Circuits Syst Signal Process (2017) 36:1393–1405
DOI 10.1007/s00034-016-0367-7

Hierarchical Stochastic Gradient Algorithm and its Performance Analysis for a Class of Bilinear-in-Parameter Systems

Feng Ding[1] · Xuehai Wang[2]

Received: 15 March 2016 / Revised: 6 July 2016 / Accepted: 8 July 2016 / Published online: 21 July 2016

Abstract This paper considers the parameter identification for a special class of nonlinear systems, i.e., bilinear-in-parameter systems. Based on the hierarchical identification principle, a hierarchical stochastic gradient (HSG) estimation algorithm is presented. The basic idea is to decompose a bilinear-in-parameter system into two subsystems and to derive the HSG identification algorithm for estimating the system parameters by replacing the unknown variables in the information vectors with their estimates obtained at the previous time. The convergence analysis of the proposed algorithm indicates that the parameter estimation errors converge to zero under persistent excitation conditions. The simulation results show that the proposed algorithm is effective.

Keywords Parameter estimation · Gradient search · Hierarchical identification · Performance analysis · Bilinear-in-parameter system

1 Introduction

Parameter estimation algorithms are often obtained through minimizing a criterion function. The gradient search, least squares search and Newton search are the use-

✉ Feng Ding
fding12@126.com; fdingab@126.com

Xuehai Wang
xuehaiwang735@163.com

1 School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, People's Republic of China

2 College of Mathematics and Information Science, Xinyang Normal University, Xinyang 464000, People's Republic of China

ful tools for solving nonlinear optimization problems [15,23,44–46]. Nonlinearities exist widely in industrial processes [21]. Typical nonlinear systems are the block-oriented systems, including input nonlinear systems [25,30,38,42], output nonlinear systems [11,41] or Wiener nonlinear systems [9], input–output (i.e., Hammerstein–Wiener) nonlinear systems [2,31] and feedback nonlinear systems [14]. When the static nonlinear part of the block-oriented systems can be expressed as a linear combination of the known basis functions, the corresponding systems are the Hammerstein systems, Wiener systems and their combinations [16,40]. A direct method of identifying the block-oriented nonlinear systems is the over-parametrization method [3]. By re-parameterizing the nonlinear systems, the output appears to be linear on the unknown parameter space so that any linear identification algorithms can be applied [4]. However, the resulting identification model contains the cross-products between the parameters in the nonlinear part and those in linear part, leading to estimate more parameters than the nonlinear system.

In the area of system identification, linear-in-parameter output error moving average systems are common, for which Wang and Tang [36] presented a recursive least squares estimation algorithm and discussed several gradient-based iterative estimation algorithms using the filtering technique [37]; Wang and Zhu [39] presented a multi-innovation parameter estimation algorithm. The system that includes the product terms of parameters is called the bilinear-in-parameter system. Bai and Liu [5] discussed the least squares solution of the normalized iterative method, the over-parametrization method and the numerical method for bilinear-in-parameter systems; Wang et al. [24] revisited the unweighted least squares solution and extended to identify the case of colored noise; Abrahamsson et al. [1] presented a two-stage method based on the approximation of a weighting matrix and discussed the applications to submarine detection. Other methods include the Kalman filtering-based identification approaches [10,23].

The convergence of identification algorithms is a basic topic for system identification and attracts much attention. Recently, an auxiliary model-based recursive least squares algorithm and an auxiliary model-based hierarchical gradient algorithm have been proposed for dual-rate state space systems [12] and for multivariable Box–Jenkins systems using the data filtering [32–34]. The modeling and multi-innovation parameter identification has been proposed for Hammerstein nonlinear state space systems using the filtering technique [35]; a recursive parameter and state estimation algorithm has been proposed for an input nonlinear state space system using the hierarchical identification principle [29]; an auxiliary model-based gradient algorithm has been reported for the time-delay systems by transforming the input–output representation into a regression model and its convergence was studied [13]. The convergence analysis of the hierarchical least squares algorithm has been analyzed for bilinear-in-parameter systems [26]. On the basis of the work in [26], this paper derives a hierarchical stochastic gradient (HSG) algorithm for bilinear-in-parameter systems based on the decomposition idea and analyzes its performances.

The rest of this paper is organized as follows. Section 2 presents an HSG algorithm for bilinear-in-parameter systems. Section 3 analyzes the performance of the HSG algorithm. Section 4 provides an illustrative example to show that the proposed algorithm is effective. Finally, a brief summary of the main contents is given in Sect. 5.

2 System Description and the HSG Algorithm

Consider the following bilinear-in-parameter systems [5,26],

$$y(t) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{b} + v(t), \tag{1}$$

where $y(t)$ is the system output, $\boldsymbol{F}(t) \in \mathbb{R}^{m\times n}$ is composed of available measurement data, $v(t)$ is a white noise sequence with zero mean and finite variance σ^2 and $\boldsymbol{a} = [a_1, a_2, \ldots, a_m]^{\mathrm{T}} \in \mathbb{R}^m$ and $\boldsymbol{b} = [b_1, b_2, \ldots, b_n]^{\mathrm{T}} \in \mathbb{R}^n$ are the unknown parameter vectors to be estimated.

For the identification model in (1), assume that m and n are known, and $y(t) = 0$, $v(t) = 0$ for $t \leqslant 0$. Note that for any pair $\lambda\boldsymbol{a}, \boldsymbol{b}/\lambda$, the system in (1) has the identical input–output relationship, so the constant λ has to be fixed. Without generality, we adopt the following assumption.

