Automatica 49 (2013) 2744–2753

Contents lists available at ScienceDirect

Automatica

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

Recursive identification of errors-in-variables Wiener systems[☆]

Bi-Qiang Mu, Han-Fu Chen [1]

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

ARTICLE INFO

Article history:
Received 23 January 2013
Received in revised form
2 May 2013
Accepted 8 June 2013
Available online 12 July 2013

Keywords:
Wiener systems
Errors-in-variables
Stochastic approximation
Recursive estimation
α-mixing
Strong consistency

ABSTRACT

This paper considers the recursive identification of errors-in-variables (EIV) Wiener systems composed of a linear dynamic system followed by a static nonlinearity. Both the system input and output are observed with additive noises being ARMA processes with unknown coefficients. By a stochastic approximation incorporated with the deconvolution kernel functions, the recursive algorithms are proposed for estimating the coefficients of the linear subsystem and for the values of the nonlinear function. 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 Wiener system composed of a dynamic linear subsystem followed by a static nonlinear function can be used to model the majority of practical systems, for example, distillation column (Zhu, 1999), pH process (Kalafatis, Arifin, Wang, & Cluett, 1995), biological cybernetics (Hunter & Korenberg, 1986), power amplifier (Kang, Cho, & Youn, 1999) and others. It is shown that the Wiener system can capture complex nonlinear phenomena in the sense that almost any nonlinear system with fading memory can be approximated by a Wiener system with an arbitrarily high accuracy (Boyd & Chua, 1985). Thus, identification of Wiener systems has received considerable attention from both theoretical researchers and engineers.

To identify the nonlinear function in a Wiener system there are parametric (Bai, 2003; Hagenblad, Ljung, & Wills, 2008; Wigren, 1994) and nonparametric approaches (Greblicki & Pawlak, 2008; Hu & Chen, 2006; Mu & Chen, 2012, 2013; Zhao & Chen, 2012a), according to the description 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. In this case identification turns to be a parametric estimation problem that can be solved by a standard optimization method such as the gradient method, Newton–Raphson method, the extended least squares and so on. The nonparametric approach is used to estimate values of the nonlinear function at any given points with the help of kernel functions. This approach requires no structural information about nonlinearity. We adopt the nonparametric method in the paper, but we consider the case where the input and output of the system are not accurately available. They may be observed but with additive noises, i.e., we intend to identify the errors-in-variables (EIV) Wiener systems.

There exist many papers on identifiability and identification of the linear EIV systems (Agüero & Goodwin, 2008; Söderström, 2007). Various estimation methods for identifying linear EIV systems, for example, instrumental variables based methods, bias-compensation approaches, the Frisch scheme, frequency domain methods, prediction error and ML methods, are well summarized in the survey paper (Söderström, 2007), but the methods mentioned there are nonrecursive. Recursive identification for linear EIV systems is considered in Chen (2007), Song and Chen (2008) and Zhao and Chen (2012b), but there is little attention paid to nonlinear EIV systems.

☆ This work was supported by the NSFC under Grants 61273193, 61120106011, and 61134013 and the National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Wei Xing Zheng under the direction of Editor Torsten Söderström.

E-mail addresses: bqmu@amss.ac.cn (B.-Q. Mu), hfchen@iss.ac.cn (H.-F. Chen).

[1] Tel.: +86 10 62579540; fax: +86 10 62587343.

0005-1098/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.automatica.2013.06.022

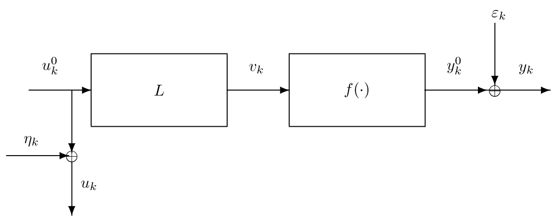

Fig. 1. EIV Wiener system.

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

$$C(z)v_k = D(z)u_k^0, \quad y_k^0 = f(v_k) \tag{1}$$

$$C(z) = 1 + c_1 z + \cdots + c_p z^p, \tag{2}$$

$$D(z) = z + d_2 z^2 + \cdots + d_q z^q, \tag{3}$$

where z is the backward shift operator: $zy_k = y_{k-1}$, $C(z)$ and $D(z)$ are polynomials with unknown coefficients but with known orders p, q, respectively. The signal v_k is not directly observed. The input and output u_k^0 and y_k^0 are observed with additive noises η_k and ε_k:

$$u_k = u_k^0 + \eta_k, \qquad y_k = y_k^0 + \varepsilon_k. \tag{4}$$

The paper is to recursively estimate $\{c_1, \ldots, c_p, d_2, \ldots, d_q\}$ of the linear part and the value of $f(x)$ at any given x on the basis of the observed data $\{u_k, y_k\}$.

The assumptions made on the system and the recursive algorithms are given in Section 2. Some auxiliary results on α-mixing are listed in Section 3. The strong consistency of the estimates for the linear part and the nonlinearity is proved in Sections 4 and 5, respectively. A numerical example is presented in Section 6, and a brief conclusion is given in Section 7.

2. Assumptions and recursive algorithms

2.1. Assumptions

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

H2 $C(z)$ and $D(z)$ are coprime and $C(z)$ is stable: $C(z) \neq 0\ \forall |z| \leq 1$.

By stability of $C(z)$ we have

$$H(z) \triangleq \frac{D(z)}{C(z)} = \sum_{i=1}^{\infty} h_i z^i, \tag{5}$$

where $|h_i| = O(e^{-ri}), r > 0, i \geq 2$, and $h_1 = 1$. The numbers $\{h_i, i \geq 1\}$ are called the impulse responses of the linear subsystem.

H3 The noises η_k and ε_k both are ARMA processes:

$$P(z)\eta_k = Q(z)\zeta_k, \qquad F(z)\varepsilon_k = G(z)\varsigma_k, \quad \text{where} \tag{6}$$

$$P(z) = 1 + p_1 z + p_2 z^2 + \cdots + p_{n_p} z^{n_p}, \tag{7}$$

$$Q(z) = 1 + q_1 z + q_2 z^2 + \cdots + q_{n_q} z^{n_q}, \tag{8}$$

$$F(z) = 1 + f_1 z + f_2 z^2 + \cdots + f_{n_f} z^{n_f}, \tag{9}$$

$$G(z) = 1 + g_1 z + g_2 z^2 + \cdots + g_{n_g} z^{n_g}. \tag{10}$$

The polynomial $P(z)$ has no common roots with $Q(z)Q(z^{-1})z^{n_q}$, and $P(z)$ and $F(z)$ are stable. The driven noises $\{\zeta_k\}$ and $\{\varsigma_k\}$ are mutually independent, and each of them is a sequence of i.i.d. zero mean random variables with probability density. Moreover, $\mathrm{E}(|\zeta_k|^{\Delta+2}) < \infty$ and $\mathrm{E}(|\varsigma_k|^{\Delta}) < \infty$ for some $\Delta > 2$.

H4 The function $f(\cdot)$ is measurable and has the left and right limits $f(x^-)$ and $f(x^+)$ at any x. As $|x| \to \infty$, $f(x)$ grows no faster than a polynomial. Further, at least one of the constants τ and ρ is nonzero, where

$$\tau \triangleq \frac{1}{\sqrt{2\pi}\sigma^3\vartheta} \int_{\mathbb{R}} x f(x) e^{-\frac{x^2}{2\sigma^2\vartheta^2}} \mathrm{d}x, \tag{11}$$

$$\rho \triangleq \frac{1}{\sqrt{2\pi}\sigma^5\vartheta} \int_{\mathbb{R}} \left(x^2 - \sigma^2\vartheta^2\right) f(x) e^{-\frac{x^2}{2\sigma^2\vartheta^2}} \mathrm{d}x, \tag{12}$$

where $\sigma^2 \triangleq \sum_{i=1}^{\infty} h_i^2$.

Remark 1. H4 implies that there are a positive number $\alpha > 0$ and an integer $\beta \geq 1$ such that

$$|f(x)| \leq \alpha(1 + |x|^{\beta}) \quad \forall x \in \mathbb{R}. \tag{13}$$

Therefore, under H4 the integrals (11)–(12) are finite.

H5 The variance ϑ^2 of the noise-free input u_k^0 is known.

H6 The driven noise $\{\zeta_k\}$ in (6) is a sequence of zero mean i.i.d. Gaussian random variables.

Let us first explain these assumptions. It is worth noting that for identifying the linear subsystem we only need H1–H4, while for estimating $f(\cdot)$ we have to additionally impose H5–H6. It is noted that H2 is a standard condition, while H4 is satisfied by a large class of nonlinear functions. The function $f(\cdot)$ is allowed to be discontinuous, and the nonzero condition for τ or ρ is not restrictive. For example, all polynomials, no matter if they are even or odd, are possible to meet the requirement. Let $f(\cdot)$ be a monic polynomial with arbitrary coefficients.

