European Journal of Control 20 (2014) 14–23

Contents lists available at ScienceDirect

European Journal of Control

journal homepage: www.elsevier.com/locate/ejcon

Recursive identification of errors-in-variables Wiener–Hammerstein systems

Bi-Qiang Mu, Han-Fu Chen *

Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, PR China

ARTICLE INFO

Article history:
Received 5 July 2013
Received in revised form
8 October 2013
Accepted 16 October 2013
Recommended by Astolfi Alessandro
Available online 26 October 2013

Keywords:
Wiener–Hammerstein systems
Errors-in-variables
Stochastic approximation
Recursive estimation
Strong consistency

ABSTRACT

This paper considers the recursive identification of errors-in-variables Wiener–Hammerstein system, which is composed of a static nonlinearity sandwiched by two linear dynamic subsystems. Both the system input and output are observed with additive noises being ARMA processes with unknown coefficients. By the stochastic approximation algorithms incorporated with the deconvolution kernel functions, the coefficients of the linear subsystems and the values of the nonlinear function are recursively estimated. All the estimates are proved to converge to the true values with probability one. A simulation example is given to verify the theoretical analysis.

1. Introduction

The block-oriented systems [14] are widely applied to model the practical nonlinear systems owing to their simple structure and excellent modeling ability. The Wiener–Hammerstein system composed of two dynamic linear subsystems with a static nonlinear function in between has a great flexibility for modeling practical systems, for example, sensor systems, electromechanical systems in robotics, mechatronics, biological and chemical systems, and others. The well-studied Hammerstein and Wiener systems can be thought of the special cases of the Wiener–Hammerstein system. Thus, the identification issue of Wiener–Hammerstein systems has received a considerable attention from both theoretical researchers and engineers.

In the early literature [5,3,16] on identification of the Wiener–Hammerstein system, the impulse responses of the two linear subsystems are connected with the correlation functions between the system input and output under the Gaussian input. Based on the maximum likelihood method, a time domain identification algorithm is proposed in [6], and a simple recursive identification technique for multi-input single-output Wiener–Hammerstein system is presented in [4] with the help of a weighted extended least squares method. Some recent work can be found in [25,29,12,18], and among others.

To identify the nonlinear function in a Wiener–Hammerstein system there are parametric [2–4,6,25] and nonparametric approaches [18,15], according to the different descriptions of the nonlinear function. The parametric approach is applied when the nonlinear function is expressed as a linear combination of basis functions such as polynomials, cubic splines functions, piecewise linear functions, neural networks with unknown coefficients, etc. In this case identification turns out to be a parameter estimation problem that can be solved by the standard optimization method such as the gradient method, Newton-Raphson method, and others. The nonparametric approach is used to estimate the values of the nonlinear function at any given point with the help of the kernel functions, requiring no structural information about the nonlinearity. For this there have been some literature [23,22] dealing with nonparametric regression by stochastic approximation involving the kernel functions. Likewise, we adopt the nonparametric method in the paper. To be specific, the stochastic approximation and the deconvolution kernel functions are together used to achieve this. Here we consider the case where the input and output of the system are not accurately available, but they are observed with additive noises, i.e., we intend to identify the errors-in-variables (EIV) Wiener–Hammerstein systems.

There exist some papers on identifiability [1] and identification [24] of the linear EIV systems. Various estimation methods for identifying linear EIV systems, for example, the instrumental variables based methods, the bias-compensation approaches, the Frisch scheme, the frequency domain methods, the prediction error and the ML methods, are well summarized in the survey paper [24], but the methods mentioned there are nonrecursive. The recursive identification for the linear EIV systems is considered under different assumptions on the system input and on the observation noise in [26,8,30,18]. There are also a few papers [28,17,19] on the identification of nonlinear EIV systems.

* Corresponding author. Tel.: +86 10 62579540.
E-mail address: hfchen@iss.ac.cn (H.-F. Chen).

0947-3580/$ - see front matter © 2013 European Control Association. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ejcon.2013.10.005

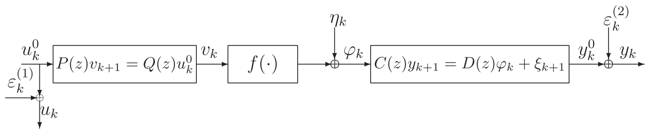

Fig. 1. EIV Wiener–Hammerstein system.

In the paper we consider the SISO EIV Wiener–Hammerstein system (see Fig. 1) described as follows:

$$P(z)v_{k+1} = Q(z)u_k^0, \tag{1}$$

$$\varphi_k = f(v_k) + \eta_k, \tag{2}$$

$$C(z)y_{k+1}^0 = D(z)\varphi_k + \xi_{k+1}, \tag{3}$$

where z denotes the backward-shift operator $zy_{k+1}^0 = y_k^0$, while $f(\cdot)$ is the unknown nonlinear function, and

$$P(z) = 1 + p_1 z + p_2 z^2 + \cdots + p_{n_p} z^{n_p}, \tag{4}$$

$$Q(z) = 1 + q_1 z + q_2 z^2 + \cdots + q_{n_q} z^{n_q}, \tag{5}$$

$$C(z) = 1 + c_1 z + c_2 z^2 + \cdots + c_{n_c} z^{n_c}, \tag{6}$$

$$D(z) = 1 + d_1 z + d_2 z^2 + \cdots + d_{n_d} z^{n_d} \tag{7}$$

are polynomials with unknown coefficients but with known orders n_p, n_q, n_c, n_d. The noise-free input ${u_k}^0$ and output ${y_k}^0$ are observed with additive noises $\varepsilon_k^{(1)}$ and $\varepsilon_k^{(2)}$:

$$u_k = u_k^0 + \varepsilon_k^{(1)}, \quad y_k = y_k^0 + \varepsilon_k^{(2)}. \tag{8}$$

Identification of the EIV Wiener–Hammerstein system is more difficult in comparison with that for the EIV Wiener system discussed in [19]:

(1) The output of the EIV Wiener system is an α-mixing with mixing coefficients decaying exponentially to zero but this is no longer true for the EIV Wiener–Hammerstein system. In [19] it is seen that the mixing property plays an important role in convergence analysis.

(2) Because of the linear subsystem at the output end, more complicated relationships relating the impulse responses and the correlation functions should be taken into account (see Lemma 1).

The goal of this paper is to recursively estimate the unknown parameters of the two linear subsystems $\{p_1, \ldots, p_{n_p}, q_1, \ldots, q_{n_q}, c_1, \ldots, c_{n_c}, d_1, \ldots, d_{n_d}\}$ and the value of $f(x)$ at any given x at the real axis on the basis of the observed data $\{u_k, y_k\}$.

The rest of the paper is arranged as follows. The system assumptions and the recursive algorithms are given in Section 2. The strong consistency of the estimates for the linear and nonlinear parts is proved in Sections 3 and 4, respectively. A numerical example is presented in Section 5, and a brief conclusion is given in Section 6.

2. Assumptions and recursive identification algorithms

2.1. Assumptions

We first give the conditions for identifying the two linear subsystems.

H1: The noise-free input $\{u_k^0\}$ is a sequence of mutually independent, identically distributed (i.i.d.) Gaussian random variables: $u_k^0 \in \mathcal{N}(0, \vartheta^2)$ with unknown $\vartheta > 0$, and is independent of the internal noises $\{\eta_k\}$ and $\{\xi_k\}$ and the observation noises $\{\varepsilon_k^{(1)}\}$ and $\{\varepsilon_k^{(2)}\}$.

H2: $P(z)$ and $Q(z)$ are coprime and $P(z)$ is stable: $P(z) \neq 0$, $\forall |z| \le 1$.

H3: $C(z)$ and $D(z)$ are coprime and both are stable: $C(z) \neq 0$ and $D(z) \neq 0$, $\forall |z| \le 1$.

By the stability of $P(z)$ and $C(z)$, we have

$$L(z) \triangleq \frac{Q(z)}{P(z)} = \sum_{i=0}^{\infty} l_i z^i, \tag{9}$$

$$H(z) \triangleq \frac{D(z)}{C(z)} = \sum_{i=0}^{\infty} h_i z^i, \tag{10}$$

where $|l_i| = O(e^{-r_1 i})$, $r_1 > 0$, $i \ge 1$ and $|h_i| = O(e^{-r_2 i})$, $r_2 > 0$, $i \ge 1$, and $l_0 = 1$ and $h_0 = 1$ since all polynomials (4)–(7) are monic. The numbers $\{l_i, i \ge 0\}$ and $\{h_i, i \ge 0\}$ are called the impulse responses of the two linear subsystems, respectively.

H4: Both the measurement noises $\{\varepsilon_k^{(1)}\}$ and $\{\varepsilon_k^{(2)}\}$ belong to the ARMA process:

$$F_1(z)\varepsilon_k^{(1)} = G_1(z)\varsigma_k^{(1)}, \quad F_2(z)\varepsilon_k^{(2)} = G_2(z)\varsigma_k^{(2)}, \tag{11}$$

where

$$F_1(z) = 1 + f_{1,1} z + f_{1,2} z^2 + \cdots + f_{1,n_{f_1}} z^{n_{f_1}}, \tag{12}$$

$$G_1(z) = 1 + g_{1,1} z + g_{1,2} z^2 + \cdots + g_{1,n_{g_1}} z^{n_{g_1}}, \tag{13}$$

$$F_2(z) = 1 + f_{2,1} z + f_{2,2} z^2 + \cdots + f_{2,n_{f_2}} z^{n_{f_2}}, \tag{14}$$

$$G_2(z) = 1 + g_{2,1} z + g_{2,2} z^2 + \cdots + g_{2,n_{g_2}} z^{n_{g_2}}. \tag{15}$$

The polynomial $F_1(z)$ has no common factor with $G_1(z)G_1(z^{-1})z^{n_{g_1}}$, and $F_1(z)$ and $F_2(z)$ are both stable. The driven noises $\{\varsigma_k^{(1)}\}$ and $\{\varsigma_k^{(2)}\}$ and the internal noises $\{\eta_k\}$ and $\{\xi_k\}$ are mutually independent, and each of them is a sequence of i.i.d. zero mean random variables with probability density. Moreover, $E(|\eta_k|^{\Delta}) < \infty$, $E(|\xi_k|^{\Delta}) < \infty$, $E(|\varsigma_k^{(1)}|^{\Delta+3}) < \infty$, and $E(|\varsigma_k^{(2)}|^{\Delta}) < \infty$ for some $\Delta > 3$.

H5: The nonlinear function $f(\cdot)$ is measurable and has both the left limit $f(x^-)$ and the right limit $f(x^+)$ at any point x. The growth rate of $f(x)$ as $|x| \to \infty$ is not faster than a polynomial. Further, at least one of the parameters ρ and κ is nonzero, where

$$\rho \triangleq \frac{1}{\sqrt{2\pi}\sigma^5\vartheta} \int_{\mathcal{R}} (x^2 - \sigma^2\vartheta^2) f(x) e^{-x^2/2\sigma^2\vartheta^2}\, dx, \tag{16}$$

$$\kappa \triangleq \frac{1}{\sqrt{2\pi}\sigma^7\vartheta} \int_{\mathcal{R}} (x^3 - 3\sigma^2\vartheta^2 x) f(x) e^{-x^2/2\sigma^2\vartheta^2}\, dx, \tag{17}$$

where $\sigma^2 = \sum_{i=0}^{\infty} l_i^2$.

