Available online at www.sciencedirect.com

ScienceDirect

Journal of the Franklin Institute 357 (2020) 1280–1298

www.elsevier.com/locate/jfranklin

Adaptive filtering-based recursive identification for time-varying Wiener output-error systems with unknown noise statistics

Zhu Wang*, Haoran An, Xionglin Luo

Department of Automation, China University of Petroleum, Beijing 102249, PR China

Received 7 November 2018; received in revised form 21 June 2019; accepted 4 November 2019
Available online 9 November 2019

Abstract

In area of control, model-based robust identification is rare, and studies in presence of unknown noise statistics are especially seldom. The robust estimation problem for time-varying Wiener output-error systems is considered in this paper. An adaptive filtering-based recursive identification scheme is proposed to distinguish nonlinear time-varying characteristics in complex noise environments. Firstly, a virtual equivalent state space model is constructed to achieve adaptive Kalman filtering. In filter design, a weighted noise estimator based on Sage-Husa principle is introduced, and is sensitive to noise changes. Secondly, the state estimates obtained by filters are used to form the unknown intermediate variables in information vectors. Then, a recursive estimation method based on multiple iterations is developed, and the convergence of identification is confirmed by martingale hyperconvergence theorem. Finally, the numerical simulation results verify the theoretical findings.

1. Introduction

Nonlinear block-oriented models can be used to describe many nonlinear dynamic behaviors [1]. Three typical nonlinear block-oriented models are Wiener models, Hammerstein models [2], and their combinations [3,4]. These types of models contain both static nonlin-

* Corresponding author.
E-mail addresses: jz21561@163.com (Z. Wang), anhaoran_9911@163.com (H. An), luoxl@cup.edu.cn (X. Luo).

https://doi.org/10.1016/j.jfranklin.2019.11.005

ear blocks and linear dynamic blocks. Among them, a Wiener model has a linear dynamic part followed by a nonlinear part [5], and it is widely encountered in engineering practices. Some biological processes [6] and chemical processes [7] are usually modeled by Wiener systems. Meanwhile, these processes are considered time-varying [8] because of environmental or human influence. Besides, if system physical laws are accurate sufficiently [9], the absence of process noise is reasonable, and such systems are known as output-error systems on system identification. Hence, this paper considers the identification of time-varying Wiener output-error systems, and aims to grasp nonlinear and time-varying characteristics of systems.

The identification issues for Wiener systems have attracted great attention. Jvoros [10] studied the identification of Wiener models that have piecewise-linear or inverse functions of nonlinear parts. The identification of non-invertibility nonlinear parts has been also studied by Hu and Chen [11]. For FIR Wiener systems, Lacy and Berntein [12] exploited simultaneous direct estimation of non-invertible, polynomial non-linearities. From another perspective, existing identification methods for Wiener systems are roughly divided to several categories, e.g., blind approaches [13,14], maximum likelihood methods [15], iterative methods [16], and recursive methods [17]. Among these methods, recursive methods are popular because they are suitable for online identification and can be used for time-varying systems [18]. For Wiener systems whose output nonlinear function is continuous and invertible, Ding et al. proposed the auxiliary-model based recursive least squares algorithm [19]. Further, multiple iterations [20–22] in recursive identification have been introduced to improve robustness of parameter estimation in noisy environments.

Kalman filter is a widely used model-based state estimator [23–25], and is applied to linear systems with Gaussian noise. For time-varying output-error system identification, the stability of Kalman filter [26] has been ensured under uniform complete observability. For Hammerstein systems [27,28], the modified Kalman smoother was derived to estimate the unknown intermediate variables in systems. For the real-time estimation of 4WD vehicle states, the extended Kalman filter [29] was combined with minimum error criterion. For actual fault diagnosis of time varying systems, the adaptive Kalman filter [30] was proposed through joint state-parameter estimation. For soft sensor maintenance, data fusion technology was introduced based on Kalman filter [31]. However, under unknown noise statistics, estimators should be redesigned to achieve adaptive filtering. In this respect, Sage-Husa maximum a posteriori estimation principle [32] has been applied into Kalman filter, in order to give statistical noise properties. Unlike Nussbaum designs [33,34] to handle non-zero-mean nonlinearities, Sage-Husa Kalman filter can recursively estimates both mean values and variances of noise.

In this paper, a linear regression form of Wiener nonlinear systems is adopted for identification. The proposed recursive estimation is used to achieve identification of time-varying parameters, and the idea of multiple iterations is introduced to improve robustness. Meanwhile, the bounded convergence of time-varying systems is harvested by using martingale theorems [35,36]. Further, to estimate the unknown intermediate variables in information vectors, a virtual equivalent state space model is proposed, and an extended Kalman filter is implemented. This adaptive Kalman filter is designed for output-error time-varying systems, and is based on the results of [26]. Besides, Sage-Husa noise estimator is also introduced into adaptive filtering. Thus, the recursive identification based on adaptive filtering is able to integrate state estimation with system identification, and can provide an effective identification method for time-varying Wiener output-error systems under complex noise environments [37].

The rest of paper is organized as follows. Section 2 gives the system description. The adaptive filtering-based recursive estimation is exploited in Section 3. Section 4 shows main

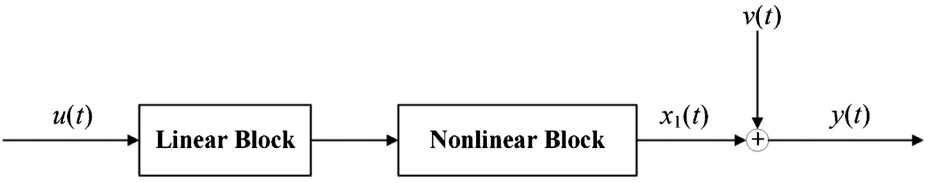

Fig. 1. Wiener output-error system structure.

performance analyses of identification. To verify effectiveness of the proposed method, numerical simulation is conducted in Section 5. Finally, Section 6 comes to a conclusion.

2. System description

Refer to the Wiener models in [16], and consider a discrete time-varying Wiener output-error system as following

$$x_1(t) = \sum_{i=1}^{p} a_i(t-1) \sum_{l=1}^{q} d_l(t-1) g_l(x_1(t-i)) + \sum_{j=1}^{p} b_j(t-1) u(t-j), \tag{1}$$

$$y(t) = x_1(t) + v(t), \tag{2}$$

where the measurement output $y(t)$ is composed of the true output $x_1(t)$ and the ambient noise $v(t)$, and $g_l(\cdot)$, $l = 1, \ldots, q$, are known nonlinear base functions, parameters $\{a_i(t-1)\}_{i=1}^{p}$, $\{d_l(t-1)\}_{l=1}^{q}$ and $\{b_j(t-1)\}_{j=1}^{p}$ are time-varying parameters that need to be identified. Note that both the order p for linear part and the number q of nonlinear base functions are known. Thus, identification should be constructed to estimate the system parameters directly. The Wiener output-error system in (1)–(2) can be depicted in Fig. 1.

Define the parameter and information vectors as

$$\begin{aligned}
&\boldsymbol{\varphi}_k{}^T(t) = \left[g_k(x_1(t-1)),\ g_k(x_1(t-2)),\ \ldots,\ g_k(x_1(t-p))\right],\quad k = 1, \ldots, q,\\
&\boldsymbol{\varphi}_u{}^T(t) = \left[u(t-1),\ \ldots,\ u(t-p)\right],\\
&\boldsymbol{\theta}_u{}^T(t-1) = \left[b_1(t-1),\ \ldots,\ b_p(t-1)\right],\\
&\boldsymbol{\theta}_k{}^T(t-1) = \left[a_1(t-1)d_k(t-1),\ \ldots,\ a_p(t-1)d_k(t-1)\right],\quad k = 1, \ldots, q,\\
&\boldsymbol{\varphi}^T(t) = \left[\boldsymbol{\varphi}_1{}^T(t),\ \ldots,\ \boldsymbol{\varphi}_q{}^T(t),\ \boldsymbol{\varphi}_u{}^T(t)\right],\\
&\boldsymbol{\theta}^T(t-1) = \left[\boldsymbol{\theta}_1{}^T(t-1),\ \ldots,\ \boldsymbol{\theta}_q{}^T(t-1),\ \boldsymbol{\theta}_u{}^T(t-1)\right],
\end{aligned} \tag{3}$$

where $\boldsymbol{\varphi}(t) \in R^{p(q+1)}$ and $\boldsymbol{\theta}(t-1) \in R^{p(q+1)}$. Thus, the regression form of (1)–(2) can be written as

$$y(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}(t-1) + v(t). \tag{4}$$

Assumptions 1 and 2. The noise $v(t)$ satisfies the following assumptions

$$\mathrm{E}[v(t)] = r(t) = 0; \tag{A1}$$

$$\mathrm{E}\left[v^2(t)\right] = R(t) \le \sigma_v{}^2 < \infty. \tag{A2}$$

The assumptions (A1) and (A2) shows that stochastic noise has properties of both zero mean and bounded variances. However, the noise can be either white or colored, and its statistical property can be non-Gaussian and time-varying. Specifically, the noise variances $R(t)$ can change with time. These assumptions reflect unknown noise statistics.

