




Hierarchical Gradient-Based Iterative Parameter Estimation Algorithms for a Nonlinear Feedback System Based on the Hierarchical Identification Principle

Dan Yang^{1,2} · Yanjun Liu¹ · Feng Ding¹  · Erfu Yang³

Received: 3 October 2022 / Revised: 26 July 2023 / Accepted: 27 July 2023 /

Published online: 17 August 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

This paper focuses on iterative parameter estimation methods for a nonlinear closed-loop system (i.e., a nonlinear feedback system) with an equation-error model for the open-loop part. By applying negative gradient search, a gradient-based iterative algorithm is constructed. To reduce the computational costs and improve the parameter estimation accuracy, the hierarchical identification principle is employed to derive a hierarchical gradient-based iterative algorithm. A simulation example is provided to test the effectiveness of the proposed algorithms.

Keywords Nonlinear feedback system · Iterative identification · Hierarchical identification · Gradient search · Parameter estimation

✉ Feng Ding
fding@jiangnan.edu.cn

Dan Yang
dyang618@126.com

Yanjun Liu
yjl@jiangnan.edu.cn

Erfu Yang
erfu.yang@strath.ac.uk

¹ Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, People's Republic of China

² Department of Electromechanical Engineering, Changzhou Vocational Institute of Textile and Garment, Changzhou 213164, People's Republic of China

³ Department of Design, Manufacturing and Engineering Management, University of Strathclyde, Glasgow G1 1XJ, UK

1 Introduction

System identification is the theory and method of establishing mathematical models of linear stochastic systems and nonlinear stochastic systems [2, 69, 79]. As most natural systems have nonlinear behaviors, the identification and modeling of nonlinear systems are meaningful [29–33]. Common nonlinear systems include Hammerstein systems, Wiener systems and Wiener-Hammerstein systems. Over the past several decades, a dozen studies have been conducted on identification methods for nonlinear systems. In [1], the estimating function approach was introduced, which can construct estimators that are asymptotically optimal with respect to a specific class of estimators for the identification of nonlinear stochastic dynamical models. Li et al. proposed a linear variable weight particle swarm method for Hammerstein-Wiener nonlinear systems with unknown time delay by utilizing the parallel search ability of the particle swarm optimization and the iterative identification technique [46].

Nonlinear systems can be constructed from dynamical linear subsystems and static nonlinear elements in many different ways, including series, parallel and feedback connections [23, 35, 58]. Here, we focus on a class of nonlinear feedback systems whose forward channels are dynamical linear subsystems and feedback channels are static nonlinearities. Recently, many publications have been devoted to studying the estimation and application of feedback nonlinear systems [20, 21]. For example, Wei et al. derived an overall recursive least squares algorithm and an overall stochastic gradient algorithm to estimate nonlinear feedback systems disturbed by white noise [93]. Based on filtering technology, a linear filter was constructed to convert the colored noise into white noise and a filtering-based multi-innovation stochastic gradient algorithm was obtained deduced for Hammerstein equation-error autoregressive systems [70] by employing the multi-innovation theory [47–51].

In the area of parameter estimation, the iterative technique is an effective tool that has been widely applied to optimize the parameters of the estimated models [41–43]. In contrast to recursive algorithms, iterative algorithms use batch data at each iteration and make full use of measurement data, which can improve parameter estimation accuracy [38–40]. By combining different search principles, various iterative algorithms can be formed, such as the least squares-based iterative algorithm, the gradient-based iterative algorithm and the Newton iteration algorithm. Several gradient-based iterative algorithms have been studied in previous works [80, 110]. For instance, Fan et al. presented a two-stage auxiliary model gradient-based iterative algorithm for an input nonlinear controlled autoregressive system with variable-gain nonlinearity [18]. Iterative identification is a class of offline identification methods. To track time-varying parameters in real time, Jiang et al. presented a moving data window gradient-based iterative algorithm for a generalized time-varying system with a measurable disturbance vector [36].

To improve the parameter estimation results, hierarchical identification methods were proposed based on the decomposition-coordination principle of large-scale systems [37, 68]. The basic idea is to reduce the scales of the estimated models through decomposition and to perform joint estimation for the models. In [96–100, 102], a separable multi-innovation Newton iterative signal modeling method was derived and implemented to estimate sine multi-frequency signals and periodic signals, in which

the measurements are sampled and used dynamically. Wang et al. proposed a two-stage gradient-based iterative algorithm by minimizing two criterion functions for a fractional-order nonlinear system with autoregressive noise [83]. Hu et al. presented a two-stage recursive least squares algorithm for nonlinear feedback systems based on the auxiliary model identification idea and the hierarchical identification principle [27].

The parameter estimation of system models is important for controller design and filter design [112–116]. Some signal processing filters are designed based on parameter estimation algorithms and some self-tuning control and model-based predicted control methods rely on parameter estimation methods [117–121]. The previous works in [104] discussed the multi-innovation gradient-based iterative identification method and the decomposition multi-innovation gradient-based iterative identification methods for nonlinear feedback systems by using the decomposition technique. This paper studies the gradient-based iterative estimation algorithm and the hierarchical gradient-based iterative estimation algorithm to solve the parameter estimation problems of feedback nonlinear systems. The main contributions of this paper are as follows.

- To make full use of measurement data to estimate the parameters of the nonlinear feedback systems, we utilize the gradient search principle as the optimization strategy and present a gradient-based iterative (GI) algorithm.
- The system to be identified can be decomposed into two subsystems based on the hierarchical identification principle. Then, a hierarchical gradient-based iterative (H-GI) algorithm is derived to improve the parameter estimation accuracy.
- The performance of the proposed algorithms is tested by a numerical example, including the computational costs and the parameter estimation accuracy.

In summary, the rest of this paper is organized as follows. Section 2 describes the identification problem for nonlinear feedback systems. A gradient-based iterative algorithm is derived in Sect. 3. Section 4 presents a hierarchical gradient-based iterative algorithm and analyzes the computational costs of the algorithms. Section 5 provides a numerical example for testing the effectiveness of the proposed algorithms. Finally, some concluding remarks are given in Sect. 6.

This paper studies two iterative estimation algorithms to solve parameter estimation problems for nonlinear feedback systems. The proposed iterative identification algorithms for the nonlinear feedback systems can be combined with other techniques and strategies to address parameter estimation problems for different systems and can be applied to other fields, such as information processing and communication.

2 System Description

Let us introduce some symbols used in this paper. The symbol \mathbf{I}_n stands for an identity matrix of order n ; $\mathbf{1}_n$ stands for an n -dimensional column vector whose elements are 1; the superscript T stands for vector/matrix transpose; and the norm of a matrix (or a column vector) \mathbf{X} is defined by $\|\mathbf{X}\|^2 := \text{tr}[\mathbf{X}\mathbf{X}^T]$.

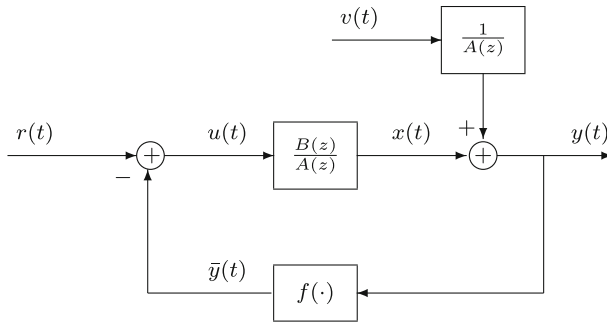


Fig. 1 The nonlinear feedback CAR system

Consider the nonlinear feedback system shown in Fig. 1, where the forward channel is described by a controlled autoregressive (CAR) model:

$$\begin{aligned}
 y(t) + a_1y(t - 1) + a_2y(t - 2) + \dots + a_{n_a}y(t - n_a) \\
 = b_1u(t - 1) + b_2u(t - 2) + \dots + b_{n_b}u(t - n_b) + v(t),
 \end{aligned}
 \tag{1}$$

$y(t)$ is the output of the system, $u(t)$ is the input of the linear part, $v(t)$ is a white noise sequence with zero mean and variance σ^2 , and $y(t) = 0$ and $u(t) = 0$ for $t \leq 0$. The feedback channel of the system is a memoryless nonlinear block whose output $\bar{y}(t)$ can be regarded as a nonlinear function of a known nonlinear basis $\mathbf{h} := (h_1, h_2, \dots, h_m)$:

$$\begin{aligned}
 \bar{y}(t) &= h(y(t)) \\
 &= q_1h_1(y(t)) + q_2h_2(y(t)) + \dots + q_mh_m(y(t)) \\
 &= \mathbf{h}(y(t))\mathbf{q},
 \end{aligned}
 \tag{2}$$

where q_i ($i = 1, 2, \dots, m$) are the unknown nonlinear coefficients and m is the nonlinear order. The input $u(t)$ of the linear part, the output $\bar{y}(t)$ of the nonlinear feedback part and the input $r(t)$ of the system have the following relation

$$u(t) := r(t) - \bar{y}(t).
 \tag{3}$$

Define the parameter vectors \mathbf{a} and \mathbf{b} of the linear part, and \mathbf{q} of the nonlinear part as

$$\mathbf{a} := \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n_a} \end{bmatrix} \in \mathbb{R}^{n_a}, \quad \mathbf{b} := \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n_b} \end{bmatrix} \in \mathbb{R}^{n_b}, \quad \mathbf{q} := \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{bmatrix} \in \mathbb{R}^m,$$

and define the output information vector $\varphi_a(t)$, the input information vector $\varphi_b(t)$, and the feedback information matrix $\mathbf{H}(t)$ as

$$\varphi_a(t) := \begin{bmatrix} -y(t-1) \\ -y(t-2) \\ \vdots \\ -y(t-n_a) \end{bmatrix} \in \mathbb{R}^{n_a}, \quad \varphi_b(t) := \begin{bmatrix} r(t-1) \\ r(t-2) \\ \vdots \\ r(t-n_b) \end{bmatrix} \in \mathbb{R}^{n_b},$$

$$\mathbf{H}(t) := \begin{bmatrix} -h_1(y(t-1)) & -h_2(y(t-1)) & \cdots & -h_m(y(t-1)) \\ -h_1(y(t-2)) & -h_2(y(t-2)) & \cdots & -h_m(y(t-2)) \\ \vdots & \vdots & & \vdots \\ -h_1(y(t-n_b)) & -h_2(y(t-n_b)) & \cdots & -h_m(y(t-n_b)) \end{bmatrix} \in \mathbb{R}^{n_b \times m}.$$

From Eqs. (1)–(3), we have

$$\begin{aligned} y(t) &= -a_1y(t-1) - a_2y(t-2) - \cdots - a_{n_a}y(t-n_a) + b_1r(t-1) + b_2r(t-2) + \cdots + b_{n_b}r(t-n_b) \\ &\quad - b_1q_1h_1(y(t-1)) - b_2q_2h_2(y(t-2)) - \cdots - b_{n_b}q_mh_m(y(t-n_b)) + v(t) \\ &= -\sum_{j=1}^{n_a} a_jy(t-j) + \sum_{j=1}^{n_b} b_jr(t-j) - \sum_{j=1}^{n_b} b_j \sum_{i=1}^m q_ih_i(y(t-j)) + v(t) \\ &= \varphi_a^T(t)\mathbf{a} + \varphi_b^T(t)\mathbf{b} + \mathbf{b}^T\mathbf{H}(t)\mathbf{q} + v(t). \end{aligned} \quad (4)$$

The parameter identification algorithms proposed in this paper are based on identification model in (4). Many identification methods are derived according to the identification models of the systems [24–26, 28, 54, 94, 95, 109], and these methods can be used to estimate the parameters of other linear systems and nonlinear systems [60–67] and can be applied to engineering areas [44, 45, 92, 101, 103–106] such as information processing and process control systems.

Remark 1 In Equation (4), there is a cross product between parameter vectors \mathbf{b} in the linear block and \mathbf{q} in the nonlinear block, which can be called the bilinear-in-parameter identification model. To identify the parameters of the linear part and the nonlinear part, we can decompose the nonlinear feedback system based on the hierarchical identification principle.

Remark 2 This article considers the feedback nonlinear systems disturbed by white noise. In future research, we will consider identification problems with colored noise, i.e., autoregressive noise, moving average noise, autoregressive moving average noise [19, 34, 111] and so on.

3 The Gradient-Based Iterative Algorithm

Based on negative gradient search, this section derives the GI algorithm for estimating the parameters of a nonlinear feedback system from the measurement data.