Remark 1. The growth rate restriction in H5 implies that there are constants $\alpha > 0$ and $\beta \ge 1$ such that

$$|f(x)| \le \alpha(1 + |x|^{\beta}) \quad \forall x \in \mathcal{R}. \tag{18}$$

Therefore, the integrals (16) and (17) are finite.

Let us explain conditions imposed here. Conditions H1 and H4 and also H6 and H7 to be introduced later concern the signals in the system, while Conditions H2, H3, and H5 are on the structure of the system. The purpose of applying the Gaussian input in H1 is to derive the explicit relationships (25)–(28) connecting the impulse responses of the two linear subsystems and the correlation functions between the observed input and output. These relationships are the basis of the proposed algorithms for estimating the impulse responses by using the observed input and output. It is clear that H2 and H3 are the standard condition on the linear subsystems. Condition H4 concerns the measurement errors, which are not negligible in consideration of the present paper. Here we allow them to be correlated. In Condition H5, the function $f(\cdot)$ is allowed to be discontinuous: it is discontinuous at x if $f(x^-) \neq f(x^+)$. Further, the assumption that at least one of the constants ρ and κ is nonzero holds for many practical nonlinearities including polynomials, preloads, saturations, deadlines, quantized functions, and many others as discussed in [18]. It is seen that the two constants characterize some kind of correlation between the input signal v_k of the nonlinearity and its output $f(v_k)$. An intuitive explanation for this assumption is that the persistently exciting property of the signal $\{v_k\}$ remains in a certain sense after passing through the nonlinearity $f(\cdot)$. At last, additional Conditions H6 and H7 are employed to obtain the nonparametric estimation of $f(\cdot)$ by the deconvolution kernel functions.

Assuming $u_k^0 = 0,\ \forall k < 0$, we have

$$v_k = \sum_{i=0}^{k-1} l_i u_{k-i-1}^0, \tag{19}$$

and hence $v_k \in \mathcal{N}(0, \sigma_k^2)$ where $\sigma_k^2 \triangleq \sum_{i=0}^{k-1} l_i^2 \underset{k\to\infty}{\longrightarrow} \sigma^2$.

From (2) and (3) it follows that

$$\begin{aligned} y_{k+1}^0 &= C^{-1}(z)D(z)f(v_k) + C^{-1}(z)D(z)\eta_k + C^{-1}(z)\xi_{k+1} \\ &= \sum_{j=0}^{k} h_j f(v_{k-j}) + \sum_{j=0}^{k} h_j \eta_{k-j} + C^{-1}(z)\xi_{k+1}. \end{aligned} \tag{20}$$

2.2. Estimation of $\{p_1, \ldots, p_{n_p}, q_1, \ldots, q_{n_q}\}$ and $\{c_1, \ldots, c_{n_c}, d_1, \ldots, d_{n_d}\}$

We first estimate the impulse responses of the two linear subsystems, and then recover their coefficients by the convolution relationship between the impulse responses and the coefficients of the linear subsystems.

According to Lemma 3.1 in [18], under H1–H5 we have the following limits:

$$E[y_{k+1}((u_{k-i-1}^0)^2 - \vartheta^2)] \underset{k\to\infty}{\longrightarrow} \rho \sum_{j=0}^{i} h_j l_{i-j}^2 \quad \forall i \geq 0, \tag{21}$$

$$E[y_{k+1} u_{k-1}^0 u_{k-i-1}^0] \underset{k\to\infty}{\longrightarrow} \rho l_i \quad \forall i \geq 1, \tag{22}$$

$$E[y_{k+1}((u_{k-i-1}^0)^3 - 3\vartheta^2 u_{k-i-1}^0)] \underset{k\to\infty}{\longrightarrow} \kappa \sum_{j=0}^{i} h_j l_{i-j}^3 \quad \forall i \geq 0, \tag{23}$$

$$E[y_{k+1}((u_{k-1}^0)^2 - \vartheta^2) u_{k-i-1}^0] \underset{k\to\infty}{\longrightarrow} \kappa l_i \quad \forall i \geq 1, \tag{24}$$

which are the basis of identifying the two linear subsystems. However, in the EIV case, the true input $\{u_k^0\}$ and its variance ϑ^2 are unknown, so (21)–(24) cannot be directly applied. We have to establish similar limits but without involving $\{u_k^0\}$ and ϑ^2.

Lemma 1. *Assume that H1–H5 hold. Then the following limits take place:*

$$E[(y_{k+1} - Ey_{k+1}) u_{k-i-1}^2] \underset{k\to\infty}{\longrightarrow} \rho \sum_{j=0}^{i} h_j l_{i-j}^2 \quad \forall i \geq 0, \tag{25}$$

$$E[(y_{k+1} - Ey_{k+1}) u_{k-1} u_{k-i-1}] \underset{k\to\infty}{\longrightarrow} \rho l_i \quad \forall i \geq 1, \tag{26}$$

$$E[(y_{k+1} - Ey_{k+1})(u_{k-i-1}^2 - 3Eu_{k-i-1}^2) u_{k-i-1}] \underset{k\to\infty}{\longrightarrow} \kappa \sum_{j=0}^{i} h_j l_{i-j}^3 \quad \forall i \geq 0, \tag{27}$$

$$\begin{aligned} &E[(y_{k+1} - Ey_{k+1})(u_{k-1}^2 - Eu_{k-1}^2) u_{k-i-1}] \\ &\quad - 2E[y_{k+1} u_{k-1}] E[u_{k-1} u_{k-i-1}] \underset{k\to\infty}{\longrightarrow} \kappa l_i \quad \forall i \geq 1. \end{aligned} \tag{28}$$

Proof. By (21) and the mutual independence of $\{u_k^0\}$, $\{\varepsilon_k^{(1)}\}$, and $\{\varepsilon_k^{(2)}\}$, we see that

$$\begin{aligned} &E((y_{k+1} - Ey_{k+1}) u_{k-i-1}^2) \\ &\quad = E(y_{k+1} - Ey_{k+1})((u_{k-i-1}^0)^2 + (\varepsilon_{k-i-1}^{(1)})^2 + 2u_{k-i-1}^0 \varepsilon_{k-i-1}^{(1)}) \\ &\quad = E((y_{k+1} - Ey_{k+1})(u_{k-i-1}^0)^2) = E(y_{k+1}((u_{k-i-1}^0)^2 - \vartheta^2)) \\ &\quad \underset{k\to\infty}{\longrightarrow} \rho \sum_{j=0}^{i} h_j l_{i-j}^2 \quad \forall i \geq 0. \end{aligned} \tag{29}$$

Similarly, by (22) and the independence of signals involved, we see that

$$\begin{aligned} &E((y_{k+1} - Ey_{k+1}) u_{k-1} u_{k-i-1}) \\ &\quad = E((y_{k+1} - Ey_{k+1})(u_{k-1}^0 u_{k-i-1}^0 + u_{k-1}^0 \varepsilon_{k-i-1}^{(1)} + \varepsilon_{k-1}^{(1)} u_{k-i-1}^0 \\ &\qquad + \varepsilon_{k-1}^{(1)} \varepsilon_{k-i-1}^{(1)})) = E((y_{k+1} - Ey_{k+1}) u_{k-1}^0 u_{k-i-1}^0) \\ &\quad = E(y_{k+1} u_{k-1}^0 u_{k-i}^0) \underset{k\to\infty}{\longrightarrow} \rho h_i \quad i \geq 1. \end{aligned} \tag{30}$$

By (23), it follows that

$$\begin{aligned} &E[(y_{k+1} - Ey_{k+1}) u_{k-i-1}(u_{k-i-1}^2 - 3Eu_{k-i-1}^2)] \\ &\quad = E[(y_{k+1} - Ey_{k+1})((u_{k-i-1}^0)^3 + u_{k-i-1}^0 (\varepsilon_{k-i-1}^{(1)})^2 \\ &\qquad - 3\vartheta^2 u_{k-i-1}^0 - 3u_{k-i-1}^0 E(\varepsilon_{k-i-1}^{(1)})^2 + 2u_{k-i-1}^0 (\varepsilon_{k-i-1}^{(1)})^2)] \\ &\quad = E[(y_{k+1} - Ey_{k+1})((u_{k-i-1}^0)^3 - 3\vartheta^2 u_{k-i-1}^0)] \\ &\quad = E[y_{k+1}((u_{k-i-1}^0)^3 - 3\vartheta^2 u_{k-i-1}^0)] \underset{k\to\infty}{\longrightarrow} \kappa \sum_{j=0}^{i} h_j l_{i-j}^3 \quad i \geq 0. \end{aligned} \tag{31}$$

By (24), we then have

$$\begin{aligned} &E[(y_{k+1} - Ey_{k+1}) u_{k-i-1}(u_{k-1}^2 - Eu_{k-1}^2)] \\ &\quad = E[(y_{k+1} - Ey_{k+1})(u_{k-i-1}^0 (u_{k-1}^0)^2 + u_{k-i-1}^0 (\varepsilon_{k-1}^{(1)})^2 - \vartheta^2 u_{k-i-1}^0 \\ &\qquad - u_{k-i-1}^0 E(\varepsilon_{k-1}^{(1)})^2 + 2u_{k-1}^0 \varepsilon_{k-1}^{(1)} \varepsilon_{k-i-1}^{(1)})] \\ &\quad = E[(y_{k+1} - Ey_{k+1})(u_{k-i-1}^0 (u_{k-1}^0)^2 - \vartheta^2 u_{k-i-1}^0 \\ &\qquad + 2u_{k-1}^0 \varepsilon_{k-1}^{(1)} \varepsilon_{k-i-1}^{(1)})] \\ &\quad = E[y_{k+1}(u_{k-i-1}^0 (u_{k-1}^0)^2 - \vartheta^2 u_{k-i-1}^0)] + 2E[y_{k+1} u_{k-1}^0 \varepsilon_{k-1}^{(1)} \varepsilon_{k-i-1}^{(1)}], \end{aligned} \tag{32}$$

which, by noticing that $E[u_{k-1} u_{k-i-1}] = E[\varepsilon_{k-1}^{(1)} \varepsilon_{k-i-1}^{(1)}] \forall i \geq 1$, implies that

$$\begin{aligned} &E[(y_{k+1} - Ey_{k+1}) u_{k-i-1}(u_{k-1}^2 - Eu_{k-1}^2)] \\ &\quad - 2E[y_{k+1} u_{k-1}] E[u_{k-1} u_{k-i-1}] \underset{k\to\infty}{\longrightarrow} \kappa l_i, \quad i \geq 1. \end{aligned} \tag{33}$$

