

---

**Research/Technical Note**

# Recursive Identification of Hammerstein Systems with Polynomial Function Approximation

Wang Jian-hong<sup>1</sup>, Tang De-zhi<sup>2</sup>, Jiang Hong<sup>1</sup>, Tang Xiao-jun<sup>1</sup>

<sup>1</sup>School of Electronic Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, China

<sup>2</sup>School of Electrical and Information Engineering, Anhui University of Technology, Ma-an-shan, China

**Email address:**

18279820758@163.com (Tang Xiao-jun)

**To cite this article:**

Wang Jian-hong, Tang De-zhi, Jiang Hong, Tang Xiao-jun. Recursive Identification of Hammerstein Systems with Polynomial Function Approximation. *International Journal of Management and Fuzzy Systems*. Vol. 3, No. 6, 2017, pp.87-94. doi: 10.11648/j.ijmfs.20170306.12

**Received:** September 25, 2017; **Accepted:** October 27, 2017; **Published:** November 20, 2017

---

**Abstract:** Nonlinear system identification is considered, where the nonlinear static function was approximated by a number of polynomial functions. It is based on a piecewise-linear Hammerstein model, which is linear in the parameters. The identification procedure is divided into two steps. Firstly we adopt the extended stochastic gradient algorithm to identify some unknown parameters. Secondly using singular value decomposition (SVD), we propose a new method to identify other parameters. The basic idea is to replace un-measurable noise terms in the information vectors by their estimates, and to compute the noise estimates based on the obtained parameter estimates. The applicability of the approach is illustrated by a simulation.

**Keywords:** Nonlinear System, Hammerstein Systems, Polynomial Functions Approximation, Recursive Identification, Singular Value Decomposition

---

## 1. Introduction

Modeling, identification and prediction are three main ubiquitous phenomena in our daily lives. Through our ideas and senses, we collect information about the world, then we interpret, predict and react actions according to our perceptions. In natural science, lots of experiments or observations guide us to formulate laws of nature, which describe different aspects of the world and let us predict all sorts of things, like planet movements or weather forecast. Also in modern technology, modeling and identification have much benefit to offer us one description corresponding to the physical object. Everywhere and everything around us, there is a need for automatic control mechanisms such as in aero-planes, cars, chemical process plants, mobiles phones, heating of houses etc. However to be able to control a system, one needs to know at least something about how it behaviors and reacts to different actions taken on it. Hence we need a model of the system. A system can informally be defined as an entity which interacts with the rest of the world through more or less well defined input and output data. A model is then an approximate description of the system. An ideal model should

be simple, accurate and general. This approximate description of the system can be constructed by system identification strategy, as the goal of system identification is to build a mathematical model of a dynamic system based on some initial information about the system and the measurement data collected from the system. According to [1], the process of system identification consists of designing and conducting the identification experiment in order to collect the measurement data, selecting the structure of the model and specifying the parameters to be identified and eventually fitting the model parameters to the obtained data [2]. Finally the quality of the obtained model is evaluated through model validation process. Generally system identification is an iterative process and if the quality of the obtained model is not satisfactory, some or all of the listed phases can be repeated in order to obtain one satisfied model.

Many nonlinear systems can be modeled by a Hammerstein model (linear time-invariant (LTI) block following some static nonlinear block), Wiener model (LTI block preceding some static nonlinear block), or Hammerstein-Wiener model (LTI block sandwiched by two nonlinear blocks) [3]. The Hammerstein model is a special kind of nonlinear systems

which has applications in many engineering problems and therefore, identification of Hammerstein models has been an active research topic for a long time. Existing methods in the literature can be roughly divided into six categories: the iterative method, the over-parameterization method, the stochastic method, the nonlinear least squares method, the separable least squares method and the blind method [4].

The Hammerstein model is a nonlinear system where a nonlinear block is followed by a linear dynamic block. The Hammerstein model has been widely used in many areas including nonlinear filtering, actuator saturations, audio-visual processes. There exists a large number of papers on the topic of the Hammerstein model identification. Reference [5] considered an extended stochastic gradient identification algorithm for Hammerstein-Wiener ARMAX systems. To improve the identification accuracy, an extended stochastic gradient algorithm with a forgetting factor is given. Reference [6] proposed three methods of separating the parameter estimates. The three methods are the averaged

method, permutation and combination method and singular value method. Reference [7] considered two identification algorithms, an iterative gradient and a recursive stochastic gradient based, for a Hammerstein nonlinear ARMAX model.

