

Research Article

Extended Stochastic Gradient Identification Method for Hammerstein Model Based on Approximate Least Absolute Deviation

Bao-chang Xu,¹ Zhong-hua Lin,¹ Ying-Dan Zhang,² and Yu-yue Xiao¹

¹Department of Automation, China University of Petroleum, Beijing 102200, China

²Beijing Urban Construction Design & Development Group Co., Limited, Beijing 100037, China

Correspondence should be addressed to Bao-chang Xu; xbcyl@cup.edu.cn

Received 24 February 2016; Revised 19 May 2016; Accepted 19 May 2016

Academic Editor: Mitsuhiro Okayasu

Copyright © 2016 Bao-chang Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to identify the parameters of nonlinear Hammerstein model which are contaminated by colored noise and peak noise, the least absolute deviation (LAD) is selected as the objective function to solve the problem of large residual square when the identification data is disturbed by the impulse noise which obeys symmetrical alpha stable (S α S) distribution. However, LAD cannot meet the need of differentiability required by most algorithms. To improve robustness and to solve the nondifferentiable problem, an approximate least absolute deviation (ALAD) objective function is established by introducing a deterministic function to replace absolute value under certain situations. The proposed method is derived from ALAD criterion and extended stochastic gradient method. Due to the differentiability of the objective function, we can get a recursive identification algorithm which is simple and easy to calculate compared with LAD. The convergence of the proposed identification method is also proved by Lyapunov stability theory, and the simulation experiments show that the proposed method has higher accuracy and stronger robustness than the least square (LS) method in the identification of Hammerstein model with colored noise and impulse noise. The impact of impulse noise can be restrained effectively.

1. Introduction

In recent years, the Hammerstein model has drawn a lot of attention because of its block-oriented nonlinear (BONL) character. It can be used to describe a variety of nonlinear systems such as nonlinear filtering, actuator saturation, signal analysis, and chemistry in biology system. So far, there have been several kinds of methods for the identification of Hammerstein model with white noise such as the iterative method [1, 2], the overparameterization method [3], the support vector machine (SVM) method [4], the subspace method [5–7], the blind method [8], the frequency domain identification method [9], and the artificial intelligence method [10–12]. In industry process control, however, the noise always includes colored noise and peak noise. It is necessary to consider the identification method for Hammerstein model when data are contaminated by colored noise and peak noise. Chang and Luus proposed an iterative method for Hammerstein model

with colored noise [13], but it cannot be used for online-identification. Ding and Chen proposed an extended stochastic gradient method based on least squares for the nonlinear Hammerstein ARMAX system and proved its convergence by the martingale convergence theorem [2]. In most researches, the LS criterion is taken as the objective function during the identification of Hammerstein model. The studies show that the LS method has better identification effect when the stochastic noise is normally distributed [14]. However, the LS method cannot work as well as the LAD method in certain conditions; for example, if the noise does not obey the normal distribution, the statistics performance of LAD estimation is better than LS estimation with irreplaceable advantages [15]. As to the square term in the objective function of LS method, a small change of the measured data will lead to a great influence on the identification results when there are outliers in the measured data [16].

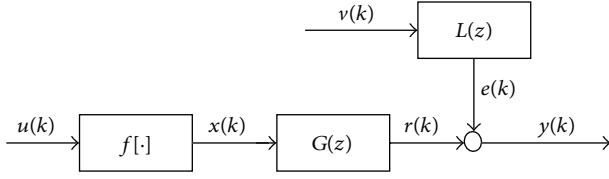


FIGURE 1: The structure of Hammerstein model.

To compensate the effect of the impulse noise and outliers on the identification accuracy, the LAD criterion is chosen to be the objective function which replaces the square terms with absolute deviation. The LAD method decreases the sensitivity to impulse noise and outliers and greatly improves the robustness because the LAD criterion only does the first power computation of the deviation. The LAD objective function is not differentiable. It needs to solve a nonsmooth optimization problem [17], which complicates the computation. The proposed method replaces the absolute deviation in LAD with a certain differentiable function and rebuilds the ALAD objective function. This paper derives the identification algorithm for Hammerstein model from ALAD objective function and the extended stochastic gradient method. To improve the identification accuracy and convergence rates, we add an inertial term to the proposed method. The convergence of the algorithm is proved by Lyapunov stability theory at the same time. The simulation experiments show that the proposed method can effectively eliminate the influence of impulse noise and outliers. Compared with the LS method, the ALAD method has stronger robustness and higher identification accuracy, demonstrating the superiority of the proposed method.

The rest of this paper is organized as follows. Section 2 describes the identification problem of nonlinear Hammerstein systems with colored noise. Section 3 derives the proposed identification algorithm from ALAD criterion. The convergence of the proposed method is also discussed in Section 3. Section 4 offers an illustrative example and compares the proposed algorithm with existing LS methods. Some concluding remarks are provided in Section 5.

2. Hammerstein Model with Colored Noise

As is shown in Figure 1, the structure of Hammerstein model is a series connection of a nonmemory static nonlinear block and a linear dynamic block. The static nonlinear block is connected with an input signal; the linear dynamic block is connected with an output signal.

In Figure 1, $u(k)$ is an input and $x(k)$ is the output of the nonlinear part and it is the input of the linear part as well. $r(k)$ is a real output. $y(k)$ is a measured output and the result of $r(k)$ which is disturbed by the colored noise $e(k)$. $v(k)$ is an additive white noise with zero mean. The colored noise $e(k)$ is the output of the linear link $L(z)$ that is driven by the white noise $v(k)$. $L(z)$ is the “noise model” which decides the property of $e(k)$, and $G(z)$ is the transfer function. Assuming

the nonlinear part $x(k)$ could be represented by a polynomial about the input $u(k)$ with known order

$$\begin{aligned} x(k) &= f(u(k)) \\ &= c_1 f_1(u(k)) + c_2 f_2(u(k)) + \cdots + c_{n_c} f_{n_c}(u(k)) \\ &= \sum_{j=1}^{n_c} c_j f_j(u(k)) \end{aligned} \quad (1)$$

the linear block of Hammerstein model is described by the model of CARMA/ARMAX (Controlled Autoregressive Moving Average/Autoregressive Moving Average exogenous), the transfer function $G(z) = B(z)/A(z)$, and $L(z) = D(z)/A(z)$. Therefore, the relationship of input and output can be written as

$$y(k) = r(k) + e(k) = \frac{B(z)}{A(z)}x(k) + \frac{D(z)}{A(z)}v(k). \quad (2)$$

From (1) and (2), it is obtained that

$$A(z)y(k) = B(z)\sum_{j=1}^{n_c} c_j f_j(u(k)) + D(z)v(k), \quad (3)$$

where $A(z)$, $B(z)$, and $D(z)$ are the polynomials of the shift operator:

$$\begin{aligned} A(z) &= 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a}; \\ B(z) &= b_1 z^{-1} + \cdots + b_{n_b} z^{-n_b}; \\ D(z) &= 1 + d_1 z^{-1} + \cdots + d_{n_d} z^{-n_d}, \end{aligned} \quad (4)$$

where a_i , b_j , c_k , and d_l ($i = 1, \dots, n_a$, $j = 1, \dots, n_b$, $k = 1, \dots, n_c$, $l = 1, \dots, n_d$) are the parameters to be identified.