The proof of the lemma is completed. □

In order to estimate the impulse responses $\{l_i, i \geq 1\}$ and $\{h_i, i \geq 1\}$, it is necessary to obtain the estimates for Ey_{k+1}, $E(y_{k+1} u_{k-1})$, and $E(u_k u_{k-i}), i \geq 0$ in (25)–(28). For this, we use the stochastic approximation algorithms with expanding truncations (SAAWETs) [7] to recursively estimate Ey_{k+1}, $E(y_{k+1} u_{k-1})$, and $E(u_k u_{k-i}), i \geq 0$, respectively:

$$\lambda_{k+1}^{(y)} = [\lambda_k^{(y)} - (1/k)(\lambda_k^{(y)} - y_{k+1})] \cdot I_{[|\lambda_k^{(y)} - (1/k)(\lambda_k^{(y)} - y_{k+1})| \leq M_{\delta_k^{(y)}}]}, \tag{34}$$

$$\delta_k^{(y)} = \sum_{j=1}^{k-1} I_{[|\lambda_j^{(y)} - (1/j)(\lambda_j^{(y)} - y_{j+1})| > M_{\delta_j^{(y)}}]}, \tag{35}$$

$$\lambda_{k+1}^{(\tau)} = [\lambda_k^{(\tau)} - (1/k)(\lambda_k^{(\tau)} - y_{k+1} u_{k-1})] \cdot I_{[|\lambda_k^{(\tau)} - (1/k)(\lambda_k^{(\tau)} - y_{k+1} u_{k-1})| \leq M_{\delta_k^{(\tau)}}]}, \tag{36}$$

$$\delta_k^{(\tau)} = \sum_{j=1}^{k-1} I_{[|\lambda_j^{(\tau)} - (1/j)(\lambda_j^{(\tau)} - y_{j+1} u_{j-1})| > M_{\delta_j^{(\tau)}}]}, \tag{37}$$

$$\lambda_{k+1}^{(i,u)} = [\lambda_k^{(i,u)} - (1/k)(\lambda_k^{(i,u)} - u_{k+1} u_{k-i+1})] \cdot I_{[|\lambda_k^{(i,u)} - (1/k)(\lambda_k^{(i,u)} - u_{k+1} u_{k-i+1})| \leq M_{\delta_k^{(i,u)}}]}, \tag{38}$$

$$\delta_k^{(i,u)} = \sum_{j=1}^{k-1} I_{[|\lambda_j^{(i,u)} - (1/j)(\lambda_j^{(i,u)} - u_{j+1}u_{j-i+1})| > M_{\delta_j^{(i,u)}}]}, \quad i \geq 0, \tag{39}$$

where $\{M_k\}$ is an arbitrarily chosen sequence of positive real numbers increasingly diverging to infinity, $\lambda_0^{(y)}$, $\lambda_0^{(\tau)}$, and $\lambda_0^{(i,u)}$ are the arbitrary initial values, respectively, and $I_{[A]}$ denotes the indicator function of a set A.

Noticing that the right-hand sides of (25) and (27) are equal to ρ and κ, respectively, when $i=0$, so we can derive the estimates for $\{l_i, i \geq 1\}$ and $\{h_j, j \geq 1\}$ by (25) and (26) if the constant $\rho \neq 0$, or by (27) and (28) if $\kappa \neq 0$. However, we only know that at least one of ρ and κ is nonzero by H5, so we design a switching mechanism between the two ways. This is realized by comparing the absolute values of the estimates for ρ and κ at each step.

We first give the estimates for ρ and κ based on (25) and (27), respectively, as follows:

$$\theta_{k+1}^{(0,\rho)} = [\theta_k^{(0,\rho)} - (1/k)(\theta_k^{(0,\rho)} - (y_{k+1} - \lambda_{k+1}^{(y)})u_{k-1}^2)] \cdot I_{[|\theta_k^{(0,\rho)} - (1/k)(\theta_k^{(0,\rho)} - (y_{k+1} - \lambda_{k+1}^{(y)})u_{k-1}^2)| \leq M_{\delta_k^{(0,\rho)}}]}, \tag{40}$$

$$\delta_k^{(0,\rho)} = \sum_{j=1}^{k-1} I_{[|\theta_j^{(0,\rho)} - (1/j)(\theta_j^{(0,\rho)} - (y_{j+1} - \lambda_{j+1}^{(y)})u_{j-1}^2)| > M_{\delta_j^{(0,\rho)}}]}, \tag{41}$$

$$\theta_{k+1}^{(0,\kappa)} = [\theta_k^{(0,\kappa)} - (1/k)(\theta_k^{(0,\kappa)} - (y_{k+1} - \lambda_{k+1}^{(y)})(u_{k-1}^2 - 3\lambda_{k-1}^{(0,u)})u_{k-1})] \cdot I_{[|\theta_k^{(0,\kappa)} - (1/k)(\theta_k^{(0,\kappa)} - (y_{k+1} - \lambda_{k+1}^{(y)})(u_{k-1}^2 - 3\lambda_{k-1}^{(0,u)})u_{k-1})| \leq M_{\delta_k^{(0,\kappa)}}]}, \tag{42}$$

$$\delta_k^{(0,\kappa)} = \sum_{j=1}^{k-1} I_{[|\theta_j^{(0,\kappa)} - (1/j)(\theta_j^{(0,\kappa)} - (y_{j+1} - \lambda_{j+1}^{(y)})(u_{j-1}^2 - 3\lambda_{j-1}^{(0,u)})u_{j-1})| > M_{\delta_j^{(0,\kappa)}}]}. \tag{43}$$

If $|\theta_{k+1}^{(0,\rho)}| \geq |\theta_{k+1}^{(0,\kappa)}|$, then the following algorithm motivated by (26) is used to estimate ρl_i, $i \geq 1$:

$$\theta_{k+1}^{(i,\rho)} = [\theta_k^{(i,\rho)} - (1/k)(\theta_k^{(i,\rho)} - (y_{k+1} - \lambda_{k+1}^{(y)})u_{k-1}u_{k-i-1})] \cdot I_{[|\theta_k^{(i,\rho)} - (1/k)(\theta_k^{(i,\rho)} - (y_{k+1} - \lambda_{k+1}^{(y)})u_{k-1}u_{k-i-1})| \leq M_{\delta_k^{(i,\rho)}}]}, \tag{44}$$

$$\delta_k^{(i,\rho)} = \sum_{j=1}^{k-1} I_{[|\theta_j^{(i,\rho)} - (1/j)(\theta_j^{(i,\rho)} - (y_{j+1} - \lambda_{j+1}^{(y)})u_{j-1}u_{j-i-1})| > M_{\delta_j^{(i,\rho)}}]}, \quad i \geq 1, \tag{45}$$

here $\theta_k^{(i,\rho)}$ is obtained from the previous step of the recursion if $|\theta_k^{(0,\rho)}| \geq |\theta_k^{(0,\kappa)}|$. Otherwise, i.e., if $|\theta_k^{(0,\rho)}| < |\theta_k^{(0,\kappa)}|$, then $\theta_k^{(i,\rho)}$ has not been computed in accordance with (44) and (45). In this case $\theta_k^{(i,\rho)}$ in (44) is set to equal $\theta_k^{(0,\rho)} l_{i,k}$. After having the estimates for ρ and ρl_i, the estimates for the impulse responses $\{l_i, i \geq 1\}$ at time $k+1$ are given by

$$l_{i,k+1} \triangleq \begin{cases} \dfrac{\theta_{k+1}^{(i,\rho)}}{\theta_{k+1}^{(0,\rho)}} & \text{if } \theta_{k+1}^{(0,\rho)} \neq 0, \\ 0 & \text{if } \theta_{k+1}^{(0,\rho)} = 0. \end{cases} \tag{46}$$

Based on (25) the following algorithms are employed to estimate $\rho \sum_{j=0}^{i} h_j l_{i-j}^2$:

$$\lambda_{k+1}^{(i,\rho)} = [\lambda_k^{(i,\rho)} - (1/k)(\lambda_k^{(i,\rho)} - (y_{k+1} - \lambda_{k+1}^{(y)})u_{k-i-1}^2)] \cdot I_{[|\lambda_k^{(i,\rho)} - (1/k)(\lambda_k^{(i,\rho)} - (y_{k+1} - \lambda_{k+1}^{(y)})u_{k-i-1}^2)| \leq M_{\gamma_k^{(i,\rho)}}]}, \tag{47}$$

$$\gamma_k^{(i,\rho)} = \sum_{j=1}^{k-1} I_{[|\lambda_j^{(i,\rho)} - (1/j)(\lambda_j^{(i,\rho)} - (y_{j+1} - \lambda_{j+1}^{(y)})u_{j-i-1}^2)| > M_{\gamma_j^{(i,\rho)}}]}, \quad i \geq 1, \tag{48}$$

where $\lambda_k^{(i,\rho)}$ is obtained from the previous step of the recursion if $|\theta_k^{(0,\rho)}| \geq |\theta_k^{(0,\kappa)}|$. Otherwise, they are set to equal $\theta_k^{(0,\rho)} \sum_{j=0}^{i} h_{j,k} l_{i-j,k}^2$. In this case the estimates for $\{h_i, i \geq 1\}$ are given by

$$h_{i,k} \triangleq \begin{cases} \dfrac{\lambda_{k+1}^{(i,\rho)}}{\theta_{k+1}^{(0,\rho)}} - \sum_{j=0}^{i-1} h_{j,k+1} l_{i-j,k+1}^2 & \text{if } \theta_k^{(0,\rho)} \neq 0, \\ 0 & \text{if } \theta_k^{(0,\rho)} = 0. \end{cases} \tag{49}$$

Conversely, if $|\theta_{k+1}^{(0,\rho)}| < |\theta_{k+1}^{(0,\kappa)}|$, then based on (28) the following algorithms are employed to estimate $E[(y_{k+1} - Ey_{k+1})(u_{k-1}^2 - Eu_{k-1}^2)u_{k-i-1}]$, $i \geq 1$:

$$\theta_{k+1}^{(i,\kappa)} = [\theta_k^{(i,\kappa)} - (1/k)(\theta_k^{(i,\kappa)} - (y_{k+1} - \lambda_{k+1}^{(y)})(u_{k-1}^2 - \lambda_{k-1}^{(0,u)})u_{k-i-1})] \cdot I_{[|\theta_k^{(i,\kappa)} - (1/k)(\theta_k^{(i,\kappa)} - (y_{k+1} - \lambda_{k+1}^{(y)})(u_{k-1}^2 - \lambda_{k-1}^{(0,u)})u_{k-i-1})| \leq M_{\delta_k^{(i,\kappa)}}]}, \tag{50}$$

$$\delta_k^{(i,\kappa)} = \sum_{j=1}^{k-1} I_{[|\theta_j^{(i,\kappa)} - (1/j)(\theta_j^{(i,\kappa)} - (y_{j+1} - \lambda_{j+1}^{(y)})(u_{j-1}^2 - \lambda_{j-1}^{(0,u)})u_{j-i-1})| > M_{\delta_j^{(i,\kappa)}}]}, \quad i \geq 1. \tag{51}$$

Similar to the previous case, $\theta_k^{(i,\kappa)}$ is derived from the previous step of the recursion if $|\theta_k^{(0,\rho)}| < |\theta_k^{(0,\kappa)}|$. Otherwise, i.e., if $|\theta_k^{(0,\rho)}| \geq |\theta_k^{(0,\kappa)}|$, then $\theta_k^{(i,\kappa)}$ has not been computed in accordance with (50) and (51). In this case, $\theta_k^{(i,\kappa)}$ in (50) is set to equal $\theta_k^{(0,\kappa)} l_{i,k} + 2\lambda_k^{(\tau)} \lambda_k^{(i,u)}$. As results, the estimates for $\{l_i, i \geq 1\}$ at time $k+1$ are given by

$$l_{i,k+1} \triangleq \begin{cases} \dfrac{\theta_{k+1}^{(i,\kappa)} - 2\lambda_{k+1}^{(\tau)} \lambda_{k+1}^{(i,u)}}{\theta_{k+1}^{(0,\kappa)}} & \text{if } \theta_{k+1}^{(0,\kappa)} \neq 0, \\ 0 & \text{if } \theta_{k+1}^{(0,\kappa)} = 0. \end{cases} \tag{52}$$