If $f(x) = x^2 + ax + b$, then $\tau = a\vartheta^2$ and $\rho = 2\vartheta^4 > 0$. If $f(x) = x^3 + ax^2 + bx + c$, then $\tau = (3\sigma^2\vartheta^2 + b)\vartheta^2$ and $\rho = 2a\vartheta^4$. Both τ and ρ equal zero only in the case where $3\sigma^2\vartheta^2 + b = 0$ and $a = 0$, which, however, can easily be violated by slightly changing the variance ϑ^2.

If $f(x) = x^4 + ax^3 + bx^2 + cx + d$, then $\tau = (3\sigma^2\vartheta^2 a + c)\vartheta^2$ and $\rho = 2\vartheta^4(6\sigma^2\vartheta^2 - b)$. Similarly, both τ and ρ are zero only in the case where $3\sigma^2\vartheta^2 a = -c$ and $6\sigma^2\vartheta^2 = b$, which can be violated by changing the variance ϑ^2. The higher order polynomials can be discussed in a similar manner. Therefore, H4 is not a difficult condition for practical systems.

Conditions H3 and H6 allow the measurement noises to be correlated.

For practical systems in operating it may not be reasonable to assume that their inputs are Gaussian. However, in the paper we aim at identifying systems, assuming the system inputs are at the user's disposal. So, H1 can be met. The purpose of applying the Gaussian input is to derive the simple relationships (15)–(17) connecting the impulse responses of the linear subsystem and the correlation functions between the observed input and output. These relationships are the basis of the algorithms for estimating the impulse responses.

We now explain the necessity unavoidability of H5 in the present approach to estimating $f(\cdot)$. As a matter of fact, $f(\cdot)$ is estimated by using the Stochastic Approximation Algorithm With Expanding Truncations (SAAWET) (Chen, 2002) with the help of the deconvolution kernel function $w(x)$. To define $w(x)$ one needs H6 (see (47) and (48)), while to estimate $w(x)$ one has to estimate the variance of e_k defined in (41). However, the observed input is $u_k = u_k^0 + \eta_k$, so the variance of e_k can be estimated on the basis of $\{u_k\}$ together with the estimate for the linear subsystem only if the variance ϑ^2 of u_k^0 is available. When a practical system is identified, maybe, its input signal cannot be designed by the user, but knowing its statistical properties may still be possible. This is the reason to impose H5.

2.2. Estimation of $\{c_1, \ldots, c_p, d_2, \ldots, d_q\}$

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

$$v_k = \sum_{i=1}^{k} h_i u_{k-i}^0, \tag{14}$$

and hence $v_k \in \mathcal{N}(0, \sigma_k^2)$ where $\sigma_k^2 \triangleq \sum_{i=1}^{k} h_i^2 \xrightarrow[k\to\infty]{} \sigma^2$.

Lemma 1. *Assume* H1–H4 *hold. Then*

$$Ey_k u_{k-i} \xrightarrow[k\to\infty]{} \tau h_i, \quad \forall i \geq 1, \tag{15}$$

$$E(y_k - Ey_k)u_{k-1}^2 \xrightarrow[k\to\infty]{} \rho, \tag{16}$$

$$E(y_k - Ey_k)u_{k-1}u_{k-i} \xrightarrow[k\to\infty]{} \rho h_i, \quad \forall i \geq 2. \tag{17}$$

Proof. Similar to Lemma 2 in Hu and Chen (2006) or Lemma 3.2 in Mu and Chen (2012), we have

$$E(f(v_k)u_{k-i}^0) = \tau_k h_i \xrightarrow[k\to\infty]{} \tau h_i,$$

where $\tau_k \triangleq \frac{1}{\sigma_k^2} E\left(f(v_k)v_k\right)$.

Since u_k^0 is independent of η_k and ε_k, we have

$$Ey_k u_{k-i} = E(f(v_k)u_{k-i}^0) \xrightarrow[k\to\infty]{} \tau h_i.$$

By Lemma 3.1 in Mu and Chen (2012), we obtain

$$\begin{aligned} &E(f(v_k)((u_{k-1}^0)^2 - \vartheta^2)) \\ &\quad = \frac{1}{\sigma_k^4} E(f(v_k)(v_k)^2) - \frac{\vartheta^2}{\sigma_k^2} Ef(v_k) \triangleq \rho_k, \quad \text{and} \end{aligned} \tag{18}$$

$$\begin{aligned} E(f(v_k)u_{k-1}^0 u_{k-j}^0) &= \left(\frac{1}{\sigma_k^4} E(f(v_k)(v_k)^2) - \frac{\vartheta^2}{\sigma_k^2} Ef(v_k)\right) h_j \\ &= \rho_k h_j, \quad j \geq 2. \end{aligned} \tag{19}$$

It is noticed that $\sigma_k^2 \xrightarrow[k\to\infty]{} \sigma^2$, by (12) we see that

$$\begin{aligned} &E(y_k - Ey_k)u_{k-1}^2 \\ &\quad = E(y_k^0 - Ey_k^0)((u_{k-1}^0)^2 + \eta_{k-1}^2 + 2u_{k-1}^0\eta_{k-1}) \\ &\quad = E[(y_k^0 - Ey_k^0)(u_{k-1}^0)^2] = Ey_k^0((u_{k-1}^0)^2 - \vartheta^2) \\ &\quad = \rho_k \xrightarrow[k\to\infty]{} \rho \quad \text{and} \end{aligned} \tag{20}$$

$$\begin{aligned} &E(y_k - Ey_k)u_{k-1}u_{k-i} \\ &\quad = E(y_k^0 + \varepsilon_k - Ey_k^0)(u_{k-1}^0 + \eta_{k-1})(u_{k-i}^0 + \eta_{k-i}) \\ &\quad = E(y_k^0 - Ey_k^0)(u_{k-1}^0 u_{k-i}^0) = Ey_k^0 u_{k-1}^0 u_{k-i}^0 \\ &\quad = \rho_k h_i \xrightarrow[k\to\infty]{} \rho h_i, \quad i \geq 2. \end{aligned} \tag{21}$$

The proof of the lemma is completed. □

The idea of estimating the coefficients of the linear subsystem consists in that we first estimate the impulse responses $\{h_i\}$ and then obtain the estimates for the coefficients $\{c_1, \ldots, c_p, d_2, \ldots, d_q\}$ by using the linear algebraic equations connecting them with $\{h_i\}$.

Let us first use SAAWET to recursively estimate Ey_k:

$$\lambda_k = \left[\lambda_{k-1} - (1/k)\left(\lambda_{k-1} - y_k\right)\right] I_{A_k}, \tag{22}$$

$$\delta_k^{(\lambda)} = \sum_{j=1}^{k-1} I_{A_j^c}, \tag{23}$$

where $A_k \triangleq \{|\lambda_{k-1} - (1/k)(\lambda_{k-1} - y_k)| \leq M_{\delta_k^{(\lambda)}}\}$ and A_k^c denotes the complement of A_k, $\{M_k\}$ is an arbitrarily chosen sequence of positive real numbers increasingly diverging to infinity, λ_0 is an arbitrary initial value, and I_A denotes the indicator function of a set A.

In the following, the notation A_k will be repeatedly used but its definition changes from place to place.

Before giving the estimates for h_i, the constants τ and ρ are needed to be estimated on the basis of (15) and (16), respectively. Their estimates are given as follows:

$$\theta_k^{(1,\tau)} = \left[\theta_{k-1}^{(1,\tau)} - (1/k)\left(\theta_{k-1}^{(1,\tau)} - y_k u_{k-1}\right)\right] I_{A_k}, \tag{24}$$

$$\delta_k^{(1,\tau)} = \sum_{j=1}^{k-1} I_{A_j^c}, \tag{25}$$

$$\theta_k^{(1,\rho)} = \left[\theta_{k-1}^{(1,\rho)} - (1/k)\left(\theta_{k-1}^{(1,\rho)} - (y_k - \lambda_k)u_{k-1}^2\right)\right] I_{A_k}, \tag{26}$$

$$\delta_k^{(1,\rho)} = \sum_{j=1}^{k-1} I_{A_j^c}, \tag{27}$$

where A_k in (24) is $\{|\theta_{k-1}^{(1,\tau)} - (1/k)(\theta_{k-1}^{(1,\tau)} - y_k u_{k-1})| \leq M_{\delta_k^{(1,\tau)}}\}$, while in (26) is $\{|\theta_{k-1}^{(1,\rho)} - (1/k)\theta_{k-1}^{(1,\rho)} - (y_k - \lambda_k)u_{k-1}^2| \leq M_{\delta_k^{(1,\rho)}}\}$.