## 2. Problem Formulation

In this paper we focus on the identification of Hammerstein model which consists of a nonlinear memory-less element followed by a linear dynamical system [8]. The true output  $x(t)$  and the inner variable  $\bar{u}(t)$  are immeasurable,  $u(t)$  is the system input,  $y(t)$  is the measurement of  $x(t)$  but is corrupted by the disturbance  $w(t)$ , the output of  $N(z)$  driven by an additive white noise  $v(t)$  with zero mean,  $G(z)$  is the transfer function of the linear part in the model,  $N(z)$  is the transfer function of the noise model [10].

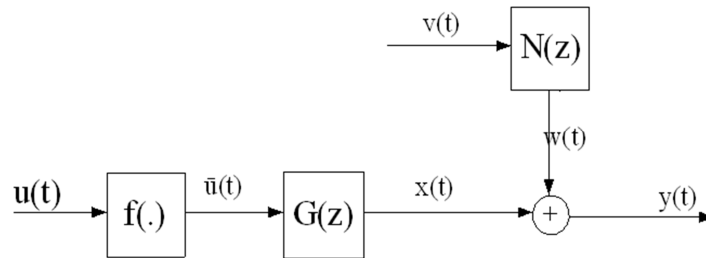


Figure 1. The discrete time SISO Hammerstein system.

Assume that the linear dynamical block in Figure 1 is described by an ARMAX model, which has the

$$\begin{aligned} y(t) &= x(t) + w(t) \\ x(t) &= G(z)\bar{u}(t) = \frac{B(z)}{A(z)}\bar{u}(t) \\ w(t) &= N(z)v(t) = \frac{D(z)}{A(z)}v(t) \end{aligned} \quad (1)$$

Here  $A(z)$ ,  $B(z)$  and  $D(z)$  are polynomials in the shift operator  $z^{-1}$   $[z^{-1}y(t) = y(t-1)]$  with

$$\begin{aligned} A(z) &= 1 + a_1z^{-1} + a_2z^{-2} + \dots + a_nz^{-n} \\ B(z) &= b_1z^{-1} + b_2z^{-2} + \dots + b_nz^{-n} \\ D(z) &= 1 + d_1z^{-1} + d_2z^{-2} + \dots + d_{n_d}z^{-n_d} \end{aligned} \quad (2)$$

Notice that in the characterization of the Hammerstein model,  $f(u)$  and  $G(z)$  are actually not unique. Any pair  $(af(u), G(z)/a)$  for some nonzero and finite constant  $a$  would produce identical input and output measurements. In other others, any identification scheme can not distinguish between  $(af(u), G(z)/a)$  and  $(f(u), G(z))$ . Therefore to

get an unique parameterization, one of the gains of  $f(u)$  and  $G(z)$  has to be fixed. There are several ways to normalize the gains.

A nonlinear static function can be approximated by a number of basis functions, connected to each other and forming a piecewise-linear function.

$$\bar{u}(t) = F(u)^T \cdot C \quad (3)$$

where  $C$  is a vector of parameters representing values  $c_j$  of the piecewise-linear function  $\bar{u}(t)$  in knots.

$$C = [c_0 \quad c_1 \quad \dots \quad c_j \quad \dots \quad c_m]^T_{(m+1) \times 1} \quad (4)$$

The knots are joints of line segments defined in terms of input signal  $u$ .

$$u = [u_0 \quad u_1 \quad \dots \quad u_j \quad \dots \quad u_m]^T_{(m+1) \times 1} \quad (5)$$

satisfying

$$u_0 < u_1 < \dots < u_j < u_{j+1} < \dots < u_m$$

Furthermore  $F(u)$  is a vector of polynomial function.

$$F(u) = \begin{bmatrix} F_0(u) & F_1(u) & \cdots & F_j(u) & \cdots & F_m(u) \end{bmatrix}_{(m+1) \times 1}^T$$

Element of the vector  $F(u)$  are defined as follows:

$$F_j(u) = u \prod_{\substack{i=0 \\ i \neq j}}^m \frac{u - u_i}{u_j - u_i}$$

substituting into equation (3), then the nonlinear part in the Hammerstein model is in the following form:

$$\bar{u}(t) = f(u) = u \sum_{j=0}^m c_j f_j(u) \quad (6)$$

with

$$f_j(u) = \prod_{\substack{i=0 \\ i \neq j}}^m \frac{u - u_i}{u_j - u_i}$$

It can be easily seen that:

$$f_j(u_i) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (7)$$

$$c_j = \frac{f(u_j)}{u_j} \quad (8)$$

Therefore the plant (1) can also be represented by the following model:

$$\begin{aligned} A(z)y(t) &= B(z)\bar{u}(t) + D(z)v(t) \\ \bar{u}(t) &= f(u) = u \sum_{j=0}^m c_j f_j(u) \\ f_j(u) &= \prod_{\substack{i=0 \\ i \neq j}}^m \frac{u - u_i}{u_j - u_i} \end{aligned} \quad (9)$$

