

# Combined state and parameter estimation for Hammerstein systems with time-delay using the Kalman filtering

Junxia Ma<sup>1</sup>, Feng Ding<sup>1\*†</sup>, Weili Xiong<sup>1</sup> and Erfu Yang<sup>2</sup>

<sup>1</sup>*Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, PR China*

<sup>2</sup>*Department of Design, Manufacture and Engineering Management, Space Mechatronic Systems Technology Laboratory, Strathclyde Space Institute, University of Strathclyde, Glasgow G1 1XJ, Scotland, United Kingdom*

## SUMMARY

This paper discusses the state and parameter estimation problem for a class of Hammerstein state space systems with time-delay. Both the process noise and the measurement noise are considered in the system. Based on the observable canonical state space form and the key term separation, a pseudo-linear regressive identification model is obtained. For the unknown states in the information vector, the Kalman filter is used to search for the optimal state estimates. A Kalman-filter based least squares iterative and a recursive least squares algorithms are proposed. Extending the information vector to include the latest information terms which are missed for the time-delay, the Kalman-filter based recursive extended least squares algorithm is derived to obtain the estimates of the unknown time-delay, parameters and states. The numerical simulation results are given to illustrate the effectiveness of the proposed algorithms. Copyright © 2016 John Wiley & Sons, Ltd.

Received . . .

**KEY WORDS:** Parameter identification; Kalman filter; State estimation; Least squares; Hammerstein state space model

## 1. INTRODUCTION

The Hammerstein model consists of an input static nonlinear block in series with a dynamic linear subsystem and can describe a wide variety of practical nonlinear systems, e.g., wind turbines [1], valve actuators [2] and power amplifiers [3]. Despite its simplicity, as a block-oriented nonlinear system, the Hammerstein model can include many different kinds of components in the nonlinear and linear blocks. The memoryless nonlinearities include polynomial, piecewise linear descriptions, saturation, preload, deadzone, backlash and so on [4, 5]. The linear subsystems may be some parametric forms, such as state space representations, transfer functions, FIR, IIR and so on [6, 7].

Much research work has been performed on the identification of Hammerstein models. For example, Wang and Zhang studied an improved least squares identification algorithm for multivariable Hammerstein output error moving average systems by using the Taylor expansion on a quadratic criterion function and defining the information vector as the derivative of the noise variable to the unified parameter vector [8]; Chen et al derived a hierarchical gradient parameter estimation algorithm for Hammerstein nonlinear systems using the key term separation principle [9];

\*Correspondence to: F. Ding, Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, PR China

†E-mail: fding@jiangnan.edu.cn

Liu and Bai presented a normalized iterative identification algorithm for Hammerstein systems which nonlinearities are with nonsmooth piecewise-linear structures [10].

In practical applications, disturbances widely exist [11, 12]. Especially in the area of signal processing and parameter estimation [13], the measured output always contain the disturbances from process environments [14]. The disturbance can be white noise or colored noise. Both the process noise and the measurement noise will bring some influence to the result of the systems to be identified. For the parameter estimation, the measurement noise has been widely discussed in the literatures [15, 16, 17], but the process noise is few considered in the model structure. For example, Hu et al. studied a parameter estimation problem for a Wiener system which is disturbed by the moving average measurement noise and derived two recursive extended least squares parameter estimation algorithms based on the over-parameterization models [18]; Li derived a maximum likelihood parameter estimation algorithm for Hammerstein systems which are disturbed by autoregressive moving average noise [19].

The state space models can describe dynamic linear and nonlinear systems [20, 21, 22] and play an important part in system identification [23] and signal filtering [24, 25]. The identification of the state space models has received much attention [26]. Chen et al. discussed the parameter and state estimation problem of a single-input single-output dual-rate system with time-delay based on the gradient search and the least squares principle [27]; Xie and Yang derived a gradient-based iterative identification algorithm for nonuniform sampling state space models [28]. In the field of nonlinear state space models, Schön et al. derived an expectation maximization (EM) algorithm under the framework of a maximum likelihood for the parameter estimation of a class of nonlinear state space dynamic systems [29]; Deng and Huang studied the identification problem of nonlinear parameter varying state space models with missing output data by using the particle filter to compute the expectation functions under the scheme of the EM algorithm [30].

For the input nonlinear state space systems, Wang and Ding derived an over-parameterization model based stochastic gradient algorithm to obtain the parameter estimates, but they did not consider the process noise and the time-delay in the model structure [31]. On the basis of the work in [31], this paper investigates the state and parameter estimation problem for the input nonlinear Hammerstein systems with time-delay. The difficulties are that the system not only contains the unknown parameters but also the unknown system states and the time-delay. The Kalman-filter based least squares iterative (LSI) algorithm and recursive least squares (RLS) algorithm are derived for the combined estimation of the state and parameter. For the unknown time-delay, a Kalman-filter based recursive extended least squares (KF-RELS) algorithm is proposed by extending the information vector and the parameter vector. The proposed algorithms are different from the least squares algorithm in [32], which decomposes the bilinear cost function into three linear functions by using the hierarchical identification principle and uses the state observer to get the estimates of the unknown states. Also, the proposed algorithms are different from the over-parameterization based recursive least squares algorithm in [33], which ignores the process noise in the model structure and the influence of the measurement noise in the process of updating the estimates of the system states. The main contributions of this paper are as follows.

- A more common model structure is considered which contains both process noise and measurement noise. By using the key term separation technique, the output of the system is expressed as a linear combination of all the unknown parameters. Then a pseudo-linear regressive identification model is obtained;
- For the known and unknown time-delay, a recursive least squares algorithm and a recursive extended least squares algorithm are derived to identify the unknown parameters;
- By using the Kalman filter, a joint state and parameter estimation algorithm is presented to obtain the estimates of the unknown parameters both in the linear subsystem and in the memoryless nonlinear block as well as the unmeasured system states.

The rest of this paper is organized as follows. Section 2 gives the identification model of Hammerstein systems with dynamic state space subsystem. Section 3 derives the Kalman-filter based LSI and RELS identification algorithms for a class of Hammerstein nonlinear systems. The

KF-RELS algorithm is proposed in Section 4. Section 5 provides two numerical examples to show the effectiveness of the proposed algorithms. Finally, Section 6 offers some concluding remarks.

## 2. SYSTEM DESCRIPTION AND IDENTIFICATION MODEL

Let us define some notation. “ $A =: X$ ” or “ $X := A$ ” stands for “ $A$  is defined as  $X$ ”;  $\hat{x}_k$  denotes the estimate of  $x$  at time  $k$ ;  $\hat{x}_k^s$  denotes the estimate of  $x_k$  at iteration  $s$ ; the symbol  $I$  ( $I_n$ ) stands for an identity matrix of appropriate sizes ( $n \times n$ ); the superscript T denotes the matrix/vector transpose;  $\mathbf{1}_n$  represents an  $n$ -dimensional column vector whose elements are 1.

Consider the following Hammerstein nonlinear system as shown in Figure 1, which is composed of a static nonlinear block followed by a linear dynamic subsystem. The linear subsystem is described by a state space model with moving average measurement noise. The time-delay is considered in the output block.

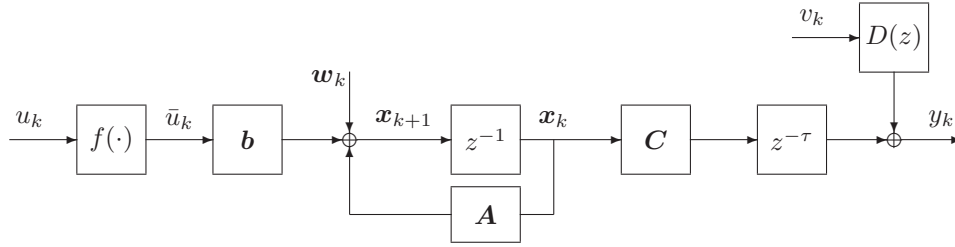


Figure 1. The Hammerstein state space model with time-delay

The Hammerstein state space model can be written as

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{b}\bar{u}_k + \mathbf{w}_k, \quad (1)$$

$$y_k = \mathbf{c}\mathbf{x}_{k-\tau} + \sum_{i=1}^{n_d} d_i v_{k-i} + v_k, \quad (2)$$

where  $y_k$  is the measured output,  $\bar{u}_k$  is the output of the static nonlinear block,  $\mathbf{w}_k := [w_{1,k}, w_{2,k}, \dots, w_{n,k}]^T \in \mathbb{R}^n$  and  $v_k$  represent mutually independent process noise and measurement noise with zero mean and variance  $\mathbf{Q}$  and  $R$ , respectively.  $\mathbf{x}_k := [x_{1,k}, x_{2,k}, \dots, x_{n,k}]^T \in \mathbb{R}^n$  is the system state vector.  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is the system parameter matrix,  $\mathbf{b} \in \mathbb{R}^n$  and  $\mathbf{c} \in \mathbb{R}^{1 \times n}$  are the parameter vectors,

$$\mathbf{A} := \begin{bmatrix} -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ -a_{n-1} & 0 & \cdots & 0 & 1 \\ -a_n & 0 & \cdots & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad \mathbf{b} := \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} \in \mathbb{R}^n,$$

$$\mathbf{c} := [1, 0, 0, \dots, 0] \in \mathbb{R}^{1 \times n}.$$

Assume that the output of the nonlinear block is a linear function of the known basis:

$$\bar{u}_k = \gamma_1 f_1(u_k) + \gamma_2 f_2(u_k) + \cdots + \gamma_m f_m(u_k) = \sum_{j=1}^m \gamma_j f_j(u_k). \quad (3)$$

where  $\gamma_i$ 's are the unknown coefficients and  $f_i(u_k)$ 's are the base functions. Let

$$\boldsymbol{\gamma} := [\gamma_1, \gamma_2, \dots, \gamma_m]^T \in \mathbb{R}^m,$$

$$\mathbf{f}(u_k) := [f_1(u_k), f_2(u_k), \dots, f_m(u_k)] \in \mathbb{R}^{1 \times m}.$$