Based on (27) we introduce the following algorithms to estimate $\kappa \sum_{j=0}^{i} h_j l_{i-j}^3$:

$$\lambda_{k+1}^{(i,\kappa)} = [\lambda_k^{(i,\kappa)} - (1/k)(\lambda_k^{(i,\kappa)} - (y_{k+1} - \lambda_{k+1}^{(y)})(u_{k-i-1}^2 - 3\lambda_{k-i-1}^{(i,u)})u_{k-i-1})] \cdot I_{[|\lambda_k^{(i,\kappa)} - (1/k)(\lambda_k^{(i,\kappa)} - (y_{k+1} - \lambda_{k+1}^{(y)})(u_{k-i-1}^2 - 3\lambda_{k-i-1}^{(i,u)})u_{k-i-1})| \leq M_{\gamma_k^{(i,\kappa)}}]}, \tag{53}$$

$$\gamma_k^{(i,\kappa)} = \sum_{j=1}^{k-1} I_{[|\lambda_j^{(i,\kappa)} - (1/j)(\lambda_j^{(i,\kappa)} - (y_{j+1} - \lambda_{j+1}^{(y)})(u_{j-i-1}^2 - 3\lambda_{j-i-1}^{(i,u)})u_{j-i-1})| > M_{\gamma_j^{(i,\kappa)}}]}, \quad i \geq 1, \tag{54}$$

where $\lambda_k^{(i,\kappa)}$ is obtained from the previous step of the recursion if $|\theta_k^{(0,\kappa)}| \geq |\theta_k^{(0,\rho)}|$. Otherwise, they are set to equal $\theta_k^{(0,\kappa)} \sum_{j=0}^{i} h_{j,k} l_{i-j,k}^3$. In this case the estimates for $\{h_i, i \geq 1\}$ are given by

$$h_{i,k} \triangleq \begin{cases} \dfrac{\lambda_{k+1}^{(i,\kappa)}}{\theta_{k+1}^{(0,\kappa)}} - \sum_{j=0}^{i-1} h_{j,k+1} l_{i-j,k+1}^3 & \text{if } \theta_k^{(0,\kappa)} \neq 0, \\ 0 & \text{if } \theta_k^{(0,\kappa)} = 0. \end{cases} \tag{55}$$

It is important to note that after establishing strong consistency of $\theta_k^{(0,\rho)}$ and $\theta_k^{(0,\kappa)}$ in Section 3, switching between the algorithms (44)–(49) and (50)–(55) ceases in a finite number of steps, because by H5 at least one of ρ and κ is nonzero and hence either $|\theta_k^{(0,\rho)}| \geq |\theta_k^{(0,\kappa)}|$ or $|\theta_k^{(0,\rho)}| < |\theta_k^{(0,\kappa)}|$ takes place for all sufficiently large k.

Remark 2. If the constant

$$\tau \triangleq \frac{1}{\sqrt{2\pi}\sigma^3 \vartheta} \int_{\mathcal{R}} x f(x) e^{-x^2/2\sigma^2\vartheta^2} \, dx$$

is nonzero, then the impulse responses of the linear subsystem at the output end can be more efficiently estimated based on the limit

$$E[y_{k+1} u_{k-i-1}] \underset{k \to \infty}{\longrightarrow} \tau \sum_{j=0}^{i} h_j l_{i-j}, \quad i \geq 0.$$

Similar to [20,21], once the estimates $\{l_{i,k}, i \ge 0\}$ for the impulse response $\{l_i, i \ge 0\}$ are obtained, then the estimates for the parameters $\{p_1, \ldots, p_{n_p}, q_1, \ldots, q_{n_q}\}$ of the linear subsystem at the input end can be derived by using the convolution relationship between the impulse responses $\{l_i, i \ge 0\}$ and the parameters $\{p_1, \ldots, p_{n_p}, q_1, \ldots, q_{n_q}\}$.

Define the matrix:

$$L \triangleq \begin{pmatrix} l_{n_q} & l_{n_q-1} & \cdots & l_{n_q-n_p+1} \\ l_{n_q+1} & l_{n_q} & \cdots & l_{n_q-n_p+2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{n_q+n_p-1} & l_{n_q+n_p-2} & \cdots & l_{n_q} \end{pmatrix}.$$

Then the matrix L is nonsingular under H2 (see [20,21]), and hence the matrix L_k obtained from L with $\{l_i, i \ge 0\}$ replaced by their estimates $\{l_{i,k}, i \ge 0\}$ and with $l_{i,k} = 0$ for $i < 0$ is also nonsingular for sufficiently large k, since $l_{i,k} \underset{k \to \infty}{\longrightarrow} l_i$ a.s. as to be shown by Theorem 1.

The estimates for $\{p_1, \ldots, p_{n_p}, q_1, \ldots, q_{n_q}\}$ are naturally defined as follows:

$$[p_{1,k}, p_{2,k}, \ldots, p_{n_p,k}]^T \triangleq -L_k^{-1}[l_{n_q+1,k}, l_{n_q+2,k}, \ldots, l_{n_q+n_p,k}]^T, \tag{56}$$

$$q_{i,k} \triangleq l_{i,k} + \sum_{j=1}^{i \wedge n_p} p_{j,k} l_{i-j,k}, \quad i = 1, 2, \ldots, n_q. \tag{57}$$

Similarly, the estimates for $\{c_1, \ldots, c_{n_c}, d_1, \ldots, d_{n_d}\}$ are obtained in the same way as that used for estimating coefficients of the linear subsystem at the input end.

For this define

$$H \triangleq \begin{pmatrix} h_{n_d} & h_{n_d-1} & \cdots & h_{n_d-n_c+1} \\ h_{n_d+1} & h_{n_d} & \cdots & h_{n_d-n_c+2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_d+n_c-1} & h_{n_d+n_c-2} & \cdots & h_{n_d} \end{pmatrix}$$

with $h_i = 0$ for $i < 0$.

The matrix H is nonsingular under H3 (see [20,21]), and hence the corresponding estimate H_k for H with $\{h_i, i \ge 0\}$ replaced by $\{h_{i,k}, i \ge 0\}$ is also nonsingular for sufficiently large k since $h_{i,k} \underset{k \to \infty}{\longrightarrow} h_i$ a.s. as to be shown by Theorem 1.

Similar to (56) and (57), the estimates for $\{c_1, \ldots, c_{n_c}, d_1, \ldots, d_{n_d}\}$ are given as follows:

$$[c_{1,k}, c_{2,k}, \ldots, c_{n_c,k}]^T \triangleq -H_k^{-1}[h_{n_d+1,k}, h_{n_d+2,k}, \ldots, h_{n_d+n_c,k}]^T, \tag{58}$$

$$d_{i,k} \triangleq h_{i,k} + \sum_{j=1}^{i \wedge n_c} c_{j,k} h_{i-j,k}, \quad i = 1, 2, \ldots, n_d. \tag{59}$$

2.3. Nonparametric estimation of $f(\cdot)$

For estimating $f(x)$ with x being an arbitrary point at the real axis, the useful sequences $\{v_k\}$ and $\{f(v_k)\}$ are not directly observed. Instead, we estimate their noisy values $\{\psi_k\}$ (see (62)) and $\{\zeta_k\}$ (see (68)), and then apply SAAWET incorporated with the deconvolution kernel functions [13,27,11] to remove the noise influence.

Let us start with estimating $\{\psi_k\}$.

Define

$$\psi_k \triangleq P^{-1}(z)Q(z)u_{k-1}, \tag{60}$$

$$e_k \triangleq P^{-1}(z)Q(z)\varepsilon_{k-1}^{(1)} = (P(z)F_1(z))^{-1}(Q(z)G_1(z))\varsigma_{k-1}^{(1)}. \tag{61}$$

According to (1), (8), and (11), we have

$$\psi_k = P^{-1}(z)Q(z)u_{k-1}^0 + [P(z)F_1(z)]^{-1}Q(z)G_1(z)\varsigma_{k-1}^{(1)} = v_k + e_k. \tag{62}$$

Define

$$P \triangleq \begin{pmatrix} -p_1 & 1 & & \\ \vdots & & \ddots & \\ \vdots & & & 1 \\ -p_s & 0 & \cdots & 0 \end{pmatrix}, \quad Q \triangleq \begin{pmatrix} 1 \\ q_1 \\ \vdots \\ q_{s-1} \end{pmatrix}, \quad G \triangleq \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

and $s \triangleq \max(n_p, n_q+1)$, where $p_i \triangleq 0$ for $i > n_p$ and $q_j \triangleq 0$ for $j > n_q$. Then, Eq. (60) connecting ψ_k and u_k can be expressed in the state space form:

$$x_{k+1} = Px_k + Qu_k, \quad \psi_{k+1} = G^T x_{k+1}. \tag{63}$$

Replacing p_i, $i = 1, \ldots, s$ and q_j, $j = 1, \ldots, s$ in P and Q with $p_{i,k}$ and $q_{j,k}$ given by (56) and (57), respectively, we obtain the estimates P_k and Q_k for P and Q at time k, and hence the estimate $\widehat{\psi}_k$ for ψ_k is given as follows:

$$\widehat{x}_{k+1} = P_{k+1}\widehat{x}_k + Q_{k+1}u_k, \quad \widehat{\psi}_{k+1} = G^T \widehat{x}_{k+1} \tag{64}$$

with an arbitrary initial value $\widehat{x}_0$.

Similarly, from (3) and (8), we have

$$C(z)y_{k+1} = D(z)(f(v_k) + \eta_k) + C(z)\varepsilon_{k+1}^{(2)} + \xi_{k+1}. \tag{65}$$

By defining

$$\zeta_k \triangleq D^{-1}(z)C(z)y_{k+1}, \quad \phi_k \triangleq D^{-1}(z)\xi_{k+1}, \quad \text{and } \chi_k \triangleq D^{-1}(z)C(z)\varepsilon_{k+1}^{(2)}, \tag{66}$$

then $f(v_k)$ in (65) can be expressed as

$$f(v_k) = D^{-1}(z)C(z)y_{k+1} - D^{-1}(z)C(z)\varepsilon_{k+1}^{(2)} - D^{-1}(z)\xi_{k+1} - \eta_k, \tag{67}$$

or

$$\zeta_k = f(v_k) + \chi_k + \phi_k + \eta_k. \tag{68}$$

The first equation in (66) can be expressed in the state space form:

$$t_{k+1} = Dt_k + Cy_{k+1}, \quad \zeta_k = \overline{G}^T t_{k+1}, \tag{69}$$

where

$$D \triangleq \begin{pmatrix} -d_1 & 1 & & \\ \vdots & & \ddots & \\ \vdots & & & 1 \\ -d_{\overline{s}} & 0 & \cdots & 0 \end{pmatrix}, \quad C \triangleq \begin{pmatrix} 1 \\ c_1 \\ \vdots \\ c_{\overline{s}-1} \end{pmatrix}, \quad \overline{G} \triangleq \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

and $\overline{s} \triangleq \max(n_d, n_c+1)$.