Assumptions 3 and 4. Define the parameter variation $\boldsymbol{\omega}(t) = \boldsymbol{\theta}(t) - \boldsymbol{\theta}(t-1)$. Assume that $\boldsymbol{\omega}(t)$ satisfies the following assumptions

$$E\left[\boldsymbol{\omega}(t)\boldsymbol{\omega}^T(s)\right] = 0, s \neq t; \mathrm{E}[v(t)\boldsymbol{\omega}(s)] = 0, \tag{A3}$$

$$\mathrm{E}\left[\|\boldsymbol{\omega}(t)\|^2\right] \leq \varepsilon\|\boldsymbol{\theta}(t-1)\|^2 \leq {\sigma_w}^2 < \infty, \tag{A4}$$

where $\varepsilon > 0$ is a very small number. That is, the parameter variation $\boldsymbol{\omega}(t)$ remains tiny in $\boldsymbol{\theta}(t-1)$. The assumptions illustrate that the variation is uncorrelated with ambient noise.

3. Adaptive filtering-based recursive identification

This section aims to employ a weighted gradient identification algorithm to estimate parameters. Adaptive filtering is introduced to harvest the optimal estimates of true $x_1(t)$.

Referring to the works in [20], multiple iterations in recursive identification can improve robustness of parameter estimation in noisy environments, and thus improve model accuracy. Because of time-varying property of parameters, damping coefficients are also introduced to cost functions. That is, different coefficients enable fast forgetting of past information, and thus enable fast tracking of current parameter changes. Define the cost function as

$$J(\boldsymbol{\theta}) = \frac{1}{2}\sum_{j=t-m+1}^{t}\left[\bar{b}^{t-j}\bar{d}_m \cdot \left(y(j) - \boldsymbol{\varphi}^T(j)\boldsymbol{\theta}\right)^2\right], \tag{5}$$

where m denotes the number of samples from epoch $t-m+1$ to epoch t at each recursion. From (5), the damping coefficients fitted by negative exponential law are selected as $\{\bar{b}^{m-1}\bar{d}_m, \ldots, \bar{b}^1\bar{d}_m, \bar{d}_m\}$, where the scalar $\bar{b}$ represents attenuation speed, and $\bar{d}_m = (1-\bar{b})/(1-\bar{b}^m)$, $0 < \bar{b} < 1$. It should be noted that the values of $\bar{b}$ and m need to be adjusted, according to situations of parameter changes. The tradeoff between robust estimation and change track needs to be solved. Thus, the proposed cost function can not only guarantee robust estimation, but also keep track of parameter changes. The quadratic criterion in (5) can be minimized by adopting the well-known Gauss-Newton technique [20]. Define

$$\begin{aligned}
&\boldsymbol{\Phi}(t) = [\boldsymbol{\varphi}(t-m+1), \ \ldots, \ \boldsymbol{\varphi}(t)]^T, \\
&\boldsymbol{Y}(t) = \left[y(t-m+1), \ \ldots, \ y(t)\right]^T, \\
&\boldsymbol{W}(t) = diag(\bar{b}^{m-1}\bar{d}_m, \ \ldots, \ \bar{b}^1\bar{d}_m, \ \bar{d}_m),
\end{aligned} \tag{6}$$

where $\boldsymbol{\Phi}(t) \in R^{m \times p(q+1)}$ and $\boldsymbol{W}(t) \in R^{m\times m}$. Then, the following equation is yielded

$$\frac{d}{d\boldsymbol{\theta}}J(\boldsymbol{\theta}) = -\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)[\boldsymbol{Y}(t) - \boldsymbol{\Phi}(t)\boldsymbol{\theta}]. \tag{7}$$

The Hessian of $J(\boldsymbol{\theta})$ is given by

$$\frac{d^2}{d\boldsymbol{\theta}^2}J(\boldsymbol{\theta}) = \boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t). \tag{8}$$

The Gauss-Newton solution of minimizing (5) can be generated by

$$\begin{aligned}\hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) + o(t)\left[-\frac{d^2 J(\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}\right]^{-1} \cdot \left.\frac{dJ(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}(t-1)} \\ &= \hat{\boldsymbol{\theta}}(t-1) + o(t)\boldsymbol{H}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\left[\boldsymbol{Y}(t) - \boldsymbol{\Phi}(t)\hat{\boldsymbol{\theta}}(t-1)\right],\end{aligned} \quad (9)$$

where $\boldsymbol{H}(t) := \boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t)$ and $o(t)$ is a sequence of positive scalar. To enhance tracking capability of process changes, $o(t)$ is usually given by

$$1/o(t) = \gamma/o(t-1) + 1, \quad (10)$$

where initial term $o(m)$ denotes a small number specified by the designer, and γ denotes forgetting factor. In addition, to prevent $\boldsymbol{H}(t)$ from being singular, $\boldsymbol{H}(t)$ can be modified according to Robbins-Monro [20] method

$$\boldsymbol{H}(t) = \boldsymbol{H}(t-1) + o(t)\left(\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t) - \boldsymbol{H}(t-1)\right). \quad (11)$$

In summary, the proposed gradient-based recursive identification (RI) method is formulated as

Gradient-based recursive identification (RI) method

Begin

1. Initialize the vectors and matrices $\hat{\boldsymbol{\theta}}(m), \boldsymbol{H}(m), o(m)$ at epoch m.

2. For $t \geq m+1$, implement the proposed recursive identification algorithm

$$\boldsymbol{\Phi}(t) = [\boldsymbol{\varphi}(t-m+1), \ \ldots, \ \boldsymbol{\varphi}(t)]^T, \quad (12)$$

$$\boldsymbol{Y}(t) = \left[y(t-m+1), \ \ldots, \ y(t)\right]^T, \quad (13)$$

$$\boldsymbol{W}(t) = diag(\bar{b}^{m-1}\bar{d}_m, \ \ldots, \ \bar{b}^1\bar{d}_m, \ \bar{d}_m), \quad (14)$$

$$1/o(t) = \gamma/o(t-1) + 1, \quad (15)$$

$$\boldsymbol{H}(t) = \boldsymbol{H}(t-1) + o(t)\left(\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t) - \boldsymbol{H}(t-1)\right), \quad (16)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + o(t)\boldsymbol{H}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\left[\boldsymbol{Y}(t) - \boldsymbol{\Phi}(t)\hat{\boldsymbol{\theta}}(t-1)\right] \quad (17)$$

3. After getting $\hat{\boldsymbol{\theta}}(t)$, the parameter estimates $\{\hat{a}_i(t)\}_{i=1}^p$, $\{\hat{d}_l(t)\}_{l=1}^q$ and $\{\hat{b}_j(t)\}_{j=1}^p$ can be further calculated. According to [21], the average method gives the following parameter estimates

$$\{\hat{a}_i(t)\}_{i=1}^p = \left[\hat{\theta}_1(t), \ \ldots, \ \hat{\theta}_p(t)\right]^T, \quad (18)$$

$$\left\{\hat{d}_l(t)\right\}_{l=1}^q = \frac{1}{p}\sum_{i=1}^p \left[1, \ \frac{\hat{\theta}_{p+i}(t)}{\hat{\theta}_i(t)}, \ \frac{\hat{\theta}_{2p+i}(t)}{\hat{\theta}_i(t)}, \ \ldots, \ \frac{\hat{\theta}_{(q-1)p+i}(t)}{\hat{\theta}_i(t)}\right]^T, \quad (19)$$

$$\left\{\hat{b}_j(t)\right\}_{j=1}^p = \left[\hat{\theta}_{pq+1}(t), \ \ldots, \ \hat{\theta}_{pq+p}(t)\right]^T, \quad (20)$$

where $\hat{d}_1(t) = 1$ ensures uniqueness of the result [21].

4. Increase t by 1 and go to Step 2.

End

It should be mentioned that the proposed RI method in (12)–(20) is based on multiple iterations. This means that the parameter updates for epoch t considers multiple errors of previous moments. Hence, this kind of "memory" ensures more accurate grasp of system characteristics, and enhances robustness of identification. Actually, robustness of identification means the ability that can distinguish system characteristics in complex noise environments. Moreover, in order to keep track of time-varying system characteristics, the attenuation speed $\bar{b}$ and the length m should be selected. That is, the balance between robustness and traceability needs to be guaranteed.

However, the difficulty in this recursive algorithm is that the information vector $\boldsymbol{\varphi}(t)$ contains unknown components $x_1(t-i)$. Hence, in this paper, an adaptive filtering method is exploited to obtain the optimal estimates of $x_1(t)$. It is well known that the extended Kalman filter [29] has been widely used to harvest state estimates for nonlinear state space models. For the system model (1), its equivalent state space model is constructed as following

$$\begin{aligned} \boldsymbol{x}(t) &= \boldsymbol{f}_{t-1}(\boldsymbol{x}(t-1)) + \boldsymbol{B}_{t-1}u(t-1), \\ y(t) &= \boldsymbol{C}\boldsymbol{x}(t) + v(t), \end{aligned} \tag{21}$$

where

$$\begin{aligned} &\boldsymbol{x}(t) = \left[x_1(t),\ x_2(t),\ \ldots,\ x_p(t)\right]^T, \\ &\boldsymbol{B}_{t-1} = \left[b_1(t-1),\ b_2(t-1),\ \ldots,\ b_p(t-1)\right]^T, \\ &\boldsymbol{C} = [1,\ 0,\ \ldots,\ 0], \\ &\boldsymbol{f}_{t-1}(\boldsymbol{x}(t-1)) = \left[f_{1(t-1)}(\boldsymbol{x}(t-1)),\ f_{2(t-1)}(\boldsymbol{x}(t-1)),\ \ldots,\ f_{p(t-1)}(\boldsymbol{x}(t-1))\right]^T, \\ &f_{k(t-1)}(\boldsymbol{x}(t-1)) = a_k(t-1)\sum_{l=1}^{q} d_l(t-1)g_l(x_1(t-1)) + x_{k+1}(t-1),\quad k = 1,\ \ldots,\ (p-1), \\ &f_{p(t-1)}(\boldsymbol{x}(t-1)) = a_n(t-1)\sum_{l=1}^{q} d_l(t-1)g_l(x_1(t-1)), \end{aligned}$$