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

$$\begin{aligned} \boldsymbol{\varphi}(t) &:= [\boldsymbol{\varphi}_a^T(t), \boldsymbol{\varphi}_b^T(t), \boldsymbol{\varphi}_m^T(t)]^T \in \mathbb{R}^n, \\ \boldsymbol{\varphi}_m(t) &:= \mathbf{H}^T(t)\mathbf{b} \in \mathbb{R}^m, \\ \boldsymbol{\vartheta} &:= [\mathbf{a}^T, \mathbf{b}^T, \mathbf{q}^T]^T \in \mathbb{R}^n, \\ n &:= n_a + n_b + m. \end{aligned}$$

The identification model for the nonlinear feedback system in (4) is given by

$$\begin{aligned} y(t) &= \boldsymbol{\varphi}_a^T(t)\mathbf{a} + \boldsymbol{\varphi}_b^T(t)\mathbf{b} + \mathbf{b}^T\mathbf{H}(t)\mathbf{q} + v(t) \\ &= [\boldsymbol{\varphi}_a^T(t), \boldsymbol{\varphi}_b^T(t), \boldsymbol{\varphi}_m^T(t)] \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{q} \end{bmatrix} + v(t) \\ &= \boldsymbol{\varphi}^T(t)\boldsymbol{\vartheta} + v(t). \end{aligned} \tag{5}$$

Let L be the data length ($L \gg n$) and define the stacked output vector $\mathbf{Y}(L)$, the stacked information matrices $\boldsymbol{\Psi}(L)$, $\boldsymbol{\Phi}_a(L)$, $\boldsymbol{\Phi}_b(L)$ and $\boldsymbol{\Phi}_m(L)$ as

$$\begin{aligned} \mathbf{Y}(L) &:= [y(1), y(2), \dots, y(L)]^T \in \mathbb{R}^L, \\ \boldsymbol{\Psi}(L) &:= [\boldsymbol{\Phi}_a^T(L), \boldsymbol{\Phi}_b^T(L), \boldsymbol{\Phi}_m^T(L)]^T \in \mathbb{R}^{L \times n}, \\ \boldsymbol{\Phi}_a(L) &:= [\boldsymbol{\varphi}_a(1), \boldsymbol{\varphi}_a(2), \dots, \boldsymbol{\varphi}_a(L)]^T \in \mathbb{R}^{L \times n_a}, \\ \boldsymbol{\Phi}_b(L) &:= [\boldsymbol{\varphi}_b(1), \boldsymbol{\varphi}_b(2), \dots, \boldsymbol{\varphi}_b(L)]^T \in \mathbb{R}^{L \times n_b}, \\ \boldsymbol{\Phi}_m(L) &:= [\boldsymbol{\varphi}_m(1), \boldsymbol{\varphi}_m(2), \dots, \boldsymbol{\varphi}_m(L)]^T \in \mathbb{R}^{L \times m}. \end{aligned}$$

Define the criterion function as

$$J_1(\boldsymbol{\vartheta}) := \frac{1}{2} \|\mathbf{Y}(L) - \boldsymbol{\Psi}(L)\boldsymbol{\vartheta}\|^2. \tag{6}$$

We make the following definitions.

- (i) $k = 1, 2, 3, \dots$ is the iteration variable.
- (ii) $\hat{\boldsymbol{\vartheta}}_k$ is the estimate of the parameter vector $\boldsymbol{\vartheta}$ at iteration k .
- (iii) $\mu_{1,k}$ is the iteration step-size.

By means of the negative gradient search, and minimizing the criterion function $J_1(\boldsymbol{\vartheta})$ with respect to $\boldsymbol{\vartheta}$, we can obtain the iterative relation:

$$\begin{aligned} \hat{\boldsymbol{\vartheta}}_k &= \hat{\boldsymbol{\vartheta}}_{k-1} - \mu_{1,k} \text{grad}[J_1(\hat{\boldsymbol{\vartheta}}_{k-1})] \\ &= \hat{\boldsymbol{\vartheta}}_{k-1} + \mu_{1,k} \boldsymbol{\Psi}^T(L)[\mathbf{Y}(L) - \boldsymbol{\Psi}(L)\hat{\boldsymbol{\vartheta}}_{k-1}], \end{aligned} \tag{7}$$

where the information matrices $\boldsymbol{\Phi}_m(L)$ in $\boldsymbol{\Psi}(L)$ are unknown. Because the information vector $\boldsymbol{\varphi}_m(t)$ in $\boldsymbol{\Phi}_m(L)$ contains the unknown parameter vector \mathbf{b} , Eq. (7) cannot yield the estimate $\hat{\boldsymbol{\vartheta}}_k$ directly. Our approach is to replace the unknown variable \mathbf{b} in $\boldsymbol{\varphi}_m(t)$

with its estimate $\hat{\mathbf{b}}_{k-1}$ at the $(k-1)$ th iteration. The estimate $\hat{\boldsymbol{\varphi}}_{m,k}(t)$ of the information vector $\boldsymbol{\varphi}_m(t)$ at iteration k is given by

$$\hat{\boldsymbol{\varphi}}_{m,k}(t) := \mathbf{H}^T(t)\hat{\mathbf{b}}_{k-1} \in \mathbb{R}^m.$$

Replacing $\boldsymbol{\Phi}_m(L)$ and $\boldsymbol{\Psi}_k(L)$ with their estimates $\hat{\boldsymbol{\Phi}}_{m,k}(L)$ and $\hat{\boldsymbol{\Psi}}_k(L)$ at k , we have

$$\begin{aligned} \hat{\boldsymbol{\Phi}}_{m,k}(L) &:= [\hat{\boldsymbol{\varphi}}_{m,k}(1), \hat{\boldsymbol{\varphi}}_{m,k}(2), \dots, \hat{\boldsymbol{\varphi}}_{m,k}(L)]^T \in \mathbb{R}^{L \times m}, \\ \hat{\boldsymbol{\Psi}}_k(L) &:= [\boldsymbol{\Phi}_a^T(L), \boldsymbol{\Phi}_b^T(L), \hat{\boldsymbol{\Phi}}_{m,k}^T(L)]^T \in \mathbb{R}^{L \times n}. \end{aligned}$$

Utilizing the estimate $\hat{\boldsymbol{\Psi}}_k(L)$ to replace the information matrix $\boldsymbol{\Psi}(L)$ in (7), we obtain

$$\begin{aligned} \hat{\boldsymbol{\theta}}_k &= \hat{\boldsymbol{\theta}}_{k-1} + \mu_{1,k} \hat{\boldsymbol{\Psi}}_k^T(L)[\mathbf{Y}(L) - \hat{\boldsymbol{\Psi}}_k(L)\hat{\boldsymbol{\theta}}_{k-1}] \\ &= [\mathbf{I}_n - \mu_{1,k} \hat{\boldsymbol{\Psi}}_k^T(L)\hat{\boldsymbol{\Psi}}_k(L)]\hat{\boldsymbol{\theta}}_{k-1} + \mu_{1,k} \hat{\boldsymbol{\Psi}}_k^T(L)\mathbf{Y}(L). \end{aligned} \tag{8}$$

Equation (8) can be seen as a discrete-time system. To ensure the convergence of $\hat{\boldsymbol{\theta}}_k$, all the eigenvalues of matrix $[\mathbf{I}_n - \mu_{1,k} \hat{\boldsymbol{\Psi}}_k^T(L)\hat{\boldsymbol{\Psi}}_k(L)]$ must be in the unit circle. Thus, the step-size $\mu_{1,k}$ can be conservatively chosen as

$$0 < \mu_{1,k} \leq \frac{2}{\lambda_{\max}[\hat{\boldsymbol{\Psi}}_k^T(L)\hat{\boldsymbol{\Psi}}_k(L)]} = 2\lambda_{\max}^{-1}[\hat{\boldsymbol{\Psi}}_k^T(L)\hat{\boldsymbol{\Psi}}_k(L)].$$

From the above derivations, we can summarize the GI algorithm for the nonlinear feedback system as

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} + \mu_{1,k} \hat{\boldsymbol{\Psi}}_k^T(L)[\mathbf{Y}(L) - \hat{\boldsymbol{\Psi}}_k(L)\hat{\boldsymbol{\theta}}_{k-1}], \tag{9}$$

$$\mu_{1,k} \leq 2\lambda_{\max}^{-1}[\hat{\boldsymbol{\Psi}}_k^T(L)\hat{\boldsymbol{\Psi}}_k(L)], \tag{10}$$

$$\mathbf{Y}(L) = [y(1), y(2), \dots, y(L)]^T, \tag{11}$$

$$\hat{\boldsymbol{\Psi}}_k(L) = [\boldsymbol{\Phi}_a^T(L), \boldsymbol{\Phi}_b^T(L), \hat{\boldsymbol{\Phi}}_{m,k}^T(L)]^T, \tag{12}$$

$$\boldsymbol{\Phi}_a(L) = [\boldsymbol{\varphi}_a(1), \boldsymbol{\varphi}_a(2), \dots, \boldsymbol{\varphi}_a(L)]^T, \tag{13}$$

$$\boldsymbol{\Phi}_b(L) = [\boldsymbol{\varphi}_b(1), \boldsymbol{\varphi}_b(2), \dots, \boldsymbol{\varphi}_b(L)]^T, \tag{14}$$

$$\hat{\boldsymbol{\Phi}}_{m,k}(L) = [\hat{\boldsymbol{\varphi}}_{m,k}(1), \hat{\boldsymbol{\varphi}}_{m,k}(2), \dots, \hat{\boldsymbol{\varphi}}_{m,k}(L)]^T, \tag{15}$$

$$\boldsymbol{\varphi}_a(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \tag{16}$$

$$\boldsymbol{\varphi}_b(t) = [r(t-1), r(t-2), \dots, r(t-n_b)]^T, \tag{17}$$

$$\hat{\boldsymbol{\varphi}}_{m,k}(t) = \mathbf{H}^T(t)\hat{\mathbf{b}}_{k-1}, \tag{18}$$

$$\hat{\boldsymbol{\varphi}}_k(t) = [\boldsymbol{\varphi}_a(t), \boldsymbol{\varphi}_b(t), \hat{\boldsymbol{\varphi}}_{m,k}(t)]^T, \tag{19}$$

$$\mathbf{H}(t) := [-\mathbf{h}^T(y(t-1)), -\mathbf{h}^T(y(t-2)), \dots, -\mathbf{h}^T(y(t-n_b))]^T, \tag{20}$$

$$\hat{\boldsymbol{\theta}}_k = [\hat{\mathbf{a}}_k^T, \hat{\mathbf{b}}_k^T, \hat{\mathbf{q}}_k^T]^T, \tag{21}$$

$$\hat{\mathbf{a}}_k = [\hat{a}_{1,k}, \hat{a}_{2,k}, \dots, \hat{a}_{n_a,k}]^T, \quad (22)$$

$$\hat{\mathbf{b}}_k = [\hat{b}_{1,k}, \hat{b}_{2,k}, \dots, \hat{b}_{n_b,k}]^T, \quad (23)$$

$$\hat{\mathbf{q}}_k = [\hat{q}_{1,k}, \hat{q}_{2,k}, \dots, \hat{q}_{m,k}]^T. \quad (24)$$

The steps of computing the parameter estimation vector $\hat{\boldsymbol{\vartheta}}_k$ by the GI algorithm in (9)–(24) are listed as follows.

1. To initialize: let $k = 1$, $\hat{\boldsymbol{\vartheta}}_0 = [\hat{a}_{1,0}, \dots, \hat{a}_{n_a,0}, \hat{b}_{1,0}, \dots, \hat{b}_{n_b,0}, \hat{q}_{1,0}, \dots, \hat{q}_{m,0}]^T$ be any real vector, $r(t-j) = 1/p_0$ and $y(t-j) = 1/p_0$, $j = 1, 2, \dots, \max[n_a, n_b]$, $t = 1, 2, \dots, L$, $p_0 = 10^6$. Set the data length L ($L \gg n$), the parameter estimation precision $\varepsilon > 0$ and the basis function $h(\cdot)$.
2. Collect the input and output data $r(t)$ and $y(t)$, $t = 1, 2, \dots, L$, and form $\mathbf{Y}(L)$ by (11).
3. Form $\mathbf{H}(t)$ by (20), form $\boldsymbol{\varphi}_a(t)$ and $\boldsymbol{\varphi}_b(t)$ by (16) and (17), $t = 1, 2, \dots, L$.
4. Form $\boldsymbol{\Phi}_a(L)$ and $\boldsymbol{\Phi}_b(L)$ by (13) and (14).
5. Compute $\hat{\boldsymbol{\varphi}}_{m,k}(t)$ by (18), form $\hat{\boldsymbol{\varphi}}_k(t)$ by (19), $t = 1, 2, \dots, l + L$.
6. Form $\hat{\boldsymbol{\Phi}}_{m,k}(L)$ by (15), form $\hat{\boldsymbol{\Psi}}_k(L)$ by (12).
7. Select a large step-size $\mu_{1,k}$ according to (10), and update the parameter estimation vector $\hat{\boldsymbol{\vartheta}}_k$ by (9).
8. Read out the parameter estimation vectors $\hat{\mathbf{a}}_k$, $\hat{\mathbf{b}}_k$ and $\hat{\mathbf{q}}_k$ from $\hat{\boldsymbol{\vartheta}}_k$ in Eq. (21).
9. Compare $\hat{\boldsymbol{\vartheta}}_k$ with $\hat{\boldsymbol{\vartheta}}_{k-1}$: if $\|\hat{\boldsymbol{\vartheta}}_k - \hat{\boldsymbol{\vartheta}}_{k-1}\| > \varepsilon$, increase k by 1 and go to Step 5; otherwise, output the parameter estimate $\hat{\boldsymbol{\vartheta}}_k$, and terminate this iterative calculation process.