Assumption 1 $\lambda = \|\boldsymbol{b}\|$, and the first element of $\boldsymbol{b}$ is positive, i.e., $b_1 > 0$, where the norm of the vector $\boldsymbol{X}$ is defined by $\|\boldsymbol{X}\|^2 := \mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$.

Define the vector $\boldsymbol{\psi}(t) := \boldsymbol{F}(t)\boldsymbol{b} \in \mathbb{R}^m$, $\boldsymbol{\varphi}(t) := \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{a} \in \mathbb{R}^n$. Then Eq. (1) can be written as

$$y(t) = \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{a} + v(t), \tag{2}$$

or

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{b} + v(t). \tag{3}$$

Define the following two cost functions:

$$J_1(\boldsymbol{a}) := \|y(t) - \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{a}\|^2,$$
$$J_2(\boldsymbol{b}) := \|y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{b}\|^2.$$

Using the negative gradient search and minimizing $J_1(\boldsymbol{a})$ and $J_2(\boldsymbol{b})$, we obtain the estimates $\hat{\boldsymbol{a}}(t)$ of $\boldsymbol{a}$ in Subsystem (2) and $\hat{\boldsymbol{b}}(t)$ of $\boldsymbol{b}$ in Subsystem (3) at time t:

$$\hat{\boldsymbol{a}}(t) = \hat{\boldsymbol{a}}(t-1) + \frac{\boldsymbol{\psi}(t)}{r_1(t)}\left[y(t) - \boldsymbol{\psi}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)\right], \tag{4}$$

$$r_1(t) = r_1(t-1) + \|\boldsymbol{\psi}(t)\|^2,\ r_1(0) = 1, \tag{5}$$

$$\hat{\boldsymbol{b}}(t) = \hat{\boldsymbol{b}}(t-1) + \frac{\boldsymbol{\varphi}(t)}{r_2(t)}\left[y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1)\right], \tag{6}$$

$$r_2(t) = r_2(t-1) + \|\boldsymbol{\varphi}(t)\|^2,\ r_2(0) = 1. \tag{7}$$

Since the vectors $\boldsymbol{\psi}(t)$ and $\boldsymbol{\varphi}(t)$ contain the unknown parameter vectors $\boldsymbol{b}$ and $\boldsymbol{a}$, the algorithm in (4)–(7) is impossible to implement. This problem can be solved by replacing $\boldsymbol{b}$ and $\boldsymbol{a}$ with their corresponding estimates $\hat{\boldsymbol{b}}(t-1)$ and $\hat{\boldsymbol{a}}(t-1)$ at time

$t-1$. Letting $\hat{\boldsymbol{\psi}}(t) := \boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1) \in \mathbb{R}^m$ and $\hat{\boldsymbol{\varphi}}(t) := \boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1) \in \mathbb{R}^n$, we have the following HSG algorithm for bilinear-in-parameter systems in (1):

$$\hat{\boldsymbol{a}}(t) = \hat{\boldsymbol{a}}(t-1) + \frac{\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)}{r_1(t)}\left[y(t) - \hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)\right], \tag{8}$$

$$r_1(t) = r_1(t-1) + \|\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)\|^2,\ r_1(0) = 1, \tag{9}$$

$$\hat{\boldsymbol{b}}(t) = \hat{\boldsymbol{b}}(t-1) + \frac{\boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)}{r_2(t)}\left[y(t) - \hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)\right], \tag{10}$$

$$r_2(t) = r_2(t-1) + \|\boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)\|^2,\ r_2(0) = 1. \tag{11}$$

The initial values are taken to be $\hat{\boldsymbol{a}}(0) = \mathbf{1}_m/p_0$, $\hat{\boldsymbol{b}}(0) = \mathbf{1}_n/p_0$, where p_0 is a large number, e.g., $p_0 = 10^6$.

3 The Convergence Analysis

Lemma 1 [8] *Assume that the nonnegative sequences $T(t)$, $\eta(t)$ and $\zeta(t)$ satisfy the inequality*

$$T(t) \leqslant T(t-1) + \eta(t) - \zeta(t)$$

and $\sum_{t=1}^{\infty}\eta(t) < \infty$, then we have $\sum_{t=1}^{\infty}\zeta(t) < \infty$ and $T(t)$ is bounded.

The proof of Lemma 1 is straightforward and hence omitted.

Theorem 1 *For the system in (1) and the HSG algorithm in (8)–(11), assume that $v(t)$ is a white noise sequence with zero mean and variances σ^2, and there exist an integer N and two positive constants c_1 and c_2 such that the following persistent excitation conditions hold:*

$$\text{(A1)}\quad \sum_{j=0}^{N-1}\frac{\hat{\boldsymbol{\psi}}(t+j)\hat{\boldsymbol{\psi}}^{T}(t+j)}{r_1(t+j)} \geqslant c_1\boldsymbol{I}_m,\ \text{a.s.,}$$

$$\text{(A2)}\quad \sum_{j=0}^{N-1}\frac{\hat{\boldsymbol{\varphi}}(t+j)\hat{\boldsymbol{\varphi}}^{T}(t+j)}{r_2(t+j)} \geqslant c_2\boldsymbol{I}_n,\ \text{a.s.,}$$