From (1), we have

$$x_{1,k+1} = -a_1 x_{1,k} + x_{2,k} + b_1 \bar{u}_k + w_{1,k}, \quad (4)$$

$$x_{2,k+1} = -a_2 x_{1,k} + x_{3,k} + b_2 \bar{u}_k + w_{2,k},$$

$$\vdots$$

$$x_{i,k+1} = -a_i x_{1,k} + x_{i+1,k} + b_i \bar{u}_k + w_{i,k}, \quad (5)$$

$$\vdots$$

$$x_{n,k+1} = -a_n x_{1,k} + b_n \bar{u}_k + w_{n,k}.$$

Multiplying  $z^{-j-1}$  on both sides of (5) gives

$$x_{i,k-j} = -a_i x_{1,k-j-1} + x_{i+1,k-j-1} + b_i \bar{u}_{k-j-1} + w_{i,k-j-1}, \quad j = 0, 1, 2, \dots, n. \quad (6)$$

Substituting (6) into (4), we have

$$\begin{aligned} x_{1,k+1} &= -a_1 x_{1,k} - a_2 x_{1,k-1} + x_{3,k-1} + b_2 \bar{u}_{k-1} + b_1 \bar{u}_k + w_{2,k-1} + w_{1,k} \\ &= -a_1 x_{1,k} - a_2 x_{1,k-1} - a_3 x_{1,k-2} + x_{4,k-2} + b_3 \bar{u}_{k-2} + b_2 \bar{u}_{k-1} + b_1 \bar{u}_k + w_{3,k-2} \\ &\quad + w_{2,k-1} + w_{1,k} \\ &= -a_1 x_{1,k} - a_2 x_{1,k-1} - \dots - a_n x_{1,k-n+1} + b_n \bar{u}_{k-n+1} + \dots + b_2 \bar{u}_{k-1} + b_1 \bar{u}_k \\ &\quad + w_{n,k-n+1} + \dots + w_{1,k} \\ &= -\sum_{i=1}^n a_i x_{1,k-i+1} + \sum_{i=1}^n b_i \sum_{j=1}^m \gamma_j f_j(u_{k-i+1}) + \sum_{i=1}^n w_{i,k-i+1}. \end{aligned} \quad (7)$$

Multiplying both sides of (7) by  $z^{-\tau-1}$  gives

$$x_{1,k-\tau} = -\sum_{i=1}^n a_i x_{1,k-\tau-i} + \sum_{i=1}^n b_i \sum_{j=1}^m \gamma_j f_j(u_{k-\tau-i}) + \sum_{i=1}^n w_{i,k-\tau-i}. \quad (8)$$

Substituting (8) into (2) gives

$$\begin{aligned} y_k &= x_{1,k-\tau} + D(z)v_k \\ &= -\sum_{i=1}^n a_i x_{1,k-\tau-i} + \sum_{i=1}^n b_i \sum_{j=1}^m \gamma_j f_j(u_{k-\tau-i}) + \sum_{i=1}^n w_{i,k-\tau-i} + \sum_{i=1}^{n_d} d_{n_d} v_{k-i} + v_k. \end{aligned} \quad (9)$$

Note that the model in (9) contains the product of the parameter vectors  $\mathbf{b}$  and  $\boldsymbol{\gamma}$  of the linear and nonlinear blocks, which increases the complexity of identification. Besides, any identification scheme cannot distinguish  $(\mathbf{b}, \boldsymbol{\gamma})$  from  $(\alpha \mathbf{b}, \boldsymbol{\gamma}/\alpha)$  for some nonzero and finite constant  $\alpha$ , because any pair  $(\alpha \mathbf{b}, \bar{u}_k/\alpha)$  would produce identical input and output measurements. Therefore, in order to get unique parameter estimates, one of the entries of  $\mathbf{b}$  and  $\boldsymbol{\gamma}$  has to be fixed. Here, adopt the key term separation technique [34, 35] and let the first element of the vector  $\mathbf{b}$  equal 1; i.e.,  $b_1 = 1$ . Then Equation (9) can be rewritten as

$$y_k = -\sum_{i=1}^n a_i x_{1,k-\tau-i} + \sum_{j=1}^m \gamma_j f_j(u_{k-\tau-i}) + \sum_{i=2}^n b_i \bar{u}_{k-\tau-i} + \sum_{i=1}^n w_{i,k-\tau-i} + \sum_{i=1}^{n_d} d_{n_d} v_{k-i} + v_k. \quad (10)$$

Let  $n_1 = 2n + m + n_d - 1$  and define parameter vectors and information vectors as

$$\boldsymbol{\vartheta} := [a_1, a_2, \dots, a_n, \boldsymbol{\gamma}^T, b_2, b_3, \dots, b_n]^T \in \mathbb{R}^{2n+m-1},$$

$$\mathbf{d} := [d_1, d_2, \dots, d_{n_d}]^T \in \mathbb{R}^{n_d},$$

$$\boldsymbol{\varphi}_k := [-x_{1,k-\tau-1}, -x_{1,k-\tau-2}, \dots, -x_{1,k-\tau-n}, \mathbf{f}(u_{k-\tau-1}), \bar{u}_{k-\tau-2}, \bar{u}_{k-\tau-3}, \dots, \bar{u}_{k-\tau-n}, \boldsymbol{\phi}_k^T]^T \in \mathbb{R}^{n_1},$$

$$\begin{aligned}\phi_k &:= [v_{k-1}, v_{k-2}, \dots, v_{k-n_d}]^T \in \mathbb{R}^{n_d}, \\ \theta &:= \begin{bmatrix} \vartheta \\ d \end{bmatrix} \in \mathbb{R}^{n_1}.\end{aligned}$$

Then Equation (10) can be written as

$$y_k = \varphi_k^T \theta + \sum_{i=1}^n w_{i,k-\tau-i} + v_k.$$

Let  $e_k := \sum_{i=1}^n w_{i,k-\tau-i} + v_k$ . Since both the process noise  $w_k$  and measurement noise  $v_k$  are Gaussian white noises, and  $w_k \sim N(\mathbf{0}, \mathbf{Q})$ ,  $v_k \sim N(0, R)$ , then the output  $y_k$  of the system in Figure 1 can be expressed by the following pseudo-linear regressive form:

$$y_k = \varphi_k^T \theta + e_k. \quad (11)$$

The objective of this paper is to derive identification algorithms to estimate the states and parameters  $a_i$ ,  $b_i$ ,  $\gamma_i$  and  $d_i$  and the time-delay (if it is unknown) for the Hammerstein state space model by using the measured input-output data  $\{u_k, y_k\}$ .

For the model in (11), the number of the unknown parameters in the vector  $\theta$  is  $n_1 := 2n + m + n_d - 1$ . In fact, there is another model which is called the over-parameterization model to deal with the product term of  $\mathbf{b}$  and  $\gamma$ . For example, the method in [33] expresses the parameter vector as

$$\begin{aligned}\theta &:= [a_1, a_2, \dots, a_n, b_1\gamma_1, b_2\gamma_1, \dots, b_n\gamma_1, b_1\gamma_2, b_2\gamma_2, \dots, b_n\gamma_2, \dots, \\ &\quad b_1\gamma_m, b_2\gamma_m, \dots, b_n\gamma_m, d_1, d_2, \dots, d_{n_d}]^T \in \mathbb{R}^{n+nm+n_d}.\end{aligned} \quad (12)$$

In that situation, the number of the unknown parameters in the vector  $\theta$  is  $n_2 := n + nm + n_d$ . Since both  $n$  and  $m$  are positive integers, the difference between  $n_1$  and  $n_2$  is  $\Delta n := n_2 - n_1 = nm - n - m + 1 \geq 0$ . That means that the dimension of the unknown parameter vector in the over-parameterization method is larger than that in the key term separation based method. When the order of the state space model becomes higher or the polynomial  $f(u_t)$  has more components, the difference  $\Delta n$  will become large. Thus the method of this paper requires lower computational load for realizing the parameter estimation algorithm.

### 3. THE KALMAN-FILTER BASED LEAST SQUARES ALGORITHMS WITH KNOWN TIME-DELAY

In process control industry, the phenomenon of time-delay often exists in the procedure of signal transmission. Based on the empirical knowledge, it is assumed that the time-delay is known. In this section, we derive the Kalman-filter based least squares algorithms to identify the Hammerstein dynamic system with known time-delay.

#### 3.1. The state estimation algorithm

The Kalman filter can give the state estimates [36]. There are two steps in the Kalman filter, one is called the time update (or prediction), the other is called the measurement update (or modification). On the prediction step, the prior estimates of the state and its covariance are predicted; on the modification step, the newest measurement and prior estimates are combined together to modify the posterior estimates.

