measurement information at each recursion, so the estimation accuracy is relatively low.

In order to make full use of the observation data, the dynamical moving window data are presented to participate the estimation operation. Define the moving window data length p. Suppose that the current observation is $y(t_k)$. The moving window data can be represented as $\{y(t_k), y(t_{k-1}), \cdots, y(t_{k-p+1})\}$. In the SG algorithm, the innovation is denoted as $e(t_k) = y(t_k) - f(\hat{\boldsymbol{\theta}}(t_{k-1}),k) \in \mathbb{R}$. It is obvious that only the current data $y(t_k)$ is used for computing the scalar innovation. When the moving window data are used in the presented algorithm, the scalar innovation should be expanded into the innovation vector and the information vector $\boldsymbol{\varphi}(t_k)$ should be expanded into the information matrix $\boldsymbol{\Phi}(p,t_k)$.

Let the integral number p denote the innovation length. Based on the SG algorithm in (14)-(17), expand the information vector $\varphi(t_k)$ to the stacked information matrix

$$\boldsymbol{\Phi}(p,t_k) := [\varphi(t_k), \varphi(t_{k-1}), \cdots, \varphi(t_{k-p+1})] \in \mathbb{R}^{n_0 \times p}.$$

Expand the scalar innovation $e(t_k)$ into the innovation vector

$$\boldsymbol{E}(p,t_k) := \begin{bmatrix} y(t_k) - f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k) \\ y(t_{k-1}) - f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_{k-1}) \\ \vdots \\ y(t_{k-p+1}) - f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_{k-p+1}) \end{bmatrix} \in \mathbb{R}^p.$$

Finally, we can obtain the multi-innovation stochastic gradient (MISG) parameter estimation algorithm

$$\hat{\boldsymbol{\theta}}(t_k) = \hat{\boldsymbol{\theta}}(t_{k-1}) - \boldsymbol{\Phi}(p,t_k)\boldsymbol{E}(p,t_k)/r(t_k), \tag{18}$$
$$\boldsymbol{E}(p,t_k) = \boldsymbol{Y}(p,t_k) - \boldsymbol{F}(p,t_k), \tag{19}$$
$$\boldsymbol{Y}(p,t_k) = [y(t_k), y(t_{k-1}),\ \cdots,\ y(t_{k-p+1})]^{\mathrm{T}}, \tag{20}$$
$$\boldsymbol{\Phi}(p,t_k) = [\varphi(t_k), \varphi(t_{k-1}),\ \cdots,\ \varphi(t_{k-p+1})], \tag{21}$$
$$\boldsymbol{F}(p,t_k) = [f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k),\ f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_{k-1}),\ \cdots,\ f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_{k-p+1})]^{\mathrm{T}}, \tag{22}$$
$$r(t_k) = r(t_{k-1}) + \|\varphi(t_k)\|^2,\ r(t_0) = 1, \tag{23}$$
$$\varphi(t_{k-i}) = \frac{\partial f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_{k-i})}{\partial \boldsymbol{\theta}},\ i = 0,\ 1,\ \cdots,\ p-1. \tag{24}$$

The GR algorithm and the MISG algorithm are of the recursive form and can be applied to online identification in real time. The steps of computing the parameter estimates $\hat{\boldsymbol{\theta}}(t_k)$ using the MISG algorithm in (18)-(24) are as follows:

1) Let $k = 1$, set the innovation length p and the error $\varepsilon > 0$, and let $\hat{\boldsymbol{\theta}}(t_0)$ be a real vector.
2) Collect the output data $y(t_k)$, compute the information vector $\varphi(t_{k-i})$ using (24) and form the stacked information matrix $\boldsymbol{\Phi}(p,t_k)$ using (21).
3) Form the stacked observation output vector $\boldsymbol{Y}(p,t_k)$ using (20), the stacked nonlinear function vector $\boldsymbol{F}(p,t_k)$ using (22) and the innovation vector $\boldsymbol{E}(p,t_k)$ using (19).
4) Compute $r(t_k)$ using (23). Update the parameter estimate $\hat{\boldsymbol{\theta}}(t_k)$ using (18).
5) If $\|\hat{\boldsymbol{\theta}}(t_k) - \hat{\boldsymbol{\theta}}(t_{k-1})\| < \varepsilon$, then terminate this recursive process and obtain the parameter estimate; otherwise, let $k := k+1$ and go to Step 2.

Remark 4: In the MISG algorithm in (18)-(24), the stacked observation output vector $\boldsymbol{Y}(p,t_k)$ in (20) contains $y(t_k)$, $y(t_{k-1})$, $\cdots$, $y(t_{t-p+1})$ in the moving data window and thus the estimation accuracy can be improved, also see the explanation for linear models in the 2nd bullet on Page 3 in [53].

5. THE GRADIENT ALGORITHM BASED ON THE DYNAMICAL LENGTH DATA

For the GR algorithm, the SG algorithm and the MISG algorithm, there are two different ways for using the output data. One is using the current data; the other is using the moving window data. In order to use more observational data to enhance the estimation accuracy, in this section a dynamical length data method is proposed to construct the criterion function to deduce the parameter estimation algorithm. In each recursion, all of the data from the first sampling time to the current sampling time are used for generating the parameter estimates. Therefore, the collected data length increases with the increasing of the sampling time and the data length varies dynamically. Define the criterion function based on the increasing length data

$$J(\boldsymbol{\theta},t_k) := J(\boldsymbol{\theta},t_{k-1}) + \frac{1}{2}[y(t_k) - f(\boldsymbol{\theta},t_k)]^2.$$