Remark 3 The parameter estimation methods proposed in this paper can be combined with other estimation approaches to study parameter estimation problems for linear stochastic systems and nonlinear stochastic systems, and they can be applied to engineering application systems, such as atomic force microscope cantilever modeling, heat exchange processes, and the identification by microwave crystal detectors.

4 The Hierarchical Gradient-Based Iterative Algorithm

Inspired by the hierarchical identification principle, we employ the H-GI algorithm for nonlinear feedback systems to improve parameter estimation accuracy. The feedback nonlinear system in (4) is decomposed into two subsystems, which contain the parameter vectors $\boldsymbol{\theta}_1 := [\mathbf{a}^T, \mathbf{b}^T]^T \in \mathbb{R}^{n_a+n_b}$ and $\boldsymbol{\theta}_m := \mathbf{q} \in \mathbb{R}^m$.

Define the system information vectors $\boldsymbol{\varphi}_1(t)$ and $\boldsymbol{\varphi}_m(t)$, and the intermediate variable y_m as

$$\begin{aligned} \boldsymbol{\varphi}_1(t) &:= [\boldsymbol{\varphi}_a^T(t), \boldsymbol{\varphi}_b^T(t) + \mathbf{q}^T \mathbf{H}^T(t)]^T \in \mathbb{R}^{n_a+n_b}, \\ \boldsymbol{\varphi}_m(t) &:= \mathbf{H}^T(t) \mathbf{b} \in \mathbb{R}^m, \\ y_m &:= y(t) - \boldsymbol{\varphi}_a^T(t) \mathbf{a} - \boldsymbol{\varphi}_b^T(t) \mathbf{b} \\ &= \boldsymbol{\varphi}_m^T(t) \boldsymbol{\theta}_m + v(t). \end{aligned} \quad (25)$$

Eq. (4) can be equivalently written as

$$y(t) = \boldsymbol{\varphi}_1^T(t)\boldsymbol{\theta}_1 + v(t). \quad (26)$$

Let L be the data length ($L \gg n$) and define the stacked output vectors $\mathbf{Y}(L)$ and $\mathbf{Y}_m(L)$, the stacked information matrices $\boldsymbol{\Phi}_1(L)$, $\boldsymbol{\Phi}_m(L)$, $\boldsymbol{\Phi}_a(L)$ and $\boldsymbol{\Phi}_b(L)$ as

$$\begin{aligned} \mathbf{Y}(L) &:= [y(1), y(2), \dots, y(L)]^T \in \mathbb{R}^L, \\ \mathbf{Y}_m(L) &:= [y_m(1), y_m(2), \dots, y_m(L)]^T \in \mathbb{R}^L, \\ \boldsymbol{\Phi}_1(L) &:= [\boldsymbol{\varphi}_1(1), \boldsymbol{\varphi}_1(2), \dots, \boldsymbol{\varphi}_1(L)]^T \in \mathbb{R}^{L \times (n_a + n_b)}, \\ \boldsymbol{\Phi}_m(L) &:= [\boldsymbol{\varphi}_m(1), \boldsymbol{\varphi}_m(2), \dots, \boldsymbol{\varphi}_m(L)]^T \in \mathbb{R}^{L \times m}, \\ \boldsymbol{\Phi}_a(L) &:= [\boldsymbol{\varphi}_a(1), \boldsymbol{\varphi}_a(2), \dots, \boldsymbol{\varphi}_a(L)]^T \in \mathbb{R}^{L \times n_a}, \\ \boldsymbol{\Phi}_b(L) &:= [\boldsymbol{\varphi}_b(1), \boldsymbol{\varphi}_b(2), \dots, \boldsymbol{\varphi}_b(L)]^T \in \mathbb{R}^{L \times n_b}. \end{aligned}$$

Define two criterion functions,

$$\begin{aligned} J_2(\boldsymbol{\theta}_1) &:= \frac{1}{2} \|\mathbf{Y}(L) - \boldsymbol{\Phi}_1(L)\boldsymbol{\theta}_1\|^2, \\ J_3(\boldsymbol{\theta}_m) &:= \frac{1}{2} \|\mathbf{Y}_m(L) - \boldsymbol{\Phi}_m(L)\boldsymbol{\theta}_m\|^2. \end{aligned}$$

Using negative gradient search and the definitions of $\mathbf{Y}(L)$ and $\mathbf{Y}_m(L)$, and minimizing the criterion functions $J_2(\boldsymbol{\theta}_1)$ and $J_3(\boldsymbol{\theta}_m)$ with respect to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_m$, we can obtain the iterative relations:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{1,k} &= \hat{\boldsymbol{\theta}}_{1,k-1} - \mu_{2,k} \text{grad}[J_2(\hat{\boldsymbol{\theta}}_{1,k-1})] \\ &= \hat{\boldsymbol{\theta}}_{1,k-1} + \mu_{2,k} \boldsymbol{\Phi}_1^T(L)[\mathbf{Y}(L) - \boldsymbol{\Phi}_1(L)\hat{\boldsymbol{\theta}}_{1,k-1}] \\ &= [\mathbf{I}_{n_a+n_b} - \mu_{2,k} \boldsymbol{\Phi}_1^T(L)\boldsymbol{\Phi}_1(L)]\hat{\boldsymbol{\theta}}_{1,k-1} + \mu_{2,k} \boldsymbol{\Phi}_1^T(L)\mathbf{Y}(L), \end{aligned} \quad (27)$$

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{m,k} &= \hat{\boldsymbol{\theta}}_{m,k-1} - \mu_{3,k} \text{grad}[J_3(\hat{\boldsymbol{\theta}}_{m,k-1})] \\ &= \hat{\boldsymbol{\theta}}_{m,k-1} + \mu_{3,k} \boldsymbol{\Phi}_m^T(L)[\mathbf{Y}_m(L) - \boldsymbol{\Phi}_m(L)\hat{\boldsymbol{\theta}}_{m,k-1}] \\ &= [\mathbf{I}_m - \mu_{3,k} \boldsymbol{\Phi}_m^T(L)\boldsymbol{\Phi}_m(L)]\hat{\boldsymbol{\theta}}_{m,k-1} + \mu_{3,k} \boldsymbol{\Phi}_m^T(L)[\mathbf{Y}_m(L) - \boldsymbol{\Phi}_a(L)\mathbf{a} - \boldsymbol{\Phi}_b(L)\mathbf{b}], \end{aligned} \quad (28)$$

where $\mu_{2,k} > 0$ and $\mu_{3,k} > 0$ are iterative step-sizes and Eqs. (27)–(28) can be seen as two discrete-time systems. To ensure the convergences of $\hat{\boldsymbol{\theta}}_{1,k}$ and $\hat{\boldsymbol{\theta}}_{m,k}$, all the eigenvalues of matrices $[\mathbf{I}_{n_a+n_b} - \mu_{2,k} \boldsymbol{\Phi}_1^T(L)\boldsymbol{\Phi}_1(L)]$ and $[\mathbf{I}_m - \mu_{3,k} \boldsymbol{\Phi}_m^T(L)\boldsymbol{\Phi}_m(L)]$ must be in the unit circle, that is to say $\mu_{2,k}$ and $\mu_{3,k}$ should satisfy $-\mathbf{I}_{n_a+n_b} \leq \mathbf{I}_{n_a+n_b} - \mu_{2,k} \boldsymbol{\Phi}_1^T(L)\boldsymbol{\Phi}_1(L) \leq \mathbf{I}_{n_a+n_b}$ and $-\mathbf{I}_m \leq \mathbf{I}_m - \mu_{3,k} \boldsymbol{\Phi}_m^T(L)\boldsymbol{\Phi}_m(L) \leq \mathbf{I}_m$, and the conservative choices of $\mu_{2,k}$ and $\mu_{3,k}$ are

$$\begin{aligned} 0 < \mu_{2,k} &\leq \frac{2}{\lambda_{\max}[\boldsymbol{\Phi}_1^T(L)\boldsymbol{\Phi}_1(L)]} = 2\lambda_{\max}^{-1}[\boldsymbol{\Phi}_1^T(L)\boldsymbol{\Phi}_1(L)], \\ 0 < \mu_{3,k} &\leq \frac{2}{\lambda_{\max}[\boldsymbol{\Phi}_m^T(L)\boldsymbol{\Phi}_m(L)]} = 2\lambda_{\max}^{-1}[\boldsymbol{\Phi}_m^T(L)\boldsymbol{\Phi}_m(L)], \end{aligned}$$

where $\lambda_{\max}[X]$ denotes the maximum eigenvalue of the square matrix X . The right-hand sides of (27)–(28) contain the unknown parameter vectors θ_1, θ_m and the unknown information matrices $\Phi_1(L)$ and $\Phi_m(L)$, so we cannot directly compute the estimates $\hat{\theta}_{1,k}$ and $\hat{\theta}_{m,k}$. The solution is to replace $\Phi_1(L)$ in (27) with their estimates $\hat{\Phi}_{1,k}(L)$, and to replace $\Phi_m(L), a$ and b in (28) with their estimates $\hat{\Phi}_{m,k}(L), \hat{a}_k$ and \hat{b}_k . We can summarize the H-GI algorithm for the nonlinear feedback system as

$$\hat{\theta}_{1,k} = \hat{\theta}_{1,k-1} + \mu_{2,k} \hat{\Phi}_{1,k}^T(L) [Y(L) - \hat{\Phi}_{1,k}(L) \hat{\theta}_{1,k-1}], \tag{29}$$

$$\mu_{2,k} \leq 2\lambda_{\max}^{-1} [\hat{\Phi}_{1,k}^T(L) \hat{\Phi}_{1,k}(L)], \tag{30}$$

$$\hat{\theta}_{m,k} = \hat{\theta}_{m,k-1} + \mu_{3,k} \hat{\Phi}_{m,k}^T(L) [Y(L) - \Phi_a(L) \hat{a}_k - \Phi_b(L) \hat{b}_k - \hat{\Phi}_{m,k}(L) \hat{\theta}_{m,k-1}], \tag{31}$$

$$\mu_{3,k} \leq 2\lambda_{\max}^{-1} [\hat{\Phi}_{m,k}^T(L) \hat{\Phi}_{m,k}(L)], \tag{32}$$

$$Y(L) = [y(1), y(2), \dots, y(L)]^T, \tag{33}$$

$$\hat{\Phi}_{1,k}(L) = [\hat{\varphi}_{1,k}(1), \hat{\varphi}_{1,k}(2), \dots, \hat{\varphi}_{1,k}(L)]^T, \tag{34}$$

$$\hat{\Phi}_{m,k}(L) = [\hat{\varphi}_{m,k}(1), \hat{\varphi}_{m,k}(2), \dots, \hat{\varphi}_{m,k}(L)]^T, \tag{35}$$

$$\Phi_a(L) = [\varphi_a(1), \varphi_a(2), \dots, \varphi_a(L)]^T, \tag{36}$$

$$\Phi_b(L) = [\varphi_b(1), \varphi_b(2), \dots, \varphi_b(L)]^T, \tag{37}$$

$$\hat{\varphi}_{1,k}(t) = [\varphi_a^T(t), \varphi_b^T(t) + \hat{q}_{k-1}^T H^T(t)]^T \tag{38}$$

$$\hat{\varphi}_{m,k}(t) = H^T(t) \hat{b}_{k-1}, \tag{39}$$

$$\varphi_a(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a)]^T, \tag{40}$$

$$\varphi_b(t) = [u(t-1), u(t-2), \dots, u(t-n_b)]^T, \tag{41}$$

$$H(t) := [-h^T(y(t-1)), -h^T(y(t-2)), \dots, -h^T(y(t-n_b))]^T, \tag{42}$$

$$\hat{\theta}_{1,k} = [\hat{a}_k^T, \hat{b}_k^T]^T, \tag{43}$$

$$\hat{\theta}_{m,k} = \hat{q}_k, \tag{44}$$

$$\hat{a}_k = [\hat{a}_{1,k}, \hat{a}_{2,k}, \dots, \hat{a}_{n_a,k}]^T, \tag{45}$$

$$\hat{b}_k = [\hat{b}_{1,k}, \hat{b}_{2,k}, \dots, \hat{b}_{n_b,k}]^T, \tag{46}$$

$$\hat{q}_k = [\hat{q}_{1,k}, \hat{q}_{2,k}, \dots, \hat{q}_{m,k}]^T. \tag{47}$$

The methods proposed in this paper can combine some statistical tools and strategies [84–87] and identification algorithms [59, 81, 82, 89–91] to study the parameter estimation algorithms of other linear and nonlinear stochastic systems and can be applied to other fields [13–17, 52–57] such as paper-making and chemical engineering systems. The steps of computing the parameter estimation vectors $\hat{\theta}_{1,k}$ and $\hat{\theta}_{m,k}$ by the H-GI algorithm in (29)–(47) are listed as follows.

