Parameter tracking of time-varying Hammerstein-Wiener Systems

Article

Accepted Version

Feng, Y. and Hong, X. ORCID: https://orcid.org/0000-0002-6832-2298 (2021) Parameter tracking of time-varying Hammerstein-Wiener Systems. International Journal of Systems Science, 52 (16). pp. 3478-3492. ISSN 0020-7721 doi: 10.1080/00207721.2021.1931546 Available at https://centaur.reading.ac.uk/99204/

To link to this article DOI: http://dx.doi.org/10.1080/00207721.2021.1931546

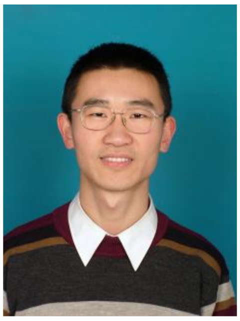
Publisher: Taylor & Francis

Taylor & Francis Word Template for journal articles

Feng Yu[a]* and Xia Hong[b]

[a] *College of Information Science and Engineering, Northeastern University, Shenyang, China, e-mail: yufeng@ise.neu.edu.cn.;*

[b] *Department of Computer Science, School of Mathematical, Physical and Computational Sciences, University of Reading, Reading, U.K.*



Feng Yu received his university education at Northeastern University, P.R. China (BSc, 2007, MSc, 2009, PhD, 2015), all in control theory and control engineering. He is currently a lecture at College of Information Science and Engineering, Northeastern University. His research interests include nonlinear systems identification, control and their applications.

Xia Hong received her university education at National University of Defense Technology, P.R. China (BSc, 1984, MSc, 1987), and University of Sheffield, UK (PhD, 1998), all in automatic control. She worked as a research assistant in Beijing Institute of Systems Engineering, Beijing, China from 1987-1993. She worked as a research fellow in the Department of Electronics and Computer Science at University of Southampton from 1997-2001. She is currently a Professor at Department of Computer Science, School of Mathematical and Physical Sciences, University of Reading. She is actively engaged in research into nonlinear systems identification, data modelling, estimation and intelligent control, neural networks, pattern recognition, learning theory and their applications. She has published over 200 research papers, and co-authored a research book. Professor Hong was awarded a Donald Julius Groen Prize by IMechE in 1999.

Parameter Tracking of Time-varying Hammerstein-Wiener Systems

Abstract: A two-stage identification algorithm is introduced for tracking the parameters in time-varying Hammerstein-Wiener systems. The Kalman filtering algorithm and parameter separation technique are employed in the proposed algorithm. The convergence analysis of this two-stage algorithm is provided. It is shown that the proposed algorithm can guarantee the boundedness of the parameter estimation error. Four simulation examples, including a practical system application of electric arc furnace, have been employed to validate the effectiveness of the proposed approaches, for a range of simulated time-varying characteristics.

Keywords: Hammerstein-Wiener systems; time-varying; two-stage identification; tracking

1 Introduction

Some complex industrial systems are composed of several subsystems as blocks, with their subsystems cascaded in series, each of which has different characteristics, either linear dynamic or nonlinear static. The systems such as Hammerstein (H) systems and Wiener (W) systems are two typical nonlinear block-oriented systems. There have been extensive researches on identification of block-oriented systems, including for H systems, the identification algorithms are reported in Wang et al. (2018), Hong et al. (2012) and (2007), Liu et al. (2019) and Zhang et al. (2017); for W systems, in Mu et al. (2013), Ding et al. (2016), Hong et al. (2013), Giri et al. (2013) and Li et al. (2017); and for the more complex Hammerstein-Wiener (H-W) systems, which are the combination of H and W systems, in Wang et al. (2012), Wills et al. (2013), Yu et al. (2013, 2014 and 2017) and Voros (2014).

The time-varying system refers to the fact that its characteristics change as operating conditions changes, e.g. with the variation of system status, working environment or operation mode (Astrom et al., 2013). One way of time-varying system modelling is to treat the system as having time-varying parameters (Chen et al., 2021). While block-oriented systems have been subjected to extensive research, much less attention has been devoted on identifying the time-varying block-oriented systems. The work of Nordsjo et al. (2001) tracked the time-varying block-oriented systems through extended Kalman filter (KF). It also considered the estimation of the variance of the measurement noise. The work of Bershad et al. (2000) identified the time-varying W systems through stochastic gradient method. The recursive least square (RLS) with forgetting factor was used in the works of Voros (2005, 2011, 2013 and 2017) to estimate the time-varying parameters of different H or W systems, such as H systems with piecewise linear block containing time-varying parameter or time-varying backlash block, W systems with time-varying hysteresis output part. The work of Kobayashi et al. (2010) studied identifying time-varying W systems. The optimization technique and neural networks were used in the identification process. These contributions all focused on tracking the simple H or W systems. For the H-W systems, there has fewer relevant research published as the problem is more challenging due to structure complexity. The work of Voros (2018) discussed on tracking a special time-varying H-W systems, of which the nonlinear part has backlash characteristics. Its parametric system model can be represented as the product form of time-varying parameters and information terms. The RLS method with forgetting factor can be used for the tracking problem. For the more general cases, the work of Yu et al. (2020) used a modified extended KF algorithm to track a more general H-W systems with time-varying parameters. It also presented a convergence analysis for the proposed algorithm. From these studies, it is suggested that the parametric model of H systems can be expressed in linear form. The RLS algorithm could be used to obtain the parameter estimates directly. For the W systems and H-W systems, nonlinear methods, such as extended KF algorithm, the method combining RLS and key term separation principle, etc., are proposed as their parametric models are nonlinear. However, these methods are all locally convergent and cannot guarantee the global convergence of parameter estimation. In addition, some methods have no strict convergence analysis. Users may not necessarily understand what conditions should be satisfied to ensure the convergence of the algorithm (Wang et al., 2019; 2021). All these restrict the application of the algorithms. It can be seen that there is still a lot of work to be done for tracking time-varying block-oriented systems.

The research on time-varying H-W systems is seldom seen, which is the topic of the paper. Tracking changes in systems is important because it not only can estimate system output in real time but also can be combined with on-line control strategies to produce adaptive control algorithms for improved performance (Voros, 2005). Compared with the general black-box modelling, such as artificial neural network modelling, the identification of block systems is more complex. This is because not only the accurate estimate of system output, but also the characteristics of each subsystem are required in the identification process. Different from other basic block systems, such as H systems and W systems, H-W systems contain two nonlinear sub blocks. Stronger nonlinear characteristic makes identification more difficult (Chen et al., 2018).

This paper studies on tracking general time-varying H-W systems with random varying parameters. Based on the parametric system model, a two-stage identification algorithm is proposed to identify the time-varying parameters. This identification strategy has been applied to off-line identifying time-invariant H-W systems (Bai et al. 1998). In this paper, it will be used to track the time-varying H-W systems. In the first stage, time-varying parameter product terms in the system model are tracked by the Kalman filtering (KF) algorithm. In the second stage, the identified estimates of parameter product terms are separated to obtain the system model parameter estimates. The average method (AVE) and the singular value decomposition (SVD) method are introduced to separate the parameter. The convergence of the algorithm is also investigated with two parts. In the first part, it is proved that the KF algorithm is global convergent for the estimation of the parameter product terms in H-W model and it can achieve tracking with bounded error. Furthermore, it is proved in the second part that the estimation error of original system parameter is bounded either AVE method or SVD method is used. Four simulation examples are given to show the validity of the proposed algorithm.

The contributions of the paper are summarized as follows:

1. The varying parameters of time-varying H-W systems can be tracked by the proposed two-stage identification algorithm.

2. The convergence of the algorithm does not depend on the initial parameter estimates. If the persistent excitation condition is satisfied, the algorithm can achieve global convergence.

3. In addition, to the case of random varying parameters, the proposed algorithm can also track the H-W systems with other kinds of time-varying parameters, which further extends its application.

The remainder of the paper is organized as follows. Section II provides H-W mathematical formulation. Section III introduces the proposed tracking algorithm, followed by convergence analysis in Section IV. Numerical examples are given in Section V and Section VI is devoted to conclusions.

2 Problem formulation

Fig. 1 shows the structure of the studied H-W systems.

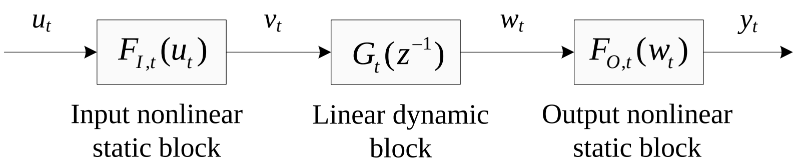


Fig .1 Time-varying H-W systems

$G_t(z^{-1})$ is the time-varying linear block transfer function; z^{-1} is the backward shift operator; $F_{I,t}(u_t)$ is the function of time-varying input nonlinear block; $F_{O,t}(w_t)$ is the function of time-varying output nonlinear block; u_t is the system input; y_t is the system output; v_t and w_t are the internal variables, both of them are immeasurable.

The following time-varying difference function with ARMA form is used to describe the dynamic characteristic of linear block:

$$w_t = -\sum_{i=1}^{r} a_{i,t} w_{t-i} + \sum_{j=1}^{m} b_{j,t} v_{t-j}, \tag{1}$$

where r and m are the system orders of linear block; $a_{i,t}$ and $b_{i,t}$ are the time-varying parameters and

$$a_{i,t+1} = a_{i,t} + \delta_{a_i,t+1}, \tag{2}$$

$$b_{i,t+1} = b_{i,t} + \delta_{b_i,t+1}, \tag{3}$$

where $\delta_{a_i,t+1}$ and $\delta_{b_i,t+1}$ are the time-variations of $a_{i,t}$ and $b_{i,t}$, respectively.

The input nonlinearity is modelled by a linear combination of basis function $f_{I,i}$ with time-varying parameters $c_{i,t}$:

$$v_t = F_{I,t}(u_t) = \sum_{i=1}^{p} c_{i,t} f_{I,i}(u_t), \tag{4}$$

where p is the number of basis function $f_{I,i}$.