If $|\theta_k^{(1,\tau)}| \geq |\theta_k^{(1,\rho)}|$, then the following algorithm based on (15) is used to estimate τh_i:

$$\theta_k^{(i,\tau)} = \left[\theta_{k-1}^{(i,\tau)} - (1/k)\left(\theta_{k-1}^{(i,\tau)} - y_k u_{k-i}\right)\right] I_{A_k}, \tag{28}$$

$$\delta_k^{(i,\tau)} = \sum_{j=1}^{k-1} I_{A_j^c}, \quad i \geq 2, \tag{29}$$

where $A_k \triangleq \{|\theta_{k-1}^{(i,\tau)} - (1/k)(\theta_{k-1}^{(i,\tau)} - y_k u_{k-i})| \leq M_{\delta_k^{(i,\tau)}}\}$. Here $\theta_{k-1}^{(i,\tau)}$ is obtained from the previous step of the recursion if $|\theta_{k-1}^{(1,\tau)}| \geq |\theta_{k-1}^{(1,\rho)}|$. Otherwise, $\theta_{k-1}^{(i,\tau)}$ in (28) is set to equal $\theta_{k-1}^{(1,\tau)} h_{i,k-1}$. After having the estimates for τ and τh_i, the estimates for the impulse responses $\{h_i, i \geq 2\}$ at time k are given by

$$h_{i,k} \triangleq \begin{cases} \dfrac{\theta_k^{(i,\tau)}}{\theta_k^{(1,\tau)}}, & \text{if } \theta_k^{(1,\tau)} \neq 0, \\ 0, & \text{if } \theta_k^{(1,\tau)} = 0. \end{cases} \tag{30}$$

Conversely, if $|\theta_k^{(1,\rho)}| > |\theta_k^{(1,\tau)}|$, then based on (17), ρh_i is estimated by the following algorithm:

$$\theta_k^{(i,\rho)} = \left[\theta_{k-1}^{(i,\rho)} - (1/k)\left(\theta_{k-1}^{(i,\rho)} - (y_k - \lambda_k)u_{k-1}u_{k-i}\right)\right] I_{A_k}, \tag{31}$$

$$\delta_k^{(i,\rho)} = \sum_{j=1}^{k-1} I_{A_j^c}, \quad i \geq 2. \tag{32}$$

where $A_k \triangleq \{|\theta_{k-1}^{(i,\rho)} - (1/k)\theta_{k-1}^{(i,\rho)} - (y_k - \lambda_k)u_{k-1}u_{k-i}| \leq M_{\delta_k^{(i,\rho)}}\}$. Similar to the previous case, $\theta_{k-1}^{(i,\rho)}$ is derived from the previous step of the recursion if $|\theta_{k-1}^{(1,\tau)}| < |\theta_{k-1}^{(1,\rho)}|$. Otherwise, $\theta_{k-1}^{(i,\rho)}$ in (31) is set to equal $\theta_{k-1}^{(1,\rho)} h_{i,k-1}$. After having the estimates for ρ and ρh_i, the estimates for the impulse responses $\{h_i, i \geq 2\}$ at time k are given by

$$h_{i,k} \triangleq \begin{cases} \dfrac{\theta_k^{(i,\rho)}}{\theta_k^{(1,\rho)}}, & \text{if } \theta_k^{(1,\rho)} \neq 0, \\ 0, & \text{if } \theta_k^{(1,\rho)} = 0. \end{cases} \tag{33}$$

It is important to note that after establishing strong consistency of $\theta_k^{(1,\tau)}$ and $\theta_k^{(1,\rho)}$ in Section 4, switching between the algorithms

(28)–(30) and (31)–(33) ceases in a finite number of steps, because by H4 at least one of τ and ρ is nonzero and hence either $\theta_k^{(1,\tau)} \geq \theta_k^{(1,\rho)}$ or $\theta_k^{(1,\tau)} < \theta_k^{(1,\rho)}$ takes place for all sufficiently large k.

Once the estimates $h_{i,k}$ for the impulse responses h_i are obtained, the parameters $\{c_1, \ldots, c_p, d_2, \ldots, d_q\}$ of the linear subsystem can be derived by the convolution relationship between $\{h_i\}$ and $\{c_1, \ldots, c_p, d_2, \ldots, d_q\}$.

In fact, from (5) it follows that

$$z + d_2 z^2 \cdots + d_q z^q = (1 + c_1 z + \cdots + c_p z^p) \cdot (z + h_2 z^2 + \cdots + h_i z^i + \cdots), \tag{34}$$

which, by identifying coefficients for the same powers of z at both sides, implies

$$d_i = \sum_{j=0}^{(i-1)\wedge p} c_j h_{i-j}, \quad \forall\, 2 \leq i \leq q, \tag{35}$$

$$h_i = -\sum_{j=1}^{(i-1)\wedge p} c_j h_{i-j}, \quad \forall\, i \geq q+1, \tag{36}$$

where $c_0 = 1$ and $a \wedge b$ denotes $\min(a, b)$. Define the Hankel matrix

$$\Gamma \triangleq \begin{bmatrix} h_q & h_{q-1} & \cdots & h_{q-p+1} \\ h_{q+1} & h_q & \cdots & h_{q-p+2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{q+p-1} & h_{q+p-2} & \cdots & h_q \end{bmatrix}, \tag{37}$$

where $h_i \triangleq 0$ for $i \leq 0$.

For h_i, $q+1 \leq i \leq q+p$, by (36) and (37) we obtain the following linear algebraic equation:

$$\Gamma [c_1, c_2, \ldots, c_p]^T = -[h_{q+1}, h_{q+2}, \ldots, h_{q+p}]^T. \tag{38}$$

Noticing that the matrix Γ is nonsingular under H2 (see Mu & Chen, 2013; Zhao & Chen, 2012a) and that $h_{i,k} \xrightarrow[k\to\infty]{} h_i$ a.s. as to be shown by Theorem 1, we see that Γ_k is nonsingular when k is sufficiently large, where Γ_k is obtained from Γ with h_i replaced by its estimates $h_{i,k}$, and $h_{i,k} = 0$ for $i \leq 0$. The estimates for $\{c_1, \ldots, c_p, d_2, \ldots, d_q\}$ are naturally defined as:

$$[c_{1,k}, c_{2,k}, \ldots, c_{p,k}]^T \triangleq -\Gamma_k^{-1} [h_{q+1,k}, h_{q+2,k}, \ldots, h_{q+p,k}]^T, \tag{39}$$

$$d_{i,k} \triangleq h_{i,k} + \sum_{j=1}^{(i-1)\wedge p} c_{j,k} h_{i-j,k}, \quad i = 2, \ldots, q. \tag{40}$$

2.3. Estimation of $f(\cdot)$

We now recursively estimate $f(x)$, where x is an arbitrary point on the real axis. Since $\{v_k\}$ is not directly available, the conventional kernel estimation method (Fan & Yao, 2003) cannot be used. We apply the deconvolution kernel functions (Davis, 1975; Fan & Truong, 1993; Stefanski & Carroll, 1990) to estimate $f(x)$. Instead of directly estimating v_k let us estimate the signal ψ_k defined below, which, in fact, is a noisy v_k.

Define

$$\psi_k \triangleq C^{-1}(z) D(z) u_k, \qquad e_k \triangleq C^{-1}(z) D(z) \eta_k. \tag{41}$$

According to (1), (4) and (6), we have

$$\begin{aligned} \psi_k &= C^{-1}(z) D(z) u_k^0 + [C(z)P(z)]^{-1} D(z) Q(z) \zeta_k \\ &= v_k + e_k. \end{aligned} \tag{42}$$

Define

$$C \triangleq \begin{bmatrix} -c_1 & 1 & & \\ \vdots & & \ddots & \\ \vdots & & & 1 \\ -c_s & 0 & \cdots & 0 \end{bmatrix},$$

$$D \triangleq \begin{bmatrix} 1 \\ d_2 \\ \vdots \\ d_s \end{bmatrix}, \quad \text{and} \quad H \triangleq \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $s \triangleq \max(p, q)$, $c_i \triangleq 0$ for $i > p$ and $d_j \triangleq 0$ for $j > q$. Then, Eq. (41) connecting ψ_k and u_k can be written as

$$\psi_k + c_1 \psi_{k-1} + \cdots + c_p \psi_{k-p} = u_{k-1} + d_2 u_{k-2} + \cdots + d_q u_{k-q},$$

or in the state space form

$$x_{k+1} = C x_k + D u_k, \qquad \psi_{k+1} = H^T x_{k+1}. \tag{43}$$

Replacing c_i and d_j in C and D with $c_{i,k}$ and $d_{j,k}$ given by (39) and (40), respectively, $i = 1, \ldots, s$, $j = 1, \ldots, s$, we obtain the estimates C_k and D_k for C and D at time k, and hence, the estimate $\widehat{\psi}_k$ for ψ_k is given as follows:

$$\widehat{x}_{k+1} = C_{k+1} \widehat{x}_k + D_{k+1} u_k, \qquad \widehat{\psi}_{k+1} = H^T \widehat{x}_{k+1} \tag{44}$$

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

In order to eliminate the influence of e_k involved in ψ_k, we will use the Sinc kernel function (Davis, 1975; Stefanski & Carroll, 1990) and its Fourier transformation

$$K(x) = \frac{\sin(x)}{\pi x}, \tag{45}$$

$$\Phi_K(t) \triangleq \int_{\mathbb{R}} e^{\iota t x} K(x) \mathrm{d}x = I_{[|t| \leq 1]} \tag{46}$$

where ι stands for the imaginary unit $\iota^2 = -1$.