1. To initialize: let $k = 1, \hat{\theta}_{1,0} = [\hat{a}_{1,0}, \dots, \hat{a}_{n_a,0}, \hat{b}_{1,0}, \dots, \hat{b}_{n_b,0}]^T, \hat{\theta}_{m,0} = [\hat{q}_{1,0}, \dots, \hat{q}_{m,0}]^T$ be any real vector, $r(t-j) = 1/p_0$ and $y(t-j) = 1/p_0, j = 1, 2, \dots, \max[n_a, n_b], t = 1, 2, \dots, L, p_0 = 10^6$. Set the data length $L (L \gg n)$, the parameter estimation precision $\varepsilon > 0$ and the basis function $h(\cdot)$.
2. Collect the input and output data $r(t)$ and $y(t), t = 1, 2, \dots, L$, and form $Y(L)$ by (33).
3. Form $H(t)$ by (42), and form $\varphi_a(t)$ and $\varphi_b(t)$ by (40) and (41), $t = 1, 2, \dots, L$.

4. Compute $\hat{\varphi}_{1,k}(t)$ and $\hat{\varphi}_{m,k}(t)$ by (38) and (39), $t = 1, 2, \dots, L$.
5. Form $\hat{\Phi}_{1,k}(L)$ and $\hat{\Phi}_{m,k}(L)$ by (34) and (35).
6. Select the large step-sizes $\mu_{2,k}$ and $\mu_{3,k}$ according to (30) and (32), and update the parameter estimation vectors $\hat{\theta}_{1,k}$ and $\hat{\theta}_{m,k}$ by (29) and (31).
7. Read out the parameter estimation vectors \hat{a}_k and \hat{b}_k from $\hat{\theta}_{1,k}$ in (43), and \hat{q}_k from $\hat{\theta}_{m,k}$ in (44).
8. If $\|\hat{\theta}_{1,k} - \hat{\theta}_{1,k-1}\| + \|\hat{\theta}_{m,k} - \hat{\theta}_{m,k-1}\| > \varepsilon$, increase k by 1 and go to step 4; otherwise, output the parameter estimates $\hat{\theta}_{1,k}$ and $\hat{\theta}_{m,k}$, terminate this iterative calculation process.

Remark 4 Previous works presented an overall recursive least squares algorithm and an overall stochastic gradient algorithm to estimate the nonlinear feedback systems disturbed by white noise [93] and proposed a two-stage recursive least squares algorithm for nonlinear feedback systems based on auxiliary model identification [27]. In contrast to the recursive algorithms proposed in [93] and [27], this paper proposes an iterative estimation method for nonlinear feedback systems and introduces the hierarchical principle to reduce the calculation costs and to improve the identification accuracy.

The computational cost for an identification algorithm can be evaluated by the number of floating point operation (flops for short). An addition or multiplication counts as a flop. The computational costs for each iteration of the GI and H-GI algorithms are listed in Tables 1, 2. The total numbers of flops of these two algorithms are $N_1 := 2n^2L + 4nL + 2L + 2mn_b - n^2 - m$ and $N_2 := 4L(n_a + n_b + m) + 4L + 2(n_a + n_b)L + 2m^2L + 2(n_a + n_b)^2L + 2mn_b - m^2 - (n_a + n_b)^2 - m$, respectively. The difference between the computation loads of these two algorithms at each iteration is $N_1 - N_2 = 2L((n_a + n_b)(2m - 1) - 1) + (n_a + n_b)^2 + m^2 > 0$, thus, the H-GI algorithm has less computation load than the GI algorithm.

5 The Simulation Example

Consider the following nonlinear feedback system:

$$\begin{aligned} A(z)y(t) &= B(z)[r(t) - \bar{y}(t)] + v(t), \\ A(z) &= 1 + a_1z^{-1} + a_2z^{-2} = 1 + 0.35z^{-1} + 0.65z^{-2}, \\ B(z) &= b_1z^{-1} + b_2z^{-2} = 1.68z^{-1} - 0.32z^{-2}, \\ \bar{y}(t) &= 0.80 \sin(y^2(t)) + 0.60 \sin(y^3(t)). \end{aligned}$$

The parameters to be identified are

$$\vartheta = [a_1, a_2, b_1, b_2, q_1, q_2]^T = [0.35, 0.65, 1.68, -0.32, 0.80, 0.60]^T.$$

The simulation of the algorithms in this paper is performed in MATLAB. In the simulation, we set the input and output data and noise as follows.

Table 1 The computational loads of the GI algorithm at each iteration

Expressions	Number of Multiplications	Number of Additions
$\hat{\boldsymbol{\phi}}_k = \hat{\boldsymbol{\phi}}_{k-1} + \mu_{1,k} \hat{\boldsymbol{\Psi}}_k^T(L) \mathbf{E}_1(L) \in \mathbb{R}^n$	$nL + L$	$nL + L - 1$
$\mathbf{E}_1(L) := \mathbf{Y}(L) - \hat{\boldsymbol{\Psi}}_k(L) \hat{\boldsymbol{\phi}}_{k-1} \in \mathbb{R}^L$	nL	nL
$\mu_{1,k} \leq 2\lambda_{\max}^{-1} [\hat{\boldsymbol{\Psi}}_k^T(L) \hat{\boldsymbol{\Psi}}_k(L)] \in \mathbb{R}^n$	$n^2 L + 1$	$n^2(L - 1)$
$\hat{\boldsymbol{\phi}}_{m,k}(t) = \mathbf{H}^T(t) \hat{\boldsymbol{\phi}}_{k-1} \in \mathbb{R}^{n_b}$	mn_b	$m(n_b - 1)$
Sum	$n^2 L + 2nL + L + mn_b + 1$	$n^2 L + 2nL + L - m - n^2 + mn_b - 1$
Total flops	$N_1 = 2n^2 L + 4nL + 2L + 2mn_b - n^2 - m$	

Table 2 The computational loads of the H-GI algorithm at each iteration

Expressions	Number of Multiplications	Number of Additions
$\hat{\theta}_{1,k} = \hat{\theta}_{1,k-1} + \mu_{2,k} \hat{\Phi}_{1,k}^T(L) \mathbf{E}_2(L) \in \mathbb{R}^{n_a+n_b}$	$(n_a + n_b)L + L$	$(n_a + n_b)L + L - 1$
$\mathbf{E}_2(L) := \mathbf{Y}(L) - \hat{\Phi}_{1,k}(L) \hat{\theta}_{1,k-1} \in \mathbb{R}^L$	$(n_a + n_b)L$	$(n_a + n_b)L$
$\mu_{2,k} \leq 2\lambda_{\max}^{-1}[\Phi_{1,k}^T(L) \Phi_{1,k}(L)] \in \mathbb{R}^{n_a+n_b}$	$(n_a + n_b)^2 L + 1$	$(n_a + n_b)^2(L - 1)$
$\hat{\theta}_{m,k} = \hat{\theta}_{m,k-1} + \mu_{3,k} \hat{\Phi}_{m,k}^T(L) \mathbf{E}_3(L) \in \mathbb{R}^m$	$(m + 1)L$	$(m + 1)L - 1$
$\mathbf{E}_3(L) := \mathbf{Y}(L) - \Phi_a(L) \hat{a}_k - \Phi_b(L) \hat{b}_k$ $-\hat{\Phi}_{m,k}(L) \hat{\theta}_{m,k-1} \in \mathbb{R}^L$	$(n_a + n_b + m)L$	$(n_a + n_b + m)L$
$\mu_{3,k} \leq 2\lambda_{\max}^{-1}[\Phi_{m,k}^T(L) \Phi_{m,k}(L)] \in \mathbb{R}^m$	$m^2 L + 1$	$m^2(L - 1)$
$\hat{\Phi}_{m,k}(t) = \mathbf{H}^T(t) \hat{b}_{k-1} \in \mathbb{R}^{n_b}$	mn_b	$m(n_b - 1)$
Sum	$2L(n_a + n_b + m)$	$2L(n_a + n_b + m)$
	$+(n_a + n_b)L + 2L$	$+(n_a + n_b)L + 2L$
	$+(n_a + n_b)^2 L$	$+(n_a + n_b)^2 L - 2$
	$+m^2 L + mn_b + 2$	$+m^2 L + mn_b - m^2$
Total flops	$N_2 = 4L(n_a + n_b + m) + 4L + 2(n_a + n_b)L + 2m^2 L$ $+2(n_a + n_b)^2 L + 2mn_b - m^2 - (n_a + n_b)^2 - m$	$-(n_a + n_b)^2 - m$

Table 3 The GI parameter estimates and errors

σ^2	k	a_1	a_2	b_1	b_2	q_1	q_2	δ (%)
1.00 ²	1	0.83711	0.85375	1.05735	0.98230	0.99603	0.97317	75.38510
	2	0.76694	0.78899	1.10964	0.94408	0.98315	0.94373	71.23351
	5	0.68499	0.73981	1.24169	0.80703	0.93997	0.86301	61.14949
	10	0.61474	0.72018	1.39875	0.59615	0.88821	0.76310	47.95324
	20	0.52017	0.69562	1.56734	0.27824	0.83486	0.64989	30.11347
	50	0.40065	0.66358	1.67901	-0.15065	0.79849	0.57169	8.49189
	100	0.36319	0.65336	1.68651	-0.28675	0.79591	0.56717	2.33042
	150	0.35916	0.65225	1.68642	-0.30137	0.79597	0.56725	1.87185
2.00 ²	200	0.35872	0.65213	1.68640	-0.30294	0.79598	0.56727	1.83204
	1	0.73415	0.84390	1.03892	0.98841	0.99541	0.98143	74.66350
	2	0.64756	0.76048	1.07450	0.95294	0.98271	0.95914	70.95090
	5	0.57346	0.70588	1.16882	0.82540	0.94330	0.89693	62.30267
	10	0.53124	0.69361	1.29506	0.63042	0.89907	0.81935	50.59384
	20	0.47499	0.67985	1.46415	0.33324	0.85685	0.72393	33.72589
	50	0.39755	0.65990	1.65022	-0.10014	0.80371	0.61333	10.76255
	100	0.36922	0.65213	1.69153	-0.26499	0.77139	0.58073	3.25076
3.00 ²	150	0.36538	0.65106	1.69428	-0.28751	0.76718	0.57776	2.62021
	200	0.36485	0.65091	1.69448	-0.29060	0.76677	0.57751	2.54960
	1	0.65859	0.83530	1.02503	0.99246	1.00160	0.99204	74.41847
	2	0.56211	0.73986	1.04826	0.96576	0.99582	0.97944	71.54020
	5	0.49649	0.68380	1.11211	0.86934	0.97428	0.94226	65.32047
	10	0.47378	0.67683	1.20460	0.71757	0.94964	0.89492	56.52162
	20	0.44304	0.67054	1.34836	0.46886	0.93091	0.83663	42.68769
	50	0.39110	0.65969	1.57469	0.02741	0.91762	0.75025	19.50050
True values	100	0.36252	0.65289	1.67840	-0.22695	0.86551	0.67383	6.44533
	150	0.35593	0.65109	1.69719	-0.28873	0.83943	0.64962	3.45201
	200	0.35437	0.65065	1.70059	-0.30350	0.83239	0.64400	2.87747
		0.35000	0.65000	1.68000	-0.32000	0.80000	0.60000	

Table 4 The H-GI parameter estimates and errors

σ^2	k	a_1	a_2	b_1	b_2	q_1	q_2	δ (%)	
1.00 ²	1	0.83977	0.85614	1.05994	0.94877	0.95027	0.66377	71.78404	
	2	0.76551	0.78663	1.12830	0.88430	0.86000	0.65560	66.09631	
	5	0.64674	0.73384	1.28394	0.70106	0.80147	0.56758	53.83765	
	10	0.56889	0.71000	1.45400	0.43495	0.81699	0.58098	38.80842	
	20	0.47081	0.68323	1.60005	0.09078	0.81468	0.57935	20.69716	
	50	0.37395	0.65642	1.67719	-0.24980	0.80000	0.56977	3.80303	
	100	0.35662	0.65160	1.68481	-0.31127	0.79666	0.56772	1.63785	
	200	0.35576	0.65136	1.68515	-0.31432	0.79648	0.56761	1.60767	
	2.00 ²	1	0.73590	0.84493	1.04214	0.96200	0.91339	0.64993	70.99104
		2	0.63713	0.76056	1.08844	0.91332	0.83094	0.66433	66.40100
5		0.54816	0.69919	1.21180	0.73790	0.73300	0.53162	55.74955	
10		0.50366	0.68590	1.36243	0.48936	0.79280	0.58188	41.80586	
20		0.44287	0.67140	1.52275	0.14428	0.82025	0.60888	23.64292	
50		0.37575	0.65393	1.65725	-0.22900	0.78536	0.58967	4.68183	
100		0.36155	0.65008	1.68840	-0.31050	0.76915	0.57930	1.93533	
200		0.36056	0.64981	1.69128	-0.31632	0.76761	0.57830	1.99102	
3.00 ²		1	0.65959	0.83578	1.03346	0.97764	1.05155	0.74357	71.98421
		2	0.55606	0.74890	1.05743	0.95643	0.79803	0.51728	67.31302
	5	0.47981	0.68012	1.15989	0.79649	0.78245	0.59306	58.60511	
	10	0.45604	0.67283	1.28479	0.58974	0.86611	0.66607	47.40159	
	20	0.42089	0.66611	1.43917	0.27101	0.92899	0.71869	31.49253	
	50	0.37179	0.65490	1.60896	-0.14591	0.88598	0.68582	10.64100	
	100	0.35686	0.65123	1.68292	-0.27991	0.84145	0.65116	3.66388	
	200	0.35480	0.65074	1.70173	-0.29961	0.83081	0.64282	2.87486	
	True values		0.35000	0.65000	1.68000	-0.32000	0.80000	0.60000	