Because the parameter matrix/vectors are unknown, we need to use the estimated parameter vector

$$\hat{\theta}_k := [\hat{a}_{1,k}, \hat{a}_{2,k}, \dots, \hat{a}_{n,k}, \hat{\gamma}_{1,k}, \hat{\gamma}_{2,k}, \dots, \hat{\gamma}_{m,k}, \hat{b}_{2,k}, \dots, \hat{b}_{n,k}, \hat{d}_{1,k}, \hat{d}_{2,k}, \dots, \hat{d}_{n_d,k}]$$

to construct the estimates  $\hat{\mathbf{A}}_k$  and  $\hat{\mathbf{b}}_k$  of  $\mathbf{A}$  and  $\mathbf{b}$ , and use the estimated parameter matrix  $\hat{\mathbf{A}}_k$  and the estimated parameter vector  $\hat{\mathbf{b}}_k$  to compute the estimate  $\hat{\mathbf{x}}_{k+1}$  of the state vector  $\mathbf{x}_{k+1}$ . Here, the

Kalman filter is used to generate the optimal state estimate. Considering the time delay  $\tau$  in the output, we can use the following Kalman filter to obtain the state estimate  $\hat{x}_{k+1}$  [23]:

$$\bar{x}_{k+1} = \hat{A}_k \hat{x}_k + \hat{b}_k \bar{u}_k, \quad \hat{x}_1 = \mathbf{1}_n / p_0 \quad (13)$$

$$\bar{P}_k = \hat{A}_k P_k \hat{A}_k^T + Q, \quad P_1 = I_n \quad (14)$$

$$L_k = \bar{P}_k c^T (c \bar{P}_k c^T + R)^{-1}, \quad (15)$$

$$\hat{x}_{k+1} = \bar{x}_{k+1} + L_k (y_{k+1+\tau} - c \bar{x}_{k+1}), \quad (16)$$

$$P_{k+1} = \bar{P}_k - L_k c \bar{P}_k, \quad (17)$$

$$\hat{A}_k = \begin{bmatrix} -\hat{a}_{1,k} & 1 & 0 & \cdots & 0 \\ -\hat{a}_{2,k} & 0 & 1 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ -\hat{a}_{n-1,k} & 0 & \cdots & 0 & 1 \\ -\hat{a}_{n,k} & 0 & \cdots & \cdots & 0 \end{bmatrix}, \quad (18)$$

$$\hat{b}_k = [1, \hat{b}_{2,k}, \hat{b}_{3,k}, \cdots, \hat{b}_{n,k}]^T, \quad (19)$$

where  $\bar{x}_{k+1}$  is a prior estimate of the state  $x_{k+1}$  for given measurements up to and including time  $k$ ;  $\hat{x}_{k+1}$  is a posterior estimate of the state  $x_{k+1}$  for given measurements up to and including time  $k+1$ ;  $\bar{P}_k$  is the variance of the prior estimation error;  $P_{k+1}$  is the variance of the posterior estimation error;  $L_k$  is the Kalman gain vector.

Another frequently used method for state estimation is to use the state observer [32] which can be used to get the approximate values of the system states. The drawback of the state observer is that it is generally suitable for deterministic systems. Thus this paper uses the parameter estimates based Kalman filter for generating the state estimates in the stochastic framework.

### 3.2. The iterative parameter estimation algorithm

Opt a set of data from  $j = k$  to  $j = k + L - 1$  ( $L$  denotes the data length) and define the stacked output vector  $\mathbf{Y}_{k,L}$  and stacked information matrix  $\Phi_{k,L}$  as

$$\mathbf{Y}_{k,L} := \begin{bmatrix} y_k \\ y_{k+1} \\ \vdots \\ y_{k+L-1} \end{bmatrix} \in \mathbb{R}^L; \quad \Phi_{k,L} := \begin{bmatrix} \varphi_k^T \\ \varphi_{k+1}^T \\ \vdots \\ \varphi_{k+L-1}^T \end{bmatrix} \in \mathbb{R}^{L \times n_1}.$$

According to (11), we define a criterion function

$$J_1(\theta) = \|\mathbf{Y}_{k,L} - \Phi_{k,L} \theta\|^2.$$

Let  $s = 1, 2, 3, \cdots$  be an iterative variable and  $\hat{\theta}_k^s$  be the estimate of  $\theta$  at iteration  $s$ . Minimizing the criterion function  $J_1(\theta)$  and letting the derivative of  $J_1(\theta)$  with respect to  $\theta$  be zero gives

$$\left. \frac{\partial J_1(\theta)}{\partial \theta} \right|_{\theta = \hat{\theta}_k^s} = -2 \Phi_{k,L}^T [\mathbf{Y}_{k,L} - \Phi_{k,L} \hat{\theta}_k^s] = \mathbf{0}.$$

Then we can obtain the least squares estimate of the parameter vector  $\theta$ :

$$\hat{\theta}_k^s = [\Phi_{k,L}^T \Phi_{k,L}]^{-1} \Phi_{k,L}^T \mathbf{Y}_{k,L}. \quad (20)$$

But we cannot obtain the estimate  $\hat{\theta}_k^s$  directly from (20), because the information vector  $\varphi_k$  in  $\Phi_{k,L}$  contains the unknown state variable  $x_{1,k-\tau-i}$ , the unknown output  $\bar{u}_{k-i}$  of the nonlinear block and the unknown noise terms  $v_{k-j}$ . The scheme here is to replace  $x_{1,k-\tau-i}$ ,  $\bar{u}_{k-i}$  and  $v_{k-j}$  in  $\varphi_k$  with their corresponding estimates  $\hat{x}_{1,k-\tau-i}^{s-1}$ ,  $\hat{u}_{k-i}^{s-1}$  and  $\hat{v}_{k-j}^{s-1}$  at iteration  $s-1$  and define

$$\hat{\varphi}_k^s = [-\hat{x}_{1,k-\tau-1}^{s-1}, -\hat{x}_{1,k-\tau-2}^{s-1}, \cdots, -\hat{x}_{1,k-\tau-n}^{s-1}, \mathbf{f}(u_{k-\tau-1}), \hat{u}_{k-\tau-2}^{s-1}, \cdots, \hat{u}_{k-\tau-n}^{s-1}, \hat{v}_{k-1}^{s-1}],$$

$$\hat{v}_{k-2}^{s-1}, \dots, \hat{v}_{k-n_d}^{s-1}]^T.$$

Replacing the unknown coefficient  $\gamma_j$  in (3) with its estimate  $\hat{\gamma}_{j,k}^s$  gives

$$\hat{u}_k^s = \hat{\gamma}_{1,k}^s f_1(u_k) + \hat{\gamma}_{2,k}^s f_2(u_k) + \dots + \hat{\gamma}_{m,k}^s f_m(u_k).$$

From (2), we have

$$v_k = y_k - \mathbf{c}\mathbf{x}_{k-\tau} - \mathbf{d}^T \boldsymbol{\phi}_k. \quad (21)$$

Replace  $\mathbf{x}_{k-\tau}$ ,  $\boldsymbol{\phi}_k$  and  $\mathbf{d}$  in (21) with their estimates  $\hat{\mathbf{x}}_{k-\tau}^s$ ,  $\hat{\boldsymbol{\phi}}_k^s$  and  $\hat{\mathbf{d}}_k^s$ , the estimate  $\hat{v}_k^s$  can be computed by

$$\hat{v}_k^s = y_k - \mathbf{c}\hat{\mathbf{x}}_{k-\tau}^s - (\hat{\mathbf{d}}_k^s)^T \hat{\boldsymbol{\phi}}_k^s.$$

Define

$$\hat{\boldsymbol{\Phi}}_{k,L}^s := [\hat{\boldsymbol{\varphi}}_k^s, \hat{\boldsymbol{\varphi}}_{k+1}^s, \dots, \hat{\boldsymbol{\varphi}}_{k+L-1}^s]^T \in \mathbb{R}^{L \times n_1}.$$

Replacing the information matrix  $\boldsymbol{\Phi}_{k,L}$  in (20) with its estimate  $\hat{\boldsymbol{\Phi}}_{k,L}^s$  and using the Kalman filter to obtain the estimates of the unknown states, we can get the Kalman-filter based least squares iterative (KF-LSI) algorithm for estimating the parameters and states of the Hammerstein state space models:

$$\hat{\boldsymbol{\theta}}_k^s = [(\hat{\boldsymbol{\Phi}}_{k,L}^s)^T \hat{\boldsymbol{\Phi}}_{k,L}^s]^{-1} (\hat{\boldsymbol{\Phi}}_{k,L}^s)^T \mathbf{Y}_{k,L}, \quad (22)$$

$$\mathbf{Y}_{k,L} = [y_k, y_{k+1}, \dots, y_{k+L-1}]^T, \quad (23)$$

$$\hat{\boldsymbol{\Phi}}_{k,L}^s = [\hat{\boldsymbol{\varphi}}_k^s, \hat{\boldsymbol{\varphi}}_{k+1}^s, \dots, \hat{\boldsymbol{\varphi}}_{k+L-1}^s]^T, \quad (24)$$

$$\hat{\boldsymbol{\varphi}}_k^s = [-\hat{x}_{1,k-\tau-1}^{s-1}, -\hat{x}_{1,k-\tau-2}^{s-1}, \dots, -\hat{x}_{1,k-\tau-n}^{s-1}, \mathbf{f}(u_{k-\tau-1}), \hat{u}_{k-\tau-2}^{s-1}, \dots, \hat{u}_{k-\tau-n}^{s-1}, \hat{v}_{k-1}^{s-1}, \dots, \hat{v}_{k-n_d}^{s-1}]^T, \quad (25)$$

$$\hat{u}_k^s = \hat{\gamma}_{1,k}^s f_1(u_k) + \hat{\gamma}_{2,k}^s f_2(u_k) + \dots + \hat{\gamma}_{m,k}^s f_m(u_k), \quad (26)$$

$$\hat{v}_k^s = y_k - \mathbf{c}\hat{\mathbf{x}}_{k-\tau}^s - (\hat{\mathbf{d}}_k^s)^T \hat{\boldsymbol{\phi}}_k^s, \quad (27)$$

$$\hat{\boldsymbol{\theta}}_k^s = [\hat{a}_{1,k}^s, \hat{a}_{2,k}^s, \dots, \hat{a}_{n,k}^s, \hat{\gamma}_{1,k}^s, \hat{\gamma}_{2,k}^s, \dots, \hat{\gamma}_{m,k}^s, \hat{b}_{2,k}^s, \dots, \hat{b}_{n,k}^s, \hat{d}_{1,k}^s, \dots, \hat{d}_{n_d,k}^s]^T, \quad (28)$$

$$\hat{\mathbf{A}}_k^s = \begin{bmatrix} -\hat{a}_{1,k}^s & 1 & 0 & \dots & 0 \\ -\hat{a}_{2,k}^s & 0 & 1 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ -\hat{a}_{n-1,k}^s & 0 & \dots & 0 & 1 \\ -\hat{a}_{n,k}^s & 0 & \dots & \dots & 0 \end{bmatrix}, \quad (29)$$

$$\hat{\mathbf{b}}_k^s = [1, \hat{b}_{2,k}^s, \hat{b}_{3,k}^s, \dots, \hat{b}_{n,k}^s]^T, \quad (30)$$

$$\hat{\mathbf{x}}_{k+1}^s = \hat{\mathbf{A}}_k^s \hat{\mathbf{x}}_k^s + \hat{\mathbf{b}}_k^s \bar{u}_k, \quad \hat{\mathbf{x}}_1^s = \mathbf{1}_n / p_0, \quad (31)$$

$$\bar{\mathbf{P}}_k = \hat{\mathbf{A}}_k^s \mathbf{P}_k (\hat{\mathbf{A}}_k^s)^T + \mathbf{Q}, \quad \mathbf{P}_1 = \mathbf{I}_n, \quad (32)$$

$$\mathbf{L}_k = \bar{\mathbf{P}}_k \mathbf{c}^T (\mathbf{c} \bar{\mathbf{P}}_k \mathbf{c}^T + R)^{-1}, \quad (33)$$

$$\hat{\mathbf{x}}_{k+1}^s = \bar{\mathbf{x}}_{k+1}^s + \mathbf{L}_k (y_{k+1+\tau} - \mathbf{c} \bar{\mathbf{x}}_{k+1}^s), \quad (34)$$

$$\mathbf{P}_{k+1} = \bar{\mathbf{P}}_k - \mathbf{L}_k \mathbf{c} \bar{\mathbf{P}}_k. \quad (35)$$

**Remark 1** The iterative algorithm is implemented off-line, it repeatedly uses a batch of observed data and can get good parameter estimates after only several iterations [37, 38, 39, 40]. In the LSI algorithm, the measured input-output data should be collected in advance.