Let C_k and D_k be the estimates obtained from C and D with entries replaced by their estimates given by (58) and (59). Then the estimate $\widehat{\zeta}_k$ for ζ_k is recursively given by the following algorithm with an arbitrary initial value $\widehat{t}_0$:

$$\widehat{t}_{k+1} = D_{k+1}\widehat{t}_k + C_{k+1}y_{k+1}, \quad \widehat{\zeta}_k = \overline{G}^T \widehat{t}_{k+1}. \tag{70}$$

In order to eliminate the influence of e_k involved in ψ_k, we apply the deconvolution kernel functions, but for this the additional assumptions are needed.

H6: The variance ϑ^2 of the noise-free input u_k^0 is known.

H7: The driven noise $\{\varsigma_k^{(1)}\}$ in (11) is a sequence of i.i.d. zero mean Gaussian random variables.

Let us introduce the Sinc kernel function [27,11]:

$$K(x) = \frac{\sin(x)}{\pi x}. \tag{71}$$

Then we have its Fourier transformation

$$\Phi_K(t) \triangleq \int_{\mathcal{R}} e^{\imath tx} K(x)\, dx = I_{[|t| \le 1]}, \tag{72}$$

where $\imath$ denotes the imaginary unit $\imath^2 = -1$.

Under H7, $\{e_k\}$ is still a sequence of zero mean Gaussian random variables, and its characteristic function is

$$\Phi_{e_k}(t) \triangleq \int_{\mathcal{R}} e^{\imath tx}\frac{1}{\sqrt{2\pi}\sigma_k(e)}e^{-x^2/2\sigma_k^2(e)}\,dx = e^{-\sigma_k^2(e)t^2/2},$$

where $\sigma_k^2(e) \triangleq Ee_k^2$ is the variance of e_k. Denote by $\sigma^2(e)$ the limit of $\sigma_k^2(e)$. It is clear that $|\sigma^2(e)-\sigma_k^2(e)| = O(e^{-r_e k})$ for some $r_e>0$.

We now introduce the deconvolution kernel $w_k(x)$, but for this we first define

$$K_k(x) \triangleq \frac{1}{2\pi}\int_{\mathcal{R}} e^{-\imath tx}\frac{\Phi_K(t)}{\Phi_{e_k}(t/b_k)}\,dt = \frac{1}{2\pi}\int_{-1}^{1} e^{-\imath tx}e^{\sigma_k^2(e)t^2/2b_k^2}\,dt, \tag{73}$$

where $b_k = (b\sigma_k^2(e)/\log\,k)^{1/2}$ is the bandwidth with $b>3$ being a constant chosen in advance.

Thus the function $w_k(x)$ is defined by

$$w_k(x) \triangleq \frac{1}{b_k}K_k\left(\frac{\psi_k - x}{b_k}\right) = \frac{1}{2\pi b_k}\int_{-1}^{1} e^{[-\imath t(\psi_k - x)/b_k]}e^{\sigma_k^2(e)t^2/2b_k^2}\,dt$$
$$= \frac{1}{2\pi b_k}\int_{-1}^{1}\cos[(\psi_k - x)t/b_k]e^{\sigma_k^2(e)t^2/2b_k^2}\,dt = \frac{1}{\pi}\int_0^{1/b_k}\cos[(\psi_k - x)t]e^{\sigma_k^2(e)t^2/2}\,dt. \tag{74}$$

Notice that $\sigma_k^2(e)$ in (73) and (74) is unknown. To obtain its estimate $\hat{\sigma}_k^2(e)$ we first estimate the spectral density of $\varepsilon_k^{(1)}$, and then derive the estimate for the spectral density of e_k by means of the estimates for the linear subsystem at the input end. Finally, the estimate $\hat{\sigma}_k^2(e)$ for $\sigma_k^2(e)$ can be obtained by the inverse Fourier transformation of the spectral density estimate for e_k.

For simplicity, we assume that the orders n_{f_1} and n_{g_1} in (12) and (13) are known in the procedure of estimating the spectral density of $\varepsilon_k^{(1)}$. When the orders n_{f_1} and n_{g_1} are unknown, their strongly consistent estimates can be derived by the method provided in [9].

The autocovariances $a_i(\varepsilon_k^{(1)}) \triangleq E(\varepsilon_k^{(1)}\varepsilon_{k-i}^{(1)})$, $i\ge 0$ of $\varepsilon_k^{(1)}$ is recursively estimated by SAAWET:

$$a_{0,k+1}(\varepsilon_k^{(1)}) = [a_{0,k}(\varepsilon_k^{(1)}) - (1/k)(a_{0,k}(\varepsilon_k^{(1)}) + \vartheta^2 - u_{k+1}^2)] \cdot I_{[|a_{0,k}(\varepsilon_k^{(1)}) - (1/k)(a_{0,k}(\varepsilon_k^{(1)}) + \vartheta^2 - u_{k+1}^2)| \le M_{\delta_k^{(0,\varepsilon_k^{(1)})}}]}, \tag{75}$$

$$\delta_k^{(0,\varepsilon_k^{(1)})} = \sum_{j=1}^{k-1} I_{[|a_{0,j}(\varepsilon_j^{(1)}) - (1/j)(a_{0,j}(\varepsilon_j^{(1)}) + \vartheta^2 - u_{j+1}^2)| > M_{\delta_j^{(0,\varepsilon_k^{(1)})}}]}, \tag{76}$$

$$a_{i,k+1}(\varepsilon_k^{(1)}) = [a_{i,k}(\varepsilon_k^{(1)}) - (1/k)(a_{i,k}(\varepsilon_k^{(1)}) - u_{k+1}u_{k-i+1})] \cdot I_{[|a_{i,k}(\varepsilon_k^{(1)}) - (1/k)(a_{i,k}(\varepsilon_k^{(1)}) - u_{k+1}u_{k-i+1})| \le M_{\delta_k^{(i,\varepsilon_k^{(1)})}}]}, \tag{77}$$

$$\delta_k^{(i,\varepsilon_k^{(1)})} = \sum_{j=1}^{k-1} I_{[|a_{i,j}(\varepsilon_k^{(1)}) - (1/j)(a_{i,j}(\varepsilon_k^{(1)}) - u_{j+1}u_{j-i+1})| > M_{\delta_j^{(i,\varepsilon_k^{(1)})}}]}, \quad i\ge 1. \tag{78}$$

It is noticed that the algorithms (77) and (78) coincide with (38) and (39) for $i\ge 1$, where $\lambda_{k+1}^{(i,u)}$ is rewritten as $a_{i,k+1}(\varepsilon_k^{(1)})$ just for consistency with notations used here.

Define the Hankel matrix

$$\Gamma_k(\varepsilon_k^{(1)}) \triangleq \begin{pmatrix} a_{n_{g_1},k}(\varepsilon_k^{(1)}) & a_{n_{g_1}-1,k}(\varepsilon_k^{(1)}) & \cdots & a_{n_{g_1}-n_{f_1}+1,k}(\varepsilon_k^{(1)}) \\ a_{n_{g_1}+1,k}(\varepsilon_k^{(1)}) & a_{n_{g_1},k}(\varepsilon_k^{(1)}) & \cdots & a_{n_{g_1}-n_{f_1}+2,k}(\varepsilon_k^{(1)}) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_{g_1}+n_{f_1}-1,k}(\varepsilon_k^{(1)}) & a_{n_{g_1}+n_{f_1}-2,k}(\varepsilon_k^{(1)}) & \cdots & a_{n_{g_1},k}(\varepsilon_k^{(1)}) \end{pmatrix} \tag{79}$$

where $a_{i,k}(\varepsilon_k^{(1)}) \triangleq a_{-i,k}(\varepsilon_k^{(1)})$ for $i<0$. Since $a_{i,k}(\varepsilon_k^{(1)}) \underset{k\to\infty}{\longrightarrow} a_i(\varepsilon_k^{(1)})$, $i\ge 0$ as to be shown in Lemma 4 and the limit of $\Gamma_k(\varepsilon_k^{(1)})$ is nonsingular under H4 [21], the matrix $\Gamma_k(\varepsilon_k^{(1)})$ is nonsingular for sufficiently large k. Therefore, at time k, the parameters $\{f_{1,1}, f_{1,2}, \ldots, f_{1,n_{f_1}}\}$ of $F_1(z)$ in (12) can be estimated by the Yule–Walker equation:

$$[f_{1,1,k}, f_{1,2,k}, \ldots, f_{1,n_{f_1},k}]^T = -\Gamma_k^{-1}(\varepsilon_k^{(1)})[a_{n_{g_1}+1,k}(\varepsilon_k^{(1)}), a_{n_{g_1}+2,k}(\varepsilon_k^{(1)}), \ldots, a_{n_{g_1}+n_{f_1},k}(\varepsilon_k^{(1)})]^T. \tag{80}$$

The spectral density $S_{\varepsilon_k^{(1)}}(z)$ of $\varepsilon_k^{(1)}$ equals

$$S_{\varepsilon_k^{(1)}}(z) \triangleq \sum_{l=-\infty}^{\infty} a_l(\varepsilon_k^{(1)})z^l = \frac{G_1(z)G_1(z^{-1})\sigma^2(\varsigma_k^{(1)})}{F_1(z)F_1(z^{-1})}, \tag{81}$$

or

$$F_1(z)F_1(z^{-1})\sum_{l=-\infty}^{\infty} a_l(\varepsilon_k^{(1)})z^l = G_1(z)G_1(z^{-1})\sigma^2(\varsigma_k^{(1)}), \tag{82}$$

where $\sigma^2(\varsigma_k^{(1)})$ denotes the variance of $\varsigma_k^{(1)}$.

By comparing the coefficients of the same order of z at both sides of (82), we derive

$$G_1(z)G_1(z^{-1})\sigma^2(\varsigma_k^{(1)}) = \sum_{l=-n_{g_1}}^{n_{g_1}}\left(\sum_{i=0}^{n_{f_1}}\sum_{j=0}^{n_{f_1}} a_{l+j-i}(\varepsilon_k^{(1)})f_{1,i}f_{1,j}\right)z^l, \tag{83}$$

where only a finite number of autocovariances $a_l(\varepsilon_k^{(1)})$, $-n_{g_1}-n_{f_1} \le l \le n_{f_1}+n_{g_1}$ are involved.

As a consequence, the estimate for $S_{\varepsilon_k^{(1)}}(z)$ is obtained as follows:

$$\hat{S}_{\varepsilon_k^{(1)}}(z) = \frac{\sum_{l=-n_{g_1}}^{n_{g_1}}(\sum_{i=0}^{n_{f_1}}\sum_{j=0}^{n_{f_1}} a_{l+j-i,k}(\varepsilon_k^{(1)})f_{1,i,k}f_{1,j,k})z^l}{(\sum_{i=0}^{n_{f_1}} f_{1,i,k}z^i)(\sum_{j=0}^{n_{f_1}} f_{1,j,k}z^{-j})}, \tag{84}$$

and by (61) the spectral density $S_{e_k}(z)$ of e_k is estimated by

$$\hat{S}_{e_k}(z) = \frac{(\sum_{i=0}^{n_q} q_{i,k}z^i)(\sum_{j=0}^{n_q} q_{j,k}z^{-j})}{(\sum_{i=0}^{n_p} p_{i,k}z^i)(\sum_{j=0}^{n_p} p_{j,k}z^{-j})}\hat{S}_{\varepsilon_k^{(1)}}(z). \tag{85}$$

Finally, the variance $\sigma_k^2(e)$ of e_k is approximated by the inverse Fourier transformation:

$$\hat{\sigma}_k^2(e) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\hat{S}_{e_k}(e^{\imath\omega})\,d\omega. \tag{86}$$

Therefore, the estimate for $w_k(x)$ at time k is given by

$$\hat{w}_k(x) \triangleq \frac{1}{\pi}\int_0^{1/\hat{b}_k}\cos[(\hat{\psi}_k - x)t]e^{\hat{\sigma}_k^2(e)t^2/2}\,dt, \tag{87}$$

where $\hat{b}_k = (b\hat{\sigma}_k^2(e)/\log\,k)^{1/2}$.