Then the parameter estimation errors converge to zero, i.e.,

$$\|\hat{\boldsymbol{a}}(t) - \boldsymbol{a}\| \to 0,\quad \|\hat{\boldsymbol{b}}(t) - \boldsymbol{b}\| \to 0.$$

Proof Define two parameter error vectors:

$$\tilde{\boldsymbol{a}}(t) := \hat{\boldsymbol{a}}(t) - \boldsymbol{a} \in \mathbb{R}^m, \tag{12}$$

$$\tilde{\boldsymbol{b}}(t) := \hat{\boldsymbol{b}}(t) - \boldsymbol{b} \in \mathbb{R}^n. \tag{13}$$

Substituting (1) and (8) into (12), we have

$$\begin{aligned}\tilde{\boldsymbol{a}}(t) &= \tilde{\boldsymbol{a}}(t-1)+\frac{\hat{\boldsymbol{\psi}}(t)}{r_1(t)}\left[y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)\right]\\ &= \tilde{\boldsymbol{a}}(t-1)+\frac{\hat{\boldsymbol{\psi}}(t)}{r_1(t)}\left[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{b}-\hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)+v(t)\right] \qquad (14)\\ &= \tilde{\boldsymbol{a}}(t-1)+\frac{\hat{\boldsymbol{\psi}}(t)}{r_1(t)}\left[-\tilde{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\tilde{\boldsymbol{b}}(t-1)+v(t)\right]\\ &=: \tilde{\boldsymbol{a}}(t-1)+\frac{\hat{\boldsymbol{\psi}}(t)}{r_1(t)}\left[-\tilde{y}_1(t)-\xi_1(t)+v(t)\right], \qquad (15)\end{aligned}$$

where

$$\tilde{y}_1(t) := \tilde{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)\in\mathbb{R}, \qquad (16)$$
$$\xi_1(t) := \boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\tilde{\boldsymbol{b}}(t-1)\in\mathbb{R}. \qquad (17)$$

Taking the norm of both sides of (15) and using (16) yield

$$\begin{aligned}\|\tilde{\boldsymbol{a}}(t)\|^2 &= \left\|\tilde{\boldsymbol{a}}(t-1)+\frac{\hat{\boldsymbol{\psi}}(t)}{r_1(t)}\left[-\tilde{y}_1(t)-\xi_1(t)+v(t)\right]\right\|^2\\ &= \|\tilde{\boldsymbol{a}}(t-1)\|^2+\frac{2\tilde{\boldsymbol{a}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\psi}}(t)}{r_1(t)}\left[-\tilde{y}_1(t)-\xi_1(t)+v(t)\right]\\ &\quad+\frac{\left\|\hat{\boldsymbol{\psi}}(t)\right\|^2}{r_1^2(t)}\left[-\tilde{y}_1(t)-\xi_1(t)+v(t)\right]^2\\ &= \|\tilde{\boldsymbol{a}}(t-1)\|^2+\frac{2\tilde{y}_1(t)}{r_1(t)}\left[-\tilde{y}_1(t)-\xi_1(t)+v(t)\right]\\ &\quad+\frac{\left\|\hat{\boldsymbol{\psi}}(t)\right\|^2}{r_1^2(t)}\left[-\tilde{y}_1(t)-\xi_1(t)+v(t)\right]^2. \qquad (18)\end{aligned}$$

Define $\tilde{y}_2(t) := \hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\tilde{\boldsymbol{b}}(t-1)\in\mathbb{R}$, $\xi_2(t) := \tilde{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\boldsymbol{b}\in\mathbb{R}$. Similarly, we have

$$\begin{aligned}\tilde{\boldsymbol{b}}(t) &= \tilde{\boldsymbol{b}}(t-1)+\frac{\hat{\boldsymbol{\varphi}}(t)}{r_2(t)}\left[-\tilde{y}_2(t)-\xi_2(t)+v(t)\right],\\ \left\|\tilde{\boldsymbol{b}}(t)\right\|^2 &= \left\|\tilde{\boldsymbol{b}}(t-1)\right\|^2+\frac{2\tilde{y}_2(t)}{r_2(t)}\left[-\tilde{y}_2(t)-\xi_2(t)+v(t)\right]\\ &\quad+\frac{\|\hat{\boldsymbol{\varphi}}(t)\|^2}{r_2^2(t)}\left[-\tilde{y}_2(t)-\xi_2(t)+v(t)\right]^2. \qquad (19)\end{aligned}$$

Let $T(t) := \|\tilde{\boldsymbol{a}}(t)\|^2 + \|\tilde{\boldsymbol{b}}(t)\|^2$. Using (18), (19), (9) and (11) gives