A regressive form can be derived as follows:

$$A(z)y(t) = B(z)\bar{u}(t) + D(z)v(t) = \sum_{i=1}^n b_i u(t-i) \sum_{j=0}^m c_j f_j(u(t-i)) + D(z)v(t)$$

From the last equality, we can easily get the following recursive equation:

$$\begin{aligned} y(t) &= -\sum_{i=1}^n a_i y(t-i) + \sum_{i=1}^n b_i u(t-i) \sum_{j=0}^m c_j f_j(u(t-i)) + \sum_{i=1}^{n_d} d_i v(t-i) + v(t) \\ y(t) &= -\sum_{i=1}^n a_i y(t-i) + \sum_{i=1}^n \sum_{j=0}^m \mu_{ij} u(t-i) f_j(u(t-i)) + \sum_{i=1}^{n_d} d_i v(t-i) + v(t) \end{aligned}$$

where

$$\mu_{ij} = b_i c_j \quad i = 1 \cdots n, j = 0, 1 \cdots m \quad (10)$$

Define the parameter vector  $\theta$  and information vector  $\phi(t)$  as

$$\theta = \begin{bmatrix} a \\ \mu_0 \\ \mu_1 \\ \vdots \\ \mu_m \\ d \end{bmatrix} \in R^{n_0}, \phi(t) = \begin{bmatrix} \varphi(t) \\ v(t-1) \\ \vdots \\ v(t-n_d) \end{bmatrix} \in R^{n_0} \quad (11)$$

$$n_0 = (m+2)n + n_d$$

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \in R^n, \mu_0 = \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \vdots \\ \mu_{n0} \end{bmatrix} \in R^n$$

$$\mu_1 = \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{n1} \end{bmatrix} \in R^n \cdots \mu_m = \begin{bmatrix} \mu_{1m} \\ \mu_{2m} \\ \vdots \\ \mu_{nm} \end{bmatrix} \in R^n$$

$$d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{n_d} \end{bmatrix} \in R^{n_d}, \varphi(t) = \begin{bmatrix} \varphi_a(t) \\ \varphi_0(t) \\ \varphi_1(t) \\ \vdots \\ \varphi_m(t) \end{bmatrix} \in R^{(m+2)n}$$

$$\varphi_a(t) = \begin{bmatrix} -y(t-1) \\ -y(t-2) \\ \vdots \\ -y(t-n) \end{bmatrix} \in R^n \quad (12)$$

$$\varphi_j(t) = \begin{bmatrix} u(t-1)f_j(u(t-1)) \\ u(t-2)f_j(u(t-2)) \\ \vdots \\ u(t-n)f_j(u(t-n)) \end{bmatrix} \in R^n, j = 0, 1 \cdots m$$

Then we have

$$y(t) = \phi(t)^T \theta + v(t)$$

Notice that  $\varphi(t)$  in  $\phi(t)$  is available but  $v(t-i) (i = 1, 2 \cdots n_d)$  in  $\phi(t)$  are unavailable. Let  $\hat{\theta}$  denotes the estimate of  $\theta$ . Since  $v(t)$  is a white noise with zero mean, then

$$\hat{y}(t) = \phi(t)^T \hat{\theta} \quad (13)$$

It is the best output prediction. Consider the quadratic output prediction error criterion

$$J(\hat{\theta}) = \sum_{i=1}^t [y(i) - \hat{y}(i)]^2 = \sum_{i=1}^t [y(i) - \phi(t)^T \hat{\theta}]^2 \quad (14)$$

The quadratic error function in (14) is one of the most common cost functions in the identification literature [9]. Many well-known Hammerstein model identification methods belong to this class and the differences lie only in the formulation of the information vector  $\phi(t)$ . Let

$$Y(t) = \begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(1) \end{bmatrix}, \Phi(t) = \begin{bmatrix} \phi^T(t) \\ \phi^T(t-1) \\ \vdots \\ \phi^T(1) \end{bmatrix} \quad (15)$$