From (3), it is obtained that

$$\begin{aligned} y(k) &= -\sum_{i=1}^{n_a} a_i y(k-i) + \sum_{i=1}^{n_b} b_i \sum_{j=1}^{n_c} c_j f_j(u(k-i)) \\ &\quad + \sum_{i=1}^{n_d} d_i v(k-i) + v(k). \end{aligned} \quad (5)$$

Define

$$\theta = \begin{bmatrix} \mathbf{a} \\ \mathbf{d} \\ \boldsymbol{\zeta} \end{bmatrix} \in R^n, \quad n = n_a + n_b \times n_c + n_d, \quad (6)$$

where

$$\boldsymbol{\xi} = \mathbf{b} \otimes \mathbf{c} = \begin{bmatrix} \mathbf{b} \times c_1 \\ \mathbf{b} \times c_2 \\ \vdots \\ \mathbf{b} \times c_{n_c} \end{bmatrix} \in R^{(n_b \times n_c) \times 1}. \quad (7)$$

a, **b**, and **d** are the parameter vectors of the linear section and **c** is the parameter vector of the nonlinear section of the Hammerstein model:

$$\begin{aligned} \mathbf{a} &= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n_a} \end{bmatrix}, \\ \mathbf{b} &= \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n_b} \end{bmatrix}, \\ \mathbf{c} &= \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{n_c} \end{bmatrix}, \\ \mathbf{d} &= \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{n_d} \end{bmatrix}. \end{aligned} \quad (8)$$

We define

$$\begin{aligned} \boldsymbol{\varphi}(k) &= [-y(k-1), -y(k-2), \dots, -y(k-n_a), \\ &\quad v(k-1), v(k-2), \dots, v(k-n_d), \boldsymbol{\psi}(k)]^T, \\ \boldsymbol{\psi}(k) &= [f_1(u(k-1)), f_1(u(k-2)), \dots, \\ &\quad f_1(u(k-n_b)), f_2(u(k-1)), f_2(u(k-2)), \dots, \\ &\quad f_2(u(k-n_b)), \dots, f_{n_c}(u(k-1)), f_{n_c}(u(k-2)), \dots, \\ &\quad f_{n_c}(u(k-n_b))]. \end{aligned} \quad (9)$$

Then, (5) can be rewritten as

$$y(k) = \boldsymbol{\varphi}^T(k) \boldsymbol{\theta} + v(k). \quad (10)$$

3. The Identification Method Based on ALAD Criterion and Extended Stochastic Gradient Method

3.1. Approximate Least Absolute Deviation. As is known, the absolute deviation is not differentiable. To overcome this shortage, we choose a certain differentiable function to

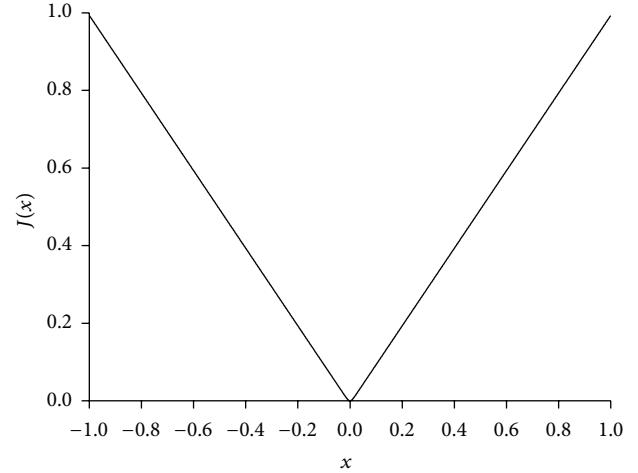


FIGURE 2: The curve of function property when $\beta = 0.01$.

approximate the LAD criterion. This requirement can be met by the following logarithmic function:

$$J(x) = \beta \ln \left[\cosh \left(\frac{x}{\beta} \right) \right]. \quad (11)$$

In (11), β is an adjustable parameter, and the nonlinear function $J(*)$ is related to β . While β is small enough, nonlinear function $\cosh(x/\beta) \approx (1/2)e^{|x|/\beta}$, and $\beta \ln[\cosh(x/\beta)] \approx |x|$; that is, $J(x) \approx |x|$, approximating the absolute value function. The property curve of the logarithmic function when $\beta = 0.01$ is shown in Figure 2.

The curve shows that while β is small enough, nonlinear function $J(x) = \beta \ln[\cosh(x/\beta)]$ can approximate the absolute value function $|x|$ effectively. Because $J(x)$ is differentiable, then the identification algorithm based on $J(x)$ can be solved by some optimization methods.

3.2. Algorithm Derivation. In (10), the input $u(k)$ and output $y(k)$ are measurement data. According to the known structure of the nonlinear function $f(*)$, we can get the value of $f(u(k))$. But $v(k-i)$ in $\boldsymbol{\varphi}(k)$ is white noise and immeasurable, which makes it impossible to solve the identification problem. Based on the idea that the noise can be replaced by its estimated value, we use $\hat{v}(k-i)$, the noise residual in the i th step, to replace $v(k-i)$ in the $i+1$ th step, so $\boldsymbol{\varphi}(k)$ can be replaced by its estimated value $\hat{\boldsymbol{\varphi}}(k)$:

$$\begin{aligned} \hat{\boldsymbol{\varphi}}(k) &= [-y(k-1), -y(k-2), \dots, -y(k-n_a), \\ &\quad \hat{v}(k-1), \hat{v}(k-2), \dots, \hat{v}(k-n_d), \boldsymbol{\psi}(k)]^T. \end{aligned} \quad (12)$$

