Some Stochastic Gradient Algorithms for Hammerstein Systems with Piecewise Linearity

Yan Pu[1,2] · **Yongqing Yang**[1] · **Jing Chen**[1]

Abstract

Some stochastic gradient (SG) algorithms for Hammerstein systems with piecewise linearity are developed in this paper. Due to the complexity of the nonlinear structure, the key term separation is used to transfer the nonlinear model into a regression model, and then, some SG algorithms are proposed for this model. Since the SG algorithm has slow convergence rate, a forgetting factor SG algorithm and an Aitken SG algorithm are provided. Compared with the forgetting factor SG algorithm, the Aitken SG algorithm has smaller variance of estimation error, which means the Aitken SG algorithm is more effective. Two simulation examples are provided to show the effectiveness of the proposed algorithms.

1 Introduction

Parameter estimation plays an important role in controller design [13,14] because the controller design of dynamic system is usually established on the premise that the parameters of dynamic systems are known [19,28]. Compared with the linear systems, nonlinear systems are more extensive in engineering practice [6,10], and they can be roughly divided into four categories: Hammerstein systems [36], Wiener sys-

✉ Yongqing Yang
    yongqingyang@163.com

    Yan Pu
    puyan@jiangnan.edu.cn

    Jing Chen
    chenjing1981929@126.com

[1] School of Science, Jiangnan University, Wuxi 214122, People's Republic of China

[2] School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, People's Republic of China

tems [18], Hammerstein–Wiener systems and Wiener–Hammerstein systems [2,35]. Recently, many identification algorithms have been developed for these nonlinear systems, such as the stochastic gradient (SG) algorithms [39], the expectation maximization algorithms and the iterative algorithms [4]. The SG algorithm updates the parameter estimates only through the latest input–output data at each sampling instant, and it does not need to compute the inverse matrix; thus, it has less computational loads. Its variants include the multi-innovation stochastic gradient algorithms [16,42] and the gradient-based iterative algorithms [12].

The idea of the gradient-based identification algorithms is first to determine the search direction and then to calculate the step size for each sampling instant. Although the computational effort of the SG algorithm is small, its convergence rate is slow because of its zigzag search directions. In general, there are two methods to improve the convergence rates. One is to obtain the optimal direction at each sampling instant. For example, for control problems with undetermined final time, Hussu provided a conjugate-gradient method [20]. The other method is to compute a suitable step size at each sampling instant. For instance, Chen and Ding introduced a convergence index into a modified stochastic gradient(M-SG) algorithm to improve the convergence rate [7]. Ma et al. [30] studied a forgetting factor stochastic gradient (FF-SG) algorithm for Hammerstein systems with saturation and preload nonlinearities. Although the M-SG and FF-SG algorithms can improve the convergence rates, they also bring some issues, such as severe oscillation when the estimates of the parameters approach to the true values [24].

One may ask whether it is feasible to develop a modified SG algorithm, which can not only quickly estimate the parameters, but also decrease the variances of the estimation errors. For this sake, the Aitken method is introduced in this paper. The Aitken method is a sequence acceleration method, used for accelerating the convergence rate of sequences. It is efficient for accelerating the convergence rate of a sequence which is converging linearly. For example, Pavaloiu et al. [34] studied an Aitken–Newton iterative method for nonlinear equations, which is more competitive than some optimization methods with the same convergence order. Bumbariu [5] developed an improved Aitken acceleration method for solving nonlinear equations which computes the solutions of the nonlinear equations with fast convergence rates. The proposed approaches of this paper have some interesting features.

1. Using the key term separation method, which can transform a complex Hammerstein systems with piecewise linearity into a simplify regression model.
2. Studying an FF-SG algorithm for this nonlinear system, which can improve the convergence rate.
3. Developing an Aitken-based SG algorithm, which has quick convergence rates and small estimation error variances.
4. Extending the proposed methods to identify the Hammerstein systems with colored noise.

In summary, this paper is listed as follows. Section 2 introduces the Hammerstein model. Section 3 presents some SG algorithms. Section 4 studies the Aitken-based SG algorithm for the piecewise linear system with colored noise. In Sect. 5, two illustrative
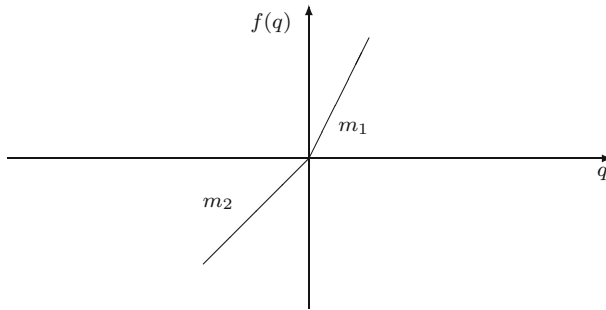
Fig. 1 The piecewise linearity

examples are provided. Section 6 gives the conclusions of this paper and the directions of future research.

2 The Hammerstein System with Piecewise Linearity

The piecewise linear system is a special kind of switching systems, which widely exists in engineering practice [27,37]. Such a system can be used to model or approximately describe the processes with different gains in different input intervals, e.g., the systems of flight control, circuits and biology [26,31].

Consider the Hammerstein system with piecewise linearity as follows:

$$A(\zeta)y(\tau) = B(\zeta)f(q(\tau)) + v(\tau), \tag{1}$$

where $q(\tau)$ is the input which is taken as a persistent excitation signal sequence with zero mean and unit variance, $y(\tau)$ is the output, $v(\tau)$ is a white noise with zero mean and variance σ^2, and a piecewise linearity $f(q(\tau))$ is shown in Fig. 1, which can be written as

$$f(q(\tau)) = \begin{cases} m_1 q(\tau), & q(\tau) \geq 0, \\ m_2 q(\tau)), & q(\tau) < 0, \end{cases} \tag{2}$$

where the corresponding segment slopes are m_1 and m_2.

The polynomials $A(\zeta)$ and $B(\zeta)$ are expressed as

$$\begin{aligned} A(\zeta) &= 1 + a_1\zeta^{-1} + a_2\zeta^{-2} + \cdots + a_n\zeta^{-n}, \\ B(\zeta) &= b_0 + b_1\zeta^{-1} + \cdots + b_{n-1}\zeta^{1-n}. \end{aligned}$$

Since the piecewise linearity is expressed by two equations, the Hammerstein system may be illustrated by two models [3]. Then, the considered Hammerstein model is equivalent to a switching model [1]. It is well known that the identification for switching models is more challenging. In order to simplify the identification process, the key term separation method is introduced [8,9].