Under H6 $\{e_k\}$ is a sequence of zero mean Gaussian random variables, and its characteristic function is

$$\Phi_{e_k}(t) \triangleq \int_{\mathbb{R}} e^{\iota t x} \frac{1}{\sqrt{2\pi}\,\sigma_k(e)} e^{-\frac{x^2}{2\sigma_k^2(e)}} \mathrm{d}t = e^{-\frac{\sigma_k^2(e) t^2}{2}},$$

where $\sigma_k^2(e) \triangleq E e_k^2$. 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 $r_e > 0$.

Define

$$\begin{aligned} K_k(x) &\triangleq \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\iota t x} \frac{\Phi_K(t)}{\Phi_{e_k}(t/b_k)} \mathrm{d}t \\ &= \frac{1}{2\pi} \int_{-1}^{1} e^{-\iota t x} e^{\frac{\sigma_k^2(e) t^2}{2 b_k^2}} \mathrm{d}t, \end{aligned} \tag{47}$$

where $b_k = \left(b \sigma_k^2(e) / \log k\right)^{1/2}$ is the bandwidth with a chosen $b > 3$. The deconvolution kernel function $w_k(x)$ is defined by

$$\begin{aligned} w_k(x) &\triangleq K_k((\psi_k - x)/b_k)/b_k \\ &= \frac{1}{2\pi b_k} \int_{-1}^{1} \cos[(\psi_k - x) t / b_k] e^{\frac{\sigma_k^2(e) t^2}{2 b_k^2}} \mathrm{d}t \\ &= \frac{1}{\pi} \int_0^{\frac{1}{b_k}} \cos[(\psi_k - x) t] e^{\frac{\sigma_k^2(e) t^2}{2}} \mathrm{d}t. \end{aligned} \tag{48}$$

We first estimate the spectral density of η_k and then the spectral density of e_k with the help of the estimates for the linear subsystem. Finally, the estimate $\widehat{\sigma}_k^2(e)$ for $\sigma_k^2(e)$ can be derived by the inverse Fourier transformation of the spectral density estimate for e_k.

For simplicity, we assume that the orders n_p and n_q in (7) and (8) are known. When they are unknown, their strongly consistent estimates can be derived by the method provided in Chen and Zhao (2010).

The autocovariances $a_i(\eta) \triangleq \mathrm{E}(\eta_k\eta_{k-i}), i \geq 0$ of η_k can be recursively estimated by SAAWET:

$$a_{0,k}(\eta) = \left[a_{0,k-1}(\eta) - (1/k)(a_{0,k-1}(\eta) + \vartheta^2 - u_k^2)\right] I_{A_k}, \tag{49}$$

$$\delta_k^{(0,\eta)} = \sum_{j=1}^{k-1} I_{A_j^c}, \tag{50}$$

$$a_{i,k}(\eta) = \left[a_{i,k-1}(\eta) - (1/k)(a_{i,k-1}(\eta) - u_k u_{k-i})\right] I_{A_k}, \tag{51}$$

$$\delta_k^{(i,\eta)} = \sum_{j=1}^{k-1} I_{A_j^c}, \quad i \geq 1. \tag{52}$$

where A_k in (49) is $\{|a_{0,k-1}(\eta) - (1/k)(a_{0,k-1}(\eta) + \vartheta^2 - u_k^2)| \leq M_{\delta_k^{(0,\eta)}}\}$ and in (51) is $\{|a_{i,k-1}(\eta) - (1/k)(a_{i,k-1}(\eta) - u_k u_{k-i})| \leq M_{\delta_k^{(i,\eta)}}\}$.

Define the Hankel matrix

$$\Gamma_k(\eta) \triangleq \begin{bmatrix} a_{n_q,k}(\eta) & a_{n_q-1,k}(\eta) & \cdots & a_{n_q-n_p+1,k}(\eta) \\ a_{n_q+1,k}(\eta) & a_{n_q,k}(\eta) & \cdots & a_{n_q-n_p+2,k}(\eta) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_q+n_p-1,k}(\eta) & a_{n_q+n_p-2,k}(\eta) & \cdots & a_{n_q,k}(\eta) \end{bmatrix},$$

where $a_{i,k}(\eta) \triangleq a_{-i,k}(\eta)$ for $i < 0$. Since $a_{i,k}(\eta) \xrightarrow[k\to\infty]{} a_i(\eta), i \geq 0$ as to be shown in Lemma 10 and the limit of $\Gamma_k(\eta)$ is nonsingular under H3 (Stoica, 1983), the matrix $\Gamma_k(\eta)$ is nonsingular for sufficiently large k. Therefore, at time k, the parameters $\{p_1, \ldots, p_{n_p}\}$ can be estimated by the Yule–Walker equation

$$[p_{1,k}, \cdots, p_{n_p,k}]^T = -\Gamma_k^{-1}(\eta) \cdot [a_{n_q+1,k}(\eta), a_{n_q+2,k}(\eta), \ldots, a_{n_q+n_p,k}(\eta)]^T. \tag{53}$$

The spectral density $S_{\eta_k}(z)$ of η_k is equal to

$$S_{\eta_k}(z) \triangleq \sum_{l=-\infty}^{\infty} a_l(\eta) z^l = \frac{Q(z)Q(z^{-1})\sigma_\zeta^2}{P(z)P(z^{-1})},$$

where σ_ζ^2 denotes the variance of ζ_k.

Identifying coefficients of the same order of z at both sides of the equation

$$P(z)P(z^{-1}) \sum_{l=-\infty}^{\infty} a_l(\eta) z^l = Q(z)Q(z^{-1})\sigma_\zeta^2,$$

we derive

$$Q(z)Q(z^{-1})\sigma_\zeta^2 = \sum_{l=-n_q}^{n_q} \left(\sum_{i=0}^{n_p} \sum_{j=0}^{n_p} a_{l+j-i}(\eta) p_i p_j \right) z^l,$$

where only a finite number of autocovariances $a_l(\eta), -n_p - n_q \leq l \leq n_p + n_q$ are involved.

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

$$\widehat{S}_{\eta_k}(z) = \frac{\sum_{l=-n_q}^{n_q} \left(\sum_{i=0}^{n_p} \sum_{j=0}^{n_p} a_{l+j-i,k}(\eta) p_{i,k} p_{j,k} \right) z^l}{\left(\sum_{i=0}^{n_p} p_{i,k} z^i\right)\left(\sum_{j=0}^{n_p} p_{j,k} z^{-j}\right)},$$

and by (41) the spectral density $S_{e_k}(z)$ of e_k is estimated by

$$\widehat{S}_{e_k}(z) = \frac{\left(\sum_{i=1}^{q} d_{i,k} z^i\right)\left(\sum_{j=1}^{q} d_{j,k} z^{-j}\right)}{\left(\sum_{i=0}^{p} c_{i,k} z^i\right)\left(\sum_{j=0}^{p} c_{j,k} z^{-j}\right)} \widehat{S}_{\eta_k}(z).$$

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

$$\widehat{\sigma}_k^2(e) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \widehat{S}_{e_k}(e^{\iota\omega}) \mathrm{d}\omega. \tag{54}$$

Therefore, $w_k(x)$ is estimated at time k by

$$\widehat{w}_k(x) \triangleq \frac{1}{\pi} \int_0^{\frac{1}{\widehat{b}_k}} \cos[(\widehat{\psi}_k - x)t] e^{\frac{\widehat{\sigma}_k^2(e)t^2}{2}} \mathrm{d}t, \tag{55}$$

where $\widehat{b}_k = \left(b\widehat{\sigma}_k^2(e)/\log k\right)^{1/2}$.