Then, the identification of $\boldsymbol{\theta}$ in (10) is to find $\boldsymbol{\theta}$ that minimizes the following criterion function (ALAD) based on input data $\hat{\boldsymbol{\varphi}}(k)$ and output data $y(k)$:

$$J(\boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(k)} = \beta \ln \cosh \left[\frac{y(k) - \boldsymbol{\varphi}^T(k) \boldsymbol{\theta}}{\beta} \right] \Bigg|_{\hat{\boldsymbol{\theta}}(k)} = \min. \quad (13)$$

For (13), it is a typical nonlinear function optimization. To reduce the computation complexity, we use the stochastic

gradient method to derive the recursive equation of $\hat{\theta}(k)$ [18, 19]. The formula of the stochastic gradient method is

$$\hat{\theta}(k) = \hat{\theta}(k-1) - \mu_k \text{grad}(J(\theta))|_{\hat{\theta}(k-1)}, \quad (14)$$

where $\text{grad}(\ast)$ represents gradient and μ_k is step size. The formula above shows that $\hat{\theta}(k)$ is corrected along the negative gradient direction until the extreme value of $J(\theta)$ is obtained.

$$\begin{aligned} J(\theta)|_{\hat{\theta}(k-1)} &= \beta \ln \cosh \left[\frac{y(k) - \hat{\varphi}^T(k) (\hat{\theta}(k-1) + \mu_k \hat{\varphi}(k) \tanh((y(k) - \hat{\varphi}(k) \hat{\theta}(k-1)) / \beta))}{\beta} \right] \\ &= \beta \ln \cosh \left[\frac{1}{\beta} \left(\hat{e}(k) - \mu_k \hat{\varphi}^T(k) \hat{\varphi}(k) \tanh\left(\frac{\hat{e}(k)}{\beta}\right) \right) \right], \end{aligned} \quad (16)$$

where $\hat{e}(k) = y(k) - \hat{\varphi}^T(k) \hat{\theta}(k-1)$.

We can get the gradient of $J(\theta)$ to μ_k from (16):

$$\frac{\partial J(\theta)}{\partial \mu_k} = -K_k \ast \tanh\left(\frac{\hat{e}(k) - \mu_k K_k}{\beta}\right), \quad (17)$$

where

$$K_k = \hat{\varphi}^T(k) \hat{\varphi}(k) \tanh\left(\frac{\hat{e}(k)}{\beta}\right). \quad (18)$$

Equation (17) is obtained based on the exact line search algorithm and always leads to the two adjacent gradient search directions being orthogonal and usually results in zigzag phenomenon. When we use (17) to compute the optimal μ_k , the convergence rate is slow because of the linear convergence order, and the identification results of the parameters may be fluctuated. For this shortcoming, an improved stochastic gradient method is applied to get μ_k where μ_k is based on the information of gradient that is k times before the current iteration step; that is,

$$K_k = \sum_{l=1}^k \hat{\varphi}^T(l) \hat{\varphi}(l) \tanh\left(\frac{\hat{e}(l)}{\beta}\right). \quad (19)$$

From $\hat{e}(k) - \mu_k K_k = 0$, it is obtained that

$$\mu_k = \frac{\hat{e}(k)}{K_k} = \frac{\hat{e}(k)}{\sum_{l=1}^k \hat{\varphi}^T(l) \hat{\varphi}(l) \tanh(\hat{e}(l) / \beta)}. \quad (20)$$

From (14), (15), and (19), it is obtained that

$$\begin{aligned} \hat{\theta}(k) &= \hat{\theta}(k-1) + \frac{\hat{e}(k)}{\sum_{l=1}^k \hat{\varphi}^T(l) \hat{\varphi}(l) \tanh(\hat{e}(l) / \beta)} \\ &\quad \cdot \hat{\varphi}(k) \tanh\left(\frac{\hat{e}(k)}{\beta}\right); \end{aligned} \quad (21)$$

then

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{\hat{\varphi}(k)}{r(k)} \tanh\left(\frac{\hat{e}(k)}{\beta}\right), \quad (22)$$

Firstly, we get the gradient of $J(\theta)$:

$$\begin{aligned} \text{grad}(J(\theta))|_{\hat{\theta}(k-1)} &= \frac{\partial J}{\partial \theta}|_{\hat{\theta}(k-1)} \\ &= -\hat{\varphi}(k) \tanh\left(\frac{y(k) - \hat{\varphi}^T(k) \hat{\theta}(k-1)}{\beta}\right); \end{aligned} \quad (15)$$

then, from (15) and (13), we obtain

where

$$r(k) = \frac{1}{\mu(k)} = \frac{\sum_{l=1}^k \hat{\varphi}^T(k) \hat{\varphi}(k) \tanh(\hat{e}(k) / \beta)}{\hat{e}(k)}. \quad (23)$$

$r(k)$ is computed by the following recursive formula:

$$\begin{aligned} r(k) &= \lambda(k) r(k-1) \\ &\quad + \frac{1}{\hat{e}(k)} \left[\hat{\varphi}^T(k) \hat{\varphi}(k) \tanh\left(\frac{\hat{e}(k)}{\beta}\right) \right], \end{aligned} \quad (24)$$

where $\lambda(k) = \hat{e}(k-1) / \hat{e}(k)$.

Equation (24) is equivalent for adding an inertial term to the recursive equation of $\mu(k)$. The weights of the inertial term are related to the iterative output error at $k-1$ and k step.

In conclusion, we can obtain the extended stochastic gradient identification method based on approximate least absolute deviation (ALADESG):

$$\begin{aligned} \hat{\theta}(k) &= \hat{\theta}(k-1) + \frac{\hat{\varphi}(k)}{r(k)} \tanh\left(\frac{\hat{e}(k)}{\beta}\right), \\ r(k) &= \lambda(k) r(k-1) + \frac{1}{\hat{e}(k)} \left[\hat{\varphi}^T(k) \hat{\varphi}(k) \right. \\ &\quad \cdot \left. \tanh\left(\frac{\hat{e}(k)}{\beta}\right) \right], \\ \hat{e}(k) &= y(k) - \hat{\varphi}^T(k) \hat{\theta}_{k-1}, \\ \hat{\varphi}(k) &= [-y(k-1), -y(k-2), \dots, -y(k-n_a), \\ &\quad \hat{v}(k-1), \hat{v}(k-2), \dots, \hat{v}(k-n_d), \psi(k)]^T, \\ \psi(k) &= [f_1(u(k-1)), f_1(u(k-2)), \dots, \\ &\quad f_1(u(k-n_b)), f_2(u(k-1)), f_2(u(k-2)), \dots, \\ &\quad f_2(u(k-n_b)), \dots, f_{n_c}(u(k-1)), f_{n_c}(u(k-2)), \dots, \\ &\quad f_{n_c}(u(k-n_b))]. \end{aligned} \quad (25)$$

3.3. *Analysis of Convergence.* Rewrite (22) as

$$\hat{\theta}(k) = \hat{\theta}(k-1) + c(k) \hat{\varphi}(k) \hat{e}(k), \quad (26)$$

where

$$c(k) = \frac{\tanh(\hat{e}(k)/\beta)}{\sum_{l=1}^k \hat{\varphi}^T(l) \hat{\varphi}(l) \tanh(\hat{e}(l)/\beta)}. \quad (27)$$

If the truth values of process parameters are noted as θ_0 , then $y(k) = \hat{\varphi}^T(k) \theta_0$.