Hence

$$J(\hat{\theta}) = [Y(t) - \Phi(t)\hat{\theta}]^T [Y(t) - \Phi(t)\hat{\theta}] = \|Y(t) - \Phi(t)\hat{\theta}\|^2$$

Provided that  $\Phi(t)$  is persistently exciting, minimizing  $J(\hat{\theta})$  gives the least-squares estimate:

$$\hat{\theta} = [\Phi^T(t)\Phi(t)]^{-1} \Phi^T(t)Y(t) \quad (16)$$

However a difficulty arises because of  $v(t-i)$ ,  $\Phi(t)$  in the expression on the right-hand side of (16) contains unknown noise terms  $v(t-i)$  ( $i=1, 2, \dots, n_d$ ), so it is impossible to compute the estimate  $\hat{\theta}$  by (16), our approach is based on the iterative identification principle [10]. Let  $k=1, 2, \dots$  the unknown variables  $v(t-i)$  are replaced by their corresponding estimate  $\hat{v}_k(t-i)$  at iteration  $k$ , and  $\phi(t)$  are replaced by  $\hat{\phi}_k(t)$  [11]. Let  $\hat{\theta}_k$  be the iterative solution of  $\theta$ . Thus the estimate of  $v(t)$  is given by

$$\hat{v}_k(t-i) = y(t-i) - \hat{\phi}_{k-1}^T(t-i)\hat{\theta}_{k-1} \quad (17)$$

$$\hat{\phi}_k(t) = \begin{bmatrix} \phi(t) \\ \hat{v}_k(t-1) \\ \vdots \\ \hat{v}_k(t-n_d) \end{bmatrix} \in R^{n_0} \quad (18)$$

$$\Phi_k(t) = \begin{bmatrix} \hat{\phi}_k^T(t) \\ \hat{\phi}_k^T(t-1) \\ \vdots \\ \hat{\phi}_k^T(1) \end{bmatrix} \quad (19)$$

Based on (16) and replace  $\Phi(t)$  by  $\Phi_k(t)$ , the iterative solution  $\hat{\theta}_k$  of  $\theta$  may be computed by:

$$\hat{\theta}_k = [\Phi_k^T(t)\Phi_k(t)]^{-1} \Phi_k^T(t)Y(t), k=1, 2, \dots \quad (20)$$

To initialize the algorithm, we take  $\hat{\theta}_0 = 0$  or some small real vector, e.g.  $\hat{\theta}_0 = 10^{-6} I_{n_0}$  with  $I_{n_0}$  being an  $n_0$ -dimensional column vector, whose elements are 1. As  $\theta$  comes in linearly, algorithm (20) turns out to be adequate parameterization to get estimates of the parameters  $a_i, \mu_{ij}$  and  $d_i$ , using the iterative least squares algorithm.

### 3. Basic Formulations for Plant Model Identification

From estimates  $\hat{\theta}(t)$  of  $\theta$ , one has to get estimates  $(\hat{b}_i, \hat{c}_j)$  of  $(b_i, c_j)$  for  $i=1, \dots, n, j=0, 1, \dots, m$ . We will first construct a procedure to go back from  $\mu_{ij}$ 's to  $b_i$ 's and  $c_i$ 's. Then relations to get  $(\hat{b}_i, \hat{c}_j)$  from  $\mu_{ij}$  will be established.

Observe that (10) can be rewritten as follows:

$$M = \begin{bmatrix} \mu_{10} & \cdots & \mu_{1m} \\ \vdots & \vdots & \vdots \\ \mu_{n0} & \cdots & \mu_{nm} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \begin{bmatrix} c_0 & c_1 & \cdots & c_m \end{bmatrix} \quad (21)$$

Notice that since  $M$  is a rank-1 matrix,  $b_i$ 's and  $c_i$ 's can not be determined uniquely from  $\mu_{ij}$ 's, unless extra conditions are imposed on  $b_i$ 's and  $c_i$ 's. Uniqueness of the solution of (21) can be achieved by imposing the following couple of conditions:

$$\sum_{i=1}^n b_i^2 = 1 \text{ and } \rho([b_1 \cdots b_n]) > 0 \quad (22)$$

where  $\rho([b_1 \cdots b_n])$  denotes the first component of the vector  $[b_1 \cdots b_n]^T$  that satisfies:

$$|\rho([b_1 \cdots b_n])| = \sup_{1 \leq j \leq n} |b_j|$$

i.e the first component with a great absolute value.

Based on the above observations [11], a procedure is designed using singular value decomposition (SVD) [10].