$$
\begin{aligned}
T(t) &= \|\tilde{\boldsymbol{a}}(t-1)\|^2 + \frac{2\tilde{y}_1(t)}{r_1(t)}\left[-\tilde{y}_1(t) - \xi_1(t) + v(t)\right] \\
&\quad + \frac{\left\|\hat{\boldsymbol{\psi}}(t)\right\|^2}{r_1^2(t)}\left[\tilde{y}_1^2(t) + \xi_1^2(t) + v^2(t) + 2\tilde{y}_1(t)\xi_1(t) - 2\tilde{y}_1(t)v(t) - 2\xi_1(t)v(t)\right] \\
&\quad + \left\|\tilde{\boldsymbol{b}}(t-1)\right\|^2 + \frac{2\tilde{y}_2(t)}{r_2(t)}\left[-\tilde{y}_2(t) - \xi_2(t) + v(t)\right] \\
&\quad + \frac{\|\hat{\boldsymbol{\varphi}}(t)\|^2}{r_2^2(t)}\left[\tilde{y}_2^2(t) + \xi_2^2(t) + v^2(t) + 2\tilde{y}_2(t)\xi_2(t) - 2\tilde{y}_2(t)v(t) - 2\xi_2(t)v(t)\right] \\
&= T(t-1) - \left[\frac{2}{r_1(t)} - \frac{\left\|\hat{\boldsymbol{\psi}}(t)\right\|^2}{r_1^2(t)}\right]\tilde{y}_1^2(t) \\
&\quad + 2\left[\frac{1}{r_1(t)} - \frac{\left\|\hat{\boldsymbol{\psi}}(t)\right\|^2}{r_1^2(t)}\right]\tilde{y}_1(t)\left[v(t) - \xi_1(t)\right] \\
&\quad + \frac{\left\|\hat{\boldsymbol{\psi}}(t)\right\|^2}{r_1^2(t)}\left[\xi_1^2(t) + v^2(t) - 2\xi_1(t)v(t)\right] - \left[\frac{2}{r_2(t)} - \frac{\|\hat{\boldsymbol{\varphi}}(t)\|^2}{r_2^2(t)}\right]\tilde{y}_2^2(t) \\
&\quad + 2\left[\frac{1}{r_2(t)} - \frac{\|\hat{\boldsymbol{\varphi}}(t)\|^2}{r_2^2(t)}\right]\tilde{y}_2(t)\left[v(t) - \xi_2(t)\right]\frac{\|\hat{\boldsymbol{\varphi}}(t)\|^2}{r_2^2(t)} \\
&\qquad \left[\xi_2^2(t) + v^2(t) - 2\xi_2(t)v(t)\right] \\
&= T(t-1) - \left[\frac{r_1(t) + r_1(t-1)}{r_1^2(t)}\right]\tilde{y}_1^2(t) + \frac{2r_1(t-1)}{r_1^2(t)}\tilde{y}_1(t)\left[v(t) - \xi_1(t)\right] \\
&\quad + \frac{\left\|\hat{\boldsymbol{\psi}}(t)\right\|^2}{r_1^2(t)}\left[\xi_1^2(t) + v^2(t) - 2\xi_1(t)v(t)\right] - \left[\frac{r_2(t) + r_2(t-1)}{r_2^2(t)}\right]\tilde{y}_2^2(t) \\
&\quad + \frac{2r_2(t-1)}{r_2^2(t)}\tilde{y}_2(t)\left[v(t) - \xi_2(t)\right] + \frac{\|\hat{\boldsymbol{\varphi}}(t)\|^2}{r_2^2(t)}\xi_2^2(t) + v^2(t) - 2\xi_2(t)v(t)] \\
&\leqslant T(t-1) - \frac{1}{r_1(t)}\tilde{y}_1^2(t) + \frac{2r_1(t-1)}{r_1^2(t)}\tilde{y}_1(t)\left[v(t) - \xi_1(t)\right] \\
&\quad + \frac{\left\|\hat{\boldsymbol{\psi}}(t)\right\|^2}{r_1^2(t)}\left[\xi_1^2(t) + v^2(t) - 2\xi_1(t)v(t)\right] - \frac{1}{r_2(t)}\tilde{y}_2^2(t) \\
&\quad + \frac{2r_2(t-1)}{r_2^2(t)}\tilde{y}_2(t)\left[v(t) - \xi_2(t)\right] + \frac{\|\hat{\boldsymbol{\varphi}}(t)\|^2}{r_2^2(t)}\left[\xi_2^2(t) + v^2(t) - 2\xi_2(t)v(t)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= T(t-1) - \gamma(t) - \frac{1}{r_1(t)}\tilde{y}_1^2(t) + \frac{2r_1(t-1)}{r_1^2(t)}\tilde{y}_1(t)v(t) \\
&\quad + \frac{\left\|\hat{\boldsymbol{\psi}}(t)\right\|^2}{r_1^2(t)}\left[\xi_1^2(t) + v^2(t)\right] \\
&\quad - \frac{1}{r_2(t)}\tilde{y}_2^2(t) + \frac{2r_2(t-1)}{r_2^2(t)}\tilde{y}_2(t)v(t) \\
&\quad + \frac{\left\|\hat{\boldsymbol{\varphi}}(t)\right\|^2}{r_2^2(t)}\left[\xi_2^2(t) + v^2(t) - 2\xi_2(t)v(t)\right],
\end{aligned} \tag{20}
$$

where

$$
\gamma(t) := \frac{2r_1(t-1)}{r_1^2(t)}\tilde{y}_1(t)\xi_1(t) + \frac{2r_2(t-1)}{r_2^2(t)}\tilde{y}_2(t)\xi_2(t).
$$