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

$$\mu_k(x) = \left[\mu_{k-1}(x) - \frac{1}{k}(\mu_{k-1}(x) - \widehat{w}_k(x))\right] I_{A_k}, \tag{56}$$

$$\delta_k^{(\mu)}(x) = \sum_{j=1}^{k-1} I_{A_j^c}, \tag{57}$$

$$\beta_k(x) = \left[\beta_{k-1}(x) - \frac{1}{k}(\beta_{k-1}(x) - \widehat{w}_k(x) y_k)\right] I_{A_k}, \tag{58}$$

$$\delta_k^{(\beta)}(x) = \sum_{j=1}^{k-1} I_{A_j^c}. \tag{59}$$

where A_k in (56) is $\{|\mu_{k-1}(x) - (1/k)(\mu_{k-1}(x) - \widehat{w}_k(x))| \leq M_{\delta_k^{(\mu)}(x)}\}$, while in (58) is $\{|\beta_{k-1}(x) - (1/k)(\beta_{k-1}(x) - \widehat{w}_k(x) y_k)| \leq M_{\delta_k^{(\beta)}(x)}\}$. As a matter of fact, $\mu_k(x)$ defined by (56)–(57) and $\beta_k(x)$ defined by (58)–(59) are applied to estimate $p(x)$ and $p(x)f(x)$ (see (87) and (88)), respectively, where $p(x) = \frac{1}{\sqrt{2\pi}\sigma\vartheta} e^{-\frac{x^2}{2\sigma^2\vartheta^2}}$ is the limit of the density function of v_k. The estimate for $f(x)$ is naturally defined as:

$$f_k(x) \triangleq \begin{cases} \dfrac{\beta_k(x)}{\mu_k(x)}, & \text{if } \mu_k(x) \neq 0 \\ 0, & \text{if } \mu_k(x) = 0. \end{cases} \tag{60}$$

3. Auxiliary results on weak dependence

We now proceed to prove strong consistency of the estimates given in Section 2.

Denote by $\mathcal{F}_i^j$ the σ-algebra generated by $\{X_s, \ 0 \leq i \leq s \leq j\}$ for a process $\{X_k, \ k = 0, 1, \ldots\}$, and define

$$\alpha_k \triangleq \sup_{n, A \in \mathcal{F}_0^n, B \in \mathcal{F}_{n+k}^\infty} |P(A)P(B) - P(AB)|.$$

The process $\{X_k\}$ is called α-mixing if $\alpha_k \xrightarrow[k\to\infty]{} 0$, and the numbers α_k are called the mixing coefficients of $\{X_k\}$ (Doukhan, 1994; Fan & Yao, 2003).

Lemma 2 (*Zhao & Chen, 2012a*). *Under Conditions* H1 *and* H2, $\{V_k\}$ *is an* α*-mixing, and the mixing coefficient* α_k *exponentially decays to zero:*

$$\alpha_k \leq d\lambda^k \quad \forall k \geq 1 \text{ for some } d > 0 \text{ and } 0 < \lambda < 1, \tag{61}$$

where $V_{k+1} \triangleq [v_{k+1}, \ldots, v_{k+2-p}, u_{k+1}^0, \ldots, u_{k+2-q}^0]^T$.

Remark 2. It is worth noting that the mixing property is hereditary (Fan & Yao, 2003) in the sense that the process $\{h(V_k)\}$ for any measurable function $h(\cdot)$ possesses the same mixing property as $\{V_k\}$ does. All the processes listed below are α-mixing with mixing coefficients exponentially tending to zero:

(1) $\{f(v_k)\}$ and $\{f(v_{k-j})u^0_{k-i-1}, \forall j = 0, 1, \ldots, i\}$, etc. under H1 and H2;
(2) $\{\eta_k\}$, $\{\varepsilon_k\}$, and $\{\eta_k\eta_{k-i}, i \ge 0\}$ under H3;
(3) $\{u_k\}$, $\{\psi_k\}$, $\{h(u_k)\}$, $\{h(\psi_k)\}$, and $\{h(\psi_k)\varepsilon_k\}$ with $h(\cdot)$ being any measurable function under H1–H3, because u^0_k, η_k, and ε_k are mutually independent.

Lemma 3 (Mu & Chen, 2012)**.** *Let $\{\varpi_k\}$ be a sequence of random variables with $\sup_k E|\varpi_k|^\delta < \infty$ for some $\delta \ge 2$, and let $\{l_i\}$ with $|l_i| = O(e^{-r_l i})$ for some $r_l > 0$ be a sequence of real numbers. Then, the process $X_k = \sum_{i=1}^k l_i \varpi_{k-i}$ has the bounded δ-th absolute moment: $\sup_k E|X_k|^\delta < \infty$.*

Lemma 4 (Mu & Chen, 2013; Zhao & Chen, 2012a)**.** *Let $\{X_k, \mathcal{F}_k\}$ be a zero mean α-mixing sequence with the mixing coefficients (α_k) exponentially decaying to zero. If $\sum_{k=1}^\infty (E|X_k|^{2+\epsilon})^{\frac{2}{2+\epsilon}} < \infty$ for some $\epsilon > 0$, then $\sum_{k=1}^\infty X_k < \infty$ a.s.*

4. Consistency of estimates for linear part

Lemma 5. *Assume that* H1–H4 *hold. Then, for any $0 \le \nu < 1/2$, the following series converge:*

$$\sum_{k=1}^\infty \frac{1}{k^{1-\nu}}(\tau h_i - \mathrm{E}y_k u_{k-i}) < \infty \quad \forall i \ge 1, \tag{62}$$

$$\sum_{k=1}^\infty \frac{1}{k^{1-\nu}}(\rho - \mathrm{E}(y_k - \mathrm{E}y_k)u^2_{k-1}) < \infty, \tag{63}$$

$$\sum_{k=1}^\infty \frac{1}{k^{1-\nu}}(\rho h_i - \mathrm{E}(y_k - \mathrm{E}y_k)u_{k-1}u_{k-i}) < \infty. \quad \forall i \ge 2. \tag{64}$$

The proof is based on the fact $|\sigma^2 - \sigma^2_k| = O(e^{-rk})$ for some $r > 0$. For details we refer to Lemma 4.3 in Mu and Chen (2012).

Lemma 6. *Assume* H1–H4 *hold. Then, λ_k defined by (22)–(23) has the following convergence rate:*

$$|\lambda_k - \mathrm{E}y_k| = o\left(\frac{1}{k^{1/2-c}}\right) \quad \forall c > 0. \tag{65}$$

Proof. By (4) and H3 we see

$$\mathrm{E}y_k = \mathrm{E}y^0_k = \mathrm{E}f(v_k)$$
$$\xrightarrow[k\to\infty]{} \frac{1}{\sqrt{2\pi}\sigma\vartheta}\int_{\mathbb{R}} f(x)e^{-\frac{x^2}{2\sigma^2\vartheta^2}}\mathrm{d}x \triangleq \bar{\lambda}, \tag{66}$$

where $\sigma^2 = \sum_{i=1}^\infty h_i^2$.

The algorithm (22) can be written as

$$\lambda_k = \left[\lambda_{k-1} - (1/k)(\lambda_{k-1} - \bar{\lambda}) - (1/k)e^{(\lambda)}_k\right] I_{A_k}, \tag{67}$$

where

$$e^{(\lambda)}_k = \bar{\lambda} - y_k = (\bar{\lambda} - \mathrm{E}y^0_k) + (\mathrm{E}y^0_k - y^0_k) - \varepsilon_k. \tag{68}$$

Since $\bar{\lambda}$ is the single root of the linear function $-(y-\bar{\lambda})$, by Theorem 3.1.1 (Chen, 2002), for proving $|\lambda_k - \bar{\lambda}| = o\left(\frac{1}{k^{1/2-c}}\right)$, it suffices to show

$$\sum_{k=1}^\infty \frac{1}{k^{1-\nu}} e^{(\lambda)}_k < \infty \quad a.s.\ \forall 0 < \nu < 1/2. \tag{69}$$

Since $|\sigma^2 - \sigma^2_k| = O(e^{-rk})$ for some $r > 0$, we have $|\bar{\lambda} - \mathrm{E}y^0_k| = O(e^{-r_\lambda k})$ for some $r_\lambda > 0$. Thus, (69) holds for the first term on the right-hand side of (68).

By Lemma 2 and Remark 2, we see that both $\{\mathrm{E}y^0_k - y^0_k\}$ and $\{\varepsilon_k\}$ are the zero mean α-mixing sequences with mixing coefficients decaying exponentially to zero. Further, by Lemma 3, we have $\mathrm{E}|y^0_k|^{2+\epsilon} < \infty$ and $\mathrm{E}|\varepsilon_k|^{2+\epsilon} < \infty$ for some $\epsilon > 0$. Thus, by Lemma 4, (69) holds for the last two terms on the right-hand side of (68).