It should be noted that $\boldsymbol{B}_{t-1} \in R^p$, $\boldsymbol{f}_{t-1}(\boldsymbol{x}(t-1)) \in R^p$ and $\boldsymbol{C} \in R^{1\times p}$. The states of this equivalent model in (21) includes true $x_1(t)$ and other virtual states $x_2(t)$, …, $x_p(t)$. Thus, (21) can be used for extended Kalman filtering. In addition, for unknown noise statistics, estimators should be designed in order to achieve adaptive filtering. Here, the design approach of noise estimator is based on the principle of Sage-Husa maximum a posteriori estimation. From (21), Sage-Husa noise estimator [32] in the framework of Kalman is shown as below

$$\hat{r}(t) = \frac{1}{t}\sum_{j=1}^{t}\left[y(j) - \boldsymbol{C}\hat{\boldsymbol{x}}(j|j-1)\right], \tag{22}$$

$$\hat{R}(t) = \frac{1}{t}\sum_{j=1}^{t}\left[\varepsilon^2(j) - \boldsymbol{C}\boldsymbol{P}(j|j-1)\boldsymbol{C}^T\right], \tag{23}$$

where $\varepsilon(j)$ is the innovation, i.e., $\varepsilon(j) := y(j) - \boldsymbol{C}\hat{\boldsymbol{x}}(j|j-1) - \hat{r}(j)$, $\hat{\boldsymbol{x}}(j|j-1) \in R^p$ is the time update state estimate for epoch j, $\boldsymbol{P}(j|j-1) \in R^{p\times p}$ is the time update error covariance estimate for epoch j, and $\hat{r}(t)$ and $\hat{R}(t)$ denote estimated mean values and estimated variances, respectively. However, according to [32], (22) and (23) are only suitable to the situation where noise prior statistics remains unchanged. If ambient noise changes rapidly and continuously, it will cause the estimator performance to degrade significantly. Based on Sage-Husa estimation, different damping coefficients are also introduced to deal with the information of noise estimations at different moments. Since the initial epoch for adaptive filtering starts from epoch m, the sequence length of damping coefficients should be $(t-m)$ instead of t. The sequence fitted by negative exponential law can be selected as $\{\tilde{b}^{t-m-1}\tilde{d}_{t-m},\ \ldots,\ \tilde{d}_{t-m}\}$, where $\tilde{d}_{t-m} = (1-\tilde{b})/(1-\tilde{b}^{t-m})$ and $0 < \tilde{b} < 1$. Hence, the weighted expressions of (22)–(23) from epoch $(m+1)$ to epoch t can be rewritten as

$$\begin{aligned}\hat{r}(t) &= \sum_{j=m+1}^{t}\left\{\tilde{b}^{t-j}\tilde{d}_{t-m}\cdot\left[y(j)-\boldsymbol{C}\hat{\boldsymbol{x}}(j|j-1)\right]\right\}\\ &= (1-\tilde{d}_{t-m})\hat{r}(t-1)+\tilde{d}_{t-m}\left[y(t)-\boldsymbol{C}\hat{\boldsymbol{x}}(t|t-1)\right],\end{aligned} \tag{24}$$

$$\begin{aligned}\hat{R}(t) &= \sum_{j=m+1}^{t}\left\{\tilde{b}^{t-j}\tilde{d}_{t-m}\cdot\left[\varepsilon^2(j)-\boldsymbol{C}\boldsymbol{P}(j|j-1)\boldsymbol{C}^T\right]\right\}\\ &= (1-\tilde{d}_{t-m})\hat{R}(t-1)+\tilde{d}_{t-m}\left[\varepsilon^2(t)-\boldsymbol{C}\boldsymbol{P}(t|t-1)\boldsymbol{C}^T\right],\end{aligned} \tag{25}$$

In (24)–(25), recent noise samples are weighted heavier than past noise samples. In this way, estimated means and covariances can follow noise changes. Moreover, the bigger the value of b is, the greater the proportion of past noise statistics is.

Because the state space model (21) is a discrete output error model, the extended Kalman filter can be established according to [26]. In order to deal with unknown noise statistics, the weighted noise estimator in (24)–(25) is also applied into the filtering algorithm. Then, parameter estimates $\{\hat{a}_i(t-1)\}_{i=1}^{p}$, $\{\hat{d}_l(t-1)\}_{l=1}^{q}$ and $\{\hat{b}_j(t-1)\}_{j=1}^{p}$ at epoch $(t-1)$ are used to form $\hat{\boldsymbol{f}}_{t-1}(\cdot)$ and $\hat{\boldsymbol{B}}_{t-1}$. Thus, the adaptive EKF algorithm for Wiener output-error systems is concluded as below

$$\hat{\boldsymbol{x}}(t|t-1)=\hat{\boldsymbol{f}}_{t-1}(\hat{\boldsymbol{x}}(t-1|t-1))+\hat{\boldsymbol{B}}_{t-1}u(t-1), \tag{26}$$

$$\boldsymbol{P}(t|t-1)=\left(\frac{\partial\hat{\boldsymbol{f}}_{t-1}}{\partial\boldsymbol{x}}|_{\hat{\boldsymbol{x}}(t-1|t-1)}\right)\boldsymbol{P}(t-1|t-1)\left(\frac{\partial\hat{\boldsymbol{f}}_{t-1}}{\partial\boldsymbol{x}}|_{\hat{\boldsymbol{x}}(t-1|t-1)}\right)^T, \tag{27}$$

$$\hat{r}(t)=(1-\tilde{d}_{t-m})\hat{r}(t-1)+\tilde{d}_{t-m}\left[y(t)-\boldsymbol{C}\hat{\boldsymbol{x}}(t|t-1)\right], \tag{28}$$

$$\varepsilon(t)=y(t)-\boldsymbol{C}\hat{\boldsymbol{x}}(t|t-1)-\hat{r}(t), \tag{29}$$

$$\hat{R}(t)=(1-\tilde{d}_{t-m})\hat{R}(t-1)+\tilde{d}_{t-m}\left[\varepsilon^2(t)-\boldsymbol{C}\boldsymbol{P}(t|t-1)\boldsymbol{C}^T\right], \tag{30}$$

$$\boldsymbol{K}(t)=\boldsymbol{P}(t|t-1)\boldsymbol{C}^T\left(\boldsymbol{C}\boldsymbol{P}(t|t-1)\boldsymbol{C}^T+\hat{R}(t)\right)^{-1}, \tag{31}$$

$$\hat{\boldsymbol{x}}(t|t)=\hat{\boldsymbol{x}}(t|t-1)+\boldsymbol{K}(t)\varepsilon(t), \tag{32}$$

$$\boldsymbol{P}(t|t)=(\boldsymbol{I}-\boldsymbol{K}(t)\boldsymbol{C})\boldsymbol{P}(t|t-1). \tag{33}$$

In (31), $\boldsymbol{K}(t)$ denotes the nonlinear gain for state updates. At initial epoch m, initialize $\hat{\boldsymbol{x}}(m|m)=\boldsymbol{x}_0$ and $\boldsymbol{P}(m|m)=\boldsymbol{P}_0$. After getting the optimal estimated state vector $\hat{\boldsymbol{x}}(t|t)$, $\hat{x}_1(t|t)$ can be extracted to construct the information vector $\boldsymbol{\varphi}(t)$. Specifically, the estimated information vectors are formed as

$$\begin{aligned}\hat{\boldsymbol{\varphi}}_k^T(t)&=\left[g_k(\hat{x}_1(t-1|t-1)),\ g_k(\hat{x}_1(t-2|t-2)),\ \dots,\ g_k(\hat{x}_1(t-p|t-p))\right],\ \ k=1,\ \dots,q,\\ \boldsymbol{\varphi}_u{}^T(t)&=\left[u(t-1),\ \dots,\ u(t-p)\right],\\ \hat{\boldsymbol{\varphi}}^T(t)&=\left[\hat{\boldsymbol{\varphi}}_1^T(t),\ \dots,\ \hat{\boldsymbol{\varphi}}_q^T(t),\ \boldsymbol{\varphi}_u{}^T(t)\right]\end{aligned} \tag{34}$$

In conclusion, the adaptive filtering algorithm and the recursive identification algorithm both start at epoch $(m+1)$. That is, the relationship can be shown as below

$$t = m + 1,$$

$$\hat{x}_1(t-1|t-1),\ \hat{\boldsymbol{\theta}}(t-1) \Rightarrow \hat{\boldsymbol{\theta}}(t),$$

$$\hat{\boldsymbol{x}}(t-1|t-1),\ \hat{\boldsymbol{\theta}}(t-1) \Rightarrow \hat{\boldsymbol{x}}(t|t) \Rightarrow \hat{x}_1(t|t),$$

$$t = t + 1.$$

Hence, the specific algorithm can be drawn as

Adaptive filtering-based recursive identification (AF-RI) method

Begin

1. Initialize the vectors and matrices $\hat{\boldsymbol{\theta}}(m), \boldsymbol{H}(m), o(m), \hat{\boldsymbol{x}}(m|m), \boldsymbol{P}(m|m)$ at epoch m.