Define the stacked output vector $\boldsymbol{Y}(t_k)$, the nonlinear function vector $\boldsymbol{F}(t_k)$ and the information matrix $\boldsymbol{\Phi}(t_k)$ as

$$\boldsymbol{Y}(t_k) := \begin{bmatrix} y(t_1) \\ y(t_2) \\ \vdots \\ y(t_k) \end{bmatrix} \in \mathbb{R}^k,$$
$$\boldsymbol{F}(t_k) := \begin{bmatrix} f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_1) \\ f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_2) \\ \vdots \\ f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k) \end{bmatrix} \in \mathbb{R}^k,$$
$$\boldsymbol{\Phi}(t_k) := [\varphi(t_1),\ \varphi(t_2),\ \cdots,\ \varphi(t_k)] \in \mathbb{R}^{n_0 \times k},$$

whose dimensions increase with the data length k increasing. Define

$$\begin{aligned}\boldsymbol{\xi}(t_k) &:= \boldsymbol{\Phi}(t_k)\boldsymbol{Y}(t_k) \\ &= \boldsymbol{\xi}(t_{k-1}) + \varphi(t_k)y(t_k) \in \mathbb{R}^{n_0}, \\ \boldsymbol{Q}(t_k) &:= \boldsymbol{\Phi}(t_k)\boldsymbol{F}(t_k) \\ &= \boldsymbol{Q}(t_{k-1}) + \varphi(t_k)f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k) \in \mathbb{R}^{n_0}.\end{aligned}$$

Minimizing the criterion function $J(\boldsymbol{\theta})$ and applying the negative gradient search deduce the stochastic gradient algorithm based on the dynamical length data (the DL-SG algorithm for short)

$$\hat{\boldsymbol{\theta}}(t_k) = \hat{\boldsymbol{\theta}}(t_{k-1}) + [\boldsymbol{\xi}(t_k) - \boldsymbol{Q}(t_k)]/r(t_k), \tag{25}$$
$$r(t_k) = r(t_{k-1}) + \|\varphi(t_k)\|^2,\ r(t_0) = 1, \tag{26}$$
$$\boldsymbol{\xi}(t_k) = \boldsymbol{\xi}(t_{k-1}) + \varphi(t_k)y(t_k),\ \boldsymbol{\xi}(t_0) = \mathbf{0}, \tag{27}$$
$$\boldsymbol{Q}(t_k) = \boldsymbol{Q}(t_{k-1}) + \varphi(t_k)f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k), \boldsymbol{Q}(t_0) = \mathbf{0}, \tag{28}$$
$$\varphi(t_k) = \frac{\partial f(\hat{\boldsymbol{\theta}}(t_{k-1}),t_k)}{\partial \boldsymbol{\theta}}. \tag{29}$$

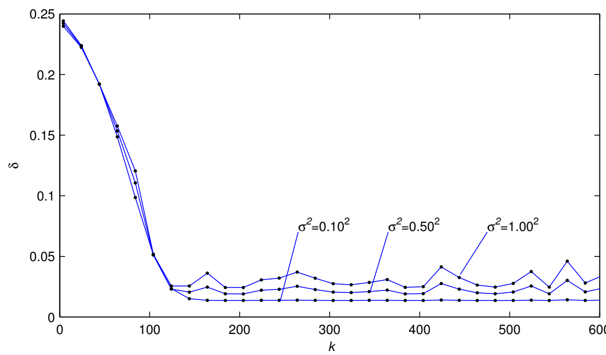

Fig. 1. The GR estimation error δ versus k.

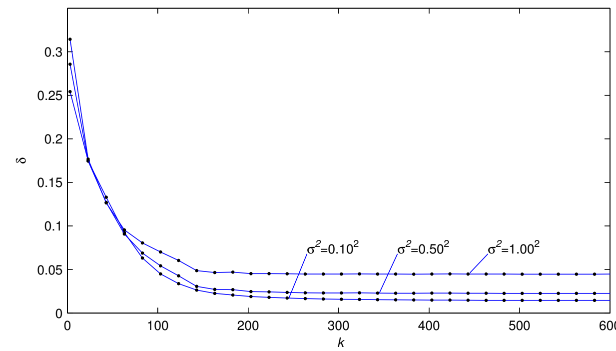

Fig. 2. The SG estimation errors δ versus k.

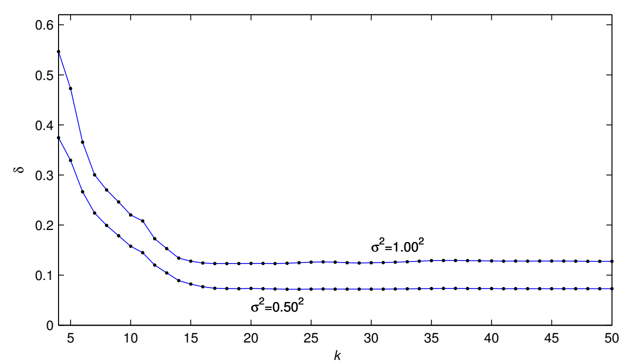

Fig. 3. The MISG estimation errors δ versus k.

The proposed approaches in this paper can combine some mathematical tools and strategies [54-59] to study new parameter estimation algorithms [60-64] and can be applied to other fields [65-70] such as information processing. The steps of computing the parameter estimates $\hat{\boldsymbol{\theta}}(t_k)$ using the DL-SG algorithm in (25)-(29) are as follows:

1) Let $k=0$, set $\hat{\boldsymbol{\theta}}(t_0)$ and a small number $\varepsilon > 0$.
2) Collect the output data $y(t_k)$. Compute $\boldsymbol{\varphi}(t_k)$ using (29). Compute $\boldsymbol{Q}(t_k)$ using (28) and $\boldsymbol{\xi}(t_k)$ using (27).
3) Compute $r(t_k)$ using (26). Update $\hat{\boldsymbol{\theta}}(t_k)$ using (25).
4) If $\|\hat{\boldsymbol{\theta}}(t_k)-\hat{\boldsymbol{\theta}}(t_{k-1})\| < \varepsilon$, then terminate this recursion process and obtain the parameter estimation vector $\hat{\boldsymbol{\theta}}(t_k)$; otherwise, let $k := k+1$ and go to Step 2.

Remark 5: The number of the sampled data used in the DL-SG algorithm increases gradually with the increasing of recursion k. It means that more observation information about the parameters to be estimated is utilized in the estimation method. Thus, the parameter estimation accuracy can be enhanced by using the dynamical data.

6. SIMULATION EXAMPLES

Here gives three simulation examples: one is the nonlinear function fitting from its observation data $y(t_k)$ and the other two are the transfer function fitting of the control systems from their output response data $y(t_k)$.

Example 1: Consider a nonlinear function

$$y(t) = 4t + 3\sin t + 2\mathrm{e}^{-3t}.$$

Here, $\theta_1 = 4$, $\theta_2 = 3$, $\theta_3 = 2$ and $\theta_4 = -3$ are the parameters to be identified.