Define a switching function,

$$s(\tau) := s[q(\tau)] = \begin{cases} 0, & q(\tau) > 0, \\ 1, & q(\tau) \le 0. \end{cases}$$

Then, the nonlinear part $f(q(\tau))$ of input is written as

$$f(q(\tau)) = m_1 s(-q(\tau))q(\tau) + m_2 s(q(\tau))q(\tau). \tag{3}$$

The nonlinear model can be written as

$$A(\zeta)y(\tau) = B(\zeta)m_1 s(-q(\tau))q(\tau) + B(\zeta)m_2 s(q(\tau))q(\tau) + v(\tau). \tag{4}$$

Define the information vector $\chi(\tau)$ and the parameter vector ξ as

$$\chi(\tau) = [-y(\tau-1), -y(\tau-2), \ldots, -y(\tau-n), s(-q(\tau))q(\tau), s(-q(\tau-1))q(\tau-1), \ldots,$$
$$s(-q(\tau-n+1))q(\tau-n+1), s(q(\tau))q(\tau), s(q(\tau-1))q(\tau-1), \ldots,$$
$$s(q(\tau-n+1))q(\tau-n+1)]^{\mathrm{T}} \in \mathbb{R}^{3n}, \tag{5}$$
$$\xi = [a_1, a_2, \ldots, a_n, b_0 m_1, b_1 m_1, \ldots, b_{n-1} m_1, b_0 m_2, b_1 m_2, \ldots, b_{n-1} m_2]^{\mathrm{T}} \in \mathbb{R}^{3n}. \tag{6}$$

Then, the nonlinear model can be simplified as a regression model:

$$y(\tau) = \chi^{\mathrm{T}}(\tau)\xi + v(\tau). \tag{7}$$

The proposed algorithms in this paper are based on this identification model. Many identification methods are derived based on the identification models of dynamical systems [29,32,33], which can be used to estimate the parameters of bilinear systems [23,38,47,48], and can be applied to fields such as chemical process control systems. From Eq. (7), it can be seen that the parameters can be estimated by all the traditional identification algorithms in the cost of heavy computational demands [17].

Remark 1　In this paper, b_0 is assumed to be equal to 1; otherwise, b_i cannot be separated from $b_i m_k$. Assume the parameter estimates are

$$\hat{\xi} = [\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_n, \hat{m}_1, \hat{b}_1 \hat{m}_1, \ldots, \hat{b}_{n-1} \hat{m}_1, \hat{m}_2, \hat{b}_1 \hat{m}_2, \ldots, \hat{b}_{n-1} \hat{m}_2]^{\mathrm{T}}$$

Once the parameter estimates have been obtained, we can get $\hat{m}_1$ and $\hat{m}_2$ first, and then, based on $\hat{m}_1$ and $\hat{m}_2$, we can get $\hat{b}_i = \frac{\hat{b}_i \hat{m}_k}{\hat{m}_k}, i = 1, \ldots, n-1, k = 1, 2$.

3 Some Stochastic Gradient Algorithms

The SG algorithm can be realized online, which updates parameters according to the latest input–output data [11]. Therefore, it has less computational efforts. However, this algorithm has slow convergence rates. In this section, some modified SG algorithms will be investigated.

3.1 The Traditional Stochastic Gradient Algorithm

Define the cost function

$$J(\boldsymbol{\xi}) = [y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)\boldsymbol{\xi}]^2.$$

Assume that the parameter estimates at time τ are $\hat{\boldsymbol{\xi}}(\tau-1)$, the key of the SG algorithm is to get a better estimate $\hat{\boldsymbol{\xi}}(\tau)$ which satisfies

$$J(\hat{\boldsymbol{\xi}}(\tau)) = [y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\xi}}(\tau)]^2 \leq J(\hat{\boldsymbol{\xi}}(\tau-1)) = [y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\xi}}(\tau-1)]^2. \quad (8)$$

$\hat{\boldsymbol{\xi}}(\tau)$ is obtained based on $\hat{\boldsymbol{\xi}}(\tau-1)$ and is written by

$$\hat{\boldsymbol{\xi}}(\tau) = \hat{\boldsymbol{\xi}}(\tau-1) + \lambda(\tau)\boldsymbol{\chi}(\tau)[y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(t)\hat{\boldsymbol{\xi}}(\tau-1)]. \quad (9)$$

In order to keep (8) holding, substituting (9) into (8) gets

$$J(\hat{\boldsymbol{\xi}}(\tau)) = \left\{ y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)(\hat{\boldsymbol{\xi}}(\tau-1) + \lambda(\tau)\boldsymbol{\chi}(\tau)[y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\xi}}(\tau-1)]) \right\}^2.$$

Keep $\hat{\boldsymbol{\xi}}(\tau-1)$ fixing and define $J(\lambda(\tau))$ as

$$J(\lambda(\tau)) = \left\{ y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)(\hat{\boldsymbol{\xi}}(\tau-1) + \lambda(\tau)\boldsymbol{\chi}(\tau)[y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\xi}}(\tau-1)]) \right\}^2.$$

Let

$$\frac{\partial J(\lambda(\tau))}{\partial \lambda(\tau)} = \frac{\partial \left[y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)(\hat{\boldsymbol{\xi}}(\tau-1) + \lambda(\tau)\boldsymbol{\chi}(\tau)[y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\xi}}(\tau-1)]) \right]^2}{\partial \lambda(\tau)}.$$

Setting the above derivative equal to zero obtains

$$\lambda(\tau) = \frac{1}{\boldsymbol{\chi}^{\mathrm{T}}(\tau)\boldsymbol{\chi}(\tau)}.$$

Then, we can get the steepest descent algorithm

$$\hat{\boldsymbol{\xi}}(\tau) = \hat{\boldsymbol{\xi}}(\tau-1) + \frac{\hat{\boldsymbol{\chi}}(\tau)}{\boldsymbol{\chi}^{\mathrm{T}}(\tau)\boldsymbol{\chi}(\tau)}\left(y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\xi}}(\tau-1) \right). \quad (10)$$

However, when $\boldsymbol{\chi}^{\mathrm{T}}(\tau)\boldsymbol{\chi}(\tau)$ is small, the correction items $\frac{\hat{\boldsymbol{\chi}}(\tau)}{\boldsymbol{\chi}^{\mathrm{T}}(\tau)\boldsymbol{\chi}(\tau)}\left(y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\xi}}(\tau-1) \right)$ would be large, which leads to the steepest descent algorithm be divergent. With this in mind, we define

$$\lambda(\tau) = \rho + \boldsymbol{\chi}^{\mathrm{T}}(\tau)\boldsymbol{\chi}(\tau), \rho \geq 0$$

Then, we get the projection algorithm

$$\hat{\boldsymbol{\xi}}(\tau) = \hat{\boldsymbol{\xi}}(\tau - 1) + \frac{\hat{\boldsymbol{\chi}}(\tau)}{\rho + \boldsymbol{\chi}^{\mathrm{T}}(\tau)\boldsymbol{\chi}(\tau)}\big(y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\xi}}(\tau - 1)\big). \tag{11}$$

Since ρ is a constant, the unchanged step size will make the estimate of the algorithm oscillate seriously when the estimates are closing to the true values. In order to solve this problem, we replace ρ by $\lambda(\tau - 1)$. Then, the SG algorithm to estimate the parameter $\boldsymbol{\xi}$ is listed as follows,

$$\hat{\boldsymbol{\xi}}(\tau) = \hat{\boldsymbol{\xi}}(\tau - 1) + \frac{\hat{\boldsymbol{\chi}}(\tau)}{\lambda(\tau)}\big(y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\xi}}(\tau - 1)\big), \tag{12}$$

$$\boldsymbol{\chi}(\tau) = [-y(\tau - 1), -y(\tau - 2), \dots, -y(\tau - n), s(-q(\tau))q(\tau), s(-q(\tau - 1))q(\tau - 1), \dots,$$
$$s(-q(\tau - n + 1))q(\tau - n + 1), s(q(\tau))q(\tau), s(q(\tau - 1))q(\tau - 1), \dots,$$
$$s(q(\tau - n + 1))q(\tau - n + 1)]^{\mathrm{T}}. \tag{13}$$

$$\lambda(\tau) = \lambda(\tau - 1) + \|\hat{\boldsymbol{\chi}}(\tau)\|^2, \lambda(0) = 1. \tag{14}$$

Remark 2 Although the traditional SG algorithm has less computational efforts, it also brings some challenging issues, e.g., slow convergence rates, especially for systems with large number of unknown parameters.

