Circuits Syst Signal Process (2017) 36:2166–2183
DOI 10.1007/s00034-016-0394-4

SHORT PAPER

Gradient-Based Recursive Identification Methods for Input Nonlinear Equation Error Closed-Loop Systems

Bingbing Shen[1] · **Feng Ding**[1,2] · **Ahmed Alsaedi**[2] · **Tasawar Hayat**[2,3]

Received: 24 May 2016 / Revised: 12 August 2016 / Accepted: 13 August 2016 /
Published online: 23 August 2016

Abstract The identification problem of closed-loop or feedback nonlinear systems is a hot topic. Based on the hierarchical identification principle, this paper presents a hierarchical stochastic gradient algorithm and a hierarchical multi-innovation stochastic gradient algorithm for feedback nonlinear systems. The simulation results show that the hierarchical multi-innovation stochastic gradient can more effectively estimate the parameters of the feedback nonlinear systems than the hierarchical stochastic gradient algorithm.

Keywords Parameter estimation · Stochastic gradient · Multi-innovation · Hierarchical identification · Nonlinear system · Closed-loop system

✉ Feng Ding
fding@jiangnan.edu.cn

Bingbing Shen
bbshen12@126.com

Ahmed Alsaedi
aalsaedi@hotmail.com

Tasawar Hayat
fmgpak@gmail.com

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 Mathematics, Nonlinear Analysis and Applied Mathematics (NAAM) Research Group, King Abdulaziz University, Jeddah 21589, Saudi Arabia

3 Department of Mathematics, Quaid-I-Azam University, Islamabad 44000, Pakistan

1 Introduction

The mathematical models are basic for controller design [40,41]. System identification is the theory and methods of establishing the mathematical models of systems. In general, the identification algorithm is derived by defining and minimizing a cost function [27–29,38,39]. Linear system identification has been mature [5,7], and nonlinear systems have received extensive attention [6,20,24]. Nonlinear phenomena and nonlinear systems are ubiquitous. Almost all systems in real world are nonlinear to a certain degree. Recently, a series of estimation algorithms have been reported for nonlinear systems [22,23,34]. There are various forms of nonlinear systems. Bai [1] studied Hammerstein–Wiener systems and proposed an optimal two-stage identification algorithm. Li et al. [14] proposed an observer-based adaptive sliding mode control method for nonlinear Markovian jump systems. Bai and Cai [2] designed the identification scheme for nonlinear parametric Wiener system under Gaussian inputs. Wang and Ding [25] presented the recursive parameter and state estimation algorithm for an input nonlinear state space system using the hierarchical identification principle. Wang and Ding [26] discussed the recursive parameter estimation algorithms and convergence for a class of nonlinear systems with colored noise. Wang and Tang [32] tackled the iterative identification problems for a class of nonlinear systems with colored noises. Recently, a new fault detection design scheme was proposed for interval type-2 Takagi–Sugeno fuzzy systems [13]. Ding et al. [8] proposed recursive least squares parameter estimation for a class of output nonlinear systems based on the model decomposition.

Closed-loop control has received much attention [31]. The closed-loop subspace identification algorithms have been applied to various fields [21]. Wei et al. [36] presented a method for sensor fault diagnosis (detection and isolation) applied to large-scale wind turbine systems. The wind turbine model was built for the wind dynamics based on the closed-loop identification technique. On-line order estimation and parameter identification algorithms were presented for linear stochastic feedback control systems [19].

The multi-innovation identification method is effective for identifying systems [35,42,43]. In comparison with the previous algorithms, Mao and Ding [17] introduced a multi-innovation stochastic gradient identification algorithm for Hammerstein controlled autoregressive autoregressive systems based on the filtering technique. The multi-innovation identification method can combine with other identification methods to solve complex problems, such as the least squares algorithm [12], the forgetting gradient algorithm [37].

On the basis of the work in [10], this paper presents a hierarchical stochastic gradient algorithm and a hierarchical multi-innovation stochastic gradient algorithm for closed-loop nonlinear systems. The basic idea is, by means of the hierarchical identification principle, to decompose a feedback nonlinear system into two subsystems and to derive a hierarchical stochastic gradient algorithm. This work differs from the hierarchical least squares algorithms in [22] and the multistage least squares-based iterative estimation for feedback nonlinear systems with moving average noises [10].

This paper is organized as follows. Section 2 derives the identification problem for feedback nonlinear systems. Section 3 describes a hierarchical stochastic gradient

algorithm. Section 4 presents a hierarchical multi-innovation stochastic gradient algorithm. Two examples are given in Sect. 5 to illustrate the effectiveness of the proposed algorithms. Finally, concluding remarks are offered in Sect. 6.

2 Problem Description

Let us define some notation. "$X := A$" or "$A =: X$" means "A is defined as X"; $\mathbf{I}$ represents an identity matrix of appropriate sizes; the superscript T denotes the matrix transpose.

Consider the following feedback nonlinear stable system in Fig. 1,

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

$$u(t) = r(t) - y(t), \tag{2}$$

$$\bar{u}(t) = g(u(t)), \tag{3}$$

where $r(t) \in \mathbb{R}$ is the reference input of the system, $u(t) \in \mathbb{R}$ is the control input of the system, $\bar{u}(t) \in \mathbb{R}$ is the output of the nonlinear block, $y(t) \in \mathbb{R}$ is the system output, $v(t) \in \mathbb{R}$ is a stochastic white noise with zero mean and variance σ^2 and $A(z)$ and $B(z)$ are the polynomials in the shift operator z^{-1}:

$$A(z) := 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a},$$
$$B(z) := b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b}.$$

Suppose that the nonlinear output $\bar{u}(t)$ is a linear combination of a known basis $\boldsymbol{g} := (g_1, g_2, \ldots, g_{n_\gamma})$ with unknown coefficients $(\gamma_1, \gamma_2, \ldots, \gamma_{n_\gamma})$ and can be expressed as

$$\begin{aligned}\bar{u}(t) = g(u(t)) &= \gamma_1 g_1(u(t)) + \gamma_2 g_2(u(t)) + \cdots + \gamma_{n_\gamma} g_{n_\gamma}(u(t)) \\ &= \gamma_1 g_1(r(t) - y(t)) + \gamma_2 g_2(r(t) - y(t)) + \cdots + \gamma_{n_\gamma} g_{n_\gamma}(r(t) - y(t)) \\ &= \boldsymbol{g}(r(t) - y(t))\boldsymbol{\gamma}. \end{aligned} \tag{4}$$