In the simulation, the white noise with zero mean and variance $\sigma^2 = 0.10^2$, $\sigma^2 = 0.50^2$ and $\sigma^2 = 1.00^2$ is added to the response $y(t)$, respectively. Using the proposed GR method to estimate the system parameters, the parameter estimates and the estimation errors are illustrated in Table 1 and the parameter estimation errors $\delta := \|\hat{\boldsymbol{\theta}}(t_k) - \boldsymbol{\theta}\|/\|\boldsymbol{\theta}\|$ versus k are shown in Fig. 1.

In the same condition, the proposed SG algorithm is used to estimate the parameters. The results are shown in Table 2 and Fig. 2.

Example 2: Consider the system impulse response

$$y(t) = 4.8\mathrm{e}^{-1.15t} + 2.7\mathrm{e}^{-1.4t}.$$

Here, the parameters to be identified are $\theta_1 = 4.8$, $\theta_2 = 1.15$, $\theta_3 = 2.7$, $\theta_4 = 1.4$. In the simulation, take the innovation length $p = 3$. The white noise with the variance $\sigma^2 = 0.50^2$ and $\sigma^2 = 1.00^2$ is added to the system outputs respectively to generate the noisy output data. Using the presented MISG estimation algorithm, the results are shown in Table 3 and Fig. 3.

Example 3: Consider the system impulse response

$$y(t) = 3\mathrm{e}^{-2.9t} + 6\mathrm{e}^{-1t}.$$

Here, the parameters to be identified are $\theta_1 = 3$, $\theta_2 = 2.9$, $\theta_3 = 6$, $\theta_4 = 1$. In the simulation experiment, the white noise is added to the system responses and the noise variance is $\sigma^2 = 0.10^2$ and $\sigma^2 = 0.50^2$, respectively. Utilizing the proposed DL-SG method to estimate the system parameters, the results are shown in Table 4 and Fig. 4.

Example 4: Suppose that a typical closed loop system with bounded oscillation is shown in Fig. 5, where the damping ratio is ξ, the undamped natural frequency is ω_n with unit rad/s and $u(t)$ is a step input with the amplitude

Table 1. The GR estimates and their estimation errors.

σ^2	k	θ_1	θ_2	θ_3	θ_4	δ (%)
0.10^2	5	4.99873	3.89873	2.46155	-2.60287	23.92924
	15	4.98645	3.88648	2.37394	-2.62491	23.16683
	100	4.26960	3.22633	2.05922	-2.91594	5.94888
	200	3.98595	3.00836	2.05132	-2.93436	1.37736
	600	3.99235	3.01200	2.05133	-2.93434	1.37157
0.50^2	5	5.00028	3.90028	2.51147	-2.59941	24.24182
	15	4.98733	3.88736	2.42076	-2.62298	23.40388
	100	4.26871	3.22851	2.10798	-2.91566	6.13867
	200	3.94952	2.98358	2.09900	-2.93690	2.09005
	600	3.95065	2.99874	2.09903	-2.93678	2.06730
1.00^2	5	5.00151	3.90151	2.55142	-2.59659	24.50864
	15	4.98803	3.88807	2.45821	-2.62145	23.60935
	100	4.26802	3.23029	2.14679	-2.91582	6.35580
	200	3.92039	2.96377	2.13693	-2.93940	2.81315
	600	3.91728	2.98814	2.13697	-2.93921	2.78341
True values		4.00000	3.00000	2.00000	-3.00000	

Table 2. The SG estimates and their estimation errors.

σ^2	k	θ_1	θ_2	θ_3	θ_4	δ (%)
0.10^2	5	4.94097	3.74098	2.78167	-3.37137	23.97047
	15	4.87800	3.67814	2.31528	-3.25892	19.17505
	60	4.52101	3.33170	1.96379	-3.02925	10.04776
	100	4.26598	3.10210	1.92813	-2.97862	4.77904
	600	4.00268	2.96244	1.92459	-2.97049	1.44878
0.50^2	5	4.93899	3.73900	3.04700	-3.36179	26.43231
	15	4.84179	3.64205	2.43730	-3.18130	18.81308
	60	4.51300	3.32375	2.12263	-2.95909	10.06159
	100	4.29701	3.12895	2.09000	-2.90931	5.64677
	200	4.07467	2.97086	2.08584	-2.89866	2.51644
	600	4.01096	2.95964	2.08582	-2.89855	2.25985
1.00^2	5	4.93741	3.73742	3.25954	-3.35100	28.70988
	15	4.81318	3.61352	2.53504	-3.10872	18.74872
	60	4.50490	3.31570	2.24697	-2.89179	10.60398
	100	4.31956	3.14821	2.21680	-2.84338	7.17483
	600	4.01748	2.95667	2.21199	-2.82997	4.47317
True values		4.00000	3.00000	2.00000	-3.00000	

a. When $\xi = 0$, the step response with bounded oscillation is $y(t) = a(1 - \cos\omega_n t)$. In this example, the proposed SG, MISG and DL-SG algorithms are employed to identify two unknown parameters a and ω_n.

1) SG and MISG test and comparison

In this case, the SG method and the MISG method are used to compare the performance for the system response with bounded oscillation. In the simulation, the innovation length is set $p = 1$, $p = 5$, $p = 10$ and $p = 30$. In the simulation, the white noise with variance $\sigma^2 = 0.60^2$ is applied to the observations and the sample period is $h = 0.01$ second. The parameter estimates and estimation errors versus recursion k are illustrated in Table 5. The parameter estimation errors versus k obtained by setting different innovation length p are shown in Fig. 6.

2) DL-SG test

In this case, we test the accuracy with different data length. The data length is set as 1000, 2000, 3000, 4000, 6000, 8000, 10000 and 12000. The parameter estimates and estimation errors are illustrated in Table 6 and the estimation errors obtained by different data length are shown in Fig. 7.

3) Performance comparison among tree methods

In this case, the SG method, the MISG method and the DL-SG method are employed to compare the estimation accuracy. In the simulation, the data length of recursion is

Table 3. The MISG estimates and their estimation errors.