3.2 Two Modified Stochastic Gradient Algorithms

In order to increase the convergence rates, two modified SG algorithms for the Hammerstein system are developed in this subsection. A forgetting factor SG (FF-SG) algorithm is first introduced,

$$\hat{\boldsymbol{\xi}}(\tau) = \hat{\boldsymbol{\xi}}(\tau - 1) + \frac{\hat{\boldsymbol{\chi}}(\tau)}{\lambda(\tau)}\big(y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\xi}}(\tau - 1)\big), \tag{15}$$

$$\boldsymbol{\chi}(\tau) = [-y(\tau - 1), -y(\tau - 2), \dots, -y(\tau - n), s(-q(\tau))q(\tau), s(-q(\tau - 1))q(\tau - 1), \dots,$$
$$s(-q(\tau - n + 1))q(\tau - n + 1), s(q(\tau))q(\tau), s(q(\tau - 1))q(\tau - 1), \dots,$$
$$s(q(\tau - n + 1))q(\tau - n + 1)]^{\mathrm{T}}. \tag{16}$$

$$\lambda(\tau) = r\lambda(\tau - 1) + \|\hat{\boldsymbol{\chi}}(\tau)\|^2, \ \lambda(0) = 1, \ 0.8 \leq r \leq 1. \tag{17}$$

Remark 3 The FF-SG algorithm introduces a forgetting factor r in the step size [15, 43,46], which will make the step size larger at each sampling instant. Therefore, the FF-SG algorithm has quicker convergence rates compared with the traditional SG algorithm.

Remark 4 Although the FF-SG algorithm can increase the convergence rates, it brings some challengings, such as large estimation error variances.

To make the variance of the estimation error smaller, another modified SG algorithm will be studied in the following, which is termed as the Aitken-based SG (A-SG) algorithm. Assume that the parameter estimate $\hat{\boldsymbol{\xi}}(\tau)$ converges to the true value $\boldsymbol{\xi}$,

which means that

$$\lim_{\tau \to \infty} [\hat{\boldsymbol{\xi}}(\tau) - \boldsymbol{\xi}] = \lim_{\tau \to \infty} [\hat{\boldsymbol{\xi}}(\tau - 1) - \boldsymbol{\xi}]. \tag{18}$$

It is equivalent to

$$\lim_{\tau \to \infty} \frac{\hat{\xi}_\varrho(\tau) - \xi_\varrho}{\hat{\xi}_\varrho(\tau - 1) - \xi_\varrho} = 1, \tag{19}$$

where ξ_ϱ is the ϱth element in the parameter vector $\boldsymbol{\xi}$, $\varrho = 1, 2, \ldots, 3n$. When τ is large enough, the equivalent expression of (19) can be written as

$$\frac{\hat{\xi}_\varrho(\tau) - \xi_\varrho}{\hat{\xi}_\varrho(\tau - 1) - \xi_\varrho} = \frac{\hat{\xi}_\varrho(\tau - 1) - \xi_\varrho}{\hat{\xi}_\varrho(\tau - 2) - \xi_\varrho}. \tag{20}$$

From (19) and (20), it follows that

$$[\hat{\xi}_\varrho(\tau) + \hat{\xi}_\varrho(\tau - 2) - 2\hat{\xi}_\varrho(\tau - 1)]\xi_\varrho = \hat{\xi}_\varrho(\tau)\hat{\xi}_\varrho(\tau - 2) - \hat{\xi}_\varrho^2(\tau - 1). \tag{21}$$

Then, the Aitken accelerated iteration formula for ξ_ϱ can be written as

$$\xi_\varrho = \hat{\xi}_\varrho(\tau - 2) - \frac{\left(\hat{\xi}_\varrho(\tau - 1) - \hat{\xi}_\varrho(\tau - 2)\right)^2}{\hat{\xi}_\varrho(\tau) + \hat{\xi}_\varrho(\tau - 2) - 2\hat{\xi}_\varrho(\tau - 1)}. \tag{22}$$

However, the parameter $\boldsymbol{\xi}$ cannot be computed by Eq. (18) because it is not a scalar, but a vector. In order to get the vector, the parameter $\boldsymbol{\xi}$ is rewritten as

$$\boldsymbol{\xi} = [a_1, a_2, \ldots, a_n, m_1, b_1 m_1, \ldots, b_{n-1} m_1, m_2, b_1 m_2, \ldots, b_{n-1} m_2]^{\mathrm{T}}$$

Then, Eq. (21) is equivalent to the $3n$ equations as follows,

$$[\hat{a}_i(\tau) + \hat{a}_i(\tau - 2) - 2\hat{a}_i(\tau - 1)]a_i = \hat{a}_i(\tau)\hat{a}_i(\tau - 2) - \hat{a}_i^2(\tau - 1), \quad i = 1, \ldots, n,$$

$$[\hat{m}_k(\tau) + \hat{m}_k(\tau - 2) - \hat{m}_k(\tau - 1)]m_k = \hat{m}_k(\tau)\hat{m}_k(\tau - 2) - \hat{m}_k^2(\tau - 1), \quad k = 1, 2,$$

$$[\hat{b}_j(\tau)\hat{m}_k(\tau) + \hat{b}_j(\tau - 2)\hat{m}_k(\tau - 2) - 2\hat{b}_j(\tau - 1)\hat{m}_k(\tau - 1)]b_j m_k$$
$$= \hat{b}_j(\tau)\hat{m}_k(\tau)\hat{b}_j(\tau - 2)\hat{m}_k(\tau - 2) - \hat{b}_j^2(\tau - 1)\hat{m}_k^2(\tau - 1), \quad j = 1, \ldots, n - 1.$$

Then, we have

$$a_i = \hat{a}_i(\tau - 2) - \frac{(\hat{a}_i(\tau - 1) - \hat{a}_i(\tau - 2))^2}{\hat{a}_i(\tau) + \hat{a}_i(\tau - 2) - 2\hat{a}_i(\tau - 1)},$$

$$m_k = \hat{m}_k(\tau - 2) - \frac{(\hat{m}_k(\tau - 1) - \hat{m}_k(\tau - 2))^2}{\hat{m}_k(\tau) + \hat{m}_k(\tau - 2) - 2\hat{m}_k(\tau - 1)},$$

$$b_j m_k = \hat{b}_j(\tau - 2)\hat{m}_k(\tau - 2)$$

$$-\frac{(\hat{b}_j(\tau-1)\hat{m}_k(\tau-1) - \hat{b}_j(\tau-2)\hat{m}_k(\tau-2))^2}{\hat{b}_j(\tau)\hat{m}_k(\tau) + \hat{b}_j(\tau-2)\hat{m}_k(\tau-2) - 2\hat{b}_j(\tau-1)\hat{m}_k(\tau-1)}.$$