In order to estimate the unavailable internal variable w_t, the inverse of the output nonlinearity is required, which is also modelled by a linear combination of basis function $f_{O,i}$ with time-varying parameters $d_{i,t}$, namely:

$$w_t = F_{O,t}^{-1}(y_t) = \sum_{i=1}^{q} d_{i,t} f_{O,i}(y_t), \tag{5}$$

where q is the number of basis function $f_{O,i}$.

Similar to the linear time-varying parameters, we have

$$c_{i,t+1} = c_{i,t} + \delta_{c_i,t+1} \tag{6}$$

$$d_{i,t+1} = d_{i,t} + \delta_{d_i,t+1} \tag{7}$$

where $\delta_{c_i,t+1}$ and $\delta_{d_i,t+1}$ are the time-variations of $c_{i,t}$ and $d_{i,t}$, respectively.

The assumptions for the H-W systems are adopted as follows:

Assumption 1. The input and output signal and the system parameters are bounded;

Assumption 2. The orders of linear block r and m are known, as well as the basic functions $f_{I,i}$ and $f_{O,i}$ and their numbers p and q in nonlinear blocks;

Assumption 3. The output nonlinearity is one-to-one within the input and output data so that the inverse of $F_{O,t}(w_t)$ exists;

Assumption 4. The values of $c_{1,t}$ and $d_{1,t}$ are set to be 1 in the system parameterization. They are not updated in the identification process;

Assumption 5. The time-variation parameters $\delta_{a_i,t}$, $\delta_{b_i,t}$, $\delta_{c_i,t}$ and $\delta_{d_i,t}$ are modelled as stochastic processes. Furthermore,

$$E\left(\delta_{i,h}\right) = 0 \tag{9}$$

and

$$E\left(\delta_{i,h}\delta_{j,k}\right) = \begin{cases} \tau^2 < \infty, & i = j \text{ and } h = k \\ 0, & otherwise \end{cases} \tag{10}$$

Assumption 1 is a traditional assumption for system identification. If the linear part is unstable, the system input must keep the output bounded. **Assumption 2** implies that the algorithm proposed in this paper aims to identify the system parameter rather than estimates the model structure (Bai et al. 2002). This shows that in the application of the algorithm, certain prior knowledge of the system structure is required. The order of the model should be known. **Assumption 3** makes the parameterization of the inverse of output nonlinearity available. This is a special assumption for the H-W system identification problem (Wang et al., 2012 and Bai et al., 1998 and 2002). The H-W systems with reversible output nonlinearity are widespread in practice, although this assumption limits the application of the algorithm for the H-W systems with irreversible output nonlinearity. **Assumption 4** guarantees the consistency of parameter estimation (Yu et al., 2013 and 2014 and Bai 1998). **Assumption 5** shows the time-varying characteristics of the system, which can be called “random walk”. It is a common method to model the time-varying parameters in the time-varying system (Ljung et al., 1990).

To facilitate the identification of model parameters, the parametric model of the H-W systems is given. Notice that $c_{1,t}$ and $d_{1,t}$ are set to be 1, define

$$\theta_t^T = \big[a_{1,t}, \cdots, a_{r,t}, b_{1,t}, \cdots, b_{m,t}, d_{2,t}, \cdots, d_{q,t}, b_{1,t}c_{2,t}, \cdots, b_{m,t}c_{2,t},$$
$$b_{1,t}c_{p,t}, \cdots, b_{m,t}c_{p,t}, a_{1,t}d_{2,t}, \cdots, a_{r,t}d_{2,t}, a_{1,t}d_{q,t}, \cdots, a_{r,t}d_{q,t} \big]$$

and

$$\varphi_t^T = \big[f_{O,1}(y_{t-1}), \cdots, f_{O,1}(y_{t-r}), f_{I,1}(u_{t-1}), \cdots, f_{I,1}(u_{t-m}), f_{O,2}(y_t), \cdots, f_{O,q}(y_t),$$
$$f_{I,2}(u_{t-1}), \cdots, f_{I,2}(u_{t-m}), \cdots, f_{I,p}(u_{t-1}), \cdots, f_{I,p}(u_{t-m}),$$
$$f_{O,2}(y_{t-1}), \cdots, f_{O,2}(y_{t-r}), \cdots, f_{O,q}(y_{t-1}), \cdots, f_{O,q}(y_{t-r}) \big].$$

Therefore, we obtain the parametric model of H-W systems as follows

$$\theta_t^T \varphi_t = -f_{O,1}(y_t). \tag{11}$$

3 Parameter estimation algorithm

In this section, a two-stage recursive identification algorithm is proposed to track the time-varying parameter vectors $a_{i,t}$, $b_{i,t}$, $c_{i,t}$ and $d_{i,t}$ in the parameterized model (11).

For the first stage, tracking the time-varying parameter vector θ_t is considered. As θ_t is not the original model parameter, a study on it is given. Define $\theta_t^T = \big[\theta_{s,t}^T, \theta_{d,t}^T \big]$, where

$$\theta_{s,t}^T = \big[a_{1,t}, \cdots, a_{r,t}, b_{1,t}, \cdots, b_{m,t}, d_{2,t}, \cdots, d_{q,t} \big];$$

$$\theta_{d,t}^T = \big[b_{1,t}c_{2,t}, \cdots, b_{m,t}c_{2,t}, \cdots, b_{1,t}c_{p,t}, \cdots, b_{m,t}c_{p,t}, a_{1,t}d_{2,t}, \cdots, a_{r,t}d_{2,t}, \cdots, a_{1,t}d_{q,t}, \cdots, a_{r,t}d_{q,t} \big].$$

From the definition of $\theta_{d,t}^T$, it is obtained:

$$\theta_{d,t}^T = \theta_{dl,t}^T \odot \theta_{dr,t}^T, \tag{12}$$

where

$$\theta_{dl,t}^T = \big[b_{1,t}, \cdots, b_{m,t}, \cdots, b_{1,t}, \cdots, b_{m,t}, a_{1,t}, \cdots, a_{r,t}, \cdots, a_{1,t}, \cdots, a_{r,t} \big];$$

$$\theta_{dr,t}^T = \big[c_{2,t}, \cdots, c_{2,t}, \cdots, c_{p,t}, \cdots, c_{p,t}, d_{2,t}, \cdots, d_{2,t}, \cdots, d_{q,t} \cdots, d_{q,t} \big].$$

Define

$$\delta_{dl,t}^T = \big[\delta_{b_1,t}, \cdots, \delta_{b_m,t}, \cdots, \delta_{b_1,t}, \cdots, \delta_{b_m,t}, \delta_{a_1,t}, \cdots, \delta_{a_r,t}, \cdots, \delta_{a_1,t}, \cdots, \delta_{a_r,t} \big]$$

and

$$\delta_{dr,t}^T = \big[\delta_{c_2,t}, \cdots, \delta_{c_2,t}, \cdots, \delta_{c_p,t}, \cdots, \delta_{c_p,t}, \delta_{d_2,t}, \cdots, \delta_{d_2,t}, \cdots, \delta_{d_q,t}, \cdots, \delta_{d_q,t} \big].$$

Therefore, we have

$$\begin{aligned}\theta_{d,t+1}^T &= \theta_{dl,t+1}^T \odot \theta_{dr,t+1}^T = (\theta_{dl,t}^T + \delta_{dl,t+1}^T) \odot (\theta_{dr,t}^T + \delta_{dr,t+1}^T) \\ &= \theta_{d,t}^T + \theta_{dl,t}^T \odot \delta_{dr,t+1}^T + \delta_{dl,t+1}^T \odot \theta_{dr,t}^T + \delta_{dl,t+1}^T \odot \delta_{dr,t+1}^T = \theta_{d,t}^T + h_{t+1}^T,\end{aligned} \tag{13}$$

where $h_{t+1}^T = \theta_{dl,t}^T \odot \delta_{dr,t+1}^T + \delta_{dl,t+1}^T \odot \theta_{dr,t}^T + \delta_{dl,t+1}^T \odot \delta_{dr,t+1}^T$.

Define $\delta_{s,t}^T = \left[\delta_{a_1,t}, \cdots, \delta_{a_r,t}, \delta_{b_1,t}, \cdots, \delta_{b_m,t}, \delta_{d_2,t}, \cdots, \delta_{d_q,t}\right]$ and $H_{t+1}^T = \left[\delta_{s,t+1}^T, h_{t+1}^T\right]$. And then

$$\theta_{t+1} = \begin{bmatrix}\theta_{s,t+1} \\ \theta_{d,t+1}\end{bmatrix} = \begin{bmatrix}\theta_{s,t} \\ \theta_{d,t}\end{bmatrix} + \begin{bmatrix}\delta_{s,t+1} \\ h_{t+1}\end{bmatrix} = \theta_t + H_{t+1} \tag{14}$$

From (11), it can be seen that θ_t is only composed of unknown parameters and φ_t only contains the items regard to the measurements of system input and output. Therefore, θ_t can be identified by the KF algorithm. The following algorithm is proposed:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + K_t e_t, \tag{15}$$

$$K_t = -(P_t + Q^2)\varphi_t[\varphi_t^T(P_t + Q^2)\varphi_t + R^2]^{-1}, \tag{16}$$

$$P_{t+1} = P_t + Q^2 + K_t\varphi_t^T(P_t + Q^2). \tag{17}$$

where $\hat{\theta}_t$ is the estimate of θ_t; e_t is the model prediction error; K_t is the update gain of $\hat{\theta}_t$; P_t is the estimate of $E\left[\tilde{\theta}_t\tilde{\theta}_t^T\right]$, where $\tilde{\theta}_t$ is the estimation error of θ_t; Q^2 is the estimate of $E\left[H_t H_t^T\right]$; R^2 is a positive real number.