σ^2	k	θ_1	θ_2	θ_3	θ_4	δ (%)
0.50^2	5	4.86538	2.78340	2.22448	2.26227	32.91651
	10	5.06735	1.87068	2.41124	1.80066	15.75908
	20	5.14550	1.21630	2.47883	1.49083	7.33680
	40	5.15522	1.06590	2.48626	1.42807	7.31225
	50	5.15495	1.07325	2.48610	1.43057	7.28873
1.00^2	5	4.26188	3.32624	1.63039	2.55880	47.27613
	10	4.56242	1.99446	1.90871	1.88415	22.02198
	20	4.66916	1.10990	2.00115	1.46419	12.33274
	40	4.68291	0.89426	2.01164	1.37452	12.83357
	50	4.68233	0.91008	2.01128	1.37990	12.74691
True values		4.80000	1.15000	2.70000	1.40000	

Table 4. The DL-SG estimates and their estimation errors of Example 3.

σ^2	k	θ_1	θ_2	θ_3	θ_4	δ (%)
0.10^2	1	4.94845	3.36033	6.46823	1.79333	17.86553
	10	4.63137	3.13475	6.22298	1.10856	9.15977
	20	4.42424	3.12052	6.20751	1.06568	6.70377
	50	4.30536	3.11715	6.20386	1.05554	5.48928
	60	4.29266	3.11697	6.20366	1.05499	5.37213
0.50^2	1	4.97021	3.39850	6.50714	1.84905	18.78879
	10	4.73803	3.23383	6.32298	1.22863	11.50408
	20	4.55310	3.22412	6.31213	1.19236	9.42411
	50	4.45480	3.22203	6.30978	1.18452	8.46997
	60	4.44414	3.22191	6.30965	1.18407	8.37423
True values		4.00000	2.90000	6.00000	1.00000	

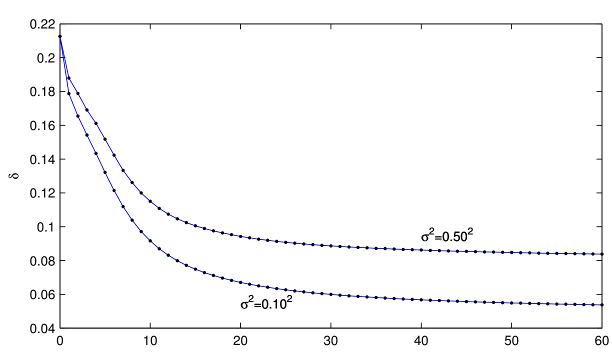

Fig. 4. The DL-SG estimation errors δ versus k.

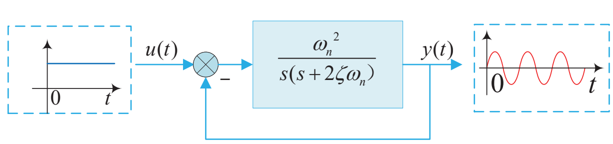

Fig. 5. The closed loop system with bounded oscillation.

Table 5. The SG and MISG estimates and their estimation errors of Example 4.

p	k	a	ω_n	δ (%)
$p=1$ (SG)	5	0.24999	0.84999	26.07798
	100	0.28989	0.87287	21.96530
	500	0.41099	0.92240	10.56167
	1000	0.41511	0.93376	9.63088
	2000	0.41393	0.94331	9.21836
$p=5$ (MISG)	5	0.24995	0.84997	26.08224
	100	0.35215	0.90866	15.54377
	500	0.47641	0.95977	4.17112
	1000	0.48061	0.97124	3.10229
	2000	0.47933	0.98072	2.52794
$p=10$ (MISG)	5	0.25142	0.85084	25.92923
	100	0.38215	0.92598	12.44763
	500	0.50701	0.97740	2.11654
	1000	0.51129	0.98895	1.41272
	2000	0.50993	0.99836	0.90016
True values		0.50000	1.00000	

6000. The noise variance is $\sigma^2 = 0.50^2$. The innovation is $p=1$, $p=3$, $p=5$ and $p=10$. It is noted that the innovation setting is used for the MISG method. When $p=1$ the MISG method becomes the SG method. The parameter estimation errors versus k with different p obtained by the MISG method and DL-SG method are demonstrated

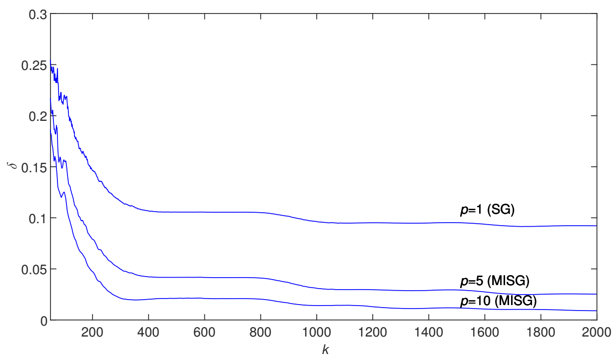

Fig. 6. The SG and MISG estimation errors δ versus k of Example 4.

Table 6. The DL-SG estimates and their estimation errors of Example 4.

k	a	ω_n	δ (%)
1000	0.40671	0.92612	10.64380
2000	0.45661	0.96540	4.96361
6000	0.49696	0.97742	2.03828
10000	0.50510	0.97991	1.85386
12000	0.50712	0.98069	1.84086
True values	0.50000	1.00000	

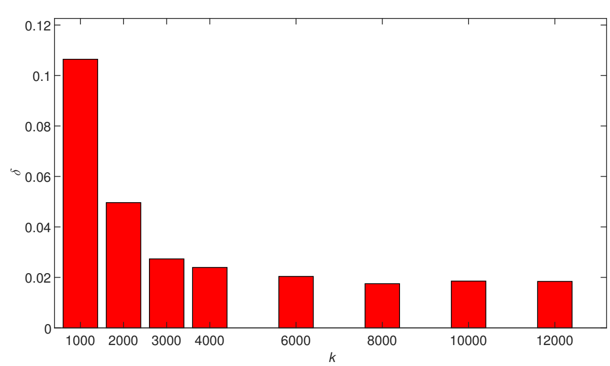

Fig. 7. The DL-SG estimation errors δ versus different data length of Example 4.