Since $|\lambda_k - \mathrm{E}y_k| \le |\lambda_k - \bar{\lambda}| + |\bar{\lambda} - \mathrm{E}y_k|$ and $|\bar{\lambda} - \mathrm{E}y_k| = O(e^{-r_\lambda k})$ for some $r_\lambda > 0$, we have

$$|\lambda_k - \mathrm{E}y_k| = o\left(\frac{1}{k^{1/2-c}}\right) \quad \forall c > 0. \quad \square$$

Lemma 7. *Assume that* H1–H4 *hold. Then, for any $0 \le \nu < 1/2$, the following series converge:*

$$\sum_{k=1}^\infty \frac{1}{k^{1-\nu}}(\mathrm{E}y_k u_{k-i} - y_k u_{k-i}) < \infty \quad a.s.\ \forall i \ge 1, \tag{70}$$

$$\sum_{k=1}^\infty \frac{1}{k^{1-\nu}}(\mathrm{E}(y_k - \mathrm{E}y_k)u^2_{k-1} - (y_k - \mathrm{E}y_k)u^2_{k-1}) < \infty \quad a.s., \tag{71}$$

$$\begin{aligned}\sum_{k=1}^\infty \frac{1}{k^{1-\nu}}(&\mathrm{E}(y_k - \mathrm{E}y_k)u_{k-1}u_{k-i}\\ &- (y_k - \mathrm{E}y_k)u_{k-1}u_{k-i}) < \infty \quad a.s.\ \forall i \ge 2,\end{aligned} \tag{72}$$

$$\sum_{k=1}^\infty \frac{1}{k^{1-\nu}}((\lambda_k - \mathrm{E}y_k)u^2_{k-1}) < \infty \quad a.s., \tag{73}$$

$$\sum_{k=1}^\infty \frac{1}{k^{1-\nu}}((\lambda_k - \mathrm{E}y_k)u_{k-1}u_{k-i}) < \infty \quad a.s.\ \forall i \ge 2. \tag{74}$$

Proof. It is noticed that u^0_k, η_k and ε_k are mutually independent, we have

$$\begin{aligned}&\mathrm{E}(y_k - \mathrm{E}y_k)u^2_{k-1} - (y_k - \mathrm{E}y_k)u^2_{k-1}\\ &= \left[\mathrm{E}(y^0_k - \mathrm{E}y^0_k)(u^0_{k-1})^2 - (y^0_k - \mathrm{E}y^0_k)(u^0_{k-1})^2\right] - \varepsilon_k u^2_{k-1}\\ &\quad - (y^0_k - \mathrm{E}y^0_k)\eta^2_{k-1} - 2(y^0_k - \mathrm{E}y^0_k)u^0_{k-1}\eta_{k-1}.\end{aligned} \tag{75}$$

It follows from (75) that

$$\begin{aligned}&\sum_{k=1}^\infty \frac{1}{k^{1-\nu}}\left[\mathrm{E}(y_k - \mathrm{E}y_k)u^2_{k-1} - (y_k - \mathrm{E}y_k)u^2_{k-1}\right]\\ &= \sum_{k=1}^\infty \frac{1}{k^{1-\nu}}\left[\mathrm{E}(y^0_k - \mathrm{E}y^0_k)(u^0_{k-1})^2 - (y^0_k - \mathrm{E}y^0_k)(u^0_{k-1})^2\right]\\ &\quad - \sum_{k=1}^\infty \frac{1}{k^{1-\nu}}\left[(y^0_k - \mathrm{E}y^0_k)\eta^2_{k-1}\right] - \sum_{k=1}^\infty \frac{1}{k^{1-\nu}}\left[\varepsilon_k u^2_{k-1}\right]\\ &\quad - 2\sum_{k=1}^\infty \frac{1}{k^{1-\nu}}\left[(y^0_k - \mathrm{E}y^0_k)u^0_{k-1}\eta_{k-1}\right].\end{aligned} \tag{76}$$

Define $z^{(1)}_k \triangleq \frac{1}{k^{1-\nu}}\mathrm{E}(y^0_k - \mathrm{E}y^0_k)(u^0_{k-1})^2 - (y^0_k - \mathrm{E}y^0_k)(u^0_{k-1})^2$. Thus, by Lemma 2, $z^{(1)}_k$ is a zero mean α-mixing sequence with the mixing coefficient decaying exponentially to zero. By Lemma 3, Cauchy–Schwarz and C_r inequalities, we have

$$\sum_{k=1}^{\infty}\left(\mathrm{E}|z_k^{(1)}|^{2+\epsilon}\right)^{\frac{2}{2+\epsilon}} \le \sum_{k=1}^{\infty}\frac{4}{k^{2(1-\nu)}}$$

$$\left(\mathrm{E}|(y_k^0-\mathrm{E}y_k^0)(u_{k-1}^0)^2|^{2+\epsilon}\right)^{\frac{2}{2+\epsilon}} \le O\left(\sum_{k=1}^{\infty}\frac{1}{k^{2(1-\nu)}}\right)<\infty.$$

Therefore, by Lemma 4, the first term on the right-hand side of (76) converges a.s. The convergence of the remaining terms on the right-hand side of (76) can be proved in a similar way, and hence (71) holds. Similarly, the assertions (70) and (72) also hold.

According to (65), we have

$$\left|\sum_{k=1}^{\infty}\frac{1}{k^{1-\nu}}((\lambda_k-\mathrm{E}y_k)u_{k-1}^2)\right|$$
$$\le \sum_{k=1}^{\infty}\frac{1}{k^{\frac{3}{2}-\nu-c}}\cdot(u_{k-1}^2-\mathrm{E}u_{k-1}^2)+\sum_{k=1}^{\infty}\frac{1}{k^{\frac{3}{2}-\nu-c}}\mathrm{E}u_{k-1}^2. \quad (77)$$

Since $\sum_{k=1}^{\infty}\frac{1}{k^{\frac{3}{2}-\nu-c}}\mathrm{E}u_{k-1}^2<\infty$, for proving $\sum_{k=1}^{\infty}\frac{1}{k^{1-\nu}}((\lambda_k-\mathrm{E}y_k)u_{k-1}^2)<\infty$ it suffices to show the first term on the right-hand side of (77) converges a.s.

By Remark 2, $z_k^{(2)} \triangleq \frac{1}{k^{\frac{3}{2}-\nu-c}}(u_{k-1}^2-\mathrm{E}u_{k-1}^2)$ is a zero mean α-mixing sequence with the mixing coefficient decaying exponentially to zero. Noticing that $\mathrm{E}|u_{k-1}^2|^{2+\epsilon}<\infty$, by the C_r inequality we have

$$\sum_{k=1}^{\infty}\left(\mathrm{E}|z_k^{(2)}|^{2+\epsilon}\right)^{\frac{2}{2+\epsilon}} \le \sum_{k=1}^{\infty}\frac{4}{k^{3-2\nu-2c}}\left(\mathrm{E}|u_{k-1}^2|^{2+\epsilon}\right)^{\frac{2}{2+\epsilon}}$$
$$=O\left(\sum_{k=1}^{\infty}\frac{1}{k^{3-2\nu-2c}}\right)<\infty.$$

Therefore, by Lemma 4, the assertion (73) holds. Similarly, (74) is also true. □

Theorem 1. *Assume that* H1–H4 *hold. Then,* $h_{i,k}$ *defined by* (30) *and* (33) *converges to* h_i $\forall i\ge 2$ *with the rate of convergence*

$$|h_{i,k}-h_i|=o(k^{-\nu}) \quad a.s. \ \forall\ \nu\in(0,1/2),\ \ i\ge 2. \quad (78)$$

As consequences, from (39)–(40) *the following convergence rates also take place:* $\forall\ \nu\in(0,1/2)$,

$$|c_{i,k}-c_i|=o(k^{-\nu}) \quad a.s.\ 1\le i\le p, \quad (79)$$
$$|d_{j,k}-d_j|=o(k^{-\nu}) \quad a.s.\ 2\le j\le q. \quad (80)$$

Proof. As pointed out before, by H4 at least one of τ and ρ is nonzero, so switching between (28)–(30) and (31)–(33) may happen only a finite number of times. Therefore, for proving (78) it suffices to show

$$|\theta_k^{(i,\tau)}-\tau h_i|=o(k^{-\nu}) \quad a.s.\ \forall\nu\in(0,1/2)\ i\ge 1, \quad (81)$$
$$|\theta_k^{(1,\rho)}-\rho|=o(k^{-\nu}) \quad a.s.\ \forall\nu\in(0,1/2), \quad (82)$$
$$|\theta_k^{(i,\rho)}-\rho h_i|=o(k^{-\nu}) \quad a.s.\ \forall\nu\in(0,1/2)\ i\ge 2. \quad (83)$$