The flowchart of computing the parameter estimates  $\hat{\boldsymbol{\theta}}_k^s$  and the state estimate  $\hat{\mathbf{x}}_k^s$  by the KF-LSI algorithm is shown in Figure 2.

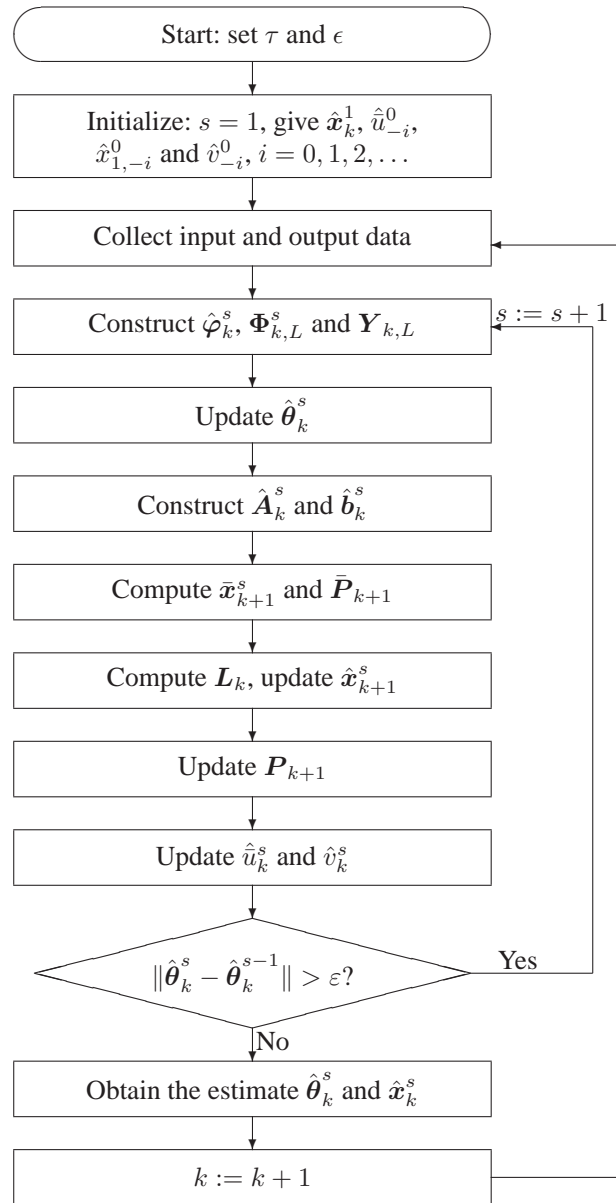


Figure 2. The flowchart of the KF-LSI algorithm for computing  $\hat{\theta}_k^s, \hat{x}_k^s$

### 3.3. The recursive parameter estimation algorithm

Unlike the LSI algorithm, the RLS algorithm may be carried out on-line. By defining a covariance matrix, the RLS algorithm uses the input and output data to update the parameter estimates at each step. In this section, a Kalman filter based RLS algorithm is derived.

Similarly, according to (11), define a quadratic criterion function

$$J_2(\theta) = \sum_{j=1}^k [y(j) - \varphi_j^T \theta]^2.$$

Minimizing  $J_2(\theta)$  using the least squares principle, letting the partial derivative of  $J_2(\theta)$  with respect to  $\theta$  be zero, we can obtain

$$\hat{\theta}_k = \hat{\theta}_{k-1} + L_{1,k} [y_k - \varphi_k^T \hat{\theta}_{k-1}], \quad (36)$$



$$\mathbf{L}_{1,k} = \mathbf{P}_{1,k-1} \boldsymbol{\varphi}_k [1 + \boldsymbol{\varphi}_k^T \mathbf{P}_{1,k-1} \boldsymbol{\varphi}_k]^{-1}, \quad (37)$$

$$\mathbf{P}_{1,k} = [\mathbf{I} - \mathbf{L}_{1,k} \boldsymbol{\varphi}_k^T] \mathbf{P}_{1,k-1}. \quad (38)$$

Because the information vector  $\boldsymbol{\varphi}_k$  contains the unmeasurable state variable  $x_{1,k-\tau-i}$ , unknown output of nonlinear block  $\bar{u}_{k-i}$  and the unknown noise  $v_{k-j}$ , the algorithm in (36)–(38) cannot be implemented. Here we replace  $x_{1,k-\tau-i}$ ,  $\bar{u}_{k-i}$  and  $v_{k-j}$  in  $\boldsymbol{\varphi}_k$  with their corresponding estimates  $\hat{x}_{1,k-\tau-i}$ ,  $\hat{u}_{k-i}$  and  $\hat{v}_{k-j}$  at time  $k$  and define

$$\hat{\boldsymbol{\varphi}}_k := [-\hat{x}_{1,k-\tau-1}, -\hat{x}_{1,k-\tau-2}, \dots, -\hat{x}_{1,k-\tau-n}, \mathbf{f}(u_{k-\tau-1}), \hat{u}_{k-\tau-2}, \dots, \hat{u}_{k-\tau-n}, \hat{v}_{k-1}, \dots, \hat{v}_{k-n_d}]^T \in \mathbb{R}^{n_1}.$$

Replacing the unknown coefficient  $\gamma_j$  in (3) with its estimate  $\hat{\gamma}_{j,k}$  gives

$$\hat{u}_k = \hat{\gamma}_{1,k} f_1(u_k) + \hat{\gamma}_{2,k} f_2(u_k) + \dots + \hat{\gamma}_{m,k} f_m(u_k). \quad (39)$$

Replacing  $\mathbf{x}_{k-\tau}$ ,  $\boldsymbol{\phi}_k$  and  $\mathbf{d}$  in (21) with their estimates  $\hat{\mathbf{x}}_{k-\tau}$ ,  $\hat{\boldsymbol{\phi}}_k$  and  $\hat{\mathbf{d}}_k$ , the estimate  $\hat{v}_k$  can be computed by

$$\hat{v}_k = y_k - \mathbf{c} \hat{\mathbf{x}}_{k-\tau} - \hat{\mathbf{d}}_k^T \hat{\boldsymbol{\phi}}_k. \quad (40)$$

Then we can summarize the following RLS algorithm for estimating the parameter vector  $\boldsymbol{\theta}$  as

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} + \mathbf{L}_{1,k} [y_k - \hat{\boldsymbol{\varphi}}_k^T \hat{\boldsymbol{\theta}}_{k-1}], \quad (41)$$

$$\mathbf{L}_{1,k} = \mathbf{P}_{1,k-1} \hat{\boldsymbol{\varphi}}_k [1 + \hat{\boldsymbol{\varphi}}_k^T \mathbf{P}_{1,k-1} \hat{\boldsymbol{\varphi}}_k]^{-1}, \quad (42)$$

$$\mathbf{P}_{1,k} = [\mathbf{I}_{n_1} - \mathbf{L}_{1,k} \hat{\boldsymbol{\varphi}}_k^T] \mathbf{P}_{1,k-1}, \quad \mathbf{P}_{1,0} = p_0 \mathbf{I}_{n_1}, \quad (43)$$

$$\hat{\boldsymbol{\varphi}}_k = [-\hat{x}_{1,k-\tau-1}, -\hat{x}_{1,k-\tau-2}, \dots, -\hat{x}_{1,k-\tau-n}, \mathbf{f}(u_{k-\tau-1}), \hat{u}_{k-\tau-2}, \dots, \hat{u}_{k-\tau-n}, \hat{v}_{k-1}, \dots, \hat{v}_{k-n_d}]^T, \quad (44)$$

$$\hat{\boldsymbol{\theta}}_k = [\hat{a}_{1,k}, \hat{a}_{2,k}, \dots, \hat{a}_{n,k}, \hat{\gamma}_{1,k}, \hat{\gamma}_{2,k}, \dots, \hat{\gamma}_{m,k}, \hat{b}_{2,k}, \dots, \hat{b}_{n,k}, \hat{d}_{1,k}, \dots, \hat{d}_{n_d,k}]^T, \quad (45)$$

$$\hat{u}_k = \hat{\gamma}_{1,k} f_1(u_k) + \hat{\gamma}_{2,k} f_2(u_k) + \dots + \hat{\gamma}_{m,k} f_m(u_k), \quad (46)$$

$$\hat{v}_k = y_k - \mathbf{c} \hat{\mathbf{x}}_{k-\tau} - \hat{\mathbf{d}}_k^T \hat{\boldsymbol{\phi}}_k. \quad (47)$$

For the unknown state  $\mathbf{x}_{1,k}$ , we also use the Kalman filter algorithm to generate its estimate  $\hat{\mathbf{x}}_{1,k}$ . Combining Equations (41)–(47) and (13)–(19), we can get the Kalman-filter based recursive least squares (KF-RLS) algorithm for identifying the Hammerstein state space systems.