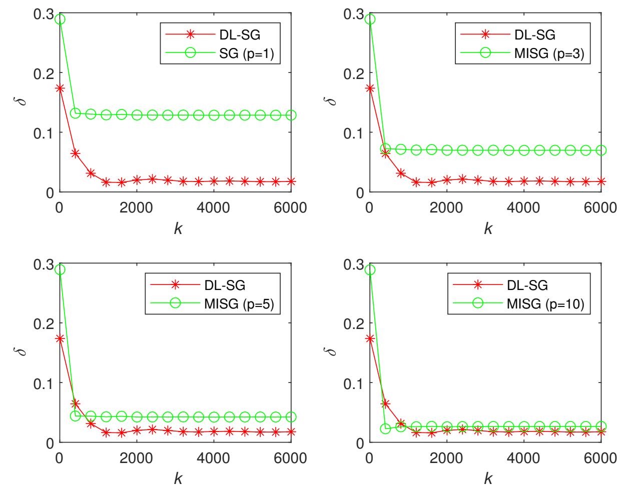

Fig. 8. The SG, MISG and DL-SG estimation errors δ versus k of Example 4.

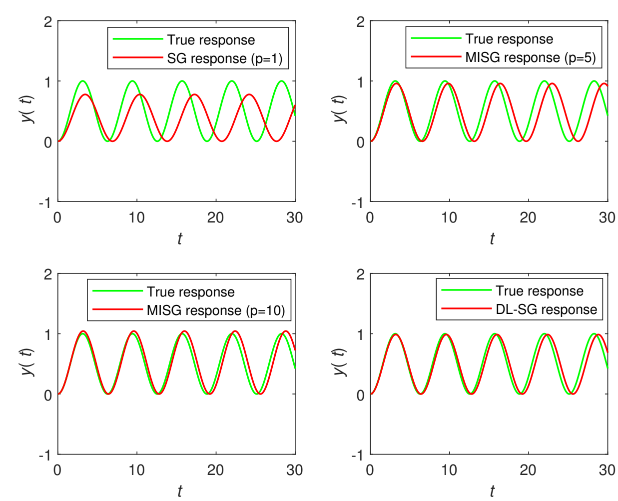

Fig. 9. The responses of the system obtained by SG, MISG and DL-SG of Example 4.

in Fig. 8.

Moreover, the system responses obtained by the SG, MISG and DL-SG methods are compared and the response curves are shown in Fig. 9.

From the numerical simulation results, we can draw the following conclusions. The parameter estimation errors obtained by the GR algorithm and the SG algorithm become smaller with the increasing of the recursion variable k under relatively low noise levels. It shows the effectiveness of the GR method and the SG method. The parameter estimation accuracy is associated with the noise variance. If the noise variance is large, the estimation accuracy is low (see Figs. 1 and 2).

Comparing the GR method with the SG method, the estimation error curves (see Fig. 2) given by the SG method are smooth while the estimation error curves (see Fig. 1) given by the GR method are not smooth for large noise variances. This simulation results demonstrate that the SG method is not sensitive to the noise and has more robust than the GR method.

The parameter estimates given by the MISG method and DL-SG method are close to the true values with the increase of the recursion k. If the noise variance is small, then the parameter estimation errors are small. Moreover, the estimation error versus k in Figs. 3 and 4) indicates that the estimation error variations gradually decrease with the increasing of recursion k. This means that the MISG method and the DL-SG method are effective.

From the performance comparison results among the

SG method, the MISG method and the DL-SG method in Fig. 8, we can see that with the increasing of the recursion k, the parameter estimation errors become smaller gradually and the parameter estimation accuracy of the DL-SG method is the best one compared with other two methods. In terms of the DL-SG method, more observations are joined into the recursion computing with the increasing of k than the SG and MISG methods. Therefore, the DL-SG has better performance and can be used for on-line identification.

From Fig. 9, we can see that the system response obtained by the DL-SG method is very close to the true systems and the response obtained by the SG method has larger errors. Moreover the system response obtained by the MISG method when $p = 10$ has good tracing performance. Therefore, the proposed MISG method and the DL-SG method are effective for identifying the systems with bounded oscillation.

7. CONCLUSIONS

In this paper, the problem of parameter estimation is considered from the system response using the dynamical optimization scheme. The GR estimation method is derived according to the nonlinear gradient search principle. On this basis, the SG method, the MISG method and the DL-SG method are presented using different dynamical observation data. The mathematical simulation results indicate that the parameter estimation accuracy given by the proposed methods are acceptable. The proposed estimation algorithms in this paper can integrate some adaptive estimation algorithms to explore identification methods for various systems [71-76] and can be applied to Control and schedule areas [77-85] such as signal processing, system modeling [76-92] and transportation systems.

CONFLICT OF INTERESTS

The author declares that there is no conflict of interests regarding the publication of this paper.

REFERENCES

[1] J. Pan, X. Jiang, X. Wan, and W. Ding, "A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems," *International Journal of Control, Automation, and Systems*, vol. 15, no. 3, pp. 1189-1197, May 2017.

[2] Y. Ji, C. Zhang, Z. Kang, and T. Yu, "Parameter estimation for block-oriented nonlinear systems using the key term separation," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 9, pp. 3727-3752, 2020.

[3] Y. Fan and X. Liu, "Two-stage auxiliary model gradient-based iterative algorithm for the input nonlinear controlled autoregressive system with variable-gain nonlinearity," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 14, pp. 5492-5509, July 2020.

[4] Y. Ji and Z. Kang, "Three-stage forgetting factor stochastic gradient parameter estimation methods for a class of nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 3, pp. 971-987, 2021.

[5] Y. M. Fan and X. M. Liu, "Auxiliary model-based multi-innovation recursive identification algorithms for an input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity," *International Journal of Adaptive Control and Signal Processing*, vol. 36, no. 3, pp. 521-540, March 2022.

[6] Y. Ji, Z. Kang, and C. Zhang, "Two-stage gradient-based recursive estimation for nonlinear models by using the data filtering," *International Journal of Control Automation and Systems*, vol. 19, no. 8, pp. 2706-2715, 2021.