Defining $\tilde{\theta}(k) = \theta_0 - \hat{\theta}(k)$, according to (26), we have

$$\begin{aligned} \hat{\theta}(k) &= \hat{\theta}(k-1) \\ &+ c(k) \hat{\varphi}(k) (y(k) - \hat{\varphi}^T(k) \hat{\theta}(k-1)); \end{aligned} \quad (28)$$

subtracting θ_0 on both sides of (28)

$$-\tilde{\theta}(k) = -\tilde{\theta}(k-1) + c(k) \hat{\varphi}(k) \hat{\varphi}^T(k) \tilde{\theta}(k-1) \quad (29)$$

then

$$\tilde{\theta}(k) = [I - c(k) \hat{\varphi}(k) \hat{\varphi}^T(k)] \tilde{\theta}(k-1). \quad (30)$$

Constructing the Lyapunov function

$$V(\tilde{\theta}(k), k) = \tilde{\theta}^T(k) \tilde{\theta}(k) > 0 \quad (31)$$

then

$$\begin{aligned} \Delta V(\tilde{\theta}(k), k) &= V(\tilde{\theta}(k), k) - V(\tilde{\theta}(k-1), k-1) \\ &= \tilde{\theta}^T(k-1) G(k) \tilde{\theta}(k-1), \end{aligned} \quad (32)$$

where

$$\begin{aligned} G(k) &= [I - c(k) \hat{\varphi}(k) \hat{\varphi}^T(k)]^T [I - c(k) \hat{\varphi}(k) \hat{\varphi}^T(k)] \\ &- I = [\hat{\varphi}^T(k) \hat{\varphi}(k) c^2(k) - 2c(k)] \hat{\varphi}(k) \hat{\varphi}^T(k). \end{aligned} \quad (33)$$

$G(k)$ is a real symmetric matrix, and the rank of $r(G) = 1$. Obviously, when the following inequality in (34) is fulfilled, the matrix $G(k)$ is negative and semidefinite, and then (28) will be asymptotically stable in a large scope at the balance point $\tilde{\theta}(k) = 0$:

$$c^2(k) \hat{\varphi}^T(k) \hat{\varphi}(k) - 2c(k) \leq 0. \quad (34)$$

According to (27),

$$\begin{aligned} c^2(k) \hat{\varphi}^T(k) \hat{\varphi}(k) - 2c(k) &= \frac{\tanh^2(\hat{e}(k)/\beta) \hat{\varphi}^T(k) \hat{\varphi}(k)}{[\sum_{l=1}^k \hat{\varphi}^T(l) \hat{\varphi}(l) \tanh(\hat{e}(l)/\beta)]^2} - \frac{2 \tanh(\hat{e}(k)/\beta)}{\sum_{l=1}^k \hat{\varphi}^T(l) \hat{\varphi}(l) \tanh(\hat{e}(l)/\beta)} \\ &= \frac{\tanh^2(\hat{e}(k)/\beta) \hat{\varphi}^T(k) \hat{\varphi}(k)}{[\sum_{l=1}^k \hat{\varphi}^T(l) \hat{\varphi}(l) \tanh(\hat{e}(l)/\beta)]^2} - \frac{2 \tanh(\hat{e}(k)/\beta) \sum_{l=1}^k \hat{\varphi}^T(l) \hat{\varphi}(l) \tanh(\hat{e}(l)/\beta)}{[\sum_{l=1}^k \hat{\varphi}^T(l) \hat{\varphi}(l) \tanh(\hat{e}(l)/\beta)]^2} \\ &= \frac{-\tanh^2(\hat{e}(k)/\beta) \sum_{j=1}^n \varphi_j^2(k) - 2 \tanh(\hat{e}(k)/\beta) \sum_{l=1}^{k-1} \hat{\varphi}^T(l) \hat{\varphi}(l) \tanh(\hat{e}(l)/\beta)}{[\sum_{l=1}^k \hat{\varphi}^T(l) \hat{\varphi}(l) \tanh(\hat{e}(l)/\beta)]^2}. \end{aligned} \quad (35)$$

The second term of (35) can be written as

$$\begin{aligned} &\frac{-2 \tanh(\hat{e}(k)/\beta) r(k-1) e(k-1)}{r^2(k) e^2(k)} \\ &= \frac{-2 \tanh(\hat{e}(k)/\beta)}{e(k)} \frac{1}{r^2(k)} \frac{e(k-1) r(k-1)}{e(k)}, \end{aligned} \quad (36)$$

where

$$\begin{aligned} \frac{\tanh(\hat{e}(k)/\beta)}{e(k)} &> 0, \\ \frac{1}{r^2(k)} &> 0. \end{aligned} \quad (37)$$

During the iterative process, if we could make sure that

$$\frac{e(k-1) r(k-1)}{e(k)} \geq 0 \quad (38)$$

then we get

$$c^2(k) \hat{\varphi}^T(k) \hat{\varphi}(k) - 2c(k) \leq 0. \quad (39)$$

Therefore, in order to guarantee the convergence of the proposed algorithm, we need to add (38) as a constraint when (24) is used for iteration.

3.4. *Separation of Parameters by Average Method.* It is noteworthy that in the description of the Hammerstein model (as shown in Figure 1), for some nonzero and finite constant a , any $af(u)$ and $G(z)/a$ can match and yield identical input and output measurements; thus, the nonlinear function $f(u)$ and the transfer function $G(z)$ are not unique. In other words, any identification scheme will not be able to distinguish between $(f(u), G(z))$ and $(af(u), G(z)/a)$. Therefore, without any loss of generality, one of the gains of $f(u)$ and $G(z)$ has to be fixed.

There are several ways to normalize these gains [20, 21]. We adopt the following method.