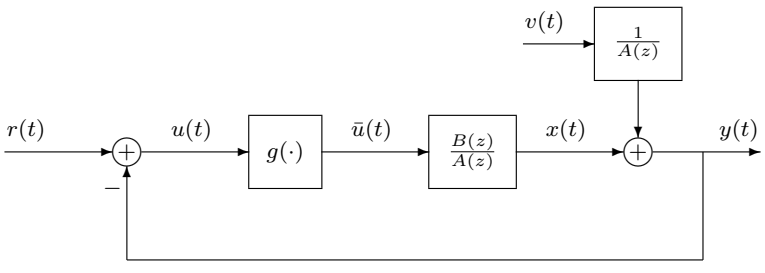

Fig. 1 The feedback nonlinear closed-loop system

Define the parameter vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ of the linear subsystems and $\boldsymbol{\gamma}$ of the nonlinear part:

$$\begin{aligned}\boldsymbol{a} &:= [a_1, a_2, \ldots, a_{n_a}]^{\mathrm{T}} \in \mathbb{R}^{n_a},\\ \boldsymbol{b} &:= [b_1, b_2, \ldots, b_{n_b}]^{\mathrm{T}} \in \mathbb{R}^{n_b},\\ \boldsymbol{\gamma} &:= [\gamma_1, \gamma_2, \ldots, \gamma_{n_\gamma}]^{\mathrm{T}} \in \mathbb{R}^{n_\gamma},\\ \boldsymbol{\vartheta} &:= \begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\\ \boldsymbol{\gamma}\end{bmatrix} \in \mathbb{R}^{n_a+n_b+n_\gamma},\end{aligned}$$

and the information vector and matrix

$$\begin{aligned}\boldsymbol{\varphi}_y(t) &:= [-y(t-1), -y(t-2), \ldots, -y(t-n_a)]^{\mathrm{T}} \in \mathbb{R}^{n_a},\\ \boldsymbol{G}(t) &:= [\boldsymbol{g}(r(t-1)-y(t-1)), \boldsymbol{g}(r(t-2)-y(t-2)), \ldots, \boldsymbol{g}(r(t-n_b)-y(t-n_b))]^{\mathrm{T}}\\ &= \begin{bmatrix} g_1(r(t-1)-y(t-1)) & g_2(r(t-1)-y(t-1)) & \cdots & g_{n_\gamma}(r(t-1)-y(t-1))\\ g_1(r(t-2)-y(t-2)) & g_2(r(t-2)-y(t-2)) & \cdots & g_{n_\gamma}(r(t-2)-y(t-2))\\ \vdots & \vdots & & \vdots\\ g_1(r(t-n_b)-y(t-n_b)) & g_2(r(t-n_b)-y(t-n_b)) & \cdots & g_{n_\gamma}(r(t-n_b)-y(t-n_b))\end{bmatrix} \in \mathbb{R}^{n_b\times n_\gamma}.\end{aligned}$$

From (1)–(4), we have

$$\begin{aligned}y(t) &= -\sum_{i=1}^{n_a} a_i y(t-i) + \sum_{i=1}^{n_b} b_i \bar{u}(t-i) + v(t)\\ &= -a_1 y(t-1) - a_2 y(t-2) - \cdots - a_{n_a}(y-n_a)\\ &\quad + \sum_{i=1}^{n_b} b_i[\gamma_1 g_1(u(t-i)) + \gamma_2 g_2(u(t-i)) + \cdots\\ &\quad + \gamma_{n_\gamma} g_{n_\gamma}(u(t-i))] + v(t)\\ &= \boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{G}(t)\boldsymbol{\gamma} + v(t). \qquad (5)\end{aligned}$$

The right-hand side of (5) contains the product of $\boldsymbol{b}$ and $\boldsymbol{\gamma}$. In order to get the unique parameter estimates, we use the standard constraint on $\boldsymbol{\gamma}$. In general, assume that the system is stable and $\|\boldsymbol{\gamma}\| = 1$ and the first entry of $\boldsymbol{\gamma}$ is positive, i.e., $\gamma_1 > 0$ where the norm of a matrix $\boldsymbol{X}$ is defined as $\|\boldsymbol{X}\|^2 := \mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$. The method proposed in the paper can be extended to identify the parameters of the system with known controller.

3 The Hierarchical Stochastic Gradient Algorithm

According to the hierarchical identification principle, Eq. (5) can be decomposed into two subsystems, and they contain the parameter vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\theta} := \begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{bmatrix}$, respectively.

Define the generalized information vectors $\boldsymbol{\phi}(\boldsymbol{\gamma}, t)$ and $\boldsymbol{\psi}(\boldsymbol{b}, t)$ and the fictitious output $y_1(t)$ as

$$\begin{aligned}\boldsymbol{\phi}(\boldsymbol{\gamma},t) &:= \begin{bmatrix}\boldsymbol{\varphi}_y(t)\\ \boldsymbol{G}(t)\boldsymbol{\gamma}\end{bmatrix} \in \mathbb{R}^{n_a+n_b},\\ \boldsymbol{\psi}(\boldsymbol{b},t) &:= \boldsymbol{G}^{\mathrm{T}}(t)\boldsymbol{b} \in \mathbb{R}^{n_\gamma},\\ y_1(t) &:= y(t) - \boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a}. \end{aligned} \tag{6}$$

From (5) and (6), we can obtain two subsystems:

$$\mathrm{S}_1:\ y_1(t) = \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{b},t)\boldsymbol{\gamma} + v(t), \tag{7}$$

$$\mathrm{S}_2:\ y(t) = \boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{\gamma},t)\boldsymbol{\theta} + v(t). \tag{8}$$

Subsystem (7) contains n_γ parameters, and Subsystem (8) contains $n_a + n_b$ parameters. The vectors $\boldsymbol{\gamma}$ in $\boldsymbol{\phi}(\boldsymbol{\gamma},t)$ and $\boldsymbol{b}$ in $\boldsymbol{\psi}(\boldsymbol{b},t)$ are the associate terms between two subsystems.

Define two criterion functions:

$$J_1(\boldsymbol{\gamma}) := \frac{1}{2}\left[y_1(t) - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{b},t)\gamma\right]^2,$$

$$J_2(\boldsymbol{\theta}) := \frac{1}{2}\left[y(t) - \boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{\gamma},t)\boldsymbol{\theta}\right]^2.$$

Let $\hat{\boldsymbol{\gamma}}(t)$ be the estimate of $\boldsymbol{\gamma}$ at time t, and $\hat{\boldsymbol{\theta}}(t) := \begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{b}}(t)\end{bmatrix}$ be the estimate of $\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{bmatrix}$ at time t. Replacing $\boldsymbol{\gamma}$ with preceding estimate $\hat{\boldsymbol{\gamma}}(t-1)$. The gradients of $J_1(\boldsymbol{\gamma})$ and $J_2(\boldsymbol{\theta})$ are given by

$$\mathrm{grad}[J_1(\hat{\boldsymbol{\gamma}}(t-1))] = \left.\frac{\partial J_1(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}(t-1)} = -\boldsymbol{\psi}(\boldsymbol{b},t)[y_1(t) - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{b},t)\hat{\gamma}(t-1)],$$

$$\mathrm{grad}[J_2(\hat{\boldsymbol{\theta}}(t-1))] = \left.\frac{\partial J_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(t-1)} = -\boldsymbol{\phi}(\boldsymbol{\gamma},t)[y(t) - \boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{\gamma},t)\hat{\boldsymbol{\theta}}(t-1)].$$

Using the gradient search and optimizing the criterion functions $J_1(\boldsymbol{\gamma})$ and $J_2(\boldsymbol{\theta})$, we can get the following recursive relations:

$$\begin{aligned}\hat{\boldsymbol{\gamma}}(t) &= \hat{\boldsymbol{\gamma}}(t-1) - \frac{1}{r_1(t)}\mathrm{grad}[J_1(\hat{\boldsymbol{\gamma}}(t-1))]\\ &= \hat{\boldsymbol{\gamma}}(t-1) + \frac{\boldsymbol{\psi}(\boldsymbol{b},t)}{r_1(t)}[y(t) - \boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a} - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{b},t)\hat{\boldsymbol{\gamma}}(t-1)],\\ \hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) - \frac{1}{r_2(t)}\mathrm{grad}[J_2(\hat{\boldsymbol{\theta}}(t-1))]\\ &= \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t),t)}{r_2(t)}[y(t) - \boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{\gamma},t)\hat{\boldsymbol{\theta}}(t-1)].\end{aligned}$$