We rewrite (26) as

$$\theta_k^{(1,\rho)}=\left[\theta_{k-1}^{(1,\rho)}-(1/k)(\theta_{k-1}^{(1,\rho)}-\rho)-(1/k)e_k^{(1,\rho)}\right]\cdot I_{A_k},$$

where

$$e_k^{(1,\rho)}=\rho-\rho_k+(\mathrm{E}(y_k-\mathrm{E}y_k)u_{k-1}^2-(y_k-\mathrm{E}y_k)u_{k-1}^2)$$
$$+(\lambda_k-\mathrm{E}y_k)u_{k-1}^2. \quad (84)$$

For (82), similar to (67), it suffices to prove

$$\sum_{k=1}^{\infty}\frac{1}{k^{1-\nu}}e_k^{(1,\rho)}<\infty \quad a.s.\ \forall\nu\in(0,1/2). \quad (85)$$

By (63), (71) and (73) we find that (85) is true for (84), and hence (82) holds. Similarly, (81) and (83) can be proved by Lemmas 5 and 7, while the assertions (79)–(80) straightforwardly follow from (78). □

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

Lemma 8 (Hu & Chen, 2006; Mu & Chen, 2012)**.** *Assume that* H1–H4 *and* H6 *hold. Then the following limits take place*

$$\frac{u_k}{k^c}\xrightarrow[k\to\infty]{a.s.}0,\qquad \frac{f(v_k)}{k^c}\xrightarrow[k\to\infty]{a.s.}0\quad\forall c>0. \quad (86)$$

The lemma can be proved by the same treatment as that used in Lemma 4 of Hu and Chen (2006) or Lemma 4.8 of Mu and Chen (2012).

Lemma 9. *Under Conditions* H1–H6*, the following limits and assertions for* $w_k(x)$ *defined by* (48) *take place*

$$\mathrm{E}[w_k(x)]\xrightarrow[k\to\infty]{}p(x), \quad (87)$$

$$\mathrm{E}[w_k(x)f(v_k)]\xrightarrow[k\to\infty]{}p(x)\widetilde{f}(x), \quad (88)$$

$$|w_k(x)|^{\delta}=O\left(k^{\frac{\delta}{2b}}(\log k)^{\frac{\delta}{2}}\right)\quad\forall\delta\ge 1, \quad (89)$$

$$|w_k(x)f(v_k)|^{\delta}=O\left(k^{\frac{\delta}{2b}+c}(\log k)^{\frac{\delta}{2}}\right)\quad\forall\delta\ge 1\ c>0, \quad (90)$$

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

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

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

Proof. By the Fubini theorem, and noticing that the density function of e_k is even, we have

$$\mathrm{E}[w_k(x)f(v_k)]$$
$$=\frac{1}{2\pi b_k}\int_{\mathbb{R}}\mathrm{E}\left(e^{[-\iota t(\psi_k-x)/b_k]}f(v_k)\right)\frac{\Phi_K(t)}{\Phi_{e_k}(t/b_k)}\mathrm{d}t$$
$$=\frac{1}{2\pi b_k}\int_{\mathbb{R}}\mathrm{E}\left(e^{[-\iota t(v_k-x)/b_k]}f(v_k)\right)\Phi_K(t)\mathrm{d}t$$
$$=\frac{1}{b_k}\int_{\mathbb{R}}\left(\frac{1}{2\pi}\int_{\mathbb{R}}e^{[-\iota t(y-x)/b_k]}\Phi_K(t)\mathrm{d}t\right)$$
$$\cdot f(y)\frac{1}{\sqrt{2\pi}\sigma_k\vartheta}e^{-\frac{y^2}{2\sigma_k^2\vartheta^2}}\mathrm{d}y$$
$$=\frac{1}{\sqrt{2\pi}\sigma_k\vartheta}e^{-\frac{x^2}{2\sigma_k^2\vartheta^2}}\int_{-\infty}^{x}K(t)f(x+b_kt)e^{-\frac{2xb_kt+b_k^2t^2}{2\sigma_k^2\vartheta^2}}\mathrm{d}t$$

$$+\frac{1}{\sqrt{2\pi}\sigma_k\vartheta}e^{-\frac{x^2}{2\sigma_k^2\vartheta^2}}\int_x^\infty K(t)f(x+b_kt)e^{-\frac{2xb_kt+b_k^2t^2}{2\sigma_k^2\vartheta^2}}\mathrm{d}t$$

$$\xrightarrow[k\to\infty]{} p(x)\left(f(x^-)\int_{-\infty}^x K(t)\mathrm{d}t+f(x^+)\int_x^\infty K(t)\mathrm{d}t\right),$$

while the limit (87) can be proved in a similar treatment.

By (48), we have

$$\begin{aligned}|w_k(x)|^\delta &\le \frac{1}{\pi^\delta}\left|\int_0^{\frac{1}{b_k}}e^{\frac{\sigma_k^2(e)t^2}{2}}\mathrm{d}t\right|^\delta\\ &\le \frac{1}{\pi^\delta}\left[\frac{1}{2}\left(1+e^{\frac{\sigma_k^2(e)}{2}\frac{\log k}{b\sigma_k^2(e)}}\right)\left(\frac{\log k}{b\sigma_k^2(e)}\right)^{\frac{1}{2}}\right]^\delta\\ &\le \frac{(\log k)^{\frac{\delta}{2}}}{(2\pi\sigma_k(e))^\delta b^{\delta/2}}k^{\frac{\delta}{2b}}=O\left(k^{\frac{\delta}{2b}}(\log k)^{\frac{\delta}{2}}\right).\end{aligned}\tag{91}$$

Similarly, the assertion (90) can be proved by noticing the second limit of (86). □

Lemma 10. *Assume that* H1, H3, H5 *and* H6 *hold. Then both* $a_{0,k}(\eta)$ *defined by* (49) *and* (50) *and* $a_{i,k}(\eta), i\ge 1$ *defined by* (51) *and* (52) *have the convergence rate*

$$|a_{i,k}(\eta)-a_i(\eta)|=o\left(\frac{1}{k^{1/2-c}}\right)\quad\forall c>0,\ i\ge 0.\tag{92}$$

The proof of the lemma is similar to that for Lemma 6.

Corollary 1. *Assume* H1–H6 *hold. Then* $\widehat{\sigma}_k^2(e)$ *defined by* (54) *has the following convergence rate:*

$$|\widehat{\sigma}_k^2(e)-\sigma_k^2(e)|=o\left(\frac{1}{k^{1/2-c}}\right)\quad\forall c>0,\tag{93}$$

Lemma 11. *Assume* H1–H6 *hold. Then there is a constant* $c>0$ *with* $\frac{1}{2}-\frac{1}{2b}-2c>0$ *such that*

$$|\psi_k-\widehat{\psi}_k|=o\left(\frac{1}{k^{\frac{1}{2}-2c}}\right),\tag{94}$$

$$|w_k(x)-\widehat{w}_k(x)|=o\left(\frac{(\log k)^{\frac{3}{2}}}{k^{\frac{1}{2}\left(1-\frac{1}{b}\right)-2c}}\right).\tag{95}$$

Proof. For (94) we refer to Theorem 2 in Hu and Chen (2006) or Lemma 4.10 in Mu and Chen (2012).

According to (48) and (55), we have

$$w_k(x)-\widehat{w}_k(x)=I_1+I_2+I_3,$$

where

$$I_1=\frac{1}{\pi}\int_{\frac{1}{\widehat{b}_k}}^{\frac{1}{b_k}}\cos[(\psi_k-x)t]e^{\frac{\sigma_k^2(e)t^2}{2}}\mathrm{d}t,\tag{96}$$

$$I_2=\frac{1}{\pi}\int_0^{\frac{1}{\widehat{b}_k}}\cos[(\psi_k-x)t]\left(e^{\frac{\sigma_k^2(e)t^2}{2}}-e^{\frac{\widehat{\sigma}_k^2(e)t^2}{2}}\right)\mathrm{d}t,\tag{97}$$

$$I_3=\frac{1}{\pi}\int_0^{\frac{1}{\widehat{b}_k}}(\cos[(\psi_k-x)t]-\cos[(\widehat{\psi}_k-x)t])e^{\frac{\widehat{\sigma}_k^2(e)t^2}{2}}\mathrm{d}t.\tag{98}$$