2. Initialize terms $\hat{x}_1(m-1|m-1),\ \ldots,\ \hat{x}_1(2-p|2-p)$ for the information matrix $\hat{\boldsymbol{\Phi}}(m+1)$.

3. For $t \geq m+1$, form the estimated information vectors by (34), and implement the proposed recursive identification algorithm

$$\hat{\boldsymbol{\Phi}}(t) = \left[\hat{\boldsymbol{\varphi}}(t-m+1),\ \ldots,\ \hat{\boldsymbol{\varphi}}(t)\right]^T, \tag{35}$$

$$\hat{\boldsymbol{H}}(t) = \hat{\boldsymbol{H}}(t-1) + o(t)\left(\hat{\boldsymbol{\Phi}}^T(t)\boldsymbol{W}(t)\hat{\boldsymbol{\Phi}}(t) - \hat{\boldsymbol{H}}(t-1)\right), \tag{36}$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + o(t)\hat{\boldsymbol{H}}^{-1}(t)\hat{\boldsymbol{\Phi}}^T(t)\boldsymbol{W}(t)\left[\boldsymbol{Y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\theta}}(t-1)\right]. \tag{37}$$

4. Harvest newest parameter estimations $\{\hat{a}_i(t)\}_{i=1}^{p}$, $\{\hat{d}_l(t)\}_{l=1}^{q}$ and $\{\hat{b}_j(t)\}_{j=1}^{p}$ by (18)–(20).

5. Construct the virtual equivalent state space model by (21).

6. For $t \geq m+1$, implement the proposed adaptive filtering algorithm by (26)–(33).

7. Extract the current optimal estimate $\hat{x}_1(t|t)$. Increase t by 1 and go to Step 3.

End

Remark 1. The sensitivity of proposed Sage-Husa Kalman filter is affected by attenuation speed $\tilde{b}$ in a large extent. When the value of $\tilde{b}$ gets bigger, attenuation becomes slower, and $\tilde{d}_{t-m}$ decreases to a small value. Then, in (28)–(30), $\hat{r}(t)$ and $\hat{R}(t)$ of noise estimator change slowly in face of noise changes. Hence, under unknown noise statistics, the bigger value of $\tilde{b}$ makes the proposed filter difficult to track noise changes, and the sensitivity of nonlinear gain $\boldsymbol{K}(t)$ decreases significantly. It should also be noted that the proposed Sage-Husa Kalman filter can deal with situations with non-zero-mean noise. The filter can recursively estimates both mean values and variances of noise.

4. Main performance analysis

Martingale convergence theorem [35] denotes the applications of Lyapunov stability theories in stochastic systems. The theorem is usually adopted to analyze convergence of time-invariant systems. Meanwhile, the martingale hyperconvergence theorem [34] is obtained to study bounded convergence of time-varying systems.

Lemma 1 [34]. *Define the non-negative function* $V(t) = V(x(t))$ *and the set*

$$R_t = [x(t) : h(x(t)) \leq \eta_{\max} < \infty, \ a.\ s.], \tag{38}$$

where $\eta_{\max}$ *denotes the upper bound. Define* R_t^c *as the complementary set of* R_t. *For* $x(t) \in R_t^c$, *if V(t) satisfies the following inequality*

$$E(V(t)) - V(t-1) \leq -b(t), \ \ a.\ s. \tag{39}$$

Meanwhile, for $x(t) \in R_t^c$, *the random variable* $b(t)$ *satisfies* $b(t) > 0$. *Then, for sufficiently large t, i.e.,* $t > t_0$, *we have* $x(t) \in R_t$, *or* $\lim_{t\to\infty} x(t) \in R_t$, $a.\ s.$

Assumptions 5 and 6. For the proposed RI method in (12)–(20) and any t ($t \geq m+1$), there exist $\alpha > 0$ and $\beta > 0$ such that the following persistent excitation condition holds

$$\alpha \boldsymbol{I} \leq \boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t) = \sum_{i=1}^{m} \left[(\bar{b}^{i-1}\bar{d}_m)\big(\boldsymbol{\varphi}(t-i+1)\boldsymbol{\varphi}^T(t-i+1)\big)\right] \leq \beta \boldsymbol{I}. \tag{A5}$$

In addition, the information vector $\boldsymbol{\varphi}(t)$ is bounded, i.e.,

$$0 \leq \|\boldsymbol{\varphi}(t)\|^2 \leq M < \infty. \tag{A6}$$

It should be noted that the assumptions (A5)–(A6) ensures system identifiability.

Theorem 1. *For* $t \geq m+1$, *consider the time-varying Wiener output-error system in* (1)–(2). *The noise* $v(t)$ *satisfies the assumptions* (A1)–(A2), *and the parameter variation* $\boldsymbol{\omega}(t)$ *satisfies the assumptions* (A3)–(A4). *Further, considering the proposed RI method in* (12)–(20), *the assumptions* (A5)–(A6) *are satisfied. Afterwards, the parameter estimation error* $\tilde{\boldsymbol{\theta}}(t)$ *is bounded, and*

$$\lim_{t\to\infty}\left\|\tilde{\boldsymbol{\theta}}(t)\right\|^2 \leq \frac{1}{\alpha(1-\gamma)}\left(\frac{m^2\bar{d}_m^2\sigma_v^2 M}{\alpha} + \frac{\gamma\sigma_w^2}{o_0 p_0} + \frac{\beta\sigma_w^2}{1-\gamma}\right), \tilde{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t). \tag{40}$$

Proof. (15) and (16) give

$$\frac{\boldsymbol{H}(t)}{o(t)} = \frac{1-o(t)}{o(t)}\boldsymbol{H}(t-1) + \boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t) = \gamma\frac{\boldsymbol{H}(t-1)}{o(t-1)} + \boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t). \tag{41}$$

Define $\boldsymbol{G}(t) = \boldsymbol{H}(t)/o(t)$. Then, we have

$$\boldsymbol{G}(t) = \gamma\boldsymbol{G}(t-1) + \boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t), \quad \boldsymbol{G}(m) = \frac{1}{o_0 p_0}\boldsymbol{I}, \quad \mathrm{o}(m) = o_0. \tag{42}$$

From the definition of $\boldsymbol{\omega}(t)$ and (17), $\tilde{\boldsymbol{\theta}}(t)$ can be written as

$$\tilde{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{G}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\Big[\boldsymbol{Y}(t) - \boldsymbol{\Phi}(t)\hat{\boldsymbol{\theta}}(t-1)\Big] - \boldsymbol{\omega}(t). \tag{43}$$

Define $\boldsymbol{\Gamma}(t) = [v(t-m+1), \cdots, v(t)]^T$ and $\boldsymbol{\Phi}(t)\tilde{\boldsymbol{\theta}}(t-1) = \tilde{\boldsymbol{Y}}(t)$, it is easy to get

$$\boldsymbol{Y}(t) = \boldsymbol{\Phi}(t)\boldsymbol{\theta}(t-1) + \boldsymbol{\Gamma}(t), \tilde{\boldsymbol{\theta}}(t) = \tilde{\boldsymbol{\theta}}(t-1) + \boldsymbol{G}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\Big(-\tilde{\boldsymbol{Y}}(t) + \boldsymbol{\Gamma}(t)\Big) - \boldsymbol{\omega}(t). \tag{44}$$

Define a non-negative definite function

$$V(t) = \tilde{\boldsymbol{\theta}}^T(t)\boldsymbol{G}(t)\tilde{\boldsymbol{\theta}}(t). \tag{45}$$

From (44)–(45), it yields

$$\begin{aligned} V(t) \ = \ & \gamma V(t-1) - \tilde{\boldsymbol{Y}}^T(t)\big(\boldsymbol{W}(t) - \boldsymbol{W}(t)\boldsymbol{\Phi}(t)\boldsymbol{G}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\big)\tilde{\boldsymbol{Y}}(t) \\ & + \boldsymbol{\Gamma}^T(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t)\boldsymbol{G}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\boldsymbol{\Gamma}(t) + \boldsymbol{\omega}^T(t)\boldsymbol{G}(t)\boldsymbol{\omega}(t) \\ & + 2\tilde{\boldsymbol{Y}}^T(t)\big(\boldsymbol{W}(t) - \boldsymbol{W}(t)\boldsymbol{\Phi}(t)\boldsymbol{G}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\big)\boldsymbol{\Gamma}(t) - 2\tilde{\boldsymbol{\theta}}^T(t-1)\boldsymbol{G}(t)\boldsymbol{\omega}(t) \\ & - 2\Big(-\tilde{\boldsymbol{Y}}^T(t) + \boldsymbol{\Gamma}^T(t)\Big)\boldsymbol{W}(t)\boldsymbol{\Phi}(t)\boldsymbol{\omega}(t). \end{aligned} \tag{46}$$

From (42), $\boldsymbol{G}(t)$ can be written as

$$\boldsymbol{G}(t) = \gamma^{t-m}\boldsymbol{G}(m) + \sum_{i=0}^{t-m-1} \gamma^i \boldsymbol{\Phi}^T(t-i)\boldsymbol{W}(t-i)\boldsymbol{\Phi}(t-i). \tag{47}$$

For $t \geq m+1$, (A5) gives the following inequality

$$\boldsymbol{I}\left(\frac{\gamma^{t-m}}{o_0 p_0} + \frac{1-\gamma^{t-m}}{1-\gamma}\alpha\right) \leq \boldsymbol{G}(t) \leq \boldsymbol{I}\left(\frac{\gamma^{t-m}}{o_0 p_0} + \frac{1-\gamma^{t-m}}{1-\gamma}\beta\right), \boldsymbol{G}^{-1}(t) \leq \frac{1-\gamma}{(1-\gamma^{t-m})\alpha}\boldsymbol{I} \leq \frac{\boldsymbol{I}}{\alpha}. \tag{48}$$