Furthermore, taking

$$r_1(t) = r_1(t-1) + \mathrm{tr}[\boldsymbol{\psi}(\boldsymbol{b},t)\boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{b},t)] = r_1(t-1) + \|\boldsymbol{\psi}(\boldsymbol{b},t)\|^2,$$

$$r_2(t) = r_2(t-1) + \mathrm{tr}[\boldsymbol{\phi}(\boldsymbol{\gamma},t)\boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{\gamma},t)] = r_2(t-1) + \|\boldsymbol{\phi}(\boldsymbol{\gamma},t)\|^2,$$

we can acquire the following stochastic gradient algorithm:

$$\hat{\boldsymbol{\gamma}}(t) = \hat{\boldsymbol{\gamma}}(t-1) + \frac{\boldsymbol{\psi}(\boldsymbol{b},t)}{r_1(t)}[y(t) - \boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a} - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{b},t)\hat{\boldsymbol{\gamma}}(t-1)], \tag{9}$$

$$r_1(t) = r_1(t-1) + \|\boldsymbol{\psi}(\boldsymbol{b},t)\|^2, \quad r_1(0) = 1, \tag{10}$$

$$e_1(t) = y(t) - \boldsymbol{\varphi}_y^{\mathrm{T}}(t)\boldsymbol{a} - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{b},t)\hat{\boldsymbol{\gamma}}(t-1), \tag{11}$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\phi}(\boldsymbol{\gamma},t)}{r_2(t)}\left[y(t) - \boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{\gamma},t)\hat{\boldsymbol{\theta}}(t-1)\right], \tag{12}$$

$$r_2(t) = r_2(t-1) + \|\boldsymbol{\phi}(\boldsymbol{\gamma},t)\|^2, \quad r_2(0) = 1, \tag{13}$$

$$e_2(t) = y(t) - \boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{\gamma},t)\hat{\boldsymbol{\theta}}(t-1). \tag{14}$$

$\mathrm{sgn}[X]$ is a sign function that extracts the sign of a real number, and we use $\mathrm{sgn}[\hat{\gamma}_1(t)]$ to stand for the sign of the first element of $\hat{\boldsymbol{\gamma}}(t)$ and normalize $\hat{\boldsymbol{\gamma}}(t)$ with the first positive element

$$\hat{\boldsymbol{\gamma}}(t) := \mathrm{sgn}[\hat{\gamma}_1(t)]\frac{\hat{\boldsymbol{\gamma}}(t)}{\|\hat{\boldsymbol{\gamma}}(t)\|}.$$

where the first entry of $\boldsymbol{\gamma}$ is positive, the initial values $\|\hat{\gamma}(0)\| = 1$.

Replacing the unknown vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in (9) and $\boldsymbol{\gamma}$ in (12) with their preceding estimates $\hat{\boldsymbol{a}}(t-1)$ and $\hat{\boldsymbol{b}}(t-1)$ and the current estimate $\hat{\boldsymbol{\gamma}}(t)$, we can obtain the following hierarchical stochastic gradient algorithm for the feedback nonlinear system (the FN-HSG algorithm for short):

$$\hat{\boldsymbol{\gamma}}(t) = \hat{\boldsymbol{\gamma}}(t-1) + \frac{\boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1),t)}{r_1(t)}e_1(t), \tag{15}$$

$$e_1(t) = y(t) - \boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1) - \boldsymbol{\psi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(t-1),t)\hat{\boldsymbol{\gamma}}(t-1), \tag{16}$$

$$r_1(t) = r_1(t-1) + \|\boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1),t)\|^2, \quad r_1(0) = 1, \tag{17}$$

$$\boldsymbol{\varphi}_y(t) = [-y(t-1), -y(t-2), \ldots, -y(t-n_a)]^{\mathrm{T}}, \tag{18}$$

$$\boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1),t) = \boldsymbol{G}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1), \tag{19}$$

$$\boldsymbol{G}(t) = \begin{bmatrix} g_1(r(t-1)-y(t-1)) & g_2(r(t-1)-y(t-1)) & \cdots & g_{n_\gamma}(r(t-1)-y(t-1)) \\ g_1(r(t-2)-y(t-2)) & g_2(r(t-2)-y(t-2)) & \cdots & g_{n_\gamma}(r(t-2)-y(t-2)) \\ \vdots & \vdots & & \vdots \\ g_1(r(t-n_b)-y(t-n_b)) & g_2(r(t-n_b)-y(t-n_b)) & \cdots & g_{n_\gamma}(r(t-n_b)-y(t-n_b)) \end{bmatrix}, \tag{20}$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t),t)}{r_2(t)}e_2(t), \tag{21}$$

$$e_2(t) = y(t) - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\boldsymbol{\gamma}}(t),t)\hat{\boldsymbol{\theta}}(t-1), \tag{22}$$

$$r_2(t) = r_2(t-1) + \|\boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t), t)\|^2, \quad r_2(0) = 1, \tag{23}$$

$$\boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t), t) = \left[\boldsymbol{\varphi}_y^{\mathrm{T}}(t), \boldsymbol{G}^{\mathrm{T}}(t)\hat{\boldsymbol{\gamma}}^{\mathrm{T}}(t)\right]^{\mathrm{T}}, \tag{24}$$

$$\hat{\boldsymbol{\gamma}}(t) = \mathrm{sgn}[\hat{\gamma}_1(t)]\frac{\hat{\boldsymbol{\gamma}}(t)}{\|\hat{\boldsymbol{\gamma}}(t)\|}, \tag{25}$$

$$\hat{\boldsymbol{\theta}}(t) = \left[\hat{\boldsymbol{a}}^{\mathrm{T}}(t), \hat{\boldsymbol{b}}^{\mathrm{T}}(t)\right]^{\mathrm{T}}. \tag{26}$$

The procedures of computing the parameter estimation vectors $\hat{\boldsymbol{\gamma}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ in (15)–(26) are listed in the following.

1. Let $t = 1$, give the data length L_e, and set the initial values $\|\hat{\boldsymbol{\gamma}}(0)\| = 1$ with $\hat{\gamma}_1(0) > 0, \hat{\boldsymbol{\theta}}(0) = \begin{bmatrix}\hat{\boldsymbol{a}}(0)\\ \hat{\boldsymbol{b}}(0)\end{bmatrix} = \mathbf{1}_{n_a+n_b}/p_0, r_1(0) = 1, r_2(0) = 1$.
2. Collect the input–output data $r(t)$ and $y(t)$, and compute $r_1(t)$ using (17).
3. Form $\boldsymbol{\varphi}_y(t)$ using (18), $\boldsymbol{G}(t)$ using (20), and compute $\boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1), t)$ using (19).
4. Update the estimate $\hat{\boldsymbol{\gamma}}(t)$ using (15), and normalize $\hat{\boldsymbol{\gamma}}(t)$ using (25).
5. Compute $r_2(t)$ using (22), and form $\boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t), t)$ using (24).
6. Update the estimate $\hat{\boldsymbol{\theta}}(t)$ using (21).
7. If $t < L_e$, increase t by 1 and go to Step 2. Otherwise, terminate the procedure and obtain the parameter estimates.

The flowchart of computing the parameter estimates $\hat{\boldsymbol{\gamma}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ is shown in Fig. 2.