Define

$$\bar{a}_i(\tau) = \hat{a}_i(\tau-2) - \frac{(\hat{a}_i(\tau-1) - \hat{a}_i(\tau-2))^2}{\hat{a}_i(\tau) + \hat{a}_i(\tau-2) - 2\hat{a}_i(\tau-1)},$$

$$\bar{m}_k(\tau) = \hat{m}_k(\tau-2) - \frac{(\hat{m}_k(\tau-1) - \hat{m}_k(\tau-2))^2}{\hat{m}_k(\tau) + \hat{m}_k(\tau-2) - 2\hat{m}_k(\tau-1)},$$

$$\bar{b}_j(\tau)\bar{m}_k(\tau) = \hat{b}_j(\tau-2)\hat{m}_k(\tau-2)$$
$$-\frac{(\hat{b}_j(\tau-1)\hat{m}_k(\tau-1) - \hat{b}_j(\tau-2)\hat{m}_k(\tau-2))^2}{\hat{b}_j(\tau)\hat{m}_k(\tau) + \hat{b}_j(\tau-2)\hat{m}_k(\tau-2) - 2\hat{b}_j(\tau-1)\hat{m}_k(\tau-1)}.$$

The Aitken-based SG algorithm is obtained as follows,

$$\bar{a}_i(\tau) = \hat{a}_i(\tau-2) - \frac{(\hat{a}_i(\tau-1) - \hat{a}_i(\tau-2))^2}{\hat{a}_i(\tau) + \hat{a}_i(\tau-2) - 2\hat{a}_i(\tau-1)}, \tag{23}$$

$$\bar{m}_k(\tau) = \hat{m}_k(\tau-2) - \frac{(\hat{m}_k(\tau-1) - \hat{m}_k(\tau-2))^2}{\hat{m}_k(\tau) + \hat{m}_k(\tau-2) - 2\hat{m}_k(\tau-1)}, \tag{24}$$

$$\bar{b}_j(\tau)\bar{m}_k(\tau) = \hat{b}_j(\tau-2)\hat{m}_k(\tau-2)$$
$$-\frac{(\hat{b}_j(\tau-1)\hat{m}_k(\tau-1) - \hat{b}_j(\tau-2)\hat{m}_k(\tau-2))^2}{\hat{b}_j(\tau)\hat{m}_k(\tau) + \hat{b}_j(\tau-2)\hat{m}_k(\tau-2) - 2\hat{b}_j(\tau-1)\hat{m}_k(\tau-1)}, \tag{25}$$

$$\hat{\boldsymbol{\xi}}(\tau) = \hat{\boldsymbol{\xi}}(\tau-1) + \frac{\boldsymbol{\chi}(\tau)}{\lambda(\tau)}e(\tau), \tag{26}$$

$$e(\tau) = y(\tau) - \boldsymbol{\chi}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\xi}}(\tau-1), \tag{27}$$

$$\boldsymbol{\chi}(\tau) = [-y(\tau-1), -y(\tau-2), \ldots, -y(\tau-n), s(-q(\tau))q(\tau),$$
$$s(-q(\tau-1))q(\tau-1), \ldots,$$
$$s(-q(\tau-n+1))q(\tau-n+1), s(q(\tau))q(\tau), s(q(\tau-1))q(\tau-1), \ldots,$$
$$s(q(\tau-n+1))q(\tau-n+1)]^{\mathrm{T}}, \tag{28}$$

$$\lambda(\tau) = \lambda(\tau-1) + \boldsymbol{\chi}^{\mathrm{T}}(\tau)\boldsymbol{\chi}(\tau). \tag{29}$$

The A-SG algorithm starts the iterations as follows.

1. To initialize: Let $\tau = 1$, $\hat{\boldsymbol{\xi}}(0) = \mathbf{1}_{3n}/p_0$, $p_0 = 10^6$ and $\lambda(0) = 1$.
2. Let $y(\tau) = 0$, $q(\tau) = 0$, $\tau \leq 0$, and give an error tolerance number ε.
3. Collect the input–output data $\{q(\tau), y(\tau)\}$.
4. Form $\boldsymbol{\chi}(\tau)$ by (28).
5. Compute $e(\tau)$ and $\lambda(\tau)$ by (27) and (29), respectively.
6. Update the estimation vector $\hat{\boldsymbol{\xi}}(\tau)$ by (26).
7. Compute each estimate $\bar{a}_i(\tau)$, $i = 1, \ldots, n$, $\bar{m}_k(t)$, $k = 1, 2$ and $\bar{b}_j(\tau)\bar{m}_k(\tau)$, $j = 1, \ldots, n-1$ by (23)–(25), and then form $\bar{\boldsymbol{\xi}}(\tau)$.
8. Compare $\bar{\boldsymbol{\xi}}(\tau)$ and $\bar{\boldsymbol{\xi}}(\tau-1)$: if $\|\bar{\boldsymbol{\xi}}(\tau) - \bar{\boldsymbol{\xi}}(\tau-1)\| \leq \varepsilon$, then obtain the $\bar{\boldsymbol{\xi}}(\tau)$ and go to the next step; otherwise, increase τ by 1 and go to step 3.

9. Compute $\bar{m}_k(\tau)$ first, and then calculate $\bar{b}_i(\tau) = \frac{\bar{b}_i(\tau)\bar{m}_k(\tau)}{\bar{m}_k(\tau)}$.

Remark 5 The A-SG algorithm utilizes three connected parameter estimates to obtain an optimal parameter estimate, which does not use a large step size to speed up the convergence. Therefore, the A-SG algorithm has quicker convergence rates but smaller estimation error variances.

4 The Identification for the Hammerstein Piecewise Linearity System with Colored Noise

In this part, the SG algorithms are developed to identify the Hammerstein system with colored noise, which contains unmeasurable noise variables in the information vector.

4.1 Problem Description and Identification Model

Consider the Hammerstein piecewise linearity system with colored noise as follows,

$$y(\tau) = \frac{B(\zeta)}{A(\zeta)}\bar{x}(\tau) + \frac{D(\zeta)}{A(\zeta)}v(\tau), \tag{30}$$

where $D(\zeta) := 1 + d_1\zeta^{-1} + d_2\zeta^{-2} + \cdots + d_n\zeta^{-n_d}$, the definitions of $A(\zeta)$, $B(\zeta)$ and the piecewise linearity part are the same as those in Section 2.