We now give the algorithms to estimate $f(x)$:

$$\mu_{k+1}(x) = [\mu_k(x) - (1/k)(\mu_k(x) - \hat{w}_k(x))] \cdot I_{[|\mu_k(x) - (1/k)(\mu_k(x) - \hat{w}_k(x))| \le M_{\delta_k^{(\mu)}(x)}]}, \tag{88}$$

$$\delta_k^{(\mu)}(x) = \sum_{j=1}^{k-1} I_{[|\mu_{j-1}(x) - (1/j)(\mu_{j-1}(x) - \hat{w}_j(x))| > M_{\delta_j^{(\mu)}(x)}]}, \tag{89}$$

$$\beta_{k+1}(x) = [\beta_k(x) - (1/k)(\beta_k(x) - \hat{w}_k(x)\hat{\zeta}_k)] \cdot I_{[|\beta_k(x) - (1/k)(\beta_k(x) - \hat{w}_k(x)\hat{\zeta}_k)| \le M_{\delta_k^{(\beta)}(x)}]}, \tag{90}$$

$$\delta_k^{(\beta)}(x) = \sum_{j=1}^{k-1} I_{[|\beta_{j-1}(x) - (1/j)(\beta_{j-1}(x) - \hat{w}_j(x)\hat{\zeta}_j)| > M_{\delta_j^{(\beta)}(x)}]}. \tag{91}$$

As a matter of fact, $\mu_k(x)$ defined by (88) and (89) and $\beta_k(x)$ defined by (90) and (91) are applied to estimate $p(x)$ and $p(x)\tilde{f}(x)$ (the definition of $\tilde{f}(x)$ is given in (112)), respectively, where $p(x) = (1/\sqrt{2\pi}\sigma\vartheta)e^{-x^2/2\sigma^2\vartheta^2}$ is the limit of the density function of v_k. The estimate for $f(x)$ is naturally defined as

$$f_{k+1}(x) \triangleq \begin{cases} \dfrac{\beta_{k+1}(x)}{\mu_{k+1}(x)} & \text{if } \mu_{k+1}(x) \neq 0 \\ 0 & \text{if } \mu_{k+1}(x) = 0. \end{cases} \tag{92}$$

3. Strong consistency of estimates for linear subsystems

The methods and the mathematical tools of proving the strong consistency of the estimates proposed in Section 2 are essentially the same as those given in [19]. We will not repeat similar derivatives, but give the new ones.

Lemma 2. *Assume that H1–H5 hold. Then,* $\lambda_k^{(y)}$, $\lambda_k^{(\tau)}$, *and* $\lambda_k^{(i,u)}$, $i \geq 0$ *defined by (34) and (35), (36) and (37), and (38) and (39), respectively, have the following convergence rate:*

$$\left|\lambda_k^{(y)} - Ey_k\right| = o\left(\frac{1}{k^{1/2-c}}\right) \quad \forall c > 0, \tag{93}$$

$$\left|\lambda_k^{(\tau)} - E(y_k u_{k-2})\right| = o\left(\frac{1}{k^{1/2-c}}\right) \quad \forall c > 0, \tag{94}$$

$$\left|\lambda_k^{(i,u)} - E(u_k u_{k-i})\right| = o\left(\frac{1}{k^{1/2-c}}\right) \quad \forall c > 0,\ i \geq 0. \tag{95}$$

The proof is similar to that for Lemma 6 in [19].

Theorem 1. *Assume that H1–H5 hold. Then,* $\{l_{i,k}, i \geq 1\}$ *and* $\{h_{j,k}, j \geq 1\}$ *generated by (40)–(55) converge to* $\{l_i, i \geq 1\}$ *and* $\{h_j, j \geq 1\}$, *respectively, with the rate of convergence*

$$|l_{i,k} - l_i| = o(k^{-\nu}) \quad a.s.\ \forall \nu \in (0, 1/2),\ i \geq 1, \tag{96}$$

$$|h_{j,k} - h_j| = o(k^{-\nu}) \quad a.s.\ \forall \nu \in (0, 1/2),\ j \geq 1. \tag{97}$$

As consequences, from (56)–(59) the following convergence rates also take place:

$$|p_{i,k} - p_i| = o(k^{-\nu}) \quad a.s.\ \forall \nu \in (0, 1/2),\ i = 1, \ldots, n_p, \tag{98}$$

$$|q_{j,k} - q_j| = o(k^{-\nu}) \quad a.s.\ \forall \nu \in (0, 1/2),\ j = 1, \ldots, n_q, \tag{99}$$

$$|c_{l,k} - c_l| = o(k^{-\nu}) \quad a.s.\ \forall \nu \in (0, 1/2),\ l = 1, \ldots, n_c, \tag{100}$$

$$|d_{m,k} - d_m| = o(k^{-\nu}) \quad a.s.\ \forall \nu \in (0, 1/2),\ m = 1, \ldots, n_d. \tag{101}$$

Proof. We first prove that $\theta_k^{(0,\rho)}$ and $\theta_k^{(0,\kappa)}$, respectively, defined by (40) and (41) and by (42) and (43) converge a.s. with the following convergence rate:

$$|\theta_k^{(0,\rho)} - \rho| = o(k^{-\nu}) \quad a.s.\ \forall \nu \in (0, 1/2), \tag{102}$$

$$|\theta_k^{(0,\kappa)} - \kappa| = o(k^{-\nu}) \quad a.s.\ \forall \nu \in (0, 1/2). \tag{103}$$

We rewrite (40) as

$$\theta_{k+1}^{(0,\rho)} = [\theta_k^{(0,\rho)} - (1/k)(\theta_k^{(0,\rho)} - \rho) - (1/k)\varepsilon_{k+1}^{(0,\rho)}] \cdot I_{[|\theta_k^{(0,\rho)} - (1/k)(\theta_k^{(0,\rho)} - \rho l_i) - (1/k)\varepsilon_{k+1}^{(0,\rho)}| \leq M_{\delta_k}^{(0,\rho)}]},$$

where

$$\begin{aligned} \varepsilon_{k+1}^{(0,\rho)} &= \rho - (y_{k+1} - \lambda_{k+1}^{(y)})u_{k-1}^2 = (\rho - E((y_{k+1} - Ey_{k+1})u_{k-1}^2)) \\ &\quad + (E((y_{k+1} - Ey_{k+1})u_{k-1}^2) - (y_{k+1} - Ey_{k+1})u_{k-1}^2)) \\ &\quad + (\lambda_{k+1}^{(y)} - Ey_{k+1})u_{k-1}^2 \end{aligned} \tag{104}$$

Since ρ is the single root of the linear function $-(y-\rho)$, by Theorem 3.1.1 in [7], for proving (102), it suffices to show

$$\sum_{k=1}^{\infty} \frac{1}{k^{1-\nu}} \varepsilon_{k+1}^{(0,\rho)} < \infty \quad a.s.\ \forall \nu \in (0, 1/2). \tag{105}$$

Noticing (29) and carrying out the treatment similar to that used in Lemma 4.3 of [18], we have $|\rho - E((y_{k+1} - Ey_{k+1})\ u_{k-1}^2)| = O(e^{-r_1 k})$ for some $r_1 > 0$ since $\sigma^2 - \sigma_k^2 = \sum_{i=k}^{\infty} l_i^2 = O(e^{-r_1 k})$ for some $r_1 > 0$. This assures that (105) holds with $\varepsilon_{k+1}^{(0,\rho)}$ replaced by the first term at the right-hands of (104).

From (20) and (29) it follows that

$$\begin{aligned} &\sum_{k=1}^{\infty} \frac{1}{k^{1-\nu}} (E((y_{k+1} - Ey_{k+1})u_{k-1}^2) - (y_{k+1} - Ey_{k+1})u_{k-1}^2) \\ &\quad = \sum_{k=1}^{\infty} \frac{1}{k^{1-\nu}} (E(y_{k+1}((u_{k-1}^0)^2 - \vartheta^2)) - y_{k+1}((u_{k-1}^0)^2 - \vartheta^2)) \\ &\quad + \sum_{k=1}^{\infty} \frac{1}{k^{1-\nu}} (Ey_{k+1})((u_{k-1}^0)^2 - \vartheta^2) + \sum_{k=1}^{\infty} \frac{\vartheta^2}{k^{1-\nu}} (Ey_{k+1} - y_{k+1}) \\ &\quad + \sum_{k=1}^{\infty} \frac{1}{k^{1-\nu}} (Ey_{k+1} - y_{k+1})(\varepsilon_{k-1}^{(1)})^2 \\ &\quad + \sum_{k=1}^{\infty} \frac{2}{k^{1-\nu}} (Ey_{k+1} - y_{k+1}) u_{k-1}^0 \varepsilon_{k-1}^{(1)}. \end{aligned} \tag{106}$$

By (4.16) and (4.20) of Lemma 4.5 in [18], we see that the first term at the right-hand side of (106) converges a.s. Since $\{(u_{k-1}^0)^2 - \vartheta^2\}$ is a sequence of zero mean i.i.d. random variables and $E(y_{k+1}) < \infty$, by the Khintchine–Kolmogorov convergence theorem (see Theorem 1 of Section 5.1 in [10]), we find that the second term at the right-hand side of (106) converges a.s. due to

$$\sum_{k=1}^{\infty} \frac{(Ey_{k+1})^2}{k^{2(1-\nu)}} E((u_{k-1}^0)^2 - \vartheta^2)^2 < \infty.$$

For the fourth term at the right-hand side of (106), by (8) and (20) we have

$$\begin{aligned} &\sum_{k=1}^{\infty} \frac{1}{k^{1-\nu}} (Ey_{k+1} - y_{k+1})(\varepsilon_{k-1}^{(1)})^2 \\ &\quad = \sum_{j=0}^{\infty} h_j \sum_{k=j}^{\infty} \frac{1}{k^{1-\nu}} (Ef(v_{k-j}) - f(v_{k-j}))(\varepsilon_{k-1}^{(1)})^2 \\ &\quad - \sum_{k=1}^{\infty} \frac{1}{k^{1-\nu}} (C^{-1}(z)D(z)\eta_k)(\varepsilon_{k-1}^{(1)})^2 \\ &\quad - \sum_{k=1}^{\infty} \frac{1}{k^{1-\nu}} (C^{-1}(z)\xi_{k+1})(\varepsilon_{k-1}^{(1)})^2 \\ &\quad - \sum_{k=1}^{\infty} \frac{1}{k^{1-\nu}} \varepsilon_{k+1}^{(2)} (\varepsilon_{k-1}^{(1)})^2. \end{aligned} \tag{107}$$