4 The Hierarchical Multi-Innovation Stochastic Gradient Algorithm

In order to improve the convergence rate and parameter estimation accuracy of the FN-HSG algorithm, based on the algorithm in (15)–(26), according to the multi-innovation identification theory, we expand the scalar innovations $e_1(t) = y(t) - \boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1) - \boldsymbol{\psi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(t-1), t)\hat{\boldsymbol{\gamma}}(t-1)$ in (16) and $e_2(t) = y(t) - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\boldsymbol{\gamma}}(t), t)\hat{\boldsymbol{\theta}}(t-1)$ in (22) to the innovation vectors:

$$\boldsymbol{E}_1(p,t) := \begin{bmatrix} y(t) - \boldsymbol{\varphi}_y^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1) - \boldsymbol{\psi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(t-1), t)\hat{\boldsymbol{\gamma}}(t-1) \\ y(t-1) - \boldsymbol{\varphi}_y^{\mathrm{T}}(t-1)\hat{\boldsymbol{a}}(t-1) - \boldsymbol{\psi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(t-1), t-1)\hat{\boldsymbol{\gamma}}(t-1) \\ \vdots \\ y(t-p+1) - \boldsymbol{\varphi}_y^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{a}}(t-1) - \boldsymbol{\psi}^{\mathrm{T}}(\hat{\boldsymbol{b}}(t-1), t-p+1)\hat{\boldsymbol{\gamma}}(t-1) \end{bmatrix} \in \mathbb{R}^p, \tag{27}$$

$$\boldsymbol{E}_2(p,t) := \begin{bmatrix} y(t) - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\boldsymbol{\gamma}}(t), t)\hat{\boldsymbol{\theta}}(t-1) \\ y(t-1) - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\boldsymbol{\gamma}}(t), t-1)\hat{\boldsymbol{\theta}}(t-1) \\ \vdots \\ y(t-p+1) - \boldsymbol{\phi}^{\mathrm{T}}(\hat{\boldsymbol{\gamma}}(t), t-p+1)\hat{\boldsymbol{\theta}}(t-1) \end{bmatrix} \in \mathbb{R}^p. \tag{28}$$

where p denotes the innovation length.

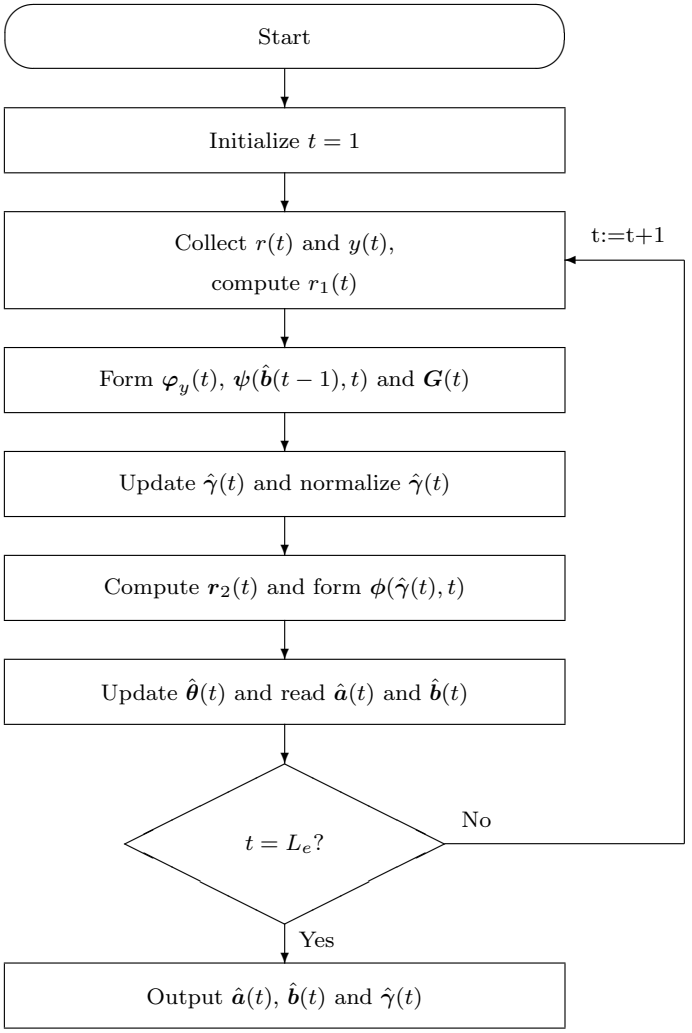

Fig. 2 The flowchart for computing the parameter estimates

Define the stacked output vector $\boldsymbol{Y}(p,t)$ and the information matrices, $\hat{\boldsymbol{\Psi}}(p,t)$ and $\hat{\boldsymbol{\Phi}}(p,t)$ as

$$
\begin{aligned}
\boldsymbol{Y}(p,t) &:= [y(t), y(t-1), \ldots, y(t-p+1)]^{\mathrm{T}} \in \mathbb{R}^{p},\\
\boldsymbol{\Phi}_y(p,t) &:= [\boldsymbol{\varphi}_y(t), \boldsymbol{\varphi}_y(t-1), \ldots, \boldsymbol{\varphi}_y(t-p+1)] \in \mathbb{R}^{n_a \times p},\\
\hat{\boldsymbol{\Psi}}(p,t) &:= [\boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1), t), \boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1), t-1), \ldots, \boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1), t-p+1)]\\
&\quad \in \mathbb{R}^{(n_a+n_b)\times p},\\
\hat{\boldsymbol{\Phi}}(p,t) &:= [\boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t), t), \boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t), t-1), \ldots, \boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t), t-p+1)] \in \mathbb{R}^{n_\gamma \times p}.
\end{aligned}
$$

Then, Eqs. (27) and (28) can be expressed as

$$\boldsymbol{E}_1(p,t) = \boldsymbol{Y}(p,t) - \boldsymbol{\Phi}_y^{\mathrm{T}}(p,t)\hat{\boldsymbol{a}}(t-1) - \hat{\boldsymbol{\Psi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\gamma}}(t-1),$$
$$\boldsymbol{E}_2(p,t) = \boldsymbol{Y}(p,t) - \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1).$$

Equations (15) and (21) can be rewritten as

$$\hat{\boldsymbol{\gamma}}(t) = \hat{\boldsymbol{\gamma}}(t-1) + \frac{\hat{\boldsymbol{\Psi}}(p,t)}{r_1(t)}\boldsymbol{E}_1(p,t),$$
$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}(p,t)}{r_2(t)}\boldsymbol{E}_2(p,t).$$