[7] J. Pan, S.D. Liu, J. Shu, and X. K. Wan, "Hierarchical recursive least squares estimation algorithm for secondorder Volterra nonlinear systems," *International Journal of Control, Automation, and Systems*, vol. 20, no. 12, pp. 3940-3950, December 2022.

[8] J. Pan, Y.Q. Liu, and J. Shu, "Gradient-based parameter estimation for an exponential nonlinear autoregressive time-series model by using the multi-innovation," *International Journal of Control, Automation, and Systems*, vol. 21, no. 1, pp. 140-150, 2023.

[9] F. Ding and T. Chen, "Combined parameter and output estimation of dual-rate systems using an auxiliary model," *Automatica*, vol. 40, no. 10, pp. 1739-1748, 2004.

[10] Y. Gu, Q. Zhu, and H. Nouri, "Identification and U-control of a state-space system with time-delay," *International Journal of Adaptive Control and Signal Processing*, vol. 36, no. 1, pp. 138-154, January 2022.

[11] H. Liu, J. Wang, and Y. Ji, "Maximum likelihood recursive generalized extended least squares estimation methods for a bilinear-parameter systems with ARMA noise based on the over-parameterization model," *International Journal of Control, Automation, and Systems*, vol. 20, no. 8, pp. 2606-2615, August 2022.

[12] X. Meng, Y. Ji, and J. Wang, "Iterative parameter estimation for photovoltaic cell models by using the hierarchical principle," *International Journal of Control, Automation, and Systems*, vol. 20, no. 8, pp. 2583-2593, August 2022.

[13] G. Chen, M. Gan, S. Wang, and C. Chen, "Insights into algorithms for separable nonlinear least squares problems," *IEEE Transactions on Image Processing*, vol. 30, pp. 1207-1218, 2021.

[14] J. Ren, J. Duan, and X. Wang, "A parameter estimation method based on random slow manifolds," *Applied Mathematical Modelling*, vol. 39, no. 13, pp. 3721-3732, 2015.

[15] F. Ding, G. Liu, and X. Liu. "Parameter estimation with scarce measurements," *Automatica*, vol. 47, no. 8, pp. 1646-1655, 2011.

[16] S. Srivastava and V. S. Pandit. "A PI/PID controller for time delay systems with desired closed loop time response and guaranteed gain and phase margins," *Journal of Process Control*, vol. 37, pp. 70-77, 2016.

[17] S. Ahmed, B. Huang, and S. L. Shah, "Novel identification method from step response," *Control Engineering Practice*, vol. 15, no. 5, pp. 545-556, 2007.

[18] P. Balaguer, V. Alfaro, and O. Arrieta, "Second order inverse response process identification from transient step response," *ISA Transactions*, vol. 50, no. 2, pp. 231-238, 2011.

[19] E. Hidayat and A. Medvedev, "Laguerre domain identification of continuous linear time-delay systems from impulse response data," *Automatica*, vol. 48, no. 11, pp. 2902-2907, 2012.

[20] L. D. Tommasi, D. Deschrijver, and T. Dhaene, "Transfer function identification from phase response data," *AEU-International Journal of Electronics and Communications*, vol. 64, no. 3, pp. 218-223, 2010.

[21] L. Xu, "Separable Newton recursive estimation method through system responses based on dynamically discrete measurements with increasing data length," *International Journal of Control, Automation, and Systems*, vol. 20, no. 2, pp. 432-443, February 2022.

[22] Y. Wang, S. Tang, and M. Deng, "Modeling nonlinear systems using the tensor network B-spline and the multi-innovation identification theory," *International Journal of Robust and Nonlinear Control*, vol. 32, no. 13, pp. 7304-7318, 2022.

[23] Y. Wang and L. Yang, "An efficient recursive identification algorithm for multilinear systems based on tensor decomposition," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 16, pp. 7920-7936, 2021.

[24] M. Kapetina, M. Rapaic, and A. Pisano, "Adaptive parameter estimation in LTI systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 10, pp. 4188-4195, 2019.

[25] Y. Ji, Z. Kang, and X. Liu, "The data filtering based multiple-stage Levenberg-Marquardt algorithm for Hammerstein nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 15, pp. 7007-7025, October 2021.

[26] H. Ma, J. Pan, and W. Ding, "Partially-coupled least squares based iterative parameter estimation for multivariable output-error-like autoregressive moving average systems," *IET Control Theory and Applications*, vol. 13, no. 18, pp. 3040-3051, December 2019.

[27] J. Wang, Y. Ji, and C. Zhang, "Iterative parameter and order identification for fractional-order nonlinear finite impulse response systems using the key term separation," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 8, pp. 1562-1577, 2021.

[28] J. Pan, H. Ma, and J. Sheng, "Recursive coupled projection algorithms for multivariable output-error-like systems with coloured noises," *IET Signal Processing*, vol. 14, no. 7, pp. 455-466, September 2020.

[29] J. Wang, Y. Ji, and X. Zhang, "Two-stage gradient-based iterative algorithms for the fractional-order nonlinear systems by using the hierarchical identification principle," *International Journal of Adaptive Control and Signal Processing*, vol. 36, no. 7, pp. 1778-1796, 2022.

[30] H. Ma, X. Zhang, and T. Hayat, "Partially-coupled gradient-based iterative algorithms for multivariable output-error-like systems with autoregressive moving average noises," *IET Control Theory and Applications*, vol. 14, no. 17, pp. 2613-2627, November 2020.

[31] E. Carvalho, J. Martinez, and F. Pisnitchenko. "On optimization strategies for parameter estimation in models governed by partial differential equations," *Mathematics and Computers in Simulation*, vol. 114, pp. 14-24, 2015.

[32] Q. Lin, R. Loxton, C. Xu, and K. L. Teo. "Parameter estimation for nonlinear time-delay systems with noisy output measurements," *Automatica*, vol. 60, pp. 48-56, 2015.

[33] A. J. Isaksson, J. Sjöberg, D. Törnqvist, L. Ljung, and M. Kok. "Using horizon estimation and nonlinear optimization for grey-box identification," *Journal of Process Control*, vol. 30, pp. 69-79, 2015.

[34] Y. Ji and A. N. Jiang, "Filtering-based accelerated estimation approach for generalized time-varying systems with disturbances and colored noises," *IEEE Transactions on Circuits and Systems-II: Express Briefs*, vol. 70, no. 1, pp. 206-210, January 2023.