Assumption 1. The first coefficient of the function $f(\cdot)$ equals 1 [22]; that is to say, in (1), $c_1 = 1$.

Based on the relationship between the parameters, \hat{c} can be computed using

$$\hat{c}_{ji}(k) = \frac{\hat{\theta}_{n_a+n_b+i}(k)}{\hat{b}_i(k)}, \quad (40)$$

$$j = 2, 3, \dots, n_c, \quad i = 1, 2, \dots, n_b.$$

From (40), we can see that the number of $\hat{c}_{ji}(k)$ is n_b , and $\hat{c}_{ji}(k)$ is the estimated value of $c_j(k)$. In order to obtain the estimated value of the parameters with much higher accuracy, we use the average method below to estimate \hat{c} :

$$\hat{c}_j(k) = \frac{1}{n_b} \sum_{i=1}^{n_b} \hat{c}_{ji}(k) = \frac{1}{n_b} \sum_{i=1}^{n_b} \frac{\hat{\theta}_{n_a+n_b+i}(k)}{\hat{b}_i(k)}, \quad (41)$$

$$j = 2, 3, \dots, n_c.$$

4. Simulation Results and Discussions

Consider the following Hammerstein model:

$$A(z)y(k) = B(z)x(k) + D(z)v(k), \quad (42)$$

where $B(z^{-1}) = b_1z^{-1} + b_2z^{-2} = 0.85z^{-1} + 0.65z^{-2}$; $D(z^{-1}) = 1 + d_1z^{-1} = 1 - 0.64z^{-1}$; $A(z^{-1}) = 1 + a_1z^{-1} + a_2z^{-2} = 1 - 1.60z^{-1} + 0.80z^{-2}$. The nonlinear static part can be described as the polynomial below:

$$x(k) = f(u(k)) = c_1u(k) + c_2u^2(k) + c_3u^3(k) \quad (43)$$

$$= u(k) + 0.5u^2(k) + 0.25u^3(k),$$

where the input sequence $u(k)$ is a series of uncorrelated, continuous, and stochastic sequences with zero mean and unit variance $\sigma_u^2 = 1$. The noise sequence $\{v(k)\}$ is a group of stochastic white noise sequences with zero mean and variance $\sigma_v^2 = 0.5$. The input and output data for the identification procedure are obtained from this model. We then apply proposed algorithm (25) to estimate the parameters of this system at different situation. The identification process and result analysis are demonstrated in the following. The simulation experiments are performed in Matlab. The value of β is taken as 0.01.

The relative error of the parameters estimation is used as the evaluation criteria of the algorithm which is defined as follows:

$$\delta = \frac{\|\hat{\theta}(k) - \theta\|}{\|\theta\|}. \quad (44)$$

In order to verify the performance of the proposed algorithm, we designed two simulation cases.

Case 1. Only colored noise exists in the measured data. This case is used to prove the impact of inertial term on the identification results of ALADESG method.

Case 2. Not only colored noise but also impulse noise exists in the measured data. This case is used to compare the identification effects of ALADESG method with that of LSESG method when measured data are contaminated by impulse noise. The robustness of ALADESG method is also verified when impulse noise with different amplitude is added.

The impulse noise in the experiment is subject to SaS distribution [23], and the probability density function of the standard SaS distribution is

$$f_\alpha(x) = \begin{cases} \frac{1}{\pi x} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \Gamma(\alpha k + 1) |x|^{-\alpha k} \sin\left(\frac{kx\pi}{2}\right), & 0 < \alpha < 1, \\ \frac{1}{\pi x} \sum_{k=1}^{\infty} \frac{(-1)^k}{2k!} \Gamma\left[\frac{(2k+1)}{\alpha}\right] x^{2k}, & 1 < \alpha \leq 2, \end{cases} \quad (45)$$

where $\Gamma(*)$ is Γ gamma function and α is characteristics exponent. The less α is, the higher chance of large amplitude sample of the random variable which is subject to SaS distribution takes place, and the stronger pulse strength is.

The way to get a random variable which obeys the standard SaS distribution follows the steps below [24]:

- (1) Get a random variable U subject to uniform distribution on $(-\pi/2, \pi/2)$.
- (2) Get a random variable E subject to exponential distribution and mean value is 1.
- (3) Get the variable subject to standard SaS distribution by the following formula:

$$Y = \frac{\sin \alpha U}{(\cos U)^{1/\alpha}} \left(\frac{\cos((1-\alpha)U)}{E} \right)^{(1-\alpha)/\alpha}, \quad (46)$$

where U and E are independent.

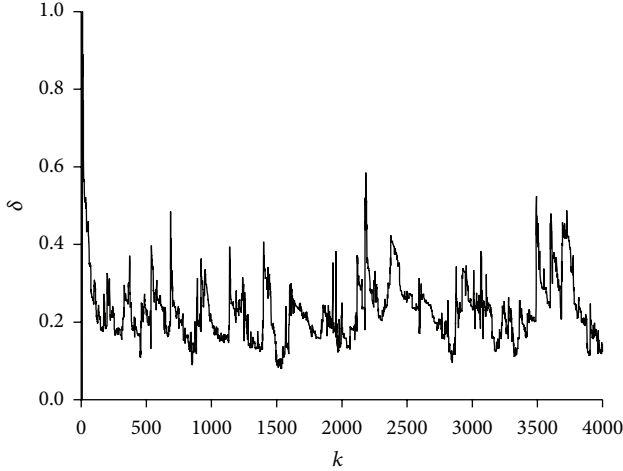
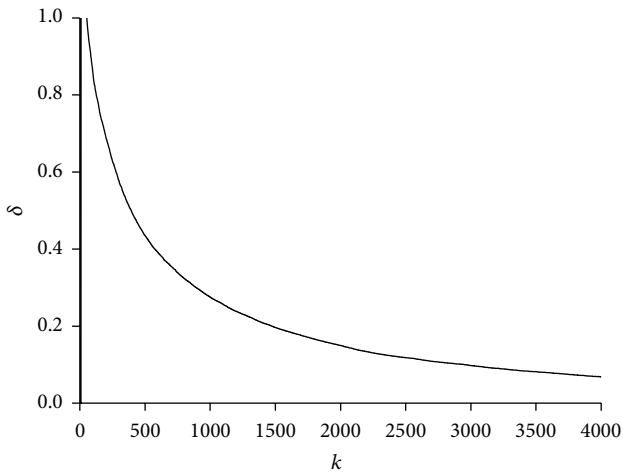
4.1. The Effect of Inertial Term $\lambda(k)$ on the Identification Results. (1) Let $\lambda(k) = 0$; namely, μ_k is computed by (17) with current gradient information only. The curve of the relative error is shown in Figure 3.