In (46), $\tilde{\boldsymbol{Y}}^T(t)$ and $\boldsymbol{\Gamma}(t)$ are uncorrelated, and $\tilde{\boldsymbol{\theta}}^T(t-1)$, $\boldsymbol{\Gamma}(t)$, $\tilde{\boldsymbol{Y}}^T(t)$ and $\boldsymbol{\omega}(t)$ are uncorrelated. Thus, taking the expectation of both sides of (46) and using the assumptions (A1) and (A3), it follows that

$$\begin{aligned} \mathrm{E}(V(t)) = & \gamma V(t-1) - \tilde{\boldsymbol{Y}}^T(t)\big(\boldsymbol{W}(t) - \boldsymbol{W}(t)\boldsymbol{\Phi}(t)\boldsymbol{G}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\big)\tilde{\boldsymbol{Y}}(t) \\ & + \mathrm{E}\big(\boldsymbol{\Gamma}^T(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t)\boldsymbol{G}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\boldsymbol{\Gamma}(t)\big) + \mathrm{E}\big(\boldsymbol{\omega}^T(t)\boldsymbol{G}(t)\boldsymbol{\omega}(t)\big), \end{aligned} \tag{49}$$

where $\mathrm{E}(V(t))$ is the expectation of $V(t)$. Because $\boldsymbol{W}(t) - \boldsymbol{W}(t)\boldsymbol{\Phi}(t)\boldsymbol{G}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t) \geq 0$, there exists $\tilde{\boldsymbol{Y}}^T(t)(\boldsymbol{W}(t) - \boldsymbol{W}(t)\boldsymbol{\Phi}(t)\boldsymbol{G}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t))\tilde{\boldsymbol{Y}}(t) \geq 0$. From (49), it gives

$$\mathrm{E}(V(t)) \leq \gamma V(t-1) + \mathrm{E}\big(\boldsymbol{\Gamma}^T(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t)\boldsymbol{G}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\boldsymbol{\Gamma}(t)\big) + \mathrm{E}\big(\boldsymbol{\omega}^T(t)\boldsymbol{G}(t)\boldsymbol{\omega}(t)\big). \tag{50}$$

Let $\boldsymbol{\rho}_i(t)$ be the ith row of $\boldsymbol{W}(t)\boldsymbol{\Phi}(t)$, i.e., $\boldsymbol{\rho}_i(t) = \bar{b}^{m-i}\bar{d}_m\boldsymbol{\varphi}^T(t-m+i)$. Then, assumptions (A2) and (A6) yields

$$\begin{aligned} \mathrm{E}\big(\boldsymbol{\Gamma}^T(t)\boldsymbol{W}(t)\boldsymbol{\Phi}(t)\boldsymbol{G}^{-1}(t)\boldsymbol{\Phi}^T(t)\boldsymbol{W}(t)\boldsymbol{\Gamma}(t)\big) & \leq m\sum_{i=1}^{m}\big[\boldsymbol{\rho}_i(t)G^{-1}(t)\boldsymbol{\rho}_i^T(t)\cdot \mathrm{E}\big(v^2(t-m+i)\big)\big] \\ & \leq \big(m^2\bar{d}_m^2\sigma_v^2 M\big)/\alpha. \end{aligned} \tag{51}$$

Further, from (A4), it yields

$$\mathrm{E}\big(\boldsymbol{\omega}^T(t)\boldsymbol{G}(t)\boldsymbol{\omega}(t)\big) \leq \left(\frac{\gamma^{t-m}}{o_0 p_0} + \frac{1-\gamma^{t-m}}{1-\gamma}\beta\right)\sigma_w^2 \leq \left(\frac{\gamma}{o_0 p_0} + \frac{\beta}{1-\gamma}\right)\sigma_w^2. \tag{52}$$

Define $\eta_{\max}$ as $\eta_{\max} = \frac{m^2\bar{d}_m^2\sigma_v^2 M}{\alpha} + (\frac{\gamma}{o_0 p_0} + \frac{\beta}{1-\gamma})\sigma_w^2$, and $b(t)$ as $b(t) = (1-\gamma)V(t-1) - \eta_{\max}$. Then, from (51)–(52), (50) can be written as

$$\mathrm{E}[V(t)] - V(t-1) \leq -b(t). \tag{53}$$

Consider the set $R_t = [\tilde{\boldsymbol{\theta}}(t) : (1-\gamma)V(t-1) \le \eta_{\max},\ \text{a. s.}]$, and define R_t^c as the complementary set of R_t. Thus, for $x(t) \in R_t^c$, there exists $b(t) > 0$. Applying martingale hyperconvergence theorem of Lemma 1, it is known that

$$\lim_{t\to\infty} \tilde{\boldsymbol{\theta}}(t) \in R_t, \quad \text{a. s.} \tag{54}$$

Therefore, we have

$$\lim_{t\to\infty} V(t) \le \frac{m^2 \bar{d}_m^2 \sigma_v^2 M}{(1-\gamma)\alpha} + \frac{\gamma\sigma_w^2}{o_0 p_o (1-\gamma)} + \frac{\beta\sigma_w^2}{(1-\gamma)^2}. \tag{55}$$

Then the parameter estimation error $\tilde{\boldsymbol{\theta}}(t)$ satisfies

$$\begin{aligned} \lim_{t\to\infty} \left\|\tilde{\boldsymbol{\theta}}(t)\right\|^2 &\le \frac{m^2 \bar{d}_m^2 \sigma_v^2 M}{(1-\gamma)\alpha^2} + \frac{\gamma\sigma_w^2}{\alpha o_0 p_0 (1-\gamma)} + \frac{\beta\sigma_w^2}{\alpha(1-\gamma)^2} \\ &= \frac{1}{\alpha(1-\gamma)}\left(\frac{m^2 \bar{d}_m^2 \sigma_v^2 M}{\alpha} + \frac{\gamma\sigma_w^2}{o_0 p_0} + \frac{\beta\sigma_w^2}{1-\gamma}\right). \end{aligned} \tag{56}$$

This proves Theorem 1.

In Theorem 1, the assumptions (A1)–(A4) specify conditions of both noise and parameter variation. The assumptions (A5) and (A6) show both persistent excitation property and boundedness of information vectors. By assuming that the information vector $\boldsymbol{\varphi}(t)$ is known, Theorem 1 shows the bounded convergence of time-varying Wiener output-error systems.

Corollary 1. *From* (4) *and* (34)*, define* $\hat{v}(t) = y(t) - \hat{\boldsymbol{\varphi}}^T(t)\boldsymbol{\theta}(t-1)$. *Assume that the error* $\hat{v}(t)$ *satisfies the assumptions* (A1)–(A2)*, and the parameter variation* $\boldsymbol{\omega}(t)$ *satisfies the assumptions* (A3)–(A4)*. Further, considering the proposed AF-RI method, assumptions* (A5) *and* (A6) *are satisfied,* i.e.*,* $\alpha \boldsymbol{I} \le \hat{\boldsymbol{\Phi}}^T(t)\boldsymbol{W}(t)\hat{\boldsymbol{\Phi}}(t) \le \beta \boldsymbol{I}$ *and* $0 \le \|\hat{\boldsymbol{\varphi}}(t)\|^2 \le M < \infty$. *Then, the parameter estimation error* $\tilde{\boldsymbol{\theta}}(t)$ *is bounded, and Eq. (40) is obtained.*

Proof. The proof of this corollary is similar to the proof of Theorem 1, and is thus omitted.

Remark 2. By selecting the suitable value of $\tilde{b}$, $\hat{r}(t)$ and $\hat{R}(t)$ of Sage-Husa noise estimator can record noise changes correctly. The nonlinear gain $\boldsymbol{K}(t)$ is sensitive to noise changes, and the adaptive Kalman filter in (26)–(33) can approximate the Kalman filter in [26]. In fact, under uniform complete observability, [26] proves that the error dynamics equation of Kalman filter is asymptotically stable for linear time-varying output-error systems. Meanwhile, recursive identification based on the Gauss-Newton technique can provide more and more accurate parameter estimates. Thus, with the improvement of parameter estimation accuracy, the proposed adaptive filter based on Sage-Husa principle can provide more accurate estimate of $\hat{\boldsymbol{\varphi}}^T(t)$. This leads to uncorrelation between $\hat{v}(t)$ and process data. Therefore, the conditions of Corollary 1 are easy to be satisfied.

Remark 3. Both the theorem and the corollary in this paper are the extensions of convergence results in [35] and [36]. Specifically, the asymptotic convergence of linear time-invariant systems is given in [35], and the bounded convergence of linear time-varying systems is obtained in [36].