[35] F. Ding and T. Chen, "Parameter estimation of dual-rate stochastic systems by using an output error method," *IEEE Transactions on Automatic Control*, vol. 50, no. 9, pp. 1436-1441, September 2005.

[36] M. Li and X. Liu, "Iterative identification methods for a class of bilinear systems by using the particle filtering technique," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 11, pp. 2056-2074, 2021.

[37] L. Xu, "Hierarchical recursive signal modeling for multifrequency signals based on discrete measured data," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 5, pp. 676-693, 2021.

[38] J. Chen, B. Huang, and C. Chen, "A novel reduced-order algorithm for rational models based on Arnoldi process and Krylov subspace," *Automatica*, vol. 129, 109663, July 2021.

[39] J. Ding, "Hierarchical least squares identification for linear SISO systems with dual-rate sampled-data," *IEEE Transactions on Automatic Control*, vol. 56, no. 11, pp. 2677-2683, November 2011.

[40] J. Hou, F. Chen, P. Li, and Z. Zhu, "Gray-box parsimonious subspace identification of Hammerstein-type systems," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 10, pp. 9941-9951, 2021.

[41] X. Zhang, "State estimation for bilinear systems through minimizing the covariance matrix of the state estimation errors," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 7, pp. 1157-1173, 2019.

[42] P. Ma and L. Wang, "Filtering-based recursive least squares estimation approaches for multivariate equation-error systems by using the multiinnovation theory," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 9, pp. 1898-1915, September 2021.

[43] X. Zhang, "Adaptive parameter estimation for a general dynamical system with unknown states," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 4, pp. 1351-1372, March 2020.

[44] M. Li and X. Liu, "Maximum likelihood hierarchical least squares-based iterative identification for dual-rate stochastic systems," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 2, pp. 240-261, 2021.

[45] X. Liu and Y. Fan, "Maximum likelihood extended gradient-based estimation algorithms for the input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 9, pp. 4017-4036, March 2021.

[46] F. Ding, Y. Liu, and B. Bao, "Gradient based and least squares based iterative estimation algorithms for multi-input multi-output systems," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 226, no. 1, pp. 43-55, 2012.

[47] Z. Kang, Y. Ji, and X. Liu, "Hierarchical recursive least squares algorithms for Hammerstein nonlinear autoregressive output-error systems," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 11, pp. 2276-2295, November 2021.

[48] J. Hou, H. Su, C. Yu, and P. Li, "Bias-correction errors-in-variables Hammerstein model identification," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 7, pp. 7268-7279, 2023.

[49] F. Ding, "Coupled-least-squares identification for multivariable systems," *IET Control Theory and Applications*, vol. 7, no. 1, pp. 68-79, January 2013.

[50] J. Hou, H. Su, C. Yu, and T. Li, "Consistent subspace identification of errors-in-variables Hammerstein systems," *IEEE Transactions on Systems Man and Cybernetics: Systems*, vol. 53, no. 4, pp. 2292-2303, 2023.

[51] C. Xu, Y. Qin, and H. Su, "Observer-based dynamic event-triggered bipartite consensus of discrete-time multi-agent systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 70, no. 3, pp. 1054-1058, 2023.

[52] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*, Prentice Hall, Englewood Cliffs, New Jersey, 1984.

[53] F. Ding and T. Chen. "Performance analysis of multi-innovation gradient type identification methods," *Automatica*, vol. 43, no. 1, pp. 1-14, 2007.

[54] F. Geng and X. Wu, "A novel kernel functions algorithm for solving impulsive boundary value problems," *Applied Mathematics Letters*, vol. 134, 108318, 2022.

[55] H. Wang, H. Fan, and J. Pan, "A true three-scroll chaotic attractor coined," *Discrete and Continuous Dynamical Systems-Series B*, vol. 27, no. 5, pp. 2891-2915, 2022.

[56] C. Yin and Y. Wen, "An extension of Paulsen-Gjessing's risk model with stochastic return on investments," *Insurance Mathematics & Economics*, vol. 52, no. 3, pp. 469-476, 2013.

[57] C. Yin and J. Zhao, "Nonexponential asymptotics for the solutions of renewal equations with applications," *Journal of Applied Probability*, vol. 43, no. 3, pp. 815-824, 2006.

[58] C. Yin and K. Yuen, "Optimality of the threshold dividend strategy for the compound Poisson model," *Statistics & Probability Letters*, vol. 81, no. 12, pp. 1841-1846, 2011.

[59] C. Yin and K. Yuen, 'Optimal dividend problems for a jump-diffusion model with capital injections and proportional transaction costs," *Journal of Industrial and Management Optimization*, vol. 11, no. 4, pp. 1247-1262, 2015.

[60] T. Cui, "Moving data window-based partially-coupled estimation approach for modeling a dynamical system involving unmeasurable states," *ISA Transactions*, vol. 128, pp. 437-452, 2022.

[61] C. Wei, "Overall recursive least squares and overall stochastic gradient algorithms and their convergence for feedback nonlinear controlled autoregressive systems," *International Journal of Robust and Nonlinear Control*, vol. 32, no. 9, pp. 5534-5554, 2022.

[62] C. Zhang, "Gradient parameter estimation of a class of nonlinear systems based on the maximum likelihood principle," *International Journal of Control, Automation, and Systems*, vol. 20, no. 5, pp. 1393-1404, 2022.

[63] J. M. Li, "A novel nonlinear optimization method for fitting a noisy Gaussian activation function," *International Journal of Adaptive Control and Signal Processing*, vol. 36, no. 3, pp. 690-707, March 2022.

[64] X. Zhang, "Hierarchical parameter and state estimation for bilinear systems,' *International Journal of Systems Science*, vol. 51, no. 2, 275-290, 2020.

[65] H. Wang, G. Ke, J. Pan, and Q. Su, "Modeling, dynamical analysis and numerical simulation of a new 3D cubic Lorenz-like system," *Scientific Reports*, vol. 13, Article number 6671, 2023.