According to theoretical analysis, there are some drawbacks when (17) is used to compute the optimal search step size, such as linear convergence order, slow convergence rate, and fluctuant estimation of parameters. As shown in Figure 3, the curve of the relative error of the parameters of Hammerstein model fluctuates severely, and the estimated parameters cannot converge to the true value. That is to say, the method cannot provide a reliable parameter estimation for the identification of Hammerstein model when $\lambda(k) = 0$.

(2) In order to improve the accuracy and convergence rate and to enhance the reliability of the identification results, we add an inertial term $\lambda(k) = \bar{e}(k-1)/\bar{e}(k)$ to the method,

TABLE 1: Identification results when $\lambda(k) = \hat{e}(k-1)/\hat{e}(k)$.

Parameters	a_1	a_2	b_1	b_2	c_2	c_3	d_1	δ (%)
Estimated value	-1.5937	0.7943	0.8173	0.6895	0.5062	0.2533	-0.5452	6.8363
True value	-1.600	0.800	0.850	0.650	0.500	0.250	-0.640	0


 FIGURE 3: Relative error curve when $\lambda(k) = 0$.

 FIGURE 4: Relative error curve when $\lambda(k) = \hat{e}(k-1)/\hat{e}(k)$.

namely, applying (24) and (38) to parameter identification. The results are shown in Table 1.

Obviously, when $\lambda(k) = \hat{e}(k-1)/\hat{e}(k)$, the deviations of the estimated values and true values of model parameters are small, and the identification accuracy is much better than that of $\lambda(k) = 0$. This shows that the inertial term coming from the improved stochastic gradient method enhances the identification accuracy effectively and insures the convergence of the ALADESG method. The relative error curve is shown in Figure 4.

Comparing Figure 3 with Figure 4 shows that the relative error curve in Figure 4 is much smoother, the estimated parameters converge to those true values steadily, and the relative error is smaller than the situation as shown in

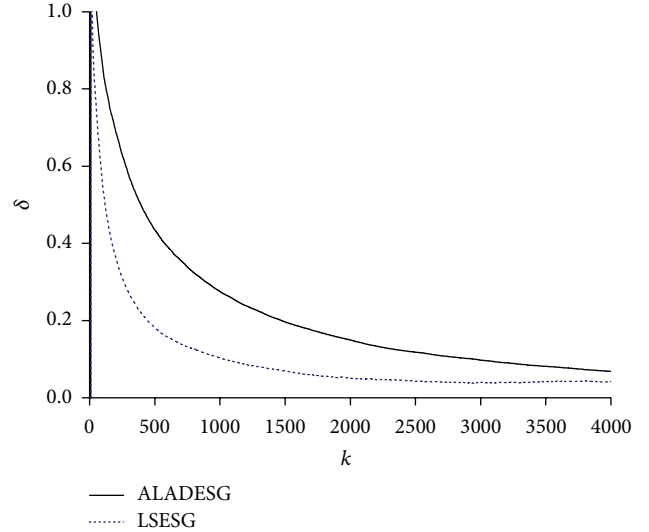


FIGURE 5: Relative error with colored noise.

Figure 3. Simulation results in Figures 3 and 4 show that the inertial term can greatly improve the identification accuracy and significantly enhance the reliability of the parameter estimation.

4.2. Comparison of ALADESG and LSESG

4.2.1. Considering the Measured Data with Colored Noise Only. According to the same idea in Section 3.2, we can get the iterative equations of LSESG also containing a similar inertial term. Then the proposed ALADESG method was compared with the LSESG method when the measured data are contaminated with colored noise. The simulation result is shown in Figure 5.

As shown in Figure 5, the identification accuracies of the ALADESG method and the LSESG method are both acceptable when the measured data are only disturbed by colored noise, but the identification accuracy of the LSESG method is higher and the convergence rate is faster. So when the measured data is contaminated with colored noise only, the identification performance of the LSESG method is better than the ALADESG method.

4.2.2. The Measured Data with Impulse Noise Subject to SaS Distribution. The measured data are contaminated with both colored noise and impulse noise. Let the characteristics exponent of the impulse noise be 1.5. The time-domain waveform of the impulse noise is shown in Figure 6. Then the measured data are identified by the ALADESG method and the LSESG method, respectively. The relative errors are shown in Figure 7.

TABLE 2: Identification results with different α .

Parameters	a_1	a_2	b_1	b_2	c_2	c_3	d_1	δ (%)
$\alpha = 1.5$	-1.5711	0.7832	0.8461	0.7040	0.4769	0.2501	-0.7541	10.2411
$\alpha = 1.2$	-1.5177	0.7147	0.8189	0.7741	0.4764	0.2580	-0.7038	11.4612
$\alpha = 0.9$	-1.4660	0.6617	0.8396	0.7817	0.4818	0.2515	-0.6394	15.1781
True value	-1.600	0.800	0.850	0.650	0.500	0.250	-0.640	0

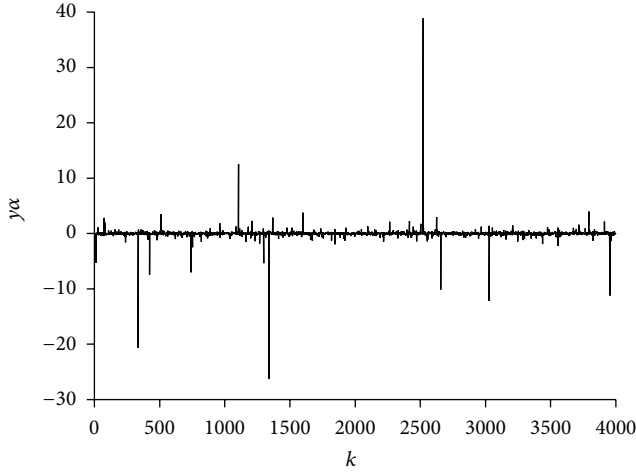
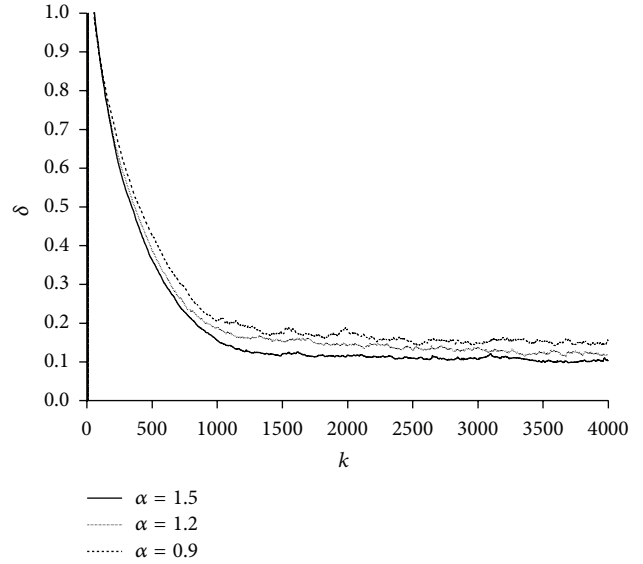
FIGURE 6: The time-domain waveform of the impulse noise when $\alpha = 1.5$.