We can drive the hierarchical multi-innovation stochastic gradient algorithm for feedback nonlinear systems (the FN-HMISG for short) as follows:

$$\hat{\boldsymbol{\gamma}}(t) = \hat{\boldsymbol{\gamma}}(t-1) + \frac{\hat{\boldsymbol{\Psi}}(p,t)}{r_1(t)}\boldsymbol{E}_1(p,t), \tag{29}$$
$$r_1(t) = r_1(t-1) + \|\boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1),t)\|^2, \quad r_1(0) = 1, \tag{30}$$
$$\boldsymbol{E}_1(p,t) = \boldsymbol{Y}(p,t) - \boldsymbol{\Phi}_y^{\mathrm{T}}(p,t)\hat{\boldsymbol{a}}(t-1) - \hat{\boldsymbol{\Psi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\gamma}}(t-1), \tag{31}$$
$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}(p,t)}{r_2(t)}\boldsymbol{E}_2(p,t), \tag{32}$$
$$r_2(t) = r_2(t-1) + \|\boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t),t)\|^2, \quad r_2(0) = 1, \tag{33}$$
$$\boldsymbol{E}_2(p,t) = \boldsymbol{Y}(p,t) - \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1), \tag{34}$$
$$\boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1),t) = \boldsymbol{G}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1), \tag{35}$$
$$\boldsymbol{Y}(p,t) = [y(t), y(t-1), \ldots, y(t-p+1)]^{\mathrm{T}}, \tag{36}$$
$$\boldsymbol{\Phi}_y(p,t) = [\boldsymbol{\varphi}_y(t), \boldsymbol{\varphi}_y(t-1), \ldots, \boldsymbol{\varphi}_y(t-p+1)], \tag{37}$$
$$\boldsymbol{\varphi}_y(t) = [-y(t-1), -y(t-2), \ldots, -y(t-n_a)]^{\mathrm{T}}, \tag{38}$$
$$\hat{\boldsymbol{\Psi}}(p,t) = [\boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1),t), \boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1),t-1), \ldots, \boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1),t-p+1)], \tag{39}$$

$$\boldsymbol{G}(t) = \begin{bmatrix} g_1(r(t-1)-y(t-1)) & g_2(r(t-1)-y(t-1)) & \cdots & g_{n_\gamma}(r(t-1)-y(t-1)) \\ g_1(r(t-2)-y(t-2)) & g_2(r(t-2)-y(t-2)) & \cdots & g_{n_\gamma}(r(t-2)-y(t-2)) \\ \vdots & \vdots & & \vdots \\ g_1(r(t-n_b)-y(t-n_b)) & g_2(r(t-n_b)-y(t-n_b)) & \cdots & g_{n_\gamma}(r(t-n_b)-y(t-n_b)) \end{bmatrix}, \tag{40}$$

$$\hat{\boldsymbol{\Phi}}(p,t) := [\boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t),t), \boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t),t-1), \ldots, \boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t),t-p+1)], \tag{41}$$
$$\boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t),t) = [\boldsymbol{\varphi}_y^{\mathrm{T}}(t), \boldsymbol{G}^{\mathrm{T}}(t)\hat{\boldsymbol{\gamma}}^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{42}$$
$$\hat{\boldsymbol{\theta}}(t) = [\hat{\boldsymbol{a}}^{\mathrm{T}}(t), \hat{\boldsymbol{b}}^{\mathrm{T}}(t)]^{\mathrm{T}}. \tag{43}$$

Table 1 The FN-HMISG parameter estimates and errors with $\sigma^2 = 0.30^2$

p	t	a_1	a_2	b_1	b_2	γ_1	γ_2	δ (%)
1	100	0.31554	0.33860	1.37106	0.64827	0.51494	0.85722	49.10921
	200	0.29826	0.31798	1.33278	0.69437	0.52623	0.85034	46.19901
	500	0.29012	0.30886	1.29540	0.73432	0.53690	0.84364	43.54742
	1000	0.29298	0.30233	1.26396	0.76460	0.54731	0.83693	41.34252
	2000	0.29682	0.30231	1.23250	0.79227	0.55521	0.83171	39.27214
	3000	0.30246	0.29701	1.21553	0.80737	0.55993	0.82854	38.08594
	4000	0.30407	0.29517	1.20445	0.81811	0.56277	0.82661	37.31223
2	100	0.31809	0.25189	0.95853	0.96821	0.59802	0.80148	25.00722
	200	0.34982	0.23405	0.94036	1.02962	0.61520	0.78837	21.45232
	500	0.36643	0.23226	0.92893	1.07396	0.60790	0.79402	18.99586
	1000	0.38290	0.23367	0.90999	1.11330	0.60639	0.79517	16.64049
	2000	0.39945	0.23843	0.89023	1.14721	0.61062	0.79193	14.51833
	3000	0.41233	0.23556	0.88307	1.16235	0.60950	0.79279	13.51048
	4000	0.41741	0.23642	0.87901	1.17361	0.61015	0.79229	12.85604
5	100	0.29724	0.20798	0.61647	1.07140	0.59661	0.80253	21.71878
	200	0.41580	0.20430	0.72925	1.19278	0.64153	0.76710	12.37638
	500	0.44237	0.21992	0.81067	1.23689	0.60671	0.79492	8.90357
	1000	0.45967	0.22919	0.81063	1.28698	0.59712	0.80215	6.18345
	2000	0.47728	0.23634	0.80061	1.32357	0.60561	0.79576	4.10524
	3000	0.49570	0.23411	0.80407	1.33188	0.60226	0.79830	3.55167
	4000	0.49722	0.23678	0.80760	1.34147	0.60254	0.79809	3.06533
7	100	0.30627	0.18988	0.59168	1.09760	0.59779	0.80165	21.20924
	200	0.45145	0.20645	0.73898	1.24343	0.64177	0.76690	9.50272
	500	0.46572	0.22122	0.84053	1.28312	0.60932	0.79293	6.67727
	1000	0.48048	0.23174	0.82491	1.33047	0.59377	0.80464	3.99013
	2000	0.49526	0.23574	0.80534	1.36248	0.60504	0.79620	2.08589
	3000	0.51522	0.23393	0.80886	1.36428	0.60135	0.79899	2.17313
	4000	0.51404	0.23647	0.81291	1.37133	0.60148	0.79889	1.87287
True values		0.50000	0.25000	0.80000	1.40000	0.60000	0.80000	

The procedures of computing the parameter estimation vectors $\hat{\boldsymbol{\gamma}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ in (29)–(43) are listed in the following.

1. Let $t = 1$, give the data length L_e, and set the initial values $\|\hat{\boldsymbol{\gamma}}(0)\| = 1$ with $\hat{\gamma}_1(0) > 0$, $\hat{\boldsymbol{\theta}}(0) = \begin{bmatrix} \hat{\boldsymbol{a}}(0) \\ \hat{\boldsymbol{b}}(0) \end{bmatrix} = \mathbf{1}_{n_a+n_b}/p_0$, $r_1(0) = 1$, $r_2(0) = 1$.
2. Collect the input–output data $r(t)$ and $y(t)$, and form $\boldsymbol{Y}(p,t)$, $\boldsymbol{\varphi}_y(t)$ and $\boldsymbol{G}(t)$ using (36), (38) and (40).
3. Form $\boldsymbol{\psi}(\hat{\boldsymbol{b}}(t-1),t)$, $\boldsymbol{\Phi}_y(p,t)$ and $\hat{\boldsymbol{\Psi}}(p,t)$ using (35), (37) and (39).
4. Compute $r_1(t)$ and $\boldsymbol{E}_1(p,t)$ using (30) and (31).
5. Update the estimate $\hat{\boldsymbol{\gamma}}(t)$ using (29).