To initialize the KF-RLS algorithm, the initial value  $\hat{\boldsymbol{\theta}}_0$  and  $\hat{\mathbf{x}}_1$  is generally taken to be a real vector, e.g.,  $\hat{\boldsymbol{\theta}}_0 = \mathbf{1}_{n_1}/p_0$  and  $\hat{\mathbf{x}}_1 = \mathbf{1}_n/p_0$  with  $p_0$  being normally a large positive number (e.g.,  $p_0 = 10^6$ ). Let  $\hat{x}_{1,-i} = 1/p_0$ ,  $\hat{u}_{-i} = 1/p_0$  and  $\hat{v}_{-j} = 1/p_0$  for  $i = 0, 1, 2, \dots, n$ ,  $j = 0, 1, 2, \dots, n_d - 1$ . The initial values of the covariance matrixes are set as  $\mathbf{P}_{1,0} = p_0 \mathbf{I}_{n_1}$ ,  $\mathbf{P}_1 = \mathbf{I}_n$ .

#### 4. THE REDUNDANT RELS ALGORITHM FOR HAMMERSTEIN SYSTEMS WITH UNKNOWN TIME-DELAY

Although we can predict the time-delay based on the empirical data in some special cases, there are many uncertain varying elements which bring the unknown time-delay in the actual systems. Besides identifying the system parameters, we also need to get the estimate of the time-delay based on the observed data.

Reconsidering (11), if we extend the information vector  $\boldsymbol{\varphi}_k$  and supplement the lost terms which are omitted because of the time-delay. The information vector  $\bar{\boldsymbol{\varphi}}_k$  is redefined as

$$\bar{\boldsymbol{\varphi}}_k := [-x_{1,k-1}, -x_{1,k-2}, \dots, -x_{1,k-\tau}, -x_{1,k-\tau-1}, -x_{1,k-\tau-2}, \dots, -x_{1,k-\tau-n}, \mathbf{f}(u_{k-\tau-1}), \bar{u}_{k-\tau-2}, \dots, \bar{u}_{k-\tau-n}, v_{k-1}, v_{k-2}, \dots, v_{k-n_d}]^T \in \mathbb{R}^{n_1+\tau}.$$

At the same time, the parameter vector  $\boldsymbol{\theta}$  is extended as the following form:

$$\bar{\boldsymbol{\theta}} := [a_{01}, a_{02}, \dots, a_{0\tau}, a_1, a_2, \dots, a_n, \gamma_1, \gamma_2, \dots, \gamma_m, b_2, b_3, \dots, b_n, d_1, d_2, \dots, d_{n_d}]^T \in \mathbb{R}^{n_1+\tau}.$$

The dimension of  $\bar{\theta}$  is larger than the dimension of the original parameter vector  $\theta$  because of the redundant parameters  $\{a_{01}, a_{02}, \dots, a_{0\tau}\}$ . In fact, the true values of the redundant parameters should be zeros to guarantee the model structure unchanged. So the number of the redundant parameters is the value of the time-delay  $\tau$ . Once the estimates of the redundant parameters lie in a given confidence interval, we can assume them as zero. So the number of the continuous approximate zero elements in the front part of the estimate of  $\bar{\theta}$  is the estimate of the time-delay  $\tau$ . Then our objective is how to obtain the estimate  $\hat{\theta}_k$  of  $\bar{\theta}_k$ .

Rewrite the model in (11) as

$$y_k = \bar{\varphi}_k^T \bar{\theta} + e_k. \quad (48)$$

Define a quadratic criterion function

$$J_3(\bar{\theta}) := \sum_{j=1}^k [y(j) - \bar{\varphi}_j^T \bar{\theta}]^2.$$

Minimizing  $J_3(\bar{\theta})$  and using the least squares principle, we can get the following relations:

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \mathbf{L}_{2,k} [y_k - \bar{\varphi}_k^T \hat{\theta}_{k-1}], \quad (49)$$

$$\mathbf{L}_{2,k} = \mathbf{P}_{2,k-1} \bar{\varphi}_k [1 + \bar{\varphi}_k^T \mathbf{P}_{2,k-1} \bar{\varphi}_k]^{-1}, \quad (50)$$

$$\mathbf{P}_{2,k} = [\mathbf{I} - \mathbf{L}_{2,k} \bar{\varphi}_k^T] \mathbf{P}_{2,k-1}. \quad (51)$$

Similarly, the information vector  $\bar{\varphi}_k$  contains the unknown  $x_{1,k-i}$ ,  $\bar{u}_{k-i}$  and  $v_{k-j}$ , the algorithm in (49)–(51) cannot be implemented. Replace  $x_{1,k-i}$ ,  $\bar{u}_{k-i}$  and  $v_{k-j}$  in  $\bar{\varphi}_k$  with their corresponding estimates  $\hat{x}_{1,k-i}$ ,  $\hat{u}_{k-i}$  and  $\hat{v}_{k-j}$  at time  $k$  and define

$$\hat{\varphi}_k := [-\hat{x}_{1,k-1}, -\hat{x}_{1,k-2}, \dots, -\hat{x}_{1,k-\tau}, -\hat{x}_{1,k-\tau-1}, -\hat{x}_{1,k-\tau-2}, \dots, -\hat{x}_{1,k-\tau-n}, \mathbf{f}(u_{k-\tau-1}), \hat{u}_{k-\tau-2}, \dots, \hat{u}_{k-\tau-n}, \hat{v}_{k-1}, \dots, \hat{v}_{k-n_d}]^T \in \mathbb{R}^{n_1+\tau}.$$

Replacing  $\bar{\varphi}_k$  in (49)–(51) with its estimate  $\hat{\varphi}_k$ , we can summarize the following recursive extended least squares (RELS) algorithm for estimating  $\bar{\theta}$ :

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \mathbf{L}_{2,k} [y_k - \hat{\varphi}_k^T \hat{\theta}_{k-1}], \quad (52)$$

$$\mathbf{L}_{2,k} = \mathbf{P}_{2,k-1} \hat{\varphi}_k [1 + \hat{\varphi}_k^T \mathbf{P}_{2,k-1} \hat{\varphi}_k]^{-1}, \quad (53)$$

$$\mathbf{P}_{2,k} = [\mathbf{I} - \mathbf{L}_{2,k} \hat{\varphi}_k^T] \mathbf{P}_{2,k-1}, \quad (54)$$

$$\hat{\varphi}_k = [-\hat{x}_{1,k-1}, -\hat{x}_{1,k-2}, \dots, -\hat{x}_{1,k-\tau}, -\hat{x}_{1,k-\tau-1}, -\hat{x}_{1,k-\tau-2}, \dots, -\hat{x}_{1,k-\tau-n}, \dots, \mathbf{f}(u_{k-\tau-1}), \hat{u}_{k-\tau-2}, \hat{u}_{k-\tau-n}, \hat{v}_{k-1}, \dots, \hat{v}_{k-n_d}]^T, \quad (55)$$

$$\hat{\theta}_k = [\hat{a}_{01,k}, \hat{a}_{02,k}, \dots, \hat{a}_{0\tau,k}, \hat{a}_{1,k}, \hat{a}_{2,k}, \dots, \hat{a}_{n,k}, \hat{\gamma}_{1,k}, \dots, \hat{\gamma}_{m,k}, \hat{b}_{2,k}, \dots, \hat{b}_{n,k}, \hat{d}_{1,k}, \dots, \hat{d}_{n_d,k}]^T, \quad (56)$$

$$\hat{u}_k = \hat{\gamma}_{1,k} f_1(u_k) + \hat{\gamma}_{2,k} f_2(u_k) + \dots + \hat{\gamma}_{m,k} f_m(u_k), \quad (57)$$

$$\hat{v}_k = y_k - \mathbf{c} \hat{\mathbf{x}}_{k-\tau} - \hat{\mathbf{d}}_k^T \hat{\varphi}_k. \quad (58)$$

Combining (52)–(58) and (13)–(19), the Kalman-filter based recursive extended least squares (KF-RELS) algorithm for estimating  $\bar{\theta}$  is obtained. Then the value of the time-delay  $\tau$  can be evaluated by counting how many elements which are close to zero in the front part of  $\hat{\theta}_k$ .

The steps for computing the state and parameter estimates  $\hat{\mathbf{x}}_k$  and  $\hat{\theta}_k$  under the KF-RELS algorithm with the increasing of  $k$  are as follows.

1. To initialize. Let  $k = 1$ , the initial value  $\hat{\theta}_0$  and  $\hat{\mathbf{x}}_1$  is generally taken to be a real vector, e.g.,  $\hat{\theta}_0 = \mathbf{1}_{n_1+\tau}/p_0$  and  $\hat{\mathbf{x}}_1 = \mathbf{1}_n/p_0$ . Let  $\hat{x}_{1,-i} = 1/p_0$ ,  $\hat{u}_{-i} = 1/p_0$  and  $\hat{v}_{-j} = 1/p_0$  ( $i = 0, 1, 2, \dots, j = 0, 1, 2, \dots, n_d - 1$ ). The initial values of the covariance matrixes are set as:  $\bar{\mathbf{P}}_0 = p_0 \mathbf{I}_{n_1+\tau}$ ,  $\mathbf{P}_1 = \mathbf{I}_n$ . Give a small positive  $\varepsilon$ .