FIGURE 8: Relative error with impulse noise of different amplitude.

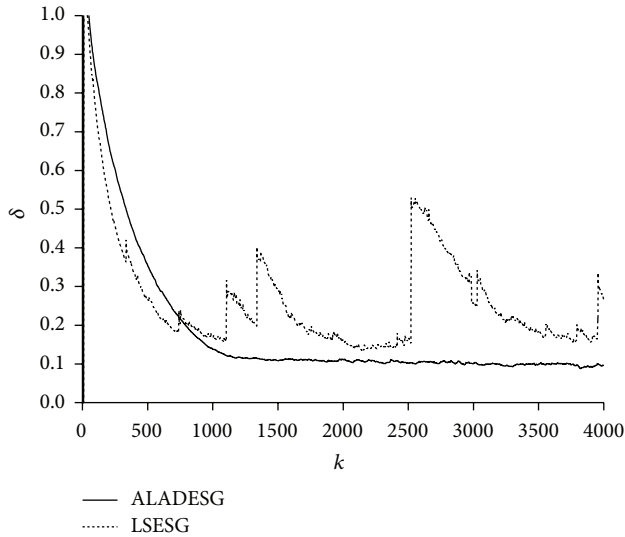
FIGURE 7: Relative error with impulse noise of $\alpha = 1.5$.

Figure 7 shows that the identification accuracy of the ALADESG method is higher than that of the LSESG method when the measured data are disturbed by colored noise and impulse noise with $\alpha = 1.5$. From Figures 6 and 7, we notice that the relative error of the LSESG method will fluctuate severely at the moment when a large impulse is added, and the larger the amplitude of the impulse is, the severer the fluctuation is, which makes the parameter estimation unable to converge steadily. In the LSESG method,

the objective function is the square of the error so the influence of impulse noise would be amplified especially when the impulse noise is large. In contrast, the ALADESG method is based on the ALAD criterion which restrains the influence of impulse noise and enhances the robustness, the parameter estimation converges to the true value quickly and steadily, and the identification accuracy is also improved. The simulation results indicate that the ALADESG method has a better identification performance than LSESG method when impulse noise is added.

4.2.3. The Influence of Different Pulse Amplitude on the Identification Results of ALADESG Method. In this part, we conduct three simulations with $\alpha = 1.5$, $\alpha = 1.2$, and $\alpha = 0.9$. The identification results of the ALADESG method are shown in Table 2, and the relative errors are shown in Figure 8. In addition, the identification results of the ALADESG method are compared with those of LSESG method when $\alpha = 1.2$ and $\alpha = 0.9$. The relative errors curves when $\alpha = 1.2$ and $\alpha = 0.9$ are shown in Figures 9 and 10, respectively.

From Table 2 and Figure 8, the identification accuracy of the ALADESG method becomes worse when α decreases, but the decrease of accuracy is small. In contrast, from Figures 7, 9, and 10, the identification accuracy of the LSESG method becomes worse when α is decreased and the decrease of accuracy is very obvious. Particularly when $\alpha = 1.2$ and $\alpha = 0.9$, the LSESG method cannot converge indeed, which suggests that the LSESG method cannot identify the parameters

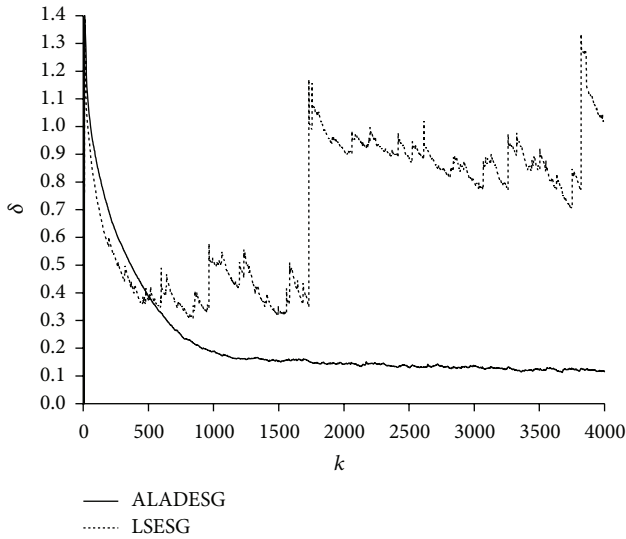


FIGURE 9: Relative error with impulse noise of $\alpha = 1.2$.

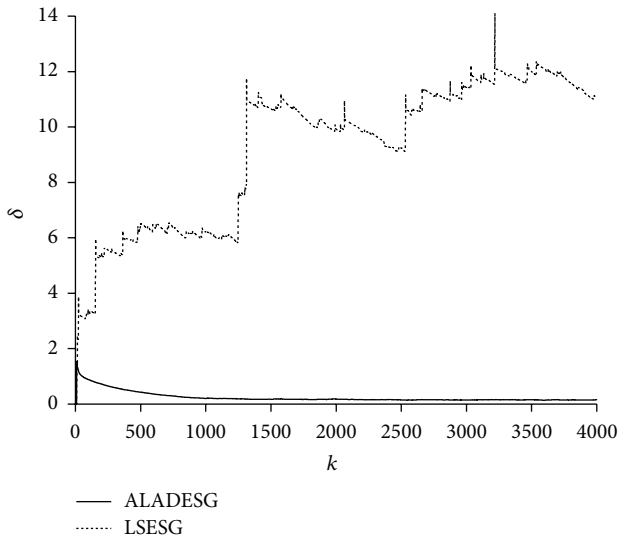


FIGURE 10: Relative error with impulse noise of $\alpha = 0.9$.

of Hammerstein model in this situation. But high accuracy and stable identification results still can be achieved by the ALADESG method. Therefore, it can be concluded that the method based on ALAD criterion overcomes the impact of impulse noise very well and has better robustness to impulse noise.