Remark 1. In the traditional identification algorithm, e_t is defined as the difference between the measurement of system output and its estimate(Chen et al., 2020a). However, there is no estimate of system output in the system model (11). Considering the special form of (11), define e_t as follows:

$$e_t = \hat{\theta}_t^T\varphi + f_{O,1}\left(y_t\right). \tag{18}$$

Remark 2. P_t is updated by (17) recursively at each sampling time. The initial value of P_t can be set to

$$P_0 = l_1 I \tag{19}$$

where I is identity matrix; l_1 is a large positive real number. It is obvious that if equation (19) is used to initialize P_0, P_t will be symmetric positive definite at any sampling time (Chen et al., 2020b). Q^2 is set as follows:

$$Q^2 = l_2 I \tag{20}$$

where l_2 is a positive real number. R^2 can be set as a small value.

From the algorithm (15)-(17), the estimate of θ_t can be obtained. However, the aim of the paper is to identify the time-varying parameter $a_{i,t}$, $b_{i,t}$, $c_{i,t}$ and $d_{i,t}$. Therefore, for the second stage, separate the parameters in θ_t is considered to obtain the original system parameter estimates $\hat{a}_t$, $\hat{b}_t$, $\hat{c}_t$ and $\hat{d}_t$. The following two methods are introduced.

Average (AVE) method

Notice that the parameter sets $a_{i,t}$, $c_{i,t}$ and $d_{i,t}$ exist independently in θ_t. Therefore, the estimates of these parameters can be obtained from $\hat{\theta}_t$ directly. The following AVE methods can be used to obtain the estimates of $b_{i,t}$:

$$\hat{b}_{i,t} = \frac{1}{p}\sum_{j=1}^{p}\frac{b_{i,t}c_{j,t}}{\hat{c}_{j,t}}, \tag{21}$$

where $i = 1, \cdots, m$.

Remark 3. $\hat{b}_{i,t}$ can also be obtain by any $b_{i,t}c_{j,t}/\hat{c}_{j,t}$. However, as there are p terms of $b_{i,t}c_{j,t}/\hat{c}_{j,t}$, different solutions may be obtained due to estimation errors of $\hat{\theta}_t$. Therefore, the average value as shown in equation (21) is used to calculate $\hat{b}_{i,t}$. On the one hand, this method can give a unique estimate; on the other hand, it eliminates some influence on $\hat{b}_{i,t}$ caused by random errors of $\hat{\theta}_t$.

Singular value decomposition (SVD) method

Note that θ_t has a special structure in terms of $a_{i,t}$, $b_{i,t}$, $c_{i,t}$ and $d_{i,t}$. The SVD method can be used to separate them. Define

$$\theta_{ad,t} = \begin{bmatrix} a_{1,t}d_{2,t} & a_{1,t}d_{3,t} & \cdots & a_{1,t}d_{q,t} \\ a_{2,t}d_{2,t} & a_{2,t}d_{3,t} & \cdots & a_{2,t}d_{q,t} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r,t}d_{2,t} & a_{r,t}d_{3,t} & \cdots & a_{r,t}d_{q,t} \end{bmatrix}$$

and

$$\theta_{bc,t} = \begin{bmatrix} b_{1,t}c_{2,t} & b_{1,t}c_{3,t} & \cdots & b_{1,t}c_{p,t} \\ b_{2,t}c_{2,t} & b_{2,t}c_{3,t} & \cdots & b_{2,t}c_{p,t} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,t}c_{2,t} & b_{m,t}c_{3,t} & \cdots & b_{m,t}c_{p,t} \end{bmatrix}.$$

The estimates of $\theta_{ad,t}$ and $\theta_{bc,t}$ can be obtained from $\hat{\theta}_t$. Let

$$\hat{\theta}_{ad,t} = \sum_{i=1}^{\min(r,q)} \sigma_i \eta_i \nu_i^T, \tag{22}$$

$$\hat{\theta}_{bc,t} = \sum_{i=1}^{\min(m,p)} \rho_i \chi_i \gamma_i^T , \tag{23}$$

where σ_i and ρ_i are the singular values of $\hat{\theta}_{ad,t}$ and $\hat{\theta}_{bc,t}$, respectively. They are arranged to be nonnegative and in order of decreasing magnitude; η_i, v_i, χ_i and γ_i are r, q, m and p dimensional singular column vectors, respectively. Then the estimates of A_t, B_t, C_t and D_t can be computed by

$$\hat{A}_t = \sigma_1 v_1 , \tag{24}$$

$$\hat{B}_t = \rho_1 \gamma_1 , \tag{25}$$

$$\hat{C}_t = \chi_1 , \tag{26}$$

$$\hat{D}_t = \eta_1 , \tag{27}$$

where $A_t^T = \left[a_{1,t}, \cdots, a_{r,t}\right]$; $B_t^T = \left[b_{1,t}, \cdots, b_{m,t}\right]$; $C_t^T = \left[c_{1,t}, \cdots, c_{p,t}\right]$; $D_t^T = \left[d_{1,t}, \cdots, d_{q,t}\right]$.

Remark 4. As an online algorithm, the complexity of the algorithm must be considered. Ave method needs less computation. It is easy to combine with KF to realize online application. The time complexity of SVD is $O(n^3)$, where n is the maximum number of rows and columns of the matrix. If n is large, it consumes a lot of computation. For the H-W systems we studied, most of the linear dynamic characteristics can be described by the model shown in equation (1) with the order less than 5. That means r and m will not be too large. In addition, a few basis functions are required to fit the nonlinear characteristics if the system works stably. Therefore, the orders of $\hat{\theta}_{ad,t[r\times q]}$ and $\hat{\theta}_{bc,t[m\times p]}$ will not be too high. SVD can also be combined with KF to realize online identification in practical application.

4 Convergence analysis

Define $\tilde{\theta}_t$ as the parameter estimation error:

$$\tilde{\theta}_t = \hat{\theta}_t - \theta_t . \tag{28}$$

In order to analyse the convergence of the algorithm strictly, the following lemma about H_t is given.

Lemma 1. If **Assumption 1** and **Assumption 5** are satisfied, then there is a positive real number $\bar{H}$ such that $E\left(H_{t+1}^T H_{t+1} \middle| \tilde{\theta}_t\right)$ is bounded via

$$E\left(H_{t+1}^T H_{t+1} \middle| \tilde{\theta}_t\right) \le \bar{H}^2. \tag{29}$$

And it also have

$$E\left(H_{t+1}^T \middle| \tilde{\theta}_t\right) = \left(0, \cdots, 0\right). \tag{30}$$

Proof:

$$\begin{aligned}
& E\left(h_{t+1}^T h_{t+1} \middle| \tilde{\theta}_t\right) \\
&= E\Big(\left[\theta_{dl,t}^T \odot \delta_{dr,t+1}^T + \delta_{dl,t+1}^T \odot \theta_{dr,t}^T + \delta_{dl,t+1}^T \odot \delta_{dr,t+1}^T\right] \\
&\quad \left[\theta_{dl,t} \odot \delta_{dr,t+1} + \delta_{dl,t+1} \odot \theta_{dr,t} + \delta_{dl,t+1} \odot \delta_{dr,t+1}\right] \Big| \tilde{\theta}_t\Big) \\
&= E\left(\left[\theta_{dl,t}^T \odot \delta_{dr,t+1}^T\right]\left[\theta_{dl,t} \odot \delta_{dr,t+1}\right] \middle| \tilde{\theta}_t\right) + E\left(\left[\theta_{dl,t}^T \odot \delta_{dr,t+1}^T\right]\left[\delta_{dl,t+1} \odot \theta_{dr,t}\right] \middle| \tilde{\theta}_t\right) \\
&\quad + E\left(\left[\theta_{dl,t}^T \odot \delta_{dr,t+1}^T\right]\left[\delta_{dl,t+1} \odot \delta_{dr,t+1}\right] \middle| \tilde{\theta}_t\right) + E\left(\left[\delta_{dl,t+1}^T \odot \theta_{dr,t}^T\right]\left[\theta_{dl,t} \odot \delta_{dr,t+1}\right] \middle| \tilde{\theta}_t\right) \\
&\quad + E\left(\left[\delta_{dl,t+1}^T \odot \theta_{dr,t}^T\right]\left[\delta_{dl,t+1} \odot \theta_{dr,t}\right] \middle| \tilde{\theta}_t\right) + E\left(\left[\delta_{dl,t+1}^T \odot \theta_{dr,t}^T\right]\left[\delta_{dl,t+1} \odot \delta_{dr,t+1}\right] \middle| \tilde{\theta}_t\right) \\
&\quad + E\left(\left[\delta_{dl,t+1}^T \odot \delta_{dr,t+1}^T\right]\left[\theta_{dl,t} \odot \delta_{dr,t+1}\right] \middle| \tilde{\theta}_t\right) + E\left(\left[\delta_{dl,t+1}^T \odot \delta_{dr,t+1}^T\right]\left[\delta_{dl,t+1} \odot \theta_{dr,t}\right] \middle| \tilde{\theta}_t\right) \\
&\quad + E\left(\left[\delta_{dl,t+1}^T \odot \delta_{dr,t+1}^T\right]\left[\delta_{dl,t+1} \odot \delta_{dr,t+1}\right] \middle| \tilde{\theta}_t\right).
\end{aligned} \tag{31}$$