Proposition 2.1. let  $M \in R^{n \times (m+1)}$  be any rank-1 real matrix. Then its SVD decomposition has the following form:

$$M = \Gamma \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \Sigma \quad (23)$$

where  $\Gamma \in R^{n \times n}$ ,  $\Sigma \in R^{(m+1) \times (m+1)}$  and  $\sigma_1$  is the unique

nonzero singular value of  $M$ . Furthermore  $M$  can be uniquely decomposed as follows:

$$M = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \begin{bmatrix} c_0 & c_1 & \cdots & c_m \end{bmatrix} \quad (24)$$

with

$$\begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix}^T = \text{sign}(r) \frac{\Gamma \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \end{bmatrix}^T}{\left\| \Gamma \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \end{bmatrix}^T \right\|} \quad (25)$$

$$\begin{bmatrix} c_0 & c_1 & \cdots & c_m \end{bmatrix}^T = \text{sign}(r) \left\| \Gamma \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \end{bmatrix}^T \right\| \times \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \Sigma$$

where

$$r = \rho(\Gamma \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \end{bmatrix})$$

The vector  $\begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix}^T$  thus obtained is the only solution of (23) that satisfies:

$$\sum_{i=1}^n b_i^2 = 1 \quad \text{and} \quad \rho(\begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix}) > 0$$

Estimates  $(\hat{b}_i, \hat{c}_j)$  of  $(b_i, c_j)$  can be recovered from  $\mu_{ij}$ 's.

Following closely proposition 2.1[12], one first considers the matrix:

$$\hat{M}(t) = \begin{bmatrix} \hat{\mu}_{10}(t) & \cdots & \hat{\mu}_{1m}(t) \\ \vdots & \vdots & \vdots \\ \hat{\mu}_{n0}(t) & \cdots & \hat{\mu}_{nm}(t) \end{bmatrix} \quad (26)$$

Proposition 2.2: let the singular values decomposition of  $\hat{M}(t)$  be written as follows:

$$\hat{M}(t) = \Gamma(t) \begin{bmatrix} \sigma_1(t) & 0 & \cdots & 0 \\ 0 & \sigma_2(t) & & \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_n(t) \end{bmatrix} \Sigma(t)$$

Then proposition 2.1 suggests the following estimates for the parameters  $(b_i, c_i)$

$$\begin{bmatrix} \hat{b}_1(t) \\ \vdots \\ \hat{b}_n(t) \end{bmatrix} = \text{sign}(r(t)) \frac{\Gamma(t) \begin{bmatrix} \sigma_1(t) & 0 & \cdots & 0 \end{bmatrix}^T}{\left\| \Gamma(t) \begin{bmatrix} \sigma_1(t) & 0 & \cdots & 0 \end{bmatrix}^T \right\|} \quad (27)$$

$$\begin{bmatrix} \hat{c}_0(t) \\ \hat{c}_1(t) \\ \vdots \\ \hat{c}_m(t) \end{bmatrix} = \text{sign}(r(t)) \left\| \Gamma(t) \begin{bmatrix} \sigma_1(t) & 0 & \cdots & 0 \end{bmatrix}^T \right\| \times \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \Sigma(t) \quad (28)$$

$$r(t) = \rho(\Gamma(t) \begin{bmatrix} \sigma_1(t) & 0 & \cdots & 0 \end{bmatrix}^T) \quad (29)$$

The estimates thus obtained are the only ones that satisfy the conditions

$$\sum_{i=1}^n b_i^2(t) = 1$$

$$\rho(\begin{bmatrix} \hat{b}_1(t) & \cdots & \hat{b}_n(t) \end{bmatrix}) > 0$$

Notice that the singular values  $\sigma_2(t) \cdots \sigma_n(t)$  have not been accounted for in the rules (27), (28) and (29). It has no effect when the  $\mu_{ij}$ 's converge to their true values, because the rank of matrix  $\hat{M}(t)$  converges to 1. This is made precisely in the next proposition.

Proposition 2.3: let  $\{\hat{M}(t)\}$  be the real matrix sequence defined by (26). Let  $\begin{bmatrix} \hat{b}_1(t) & \cdots & \hat{b}_n(t) \end{bmatrix}$  and  $\begin{bmatrix} \hat{c}_0(t) & \cdots & \hat{c}_m(t) \end{bmatrix}$  be the vectors obtained from  $\hat{M}(t)$  according to the rules (27), (28) and (29). If the  $\mu_{ij}$ 's converge to their true values  $\mu_{ij}$ , then the estimates  $(\hat{b}_i(t), \hat{c}_j(t))$  will converge to their true values  $(b_i, c_j)$ .