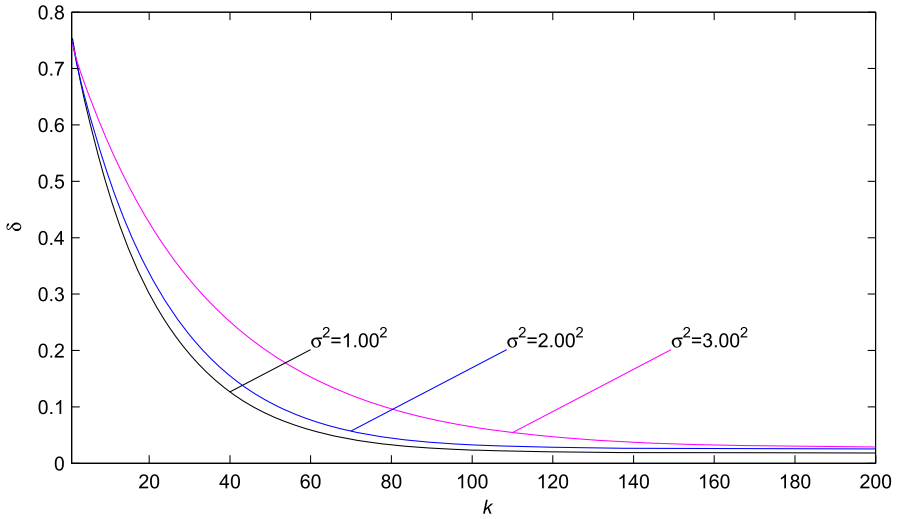


Fig. 2 The GI parameter estimation errors δ versus k with different σ^2

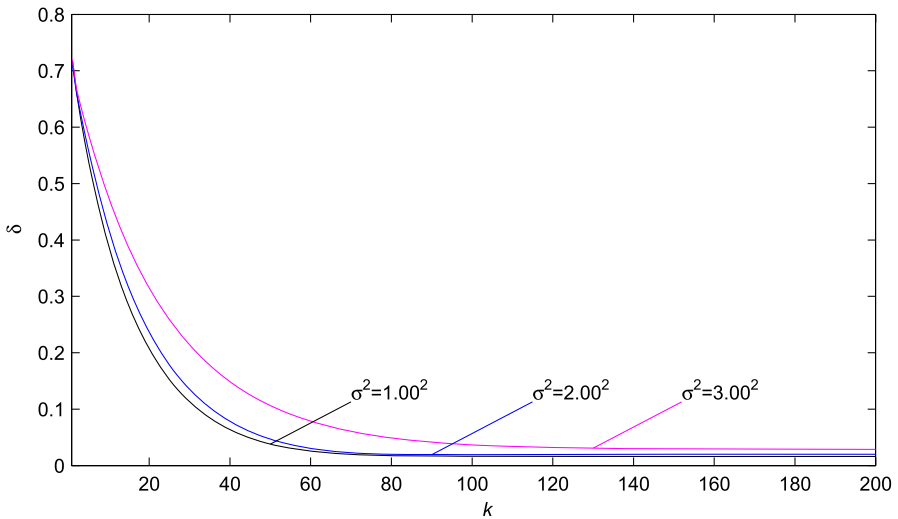


Fig. 3 The H-GI parameter estimation errors δ versus k with different σ^2

- (1) Set the input signal $\{r(t)\}$ with a persistent excitation sequence with zero mean and unit variance.
- (2) Set the white noise sequence $\{v(t)\}$ with zero mean and variances σ^2 .
- (3) To perform the following identification algorithm simulation, the corresponding output signals are generated based on the given model parameters.
- (4) Set the first $L_e = 3000$ as the data length for parameter estimation and the remaining $L_r = 200$ as the data length for model validation.

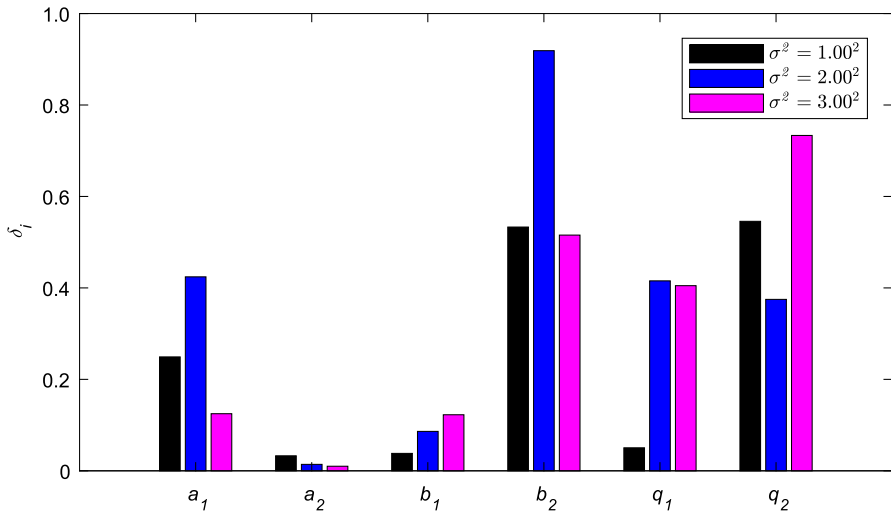


Fig. 4 The GI estimation errors δ_i under different σ^2

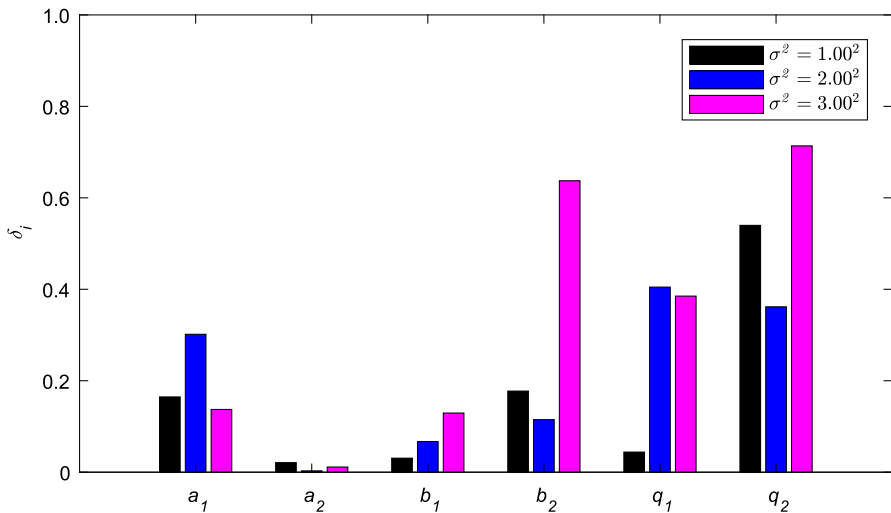


Fig. 5 The H-GI estimation errors δ_i under different σ^2

Take the noise variances $\sigma^2 = 1.00^2$, $\sigma^2 = 2.00^2$ and $\sigma^2 = 3.00^2$, respectively, and apply the GI and H-GI algorithms to estimate the parameters of this example system. The parameter estimates and the corresponding errors are given in Tables 3, 4, and the parameter estimation errors $\delta := \|\hat{\vartheta}_k - \vartheta\| / \|\vartheta\|$ versus k are shown in Figs. 2, 3. In addition, to show the estimation results of each parameter under each algorithm more clearly, define $\delta_i := |\hat{\vartheta}_{i,k} - \vartheta_i| / |\vartheta_i|$. Hence, under different noise variances σ^2 , the estimation errors δ_i of each parameter at $k = 200$ as shown in Figs. 4, 5.

Apply the H-GI algorithm with the noise variance $\sigma^2 = 1.00^2$ and the data lengths $L = 1000$, $L = 2000$ and $L = 3000$ to identify this example system, respectively.

Table 5 The H-GI parameter estimates and errors ($\sigma^2 = 1.00^2$)

L	k	a_1	a_2	b_1	b_2	q_1	q_2	δ (%)	
1000	1	0.84191	0.84968	1.05997	0.94492	0.89082	0.62616	71.34583	
	2	0.74276	0.78467	1.13394	0.88525	0.88114	0.61449	65.71028	
	5	0.63991	0.72522	1.29263	0.71238	0.75447	0.54024	54.16166	
	10	0.56356	0.70696	1.46819	0.45851	0.77782	0.55439	39.64039	
	20	0.46855	0.68746	1.61432	0.12807	0.79250	0.55568	22.31405	
	50	0.37494	0.66669	1.67527	-0.19771	0.80164	0.55553	6.32067	
	100	0.35875	0.66296	1.67631	-0.25415	0.80398	0.55650	3.81430	
	200	0.35802	0.66279	1.67622	-0.25671	0.80412	0.55657	3.70972	
	2000	1	0.83950	0.85891	1.06127	0.95023	0.96018	0.67279	71.89707
		2	0.76445	0.78694	1.13152	0.88490	0.90332	0.65629	66.16827
5		0.64727	0.73719	1.28831	0.70600	0.81059	0.57338	53.98981	
10		0.57031	0.71306	1.45975	0.44487	0.82385	0.58256	39.20211	
20		0.47423	0.68658	1.60491	0.10804	0.81792	0.57612	21.49482	
50		0.37886	0.66037	1.67480	-0.22916	0.80355	0.56569	4.82556	
100		0.36108	0.65550	1.67892	-0.29257	0.80133	0.56430	2.20989	
200		0.36014	0.65524	1.67900	-0.29593	0.80124	0.56425	2.10995	
3000		1	0.83977	0.85614	1.05994	0.94877	0.95027	0.66377	71.78404
		2	0.76551	0.78663	1.12830	0.88430	0.86000	0.65560	66.09631
	5	0.64674	0.73384	1.28394	0.70106	0.80147	0.56758	53.83765	
	10	0.56889	0.71000	1.45400	0.43495	0.81699	0.58098	38.80842	
	20	0.47081	0.68323	1.60005	0.09078	0.81468	0.57935	20.69716	
	50	0.37395	0.65642	1.67719	-0.24980	0.80000	0.56977	3.80303	
	100	0.35662	0.65160	1.68481	-0.31127	0.79666	0.56772	1.63785	
	200	0.35576	0.65136	1.68515	-0.31432	0.79648	0.56761	1.60767	
	True values		0.35000	0.65000	1.68000	-1.32000	0.80000	0.60000	

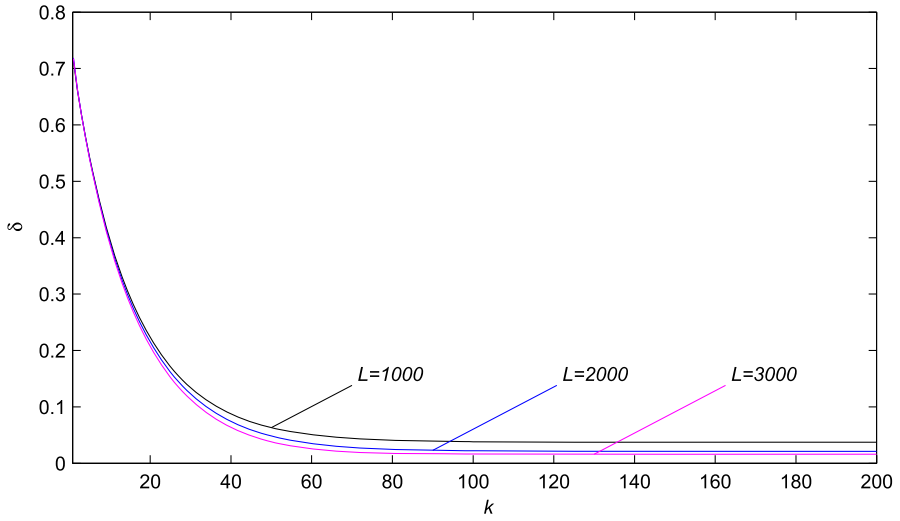


Fig. 6 The H-GI parameter estimation errors δ versus k with different data length L ($\sigma^2 = 1.00^2$)

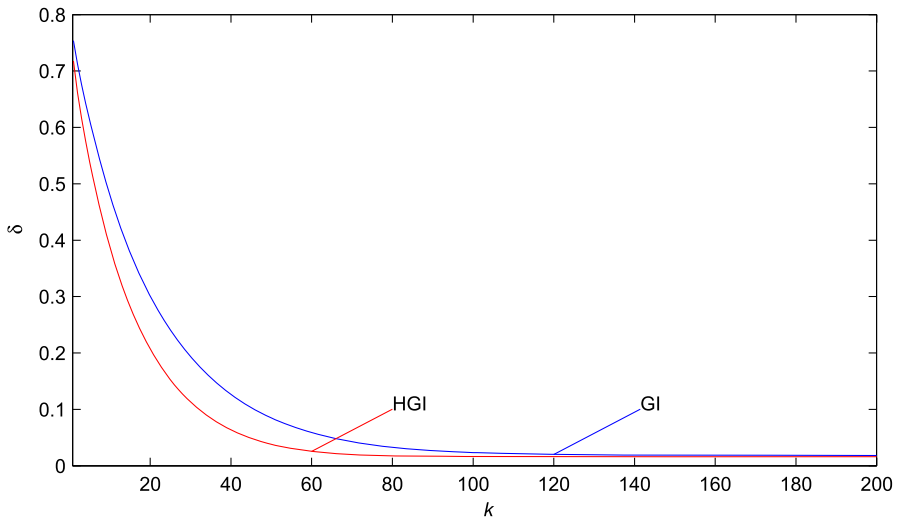


Fig. 7 The GI and H-GI estimation errors δ versus k ($\sigma^2 = 1.00^2$)

The parameter estimates and the corresponding errors are shown in Table 5, and the parameter estimation errors δ versus k are shown in Fig. 6.