5. Numerical simulation

In this section, the time-varying Wiener output-error systems are considered for identification. Three examples are used to demonstrate the effectiveness of proposed AF-RI method. Meanwhile, by replacing unknown $x_1(t-i)$ in $\boldsymbol{\varphi}(t)$ with measurements $y(t-i)$, the proposed RI method in (12)–(20) is also implemented. These two identification methods are used to illustrate the following aspects: (i) the effectiveness of proposed recursive identification; (ii) the better performance of AF-RI method; (iii) the effectiveness of dealing with unknown noise statistics. Here, the same system structure is used for examples 1 and 2, and it is taken the following form

$$x_1(t)=\sum_{i=1}^{2}a_i(t-1)\sum_{l=1}^{3}d_l(t-1)g_l(x_1(t-i))+\sum_{j=1}^{2}b_j(t-1)u(t-j), \tag{57}$$

$$y(t)=x_1(t)+v(t), \tag{58}$$

$$\boldsymbol{\theta}(t)=[a_1(t),a_2(t),d_1(t),d_2(t),d_3(t),b_1(t),b_2(t)]^T, \tag{59}$$

where the vector $[g_1(y(t-i)),\ g_2(y(t-i)),\ g_3(y(t-i))]^T$ contains nonlinear basis functions, and the vector $\boldsymbol{\theta}(t)$ denotes the parameters of system. In simulation tests, the attenuation speed $\bar{b}=\tilde{b}=0.9$, the sample number $m=20$, and the forgetting factor $\gamma=0.99$. Furthermore, the initial parameters and states are taken as $\hat{\boldsymbol{x}}(m|m)=\boldsymbol{x}_0=10^{-6}*[1,1]^T$ and $\hat{\boldsymbol{\theta}}(m)=10^{-6}*[1,\dots,1]^T$, respectively. For the proposed Kalman filter, there exist $\boldsymbol{P}(m|m)=\boldsymbol{P}_0=10*\boldsymbol{I}_{2\times2}$, $\hat{R}(m)=10^6$ and $\hat{r}(m)=10^6$.

Example 1. In this example, two time-varying systems are adopted to explain the effectiveness of proposed AF-RI method. These two systems are shown in Situation 1 and Situation 2, respectively. Then, in Situation 1, the fluctuations of system parameters are set to be extremely small. Meanwhile, in Situation 2, the slow time-varying characteristics of system are shown, and the linear changes on system parameters are introduced. It should also be noted that the identified models have same structures as the systems in both situations.

Situation 1. Consider the following time-varying Wiener output-error system

$$x_1(t)=\sum_{i=1}^{2}a_i\big[d_1x_1(t-i)+d_2(x_1(t-i))^2+d_3(x_1(t-i))^3\big]+\sum_{j=1}^{2}b_ju(t-j), \tag{60}$$

$$y(t)=x_1(t)+v_1(t), \tag{61}$$

$$\boldsymbol{\theta}=[a_1,a_2,b_1,b_2,d_1,d_2,d_3]^T=\begin{bmatrix}N(0.18,\ 0.001^2)\\N(0.23,\ 0.001^2)\\N(-0.3,\ 0.001^2)\\N(1.0,\ 0.001^2)\\N(1.0,\ 0.001^2)\\N(0.2,\ 0.001^2)\\N(-0.33,\ 0.001^2)\end{bmatrix}, \tag{62}$$

where the true parameters conform to normal distributions, and fluctuate slightly around certain values. The inputs $\{u(t)\}$ are taken as persistent excitation signals with zero mean and unit variance. The noise $\{v_1(t)\}$ is taken as the three-term mixture Gaussian noise

$$v_1(t)\sim 0.8*N\big(0,\ 0.2^2\big)+0.15*N\big(0,\ 50*0.2^2\big)+0.05*N\big(0,\ 100*0.2^2\big), \tag{63}$$

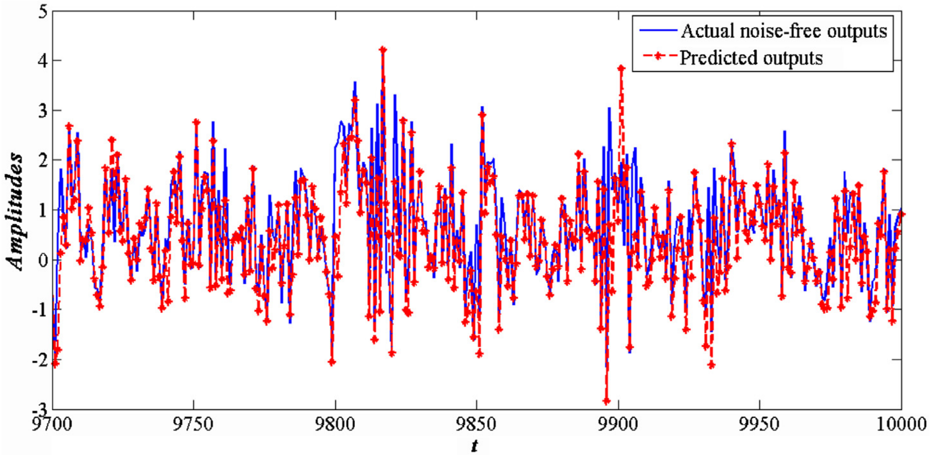

Fig. 2. The predicted output quality of RI method in Situation 1 of Example 1.

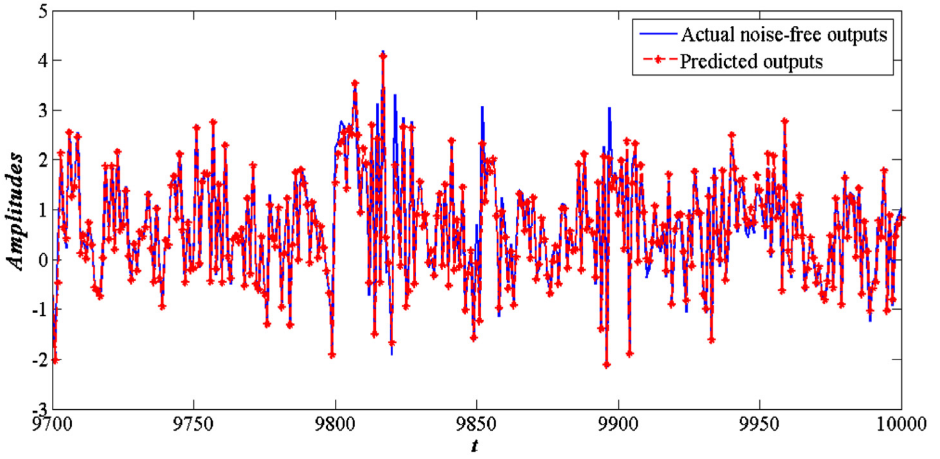

Fig. 3. The predicted output quality of AF-RI method in Situation 1 of Example 1.

where $v_1(t)$ denotes the non-Gaussian white noise. In (63), the term $N(0,\ 0.2^2)$ represents nominal ambient noise, and the terms $N(0,\ 50 * 0.2^2)$ and $N(0,\ 100 * 0.2^2)$ represent impulsive components. The probabilities that the impulses occur are 0.15 and 0.05. For the models in (60)–(63), noise and parameter variations satisfy the assumptions (A1)–(A4) approximately.

Further, the proposed AF-RI and RI methods are applied to estimate parameters. Set the total data length $L = 10{,}000$, and assume $\delta(t) := \|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)\| / \|\boldsymbol{\theta}(t)\| \times 100\%$ as the relative estimation error. The predicted outputs produced by two methods are also compared with the actual noise-free outputs in Figs. 2–3. From the figures, the predicted outputs obtained by AF-RI method are closer to the actual noise-free outputs. Fig. 4 describes the change processes of relative estimation errors for both RI and AF-RI methods. From Figs. 2–4, we can get the following conclusions

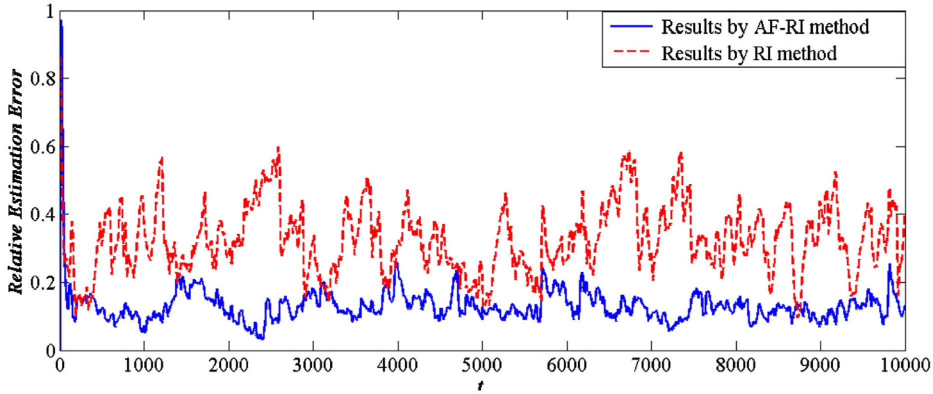

Fig. 4. The relative estimation error of parameters in Situation 1 of Example 1.

i) The relative estimation errors are fluctuating.
ii) AF-RI method can generate more accurate estimations of system parameters.
iii) AF-RI method generates the more accurate model to track time-varying system.