From equation (10) in **Assumption 5**, it is obtained that for any $i \ne j$ and $h \ne k$, $E\left(\delta_{i,h}\delta_{j,k}\right) = 0$. It is also noticed that $\theta_{dr,t}$ does not depend on $\delta_{dl,t+1}$ and $\theta_{dl,t}$ does not depend on $\delta_{dr,t+1}$, namely $E(\delta_{dl,t+1} \odot \theta_{dr,t} | \tilde{\theta}_t) = 0$ and $E(\delta_{dr,t+1} \odot \theta_{dl,t} | \tilde{\theta}_t) = 0$. Therefore, the following terms

$$E([\theta_{dl,t}^T \odot \delta_{dr,t+1}^T][\delta_{dl,t+1} \odot \theta_{dr,t}] | \tilde{\theta}_t),$$

$$E([\theta_{dl,t}^T \odot \delta_{dr,t+1}^T][\delta_{dl,t+1} \odot \delta_{dr,t+1}] | \tilde{\theta}_t),$$

$$E([\delta_{dl,t+1}^T \odot \theta_{dr,t}^T][\theta_{dl,t} \odot \delta_{dr,t+1}] | \tilde{\theta}_t),$$

$$E([\delta_{dl,t+1}^T \odot \theta_{dr,t}^T][\delta_{dl,t+1} \odot \delta_{dr,t+1}] | \tilde{\theta}_t),$$

$$E([\delta_{dl,t+1}^T \odot \delta_{dr,t+1}^T][\theta_{dl,t} \odot \delta_{dr,t+1}] | \tilde{\theta}_t)$$

and

$$E([\delta_{dl,t+1}^T \odot \delta_{dr,t+1}^T][\delta_{dl,t+1} \odot \theta_{dr,t}] | \tilde{\theta}_t)$$

are all equal to 0. Therefore,

$$
\begin{aligned}
&E(h_{t+1}^T h_{t+1} \big| \tilde{\theta}_t) \\
&= E([\theta_{dl,t}^T \odot \delta_{dr,t+1}^T][\theta_{dl,t} \odot \delta_{dr,t+1}] \big| \tilde{\theta}_t) + E([\delta_{dl,t+1}^T \odot \theta_{dr,t}^T][\delta_{dl,t+1} \odot \theta_{dr,t}] \big| \tilde{\theta}_t) \\
&\quad + E([\delta_{dl,t+1}^T \odot \delta_{dr,t+1}^T][\delta_{dl,t+1} \odot \delta_{dr,t+1}] \big| \tilde{\theta}_t) \\
&= \tau^2 \theta_{dl,t}^T \theta_{dl,t} + \tau^2 \theta_{dr,t}^T \theta_{dr,t} + [mp + (q-1)r]\tau^4.
\end{aligned} \tag{32}
$$

According to the definition of h_t and **Assumption 5**, it follows that

$$E(h_{t+1}^T \big| \tilde{\theta}_t) = (0, \cdots, 0). \tag{33}$$

In addition, we have

$$
\begin{aligned}
&E(\delta_{s,t+1}^T \delta_{s,t+1} \big| \tilde{\theta}_t) \\
&= E([\delta_{a_1,t+1}, \cdots, \delta_{a_r,t+1}]^T [\delta_{a_1,t+1}, \cdots, \delta_{a_r,t+1}] \big| \tilde{\theta}_t) \\
&\quad + E([\delta_{b_1,t+1}, \cdots, \delta_{b_m,t+1}]^T [\delta_{b_1,t+1}, \cdots, \delta_{b_m,t+1}] \big| \tilde{\theta}_t) \\
&\quad + E([\delta_{d_2,t+1}, \cdots, \delta_{d_q,t+1}]^T [\delta_{d_2,t+1}, \cdots, \delta_{d_q,t+1}] \big| \tilde{\theta}_t) \\
&= (r + m + q - 1)\tau^2.
\end{aligned} \tag{34}
$$

According to the definition of h_t and equation (9) in **Assumption 5**, it is obvious that

$$E(\delta_{s,t+1}^T \big| \tilde{\theta}_t) = (0, \cdots, 0). \tag{35}$$

From (32) and (34), it follows that

$$
\begin{aligned}
&E(H_{t+1}^T H_{t+1} \big| \tilde{\theta}_t) \\
&= E(\delta_{s,t+1}^T \delta_{s,t+1} \big| \tilde{\theta}_t) + E(h_{t+1}^T h_{t+1} \big| \tilde{\theta}_t) \\
&= (\theta_{dl,t}^T \theta_{dl,t} + \theta_{dr,t}^T \theta_{dr,t} + r + m + q - 1)\tau^2 + [mp + (q-1)r]\tau^4.
\end{aligned} \tag{36}
$$

According to **Assumption 1** and **Assumption 5**, τ^2 and all the system parameter are finite. Therefore, there must be a positive real number $\bar{H}$ such that

$$\bar{H}^2 \geq (\theta_{dl,t}^T \theta_{dl,t} + \theta_{dr,t}^T \theta_{dr,t} + r + m + q - 1)\tau^2 + [mp + (q-1)r]\tau^4. \tag{37}$$

The conclusion of (29) is obtained.

From (33) and (35), the conclusion of (30) is obtained.□

For further analysis, the following two lemmas are given.

Lemma 2 (Deyst et al., 1968). Assume there is a stochastic process $L_t(\tilde{\theta}_t)$ and the following positive real numbers β, $\underline{l}$, $\bar{l}$ and $0 < \alpha < 1$ such that

$$\underline{l} \,\| \tilde{\theta}_t \|_2^2 \leq L_t(\tilde{\theta}_t) \leq \bar{l} \,\| \tilde{\theta}_t \|_2^2 \tag{38}$$

and

$$E(L_{t+1} \big| \tilde{\theta}_t) - L_t \leq -\alpha L_t + \beta \tag{39}$$

are satisfied for all the solution of (15). Then the stochastic process is exponentially bounded in mean square and bounded with probability one.

Lemma 3 (Reif et al., 1999). Let the following conditions hold.

1) There are positive real numbers $\underline{r}$, $\bar{r}$, $\underline{q}$ and $\bar{q}$ such that R^2 and Q^2 in (16) and (17) are bounded by

$$\underline{r} \le R^2 \le \bar{r} \tag{40}$$

$$\underline{q}I \le Q^2 \le \bar{q}I, \tag{41}$$

2) There is a finite positive integer N and positive real numbers $\underline{\varphi}$ and $\bar{\varphi}$ such that the persistent excitation condition

$$\underline{\varphi}I \le \sum_{i=t-N+1}^{t} \varphi_t \varphi_t^T \le \bar{\varphi}I \tag{42}$$

are satisfied for all $t \ge N$.

Then there are positive real numbers $\underline{p}$ and $\bar{p}$ such that P_t is bounded by

$$\underline{p}I \le P_t \le \bar{p}I. \tag{43}$$

Based on the three lemmas, the main convergence analysis is given. Since the identification algorithm consists of two stages, the convergence analysis is divided into two parts.

Theorem 1. When the algorithm (15)-(17) is used to system (11), if **Assumptions 1-5** and conditions (40), (41) and (42) are satisfied, then $\tilde{\theta}_t$ is exponentially bounded in mean square and bounded with probability one.

Proof: From (14), (15) and (28), it follows that

$$\tilde{\theta}_{t+1} = \tilde{\theta}_t + K_t e_t - H_{t+1}. \tag{44}$$

Define the following Lyapunov function (Kalman et al., 1960):

$$L_{t+1} = \tilde{\theta}_{t+1}^T P_{t+1}^{-1} \tilde{\theta}_{t+1}. \tag{45}$$

From (16) and (17), it can be obtained

$$P_{t+1}^{-1} = (P_t + Q^2)^{-1} + R^{-2}\varphi_t \varphi_t^T \tag{46}$$

and

$$K_t = -R^{-2} P_{t+1} \varphi_t. \tag{47}$$

By using (15) and (47), (45) becomes

$$L_{t+1}=\left(\tilde{\theta}_t-R^{-2}P_{t+1}\varphi_t e_t-H_{t+1}\right)^T P_{t+1}^{-1}\left(\tilde{\theta}_t-R^{-2}P_{t+1}\varphi_t e_t-H_{t+1}\right)$$
$$=\left(\tilde{\theta}_t-H_{t+1}\right)^T P_{t+1}^{-1}\left(\tilde{\theta}_t-H_{t+1}\right)-2R^{-2}\left(\tilde{\theta}_t-H_{t+1}\right)^T\varphi_t e_t+R^{-4}\varphi_t^T P_{t+1}\varphi_t e_t^2. \quad (48)$$

Substituting (46) into (48), it follows that

$$L_{t+1}=\left(\tilde{\theta}_t-H_{t+1}\right)^T\left[\left(P_t+Q^2\right)^{-1}+R^{-2}\varphi_t\varphi_t^T\right]\left(\tilde{\theta}_t-H_{t+1}\right)$$
$$-2R^{-2}\left(\tilde{\theta}_t-H_{t+1}\right)^T\varphi_t e_t+R^{-4}\varphi_t^T P_{t+1}\varphi_t e_t^2$$
$$=\left(\tilde{\theta}_t-H_{t+1}\right)^T\left(P_t+Q^2\right)^{-1}\left(\tilde{\theta}_t-H_{t+1}\right)+R^{-2}\tilde{\theta}_t^T\varphi_t\varphi_t^T\tilde{\theta} \quad (49)$$
$$-2R^{-2}\tilde{\theta}_t^T\varphi_t\varphi_t^T H_{t+1}+R^{-2}H_{t+1}^T\varphi_t\varphi_t^T H_{t+1}$$
$$-2R^{-2}\tilde{\theta}_t^T\varphi_t e_t+2R^{-2}H_{t+1}^T\varphi_t e_t+R^{-4}\varphi_t^T P_{t+1}\varphi_t e_t^2.$$

From (11) and (18),

$$e_t=\hat{\theta}_t^T\varphi+f_{O,1}\left(y_t\right)-\theta_t^T\varphi-f_{O,1}\left(y_t\right)=\tilde{\theta}_t^T\varphi. \quad (50)$$