Taking the noise variance $\sigma^2 = 1.00^2$, the parameter estimation errors δ versus k with different algorithms are plotted in Fig. 7. The GI and H-GI estimates of the parameters a_1, a_2, b_1, b_2, q_1 and q_2 versus k are shown in Figs. 8, 9.

For model validation, we use the remaining $L_r = 200$ observations from $t = L_e + 1$ to $t = L_e + L_r$ and the models predicted by the GI and H-GI algorithms with $\sigma^2 =$

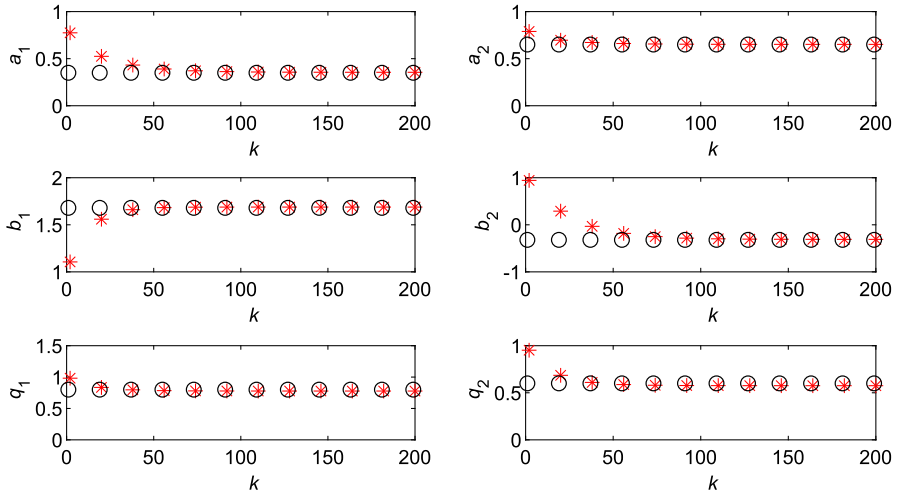


Fig. 8 The GI parameter estimates versus k ($\sigma^2 = 1.00^2$). Asterisk: the parameter estimates, Circle: the true values

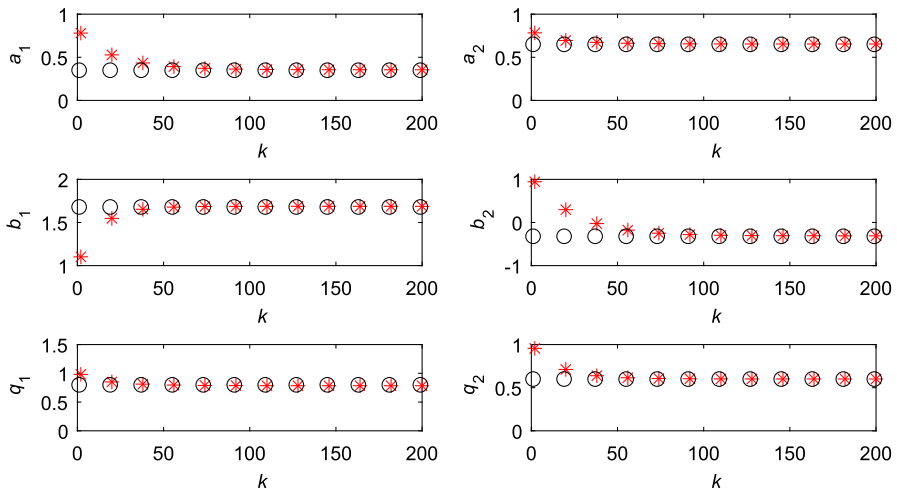


Fig. 9 The H-GI parameter estimates versus k ($\sigma^2 = 1.00^2$). Asterisk: the parameter estimates, Circle: the true values

1.00^2 . The predicted outputs $\hat{y}_i(t)$ ($i = 1, 2$), the actual outputs $y(t)$ and their errors are plotted in Figs. 10, 11.

The predicted outputs are

$$\hat{y}_i(t) = [1 - \hat{A}(z)]y(t) + \hat{B}(z)[r(t) - \hat{y}(t)]. \tag{48}$$

Use the GI estimates in Table 3 with the noise variance $\sigma^2 = 1.00^2$ and $k = 200$ to construct the GI estimated model:

$$\begin{aligned}\hat{y}_1(t) &= [1 - \hat{A}(z)]y(t) + \hat{B}(z)[r(t) - \hat{y}(t)], \\ \hat{A}(z) &= 1 + 0.35872z^{-1} + 0.65213z^{-2}, \\ \hat{B}(z) &= 1.68640z^{-1} - 0.30294z^{-2}, \\ \hat{y}(t) &= 0.79598 \sin(y^2(t)) + 0.56727 \sin(y^3(t)), \\ \hat{\boldsymbol{\theta}} &= [\hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2, \hat{q}_1, \hat{q}_2]^T \\ &= [0.35872, 0.65213, 1.68640, -0.30294, 0.79598, 0.56727]^T.\end{aligned}$$

Use the H-GI estimates in Table 4 with the noise variance $\sigma^2 = 1.00^2$ and $k = 200$ to construct the H-GI estimated model:

$$\begin{aligned}\hat{y}_2(t) &= [1 - \hat{A}(z)]y(t) + \hat{B}(z)[r(t) - \hat{y}(t)], \\ \hat{A}(z) &= 1 + 0.35576z^{-1} + 0.65136z^{-2}, \\ \hat{B}(z) &= 1.68515z^{-1} - 0.31432z^{-2}, \\ \hat{y}(t) &= 0.79648 \sin(y^2(t)) + 0.56761 \sin(y^3(t)), \\ \hat{\boldsymbol{\theta}} &= [\hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2, \hat{q}_1, \hat{q}_2]^T \\ &= [0.35576, 0.65136, 1.68515, -0.31432, 0.79648, 0.56761]^T.\end{aligned}$$

To evaluate the prediction performance, we define and compute the root-mean-square errors (RMSEs):

$$\begin{aligned}Error(\hat{y}_1(t)) &:= \left[\frac{1}{L_r} \sum_{t=L_e+1}^{L_e+L_r} [\hat{y}_1(t) - y(t)]^2 \right]^{\frac{1}{2}} = 1.00249, \\ Error(\hat{y}_2(t)) &:= \left[\frac{1}{L_r} \sum_{t=L_e+1}^{L_e+L_r} [\hat{y}_2(t) - y(t)]^2 \right]^{\frac{1}{2}} = 1.00206.\end{aligned}$$

From Tables 3, 4, 5 and Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, we can draw the following conclusions.

- Throughout the entire simulation process, shown in Tables 3, 4 and Figs. 2, 3, 4, 5, we find that the parameter estimation errors of the GI and H-GI algorithms decrease as the iteration number k increases, and the estimation accuracy improves as the noise level reduces. However, the H-GI algorithm yields more exact estimates than the GI algorithm under the same iteration and noise variance—see Fig. 7.
- For the same noise variance, the parameter estimation errors of the H-GI algorithm become smaller as the data length L and iteration k increase—see Table 5 and Fig. 6.

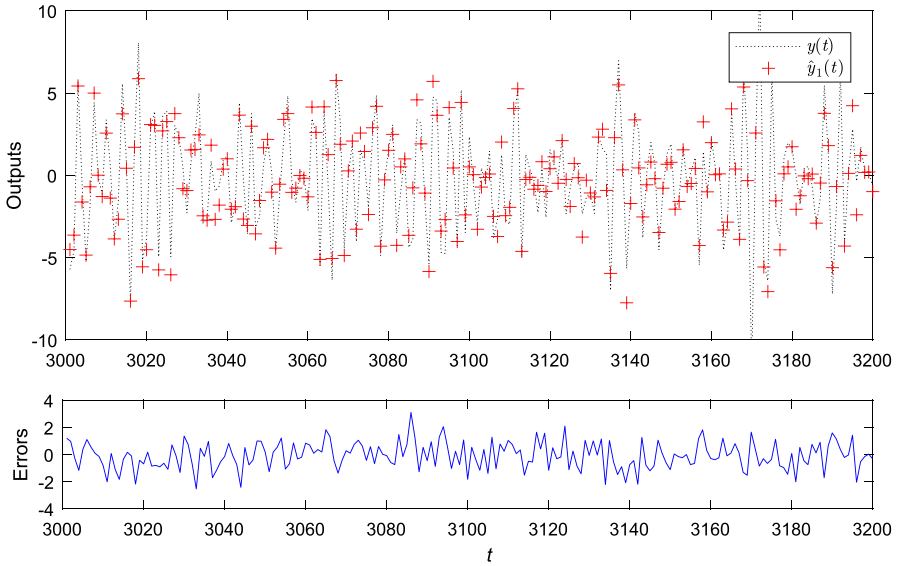


Fig. 10 The predicted outputs $\hat{y}_1(t)$, the actual outputs $y(t)$ and their errors versus t based on the GI estimates

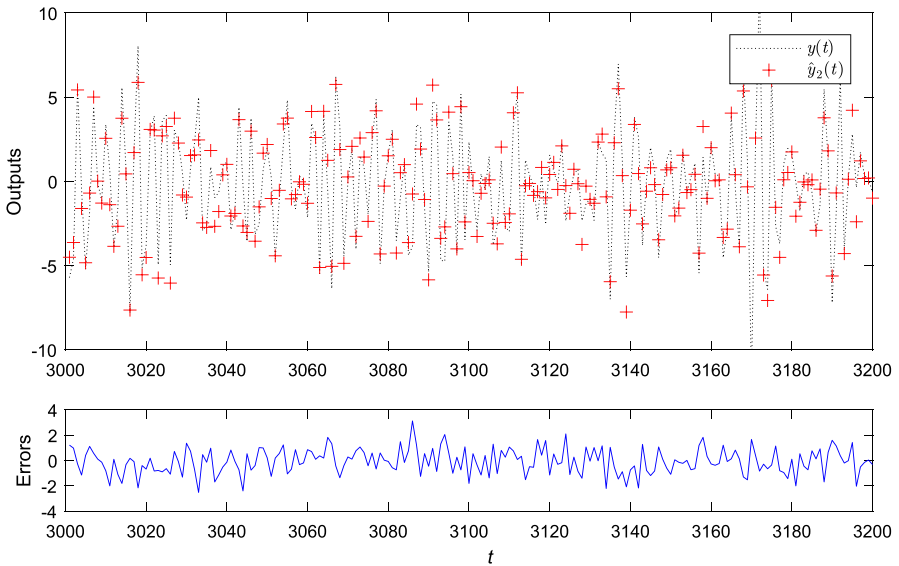


Fig. 11 The predicted outputs $\hat{y}_2(t)$, the actual outputs $y(t)$ and their errors versus t based on the H-GI estimates

- For sufficiently large data length, the parameter estimates of the GI and H-GI algorithms gradually approach the true values with the increase of the iteration k , which verifies the effectiveness of the GI and H-GI algorithms—see Figs. 8, 9.
- The predicted outputs of the GI and H-GI algorithms verge on the true values and the differences between them are small—see Figs. 10, 11. This shows that the estimated models based on the GI and H-GI algorithms have good prediction performance and can acquire the dynamics of the feedback nonlinear systems well.

6 Conclusions

This paper studies the parameter identification problem for nonlinear feedback system. By negative gradient search, a GI algorithm is derived for estimating the unknown parameters. Inspired by the hierarchical identification principle, the nonlinear feedback system is decomposed into two subsystems and an H-GI algorithm is proposed for the nonlinear feedback system. Compared with the GI algorithm, the H-GI algorithm has a higher computational efficiency and a comparable parameter estimation accuracy. The simulation results demonstrate the performance of the proposed algorithms. The approaches presented in this paper can be extended to investigate the modeling and optimization of some production and process systems [22, 88, 107, 108] and can be applied to other control and schedule areas [3–12] such as information processing systems and transportation communication systems [71–78] and so on.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 62273167, 62076110), the 111 Project (B23008), the Natural Science Foundation of Shanghai (No. 22ZR1445300) and the Academic Research Fund (Applied Technology) Project (CFK201808) of Changzhou Vocational Institute of Textile and Garment and the Fundamental Research Funds for the Central Universities under Grant (JUSRP221027).