4.3. Identification Results Analysis. According to the identification results shown in Sections 4.1 and 4.2, the following conclusions are summarized.

- (1) When the measured data are contaminated by colored noise, the Hammerstein model can be identified effectively by the proposed method.
- (2) By applying inertial term $\lambda(k)$ which contains the gradient information of past time, the identification

accuracy and convergence rates are improved significantly.

- (3) The identification performance of the LSESG method is better than that of the ALADESG method when only colored noise exists.
- (4) The ALADESG method has better accuracy and convergent rate than LSESG method when both colored noise and impulse noise exist in measured data. Particularly when the amplitude of impulse noise is very large (such as when $\alpha = 0.9$), the ALADESG method still can get an acceptable identification result, which proves the adaptability of the method to more serious noise. The robustness of the method is also confirmed by simulation results.

5. Conclusions

In this paper, we expand the application of the LAD technique to the field of nonlinear identification. A new algorithm is proposed based on ALAD criterion and improved stochastic gradient search algorithm for the Hammerstein model, and the convergence is proven by Lyapunov stability theory. With the application of ALAD criterion, the proposed method overcomes the influence of impulse noise and colored noise on the identification results and improves the robustness to a large extent. The proposed algorithm is easy to implement. The simulation results show that the proposed algorithm yields better robustness against impulse noises and a faster convergence rate than that by the LSESG method.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by Important National Science & Technology Specific Projects of China (no. 2011ZX05027-005).

References

- [1] K. S. Narendra and P. G. Gallman, "An iterative method for the identification of nonlinear systems using a hammerstein model," *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 546–550, 1966.
- [2] F. Ding and T. Chen, "Identification of Hammerstein nonlinear ARMAX systems," *Automatica*, vol. 41, no. 9, pp. 1479–1489, 2005.
- [3] F. Ding, Y. Shi, and T. Chen, "Auxiliary model-based least-squares identification methods for Hammerstein output-error systems," *Systems & Control Letters*, vol. 56, no. 5, pp. 373–380, 2007.
- [4] I. Goethals, K. Pelckmans, J. A. Suykens, and B. De Moor, "Identification of MIMO Hammerstein models using least squares support vector machines," *Automatica*, vol. 41, no. 7, pp. 1263–1272, 2005.
- [5] L. Bako, G. Mercère, S. Lecoche, and M. Lovera, "Recursive subspace identification of Hammerstein models based on least

- squares support vector machines,” *IET Control Theory & Applications*, vol. 3, no. 9, pp. 1209–1216, 2009.
- [6] Z. Liao, Z. Zhu, S. Liang, C. Peng, and Y. Wang, “Subspace identification for fractional order hammerstein systems based on instrumental variables,” *International Journal of Control, Automation and Systems*, vol. 10, no. 5, pp. 947–953, 2012.
 - [7] M. Verhaegen and D. Westwick, “Identifying MIMO Hammerstein systems in the context of subspace model identification methods,” *International Journal of Control*, vol. 63, no. 2, pp. 331–349, 1996.
 - [8] L. Vanbeylen, R. Pintelon, and J. Schoukens, “Blind maximum likelihood identification of Hammerstein systems,” *Automatica*, vol. 44, no. 12, pp. 3139–3146, 2008.
 - [9] B. Er-Wei, “Frequency domain identification of Wiener models,” *Automatica*, vol. 44, no. 5, pp. 1451–1455, 2008.
 - [10] X.-Y. Li and Z.-G. Han, “Identification approach for nonlinear systems based on particle swarm optimization,” *Control and Decision*, vol. 26, no. 11, pp. 1627–1631, 2011.
 - [11] X. Xiaoping, Q. Fucai, and W. Feng, “Identification method for Hammerstein models based on hybrid particle swarm optimization,” *Journal of Engineering Mathematics*, vol. 27, no. 1, pp. 47–52, 2010.
 - [12] X. Xiaoping, Q. Fucai, W. Feng, and L. Hongyan, “Identification method for Hammerstein models based on advanced particle swarm optimization algorithm,” *Computer Engineering*, vol. 34, no. 14, pp. 200–203, 2008.
 - [13] F. Chang and R. Luus, “A noniterative method for identification using Hammerstein model,” *IEEE Transactions on Automatic Control*, vol. 16, no. 5, pp. 464–468, 1971.
 - [14] C. Z. Fang and D. Y. Xiao, *Process Identification*, Tsinghua-University Press, 1988.
 - [15] W. Xiang and Z.-H. Chen, “New identification method of nonlinear systems based on Hammerstein models,” *Control Theory and Applications*, vol. 24, no. 1, pp. 143–147, 2007.
 - [16] B.-C. Xu and X.-L. Liu, “Identification algorithm based on the approximate least absolute deviation criteria,” *International Journal of Automation and Computing*, vol. 9, no. 5, pp. 501–505, 2012.
 - [17] D. Jian and X. Kaigui, “Research of the non-linear regress models based on the least absolute criteria,” *Journal of Chongqing Normal University (Natural Science Edition)*, vol. 18, no. 4, pp. 71–74, 2001.
 - [18] Y. Liu, Y. Xiao, and X. Zhao, “Multi-innovation stochastic gradient algorithm for multiple-input single-output systems using the auxiliary model,” *Applied Mathematics and Computation*, vol. 215, no. 4, pp. 1477–1483, 2009.
 - [19] D. Feng, L. Jingfan, and X. Yongsong, “Parameter estimation for a class of nonlinear systems,” *Control Engineering of China*, vol. 18, no. 3, pp. 373–376, 2011.
 - [20] F. Ding, X. P. Liu, and G. Liu, “Identification methods for Hammerstein nonlinear systems,” *Digital Signal Processing*, vol. 21, no. 2, pp. 215–238, 2011.
 - [21] F. Wei and D. Feng, “Three methods for Hammerstein models’ parameter separation,” *Science Technology and Engineering*, vol. 8, no. 6, pp. 1586–1589, 2008.
 - [22] E.-W. Bai, “A blind approach to the Hammerstein-Wiener model identification,” *Automatica*, vol. 38, no. 6, pp. 967–979, 2002.
 - [23] M. J. Coates and E. E. Kuruoğlu, “Time-frequency-based detection in impulsive noise environments using α -stable noise models,” *Signal Processing*, vol. 82, no. 12, pp. 1917–1925, 2002.
 - [24] L. Deroye, *Non-Uniform Random Variant Generation*, Springer, New York, NY, USA, 1986.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