Therefore, (49) becomes

$$L_{t+1}=\left(\tilde{\theta}_t-H_{t+1}\right)^T\left(P_t+Q^2\right)^{-1}\left(\tilde{\theta}_t-H_{t+1}\right)+R^{-2}e_t^2$$
$$-2R^{-2}e_t\varphi_t^T H_{t+1}+R^{-2}H_{t+1}^T\varphi_t\varphi_t^T H_{t+1}$$
$$-2R^{-2}e_t^2+2R^{-2}H_{t+1}^T\varphi_t e_t+R^{-4}\varphi_t^T P_{t+1}\varphi_t e_t^2 \quad (51)$$
$$=\tilde{\theta}_t^T\left(P_t+Q^2\right)^{-1}\tilde{\theta}_t-2\tilde{\theta}_t^T\left(P_t+Q^2\right)^{-1}H_{t+1}+H_{t+1}^T P_{t+1}^{-1}H_{t+1}$$
$$+R^{-2}\left(R^{-2}\varphi_t^T P_{t+1}\varphi_t-1\right)e_t^2.$$

According to **Lemma 1**, $E\left(H_{t+1}^T\middle|\tilde{\theta}_t\right)=\left(0,\cdots,0\right)$. In addition, since neither P_t^{-1} nor $\tilde{\theta}_t$ depend on H_{t+1}. Therefore, $E[\tilde{\theta}_t^T(P_t+Q^2)^{-1}H_{t+1}\mid\tilde{\theta}_t]=0$. And then, given that

$$E(L_{t+1}\mid\tilde{\theta}_t)-L_t$$
$$=\tilde{\theta}_t^T\left(P_t+Q^2\right)^{-1}\tilde{\theta}_t-\tilde{\theta}_t^T P_t^{-1}\tilde{\theta}_t-2E\left[\tilde{\theta}_t^T(P_t+Q^2)^{-1}H_{t+1}\mid\tilde{\theta}_t\right]+E(H_{t+1}^T P_{t+1}^{-1}H_{t+1}\mid\tilde{\theta}_t)$$
$$+R^{-2}E\left[\left(R^{-2}\varphi_t^T P_{t+1}\varphi_t-1\right)e_t^2\mid\tilde{\theta}_t\right]. \quad (52)$$
$$=\tilde{\theta}_t^T\left[\left(P_t+Q^2\right)^{-1}-P_t^{-1}\right]\tilde{\theta}_t+E(H_{t+1}^T P_{t+1}^{-1}H_{t+1}\mid\tilde{\theta}_t)+R^{-2}E\left[\left(R^{-2}\varphi_t^T P_{t+1}\varphi_t-1\right)e_t^2\mid\tilde{\theta}_t\right].$$

Considering that P_t and Q^2 are symmetric positive, from (46), it is obviously that

$$R^{-2}\varphi_t^T P_{t+1}\varphi_t=R^{-2}\varphi_t^T\left[(P_t+Q^2)^{-1}+R^{-2}\varphi_t\varphi_t^T\right]^{-1}\varphi_t<1. \quad (53)$$

Therefore,

$$R^{-2}E\left[\left(R^{-2}\varphi_t^T P_{t+1}\varphi_t-1\right)e_t^2\mid\tilde{\theta}_t\right]\le 0. \quad (54)$$

Applying **Lemma 1** and **Lemma 3**, it is obtained

$$E(H_{t+1}^T P_{t+1}^{-1} H_{t+1} \mid \tilde{\theta}_t) \le \bar{H}^2 \underline{p}^{-1}. \tag{55}$$

According to (41) and (43), we have

$$\left(\bar{p}+\underline{q}\right)P_t \le \bar{p}\left(P_t + Q^2\right). \tag{56}$$

Therefore,

$$\left(P_t + Q^2\right)^{-1} \le \frac{\bar{p}}{\bar{p}+\underline{q}} P_t^{-1}. \tag{57}$$

Considering (54), (55) and (57), (52) yields

$$E(L_{t+1} \mid \tilde{\theta}_t) - L_t \le \tilde{\theta}_t^T \left(\frac{\bar{p}}{\bar{p}+\underline{q}} P_t^{-1} - P_t^{-1} \right) \tilde{\theta}_t + \bar{H}^2 \underline{p}^{-1} = -\frac{\underline{q}}{\bar{p}+\underline{q}} L_t + \bar{H}^2 \underline{p}^{-1}. \tag{58}$$

As $0 < \underline{q}\left(\bar{p}+\underline{q}\right)^{-1} < 1$, let $\underline{q}\left(\bar{p}+\underline{q}\right)^{-1}$ be α and $\bar{H}^2 \underline{p}^{-1}$ be β. From **Lemma 2**, the conclusion of the theorem is obtained. □

Based on the conclusion of **Theorem 1**, the second stage of the convergence analysis is given.

Theorem 2. Consider the solutions of $\hat{A}_t$, $\hat{B}_t$, $\hat{C}_t$ and $\hat{D}_t$ from AVE method or SVD method, if $\tilde{\theta}_t$ is bounded and $\hat{c}_{i,t} \ne 0$ ($i = 2, \cdots, p$), then $\hat{A}_t$, $\hat{B}_t$, $\hat{C}_t$ and $\hat{D}_t$ are bounded.

Proof: Firstly, the convergence analysis for AVE method is given. If $\tilde{\theta}_t$ is bounded, $\hat{\theta}_t$ is bounded. Since $\hat{A}_t$, $\hat{C}_t$ and $\hat{D}_t$ are obtained from $\hat{\theta}_t$ directly, they are also bounded. $\hat{B}_t$ is obtained from (21), it follows that

$$\hat{b}_{i,t} = \frac{1}{p}\sum_{j=1}^{p} \frac{b_{i,t} c_{j,t}}{\hat{c}_{j,t}} = \frac{1}{p}\sum_{j=1}^{p} \frac{b_{i,t} c_{j,t} + b_{i,t} c_{j,t}}{c_{j,t} + \tilde{c}_{j,t}} \tag{59}$$

It can be seen that if $\hat{c}_{i,t} \ne 0$ as well as $b_{i,t} c_{j,t}$ and $\tilde{c}_j$ are bounded, $\hat{B}_t$ is bounded either.

Secondly, the convergence analysis for SVD method is studied. If $\tilde{\theta}_t$ is bounded, $\hat{\theta}_{ab,t}$ and $\hat{\theta}_{bc,t}$ are bounded. According to the calculation method of singular value and singular column vector, if a matrix is bounded, all its singular values and singular column vectors are bounded. Therefore, $\hat{A}_t$, $\hat{B}_t$, $\hat{C}_t$ and $\hat{D}_t$ are also bounded. The conclusion of **Theorem 2** is obtained. □

Remark 5. For the AVE method, the additional condition $\hat{c}_{i,t} \ne 0$ is required to guarantee the boundedness of $\hat{b}_{i,t}$. This assumption is reasonable because it is almost impossible for $\hat{c}_{i,t}$ to be exactly zero in the identification process. However, in order to eliminate this situation completely, $\hat{c}_{i,t}$ can be set as a small number if it equals to zero.

Remark 6. From the conclusions of **Theorem 1** and **Theorem 2**, it is clear that the algorithm can achieve global convergence. Compared with the local convergence algorithm, such as the algorithm proposed in Yu et al. (2020), the global convergence algorithm can make the parameter estimate converge to the true value even when the parameter estimation error is large. This property is very important, especially for the identification of systems without prior knowledge. We will further show this in the simulation experiments.

In addition, the variation of the parameters is modelled as a stochastic process shown in **Assumption 5** in the convergence analysis of the algorithm. The influence of noise on parameter estimation is also not considered. However, the algorithm can still be applied to track the H-W systems with other kinds of time-varying parameters under noisy environments. This property will also be shown through simulation examples.

5 Simulation examples

Four simulation examples will be shown to verify the validity of the proposed algorithm in this section.

Example 1. Consider the following time-varying H-W system. The input nonlinear block is represented by a polynomial function:

$$v_t = F_I(u_t) = u_t + 0.4u_t^2 - 0.7u_t^3,$$

the linear block is represented by

$$w_t = 0.6w_{t-1} - 0.4w_{t-2} + 0.7v_{t-1} - 0.36v_{t-2},$$

and the inverse of the output block is also represented by a one-to-one polynomial function:

$$w_t = F_O^{-1}(y_t) = y_t - 0.4y_t^2 + 0.6y_t^3.$$

All of the time-varying terms $\delta_{i,t}$ obey the statistical laws $\delta_{i,t} \sim N\left(0,\quad 0.002\right)$.

P_1 is initialized with $10^8 I$. Q^2 is defined as the estimate of $E\left[H_t H_t^T\right]$. However, unbiased estimation of this matrix is not necessary from **Theorem 1**. According to (58), we have

$$E(L_{t+1} \mid \tilde{\theta}_t) \le \frac{\bar{p}}{\bar{p}+\underline{q}} L_t + \bar{H}^2 \underline{p}^{-1}. \tag{60}$$

It can be seen that appropriately increasing Q^2 may reduce the parameter estimation error. Therefore, Q^2 is set to $10I$. $\hat{\theta}_1$ is initialized with $0.01I$. The white Gaussian sequence with zero mean and variance 1 is taken as the system input. In the second stage of the algorithm, both of the AVE method and SVD method are used. The simulation goes through 5000 samples. The convergence results of the parameter estimation errors are shown in Fig. 2.

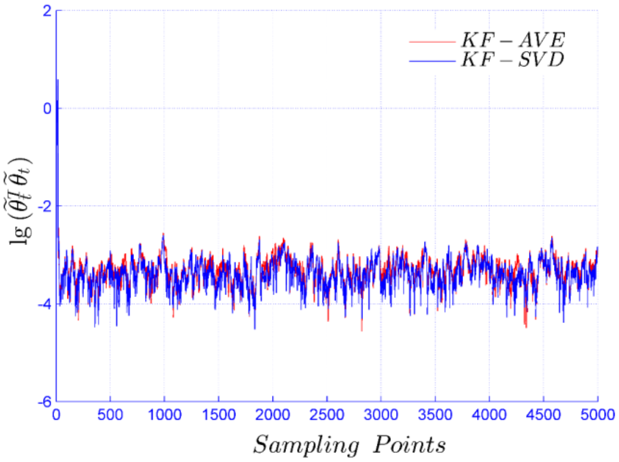


Fig. 2 Convergence results of parameter estimation error

In Fig. 3 and Fig. 4, the estimation results of each parameter from the two parameter separation methods are shown for contrasting.