When $\xi_1^2 > \varepsilon$ or $\xi_2^2 > \varepsilon$ or $\gamma(t) < 0$ (ε is a given positive number), we let $\tilde{\boldsymbol{a}}(t) := \tilde{\boldsymbol{a}}(t-1)$ and $\tilde{\boldsymbol{b}}(t) := \tilde{\boldsymbol{b}}(t-1)$, and thus we have $T(t) = T(t-1)$. When $\xi_1^2 \leqslant \varepsilon$ and $\xi_2^2 \leqslant \varepsilon$ and $\gamma(t) \geqslant 0$, since $v(t)$ is a white noise with zero mean and variance σ^2, and $\boldsymbol{F}(t)$, $\hat{\boldsymbol{a}}(t-1)$, $\hat{\boldsymbol{b}}(t-1)$, $r_1(t)$, $r_2(t)$, $\xi_1(t)$ and $\xi_2(t)$ are independent of $v(t)$, taking expectation of both sides of (20), we have

$$
\begin{aligned}
\mathrm{E}[T(t)] \leqslant \mathrm{E}[T(t-1)] - \mathrm{E}\left[\frac{\tilde{y}_1^2(t)}{r_1(t)} + \frac{\tilde{y}_2^2(t)}{r_2(t)}\right] \\
+\mathrm{E}\left[\frac{\left\|\hat{\boldsymbol{\psi}}(t)\right\|^2}{r_1^2(t)} + \frac{\left\|\hat{\boldsymbol{\varphi}}(t)\right\|^2}{r_2^2(t)}\right](\sigma^2 + \varepsilon),
\end{aligned} \tag{21}
$$

From (9), we have

$$
\begin{aligned}
\sum_{t=1}^{\infty}\frac{\left\|\hat{\boldsymbol{\psi}}(t)\right\|^2}{r_1^2(t)} &\leqslant \sum_{t=1}^{\infty}\frac{\left\|\hat{\boldsymbol{\psi}}(t)\right\|^2}{r_1(t)r_1(t-1)} = \sum_{t=1}^{\infty}\frac{r_1(t) - r_1(t-1)}{r_1(t)r_1(t-1)} \\
&= \sum_{t=1}^{\infty}\left[\frac{1}{r_1(t-1)} - \frac{1}{r_1(t)}\right] = \frac{1}{r_1(0)} - \frac{1}{r_1(\infty)} < \infty, \ \text{a.s.}
\end{aligned}
$$

Similarly, from (11), we have

$$
\sum_{t=1}^{\infty}\frac{\left\|\hat{\boldsymbol{\varphi}}(t)\right\|^2}{r_2^2(t)} < \infty, \ \text{a.s.}
$$

Hence, summation of the last term of the right-hand side of (21) from $t = 1$ to ∞ is finite. Applying Lemma 1 to (21), we conclude that $\mathrm{E}[T(t)]$ converges to a constant.

So there exist a constant $C > 0$ and t_0 such that $\mathrm{E}[T(t)] \leqslant C$ for $t > t_0$. From (21), it follows that

$$\sum_{t=1}^{\infty}\left[\frac{\tilde{y}_1^2(t)}{r_1(t)}+\frac{\tilde{y}_2^2(t)}{r_2(t)}\right]<\infty.$$

Note that $r_1(t) > 0$ and $r_2(t) > 0$, we have

$$\sum_{t=1}^{\infty}\frac{\tilde{y}_1^2(t)}{r_1(t)}<\infty,\quad \sum_{t=1}^{\infty}\frac{\tilde{y}_2^2(t)}{r_2(t)}<\infty,\quad \lim_{t\to\infty}\frac{\tilde{y}_1^2(t)}{r_1(t)}=0,\quad \lim_{t\to\infty}\frac{\tilde{y}_2^2(t)}{r_2(t)}=0. \tag{22}$$

Define the identification innovation

$$e(t) := y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)\in\mathbb{R},$$

From (14), we have

$$\tilde{\boldsymbol{a}}(t)=\tilde{\boldsymbol{a}}(t-1)+\frac{\hat{\boldsymbol{\psi}}(t)}{r_1(t)}e(t). \tag{23}$$

Replacing t in (23) with $t + j$ and successive substitutions give

$$\tilde{\boldsymbol{a}}(t+j)=\tilde{\boldsymbol{a}}(t)+\sum_{i=1}^{j}\frac{\hat{\boldsymbol{\psi}}(t+i)}{r_1(t+i)}e(t+i). \tag{24}$$

Using (16), it follows that

$$\begin{aligned}\tilde{y}_1(t)&=\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)\tilde{\boldsymbol{a}}(t-1),\\ \tilde{y}_1(t+j)&=\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t+j)\tilde{\boldsymbol{a}}(t+j-1).\end{aligned} \tag{25}$$

Substituting (24) into (25) gives

$$\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t+j)\tilde{\boldsymbol{a}}(t)=\tilde{y}_1(t+j)-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t+j)\sum_{i=1}^{j-1}\frac{\hat{\boldsymbol{\psi}}(t+i)}{r_1(t+i)}e(t+i), \tag{26}$$