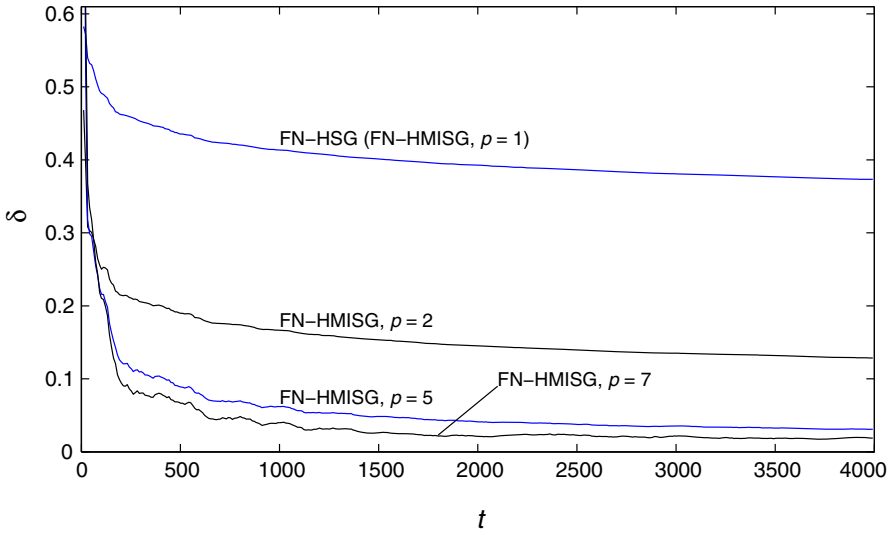

Fig. 3 The FN-HMISG parameter estimation errors σ versus t ($\sigma^2 = 0.30^2$)

6. Form $\hat{\boldsymbol{\Phi}}(p,t)$ and $\boldsymbol{\phi}(\hat{\boldsymbol{\gamma}}(t),t)$ using (41) and (42), and compute $r_2(t)$ and $\boldsymbol{E}_2(p,t)$ using (33) and (34).
7. Update the estimate $\hat{\boldsymbol{\theta}}(t)$ using (32).
8. If $t < L_e$, increase t by 1 and go to Step 2. Otherwise, terminate the procedure and obtain the parameter estimates.

Equations (29)–(43) form the FN-HMISG algorithm for the feedback nonlinear system. Obviously, when $p = 1$, the FN-HMISG algorithm reduces into the FN-HSG algorithm.

5 Examples

In this section, two examples are given to test the FN-HMISG algorithm.

Example 1 Consider the following nonlinear system:

$$\begin{aligned}
y(t) &= (1 - A(z))y(t) + B(z)g(r(t) - y(t)) + v(t),\\
A(z) &= 1 + a_1 z^{-1} + a_2 z^{-2} = 1 + 0.50z^{-1} + 0.25z^{-2},\\
B(z) &= b_1 z^{-1} + b_2 z^{-2} = 0.80z^{-1} + 1.40z^{-2},\\
g(r(t) - y(t)) &= 0.60\sin(r(t) - y(t)) + 0.80\sin((r(t) - y(t))^2),\\
\boldsymbol{\theta} &= [a_1, a_2, b_1, b_2]^{\mathrm{T}} = [0.50, 0.25, 0.80, 1.40]^{\mathrm{T}},\\
\boldsymbol{\gamma} &= [\gamma_1, \gamma_2]^{\mathrm{T}} = [0.60, 0.80]^{\mathrm{T}}.
\end{aligned}$$

In simulation, the input $r(t)$ is taken as uncorrelated stochastic sequence with zero mean and $v(t)$ as a white noise sequence zero and variance $\sigma^2 = 0.30^2$. Take the data

Table 2 The parameter estimates based on the 20 Monte Carlo runs ($\sigma^2 = 0.30^2$)

t	a_1	a_2	b_1	b_2	γ_1	γ_2
100	0.43713 ± 0.19243	0.19053 ± 0.12166	0.74112 ± 0.17895	1.22179 ± 0.23846	0.62570 ± 0.12214	0.77656 ± 0.11268
200	0.46562 ± 0.08160	0.22815 ± 0.05858	0.77805 ± 0.09119	1.29709 ± 0.13534	0.61637 ± 0.08467	0.78620 ± 0.06276
500	0.48742 ± 0.05940	0.22467 ± 0.04346	0.78770 ± 0.05283	1.34382 ± 0.08874	0.60031 ± 0.03549	0.79952 ± 0.02767
1000	0.50060 ± 0.03697	0.23138 ± 0.03025	0.79268 ± 0.04349	1.36643 ± 0.05737	0.60321 ± 0.01596	0.79751 ± 0.01189
2000	0.50292 ± 0.02491	0.23906 ± 0.01744	0.79669 ± 0.02862	1.38629 ± 0.04753	0.60329 ± 0.01480	0.79746 ± 0.01135
3000	0.50827 ± 0.02248	0.23698 ± 0.01272	0.79820 ± 0.02678	1.39163 ± 0.02945	0.60089 ± 0.01020	0.79931 ± 0.00759
4000	0.51132 ± 0.01423	0.23844 ± 0.01693	0.79983 ± 0.01832	1.39624 ± 0.02595	0.60306 ± 0.01025	0.79768 ± 0.00766
True values	0.50000	0.25000	0.80000	1.40000	0.60000	0.80000

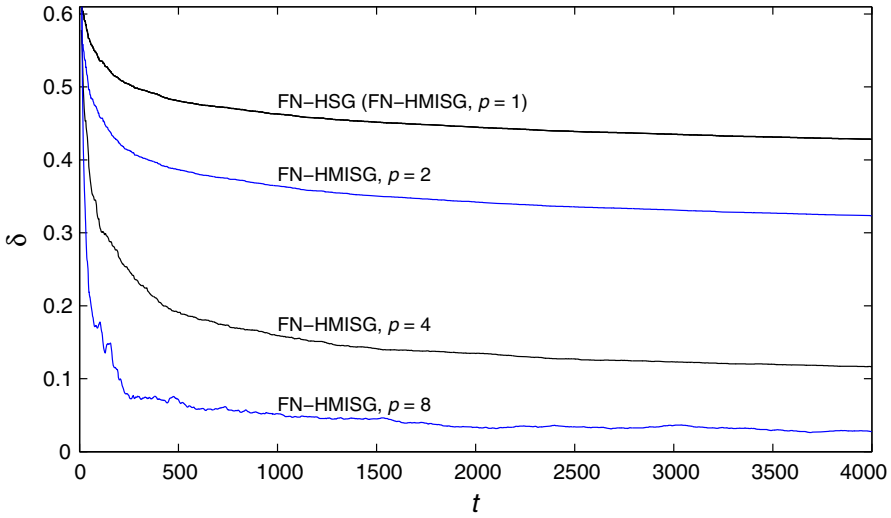

Fig. 4 The FN-HMISG parameter estimation errors σ versus t ($\sigma^2 = 0.30^2$)

length $L_e = 4000$ and $\|\boldsymbol{\gamma}\|^2 = \gamma_1^2 + \gamma_2^2 = 1$. Applying the FN-HMISG algorithm to estimate the parameters of this nonlinear system, the parameter estimates and their errors are shown in Table 1 with $p = 1$, $p = 2$, $p = 5$ and $p = 7$. The estimation errors $\delta := \|\hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}\|/\|\boldsymbol{\vartheta}\|$ versus t are shown in Fig. 3.

Furthermore, using the Monte Carlo simulations with 20 sets of noise realizations, the parameter estimates and the variances of the FN-HMISG algorithm are shown in Table 2 with $\sigma^2 = 0.30^2$, $p = 7$ and $L_e = 4000$. From Table 2, we can see that the average values of the parameter estimates are very close to the true parameters.