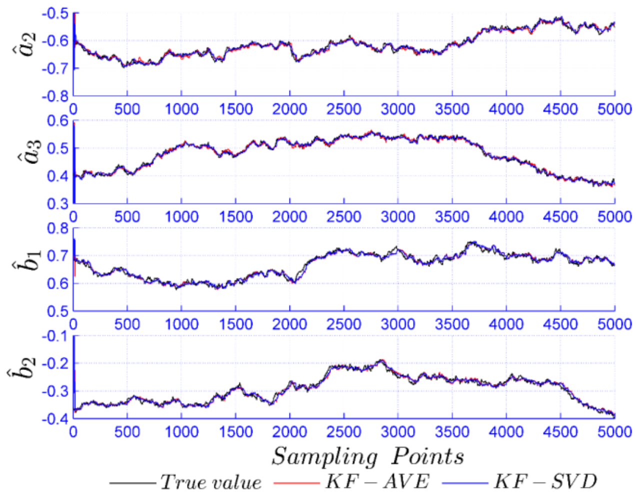


Fig. 3 Tracking results of parameter a_i, b_i

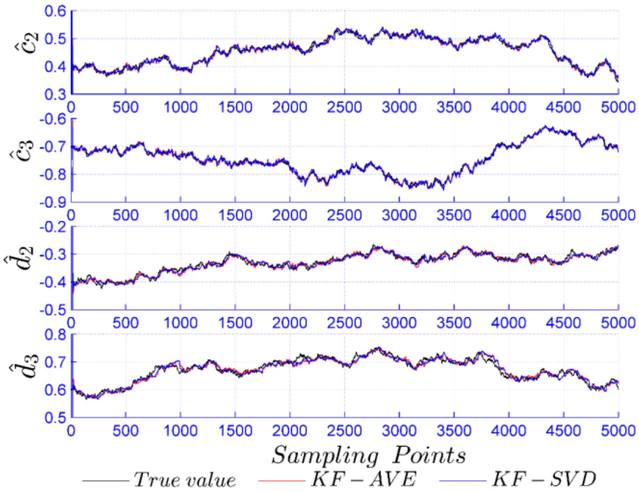


Fig. 4 Tracking results of parameter c_i, d_i

From the identification results, it can be seen that the parameter estimation errors of the two algorithms converge to a small range rapidly. All the parameter estimates can track the true values. Compared with AVE method, SVD method has higher tracking accuracy. But the difference between the two results is not obvious.

Example 2. In this simulation example, the two-stage algorithm proposed in the paper and a modified EKF algorithm proposed in Yu et al. (2020) are used to carry out a comparative experiment. The simulation example contains three experiments. The identification object in the three experiments is the same as **Example 1**. In order to test the convergence of different algorithms, different initial conditions of $\hat{\theta}_1$ are used. In the first experiment, $\hat{\theta}_1 = 0.01 \times (1, \ldots, 1)^T$, which is the same as **Example 1**. In the second experiment, $\hat{\theta}_1 = 10 \times (1, \ldots, 1)^T$. In the third experiment, $\hat{\theta}_1 = 1000 \times (1, \ldots, 1)^T$. Same process is used for the two algorithms in each experiments in order to have a fair comparison. Convergence curves of parameter estimates are shown in Fig.5, Fig.6 and Fig.7.

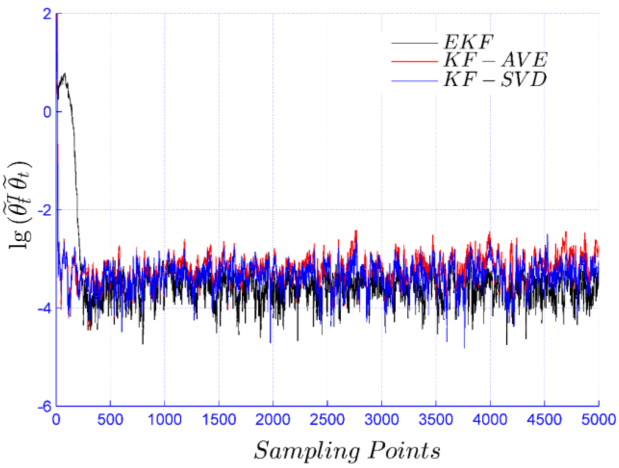


Fig. 5 Convergence results of estimation error with initial condition $\hat{\theta}_1 = 0.01 \times (1, \ldots, 1)^T$

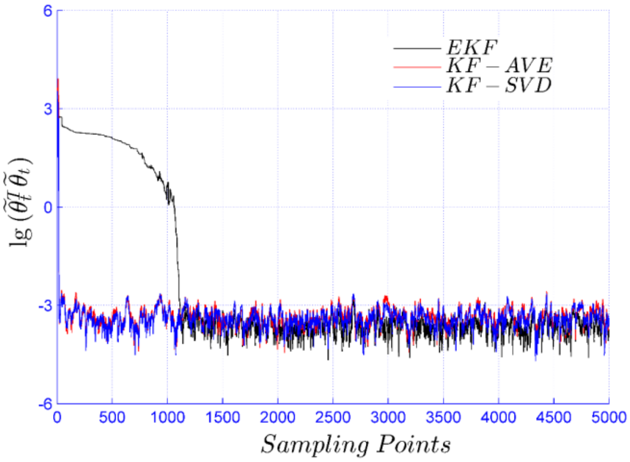


Fig. 6 Convergence results of estimation error with initial condition $\hat{\theta}_1 = 10 \times (1, \ldots, 1)^T$

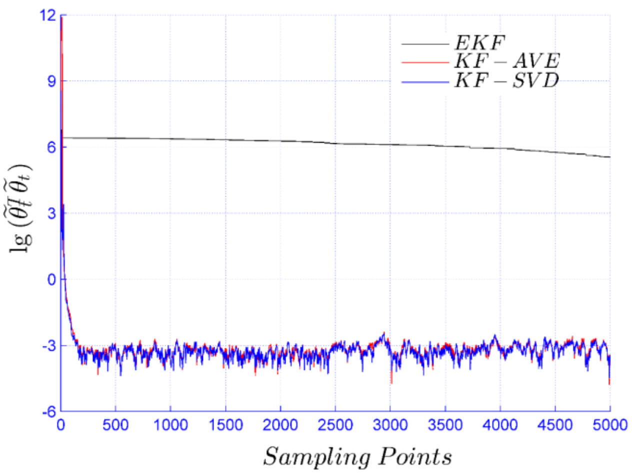


Fig. 7 Convergence results of estimation error with initial condition $\hat{\theta}_1 = 1000 \times (1, \ldots, 1)^T$

It can be seen from Fig.5 that the convergence rate of EKF algorithm is slower than that of two-stage method. This phenomenon is more obvious in Fig. 6. It is because the first order Taylor approximation is no longer reasonable with the increase of parameter estimation error. At the same time, too large parameter estimation error will make the EKF algorithm unable to converge, such as the results shown in Fig. 7. However, the two-stage method can achieve fast convergence no matter how large the initial parameter estimation error is.

Example 3. In this simulation example, the initialization form of the model to be identified is also the same as that in **Example 1**. However, the variation properties of the parameters are different. For parameter $a_{i,t}$, $\delta_{a_i,t}$ are constants, setting $\delta_{a_1,t} = 0.00005$ and

$\delta_{a_2,t} = -0.00004$. It is obvious that the variation of $a_{i,t}$ is monotone linear. The change of parameters $b_{i,t}$ obeys the following rules:

$$b_{1,t} = 0.7 + 0.0003\cos(0.005t),$$

$$b_{2,t} = -0.36 + 0.0005\sin(0.003t).$$

The characteristic of the time-varying parameters C_t is the same as that in **Example 1**, namely $\delta_{c_i,t} \sim N(0,\quad 0.002)$. The parameter $d_{i,t}$ have a jump at the 1000th sampling point. The amplitudes of the jump for $d_{2,t}$ and $d_{3,t}$ are 0.1 and -0.1, respectively.

In addition, measurement noise is added to the system output with 20dB signal-to-noise ratio. The simulation also goes through 5000 samples. The convergence results of the parameter estimation errors for the two parameter separation methods are shown in Fig.8.

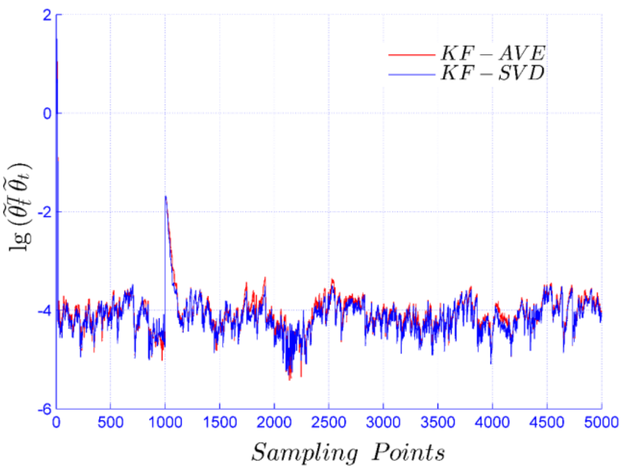


Fig. 8 Convergence results of estimation error

In Fig. 9 and Fig. 10, the estimation results of each parameter are shown for contrasting.