## 4. Recursive Identification Algorithm

In this section, we derive a recursive identification algorithm which can be on-line implemented. We rewrite the equation.

$$y(t) = \phi(t)^T \theta + v(t) \quad (30)$$

Equation.(27) is a pseudo-linear regression identification model for the Hammerstein system. Note that  $\phi(t)$  in  $\phi(t)$  is available but  $v(t-i)$ ,  $i = 1, 2, \dots, n_d$  in  $\phi(t)$  are unavailable. Let  $E$  denote the expectation operator,  $\hat{\theta}(t)$  the estimate of  $\theta$  at time  $t$ ,  $\|x\| = \text{tr}[xx^T]$  the norm of the matrix  $x$ . Since  $v(t)$  is a white noise, forming a quadratic cost function

$$J(\theta) = E \left[ \|y(t) - \phi(t)\theta\|^2 \right]$$

minimizing  $J(\theta)$  leads to the following stochastic gradient algorithm of estimating  $\theta$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\phi(t)}{r(t)} \left[ y(t) - \phi(t)^T \hat{\theta}(t-1) \right] \quad (31)$$

$$r(t) = r(t-1) + \|\phi(t)\|^2, \quad r(0) = 1 \quad (32)$$

However the algorithm in (31) and (32) is impossible to realize because the information vector  $\phi(t)$  on the right-hand side contains unknown noise terms  $v(t-k)$ . The solution is to replace the unknown variables  $v(t-i)$  with their corresponding estimates  $\hat{v}(t-k)$ , and further define:

$$\hat{\phi}(t) = [\varphi(t) \quad \hat{v}(t-1) \quad \dots \quad \hat{v}(t-n_d)]^T \quad (33)$$

From (33), we have

$$v(t) = y(t) - \phi(t)^T \theta$$

replacing  $\phi(t)$  and  $\theta$  in the above equation with  $\hat{\phi}(t)$  and  $\hat{\theta}(t)$ , the estimated residual can be computed by

$$\hat{v}(t) = y(t) - \hat{\phi}(t)^T \hat{\theta}(t) \quad (34)$$

replacing  $\phi(t)$  in (31) and (32) with  $\hat{\phi}(t)$ , we can obtain the extended stochastic gradient [ESG] identification algorithm of estimating  $\theta$

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) + \frac{\hat{\phi}(t)}{r(t)} [y(t) - \hat{\phi}(t)^T \hat{\theta}(t-1)] \\ r(t) &= r(t-1) + \|\hat{\phi}(t)\|^2, r(0) = 1 \end{aligned} \quad (35)$$

$$\hat{\phi}(t) = \begin{bmatrix} \varphi(t) \\ \hat{v}(t-1) \\ \vdots \\ \hat{v}(t-n_d) \end{bmatrix} \quad \hat{\theta}(t) = \begin{bmatrix} \hat{a}(t) \\ \hat{\mu}_0(t) \\ \vdots \\ \hat{\mu}_m(t) \\ \hat{d}(t) \end{bmatrix} \quad (36)$$

To initialize this ESG algorithm  $\hat{\theta}(0)$  is generally taken to be some small real vector, e.g.  $\hat{\theta}_0 = 10^{-6} I_{n_0}$  with  $I_{n_0}$  being an  $n_0$ -dimensional column vector whose elements are 1.

The ESG algorithm has low computation, but its convergence is relatively slow. In order to improve the tracking performance of the ESG algorithm, we introduce a forgetting factor  $\lambda$  in the ESG algorithm to get the ESG algorithm with a forgetting factor, which is referred to the EFG algorithm.

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) + \frac{\hat{\phi}(t)}{r(t)} [y(t) - \hat{\phi}(t)^T \hat{\theta}(t-1)] \\ r(t) &= \lambda r(t-1) + \|\hat{\phi}(t)\|^2, r(0) = 1, 0 < \lambda < 1 \end{aligned} \quad (37)$$

when  $\lambda = 1$ , the EFG algorithm reduces to the ESG algorithm.

## 5. Simulation

A simulation is given to demonstrate the effectiveness of the proposed algorithms.