Squaring and summing for j from $j = 1$ to $j = N - 1$, dividing by $r_1(t + j)$, and using (A1), (24) and (26), we have

$$\begin{aligned}c_1\|\tilde{\boldsymbol{a}}(t)\|^2 &\leqslant \tilde{\boldsymbol{a}}^{\mathrm{T}}(t)\left[\sum_{j=1}^{N-1}\frac{\hat{\boldsymbol{\psi}}(t+j)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t+j)}{r_1(t+j)}\right]\tilde{\boldsymbol{a}}(t)\\ &=\sum_{j=1}^{N-1}\frac{\tilde{\boldsymbol{a}}^{\mathrm{T}}(t)\hat{\boldsymbol{\psi}}(t+j)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t+j)\tilde{\boldsymbol{a}}(t)}{r_1(t+j)}\end{aligned}$$

$$\leqslant \sum_{j=1}^{N-1}\left[\frac{2\tilde{y}_1^2(t+j)}{r_1(t+j)}+\frac{2\left\|\hat{\boldsymbol{\psi}}(t+j)\right\|^2}{r_1(t+j)}\left\|\sum_{i=1}^{j-1}\frac{\hat{\boldsymbol{\psi}}(t+i)}{r_1(t+i)}e(t+i)\right\|^2\right]$$
$$=\sum_{j=1}^{N-1}\left[\frac{2\tilde{y}_1^2(t+j)}{r_1(t+j)}+\frac{2\left\|\hat{\boldsymbol{\psi}}(t+j)\right\|^2}{r_1(t+j)}\|\tilde{\boldsymbol{a}}(t+j-1)-\tilde{\boldsymbol{a}}(t)\|^2\right]$$
$$\leqslant \sum_{j=1}^{N-1}\left[\frac{2\tilde{y}_1^2(t+j)}{r_1(t+j)}+\frac{4\left\|\hat{\boldsymbol{\psi}}(t+j)\right\|^2}{r_1(t+j)}\left(\|\tilde{\boldsymbol{a}}(t+j-1)\|^2+\|\tilde{\boldsymbol{a}}(t)\|^2\right)\right], \tag{27}$$

Since $\mathrm{E}[T(t)]=\mathrm{E}[\|\tilde{\boldsymbol{a}}(t)\|^2+\|\tilde{\boldsymbol{b}}(t)\|^2]\leqslant C$, we have $\mathrm{E}[\|\tilde{\boldsymbol{a}}(t)\|^2]\leqslant C$. Taking the expectation and the limit of both sides of (27), it follows

$$\lim_{t\to\infty}\mathrm{E}\left[\|\tilde{\boldsymbol{a}}(t)\|^2\right]\leqslant \lim_{t\to\infty}\frac{1}{c_1}\mathrm{E}\left\{\sum_{j=1}^{N-1}\left[\frac{2\tilde{y}_1^2(t+j)}{r_1(t+j)}+\frac{8C\left\|\hat{\boldsymbol{\psi}}(t+j)\right\|^2}{r_1(t+j)}\right]\right\}.$$

Assume that $\lim_{t\to\infty}\|\hat{\boldsymbol{\psi}}(t+j)\|^2/r_1(t+j)=0$. Using (22) gives $\lim_{t\to\infty}\mathrm{E}\left[\|\tilde{\boldsymbol{a}}(t)\|^2\right]=0$. Similarly, we can obtain $\lim_{t\to\infty}\mathrm{E}[\|\tilde{\boldsymbol{b}}(t)\|^2]=0$. This completes the proof. □

In order to improve the convergence rate of the HSG algorithm, we introduce a forgetting factor λ ($0\leqslant\lambda\leqslant 1$) in (8)–(11) and the corresponding algorithm is called the forgetting factor HSG (FF-HSG) algorithm, which is as follows:

$$\hat{\boldsymbol{a}}(t)=\hat{\boldsymbol{a}}(t-1)+\frac{\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)}{r_1(t)}\left[y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)\right], \tag{28}$$

$$r_1(t)=\lambda r_1(t-1)+\left\|\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)\right\|^3,\ r_1(0)=1, \tag{29}$$

$$\hat{\boldsymbol{b}}(t)=\hat{\boldsymbol{b}}(t-1)+\frac{\boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)}{r_3(t)}\left[y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{b}}(t-1)\right], \tag{30}$$

$$r_3(t)=\lambda r_3(t-1)+\left\|\boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)\right\|^3,\ r_3(0)=1. \tag{31}$$

Obviously, when the forgetting factor $\lambda=1$, the FF-HSG algorithm is reduced to the HSG algorithm; when $\lambda=0$, the FF-HSG algorithm is degenerated to the hierarchical projection algorithm.

4 Example

Consider the following bilinear-in-parameter system with $m = 2$ and $n = 3$,

$$\begin{aligned} y(t) &= \boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{b} + v(t), \\ \boldsymbol{F}(t) &= \begin{bmatrix} \boldsymbol{f}(u(t-1)) \\ \boldsymbol{f}(u(t-2)) \end{bmatrix} = \begin{bmatrix} u(t-1)\ u^2(t-1)\ u^3(t-1) \\ u(t-2)\ u^2(t-2)\ u^3(t-2) \end{bmatrix}. \\ \boldsymbol{a} &= [2.06, 1.00]^{\mathrm{T}}, \quad \boldsymbol{b} = [0.70, \sqrt{0.02}, 0.70]^{\mathrm{T}}, \\ \boldsymbol{\theta} &= [\boldsymbol{a}, \boldsymbol{b}]^{\mathrm{T}} = [2.06, 1.00, 0.70, \sqrt{0.02}, 0.70]^{\mathrm{T}}, \end{aligned}$$

where $\|\boldsymbol{b}\| = 1$. In simulation, we generate a persistent excitation sequence with zero mean and unit variance as the input $u(t)$ and take $v(t)$ to be an uncorrelated noise sequence with zero mean and variance $\sigma^2 = 0.10^2$. Taking the data length $L = 3000$ and using the HSG algorithm to generate the parameter estimates $\hat{\boldsymbol{a}}(t)$ and $\hat{\boldsymbol{b}}(t)$ from the input–output data $\{y(t), \boldsymbol{F}(t)\colon t = 1, 2, 3 \ldots\}$, the parameter estimates and their estimation errors are given in Tables 1, 2 and 3, and the estimation error $\delta := \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|/\|\boldsymbol{\theta}\|$ versus t is shown in Fig. 1.