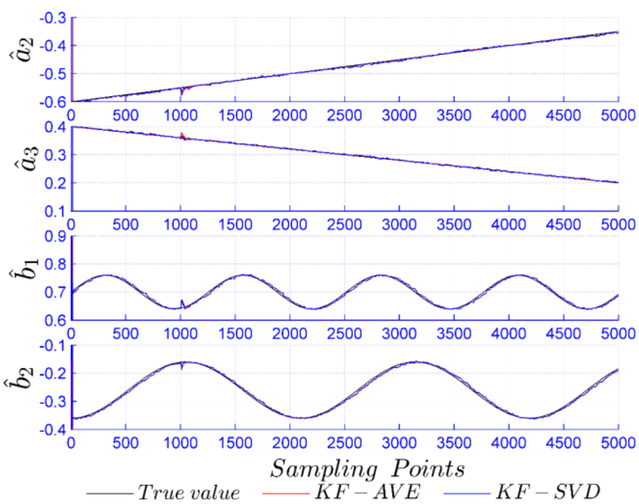


Fig. 9 Tracking results of parameter a_i and b_i

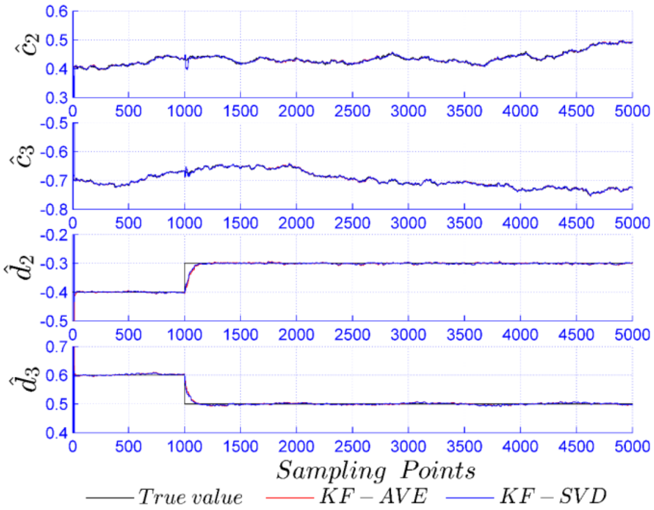


Fig. 10 Tracking results of parameter c_i and d_i

From the results, it can be seen that although the time-variation properties of parameters are different from **Assumption 5**, all the parameter estimates track the true values very well for both of the two separation methods. At the same time, it is also noted that the noise does not cause obvious interference to the identification accuracy. At the 1000th sampling point, the parameter estimation error has a great change. It is because $d_{2,t}$ and $d_{3,t}$ jumped at that time. Although there is a sudden change, the parameter estimation error converged to the normal level quickly.

Example 4. In this simulation, the identification of an electric arc furnace (EAF) system is given (Yu et al. 2014). EAF is a device for steel production by melting scrap. The model of arc furnace consists of proportional valve, hydraulic system and arc characteristics in series. The structure of EAF model is shown in Fig.11,

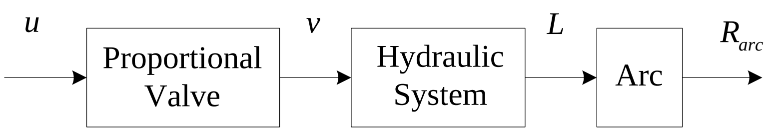


Fig. 11 Structure of EAF system

where u is input signal; v is voltage value, V; L is the length of arc, cm; R_{arc} is arc resistance, mΩ. The proportional valve converts input signal into voltage value and drives the hydraulic system to change the length of arc, so as to realize the control of arc resistance.

The proportional valve contains a dead-zone nonlinearity, which can be expressed as:

$$v = \begin{cases} u - p_1 & 10 \geq u > p_1 \\ 0 & p_2 \leq u \leq p_1 \\ u - p_2 & -10 \leq u < p_2 \end{cases} \tag{61}$$

where p_1 and p_2 are dead-zone parameters. The characteristic of hydraulic system can be expressed by a linear model. The discrete transfer function is

$$G(z^{-1}) = \frac{b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3}}{1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3}}. \tag{62}$$

The EAF power supply system can be modelled by the following equivalent R-L circuit:

$$U_p{}^2 = (R_d I_{arc} + U_{arc})^2 + (X_d I_{arc})^2 \tag{63}$$

where U_p is the transformer secondary voltage, V; R_d is the circuit resistance, mΩ; X_d is the circuit inductance, mΩ; I_{arc} is the arc current, A; U_{arc} is the arc voltage, V; and we have:

$$U_{arc} = a_{arc} + b_{arc,t} L = I_{arc} R_{arc} \tag{64}$$

where a_{are} is the voltage drop of arc anode and cathode, V; $b_{are,t}$ is the arc voltage drop gradient, V/cm. It is a time-varying parameter and decreases with the increase of furnace temperature. It can be seen that the relationship between R_{are} and L is one-to-one, time-varying and nonlinear. In the simulation, we set $p_1 = 1, p_2 = -1$ for proportional valve; $a_1 = -1.637$, $a_2 = 1.044$, $a_3 = -0.4066$, $b_1 = 0.0378$, $b_2 = 0.0333$, $b_3 = 0.0097$ for hydraulic system; $U_p = 320$, $R_d = 0.4$, $X_d = 0.7$, $a_{arc} = 30$, $b_{are,t}$ decreases from 6 to 3 for power supply system.

The characteristics of EAF can be expressed by an H-W model. Rewritten equation (61) with the form of equation (4):

$$\begin{aligned} v_t = & \left[h_1(u_t) - h_1(u_t) h_1(p_1 - u_t) + h_2(u_t) - h_2(u_t) h_2(p_2 - u_t) \right] u_t \\ & + \left[h_1(u_t) h_1(p_1 - u_t) - h_1(u_t) \right] p_1 + \left[h_2(u_t) h_2(p_2 - u_t) - h_2(u_t) \right] p_2 \end{aligned} \tag{65}$$

As the characteristic of arc is one-to-one, its inverse is approximated by a polynomial function

$$L_t = F_{O,t}^{-1}\left(R_{arc,t}\right) = \sum_{i=1}^{q} d_{i,t} R_{arc,t}^{i}, \tag{66}$$

The purpose of identification is to estimate the parameters p_1, p_2, a_1-a_3 and b_1-b_3 and track the change of arc characteristics by the measurable input and output data u and R_{arc} for obtaining accurate system output estimation. In the simulation, P_1 is initialized with $10^8 I$. Q^2 is set to I. $\hat{\theta}_1$ is initialized with 0.01I. The uniformly distributed sequences within [-3, 3] around the working point is taken as the system input. The simulation goes through 5000 samples. The parameter $b_{are,t}$ changes uniformly in each sampling point from 6 to 3. Three different polynomial functions with orders q equal to 4, 5 and 6 are used. The AVE method is adopted to obtain the original system parameters. The identification results are shown in Fig. 12 and Fig. 13. Fig. 12 shows the sum of squares of estimation errors of parameters p_i, a_i and b_i for the models with different orders, where

$$Er = \sum\left(\hat{p}_{i,t} - p_i\right)^2 + \sum\left(\hat{a}_{i,t} - a_i\right)^2 + \sum\left(\hat{b}_{i,t} - b_i\right)^2. \tag{67}$$

Fig.13 shows the estimation error e_t of system output.

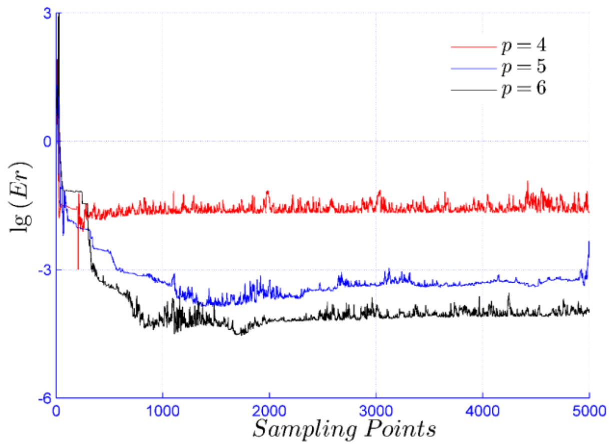


Fig. 12 Parameter estimation error

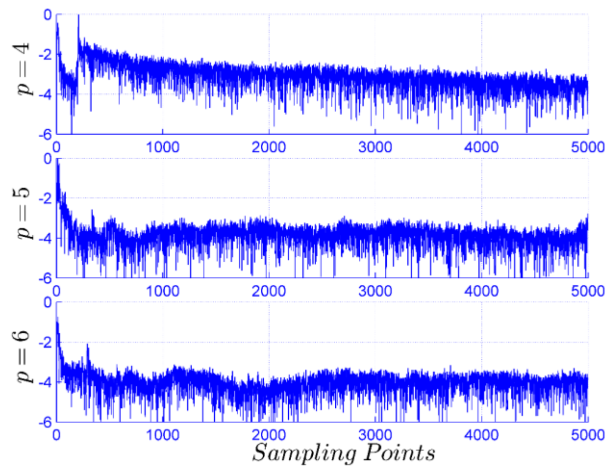


Fig. 13 System output estimation error

From Fig. 12 and Fig. 13, it can be seen that both of the parameter estimates and system output estimates converged for all the three models. However, with the increase of model order p, estimation errors of parameter and system output decrease gradually. It can be seen that in the application of the algorithm, appropriately increasing the number of basis function can improve the accuracy of identification.

6 Conclusions

This paper has introduced a new two-stage recursive algorithm aimed at tracking the time-varying H-W systems. In the first stage, KF algorithm is used for identifying the parameter vector θ_t. In the second stage, AVE method and SVD method are proposed to separate the parameter product terms in θ_t to obtain the system parameter estimates. The convergence analysis for the proposed method also includes two parts. In the first part, it is proved that KF algorithm can guarantee that $\tilde{\theta}_t$ is bounded under the condition of sufficient excitation. Based on the conclusion of the first part, it is proved in the second part that both parameter separation methods can ensure the boundness of parameter estimation errors. The performance of the proposed algorithms has been demonstrated via four simulated examples with experimental design of a variety of system varying characteristics typical in complex systems, including a practical EAF system case. It is shown that the proposed algorithm is robust to measurement noise and can be used to tracking the H-W systems with deterministic trend change, jump change.