Consider the following system:

$$A(z)y(t) = B(z)\bar{u}(t) + D(z)v(t)$$

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} = 1 - 1.60z^{-1} + 0.80z^{-2}$$

$$B(z) = b_1 z^{-1} + b_2 z^{-2} = 0.80z^{-1} + 0.60z^{-2}$$

where

$$0.80^2 + 0.60^2 = 1$$

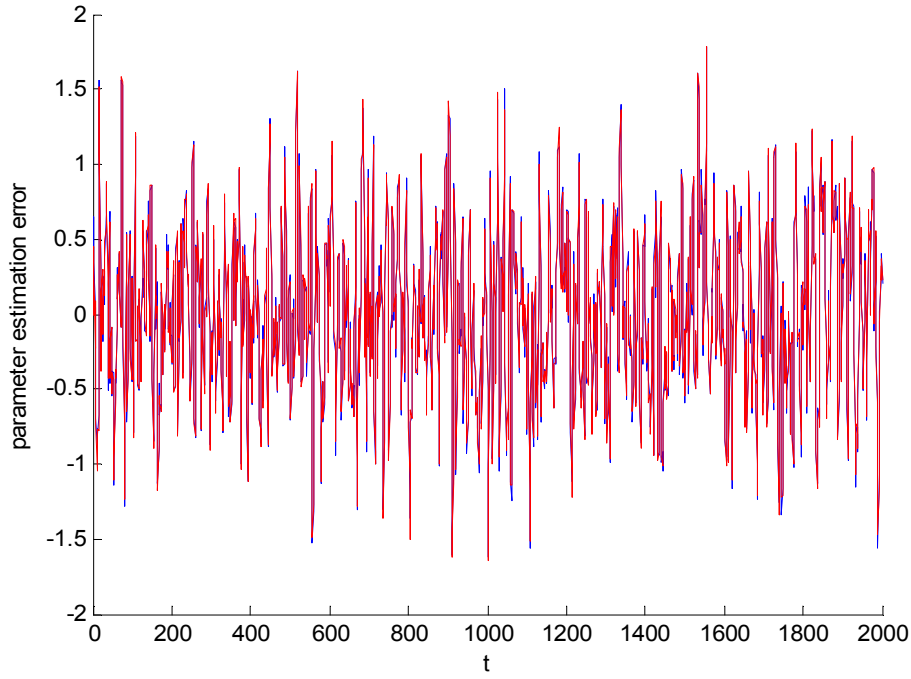
$$D(z) = 1 + d_1 z^{-1} = 1 - 0.64z^{-1}$$

$$\begin{aligned} \bar{u}(t) &= u [c_0 f_0(u) + c_1 f_1(u) + c_2 f_2(u)] \\ &= u [0.20 f_0(u) + 0.40 f_1(u) + 0.60 f_2(u)] \end{aligned}$$

where

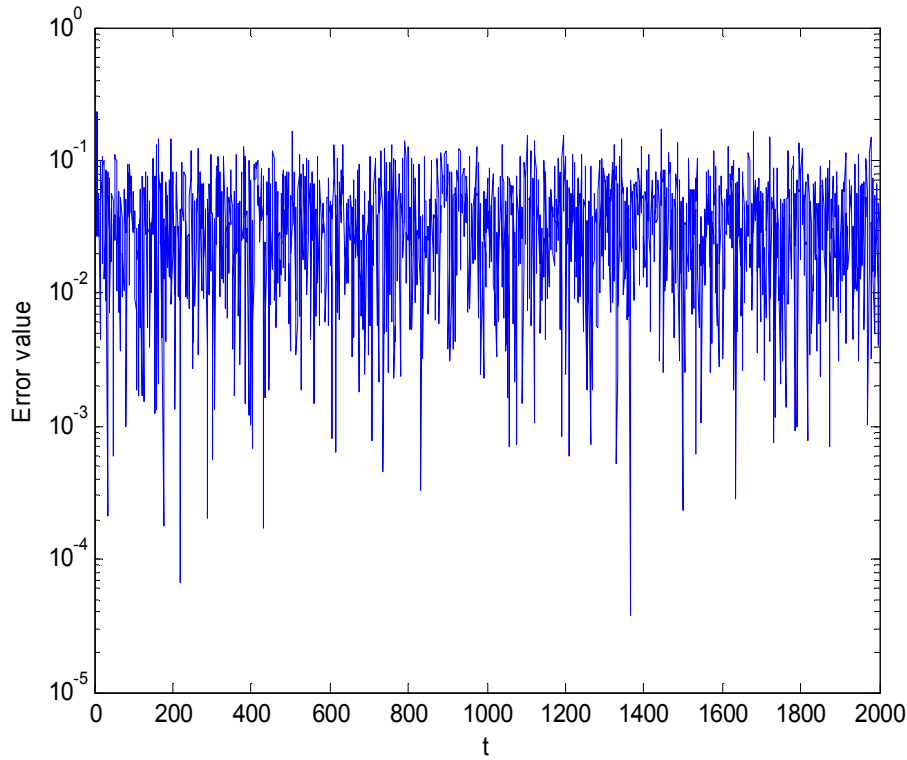
$$F_j(u) = u \prod_{\substack{i=0 \\ i \neq j}}^2 \frac{u - u_i}{u_j - u_i}$$