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

Declarations

Conflict of interest The authors declare that they have no conflicts of interest

References

1. M.R.H. Abdalmoaty, H. Hjalmarsson, Identification of stochastic nonlinear models using optimal estimating functions. *Automatica* **119**, 109055 (2020)
2. V. Bezzubov, A. Bobtsov, D. Efimov, R. Ortega, N. Nikolaev, Adaptive state observation of linear time-varying systems with delayed measurements and unknown parameters. *Int. J. Robust Nonlinear Control* **33**(2), 1203–1213 (2023)
3. Y. Cao, Y. An, S. Su et al., A statistical study of railway safety in China and Japan 1990–2020. *Accid. Anal. Prevent.* **175**, 106764 (2022)
4. Y. Cao, Y.S. Ji, Y.K. Sun, S. Su, The fault diagnosis of a switch machine based on deep random forest fusion. *IEEE Intell. Transp. Syst. Mag.* **15**(1), 437–452 (2023)
5. Y. Cao, L.C. Ma, S. Xiao et al., Standard analysis for transfer delay in CTCS-3. *Chinese J. Electron.* **26**(5), 1057–1063 (2017)

6. Y. Cao, Y.K. Sun, G. Xie, P. Li, 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.* **23**(8), 12074–12083 (2022)
7. Y. Cao, Y.K. Sun, G. Xie, T. Wen, Fault diagnosis of train plug door based on a hybrid criterion for IMF's selection and fractional wavelet package energy entropy. *IEEE Trans. Veh. Technol.* **68**(8), 7544–7551 (2019)
8. Y. Cao, Z. Wang, F. Liu, P. Li, G. Xie, Bio-inspired speed curve optimization and sliding mode tracking control for subway trains. *IEEE Trans. Veh. Technol.* **68**(7), 6331–6342 (2019)
9. Y. Cao, J.K. Wen, A. Hobiny, P. Li, T. Wen, Parameter-varying artificial potential field control of virtual coupling system with nonlinear dynamics. *Fractals* **30**(2), 2240099 (2022)
10. Y. Cao, J.K. Wen, L.C. Ma, Tracking and collision avoidance of virtual coupling train control system. *Alex. Eng. J.* **60**(2), 2115–2125 (2021)
11. Y. Cao, Y.R. Yang, L.C. Ma et al., Research on virtual coupled train control method based on GPC & VAPF. *Chinese J. Electron.* **31**(5), 897–905 (2022)
12. Y. Cao, Z.X. Zhang, F.L. Cheng, S. Su, Trajectory optimization for high-speed trains via a mixed integer linear programming approach. *IEEE Trans. Intell. Transp. Syst.* **23**(10), 17666–17676 (2022)
13. T. Cui, Moving data window-based partially-coupled estimation approach for modeling a dynamical system involving unmeasurable states. *ISA Trans.* **128**, 437–452 (2022)
14. F. Ding, Least squares parameter estimation and multi-innovation least squares methods for linear fitting problems from noisy data. *J. Comput. Appl. Math.* **426**, 115107 (2023)
15. F. Ding, H. Ma, J. Pan, E.F. Yang, Hierarchical gradient- and least squares-based iterative algorithms for input nonlinear output-error systems using the key term separation. *J. Franklin Inst.* **358**(9), 5113–5135 (2021)
16. F. Ding, L. Xu, X. Zhang, Y.H. Zhou, Filtered auxiliary model recursive generalized extended parameter estimation methods for Box-Jenkins systems by means of the filtering identification idea. *Int. J. Robust Nonlinear Control* **33**(10), 5510–5535 (2023)
17. J.L. Ding, W.H. Zhang, Finite-time adaptive control for nonlinear systems with uncertain parameters based on the command filters. *Int. J. Adapt. Control Signal Process.* **35**(9), 1754–1767 (2021)
18. Y.M. Fan, X.M. Liu, Two-stage auxiliary model gradient-based iterative algorithm for the input nonlinear controlled autoregressive system with variable-gain nonlinearity. *Int. J. Robust Nonlinear Control* **30**(14), 5492–5509 (2020)
19. Y.M. Fan, X.M. Liu, 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.* **36**(3), 521–540 (2022)
20. R. Goel, T. Garg, S.B. Roy, Closed-Loop reference model based distributed MRAC using cooperative initial excitation and distributed reference input estimation. *IEEE Trans. Control Net. Syst.* **9**(1), 37–49 (2021)
21. R. Goel, S.B. Roy, Closed-loop reference model based distributed model reference adaptive control for multi-agent systems. *IEEE Control Syst. Lett.* **5**(5), 1837–1842 (2021)
22. Y. Gu, Q.M. Zhu, H. Nouri, Identification and U-control of a state-space system with time-delay. *Int. J. Adapt. Control Signal Process.* **36**(1), 138–154 (2022)
23. C.S. Hems, C.M. Panazio, Sparse estimation technique for digital pre-distortion of impedance-mismatched power amplifiers. *Circuits Syst. Signal Process.* **40**(8), 3727–3755 (2021)
24. J. Hou, F. Chen, P. Li et al., Gray-box parsimonious subspace identification of Hammerstein-type systems. *IEEE Trans. Ind. Electron.* **68**(10), 9941–9951 (2021)
25. J. Hou, H. Su, C. Yu et al., Bias-correction errors-in-variables Hammerstein model identification. *IEEE Trans. Ind. Electron.* **70**(7), 7268–7279 (2023)
26. J. Hou, H. Su, C.P. Yu et al., Consistent subspace identification of errors-in-variables Hammerstein systems. *IEEE Trans. Syst. Man Cybern. Syst.* **53**(4), 2292–2303 (2023)
27. P.P. Hu, F. Ding, J. Sheng, Auxiliary model based least squares parameter estimation algorithm for feedback nonlinear systems using the hierarchical identification principle. *J. Frankl. Inst.* **350**(10), 3248–3259 (2013)
28. C. Hu, Y. Ji, C.Q. Ma, Joint two-stage multi-innovation recursive least squares parameter and fractional-order estimation algorithm for the fractional-order input nonlinear output-error autoregressive model. *Int. J. Adapt. Control Signal Process.* **37**(7), 1650–1670 (2023)
29. G. Huang, A. Liu, M.J. Zhao, Two-stage adaptive and compressed CSI feedback for FDD massive MIMO. *IEEE Trans. Veh. Technol.* **70**(9), 9602–9606 (2021)

30. Y. Ji, A.N. Jiang, Filtering-based accelerated estimation approach for generalized time-varying systems with disturbances and colored noises. *IEEE Trans. Circuits Syst. II Express Briefs.* **70**(1), 206–210 (2023)
31. Y. Ji, X.K. Jiang, L.J. Wan, Hierarchical least squares parameter estimation algorithm for two-input Hammerstein finite impulse response systems. *J. Frankl. Inst.* **357**(8), 5019–5032 (2020)
32. Y. Ji, Z. Kang, Three-stage forgetting factor stochastic gradient parameter estimation methods for a class of nonlinear systems. *Int. J. Robust Nonlinear Control* **31**(3), 971–987 (2021)
33. Y. Ji, Z. Kang, X.M. Liu, The data filtering based multiple-stage Levenberg-Marquardt algorithm for Hammerstein nonlinear systems. *Int. J. Robust Nonlinear Control* **31**(15), 7007–7025 (2021)
34. Y. Ji, Z. Kang, C. Zhang, Two-stage gradient-based recursive estimation for nonlinear models by using the data filtering. *Int. J. Control Autom. Syst.* **19**(8), 2706–2715 (2021)
35. Y. Ji, C. Zhang, Z. Kang, T. Yu, Parameter estimation for block-oriented nonlinear systems using the key term separation. *Int. J. Robust Nonlinear Control* **30**(9), 3727–3752 (2020)
36. A.N. Jiang, Y. Ji, L.J. Wan, Iterative parameter identification algorithms for the generalized time-varying system with a measurable disturbance vector. *Int. J. Robust Nonlinear Control* **32**(6), 3527–3548 (2020)
37. Z. Kang, Y. Ji, X.M. Liu, Hierarchical recursive least squares algorithms for Hammerstein nonlinear autoregressive output-error systems. *Int. J. Adapt. Control Signal Process.* **35**(11), 2276–2295 (2021)
38. J.M. Li, A novel nonlinear optimization method for fitting a noisy Gaussian activation function. *Int. J. Adapt. Control Signal Process.* **36**(3), 690–707 (2022)
39. M.H. Li, X.M. Liu, Maximum likelihood hierarchical least squares-based iterative identification for dual-rate stochastic systems. *Int. J. Adapt. Control Signal Process.* **35**(2), 240–261 (2021)
40. M.H. Li, X.M. Liu, Iterative identification methods for a class of bilinear systems by using the particle filtering technique. *Int. J. Adapt. Control Signal Process.* **35**(10), 2056–2074 (2021)
41. M.H. Li, X.M. Liu, Particle filtering-based iterative identification methods for a class of nonlinear systems with interval-varying measurements. *Int. J. Control Autom. Syst.* **20**(7), 2239–2248 (2022)
42. M.H. Li, X.M. Liu, The filtering-based maximum likelihood iterative estimation algorithms for a special class of nonlinear systems with autoregressive moving average noise using the hierarchical identification principle. *Int. J. Adapt. Control Signal Process.* **33**(7), 1189–1211 (2019)
43. L.W. Li, F.X. Wang, H.L. Zhang, X.M. Ren, Adaptive model recovery scheme for multivariable system using error correction learning. *IEEE Trans. Instrum. Meas.* **70**, 3108569 (2021)
44. L.H. Li, G.C. Yang, Y. Li et al., Abnormal sitting posture recognition based on multi-scale spatiotemporal features of skeleton graph. *Eng. Appl. Artif. Intell.* **123**, 106374 (2023)
45. Y. Li, G. Yang, Z. Su, Y. Wang, Human activity recognition based on multienvironment sensor data. *Inf. Fusion* **91**, 47–63 (2023)
46. J.H. Li, T.C. Zong, G.P. Lu, Parameter identification of Hammerstein-Wiener nonlinear systems with unknown time delay based on the linear variable weight particle swarm optimization. *ISA Trans.* **120**, 89–98 (2022)
47. Q.L. Liu, Recursive least squares estimation methods for a class of nonlinear systems based on non-uniform sampling. *Int. J. Adapt. Control Signal Process.* **35**(8), 1612–1632 (2021)
48. Q.L. Liu, Gradient-based recursive parameter estimation for a periodically nonuniformly sampled-data Hammerstein-Wiener system based on the key-term separation. *Int. J. Adapt. Control Signal Process.* **35**(10), 1970–1989 (2021)
49. S.Y. Liu, Hierarchical principle-based iterative parameter estimation algorithm for dual-frequency signals. *Circuits Syst Signal Process.* **38**(7), 3251–3268 (2019)
50. X.M. Liu, Y.M. Fan, Maximum likelihood extended gradient-based estimation algorithms for the input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity. *Int. J. Robust Nonlinear Control* **31**(9), 4017–4036 (2021)
51. H.B. Liu, J.W. Wang, 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. *Int. J. Control Autom. Syst.* **20**(8), 2606–2615 (2022)
52. P. Ma, New gradient based identification methods for multivariate pseudo-linear systems using the multi-innovation and the data filtering. *J. Franklin Inst.* **354**(3), 1568–1583 (2017)
53. J.X. Ma, Filtering-based multistage recursive identification algorithm for an input nonlinear output-error autoregressive system by using the key term separation technique. *Circuits Syst. Signal Process.* **36**(2), 577–599 (2017)