Situation 2. Consider the following slow time-varying Wiener output-error system

$$x_1(t) = \sum_{i=1}^{2} a_i[d_1x_1(t-i) + d_2sin(x_1(t-i)) + d_3cos(x_1(t-i))] + \sum_{j=1}^{2} b_ju(t-j), \tag{64}$$

$$y(t) = x_1(t) + v_1(t), \tag{65}$$

$$\boldsymbol{\theta} = [a_1, a_2, b_1, b_2, d_1, d_2, d_3]^T = \begin{bmatrix} 0.25 + 0.0001 * t \\ 0.28 + 0.0001 * t \\ -0.3 - 0.0001 * t \\ 1.0 + 0.0001 * t \\ 1.0 \\ -0.5 + 0.0001 * t \\ -0.3317 + 0.0002 * t \end{bmatrix}, \tag{66}$$

where the parameters to be identified change slowly and linearly. In order to ensure uniqueness of identification, the parameter d_1 of nonlinear part is equal to 1, and does not participate in identification. The inputs $\{u(t)\}$ are taken as persistent excitation signals with zero mean and unit variance, and the noise $\{v_1(t)\}$ remains same as in (63). Thus, the assumptions (A1)–(A4) are approximately satisfied. With $L = 2000$, RI and AF-RI methods are applied to estimate parameters $\{a_1,\ a_2,\ b_1,\ b_2,\ d_2,\ d_3\}$. The performances of tracking time-varying parameters are shown in Figs. 5–6. From the figures, the following conclusions are got

i) The tracking performances of $\{a_1,\ a_2,\ b_1,\ b_2\}$ of linear part are poor by RI method.
ii) The convergence speed of $\{d_2,\ d_3\}$ of nonlinear part is slow by RI method.
iii) The convergence speed of parameter estimation is faster by AF-RI method.
iv) The better performances of tracking time-varying parameters are got by AF-RI method.

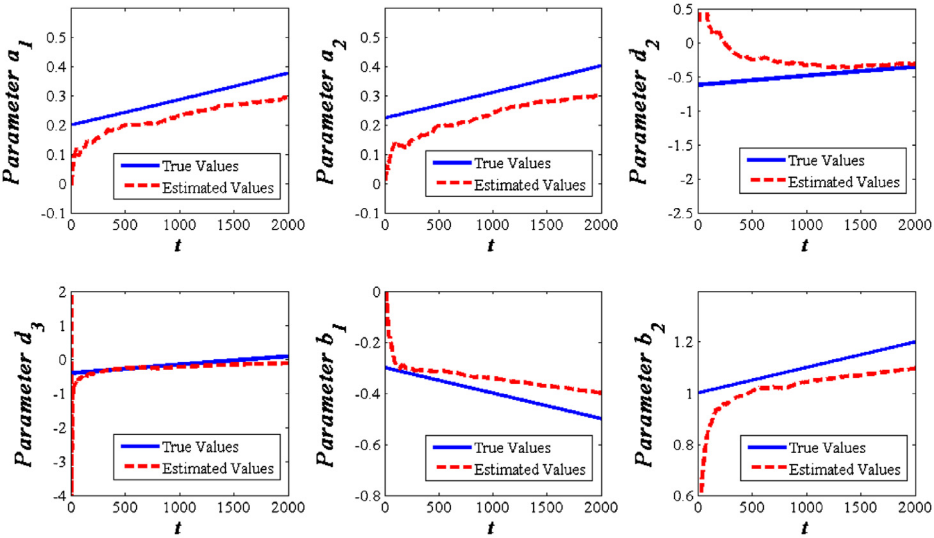

Fig. 5. The performances of tracking time-varying parameters by RI method.

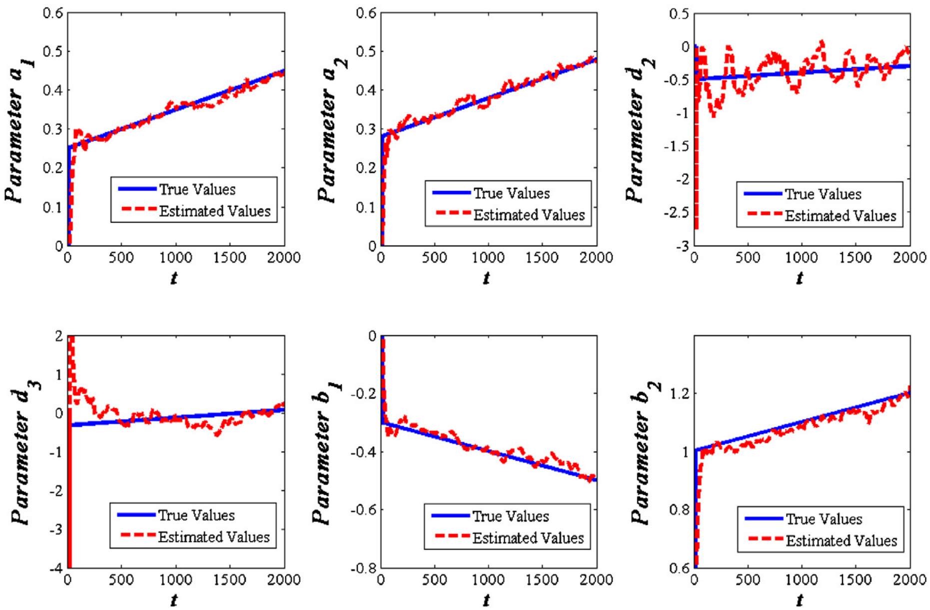

Fig. 6. The performances of tracking time-varying parameters by AF-RI method.

Example 2. In this example, robustness of identification is tested under different noise environments. The system structure adopts the same form described in (64)–(66). Both RI and AF-RI methods are applied into identification. The inputs $\{u(t)\}$ are taken as persistent excitation signals with zero mean and unit variance. Besides, the following root mean square

Table 1
Mean values and standard deviations of RMSE under colored impulsive noise.

Noise	Algorithms	RMSE
$v_2(t)$	RI	0.4714 ± 0.0071
	AF-RI	0.1937 ± 0.0064

error (RMSE) index is taken to evaluate robustness

$$\mathrm{RMSE} = \sqrt{\sum_{t=1}^{L} \big(x_1(t) - \hat{y}(t)\big)^2 / L}, \tag{67}$$

where $x_1(t)$ denotes the actual noise-free output, $\hat{y}(t)$ is the predicted output. With $L = 2000$, the noise sequences $\{v(t)\}$ are taken in the following different situations.

Situation I. The colored impulsive noise is adopted to verify the effectiveness of identification methods. That is, the mixture Gaussian noise $v_1(t)$ in (63) can be seen as white noise, and the colored impulsive noise $v_2(t)$ is produced by moving average (MA) process. The expressions of $v_2(t)$ is shown as below.

$$v_2(t) = C\big(z^{-1}\big)v_2(t) = \big(1 + c_1 * z^{-1} + c_2 * z^{-2}\big)v_1(t), \tag{68}$$

where $c_1 = 0.7,\ \ c_2 = -0.5$, and the measurement outputs $y(t) = x_1(t) + v_2(t)$. Then, the recommended RMSE index is used, and RI and AF-RI methods are run 50 times, respectively. Both mean values and standard deviations of RMSE are shown in Table 1. From the table, it is known that these two identification methods are effective for colored impulsive noise. More accurate predicted outputs are obtained by AF-RI method.

Situation II. Besides colored impulsive noise, the noise estimator based on Sage-Husa principle can also handle noise with time-varying variances. Hence, the noise form is given as

$$v_x(t) \sim 0.8 * N\big(0,\ \bar{\sigma}_v^2(t)\big) + 0.15 * N\big(0,\ 50 * \bar{\sigma}_v^2(t)\big) + 0.05 * N\big(0,\ 100 * \bar{\sigma}_v^2(t)\big), \tag{69}$$

where $\bar{\sigma}_v^2(t)$ denotes the variance function. The expression of $\bar{\sigma}_v^2(t)$ is

$$\bar{\sigma}_v^2(t) = \begin{cases} (r + k * t)^2, & t \le t_0 \\ \bar{\sigma}_v^2(t_0), & t > t_0 \end{cases} \tag{70}$$

where $k = 0.00015$, and t_0 is bigger than the data length L. The noise equations in (69)–(70) satisfy (A1)–(A2). When the term r in (70) is taken as 0.07, 0.1, 0.2, 0.3, four time-varying noise sequences $v_3(t), v_4(t), v_5(t), v_6(t)$ are generated to test identification performance. RI and AF-RI methods are run 50 times, respectively. Both mean values and standard deviations of RMSE are shown in Table 2. From the table, it is known that both RI and AF-RI methods can deal with time-varying impulsive noise effectively. Meanwhile, AF-RI method can provide more accurate identification results under reasonable noise intensities.

Situation III. In order to observe the performance of proposed Sage-Husa Kalman filter, non-random noise is introduced. Consider the following time-varying multi-sinusoidal noise

Table 2
Mean values and standard deviations of RMSE under noise with time-varying variances.