$\theta = [-1.60, 0.80, 0.16, 0.24, 0.32, 0.24, 0.48, 0.36, -0.61]^T$   
 $\{u(t)\}$  is taken as a persistent excitation signal sequence with zero mean and unit variance  $\sigma_u^2 = 1.00^2$ , and  $\{v(t)\}$  as a white noise sequence with zero mean and constant variance  $\sigma_v^2$ . Apply the proposed algorithms in (28)-(34) and (25)-(26) respectively, to estimate the parameters of this system. The parameter errors  $\delta$  versus  $t$  are shown in Figure 2, where the data length is 2000, the forgetting factor  $\lambda = 0.98$ , the absolute parameter estimation error  $\delta = \|\hat{\theta}(t) - \theta(t)\|$ , and the noise-to-signal ratio  $\delta_{ns} = 27.35\%$ . When changing  $\lambda = 0.90$ , the parameter error  $\delta$  versus  $t$  is shown in Figure 3.



Solid line: the ESG algorithm; dots: the EFG algorithm. ( $\lambda = 0.98$ )

**Figure 2.** The parameter estimation error  $\delta$  versus  $t$ .



**Figure 3.** The parameter estimation error  $\delta$  versus  $t$  of the EFG algorithm. ( $\lambda = 0.90$ ).

## 6. Conclusion

We have considered system identification based on Hammerstein model, where the nonlinear element is defined by a piecewise-linear function. We have designed an identification scheme that determines precisely the unknown nonlinear parameters  $(b_i, c_j)$   $i = 1 \dots n, j = 0, 1 \dots m$  and

those of the linear transfer function. An iterative algorithm based on replacing un-measurable noise variables by their estimates are derived for Hammerstein nonlinear models.

## Acknowledgements

This work was supported by the Grants from the National Science Foundation of China (no. 61563022), the Jiangxi

Provincial National Science Foundation (no. 20142BAB206020, no. GJJ150889), and the National Science Foundation of the Colleges and University in Anhui Provincial (no.KJ2016A094).

## References

- [1] Anna Hagenblad, Lennart Ljung, Adrian Wills. Maximum likelihood identification of Wiener models[J]. Automatica, 2008.44(11):2697-2705.
- [2] Lennart Ljung. System Identification: Theory for the user [M], Prentice-Hall, Upper Saddle River, 1999.
- [3] Martin Enqvist, Lennart Ljung. Linear approximations of nonlinear FIR systems for separable input processes. Automatica [J], 2005. 41(3):459-473.
- [4] J. J. Bussgang. Cross correlate on functions of amplitude-distorted Gaussian signals. Technical Report Technical report 216, MIT Laboratory of Electronics, 1952.
- [5] Bai E-W. Frequency domain identification of Hammerstein models[J]. IEEE transactions on automatic control. 2003. 48(4):530-541.
- [6] Bai E-W. A random least-trimmed-squares identification algorithm [J]. Automatica. 2003. 39(9): 1651-1659.
- [7] Bai E-W. Identification of linear systems with hard input nonlinearities of known structure [J]. Automatica, 2002. 38(5): 853-860.
- [8] Lennart Ljung. Estimating linear time invariant models of non-linear time-varying system [J]. European Journal of Control, 2001. 7 (2):203-219.
- [9] R. Pintelon, J. Schoukens. Fast approximation identification of nonlinear systems [J]. Automatica, 2003. 39(7):1267-1273.
- [10] Ding F, Tongwen Chen. Identification of Hammerstein nonlinear ARMAX systems [J]. Automatica, 2005. 41 (9):1479-1489.
- [11] Ding F, Tongwen Chen. Performance analysis of multi-innovation gradient type identification methods [J]. Automatica, 2007. 43(1):1-14.
- [12] Wang DQ, Ding F. Extended stochastic gradient identification algorithms for Hammerstein-Wiener ARMAX systems [J]. Computers & Mathematics with Applications, 2008. 56(12): 3157-3164.