Define $z_k^{(j)} \triangleq (1/k^{1-\nu})(Ef(v_{k-j}) - f(v_{k-j}))(\varepsilon_{k-1}^{(1)})^2$. Then by Remark 2 in [19] and by the hereditary property of mixing processes, $\{f(v_k)\}$ and $\{(\varepsilon_k^{(1)})^2\}$ are α-mixing sequences with the mixing coefficients decaying exponentially to zero. Since $\{v_k\}$ and $\{\varepsilon_k^{(1)}\}$ are mutually independent, $z_k^{(j)}$ is a zero mean α-mixing sequence with the mixing coefficients decaying exponentially to zero. By H5 and Lemma 3 in [19], we find that $E|f(v_{k-j})|^{2+\varepsilon} < \infty$ and $E|(\varepsilon_{k-1}^{(1)})|^{2(2+\varepsilon)} < \infty$ for some $\varepsilon > 0$. By applying the C_r inequalities, we have

$$\begin{aligned} &\sum_{k=j}^{\infty} (E|z_k^{(j)}|^{2+\varepsilon})^{2/(2+\varepsilon)} \leq \sum_{k=j}^{\infty} \frac{4}{k^{2(1-\nu)}} (E|f(v_{k-j})|^{2+\varepsilon})^{2/(2+\varepsilon)} \\ &\quad \cdot (E|(\varepsilon_{k-1}^{(1)})|^{2(2+\varepsilon)})^{2/(2+\varepsilon)} \leq O\left(\sum_{k=j}^{\infty} \frac{1}{k^{2(1-\nu)}}\right) < \infty. \end{aligned}$$

Therefore, by Lemma 4 in [19] we obtain

$$\sum_{k=j}^{\infty} z_k^{(j)} = \sum_{k=j}^{\infty} \frac{1}{k^{1-\nu}} (Ef(v_{k-j}) - f(v_{k-j}))(\varepsilon_{k-1}^{(1)})^2 < \infty \quad a.s.,$$

which implies that

$$\sum_{k=2j}^{\infty} \frac{1}{k^{1-\nu}} (Ef(v_{k-j}) - f(v_{k-j}))(\varepsilon_{k-1}^{(1)})^2 \underset{j \to \infty}{\longrightarrow} 0 \quad a.s.,$$

and

$$\left|\sum_{k=j}^{2j-1}\frac{1}{k^{1-\nu}}(Ef(v_{k-j})-f(v_{k-j}))(\varepsilon_{k-1}^{(1)})^2\right|$$

$$\le \sum_{k=j}^{2j-1}\frac{1}{j^{1-\nu}}o(j^c)o(j^{1/3})=o(j^{1/3+c+\nu})\quad a.s.$$

Since $|f(v_{k-j})|/j^c \underset{j\to\infty}{\longrightarrow} 0$ *a.s.* for any $c>0$ and $(\varepsilon_{k-1}^{(1)})^2/j^{1/3}\underset{j\to\infty}{\longrightarrow} 0$ *a.s.* by the Borel-Cantelli lemma [10], H5 and H4, when $j\le k<2j$. Thus we have

$$\left|\sum_{k=j}^{\infty}\frac{1}{k^{1-\nu}}(Ef(v_{k-j})-f(v_{k-j}))(\varepsilon_{k-1}^{(1)})^2\right| = o(j^{1/3+c+\nu})\quad a.s.$$

By noticing that $|h_j|=O(e^{-r_2 j}),\ r_2>0,\ j\ge 1$, we have

$$\sum_{j=0}^{\infty}h_j\sum_{k=j}^{\infty}\frac{1}{k^{1-\nu}}(Ef(v_{k-j})-f(v_{k-j}))(\varepsilon_{k-1}^{(1)})^2<\infty\quad a.s.$$

The convergence of the remaining terms at the right-hand side of (107) can be proved in a similar way, and hence the fourth term at the right-hand side of (106) converges a.s. Likewise, the rest at the right-hand side of (106) can be shown to be convergent a.s., and hence the second term at the right-hands of (104) also satisfies (105) with $\varepsilon_{k+1}^{(0,\rho)}$ replaced by it.

Noticing (93) and applying the method similar to that used to verify (73) in Lemma 7 of [19], we see that the last term at the right-hands of (104) satisfies (105) with $\varepsilon_{k+1}^{(0,\rho)}$ replaced by it.

Therefore, we have proved (102), while (103) can be similarly proved.

As pointed out before, by H5 at least one of ρ and κ is nonzero, so switching between (44)–(49) and (50)–(55) may happen only a finite number of times. Therefore, noticing (94) and (95), for proving (96) and (97) it suffices to show $\forall \nu\in(0,1/2)$,

$$|\theta_k^{(i,\rho)}-\rho l_i|=o(k^{-\nu})\quad a.s.\ i\ge 1, \tag{108}$$

$$\left|\lambda_k^{(i,\rho)}-\rho\sum_{j=0}^{i}h_j l_{i-j}^2\right|=o(k^{-\nu})\quad a.s.\ i\ge 1, \tag{109}$$

$$|\theta_k^{(i,\kappa)}-E[(y_{k+1}-Ey_{k+1})(u_{k-1}^2-Eu_{k-1}^2)u_{k-i-1}]|=o(k^{-\nu})\quad a.s.\ i\ge 1, \tag{110}$$

$$\left|\lambda_k^{(i,\kappa)}-\kappa\sum_{j=0}^{i}h_j l_{i-j}^3\right|=o(k^{-\nu})\quad a.s.\ i\ge 1. \tag{111}$$

Similar to proving (102), we can show (108)–(111), while the assertions (98)–(101) straightforwardly follow from (96) and (97). □

4. Strong consistency of estimates for $f(\cdot)$

Lemma 3. *Under Conditions H1–H7, the following assertions for $w_k(x)$ defined by (74) take place:*

$$E[w_k(x)]\underset{k\to\infty}{\longrightarrow} p(x),\quad E[w_k(x)f(v_k)]\underset{k\to\infty}{\longrightarrow} p(x)\tilde f(x), \tag{112}$$

where $p(x)=(1/\sqrt{2\pi}\sigma\vartheta)e^{-x^2/2\sigma^2\vartheta^2}$, $\sigma^2=\sum_{i=0}^{\infty}l_i^2$, and

$$\tilde f(x)=f(x^-)\int_{-\infty}^{x}K(t)\,\mathrm{d}t+f(x^+)\int_{x}^{\infty}K(t)\,\mathrm{d}t,$$

which equals f(x) for any x where $f(\cdot)$ is continuous.

For the proof, we refer to Lemma 9 in [19].

Lemma 4. *Assume that H1, H4, H6, and H7 hold. Then both $a_{0,k}(\varepsilon_k^{(1)})$ defined by (75) and (76) and $a_{i,k}(\varepsilon_k^{(1)}), i\ge 1$ defined by (77) and (78) have the convergence rate*

$$\left|a_{i,k}(\varepsilon_k^{(1)})-a_i(\varepsilon_k^{(1)})\right|=o\left(\frac{1}{k^{1/2-c}}\right)\quad \forall c>0,\ i\ge 0. \tag{113}$$

The proof is similar to that for Lemma 10 in [19].

Corollary 1. *Assume that H1–H7 hold. Then $\hat\sigma_k^2(e)$ defined by (86) converges to $\sigma_k^2(e)$ a.s. with the following convergence rate:*

$$\left|\hat\sigma_k^2(e)-\sigma_k^2(e)\right|=o\left(\frac{1}{k^{1/2-c}}\right)\quad \forall c>0. \tag{114}$$

Lemma 5. *Assume H1–H7 hold. Then there is a constant $c>0$ with $1/6-1/2b-2c>0$ such that*

$$\left|\zeta_k-\hat\zeta_k\right|=o\left(\frac{1}{k^{1/6-c}}\right), \tag{115}$$

$$|\psi_k-\hat\psi_k|=o\left(\frac{1}{k^{1/2-2c}}\right), \tag{116}$$

$$|w_k(x)-\hat w_k(x)|=o\left(\frac{(\log k)^{3/2}}{k^{1/2-1/2b-2c}}\right). \tag{117}$$

For (115), we refer to Lemma 4.10 in [18], while for (116) and (117), we refer to Lemma 11 in [19].

Theorem 2. *Assume that H1–H7 hold. Then $\mu_k(x)$ defined by (88) and (89) and $\beta_k(x)$ defined by (90) and (91) are convergent:*

$$\mu_k(x)\underset{k\to\infty}{\longrightarrow} p(x)\quad a.s., \tag{118}$$

$$\beta_k(x)\underset{k\to\infty}{\longrightarrow} p(x)\tilde f(x)\quad a.s. \tag{119}$$

As a consequence, $f_k(x)$ defined by (92) is strongly consistent

$$f_k(x)\underset{k\to\infty}{\longrightarrow}\tilde f(x)\quad a.s. \tag{120}$$

Proof. The algorithm (88) can be rewritten as

$$\mu_{k+1}(x)=[\mu_k(x)-(1/k)(\mu_k(x)-p(x))-(1/k)\bar e_{k+1}(x)]\cdot I_{[|\mu_k(x)-(1/k)(\mu_k(x)-p(x))-(1/k)\bar e_{k+1}(x)|\le M_{\delta_k^{(\mu)}(x)}]},$$

where

$$\bar e_{k+1}(x)=p(x)-\hat w_k(x)=[p(x)-Ew_k(x)]+[Ew_k(x)-w_k(x)]+[w_k(x)-\hat w_k(x)]. \tag{121}$$

Since $p(x)$ is the single root of the linear function $-(y-p(x))$, by Theorem 2.1.1 in [7], for convergence of $\mu_k(x)$ it suffices to show

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left|\sum_{j=n_k}^{m(n_k,T_k)}(1/j)\bar e_{j+1}(x)\right|=0\quad \forall T_k\in[0,T] \tag{122}$$

for any convergent subsequence $\mu_{n_k}(x)$, where $m(k,T)\triangleq\max\{m:\sum_{j=k}^{m}(1/j)\le T\}$. By the first limit in (112) and (117), it follows that (122) holds for the first and last terms at the right-hand side of (121), respectively. By the method similar to that used to verify (102) in [19], it is shown that the second term at the right-hand side of (121) satisfies (122). The proof of (119) can be similarly carried out. Therefore, the estimate (92) is strongly consistent. □

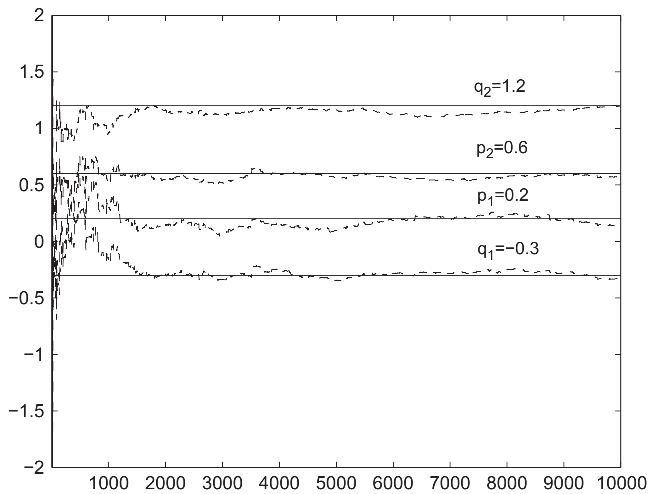

Fig. 2. Estimates for p_1, p_2, q_1, q_2.