2. Collect the input-output data, construct the information vector  $\hat{\varphi}_k$  by (55).
3. Compute the gain vector  $L_{2,k}$  and the covariance matrix  $P_{2,k}$  by (53) and (54).
4. Update the parameter estimates  $\hat{\theta}_k$  by (52).
5. Construct the matrix  $\hat{A}_k$  by (18), and the vector  $\hat{b}_k$  by (19).
6. Compute the prior estimate  $\bar{x}_{k+1}$  and the covariance matrix  $\bar{P}_k$  by (13) and (14).
7. Compute the Kalman gain  $L_k$  by (15), update the posterior state estimate  $\hat{x}_{k+1}$  by (16) and the posterior covariance matrix  $P_{k+1}$  by (17).
8. Update the estimates  $\hat{u}_k$  and  $\hat{v}_k$  by (57) and (58).
9. Evaluate the relative error of parameter estimates, if

$$\delta = \|\hat{\theta}_k - \hat{\theta}_{k-1}\| \leq \varepsilon$$

for some pre-set small  $\varepsilon$ , then terminate this procedure and obtain the parameter estimate  $\hat{\theta}_k$  and the state estimate  $\hat{x}_{k+1}$ ; otherwise, increase  $k$  by 1 and go to Step 2.

## 5. EXAMPLE

**Example 1:** Consider the following Hammerstein system with state space model:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{b}\bar{u}_k + \mathbf{w}_k, \\ y_k &= \mathbf{c}\mathbf{x}_{t-\tau} - 0.30v_{k-1} + v_k, \\ \bar{u}_k &= \gamma_1 u_k + \gamma_2 u_k^2 + \gamma_3 u_k^3, \\ \mathbf{A} &= \begin{bmatrix} -0.50 & 1 \\ -0.26 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1.50 \end{bmatrix}, \\ \mathbf{c} &= [1, 0], \\ \gamma &= [0.25, 0.60, 0.76]. \end{aligned}$$

In simulation, the input  $\{u_k\}$  is taken as a persistent excitation sequence with zero mean and unit variance, and  $\{\mathbf{w}_k\}$ ,  $\{v_k\}$  as uncorrelated noise sequences with zero mean and variances  $\mathbf{Q} := \begin{bmatrix} \sigma_w^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}$  and  $R := \sigma_v^2$ , respectively. The output time-delay  $\tau = 2$ . Using the KF-LSI algorithm and KF-RLS algorithm to estimate the parameters of this model, the parameter estimates and their estimation errors  $\delta := \|\hat{\theta}_k - \theta\|/\|\theta\|$  with different noise variances are shown in Tables I–II. In the KF-LSI algorithm, the data length  $L = 1000$ .

To show the advantage of Kalman filter in obtaining the estimates of unknown states, the state observer (SO) is also used to get the estimates of the unknown system states. Combining the SO with the LSI and RLS algorithm to get the SO-LSI estimates and the SO-RLS estimates. The estimation errors versus  $s$  or  $k$  are plotted in Figure 3 and Figure 4.

From Tables I–III and Figures 3–4, we can draw the following conclusions.

1. It is clear see that the estimation errors  $\delta$  are becoming smaller (in general) as  $s$  or  $k$  increasing. This indicates that the proposed KF-LSI and KF-RLS algorithms are effective for Hammerstein state space systems.
2. A smaller noise variance leads to smaller parameter estimation errors under the same data length in the KF-LSI and KF-RLS algorithm.
3. The proposed KF-LSI and KF-RLS algorithm have better performances on estimation accuracy than the SO-LSI and SO-RLS estimation algorithm.

For comparison, we use the state observer based hierarchical stochastic gradient (SO-HSG) algorithm and the hierarchical multi-innovation stochastic gradient (SO-HMISG) algorithm in [31] and the state observer based hierarchical least squares (SO-HLS) algorithm in [32] to identify this model. In order to acquire the unique estimate, they assume that the norm of the coefficient vector  $\gamma$  is unity and the first coefficient is positive, i.e.,  $\|\gamma\|^2 = 1$  and  $\gamma_1 > 0$ . The root mean square error

(RMSE),

$$\text{RMSE} = \sqrt{\sum_{j=1}^L [y_j - \hat{y}_j]^2 / L}$$

is employed to reflect the errors between the predicted values and true values. Using the Monte Carlo simulations with 100 set of noise realizations, the RMSE and its bias of the five different algorithms are summarized in Table III. The results of self-validation and cross-validation of the discussed different algorithms are shown in Figures 5 and 6.

From Figures 5–6 and Table III, we can see that 1) the KF-LSI algorithm has the smallest RMSE in the self-validation and cross validation, 2) the KF-RLS algorithm has better performances in prediction than other algorithms, 3) the RMSE of the SO-HSG algorithm is obvious larger than those of other four algorithms, by extending the length of the innovation, the SO-HMISG algorithm can enhance the accuracy of parameter estimation.

Table I. The KF-LSI parameter estimates and errors versus  $s$

$\sigma^2$	$s$	$a_1$	$a_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$b_2$	$d_1$	$\delta$ (%)
$\sigma_w^2 = 0.02^2$ $\sigma_v^2 = 0.2^2$	1	0.50024	0.26313	0.24944	0.59815	0.75849	1.49801	-0.00523	15.41463
	2	0.51093	0.26052	0.24857	0.59917	0.75988	1.50482	-0.25617	2.37727
	5	0.50822	0.26122	0.24882	0.59880	0.75964	1.50185	-0.29186	0.62240
	10	0.50822	0.26121	0.24882	0.59875	0.75964	1.50184	-0.29185	0.62308
	15	0.50822	0.26121	0.24882	0.59876	0.75964	1.50184	-0.29185	0.62308
$\sigma_w^2 = 0.05^2$ $\sigma_v^2 = 1.00^2$	1	0.50088	0.27553	0.24811	0.59074	0.75278	1.49015	-0.02357	14.49929
	2	0.50629	0.28097	0.24377	0.58957	0.75995	1.50340	-0.27305	1.93178
	5	0.50402	0.27903	0.24510	0.59069	0.75864	1.49624	-0.29440	1.21077
	10	0.50402	0.27897	0.24509	0.59063	0.75863	1.49619	-0.29440	1.21021
	15	0.50402	0.27897	0.24509	0.59064	0.75863	1.49619	-0.29440	1.21016
True values		0.50000	0.26000	0.25000	0.60000	0.76000	1.50000	-0.30000	

Table II. The KF-RLS parameter estimates and errors versus  $k$

$\sigma^2$	$k$	$a_1$	$a_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$b_2$	$d_1$	$\delta$ (%)
$\sigma_w^2 = 0.02^2$ $\sigma_v^2 = 0.2^2$	100	0.51279	0.24805	0.25026	0.57382	0.74856	1.52813	-0.25841	3.15621
	200	0.50853	0.26082	0.24446	0.59100	0.76103	1.51115	-0.24533	3.00345
	500	0.51223	0.25741	0.24892	0.60433	0.76192	1.49441	-0.24093	3.18044
	1000	0.50970	0.25894	0.25265	0.60038	0.75789	1.49877	-0.24867	2.73861
	2000	0.50929	0.26090	0.25178	0.59864	0.75763	1.50310	-0.26989	1.66526
	3000	0.50774	0.25906	0.24794	0.60148	0.76008	1.50219	-0.27050	1.60483
$\sigma_w^2 = 0.05^2$ $\sigma_v^2 = 1.00^2$	100	0.49437	0.15277	-0.33553	0.65580	1.00088	1.36885	-0.12603	35.57776
	200	0.53590	0.20030	0.13232	0.61041	0.76896	1.43699	-0.21827	8.98682
	500	0.50483	0.24569	0.22204	0.60475	0.74398	1.42952	-0.25032	4.88400
	1000	0.49799	0.24748	0.24848	0.61428	0.75588	1.43737	-0.23780	4.72763
	2000	0.49988	0.26124	0.23658	0.61713	0.76864	1.46757	-0.25904	2.99454
	3000	0.49521	0.25721	0.25604	0.59299	0.75929	1.47596	-0.27036	2.07397
True values		0.50000	0.26000	0.25000	0.60000	0.76000	1.50000	-0.30000	

**Example 2:** Consider a following Hammerstein system with unknown time-delay:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{b}\bar{u}_k + \mathbf{w}_k, \\ y_k &= \mathbf{c}\mathbf{x}_{k-\tau} - 0.18v_{k-1} + v_k, \\ \bar{u}_k &= \gamma_1 \sin(u_k) + \gamma_2 \cos(u_k), \\ \mathbf{A} &= \begin{bmatrix} -0.45 & 1 \\ -0.30 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1.50 \end{bmatrix}, \quad \mathbf{c} = [1, 0], \\ \gamma &= [1.20, -0.32]. \end{aligned}$$

Table III. The RMSE and its bias of different algorithms based on 100 Monte Carlo runs ( $\sigma_w^2 = 0.02^2$ ,  $\sigma_v^2 = 0.20^2$ )

Algorithm	RMSE	
	Self-validation	Cross-validation
KF-LSI	0.22374±0.00706	0.22602±0.02726
KF-RLS	0.23852±0.00708	0.26681±0.02815
SO-HSG	1.28185±0.24508	1.12947±0.13251
SO-HMISG(p=6)	0.30017±0.29869	0.35248±0.10866
SO-HLS	0.27849±0.10688	0.36280±0.05659

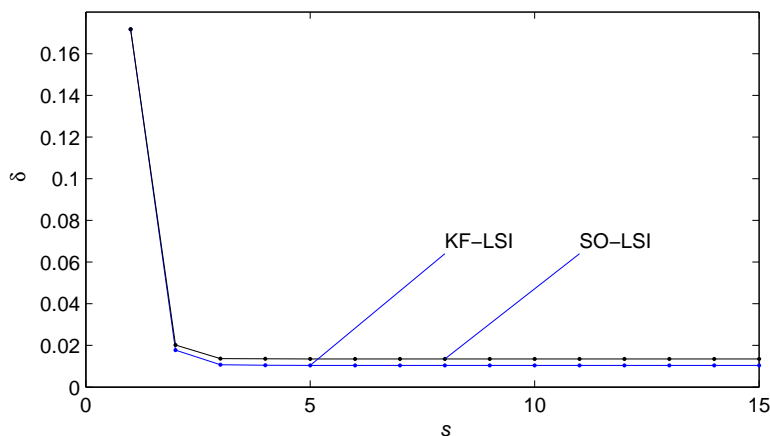


Figure 3. The SO-LSI and KF-LSI estimation errors  $\delta$  versus  $s$  ( $\sigma_w^2 = 0.02^2$ ,  $\sigma_v^2 = 0.20^2$ )