From Tables 1, 2, 3 and Fig. 1, we can draw the following conclusions.

1. The estimation errors become smaller with time t increasing—see Tables 1, 2 and 3.

Table 1 HSG parameter estimates and errors

t	a_1	a_2	b_1	b_2	b_3	δ (%)
100	2.47333	0.63015	−0.06240	−0.37051	0.96447	44.22049
200	2.34054	0.65929	−0.06626	−0.31078	0.87249	40.34286
500	2.15692	0.72103	−0.06489	−0.24159	0.77122	36.32918
1000	2.15413	0.76091	−0.05511	−0.21729	0.78665	35.17235
2000	2.14295	0.81131	−0.04479	−0.18204	0.80072	33.76856
3000	2.14314	0.84939	−0.03665	−0.15628	0.81488	32.85750
True values	2.06000	1.00000	0.70000	0.14142	0.70000	

Table 2 FF-HSG parameter estimates and errors ($\lambda = 0.98$)

t	a_1	a_2	b_1	b_2	b_3	δ (%)
100	2.33101	0.74668	−0.13624	−0.33429	0.78811	41.41640
200	2.10318	0.90948	−0.09174	−0.19064	0.78149	34.74696
500	2.02914	0.98394	−0.00056	0.04603	0.81754	28.71770
1000	2.08670	1.05831	0.15974	0.13009	0.86222	22.72502
2000	2.06186	1.01743	0.33061	0.13268	0.81696	15.52644
3000	2.03174	0.99991	0.43990	0.14740	0.77379	10.88153
True values	2.06000	1.00000	0.70000	0.14142	0.70000	

Table 3 FF-HSG parameter estimates and errors ($\lambda = 0.95$)

t	a_1	a_2	b_1	b_2	b_3	δ (%)
100	1.99873	0.96017	−0.21533	−0.25200	0.71296	39.98278
200	2.09083	1.15735	−0.08182	−0.00391	0.82813	32.86776
500	2.03160	1.01576	0.10155	0.17291	0.78998	24.28724
1000	2.10602	1.07287	0.38946	0.15304	0.82141	13.79053
2000	2.07128	1.01429	0.58451	0.13520	0.74008	4.95243
3000	2.06078	1.00344	0.65243	0.13840	0.70962	1.95123
True values	2.06000	1.00000	0.70000	0.14142	0.70000	

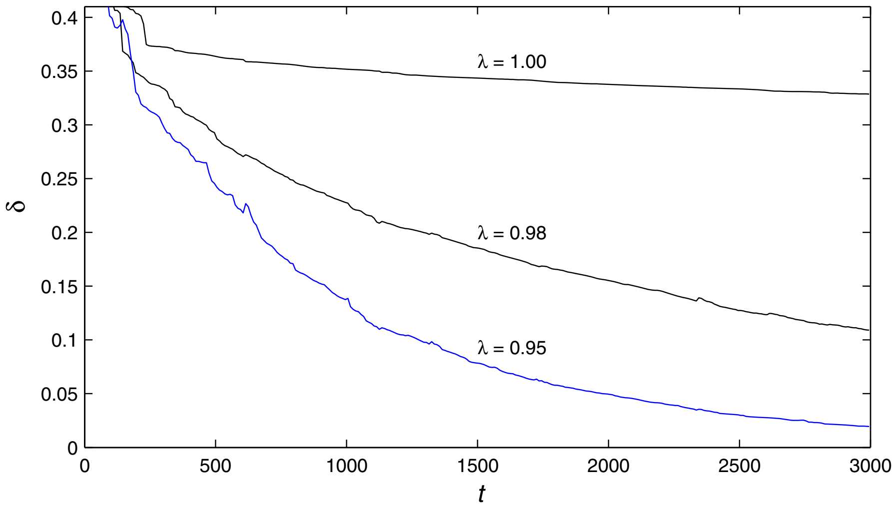

Fig. 1 HSG estimation errors δ versus t with different forgetting factors

2. The FF-HSG algorithm has faster convergence rates than the HSG algorithm, and the convergence rates increase for appropriate small forgetting factors—see Fig. 1.

5 Conclusions

This paper investigates the performances of the HSG algorithm for bilinear-in-parameter systems. The theoretical analysis shows that the estimates converge to the true values under the persistent excitation conditions, and the simulation results verify the proposed convergence theorem. The method used in this paper can be extended to analyze the convergence of the identification algorithms for linear or nonlinear control systems [7,19,20,43] and applied to hybrid switching-impulsive dynamical networks [18] and uncertain chaotic delayed nonlinear systems [17] or applied to other fields [6,27,28].

Acknowledgements This work was supported by the National Natural Science Foundation of China (Nos. 61164015, 60474039) and the Key Research Project of Henan Higher Education Institutions (No. 16A120010).