By utilizing the key term separation technique, the system can be transformed into

$$A(\zeta)y(\tau) = B(\zeta)m_1s(-q(\tau))q(\tau) + B(\zeta)m_2s(q(\tau))u(\tau) + D(\zeta)v(\tau). \tag{31}$$

Then, the system is written by

$$y(\tau) = -\sum_{i=1}^{n}a_iy(\tau-i) + m_1s(-q(\tau))q(\tau) + \sum_{i=1}^{n-1}m_1b_is(-q(\tau-i))q(\tau-i)$$

$$+m_2s(q(\tau))q(\tau) + \sum_{i=1}^{n-1}m_2b_is(q(\tau-i))q(\tau-i) - \sum_{i=1}^{n_d}d_iv(\tau-i) + v(\tau). \tag{32}$$

Define the information vector $\boldsymbol{\psi}(\tau)$ and the parameter vector $\boldsymbol{\vartheta}$ as

$$\boldsymbol{\psi}(\tau) := [-y(\tau-1), -y(\tau-2), \ldots, -y(\tau-n), s(-q(\tau))q(\tau),$$
$$s(-q(\tau-1))q(\tau-1), \ldots, s(-q(\tau-n+1))q(\tau-n+1), s(q(\tau))u(\tau),$$
$$s(q(\tau-1))q(\tau-1), \ldots, s(q(\tau-n+1))q(\tau-n+1),$$
$$v(\tau-1), v(\tau-2), \ldots, v(\tau-n_d),]^{\mathrm{T}} \in \mathbb{R}^{3n+n_d}, \tag{33}$$
$$\boldsymbol{\vartheta} := [a_1, a_2, \ldots, a_n, m_1, b_1m_1, \ldots, b_{n-1}m_1, m_2, b_1m_2, \ldots, b_{n-1}m_2,$$
$$d_1, d_2, \ldots, d_{n_d}]^{\mathrm{T}} \in \mathbb{R}^{3n+n_d}. \tag{34}$$

Then, the nonlinear system can be expressed as a simple form,

$$y(\tau) = \boldsymbol{\psi}^{\mathrm{T}}(\tau)\boldsymbol{\vartheta} + v(\tau). \tag{35}$$

4.2 The Aitken Stochastic Gradient Algorithm

Since the information vector in the Hammerstein piecewise linearity system with colored noise contains the unmeasured noise variables $v(\tau - i)$, we denote $\hat{v}(\tau)$ and $\hat{\boldsymbol{\psi}}(\tau)$ as the estimates of the $v(\tau)$ and $\boldsymbol{\psi}(\tau)$ at time τ, respectively. Let $\hat{\boldsymbol{\vartheta}}(\tau)$ be the estimate of $\boldsymbol{\vartheta}$ at time τ and define the innovation $e(\tau)$ at time τ as follows,

$$e(\tau) := y(\tau) - \hat{\boldsymbol{\psi}}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\vartheta}}(\tau - 1), \tag{36}$$

where

$$\begin{aligned}
\hat{\boldsymbol{\psi}}(\tau) = [&-\hat{y}(\tau - 1), -\hat{y}(\tau - 2), \ldots, -\hat{y}(\tau - n), \hat{s}(-q(\tau))\hat{q}(\tau), \\
&\hat{s}(-q(\tau - 1))\hat{q}(\tau - 1), \ldots, \hat{s}(-q(\tau - n + 1))\hat{q}(\tau - n + 1), \hat{s}(q(\tau))\hat{q}(\tau), \\
&\hat{s}(q(\tau - 1))\hat{q}(\tau - 1), \ldots, \hat{s}(q(\tau - n + 1))\hat{q}(\tau - n + 1), \\
&e(\tau - 1), e(\tau - 2), \ldots, e(\tau - n_d)]^{\mathrm{T}} \in \mathbb{R}^{3n + n_d}.
\end{aligned} \tag{37}$$

Remark 6 Since the information vector $\boldsymbol{\psi}(\tau)$ contains the unmeasurable variables $v(\tau - i)$, their estimates $e(\tau - i)$ can be used to replace these unknown noise variables $v(\tau - i)$ in the information vector.

By using the Aitken accelerated iteration technique, the Aitken SG (A-SG) algorithm for the Hammerstein system with colored noise is developed as follows,

$$\bar{a}_i(\tau) = \hat{a}_i(\tau - 2) - \frac{(\hat{a}_i(\tau - 1) - \hat{a}_i(\tau - 2))^2}{\hat{a}_i(\tau) + \hat{a}_i(\tau - 2) - 2\hat{a}_i(\tau - 1)}, \tag{38}$$

$$\bar{m}_k(\tau) = \hat{m}_k(\tau - 2) - \frac{(\hat{m}_k(\tau - 1) - \hat{m}_k(\tau - 2))^2}{\hat{m}_k(\tau) + \hat{m}_k(\tau - 2) - 2\hat{m}_k(\tau - 1)}, \tag{39}$$

$$\begin{aligned}
\bar{b}_j(\tau)\bar{m}_k(\tau) = &\hat{b}_j(\tau - 2)\hat{m}_k(\tau - 2) \\
&- \frac{(\hat{b}_j(\tau - 1)\hat{m}_k(\tau - 1) - \hat{b}_j(\tau - 2)\hat{m}_k(\tau - 2))^2}{\hat{b}_j(\tau)\hat{m}_k(\tau) + \hat{b}_j(\tau - 2)\hat{m}_k(\tau - 2) - 2\hat{b}_j(\tau - 1)\hat{m}_k(\tau - 1)}, \tag{40}
\end{aligned}$$

$$\bar{d}_i(\tau) = \hat{d}_i(\tau - 2) - \frac{(\hat{d}_i(\tau - 1) - \hat{d}_i(\tau - 2))^2}{\hat{d}_i(\tau) + \hat{d}_i(\tau - 2) - 2\hat{d}_i(\tau - 1)}, \tag{41}$$

$$\begin{aligned}
\bar{\boldsymbol{\vartheta}}(\tau) = [&\bar{a}_1(\tau), \bar{a}_2(\tau), \ldots, \bar{a}_n(\tau), \bar{m}_1(\tau), \bar{b}_1(\tau)\bar{m}_1(\tau), \ldots, \bar{b}_{n-1}(\tau)\bar{m}_1(\tau), \\
&\bar{m}_2(\tau), \bar{b}_1(\tau)\bar{m}_2(\tau), \ldots, \bar{b}_{n-1}(\tau)\bar{m}_2(\tau)]^{\mathrm{T}}, \tag{42}
\end{aligned}$$

$$\hat{\boldsymbol{\vartheta}}(\tau) = \hat{\boldsymbol{\vartheta}}(\tau - 1) + \frac{\hat{\boldsymbol{\psi}}(\tau)}{\lambda(\tau)}e(\tau), \tag{43}$$

$$e(\tau) = y(\tau) - \hat{\boldsymbol{\psi}}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\vartheta}}(\tau - 1), \tag{44}$$

$$\hat{\boldsymbol{\psi}}(\tau) = [-\hat{y}(\tau - 1), -\hat{y}(\tau - 2), \dots, -\hat{y}(\tau - n), \hat{s}(-q(\tau))\hat{q}(\tau),$$
$$\hat{s}(-q(\tau - 1))\hat{q}(\tau - 1), \dots, \hat{s}(-q(\tau - n + 1))\hat{q}(\tau - n + 1),$$
$$\hat{s}(q(\tau))\hat{q}(\tau), \hat{s}(q(\tau - 1))\hat{q}(\tau - 1), \dots,$$
$$\hat{s}(q(\tau - n + 1))\hat{q}(\tau - n + 1), e(\tau - 1), e(\tau - 2), \dots, e(\tau - n_d)]^{\mathrm{T}}, \tag{45}$$

$$\lambda(\tau) = \lambda(\tau - 1) + \hat{\boldsymbol{\psi}}^{\mathrm{T}}(\tau)\hat{\boldsymbol{\psi}}(\tau). \tag{46}$$

The flowchart of the A-SG algorithm is presented in Fig. 2. The proposed methods in this paper can combine other identification methods [40,41] to study the parameter estimation problems of different systems with colored noises such as nonlinear systems [21,22] and can be applied to other studies such as signal modeling and communication networked systems.