Noise	Algorithms	RMSE1
$v_3(t)$	RI	0.6239 ± 0.0201
	AF-RI	0.3469 ± 0.0305
$v_4(t)$	RI	0.6613 ± 0.0198
	AF-RI	0.4004 ± 0.0464
$v_5(t)$	RI	0.8288 ± 0.0348
	AF-RI	0.5054 ± 0.0606
$v_6(t)$	RI	0.9739 ± 0.0198
	AF-RI	0.6135 ± 0.0570

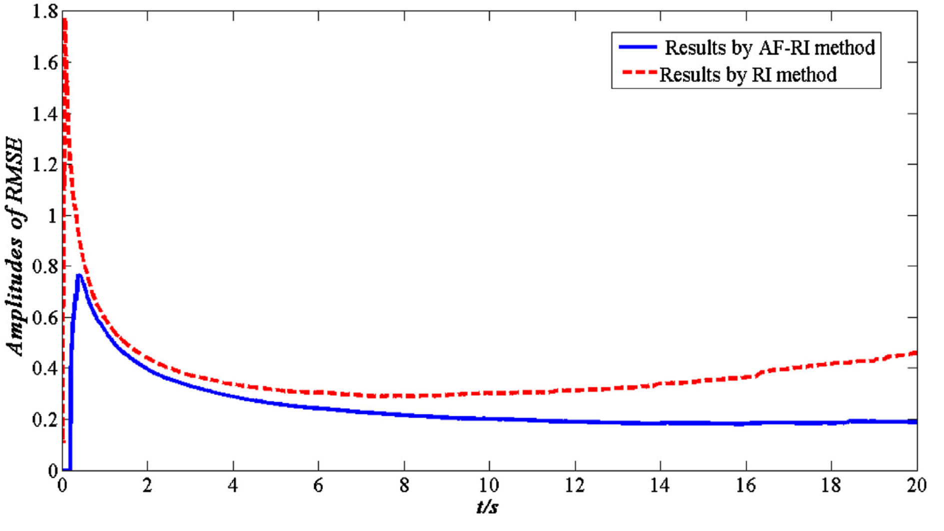

Fig. 7. The changing process of RMSE index under time-varying multi-sinusoidal noise.

$$v_7(t) = (c_1 + k * t)\sin(2\pi f_1 t) + c_2 \sin(2\pi f_2 t) + c_3 \sin(2\pi f_3 t), \tag{71}$$

where $c_1 = 0.25$, $c_2 = 0.2$, $c_3 = 0.1$, $k = 0.01$, and $f_1 = 5Hz$, $f_2 = 12Hz$, $f_3 = 18Hz$. Assuming that the sampling time $T_s = 0.01s$, the changing process of RMSE index is shown in Fig. 7. From the figure, the RMSE values obtained by AF-RI method finally stabilize around 0.19, and the RMSE values by RI method are gradually increasing after $t = 12$s. This illustrates that the proposed Sage-Husa Kalman filter is effective. Further, the performance of AF-RI method is better than the performance of RI method under time-varying multi-sinusoidal noise.

6. Conclusions

The combination of state estimation and system identification provides an effective tool to analyze time-varying and nonlinear behaviors under complex noise environments. This paper proposes AF-RI method to achieve robust identification for time-varying Wiener output-error systems. Meanwhile, this paper tries to solve the problems below.

- The appropriate attenuation speed determines the sensitivity of proposed Sage-Husa Kalman filter, and the sensitive filter can keep track of noise changes.
- The virtual equivalent state space model is constructed to implement adaptive filtering.
- The unknown variables in information vectors are associated with optimal state estimates.
- The bounded convergence for time-varying nonlinear systems is exploited.

The proposed adaptive filtering-based recursive method can also be extended to the identification of other time-varying nonlinear systems under unknown noise statistics.

Acknowledgements

This work is supported by National Natural Science Foundation of China (No. 61703434) and Science Foundation of China University of Petroleum, Beijing (No. 2462016YJRC028).

References

[1] M. Schoukens, K. Tiels, Identification of block-oriented nonlinear systems starting from linear approximations: a survey, Automatica 85 (2017) 272–292.
[2] D. Wang, F. Ding, Parameter estimation algorithms for multivariable Hammerstein CARMA systems, Inform. Sciences 355-356 (2016) 237–248.
[3] A. Wills, B. Ninness, Generalised Hammerstein-Wiener system estimation and a benchmark application, Control Eng. Pract. 20 (11) (2012) 1097–1108.
[4] T. Katayama, H. Ase, Linear approximation and identification of MIMO Wiener-Hammerstein systems, Automatica 71 (2016) 118–124.
[5] Q. Jin, J. Dou, F. Ding, L. Cao, A novel identification method for Wiener systems with the limited information, Math. Comput. Model. 58 (7–8) (2013) 1531–1539.
[6] M.M.D. Silva, T. Wigren, T. Mendonca, Nonlinear identification of a minimal neuromuscular blockade model in anesthesia, IEEE Trans. Control Syst. Technol. 20 (1) (2012) 181–188.
[7] G. Shafiee, M.M. Arefi, M.R. Jahed-Motlagh, A.A. Jalali, Nonlinear predictive control of a polymerization reactor based on piecewise linear Wiener model, Chem. Eng. J. 143 (1–3) (2008) 282–292.
[8] Y. Zhang, B. Fidan, P.A. Ioannou, Backstepping control of linear time-varying systems with known and unknown parameters, IEEE Trans. Autom. Control 48 (11) (2003) 1908–1925.
[9] B. Ni, Q. Zhang, Stability of the Kalman filter for continuous time output error systems, Syst. Control Lett. 94 (2016) 172–180.
[10] J. Vörös, Parameter identification of Wiener systems with multi-segment piecewise-linear nonlinearities, Syst. Control Lett. 56 (2) (2007) 99–105.
[11] X. Hu, H. Chen, Strong consistence of recursive identification for Wiener systems, Automatica 41 (11) (2005) 1905–1916.
[12] S.L. Lacy, D.S. Berntein, Identification of FIR Wiener systems with unknown, non-invertible, polynomial non–linearities, Int. J. Control 76 (15) (2003) 1500–1507.
[13] E.W. Bai, Frequency domain identification of Wiener models, Automatica 39 (9) (2003) 1521–1530.
[14] E.W. Bai, J. Reyland, Towards identification of Wiener systems with the least amount of a priori information: iir cases, Automatica 45 (4) (2009) 956–964.
[15] A. Hagenblad, L. Ljung, A. Wills, Maximum likelihood identification of Wiener models, Automatica 44 (11) (2008) 2697–2705.
[16] D. Wang, F. Ding, Least squares based and gradient based iterative identification for Wiener nonlinear systems, Signal Process 91 (5) (2011) 1182–1189.
[17] H. Chen, in: Recursive identification of errors-in-variables Wiener systems, 25th Chinese Control Conference (CCC 2006), August, 2006, pp. 383–387.

[18] Z. Wang, Q. Jin, X. Liu, Recursive least squares identification of hybrid Box-Jenkins model structure in open-loop and closed-loop, J. Franklin I. 353 (2) (2016) 265–278.
[19] F. Ding, X. Liu, M. Liu, The recursive least squares identification algorithm for a class of Wiener nonlinear systems, J. Franklin I. 353 (7) (2016) 1518–1526.
[20] Z. Xu, Y. Zhu, K. Han, J. Qian, A multi-iteration pseudo-linear regression method and an adaptive disturbance model for MPC, J. Process Contr. 20 (4) (2010) 384–395.
[21] S. Cheng, Y. Wei, D. Sheng, Y. Chen, Y. Wang, Identification for Hammerstein nonlinear ARMAX systems based on multi-innovation fractional order stochastic gradient, Signal Process. 142 (2018) 1–10.
[22] F. Ding, P.X. Liu, G. Liu, Multi-innovation least-squares identification for system modeling, IEEE Trans. Syst. Man Cybern. B 40 (3) (2010) 767–778.
[23] J. Wang, J. Wang, D. Zhang, X. Shao, G. Chen, Kalman filtering through the feedback adaption of prior error covariance, Signal Process 152 (2018) 47–53.
[24] D. Marelli, M. Zamani, M. Fu, B. Ninness, Distributed Kalman filter in a network of linear systems, Syst. Control Lett. 116 (2018) 71–77.
[25] M. Bohner, N. Wintz, The Kalman filter for linear systems on time scales, J. Math. Anal. Appl. 406 (2) (2013) 419–436.
[26] Q. Zhang, On stability of the Kalman filter for discrete time output error systems, Syst. Control Lett. 107 (2017) 84–91.
[27] J. Ma, O. Wu, B. Huang, F. Ding, Expectation maximization estimation for a class of input nonlinear state space systems by using the Kalman smoother, Signal Process 145 (2018) 295–303.
[28] J. Ma, W. Xiong, J. Chen, F. Ding, Hierarchical identification for multivariate Hammerstein systems by using the modified Kalman filter, Digit. Signal Process 11 (6) (2017) 857–869.
[29] W. Liu, H.W. He, F.C. Sun, Vehicle state estimation based on minimum model error criterion combining with extended Kalman filter, J. Franklin I. 353 (4) (2016) 834–856.
[30] Q. Zhang, Adaptive Kalman filter for actuator fault diagnosis, Automatica 93 (1) (2018) 333–342.
[31] Y. Wu, X. Luo, A novel calibration approach of soft sensor based on multi-rate data fusion technology, J. Process Contr. 10 (2010) 1252–1260.
[32] A. Sage, G. Husa, Adaptive filtering with unknown prior statistics,, in: The 10th Joint Automatic Control Conference (JACC 1969), 1969, pp. 760–769.
[33] C. Chen, Z. Liu, K. Xie, Y. Zhang, C.L. Philip Chen, Asymptotic adaptive control of nonlinear systems with elimination of overparametrization in a Nussbaum-like design, Automatica 98 (2018) 277–284.
[34] C. Chen, K. Xie, F.L. Lewis, S. Xie, A. Davoudi, Fully distributed resilience for adaptive exponential synchronization of heterogeneous multi-agent systems against actuator faults, IEEE Trans. Autom. Control. doi:10.1109/TAC.2018.2881148.
[35] Q. Jin, Z. Wang, X. Liu, Auxiliary model-based interval-varying multi-innovation least squares identification for multivariable OE-like systems with scarce measurements, J. Process Contr. 35 (2015) 154–168.
[36] F. Ding, Martingale hyperconvergence theorem and the convergence of forgetting factor least squares algorithm, Control Theory Appl 14 (1) (1997) 90–95 (in Chinese).
[37] Z. Wang, Q. Jin, X. Liu, Iteratively reweighted correlation analysis method for robust parameter identification of multiple-input multiple-output discrete-time systems, IET Signal Process 10 (5) (2016) 549–556.