For the identification of time-varying H-W systems, there are still a lot of important issues to be investigated. The further research should focus on structure identification, e .g, on adaptively estimating of the order r and m, the numbers of basis function p and q and determining the type of the basis function for the nonlinear parts.

Acknowledgments

This work was supported by the National Natural Science Foundation of China Grant (No. 61703085, No. 61773105), the Fundamental Research Funds for the Central Universities (No. 170404021) and the China Scholarship Council (No. 201906085064)

References

Astrom, K. J., & Wittenmark, B. (2013). Adaptive control, Courier Dover Publications.

Bai, E. W. (1998). An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica*, 34(3), 333-338. https://doi.org/10.1016/S0005-1098(97)00198-2

Bai, E. W. (2002). A blind approach to the Hammerstein-Wiener model identification. *Automatica*, 38(6), 967-979. https://doi.org/10.1016/S0005-1098(01)00292-8

Bershad, N. J., Celka, P., & Vesin, J. (2000). Analysis of stochastic gradient tracking of time-varying polynomial Wiener systems. *IEEE Transactions on Signal Processing*, 48(6), 1676-1686. https://doi.org/10.1109/78.845925

Chen, J., Huang, B., Ding, F., & Gu, Y. (2018). Variational Bayesian approach for ARX systems with missing observations and varying time-delays. *Automatica*, 94, 194-204. https://doi.org/10.1016/j.automatica.2018.04.003

Chen, J., Liu, Y. J., Ding, F., Zhu, Q. M. (2020a) Gradient-based particle filter algorithm for an ARX model with nonlinear communication output. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, 50(6), 2198-2207. https://doi.org/10.1109/TSMC.2018.2810277

Chen, J., Zhu, Q. M., & Liu, Y. J. (2020b). Modified Kalman filtering based multi-step-length gradient iterative algorithm for ARX models with random missing outputs. *Automatica*, 118, 109034. https://doi.org/10.1016/j.automatica.2020.109034

Chen, M., Wang, H. Q. & Liu, X. P. (2021). Adaptive Fuzzy Practical Fixed-Time Tracking Control of Nonlinear Systems. *IEEE Transactions on Fuzzy Systems*, 29(3), 664-673. https://doi.org/10.1109/TFUZZ.2019.2959972.

Deyst, J., & Price, C. (1968). Conditions for asymptotic stability of the discrete minimum-variance linear estimator. *IEEE Transactions on Automatic Control*, 13(6), 702-705. https://doi.org/10.1109/TAC.1968.1099024

Ding, F., Liu, X. M., & Liu, M. M. (2016). The recursive least squares identification algorithm for a class of Wiener nonlinear systems. *Journal of the Franklin Institute*, 353(7), 1518-1526. https://doi.org/10.1016/j.jfranklin.2016.02.013

Giri, F., Rochdi, Y., Radouane, A., Brouri, A., & Chaoui, F. Z. (2013). Frequency identification of nonparametric Wiener systems containing backlash nonlinearities. *Automatica*, 49(1), 124-137. https://doi.org/10.1016/j.automatica.2012.08.043

Hong, X., & Mitchell, R. J. (2007). Hammerstein model identification algorithm using Bezier–Bernstein approximation. *IET Control Theory & Applications*, 1(4), 1149-1159. https://doi.org/10.1049/iet-cta:20060018

Hong, X., Mitchell R. J., & Chen S. (2012). Modelling and control of Hammerstein system using B-spline approximation and the inverse of De Boor algorithm. *International journal of systems sciences*, 43(10), 1976-1984. https://doi.org/10.1080/00207721.2011.564320

Hong, X., Mitchell, R. J., & Chen, S. (2013). System identification of Wiener systems with B-spline functions using De Boor recursion. *International Journal of Systems Science*, 44(9), 1666-1674. https://doi.org/10.1080/00207721.2012.669863

Kalman, R. & Bertram, J. (1960). Control System Analysis and Design Via the “Second Method” of Lyapunov: II—Discrete-Time Systems, ASME. Journal of Fluids Engineering, 82(2), 394–400. https://doi.org/10.1115/1.3662605

Kobayashi, Y., Shiotani, Y., Hikita, S., & Nomura, K. (2010). Identification of time-varying Wiener systems with unknown parameters. *Electronics & Communications in Japan*, 93(4), 1-9. https://doi.org/10.1002/ecj.10247

Li, L. W., & Ren, X. M. (2017). Decomposition-based recursive least-squares parameter estimation algorithm for Wiener-Hammerstein systems with dead-zone nonlinearity. *International Journal of Systems Science*, 48(11), 2405-2414. https://doi.org/10.1080/00207721.2017.1320455

Liu, Q.Y., Ding, F., Zhu, Q.M., & Hayat, T. (2019). Two-stage multi-innovation stochastic gradient algorithm for multivariate output-error ARMA systems based on the auxiliary model. *International journal of systems sciences*, 50(15), 2870-2884. https://doi.org/10.1080/00207721.2019.1690720

Ljung, L & Gunnarsson, S. (1990). Adaptation and Tracking in System Identification-A Survey. *Automatica*, 26(1), 7-21. https://doi.org/10.1016/0005-1098(90)90154-A

Mu, B. Q., & Chen, H. F. (2013). Recursive Identification of MIMO Wiener Systems. *IEEE Transactions on Automatic Control*, 58(3), 802-808. https://doi.org/10.1109/TAC.2012.2215539

Nordsjo, A. E., & Zetterberg, L. H. (2001). Identification of certain time-varying nonlinear Wiener and Hammerstein systems. *IEEE Transactions on Signal Processing*, 49(3), 577-592. https://doi.org/10.1109/78.905884

Reif, K., Gunther, S., Yaz, E., & Unbehauen, R. (1999). Stochastic stability of the discrete-time extended Kalman filter. *IEEE Transactions on Automatic Control*, 44(4),714-728. https://doi.org/10.1109/9.754809

Voros, J. (2005). Identification of Hammerstein systems with time-varying piecewise-linear characteristics. *IEEE Transactions on Circuits & Systems II Express Briefs*, 52(12), 865-869. https://doi.org/10.1109/TCSII.2005.853339

Voros, J. (2011). Identification of Nonlinear Cascade Systems with Time-Varying Backlash. *Journal of Electrical Engineering*, 62(2), 87-92. https://doi.org/10.2478/v10187-011-0014-2

Voros, J. (2013). Recursive Identification of Nonlinear Cascade Systems With Time-Varying General Input Backlash. *Journal of Dynamic Systems, Measurement and Control*, 135(1), 014504. https://doi.org/10.1115/1.4006630

Voros, J. (2014). Identification of nonlinear dynamic systems with input saturation and output backlash using three-block cascade models. *Journal of the Franklin Institute*, 351(12), 5455-5466. https://doi.org/10.1016/j.jfranklin.2014.09.025

Voros, J. (2017). Recursive identification of discrete-time nonlinear cascade systems with time-varying output hysteresis. *Nonlinear Dynamics*, 87, 1427-1434. https://doi.org/10.1007/s11071-016-3124-3

Voros, J. (2018). Recursive identification of time-varying non-linear cascade systems with static input and dynamic output non-linearities. *Transactions of the Institute of Measurement & Control*, 40(3), 896-902. https://doi.org/10.1177/0142331216670485

Wang, D. Q. & Ding, F. (2012). Hierarchical Least Squares Estimation Algorithm for Hammerstein-Wiener Systems. *IEEE Signal Processing Letters*, 19(12), 825-828. https://doi.org/10.1109/LSP.2012.2221704

Wang, H. Q., Bai, W., Zhao, X. D. & Liu, P. X. (2021). Finite-Time-Prescribed Performance-Based Adaptive Fuzzy Control for Strict-Feedback Nonlinear Systems with Dynamic Uncertainty and

Actuator Faults. *IEEE Transactions on Cybernetics*, 1-13. https://doi.org/10.1109/TCYB.2020.3046316

Wang, H. Q., Bai, W., & Liu, P. X. (2019). Finite-time adaptive fault-tolerant control for nonlinear systems with multiple faults. *IEEE/CAA Journal of Automatica Sinica*, 6(6), 1417-1427. https://doi.org/10.1109/JAS.2019.1911765.

Wang, J., & B. C., Ding. (2018). Two-step output feedback predictive control forHammerstein systems with networked-induced time delays. *International journal of systems sciences*, 49(13), 2753-2762. https://doi.org/10.1080/00207721.2018.1511872

Wills, A., Schon, T. B., Ljung, L., & Ninness, B. (2013). Identification of Hammerstein-Wiener models. *Automatica*, 49(1), 70-81. https://doi.org/10.1016/j.automatica.2012.09.018

Yu, F., Mao, Z. Z., & Jia, M. X. (2013). Recursive identification for Hammerstein–Wiener systems with dead-zone input nonlinearity. *Journal of Process Control*, 23(8), 1108-1115. https://doi.org/10.1016/j.jprocont.2013.06.014

Yu, F., Mao, Z. Z., Jia, M. X., & Yuan, P. (2014). Recursive parameter identification of Hammerstein–Wiener systems with measurement noise. *Signal Processing*, 105, 137-147. https://doi.org/10.1016/j.sigpro.2014.05.030

Yu, F., Mao, Z. Z., Yuan, P., He, D. K., & Jia, M. X. (2017). Recursive parameter estimation for Hammerstein-Wiener systems using modified EKF algorithm. *ISA transactions*, 70, 104-115. https://doi.org/10.1016/j.isatra.2017.05.012

Yu, F., Mao, Z. Z., & He, D. K. (2020). Identification of Time-Varying Hammerstein-Wiener Systems. *IEEE Access*, https://doi.org/doi: 10.1109/ACCESS.2020.3011608.

Zhang, J., Yu, F., & Chin, K. S. (2017). Recursive identification of Hammerstein systems with dead-zone nonlinearity in the presence of bounded noise. *International journal of systems sciences*, 48(11), 2394-2404. https://doi.org/10.1080/00207721.2017.1316427