5 Numerical Examples

Example 1 Consider the following Hammerstein model,

$$A(\zeta)y(\tau) = B(\zeta)f(q(\tau)) + v(\tau),$$
$$y(\tau) = -a_1 y(\tau - 1) - a_2 y(\tau - 2) + f(q(\tau)) + b_1 f(q(\tau - 1)) + v(\tau)$$
$$= -0.15 y(\tau - 1) - 0.46 y(\tau - 2) + f(q(\tau)) + 0.9 f(q(\tau - 1)) + v(\tau),$$
$$f(q(\tau)) = \begin{cases} 0.3q(\tau), \ 0 \le q(\tau), \\ 0.2q(\tau), \ q(\tau) \le 0, \end{cases}$$
$$\boldsymbol{\xi} = [a_1, a_2, m_1, b_1 m_1, m_2, b_1 m_2]^{\mathrm{T}} = [0.15, 0.46, 0.3, 0.27, 0.2, 0.18]^{\mathrm{T}},$$
$$\boldsymbol{\chi}(\tau) = [-y(\tau - 1), -y(\tau - 2), s(-q(\tau))q(\tau), s(-q(\tau - 1))q(\tau - 1), s(q(\tau))q(\tau),$$
$$s(q(\tau - 1))q(\tau - 1)]^{\mathrm{T}},$$

where $\{v(\tau)\}$ is taken as a white noise sequence with zero mean and variance $\sigma^2 = 0.10^2$, and $\{q(\tau)\}$ is an input sequence with zero mean and unit variance.

The SG, the FF-SG and the A-SG algorithms are applied to estimate the parameters of the piecewise linear system. The estimation errors $\delta := \|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|/\|\boldsymbol{\xi}\|$ or $\delta := \|\bar{\boldsymbol{\xi}} - \boldsymbol{\xi}\|/\|\boldsymbol{\xi}\|$ versus τ are shown in Fig. 3 and Tables 1, 2, 3. The means and variances of these three algorithms are given in Table 4.

Example 2 Consider the following Hammerstein model with colored noise,

$$y(\tau) = \frac{B(\zeta)}{A(\zeta)}\bar{x}(\tau) + \frac{D(\zeta)}{A(\zeta)}v(\tau),$$
$$y(\tau) = -a_1 y(\tau - 1) - a_2 y(\tau - 2) + f(q(\tau))$$
$$+ b_1 f(q(\tau - 1)) + v(\tau) + d_1 v(\tau - 1)$$
$$= -0.21 y(\tau - 1) - 0.10 y(\tau - 2) + f(q(\tau))$$
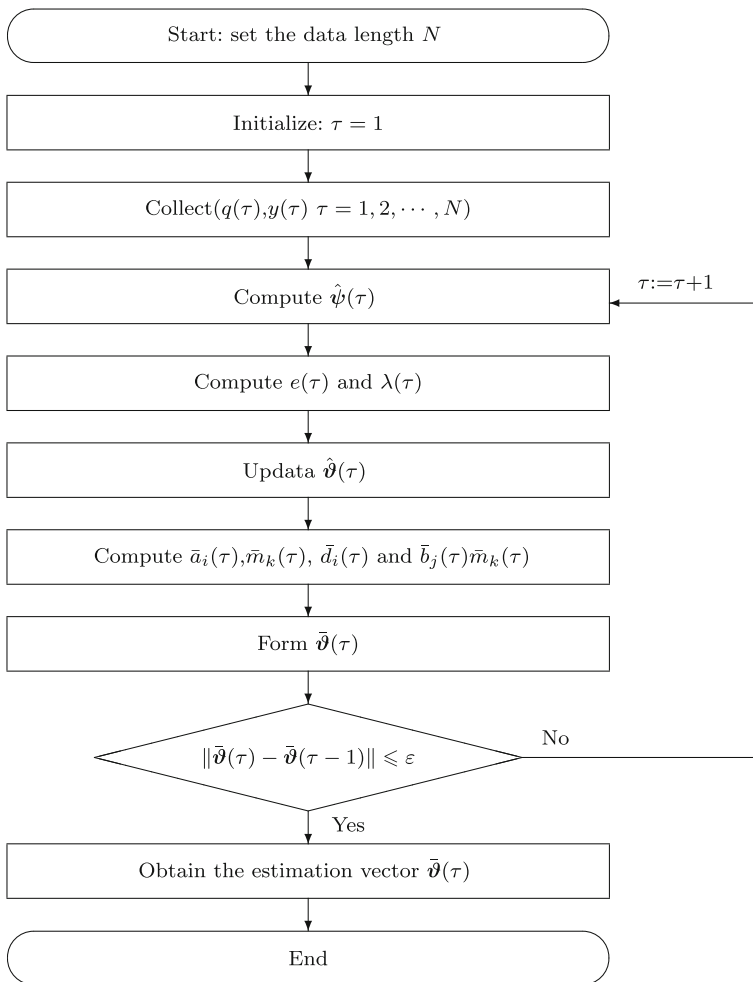$$+ 0.5 f(q(\tau - 1)) + v(\tau) - 0.38 v(\tau - 1),$$

Fig. 2 The flowchart of the A-SG algorithm for computing $\bar{\bar{\vartheta}}(\tau)$

$$f(q(\tau)) = \begin{cases} 2.0q(\tau), & 0 \le q(\tau), \\ 1.4q(\tau), & q(\tau) \le 0, \end{cases}$$

$$\vartheta = [a_1, a_2, m_1, b_1m_1, m_2, b_1m_2, d_1]^{\mathrm{T}} = [0.21, 0.1, 2.1, 1.0, 1.4, 0.7, -0.38]^{\mathrm{T}},$$

$$\psi(\tau) = [-y(\tau - 1), -y(\tau - 2), s(-q(\tau))q(\tau), s(-q(\tau - 1))q(\tau - 1),$$
$$s(q(\tau))q(\tau), s(q(\tau - 1))q(\tau - 1), v(\tau - 1)]^{\mathrm{T}},$$

where $\{v(\tau)\}$ is taken as a white noise sequence with zero mean and variance $\sigma^2 = 0.10^2$, and $\{q(\tau)\}$ is an input sequence with zero mean and unit variance.

The SG, the FF-SG and the A-SG algorithms are applied to estimate the parameters of the piecewise linear system with colored noise, and the estimation errors $\delta :=\|\hat{\vartheta} - \vartheta\|/\|\vartheta\|$ or $\delta := \|\bar{\vartheta} - \vartheta\|/\|\vartheta\|$ versus τ are shown in Fig. 4.