References

1. R. Abrahamssona, S.M. Kay, P. Stoica, Estimation of the parameters of a bilinear model with applications to submarine detection and system identification. Digit. Signal Process. **17**(4), 756–773 (2007)
2. A. Atitallah, S. Bedoui, K. Abderrahim, Identification of wiener time delay systems based on hierarchical gradient approach. in *The 8th Vienna International Conference on Mathematical Modelling—MATHMOD, IFAC-Papers OnLine* **48**(1), 403–408 (2015)
3. E.W. Bai, An optimal two-stage identification algorithm for Hammerstein–Wiener nonlinear systems. Automatica **34**(3), 333–338 (1998)
4. E.W. Bai, A blind approach to the Hammerstein–Wiener model identification. Automatica **38**(6), 967–979 (2002)
5. E.W. Bai, Y. Liu, Least squares solutions of bilinear equations. Syst. Control Lett. **55**(6), 466–472 (2006)
6. 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
7. 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
8. F. Ding, G.J. Liu, X.P. Liu, Parameter estimation with scarce measurements. Automatica **47**(8), 1646–1655 (2011)
9. 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)
10. 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)
11. 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. (2016). doi:10.1007/s00034-015-0190-6
12. 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)
13. F. Ding, Y. Gu, Performance analysis of the auxiliary model-based stochastic gradient parameter estimation algorithm for state-space systems with one-step state delay. Circuits Syst. Signal Process. **32**(2), 585–599 (2013)
14. M. Gilson, P. Van den Hof, Instrumental variable methods for closed-loop system identification. Automatica **41**(2), 241–249 (2005)
15. G.C. Goodwin, K.S. Sin, *Adaptive Filtering Prediction and Control* (Prentice-Hall, Englewood Cliffs, 1984)
16. A. Haryanto, K.S. Hong, Maximum likelihood identification of Wiener–Hammerstein models. Mech. Syst. Signal Process. **41**(1–2), 54–70 (2013)
17. Y. Ji, X.M. Liu, F. Ding, New criteria for the robust impulsive synchronization of uncertain chaotic delayed nonlinear systems. Nonlinear Dyn. **79**(1), 1–9 (2015)
18. Y. Ji, X.M. Liu, Unified synchronization criteria for hybrid switching-impulsive dynamical networks. Circuits Syst. Signal Process. **34**(5), 1499–1517 (2015)
19. 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
20. 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)
21. H. Li, Y. Shi, Robust H-infinity filtering for nonlinear stochastic systems with uncertainties and random delays modeled by Markov chains. Automatica **48**(1), 159–166 (2012)
22. L. Ljung, *System Identification: Theory for the User*, 2nd edn. (Prentice Hall, Englewood Cliffs, 1999)
23. J. Pan, X.H. Yang, H.F. Cai, B.X. Mu, Image noise smoothing using a modified Kalman filter. Neurocomputing **173**, 1625–1629 (2016)

24. J.D. Wang, Q.H. Zhang, L. Ljung, Revisiting Hammerstein system identification through the two-stage algorithm for bilinear parameter estimation. Automatica **45**(11), 2627–2633 (2009)
25. 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)
26. X.H. Wang, F. Ding, F.E. Alsaadi, T. Hayat, Convergence analysis of the hierarchical least squares algorithm for bilinear-in-parameter systems. Circuits Syst. Signal Process. (2016). doi:10.1007/s00034-016-0278-7
27. 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)
28. 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)
29. 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)
30. D.Q. Wang, F. Ding, Parameter estimation algorithms for multivariable Hammerstein CARMA systems. Inf. Sci. **355**, 237–248 (2016)
31. 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)
32. 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)
33. Y.J. Wang, F. Ding, The filtering based iterative identification for multivariable systems. IET Control Theory Appl. **10**(8), 894–902 (2016)
34. 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)
35. X.H. Wang, F. Ding, Modelling and multi-innovation parameter identification for Hammerstein nonlinear state space systems using the filtering technique. Math. Comput. Modell. Dyn. Syst. **22**(2), 113–140 (2016)
36. C. Wang, T. Tang, Recursive least squares estimation algorithm applied to a class of linear-in-parameters output error moving average systems. Appl. Math. Lett. **29**, 36–41 (2014)
37. 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)
38. D.Q. Wang, W. Zhang, Improved least squares identification algorithm for multivariable Hammerstein systems. J. Franklin Inst. **352**(11), 5292–5370 (2015)
39. 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)
40. A. Wills, T.B. Schön, L. Ljung et al., Identification of Hammerstein–Wiener models. Automatica **49**(1), 70–81 (2013)
41. W.L. Xiong, J.X. Ma, R.F. Ding, An iterative numerical algorithm for modeling a class of Wiener nonlinear systems. Appl. Math. Lett. **26**(4), 487–493 (2013)
42. X.P. Xu, F. Wang, G.J. Liu, Identification of Hammerstein systems using key-term separation principle, auxiliary model and improved particle swarm optimisation algorithm. IET Signal Process. **7**(8), 766–773 (2013)
43. L. Xu, A proportional differential control method for a time-delay system using the Taylor expansion approximation. Appl. Math. Comput. **236**, 391–399 (2014)
44. L. Xu, Application of the Newton iteration algorithm to the parameter estimation for dynamical systems. J. Comput. Appl. Math. **288**, 33–43 (2015)
45. 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)
46. L. Xu, The damping iterative parameter identification method for dynamical systems based on the sine signal measurement. Signal Process. **120**, 660–667 (2016)