Example 2 Consider the following nonlinear system:

$$\begin{aligned}
y(t) &= (1 - A(z))y(t) + B(z)g(r(t) - y(t)) + v(t),\\
A(z) &= 1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} = 1 + 0.30z^{-1} + 0.48z^{-2} + 0.45z^{-3},\\
B(z) &= b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3} = 0.79z^{-1} + 0.84z^{-3},\\
g(r(t) - y(t)) &= 0.48\sin(r(t) - y(t)) + 0.87\sin((r(t) - y(t))^2)\\
&\quad +0.096\cos(r(t) - y(t)),\\
\boldsymbol{\theta} &= [a_1, a_2, a_3, b_1, b_2, b_3]^{\mathrm{T}} = [0.30, 0.48, 0.45, 0.79, 0.00, 0.84]^{\mathrm{T}},\\
\boldsymbol{\gamma} &= [\gamma_1, \gamma_2, \gamma_3]^{\mathrm{T}} = [0.48, 0.87, 0.096]^{\mathrm{T}},\ \|\boldsymbol{\gamma}\|^2 = \gamma_1^2 + \gamma_2^2 + \gamma_3^2 = 1.
\end{aligned}$$

The simulation conditions are similar to those in Example 1. We apply the FN-HMISG algorithm to estimate the parameters of this nonlinear system, the parameter estimates and their errors are shown in Table 3 with $p = 1$, $p = 2$, $p = 4$ and $p = 8$, and the estimation errors δ versus t are shown in Fig. 4.

From Tables 1 and 3 and Figs. 3 and 4, we can draw the following conclusions.

Table 3 The FN-HMISG parameter estimates and errors with $\sigma^2 = 0.30^2$

p	t	a_1	a_2	a_3	b_1	b_2	b_3	γ_1	γ_2	γ_3	δ (%)
1	100	0.13751	0.84283	0.82143	0.30668	0.01266	1.27061	0.57873	0.72573	0.37200	53.71065
	200	0.14069	0.82657	0.78609	0.34309	0.02705	1.27024	0.55134	0.74445	0.37658	51.08746
	500	0.16818	0.80530	0.76621	0.36964	0.01392	1.26731	0.53488	0.77007	0.34769	48.09448
	1000	0.17410	0.79102	0.74738	0.39172	0.01695	1.26790	0.52083	0.78420	0.33729	46.28168
	2000	0.18778	0.78291	0.73244	0.40723	0.01176	1.25883	0.50909	0.79806	0.32239	44.47125
	3000	0.19632	0.77863	0.72356	0.41581	0.00728	1.25537	0.50561	0.80400	0.31296	43.51317
	4000	0.20253	0.77505	0.71884	0.42218	0.00442	1.25180	0.50390	0.80749	0.30666	42.82710
2	100	0.25190	0.79499	0.70399	0.40596	−0.10176	1.30996	0.54798	0.77766	0.30815	46.00138
	200	0.25126	0.77983	0.65761	0.45795	−0.07166	1.29112	0.50996	0.79945	0.31752	42.27640
	500	0.28503	0.75241	0.64576	0.50298	−0.08327	1.27656	0.49006	0.82805	0.27234	38.65031
	1000	0.28598	0.73815	0.62840	0.53103	−0.07357	1.26479	0.47632	0.84101	0.25654	36.43490
	2000	0.29842	0.73137	0.61479	0.55290	−0.07326	1.24183	0.46494	0.85335	0.23584	34.21094
	3000	0.30607	0.72885	0.60446	0.56499	−0.07531	1.23222	0.46505	0.85650	0.22392	33.13048
	4000	0.31130	0.72417	0.60201	0.57417	−0.07500	1.22464	0.46494	0.85827	0.21727	32.34498
4	100	0.50116	0.58697	0.40689	0.57195	0.22491	0.83537	0.67325	0.66568	0.32188	31.46271
	200	0.44991	0.59252	0.41923	0.61514	0.21945	0.81147	0.62928	0.71762	0.29838	26.63277
	500	0.43777	0.57105	0.45083	0.67771	0.17497	0.81625	0.57295	0.78736	0.22759	19.09121
	1000	0.41388	0.56076	0.46385	0.69900	0.15927	0.82200	0.53493	0.81873	0.20865	15.89606
	2000	0.40290	0.56080	0.47441	0.71106	0.13783	0.82744	0.51571	0.83859	0.17551	13.46840
	3000	0.40230	0.56285	0.46844	0.72191	0.12265	0.83225	0.51142	0.84432	0.15992	12.30540
	4000	0.39745	0.55877	0.47442	0.72845	0.11745	0.83545	0.51021	0.84619	0.15378	11.65769
8	100	0.31498	0.36554	0.30682	0.72215	−0.16035	0.75389	0.53477	0.82120	0.19910	17.53175
	200	0.24520	0.45157	0.37207	0.76453	−0.04944	0.74866	0.52615	0.83638	0.15373	9.89806
	500	0.25388	0.45609	0.41349	0.84424	−0.06019	0.78816	0.51016	0.85519	0.09162	7.11726

Table 3 continued

p	t	a_1	a_2	a_3	b_1	b_2	b_3	γ_1	γ_2	γ_3	δ (%)
	1000	0.25457	0.44876	0.42858	0.82104	−0.04842	0.81071	0.48110	0.87141	0.09584	5.18805
	2000	0.26757	0.47475	0.43083	0.80196	−0.03505	0.82752	0.47192	0.87735	0.08687	3.33898
	3000	0.28308	0.48839	0.41644	0.80412	−0.04147	0.83408	0.48348	0.87161	0.08093	3.59351
	4000	0.28147	0.48941	0.43226	0.80283	−0.03357	0.83357	0.48471	0.87002	0.09007	2.73871
True values		0.30000	0.48000	0.45000	0.79000	0.00000	0.84000	0.48000	0.87000	0.09600	0.00000

1. The parameter estimation errors given by the FN-HMISG algorithm are becoming smaller with the innovation length p increasing.
2. The proposed FN-HMISG algorithm becomes more accurate as the data length t increases.
3. The parameter estimates of the FN-HMISG algorithm with $p > 1$ converge faster to their true values than the parameter estimates of the FN-HSG algorithm.
4. These results confirm that the FN-HMISG algorithm with $p > 1$ can estimate the parameters more accurately than the FN-HSG algorithm.

6 Conclusions

This paper proposes a hierarchical stochastic gradient algorithm and a hierarchical multi-innovation stochastic gradient algorithm for feedback nonlinear systems. The simulation results indicate that the multi-innovation stochastic gradient algorithm can improve the parameter estimation accuracy compared with the stochastic gradient algorithm. The method used in this paper can be extended to study the identification problems of nonlinear control systems with colored noise [15,16] and applied to hybrid switching-impulsive dynamical networks [11] or applied to other fields [3,4,9,18,30,33].

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 61273194) and the 111 Project (B12018).