Table 1 The SG algorithm estimates and errors

τ	a_1	a_2	m_1	b_0m_1	m_2	b_1m_2	δ (%)
200	0.04083	0.16651	0.26437	0.17274	0.00366	0.00348	61.66538
300	0.04830	0.20358	0.26153	0.18991	0.00494	0.00702	56.89736
1000	0.10378	0.41049	0.29093	0.24533	0.01641	0.01982	37.10998
2000	0.12327	0.46443	0.30027	0.26422	0.02391	0.02621	34.36189
3000	0.13360	0.47360	0.30331	0.26445	0.02896	0.03104	33.26419
4000	0.13824	0.46949	0.30147	0.26559	0.03473	0.03676	32.00474
True values	0.15000	0.46000	0.30000	0.27000	0.20000	0.18000	

Table 2 The FF-SG algorithm estimates and errors

τ	a_1	a_2	m_1	b_0m_1	m_2	b_1m_2	δ (%)
200	0.13891	0.46556	0.29009	0.26591	0.02386	0.02229	34.59102
300	0.15193	0.48521	0.30785	0.27068	0.03740	0.03271	32.25484
1000	0.13960	0.47354	0.30659	0.27659	0.11953	0.12364	14.61765
2000	0.13406	0.47258	0.29529	0.26013	0.16786	0.18061	5.77445
3000	0.15447	0.46863	0.30856	0.26935	0.17421	0.19053	4.48426
4000	0.14725	0.45795	0.30604	0.26156	0.19427	0.20173	3.64792
True values	0.15000	0.46000	0.30000	0.27000	0.20000	0.18000	

Table 3 The A-SG algorithm estimates and errors

τ	a_1	a_2	m_1	b_0m_1	m_2	b_1m_2	δ (%)
200	0.11453	0.40492	0.28738	0.24380	0.01903	0.01315	37.41977
300	0.12956	0.44890	0.30304	0.25569	0.02803	0.02165	34.35594
1000	0.13892	0.48233	0.30700	0.26897	0.09372	0.07633	21.99768
2000	0.13399	0.47318	0.29746	0.26644	0.14094	0.11049	13.66729
3000	0.14170	0.47315	0.30640	0.26826	0.16830	0.13107	8.85996
4000	0.14612	0.46246	0.30048	0.26475	0.20241	0.15626	3.62995
True values	0.15000	0.46000	0.30000	0.27000	0.20000	0.18000	

Table 4 The means and variances of the parameter estimation errors

Algorithms	Means	Variances
SG	0.22881	0.21693
FF-SG	0.11682	0.44864
A-SG	0.10235	0.14723
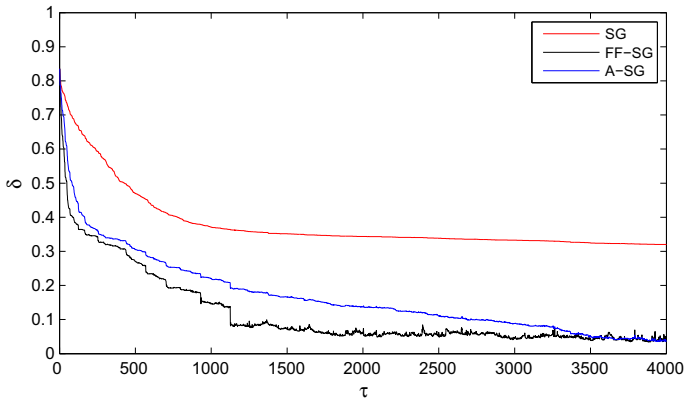
Fig. 3 The SG, FF-SG and A-SG estimation errors δ versus τ of Example 1
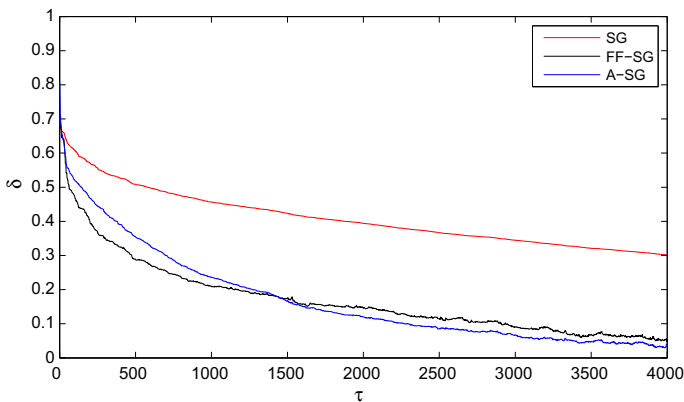

Fig. 4 The SG, FF-SG and A-SG estimation errors δ versus τ of Example 2

From these two examples, we can get the following finds.

1. Tables 1, 2, 3 show that the FF-SG algorithm and the A-SG algorithm are better than the SG algorithm.
2. Figures 3 and 4 show that the estimation error curve of the FF-SG algorithm oscillates seriously when the errors converge to zero, but the estimation error curve of the A-SG algorithm is relatively smooth.
3. Table 4 shows that the A-SG algorithm is the most effective algorithm among these three algorithms.
4. The algorithms proposed in this paper can not only identify the Hammerstein system with white noise, but also the Hammerstein system with colored noise.

6 Conclusions

In this paper, some SG algorithms are proposed for Hammerstein systems with piecewise linearity. The key term separation method is used to transform the nonlinear model into a regression model. In order to accelerate the convergence rate of the SG algorithm, an FF-SG algorithm and an A-SG algorithm are studied. Compared with the FF-SG algorithm, the A-SG algorithm has almost the same estimation error mean but smaller estimation error variance. Therefore, the A-SG algorithm has a wider application prospect in system identification.

The purpose of this paper is to develop two accelerated SG algorithms for nonlinear systems. These methods can combine other identification algorithms, e.g., recursive least squares algorithm, expectation–maximization algorithm, to study the parameter estimation issues of time-delay systems, switching systems and neural network learning systems [25,44,45].

Data Availability Statement All data generated or analyzed during this study are included in this article.

References

1. M. Ahmadi, H. Mojallali, Identification of multiple-input single-output Hammerstein models using Bezier curves and Bernstein polynomials. Appl. Math. Modell. **35**(4), 1969–1982 (2011)
2. E.W. Bai, An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. Automatica **34**(3), 333–338 (1998)
3. E.W. Bai, Identification of linear systems with hard input nonlinearities of known structure. Automatica **38**(5), 853–860 (2002)
4. G. Bottegal, A.Y. Aravkin, H. Hjalmarsson, G. Pillonetto, Robust EM kernel-based methods for linear system identification. Automatica **67**, 114–126 (2016)
5. O. Bumbariu, A new Aitken type method for accelerating iterative sequences. Appl. Math. Comput. **219**(1), 78–82 (2012)
6. G.Y. Chen, M. Gan, G.L. Chen, Generalized exponential autoregressive models for nonlinear time series: Stationarity, estimation and applications. Inf. Sci. **438**, 46–57 (2018)
7. J. Chen, Modified stochastic gradient algorithms with fast convergence rates. J. Vib. Control **17**(9), 1281–1286 (2011)
8. J. Chen, Y.J. Liu, Q.M. Zhu, Multi-step-length gradient iterative algorithm for equation-error type models. Syst. Control Lett. **115**, 15–21 (2018)
9. J. Chen, X.P. Wang, R. Ding, Gradient based estimation algorithm for Hammerstein systems with saturation and dead-zone nonlinearities. Appl. Math. Modell. **36**, 238–243 (2012)
10. J. Chen, Q.M. Zhu, J. Li, Biased compensation recursive least squares algorithm for rational models. Nonlinear Dyn. **91**(2), 797–807 (2018)
11. F. Ding, X.P. Liu, G. Liu, Identification methods for Hammerstein nonlinear systems. Digit. Signal Process **21**(2), 215–238 (2011)
12. F. Ding, Y.J. Liu, B. Bao, Gradient based and least squares based iterative estimation algorithms for multi-input multi-output systems. Proc. Inst. Mech. Eng. Part I: J. Syst. Control Eng. **226**(1), 43–55 (2012)
13. F. Ding, L. Lv, J. Pan, X.K. Wan, X.B. Jin, Two-stage gradient-based iterative estimation methods for controlled autoregressive systems using the measurement data. Int. J. Control Autom. Syst. **18**(4), 886–896 (2020)