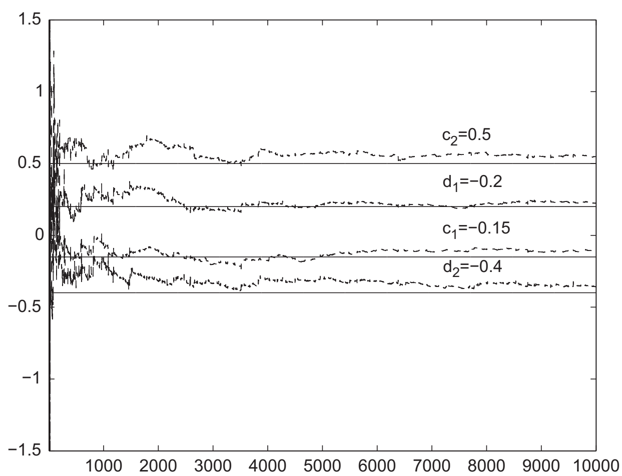

Fig. 3. Estimates for c_1, c_2, d_1, d_2.

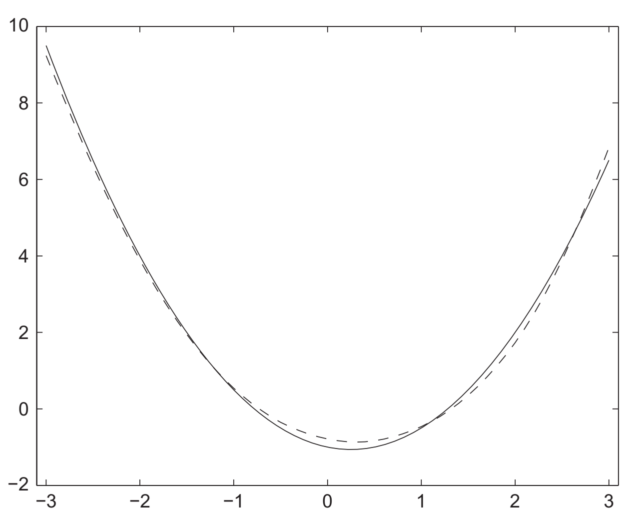

Fig. 4. Estimates for $f(x)=x^2-0.5x-1$.

5. Example

Let the linear subsystem at the input end be

$$v_k+p_1v_{k-1}+p_2v_{k-2}=u^0_{k-1}+q_1u^0_{k-2}+q_2u^0_{k-3},$$

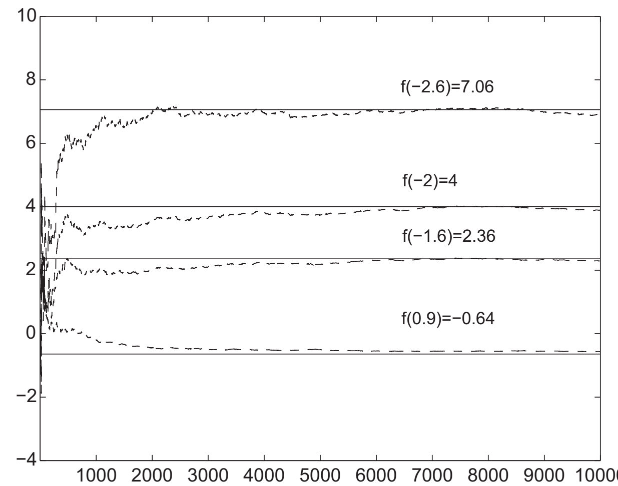

Fig. 5. Estimates for $f(x)$ at some fixed points.

where $p_1=0.2$, $p_2=0.6$, $q_1=-0.3$ and $q_2=1.2$, and let the linear subsystem at the output end be given by

$$y^0_{k+1}+c_1y^0_k+c_2y^0_{k-1}=\varphi_k+d_1\varphi_{k-1}+d_2\varphi_{k-2}+\xi_{k+1},$$

where $c_1=-0.15$, $c_2=0.5, d_1=0.2$ and $d_2=-0.4$. Let the nonlinear function be

$$f(x)=x^2-0.5x-1.$$

The input signal $\{u^0_k\}$ is a sequence of Gaussian random variables: $u^0_k\in\mathcal{N}(0,1.2^2)$. The driven noises $\{\varsigma^{(1)}_k\}$ and $\{\varsigma^{(2)}_k\}$ and the internal noises $\{\eta_k\}$ and $\{\xi_k\}$ all are sequences of mutually independent Gaussian random variables: $\mathcal{N}(0,0.3^2)$. The measurement noises $\{\varepsilon^{(1)}_k\}$ and $\{\varepsilon^{(2)}_k\}$ are the following ARMA processes, respectively:

$$\varepsilon^{(1)}_k-0.7\varepsilon^{(1)}_{k-1}=\varsigma^{(1)}_k+0.5\varsigma^{(1)}_{k-1},$$
$$\varepsilon^{(2)}_k+0.4\varepsilon^{(2)}_{k-1}=\varsigma^{(2)}_k-0.6\varsigma^{(2)}_{k-1}.$$

In the figures listed below, the solid lines represent the true values of the system, while the dashed lines denote the corresponding estimates. Figs. 2 and 3 demonstrate the recursive estimates for the coefficients of the two linear subsystems, respectively, while the estimate for the nonlinear function at time $k=10\,000$ is given in Fig. 4. The behavior of the estimates for the nonlinearity at points $\{-2.6,-2,-1.6,0.9\}$ versus time is demonstrated in Fig. 5.

6. Conclusion

The recursive algorithms for identifying the EIV Wiener–Hammerstein systems are proposed in the paper. The estimation is carried out by the stochastic approximation algorithms incorporated with the deconvolution kernel. The estimates for the two linear subsystems as well as for the nonlinearity are shown to be convergent to the true values with probability one.

For further research it is of interest to consider identification of other EIV nonlinear systems, for example, the MIMO EIV Wiener–Hammerstein systems.

Acknowledgments

This work was supported by NSFC under Grants 61273193, 61120106011, 61134013 and the National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences.

References

[1] J.C. Agüero, G.C. Goodwin, Identifiability of errors in variables dynamic systems, Automatica 44 (2) (2008) 371–382.

[2] E.W. Bai, Z.J. Cai, S. Dudley-Javorosk, R.K. Shields, Identification of a modified Wiener–Hammerstein system and its application in electrically stimulated paralyzed skeletal muscle modeling, Automatica 45 (3) (2009) 736–743.

[3] S.A. Billings, S.Y. Fakhouri, Identification of systems containing linear dynamic and static nonlinear elements, Automatica 18 (1) (1982) 15–26.

[4] M. Boutayeb, M. Darouach, Recursive identification method for MISO Wiener–Hammerstein model, IEEE Trans. Autom. Control 40 (February (2)) (1995) 287–291.

[5] D.R. Brillinger, The identification of a particular nonlinear time series system, Biometrika 64 (3) (1977) 509–515.

[6] C.H. Chen, S.D. Fassois, Maximum likelihood identification of stochastic Wiener–Hammerstein-type nonlinear systems, Mech. Syst. Signal Process. 6 (2) (1992) 135–153.

[7] H.F. Chen, Stochastic Approximation and its Applications, Kluwer, Dordrecht, The Netherlands, 2002.

[8] H.F. Chen, Recursive identification for multivariate errors-in-variables systems, Automatica 43 (7) (2007) 1234–1242.

[9] H.F. Chen, W.X. Zhao, New method of order estimation for ARMA/ARMAX processes, SIAM J. Control Optim. 48 (6) (2010) 4157–4176.

[10] Y.S. Chow, H. Teicher, Probability Theory: Independence, Interchangeability, Martingales, 3rd ed., Springer-Verlag, New York, 2003.

[11] K.B. Davis, Mean square error properties of density estimates, Ann. Stat. 3 (4) (1975) 1025–1030.

[12] T. Falck, P. Dreesen, K.D. Brabanter, K. Pelckmans, B.D. Moor, J.A.K. Suykens, Least-squares support vector machines for the identification of Wiener–Hammerstein systems, Control Eng. Pract. 20 (11) (2012) 1165–1174.

[13] J. Fan, Y.K. Truong, Nonparametric regression with errors in variables, Ann. Stat. 21 (4) (1993) 1900–1925.

[14] F. Giri, E.W. Bai, Block-oriented nonlinear system identification, in: T. Manfred, A. Frank, M. Manfred (Eds.), Lecture Notes in Control and Information Sciences, vol. 404, Springer, London, 2010.

[15] W. Greblicki, M. Pawlak, Nonparametric System Identification, Cambridge University Press, New York, 2008.

[16] I.W. Hunter, M.J. Korenberg, The identification of nonlinear biological system: Wiener and Hammerstein cascade models, Biol. Cybern. 55 (2–3) (1986) 135–144.

[17] B.E. Jun, D.S. Bernstein, Extended least-correlation estimates for errors-in-variables non-linear models, Int. J. Control 80 (2) (2007) 256–267.

[18] B.-Q. Mu, H.-F. Chen, Recursive identification of Wiener–Hammerstein systems, SIAM J. Control Optim. 50 (5) (2012) 2621–2658.

[19] B.-Q. Mu, H.-F. Chen, Recursive identification of errors-in-variables Wiener systems, Automatica 49 (9) (2013) 2744–2753.

[20] B.-Q. Mu, H.-F. Chen, Recursive identification of MIMO Wiener systems, IEEE Trans. Autom. Control 58 (March (3)) (2013) 802–808.

[21] B.-Q. Mu, H.-F. Chen, Hankel matrices for system identification, J. Math. Anal. Appl. 409 (1) (2014) 494–508.

[22] A.V. Nazin, B.T. Polyak, A.B. Tsybakov, Passive stochastic approximation, Autom. Remote Control 50 (3) (1989) 1563–1569.

[23] P. Révész, How to apply the method of stochastic approximation in the non-parametric estimation of a regression function, Ser. Stat. 8 (1) (1977) 119–126.

[24] T. Söderström, Errors-in-variables methods in system identification, Automatica 43 (6) (2007) 939–958.

[25] J. Sjöberg, J. Schoukens, Initializing Wiener–Hammerstein models based on partitioning of the best linear approximation, Automatica 48 (2) (2012) 353–359.

[26] Q.J. Song, H.F. Chen, Identification of errors-in-variables systems with ARMA observation noise, Syst. Control Lett. 57 (5) (2008) 420–424.

[27] L. Stefanski, R.J. Carroll, Deconvoluting kernel density estimators, Statistics 21 (2) (1990) 169–184.

[28] I. Vajk, J. Hetthéssy, Identification of nonlinear errors-in-variables models, Automatica 39 (12) (2003) 2099–2107.

[29] D.T. Westwick, J. Schoukens, Initial estimates of the linear subsystems of Wiener–Hammerstein models, Automatica 48 (11) (2012) 2931–2936.

[30] W.X. Zhao, H.F. Chen, Stochastic approximation based PCA and its application to identification of EIV systems, in: Proceedings of the 10-th World Congress on Intelligent Control and Automation, Beijing, 2012, pp. 3276–3280.