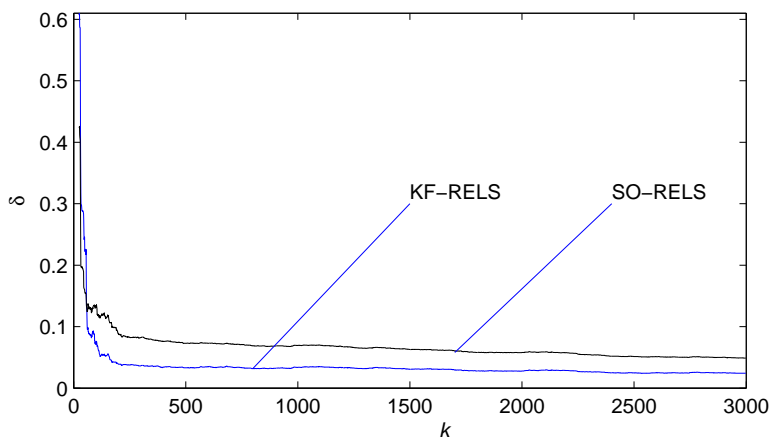


Figure 4. The SO-RLS and KF-RLS estimation errors  $\delta$  versus  $k$  ( $\sigma_w^2 = 0.02^2$ ,  $\sigma_v^2 = 0.20^2$ )

Take the same simulation conditions as those in Example 1 and let  $\sigma_w^2 = 0.02^2$  and  $\sigma_v^2 = 0.20^2$ . Using the KF-RELS algorithm to estimate the parameters, states and time-delay of this system, the parameter estimates and their estimation errors are shown in Table IV. Give a confidence interval  $[-0.005, 0.005]$  for the redundant parameters. Checking the last second line in Table IV, we can find there are two elements which can be assumed as zero. So the time-delay  $\tau$  equals 2.

As a comparison, use the state observer to estimate the unknown system states again. Combining it with the RELS algorithm to get the SO-RELS estimates. The parameter estimates and their

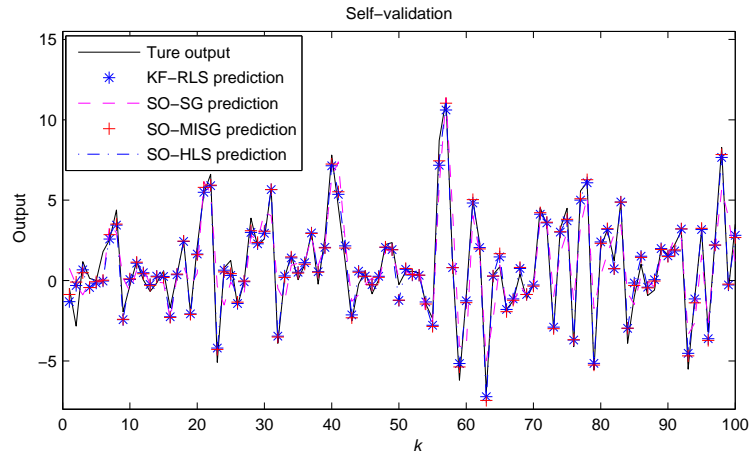


Figure 5. A section of the self-validation of different algorithms ( $\sigma_w^2 = 0.02^2$ ,  $\sigma_v^2 = 0.20^2$ )

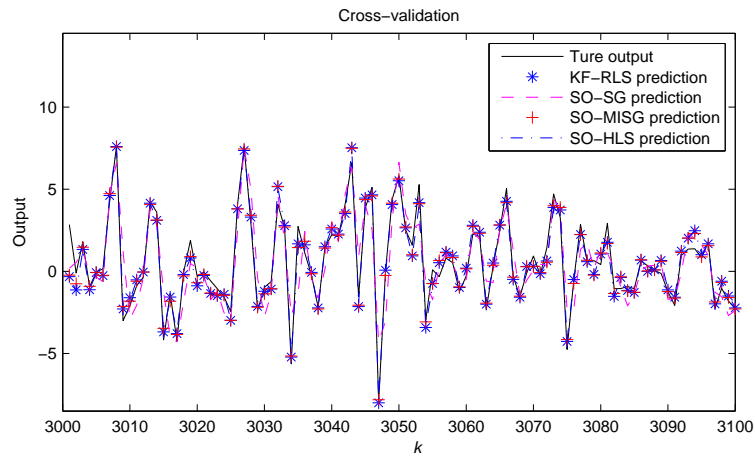


Figure 6. A section of the cross-validation of different algorithms ( $\sigma_w^2 = 0.02^2$ ,  $\sigma_v^2 = 0.20^2$ )

estimation errors are also shown in Table IV. From Table IV, we can see that the KF-RELS algorithm generates better estimates than the SO-RELS algorithm.

For model validation, a different dataset ( $L_e = 200$  samples from  $t = 3001$  to  $3200$ ) and the estimated model obtained by the KF-RELS algorithm are used. The predicted output and the true output under the KF-RELS algorithm are plotted in Figure 7. The true states and the estimated states are compared in Figure 8. From Figures 7–8, we can draw the following conclusions.

1. The model outputs can fit the true outputs well, the KF-RELS algorithm can capture the dynamic system well.
2. The estimated states are very close to the true states, the parameter estimates based Kalman filter can generate highly accurate state estimates.

## 6. CONCLUSIONS

A Kalman filter based least squares iterative algorithm and recursive least squares algorithm are developed for Hammerstein systems with dynamic state space model. The process noise is assumed to be white noise, the measurement noise is fitted by the moving average noise. For the unknown

Table IV. The SO-RELS and KF-RELS parameter estimates and errors versus  $k$

Algorithm	$k$	$a_{01}$	$a_{02}$	$a_1$	$a_2$	$\gamma_1$	$\gamma_2$	$b_2$	$d_1$	$\delta$ (%)
SO-RELS	100	0.02795	0.18176	0.48168	-0.20607	1.45273	-0.47187	1.51377	-0.27369	16.61618
	200	0.01253	0.15293	0.47055	-0.23135	1.43085	-0.41217	1.45089	-0.27925	13.90675
	500	0.00365	0.12384	0.47863	-0.25194	1.39700	-0.38491	1.40655	-0.28999	12.30822
	1000	-0.00232	0.10862	0.48803	-0.26634	1.36120	-0.36924	1.41128	-0.29300	10.67716
	2000	0.00346	0.08652	0.47877	-0.27692	1.33084	-0.35337	1.39472	-0.26988	8.79811
	3000	0.00633	0.07844	0.48158	-0.28087	1.32018	-0.35446	1.40312	-0.25708	7.95681
KF-RELS	100	0.03284	0.03973	0.49893	-0.20725	1.21083	-0.45434	1.30588	-0.32613	13.40018
	200	0.02832	0.04168	0.49727	-0.27377	1.26652	-0.37328	1.39482	-0.23716	6.44682
	500	0.00907	0.04078	0.49278	-0.29927	1.27048	-0.34336	1.42444	-0.20737	4.69370
	1000	-0.00230	0.02711	0.49244	-0.30931	1.23701	-0.32881	1.43377	-0.19160	3.06655
	2000	0.00198	0.00730	0.48081	-0.31664	1.21497	-0.31732	1.41960	-0.17097	2.49090
	3000	0.00442	-0.00050	0.48446	-0.31938	1.20665	-0.31978	1.42942	-0.16938	2.32384
True values		0.00000	0.00000	0.45000	-0.30000	1.20000	-0.32000	1.50000	-0.18000	

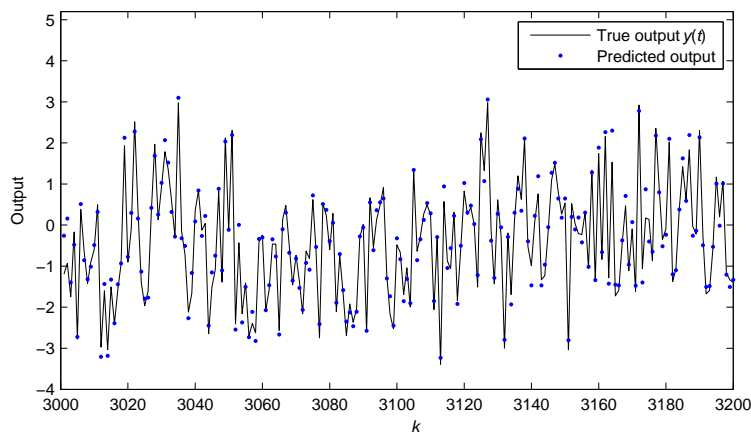


Figure 7. The KF-RELS predicted output and true output versus  $k$

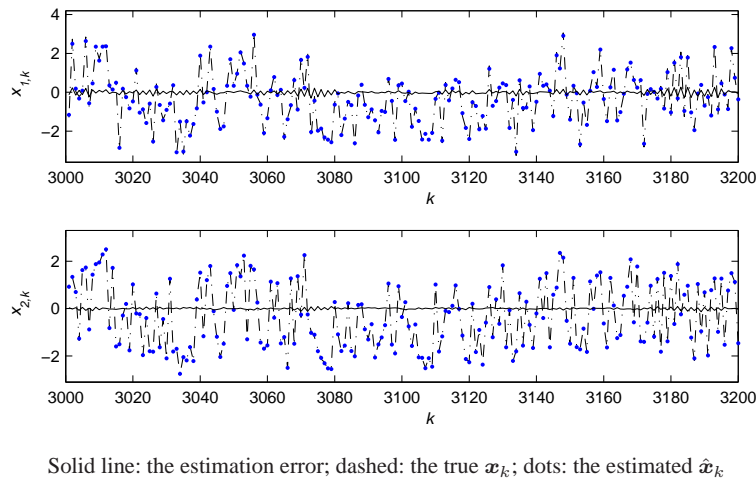


Figure 8. The state  $x_k$  and estimated estimate  $\hat{x}_k$  versus  $k$

time-delay in the output block, a Kalman filter based recursive extended least squares algorithm is derived. Based on the proposed algorithms, the combined state and parameter estimation are



obtained. The proposed algorithms can be extended to other system identification models, e.g., the linear state space models, and Wiener models [41, 42] and Hammerstein-Wiener systems [43, 44] or applied to nonlinear systems [45, 46] and other fields [47, 48, 49].