14. F. Ding, L. Xu, D.D. Meng et al., Gradient estimation algorithms for the parameter identification of bilinear systems using the auxiliary model. J. Comput. Appl. Math. **369**, 112575 (2020)
15. F. Ding, L. Xu, Q.M. Zhu, Performance analysis of the generalised projection identification for time-varying systems. IET Control Theory Appl. **10**(18), 2506–2514 (2016)
16. F. Ding, X. Zhang, L. Xu, The innovation algorithms for multivariable state-space models. Int. J. Adapt. Control Signal Process. **33**(11), 1601–1608 (2019)
17. M. Gan, C.L.P. Chen, G.Y. Chen, L. Chen, On some separated algorithms for separable nonlinear squares problems. IEEE Trans. Cybern. **48**(10), 2866–2874 (2018)
18. A. Hagenblad, L. Ljung, A. Wills, Maximum likelihood identification of Wiener models. Automatica **44**(11), 2697–2705 (2008)
19. J.T. Hu, G.X. Sui, X.X. Lv, X.D. Li, Fixed-time control of delayed neural networks with impulsive perturbations. Nonlinear Anal.: Modell. Control **23**(6), 904–920 (2018)
20. A. Hussu, The conjugate-gradient method for optimal control problems with undetermined final time. Int. J. Control **15**(1), 79–82 (1972)
21. 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)
22. 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)
23. M.H. Li, X.M. Liu, Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique. Int. J. Control Autom. Syst. **18**(6), 1581–1592 (2020)
24. J.S. Li, Y.Y. Zheng, Z.P. Lin, Recursive identification of time-varying systems: Self-tuning and matrix RLS algorithms. Syst. Control Lett. **66**, 104–110 (2014)
25. X. Li, D. O'Regan, H. Akca, Global exponential stabilization of impulsive neural networks with unbounded continuously distributed delays. IMA J. Appl. Math. **80**(1), 85–99 (2015)
26. X. Liu, J. Cao, W. Yu, Q. Song, Nonsmooth finite-time synchronization of switched coupled neural networks. IEEE Trans. Cybern. **46**(10), 2360–2371 (2016)
27. X. Liu, J. Lam, W. Yu, G. Chen, Finite-time consensus of multiagent systems with a switching protocol. IEEE Trans. Neural Netw. Learn. Syst. **27**(4), 853–862 (2016)
28. X.Y. Liu, H.S. Su, M.Z.Q. Chen, A switching approach to designing finite-time synchronization controllers of coupled neural networks. IEEE Trans. Neural Netw. Learn. Syst. **27**(2), 471–482 (2016)
29. H. Ma, J. Pan et al., Partially-coupled least squares based iterative parameter estimation for multivariable output-error-like autoregressive moving average systems. IET Control Theory Appl. **13**(18), 3040–3051 (2019)
30. J.X. Ma, W.L. Xiong et al., Data filtering based forgetting factor stochastic gradient algorithm for Hammerstein systems with saturation and preload nonlinearities. J. Frankl. Inst. **353**(16), 4280–4299 (2016)
31. H. Oktem, A survey on piecewise-linear models of regulatory dynamical systems. Nonlinear Anal.: Theory, Method Appl. **63**(3), 336–349 (2005)
32. 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)
33. 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)
34. I. Pavaloiu, E. Catinas, On a robust Aitken–Newton method based on the Hermite polynomial. Appl. Math. Comput. **287**, 224–231 (2016)
35. C. Philippe, S.C. Johan, Hammerstein–Wiener system estimator initialization. Automatica **40**(9), 1543–1550 (2004)
36. H. Salhi, S. Kamoun, A recursive parametric estimation algorithm of multivariable nonlinear systems described by Hammerstein mathematical models. Appl. Math. Modell. **39**(16), 4951–4962 (2015)
37. J. Vörös, Parameter identification of Wiener systems with multisegment piecewise-linear nonlinearities. Syst. Control Lett. **56**(2), 99–105 (2007)
38. L.J. Wang, Y. Ji, L.J. Wan, N. Bu, Hierarchical recursive generalized extended least squares estimation algorithms for a class of nonlinear stochastic systems with colored noise. J. Frankl. Inst. **356**(16), 10102–10122 (2019)
39. X.H. Wang, T. Hayat, A. Alsaedi, Combined state and multi-innovation parameter estimation for an input nonlinear state space system using the key term separation. IET Control Theory Appl. **10**(13), 1503–1512 (2016)

40. L. Xu, The damping iterative parameter identification method for dynamical systems based on the sine signal measurement. Signal Process. **120**, 660–667 (2016)
41. L. Xu, F. Ding, Iterative parameter estimation for signal models based on measured data. Circuits Syst. Signal Process. **37**(7), 3046–3069 (2018)
42. L. Xu, F. Ding, Recursive least squares and multi-innovation stochastic gradient parameter estimation methods for signal modeling. Circuits Syst. Signal Process. **36**(4), 1735–1753 (2017)
43. L. Xu, W.L. Xiong, A. Alsaedi, T. Hayat, Hierarchical parameter estimation for the frequency response based on the dynamical window data. Int. J. Control Autom. Syst. **16**(4), 1756–1764 (2018)
44. X. Yang, X. Li, Q. Xi, P. Duan, Review of stability and stabilization for impulsive delayed systems. Math. Biosci. Eng. **15**(6), 1495–1515 (2018)
45. X. Zhang, F. Ding, Adaptive parameter estimation for a general dynamical system with unknown states. Int. J. Robust Nonlinear Control **30**(4), 1351–1372 (2020)
46. X. Zhang, F. Ding, L. Xu, Recursive parameter estimation methods and convergence analysis for a special class of nonlinear systems. Int. J. Robust Nonlinear Control **30**(4), 1373–1393 (2020)
47. X. Zhang, F. Ding, L. Xu, E.F. Yang, Highly computationally efficient state filter based on the delta operator. Int. J. Adapt. Control Signal Process. **33**(6), 875–889 (2019)
48. X. Zhang, Q.Y. Liu et al., Recursive identification of bilinear time-delay systems through the redundant rule. J. Frankl. Inst. **357**(1), 726–747 (2020)