[66] F. Ding, X.M. Liu, and H.B. Chen, "Hierarchical gradient based and hierarchical least squares based iterative parameter identification for CARARMA systems," *Signal Processing*, vol. 97, pp. 31-39, April 2014.

[67] N. Zhao, A. Wu, Y. Pei, and D. Niyato, "Spatial-temporal aggregation graph convolution network for efficient mobile cellular traffic prediction," *IEEE Communications Letters*, vol. 26, no. 3, pp. 587-591, 2022.

[68] Y. Chen, C. Zhang, C. Liu, Y. Wang, and X. Wan, "Atrial fibrillation detection using feedforward neural network," *Journal of Medical and Biological Engineering*, vol. 242, no. 1, pp. 63-73, February 2022.

[69] F. Ding, H. Yang, and F. Liu, "Performance analysis of stochastic gradient algorithms under weak conditions," *Science in China Series F - Information Sciences*, vol. 51, no. 9, pp. 1269-1280, 2008.

[70] H. Wang, G. Ke, and J. Pan, "Two pairs of heteroclinic orbits coined in a new sub-quadratic Lorenz-like system," *European Physical Journal B*, vol. 96, no. 3, p. 28, 2023.

[71] Y. Wang, "Recursive parameter estimation algorithm for multivariate output-error systems," *Journal of the Franklin Institute*, vol. 355, no. 12, pp. 5163-5181, 2018.

[72] J. Ding and W. Zhang, "Finite-time adaptive control for nonlinear systems with uncertain parameters based on the command filters," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 9, pp. 1754-1767, September 2021.

[73] F. Ding, G. Liu, and X. Liu, "Partially coupled stochastic gradient identification methods for non-uniformly sampled systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1976-1981, August 2010.

[74] J. Pan, W. Li, and H. Zhang, "Control algorithms of magnetic suspension systems based on the improved double exponential reaching law of sliding mode control," *International Journal of Control, Automation, and Systems*, vol. 16, no. 6, pp. 2878-2887, December 2018.

[75] J. Xiong, J. Pan, and G. Chen, "Sliding mode dual-channel disturbance rejection attitude control for a quadrotor," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 10, pp. 10489-10499, 2022.

[76] J. Pan, Q. Chen, J. Xiong, and G. Chen, "A novel quadruple boost nine level switched capacitor inverter," *Journal of Electrical Engineering & Technology*, vol. 18, no. 1, pp. 467-480, 2023.

[77] Y. Cao, Y. Yang, and J. Wen, "Research on virtual coupled train control method based on GPC & VAPF," *Chinese Journal of Electronics*, vol. 31, no. 5, pp. 897-905, 2022.

[78] Y. Cao, Y. Sun, G. Xie, and P. Li, "A sound-based fault diagnosis method for railway point machines based on two-stage feature selection strategy and ensemble classifier," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 12074-12083, 2022.

[79] Y. Cao, J. Wen, A. Hobiny, and T. Wen, "Parameter-varying artificial potential field control of virtual coupling system with nonlinear dynamics," *Fractals*, vol. 30, no. 2, 2240099, 2022.

[80] Y. Cao, L. Ma, S. Xiao, and W. Xu, "Standard analysis for transfer delay in CTCS-3," *Chinese Journal of Electronics*, vol. 26, no. 5, pp. 1057-1063, September 2017.

[81] Y. Cao, J. Wen, and L. Ma, "Tracking and collision avoidance of virtual coupling train control system," *Alexandria Engineering Journal*, vol. 60, no. 2, pp. 2115-2125, 2021.

[82] Y. Li, G. Yang, Z. Su, S. Li, and Y. Wang, "Human activity recognition based on multienvironment sensor data," *Information Fusion*, vol. 91, pp. 47-63, March 2023.

[83] Y. Wang and G. Yang, "Arrhythmia classification algorithm based on multi-head self-attention mechanism," *Biomedical Signal Processing and Control*, vol. 79, p. 104206, 2023.

[84] J. Lin, Y. Li, and G. C. Yang, "FPGAN: Face de-identification method with generative adversarial networks for social robots," *Neural Networks*, vol. 133, pp. 132-147, January 2021.

[85] G. C. Yang, Z. J. Chen, Y. Li, and Z. D. Su, "Rapid relocation method for mobile robot based on improved ORB-SLAM2 algorithm," *Remote Sensing*, vol. 11, no. 2, 149, 2019.

[86] F. Ding, "Least squares and multi-innovation least squares methods," *Journal of Computational and Applied Mathematics*, vol. 426, p. 115107, July 2023.

[87] F. Z. Geng and X. Y. Wu, "Reproducing kernel-based piecewise methods for efficiently solving oscillatory systems of second-order initial value problems," *Calcolo*, vol. 60, no. 2, p. 20, June 2023.

[88] X. Y. Li, and X. Y. Liu, "A hybrid kernel functions collocation approach for boundary value problems with Caputo fractional derivative," *Applied Mathematics Letters*, vol. 142, 108636, 2023.

[89] L. Xu, "Separable synthesis estimation methods and convergence analysis for multivariable systems, *Journal of Computational and Applied Mathematics*, vol. 427, p. 115104, August 2023.

[90] X. Y. Li and B. Y. Wu, "A kernel regression approach for identification of first order differential equations based on functional data," *Applied Mathematics Letters*, vol. 127, p. 107832, May 2022.

[91] F. Ding, "Filtered auxiliary model recursive generalized extended parameter estimation methods for Box-Jenkins systems for Box-Jenkins systems by means of the filtering identification idea," *International Journal of Robust and Nonlinear Control*, vol. 33, 2023. DOI: 10.1002/rnc.6657

[92] Z. Shi, H. Yang, and M. Dai, "The data-filtering based bias compensation recursive least squares identification for multi-input single-output systems with colored noises," *Journal of the Franklin Institute*, vol. 360, no. 7, pp. 4753-4783, 2023.

Ling Xu was born in Tianjin, China. She received her master's and Ph.D. degrees from the Jiangnan University (Wuxi, China), in 2005 and 2015, respectively. She was a Post-Doctoral Fellow at the Jiangnan University from 2016 to 2020 and is currently a Professor. She is a Colleges and Universities "Blue Project" Young Teacher (Jiangsu, China). Her research interests include process control, parameter estimation, and signal modeling.