## 7. ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (No. 61273194), the Graduate Education Innovation Program of Jiangsu Province (No. KYLX16.0780) and the 111 Project (B12018).

## REFERENCES

1. van der Veen G, van Wingerden JW, Verhaegen M. Global identification of wind turbines using a Hammerstein identification method. *IEEE Transactions on Control Systems Technology* 2013; **21**(4): 1471–1478.
2. Wang J, Zhang Q. Detection of asymmetric control valve stiction from oscillatory data using an extended Hammerstein system identification method. *Journal of Process Control* 2014; **24**(1): 1–12.
3. Kim J, Konstantinou K. Digital predistortion of wideband signals based on power amplifier model with memory. *IEEE Electronics Letters* 2001; **37**: 1417–1418.
4. Zhang LM, Hua CC, Guan XP. Structure and parameter identification for Bayesian Hammerstein system. *Nonlinear Dynamics* 2015; **79**(3): 1847–1861.
5. Chen J, Wang XP. Identification of Hammerstein systems with continuous nonlinearity. *Information Processing Letters* 2015; **115**(11): 822–827.
6. Zhao S, Huang B, Liu F. Minimum variance unbiased FIR filter for discrete time-variant systems. *Automatica* 2015; **53**(3): 355–361.
7. Zhao S, Shmaliy YS, Liu F. Fast Kalman-like optimal unbiased FIR filtering with applications. *IEEE Transactions on Signal Processing* 2016; **64**(9): 2284–2297.
8. Wang DQ, Zhang W. Improved least squares identification algorithm for multivariable Hammerstein systems. *Journal of the Franklin Institute* 2015; **352**(11): 5292–5307.
9. Chen HB, Xiao YS, Ding F. Hierarchical gradient parameter estimation algorithm for Hammerstein nonlinear systems using the key term separation principle. *Applied Mathematics and Computation* 2014; **247**: 1202–1210.
10. Liu Y, Bai EW. Iterative identification of Hammerstein systems. *Automatica* 2007; **43**(2): 346–354.
11. Wang XX, Liang Y, Pan Q, Zhao CH, Yang F. Nonlinear gaussian smoothers with colored measurement noise. *IEEE Transactions on Automatic Control* 2015; **60**(3): 870–876.
12. Roopa S, Narasimhan SV, Babloo B. Steiglitz–McBride adaptive notch filter based on a variable-step-size LMS algorithm and its application to active noise control. *International Journal of Adaptive Control and Signal Processing* 2016; **30**(1): 16–30.
13. Xu L, Ding F. Recursive least squares and multi-innovation stochastic gradient parameter estimation methods for signal modeling. *Circuits, Systems and Signal Processing* 2017; **36**. doi: 10.1007/s00034-016-0378-4
14. Huang J, Shi Y, Huang HN, Li Z. 1-2-1-infinity filtering for multirate nonlinear sampled-data systems using T-S fuzzy models. *Digital Signal Processing* 2013; **23**(1): 418–426.
15. Wang YJ, Ding F. Novel data filtering based parameter identification for multiple-input multiple-output systems using the auxiliary model. *Automatica* 2016; **71**: 308–313.
16. Wang YJ, Ding F. The filtering based iterative identification for multivariable systems. *IET Control Theory and Applications* 2016; **10**(8): 894–902.
17. Wang YJ, Ding F. The auxiliary model based hierarchical gradient algorithms and convergence analysis using the filtering technique. *Signal Processing* 2016; **128**: 212–221.
18. Hu YB, Liu BL, Zhou Q, Yang C. Recursive extended least squares parameter estimation for Wiener nonlinear systems with moving average noises. *Circuits, Systems, and Signal Processing* 2014; **33**(2): 655–664.
19. Li JH. Parameter estimation for Hammerstein CARARMA systems based on the Newton iteration. *Applied Mathematics Letters* 2013; **26**(1): 91–96.
20. Ding J, Lin JX. Modified subspace identification for periodically non-uniformly sampled systems by using the lifting technique. *Circuits, Systems, and Signal Processing* 2014; **33**(5): 1439–1449.
21. Uzinski JC, Paiva HM, Duarte MA, Galvão RK, Villarreal F. A state-space description for perfect-reconstruction wavelet FIR filter banks with special orthonormal basis functions. *Journal of Computational and Applied Mathematics* 2015; **290**: 290–297.
22. Yan LP, Jiang L, Xia YQ, Fu MY. State estimation and data fusion for multirate sensor networks. *International Journal of Adaptive Control and Signal Processing* 2016; **30**(1): 3–15.
23. Ding F, Liu XM, Ma XY. Kalman state filtering based least squares iterative parameter estimation for observer canonical state space systems using decomposition. *Journal of Computational and Applied Mathematics* 2016; **301**: 135–143.
24. Zhang H, Shi Y, Wang JM. On energy-to-peak filtering for nonuniformly sampled nonlinear systems: a Markovian jump system approach. *IEEE Transactions on Fuzzy Systems* 2014; **22**(1): 212–222.
25. Zhao S, Huang B, Liu F. Linear optimal unbiased filter for time-variant systems without a priori information on initial condition. *IEEE Transactions on Automatic Control* 2016; doi: 10.1109/TAC.2016.2557999.



26. Ding F, Liu XM, Gu Y. An auxiliary model based least squares algorithm for a dual-rate state space system with time-delay using the data filtering. *Journal of the Franklin Institute* 2016; **353**(2): 398–408.
27. Chen L, Han LL, Huang B, Liu F. Parameter estimation for a dual-rate system with time delay. *ISA transactions* 2014; **53**(5): 1368–1376.
28. Xie L, Yang HZ. Gradient based iterative identification for nonuniform sampling output error systems, *Journal of Vibration and Control* 2011; **17**(3): 471–478.
29. Schön TB, Wills A, Ninness B. System identification of nonlinear state-space models. *Automatica* 2011; **47**(1): 39–49.
30. Deng J, Huang B. Identification of nonlinear parameter varying systems with missing output data. *AIChE Journal* 2012; **58**(11): 3454–3467.
31. Wang XH, Ding F. Recursive parameter and state estimation for an input nonlinear state space system using the hierarchical identification principle. *Signal Processing* 2015; **117**: 208–218.
32. Wang DQ, Ding F, Liu XM. Least squares algorithm for an input nonlinear system with a dynamic subspace state space model. *Nonlinear Dynamics* 2014; **75**(1-2): 49–61.
33. Wang XH, Ding F. Joint estimation of states and parameters for an input nonlinear state-space system with colored noise using the filtering technique. *Circuits, Systems, and Signal Processing* 2015; **35**(2): 481–500.
34. Vörös J. Modeling and parameter identification of systems with multi-segment piecewise-linear characteristics. *IEEE Transactions on Automatic Control* 2002; **47**(1): 184–188.
35. Vörös J. Recursive identification of Hammerstein systems with discontinuous nonlinearities containing dead-zones. *IEEE Transactions on Automatic Control* 2003; **48**(12): 2203–2206.
36. Pan J, Yang XH, Cai HF, Mu BX. Image noise smoothing using a modified Kalman filter. *Neurocomputing* 2016; **173**: 1625–1629.
37. Xu L. The damping iterative parameter identification method for dynamical systems based on the sine signal measurement, *Signal Processing* 120 (2016) 660–667.
38. Xu L. Application of the Newton iteration algorithm to the parameter estimation for dynamical systems. *Journal of Computational and Applied Mathematics* 2015; **288**: 33–43.
39. Xu L, Chen L, Xiong WL. Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration. *Nonlinear Dynamics* 2015; **79**(3): 2155–2163.
40. Xu L. A proportional differential control method for a time-delay system using the Taylor expansion approximation. *Applied Mathematics and Computation* 2014; **236**: 391–399.
41. Ding F, Liu XM, Liu MM. The recursive least squares identification algorithm for a class of Wiener nonlinear systems. *Journal of the Franklin Institute* 2016; **353**(7): 1518–1526.
42. Ding F, Wang XH, Chen QJ, Xiao YS. Recursive least squares parameter estimation for a class of output nonlinear systems based on the model decomposition. *Circuits, Systems and Signal Processing* 2016; **35**(9): 3323–3338.
43. Wang YJ, Ding F. Recursive least squares algorithm and gradient algorithm for Hammerstein-Wiener systems using the data filtering. *Nonlinear Dynamics* 2016; **84**(2): 1045–1053.
44. Wang YJ, Ding F. Recursive parameter estimation algorithms and convergence for a class of nonlinear systems with colored noise. *Circuits, Systems and Signal Processing* 2016; **35**(10): 3461–3481.
45. Li H, Shi Y, Yan W. On neighbor information utilization in distributed receding horizon control for consensus-seeking. *IEEE Transactions on Cybernetics* 2016; doi: 10.1109/TCYB.2015.2459719
46. Li H, Shi Y, Yan W. Distributed receding horizon control of constrained nonlinear vehicle formations with guaranteed  $\gamma$ -gain stability. *Automatica* 2016; **68**: 148–154.
47. Wang TZ, Qi J, Xu H, et al. Fault diagnosis method based on FFT-RPCA-SVM for cascaded-multilevel inverter. *ISA Transactions* 2016; **60**: 156–163.
48. Wang TZ, Wu H, Ni MQ, et al. An adaptive confidence limit for periodic non-steady conditions fault detection. *Mechanical Systems and Signal Processing* 2016; **72-73**: 328–345.
49. Feng L, Wu MH, Li QX, et al. Array factor forming for image reconstruction of one-dimensional nonuniform aperture synthesis radiometers. *IEEE Geoscience and Remote Sensing Letters* 2016; **13**(2): 237–241.