References

1. E.W. Bai, An optimal two-stage identification algorithm for Hammerstein–Wiener nonlinear systems. Automatica **34**(3), 333–338 (1998)
2. E. Bai, Z. Cai, How nonlinear parametric Wiener system identification is under Gaussian inputs? IEEE Trans. Automat. Control **57**(3), 738–742 (2012)
3. X. Cao, D.Q. Zhu, S.X. Yang, Multi-AUV target search based on bioinspired neurodynamics model in 3-D underwater environments. IEEE Trans. Neural Netw. Learn. Syst. (2016). doi:10.1109/TNNLS.2015.2482501
4. Z.Z. Chu, D.Q. Zhu, S.X. Yang, Observer-based adaptive neural network trajectory tracking control for remotely operated Vehicle. IEEE Trans. Neural Netw. Learn. Syst. (2016). doi:10.1109/TNNLS.2016.2544786
5. F. Ding, X.M. Liu, Y. Gu, An auxiliary model based least squares algorithm for a dual-rate state space system with time-delay using the data filtering. J. Franklin Inst. **353**(2), 398–408 (2016)
6. F. Ding, X.M. Liu, M.M. Liu, The recursive least squares identification algorithm for a class of Wiener nonlinear systems. J. Franklin Inst. **353**(7), 1518–1526 (2016)
7. F. Ding, X.M. Liu, X.Y. Ma, Kalman state filtering based least squares iterative parameter estimation for observer canonical state space systems using decomposition. J. Comput. Appl. Math. **301**, 135–143 (2016)
8. F. Ding, X.H. Wang, Q.J. Chen, Y.S. Xiao, Recursive least squares parameter estimation for a class of output nonlinear systems based on the model decomposition. Circuits Syst. Signal Process. **35**(9), 3323–3338 (2016)
9. L. Feng, M.H. Wu, Q.X. Li et al., Array factor forming for image reconstruction of one-dimensional nonuniform aperture synthesis radiometers. IEEE Geosci. Remote Sens. Lett. **13**(2), 237–241 (2016)
10. P.P. Hu, F. Ding, Multistage least squares based iterative estimation for feedback nonlinear systems with moving average noises using the hierarchical identification principle. Nonlinear Dyn. **73**(1–2), 583–592 (2013)

11. Y. Ji, X.M. Liu, Unified synchronization criteria for hybrid switching-impulsive dynamical networks. Circuits Syst. Signal Process. **34**(5), 1499–1517 (2015)
12. Q.B. Jin, Z. Wang, X.P. Liu, Auxiliary model-based interval-varying multi-innovation least squares identification for multivariable OE-like systems with scarce measurements. J. Process Control **35**(11), 154–168 (2015)
13. H. Li, Y. Gao, P. Shi, H.K. Lam, Observer-based fault detection for nonlinear systems with sensor fault and limited communication capacity. IEEE Trans. Automat. Control (2016). doi:10.1109/TAC.2015.2503566
14. H. Li, P. Shi, D. Yao, L. Wu, Observer-based adaptive sliding mode control for nonlinear Markovian jump systems. Automatica **64**, 133–142 (2016)
15. H. Li, Y. Shi, W. Yan, On neighbor information utilization in distributed receding horizon control for consensus-seeking. IEEE Trans. Cybern. (2016). doi:10.1109/TCYB.2015.2459719
16. H. Li, Y. Shi, W. Yan, Distributed receding horizon control of constrained nonlinear vehicle formations with guaranteed γ-gain stability. Automatica **68**, 148–154 (2016)
17. Y.W. Mao, F. Ding, Multi-innovation stochastic gradient identification for Hammerstein controlled autoregressive autoregressive systems based on the filtering technique. Nonlinear Dyn. **79**(3), 1745–1755 (2015)
18. J. Pan, X.H. Yang, H.F. Cai, B.X. Mu, Image noise smoothing using a modified Kalman filter. Neurocomputing **173**, 1625–1629 (2016)
19. R.Y. Ruan, C.L. Yang, H.X. Chen, B. Li, On-line order estimation and parameter identification for linear stochastic feedback control systems. Automatica **39**(2), 243–253 (2003)
20. C. Sun, F.L. Wang, X.Q. He, Robust fault estimation for takagi-sugeno nonlinear systems with time-varying state delay. Circuits Syst. Signal Process. **34**(2), 641–661 (2015)
21. J. van Wingerden, M. Verhaegen, Subspace identification of bilinear and LPV systems for open-and closed-loop data. Automatica **45**(2), 372–381 (2009)
22. D.Q. Wang, Hierarchical parameter estimation for a class of MIMO Hammerstein systems based on the reframed models. Appl. Math. Lett. **57**, 13–19 (2016)
23. D.Q. Wang, F. Ding, Parameter estimation algorithms for multivariable Hammerstein CARMA systems. Inf. Sci. **355–356**(10), 237–248 (2016)
24. Y.J. Wang, F. Ding, Recursive least squares algorithm and gradient algorithm for Hammerstein–Wiener systems using the data filtering. Nonlinear Dyn. **84**(2), 1045–1053 (2016)
25. X.H. Wang, F. Ding, Recursive parameter and state estimation for an input nonlinear state space system using the hierarchical identification principle. Signal Process. **117**, 208–218 (2015)
26. Y.J. Wang, F. Ding, Recursive parameter estimation algorithms and convergence for a class of nonlinear systems with colored noise. Circuits Syst. Signal Process. **35**(10), 3461–3481 (2016)
27. Y.J. Wang, F. Ding, Novel data filtering based parameter identification for multiple-input multiple-output systems using the auxiliary model. Automatica **71**, 308–313 (2016)
28. Y.J. Wang, F. Ding, The filtering based iterative identification for multivariable systems. IET Control Theory Appl. **10**(8), 894–902 (2016)
29. Y.J. Wang, F. Ding, The auxiliary model based hierarchical gradient algorithms and convergence analysis using the filtering technique. Signal Process. **128**, 212–221 (2016)
30. T.Z. Wang, J. Qi, H. Xu et al., Fault diagnosis method based on FFT-RPCA-SVM for cascaded-multilevel inverter. ISA Trans. **60**, 156–163 (2016)
31. J. Wang, A. Sano, D. Shook, T. Chen, B. Huang, A blind approach to closed-loop identification of Hammerstein systems. Int. J. Control **80**(2), 302–313 (2007)
32. C. Wang, T. Tang, Several gradient-based iterative estimation algorithms for a class of nonlinear systems using the filtering technique. Nonlinear Dyn. **77**(3), 769–780 (2014)
33. T.Z. Wang, H. Wu, M.Q. Ni et al., An adaptive confidence limit for periodic non-steady conditions fault detection. Mech. Syst. Signal Process. **72–73**, 328–345 (2016)
34. D.Q. Wang, W. Zhang, Improved least squares identification algorithm for multivariable Hammerstein systems. J. Franklin Inst. **352**(11), 5292–5307 (2015)
35. C. Wang, L. Zhu, Parameter identification of a class of nonlinear systems based on the multi-innovation identification theory. J. Franklin Inst. **352**(10), 4624–4637 (2015)
36. X.K. Wei, M. Verhaegen, T. van Engelen, Sensor fault detection and isolation for wind turbines based on subspace identification and Kalman filter techniques. Int. J. Adapt. Control Signal Process. **24**(8), 687–707 (2010)

37. D.H. Wu, Y.Y. Li, Fault diagnosis of variable pitch for wind turbines based on the multi-innovation forgetting gradient identification algorithm. Nonlinear Dyn. **79**(3), 2069–2077 (2014)
38. L. Xu, The damping iterative parameter identification method for dynamical systems based on the sine signal measurement. Signal Process. **120**, 660–667 (2016)
39. L. Xu, Application of the Newton iteration algorithm to the parameter estimation for dynamical systems. J. Comput. Appl. Math. **288**, 33–43 (2015)
40. L. Xu, A proportional differential control method for a time-delay system using the Taylor expansion approximation. Appl. Math. Comput. **236**, 391–399 (2014)
41. L. Xu, L. Chen, W.L. Xiong, Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration. Nonlinear Dyn. **79**(3), 2155–2163 (2015)
42. L. Xu, F. Ding, Recursive least squares and multi-innovation stochastic gradient parameter estimation methods for signal modeling. Circuits Syst. Signal Process (2017). doi:10.1007/s00034-016-0378-4
43. G. Zhang, X. Zhang, H. Pang, Multi-innovation auto-constructed least squares identification for 4 DOF ship manoeuvring modelling with full-scale trial data. ISA Trans. **58**, 186–195 (2015)