54. H. Ma, J. Pan, W. Ding, Partially-coupled least squares based iterative parameter estimation for multi-variable output-error-like autoregressive moving average systems. *IET Control Theory Appl.* **13**(18), 3040–3051 (2019)
55. P. Ma, L. Wang, Filtering-based recursive least squares estimation approaches for multivariate equation-error systems by using the multiinnovation theory. *Int. J. Adapt. Control Signal Process.* **35**(9), 1898–1915 (2021)
56. J.X. Ma, W.L. Xiong, J. Chen, Hierarchical identification for multivariate Hammerstein systems by using the modified Kalman filter. *IET Control Theory Appl.* **11**(6), 857–869 (2017)
57. Y.W. Mao, C. Xu, J. Chen, Y. Pu, Q.Y. Hu, Auxiliary model-based iterative estimation algorithms for nonlinear systems using the covariance matrix adaptation strategy. *Circuits Syst. Signal Process.* **41**(12), 6750–6773 (2022)
58. S. Marzougui, S. Bedoui, A. Atitallah, K. Abderrahim, Parameter and state estimation of nonlinear fractional-order model using Luenberger observer. *Circuits Syst. Signal Process.* **41**(10), 5366–5391 (2022)
59. X. Meng, Y. Ji, J. Wang, Iterative parameter estimation for photovoltaic cell models by using the hierarchical principle. *Int. J. Control Autom. Syst.* **20**(8), 2583–2593 (2022)
60. J. Pan, Q. Chen, J. Xiong, G. Chen, A novel quadruple-boost nine-level switched capacitor inverter. *J. Electr. Eng. Technol.* **18**(1), 467–480 (2023)
61. J. Pan, X. Jiang, X.K. Wan, W. Ding, A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems. *Int. J. Control Autom. Syst.* **15**(3), 1189–1197 (2017)
62. J. Pan, W. Li, H.P. Zhang, Control algorithms of magnetic suspension systems based on the improved double exponential reaching law of sliding mode control. *Int. J. Control Autom. Syst.* **16**(6), 2878–2887 (2018)
63. J. Pan, Y.Q. Liu, J. Shu, Gradient-based parameter estimation for an exponential nonlinear autoregressive time-series model by using the multi-innovation. *Int. J. Control Autom. Syst.* **21**(1), 140–150 (2023)
64. J. Pan, S.D. Liu, J. Shu, X.K. Wan, Hierarchical recursive least squares estimation algorithm for secondorder Volterra nonlinear systems. *Int. J. Control Autom. Syst.* **20**(12), 3940–3950 (2022)
65. J. Pan, H. Ma, X. Zhang et al., Recursive coupled projection algorithms for multivariable output-error-like systems with coloured noises. *IET Signal Process.* **14**(7), 455–466 (2020)
66. J. Pan, B. Shao, J.X. Xiong, Q. Zhang, Attitude control of quadrotor UAVs based on adaptive sliding mode. *Int. J. Control Autom. Syst.* **21**(8), 2698–2707 (2023)
67. J. Pan, H. Zhang, H. Guo, S. Liu, Y. Liu, Multivariable CAR-like system identification with multi-innovation gradient and least squares algorithms. *Int. J. Control Autom. Syst.* **21**(5), 1455–1464 (2023)
68. Y. Pu, Y.J. Rong, J. Chen, Y.W. Mao, Accelerated identification algorithms for exponential nonlinear models: two-stage method and particle swarm optimization method. *Circuits Syst. Signal Process.* **41**(5), 2636–2652 (2022)
69. F. Shahriari, M.M. Arefi, H. Luo, S. Yin, Multistage parameter estimation algorithms for identification of bilinear systems. *Nonlinear Dyn.* **110**(3), 2635–2655 (2022)
70. B.B. Shen, F. Ding, L. Xu, T. Hayat, Data filtering based multi-innovation gradient identification methods for feedback nonlinear systems. *Int. J. Control Autom. Syst.* **16**(5), 2225–2234 (2018)
71. S. Su, J. She, K. Li et al., A nonlinear safety equilibrium spacing based model predictive control for virtually coupled train set over gradient terrains. *IEEE Trans. Transp. Electrification.* **8**(2), 2810–2824 (2022)
72. S. Su, T. Tang, J. Xun et al., Design of running grades for energy-efficient train regulation: a case study for Beijing Yizhuang line. *IEEE Intell. Transp. Syst. Mag.* **13**(2), 189–200 (2021)
73. S. Su, X. Wang, Y. Cao, J.T. Yin, An energy-efficient train operation approach by integrating the metro timetabling and eco-driving. *IEEE Trans. Intell. Transp. Syst.* **21**(10), 4252–4268 (2020)
74. S. Su, X. Wang, T. Tang et al., Energy-efficient operation by cooperative control among trains: a multi-agent reinforcement learning approach. *Control Eng. Pract.* **116**, 104901 (2021)
75. S. Su, Q. Zhu, J. Liu et al., Eco-driving of trains with a data-driven iterative learning approach. *IEEE Trans. Ind. Inf.* (2023). <https://doi.org/10.1109/TII.2022.3195888>
76. Y.K. Sun, Y. Cao, P. Li, Contactless fault diagnosis for railway point machines based on multi-scale fractional wavelet packet energy entropy and synchronous optimization strategy. *IEEE Trans. Veh. Technol.* **71**(6), 5906–5914 (2022)

77. Y.K. Sun, Y. Cao, L.C. Ma, A fault diagnosis method for train plug doors via sound signals. *IEEE Intell. Transp. Syst. Mag.* **13**(3), 107–117 (2021)
78. Y.K. Sun, Y. Cao, G. Xie, T. Wen, Sound based fault diagnosis for RPMs based on multi-scale fractional permutation entropy and two-scale algorithm. *IEEE Trans. Veh. Technol.* **70**(11), 11184–11192 (2021)
79. S.S. Tabatabaei, M. Tavakoli, H.A. Talebi, A finite-time adaptive order estimation approach for non-integer order nonlinear systems. *ISA Trans.* **127**, 383–394 (2022)
80. J.W. Wang, Y. Ji, C. Zhang, Iterative parameter and order identification for fractional-order nonlinear finite impulse response systems using the key term separation. *Int. J. Adapt. Control Signal Process.* **35**(8), 1562–1577 (2021)
81. X.H. Wang, Modified particle filtering-based robust estimation for a networked control system corrupted by impulsive noise. *Int. J. Robust Nonlinear Control* **32**(2), 830–850 (2022)
82. Y.J. Wang, Recursive parameter estimation algorithm for multivariate output-error systems. *J. Franklin Inst.* **355**(12), 5163–5181 (2018)
83. J.W. Wang, Y. Ji, X. Zhang, L. Xu, Two-stage gradient-based iterative algorithms for the fractional-order nonlinear systems by using the hierarchical identification principle. *Int. J. Adapt. Control Signal Process.* **36**(7), 1778–1796 (2022)
84. H. Wang, G. Ke, J. Pan, Q. Su, Modeling, dynamical analysis and numerical simulation of a new 3D cubic Lorenz-like system. *Sci. Rep.* **13**, 6671 (2023)
85. H. Wang, G. Ke, J. Pan, F. Hu, H. Fan, Q. Su, Two pairs of heteroclinic orbits coined in a new sub-quadratic Lorenz-like system. *Eur. Phys. J. B* **96**(3), 28 (2023)
86. H. Wang, G. Ke, J. Pan, Q. Su, Conjoined Lorenz-like attractors coined. *Miskolc Math. Note* (2023)
87. H. Wang, G. Ke, J. Pan, Q. Su, G. Dong, H. Fan, Revealing the true and pseudo-singularly degenerate heteroclinic cycles. *Indian J. Phys.* (2023). <https://doi.org/10.1007/s12648-023-02689-w>
88. X. Wang, S. Su, Y. Cao, X.L. Wang, Robust control for dynamic train regulation in fully automatic operation system under uncertain wireless transmissions. *IEEE Trans. Intell. Transp. Syst.* **23**(11), 20721–20734 (2022)
89. Y.J. Wang, S.H. Tang, M.Q. Deng, Modeling nonlinear systems using the tensor network B-spline and the multi-innovation identification theory. *Int. J. Robust Nonlinear Control* **32**(13), 7304–7318 (2022)
90. Y.J. Wang, S.H. Tang, X.B. Gu, Parameter estimation for nonlinear Volterra systems by using the multi-innovation identification theory and tensor decomposition. *J. Franklin Inst.* **359**(2), 1782–1802 (2022)
91. Y.J. Wang, L. Yang, An efficient recursive identification algorithm for multilinear systems based on tensor decomposition. *Int. J. Robust Nonlinear Control* **31**(16), 7920–7936 (2021)
92. Y. Wang, G. Yang, Arrhythmia classification algorithm based on multi-head self-attention mechanism. *Biomed. Signal Process. Control* **79**, 104206 (2023)
93. C. Wei, Overall recursive least squares and overall stochastic gradient algorithms and their convergence for feedback nonlinear controlled autoregressive systems. *Int. J. Robust Nonlinear Control* **32**(9), 5534–5554 (2022)
94. J.X. Xiong, J. Pan, G.Y. Chen et al., Sliding mode dual-channel disturbance rejection attitude control for a quadrotor. *IEEE Trans. Ind. Electron.* **69**(10), 10489–10499 (2022)
95. H. Xu, Joint parameter and time-delay estimation for a class of nonlinear time-series models. *IEEE Signal Process. Lett.* **29**, 947–951 (2022)
96. L. Xu, Separable multi-innovation Newton iterative modeling algorithm for multi-frequency signals based on the sliding measurement window. *Circuits Syst. Signal Process.* **41**(2), 805–830 (2022)
97. L. Xu, Parameter estimation for nonlinear functions related to system responses. *Int. J. Control Autom. Syst.* **21**(6), 1780–1792 (2023)
98. L. Xu, Separable Newton recursive estimation method through system responses based on dynamically discrete measurements with increasing data length. *Int. J. Control Autom. Syst.* **20**(2), 432–443 (2022)
99. L. Xu, Hierarchical recursive signal modeling for multi-frequency signals based on discrete measured data. *Int. J. Adapt. Control Signal Process.* **35**(5), 676–693 (2021)
100. L. Xu, Separable synthesis estimation methods and convergence analysis for multivariable systems. *J. Comput. Appl. Math.* **427**, 115104 (2023)
101. C. Xu, Y. Qin, H. Su, Observer-based dynamic event-triggered bipartite consensus of discrete-time multi-agent systems. *IEEE Trans. Circuits Syst. II: Express Briefs* **70**(3), 1054–1058 (2023)

102. L. Xu, G.L. Song, A recursive parameter estimation algorithm for modeling signals with multi-frequencies. *Circuits Syst. Signal Process.* **39**(8), 4198–4224 (2020)
103. C. Xu, H. Xu, Z.H. Guan, Y. Ge, Observer-based dynamic event-triggered semi-global bipartite consensus of linear multi-agent systems with input saturation. *IEEE Trans. Cybern.* **53**(5), 3139–3152 (2023)
104. D. Yang, F. Ding, Multi-innovation gradient-based iterative identification methods for feedback nonlinear systems by using the decomposition technique. *Int. J. Robust Nonlinear Control* **33**(13), 7755–7773 (2023)
105. G. Yang, S. Li, L. He, Short-term prediction method of blood glucose based on temporal multi-head attention mechanism for diabetic patients. *Biomed. Signal Process. Control* **82**, 104552 (2023)
106. G. Yang, S. Yang, K. Luo, S. La, L. He, Y. Li, Detection of non-suicidal self-injury based on spatiotemporal features of indoor activities. *IET Biometrics* **12**, 91–101 (2023)
107. J. You, C. Yu, J. Sun, J. Chen, Generalized maximum entropy based identification of graphical ARMA models. *Automatica* **141**, 110319 (2022)
108. C. Yu, Y. Li, H. Fang, J. Chen, System identification approach for inverse optimal control of finite-horizon linear quadratic regulators. *Automatica* **129**, 109636 (2021)
109. X. Zhang, Hierarchical parameter and state estimation for bilinear systems. *Int. J. Syst. Sci.* **51**(2), 275–290 (2020)
110. C. Zhang, H.B. Liu, Y. Ji, Gradient parameter estimation of a class of nonlinear systems based on the maximum likelihood principle. *Int. J. Control Autom. Syst.* **20**(5), 1393–1404 (2022)
111. E.L. Zhang, R. Pintelon, Identification of dynamic errors-in-variables systems with quasi-stationary input and colored noise. *Automatica* **123**, 109344 (2021)
112. T.Y. Zhang, S.Y. Zhao, X.L. Luan, F. Liu, Bayesian inference for state-space models with student-t mixture distributions. *IEEE Trans. Cybern.* **53**(7), 4435–4445 (2023)
113. S.Y. Zhao, B. Huang, Trial-and-error or avoiding a guess? Initialization of the Kalman filter. *Automatica* **121**, 109184 (2020)
114. S.Y. Zhao, B. Huang, C.H. Zhao, Online probabilistic estimation of sensor faulty signal in industrial processes and its applications. *IEEE Trans. Ind. Electron.* **68**(9), 8858–8862 (2021)
115. S.Y. Zhao, K. Li, C. Ahn, B. Huang, F. Liu, Tuning-free Bayesian estimation algorithms for faulty sensor signals in state-space. *IEEE Trans. Ind. Electron.* **70**(1), 921–929 (2023)
116. S.Y. Zhao, Y.S. Shmaliy, C.K. Ahn, F. Liu, Self-tuning unbiased finite impulse response filtering algorithm for processes with unknown measurement noise covariance. *IEEE Trans. Control Syst. Technol.* **29**(3), 1372–1379 (2021)
117. S.Y. Zhao, Y.S. Shmaliy, C.K. Ahn, L.J. Luo, An improved iterative FIR state estimator and its applications. *IEEE Trans. Ind. Inf.* **16**(2), 1003–1012 (2020)
118. S.Y. Zhao, Y.S. Shmaliy, C.K. Ahn, C.H. Zhao, Probabilistic monitoring of correlated sensors for nonlinear processes in state space. *IEEE Trans. Ind. Electron.* **67**(3), 2294–2303 (2020)
119. S.Y. Zhao, Y.S. Shmaliy, J.A. Andrade-Lucio, F. Liu, Multipass optimal FIR filtering for processes with unknown initial states and temporary mismatches. *IEEE Trans. Ind. Inf.* **17**(8), 5360–5368 (2021)
120. S.Y. Zhao, Y.S. Shmaliy, F. Liu, Batch optimal FIR smoothing: increasing state informativity in nonwhite measurement noise environments. *IEEE Trans. Ind. Inf.* **19**(5), 6993–7001 (2023)
121. S.Y. Zhao, J.F. Wang, Y.S. Shmaliy, F. Liu, Discrete time q-lag maximum likelihood FIR smoothing and iterative recursive algorithm. *IEEE Trans. Signal Process.* **69**, 6342–6354 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.