Since $|\widehat{\sigma}_k^2(e)-\sigma_k^2(e)|=o\left(\frac{1}{k^{1/2-c}}\right)$ and $k^{\sigma_k^2(e)/(2b\widehat{\sigma}_k^2(e))}=o(k^{\frac{1}{2b}+c})$ for any $c>0$, we have

$$\begin{aligned}|I_1| &\le \frac{1}{2\pi}\left(e^{\frac{\sigma_k^2(e)}{2}\frac{1}{b_k^2}}+e^{\frac{\sigma_k^2(e)}{2}\frac{1}{\widehat{b}_k^2}}\right)\left|\frac{1}{b_k}-\frac{1}{\widehat{b}_k}\right|\\ &\le \frac{1}{2\pi}\left(e^{\frac{\sigma_k^2(e)}{2}\frac{\log k}{b\sigma_k^2(e)}}+e^{\frac{\sigma_k^2(e)}{2}\frac{\log k}{b\widehat{\sigma}_k^2(e)}}\right)\left(\frac{\log k}{b}\right)^{\frac{1}{2}}\\ &\quad\cdot\frac{|\widehat{\sigma}_k^2(e)-\sigma_k^2(e)|}{\widehat{\sigma}_k(e)\sigma_k(e)(\widehat{\sigma}_k(e)+\sigma_k(e))}\\ &\le \frac{1}{2\pi}\left(k^{\frac{1}{2b}}+k^{\frac{\sigma_k^2(e)}{2b\widehat{\sigma}_k^2(e)}}\right)o\left(\frac{(\log k)^{\frac{1}{2}}}{k^{\frac{1}{2}-c}}\right)\\ &= o\left(\frac{(\log k)^{\frac{1}{2}}}{k^{\frac{1}{2}\left(1-\frac{1}{b}\right)-2c}}\right).\end{aligned}\tag{99}$$

By the mean value theorem, there is an $\bar{s}\in(\widehat{\sigma}_k^2(e),\sigma_k^2(e))$ or $\bar{s}\in(\sigma_k^2(e),\widehat{\sigma}_k^2(e))$ such that

$$\left|e^{\frac{\sigma_k^2(e)t^2}{2}}-e^{\frac{\widehat{\sigma}_k^2(e)t^2}{2}}\right|=\frac{t^2}{2}e^{\frac{\bar{s}t^2}{2}}|\sigma_k^2(e)-\widehat{\sigma}_k^2(e)|.$$

Again by $|\widehat{\sigma}_k^2(e)-\sigma_k^2(e)|=o\left(\frac{1}{k^{1/2-c}}\right)$ and $k^{\bar{s}/(2b\widehat{\sigma}_k^2(e))}=o(k^{\frac{1}{2b}+c})$ for any $c>0$, we have

$$\begin{aligned}|I_2| &= \frac{1}{\pi}\int_0^{\frac{1}{\widehat{b}_k}}\frac{t^2}{2}e^{\frac{\bar{s}t^2}{2}}dt|\sigma_k^2(e)-\widehat{\sigma}_k^2(e)|\\ &\le \frac{1}{2\pi}\left(\frac{\log k}{2b\widehat{\sigma}_k^2(e)}e^{\frac{\bar{s}}{2}\frac{\log k}{b\widehat{\sigma}_k^2(e)}}\right)\left(\frac{\log k}{b\widehat{\sigma}_k^2(e)}\right)^{\frac{1}{2}}|\sigma_k^2(e)-\widehat{\sigma}_k^2(e)|\\ &= \frac{1}{4\pi}\left(\frac{\log k}{b\widehat{\sigma}_k^2(e)}\right)^{\frac{3}{2}}k^{\frac{\bar{s}}{2b\widehat{\sigma}_k^2(e)}}|\sigma_k^2(e)-\widehat{\sigma}_k^2(e)|\\ &= o\left(\frac{(\log k)^{\frac{3}{2}}}{k^{\frac{1}{2}\left(1-\frac{1}{b}\right)-2c}}\right).\end{aligned}\tag{100}$$

From (94) it follows that

$$\begin{aligned}|I_3| &\le \frac{1}{\pi}\int_0^{\frac{1}{\widehat{b}_k}}\left|-2\sin\left(\frac{(\psi_k+\widehat{\psi}_k)t-2xt}{2}\right)\right.\\ &\quad\cdot\left.\sin\left(\frac{(\psi_k-\widehat{\psi}_k)t}{2}\right)\right|e^{\frac{\widehat{\sigma}_k^2(e)t^2}{2}}\mathrm{d}t\\ &\le \frac{1}{\pi}\int_0^{\frac{1}{\widehat{b}_k}}te^{\frac{\widehat{\sigma}_k^2(e)t^2}{2}}\mathrm{d}t\,|\psi_k-\widehat{\psi}_k|\\ &= \frac{1}{\pi\widehat{\sigma}_k^2(e)}(k^{\frac{1}{2b}}-1)\,|\psi_k-\widehat{\psi}_k|=o\left(\frac{1}{k^{\frac{1}{2}-\frac{1}{2b}-2c}}\right).\end{aligned}\tag{101}$$

By (99)–(101), we have

$$|w_k(x)-\widehat{w}_k(x)|=o\left(\frac{(\log k)^{\frac{3}{2}}}{k^{\frac{1}{2}\left(1-\frac{1}{b}\right)-2c}}\right).\quad\square$$

Lemma 12. *Assume* H1–H6 *hold. The following series converge a.s.*

$$\sum_{k=1}^{\infty} \frac{1}{k} (\mathrm{E}w_k(x) - w_k(x)) < \infty, \tag{102}$$

$$\sum_{k=1}^{\infty} \frac{1}{k} (\mathrm{E}w_k(x)f(v_k) - w_k(x)f(v_k)) < \infty, \tag{103}$$

$$\sum_{k=1}^{\infty} \frac{1}{k} (w_k(x) - \widehat{w}_k(x))\varepsilon_k < \infty, \qquad \sum_{k=1}^{\infty} \frac{1}{k} w_k(x)\varepsilon_k < \infty. \tag{104}$$

Proof. By Remark 2, $z_k^{(3)} \triangleq \frac{1}{k}(\mathrm{E}w_k(x) - w_k(x))$ is a zero mean α-mixing sequence with mixing coefficients decaying exponentially to zero. Noticing $\mathrm{E}|w_k(x)|^{2+\epsilon} = O\left(k^{\frac{2+\epsilon}{2b}} (\log k)^{\frac{2+\epsilon}{2}}\right)$ by (89), and by the C_r inequality we have

$$\sum_{k=1}^{\infty} \left(\mathrm{E}|z_k^{(3)}|^{2+\epsilon}\right)^{\frac{2}{2+\epsilon}} = O\left(\sum_{k=1}^{\infty} \frac{\log k}{k^{2-\frac{1}{b}}}\right) < \infty.$$

Therefore, by Lemma 4 we have proved (102), while (103) can be verified in a similar way.

The convergence of the first series in (104) can be proved by the treatment similar to that used for proving (73).

By Remark 2, $z_k^{(4)} \triangleq \frac{1}{k} w_k(x)\varepsilon_k$ is a zero mean α-mixing sequence with mixing coefficients decaying exponentially to zero. Noticing $\mathrm{E}|w_k(x)|^{2+\epsilon} = O\left(k^{\frac{2+\epsilon}{2b}} (\log k)^{\frac{2+\epsilon}{2}}\right)$ by (89), and $\mathrm{E}|\varepsilon_k|^{2+\epsilon} < \infty$ by Lemma 3, we have

$$\begin{aligned}\sum_{k=1}^{\infty} \left(\mathrm{E}|z_k^{(4)}|^{2+\epsilon}\right)^{\frac{2}{2+\epsilon}} &\le \sum_{k=1}^{\infty} \frac{1}{k^2} \left(\mathrm{E}|w_k(x)|^{2+\epsilon}\right)^{\frac{2}{2+\epsilon}} \cdot \left(\mathrm{E}|\varepsilon_k|^{2+\epsilon}\right)^{\frac{2}{2+\epsilon}} \\ &= O\left(\sum_{k=1}^{\infty} \frac{\log k}{k^{2-\frac{1}{b}}}\right) < \infty.\end{aligned}$$

Therefore, by Lemma 4 we have proved the convergence of the last series of (104). $\square$

Theorem 2. *Assume* H1–H6 *hold. Then*

$$\mu_k(x) \xrightarrow[k\to\infty]{} p(x) \quad a.s., \tag{105}$$

$$\beta_k(x) \xrightarrow[k\to\infty]{} p(x)\widetilde{f}(x) \quad a.s., \tag{106}$$

$$f_k(x) \xrightarrow[k\to\infty]{} \widetilde{f}(x) \quad a.s., \tag{107}$$

where $\mu_k(x)$, $\beta_k(x)$, and $f_k(x)$ are defined by (56)–(60), respectively.

Proof. The algorithm (56) can be rewritten as

$$\mu_k(x) = \left[\mu_{k-1}(x) - \frac{1}{k}(\mu_{k-1}(x) - p(x)) - \frac{1}{k}\bar{e}_k(x)\right] I_{A_k},$$

where

$$\begin{aligned}\bar{e}_k(x) &= p(x) - \widehat{w}_k(x) = [p(x) - \mathrm{E}w_k(x)] \\ &\quad + [\mathrm{E}w_k(x) - w_k(x)] + [w_k(x) - \widehat{w}_k(x)]\,. \end{aligned} \tag{108}$$

Noticing that $\sum_{k=1}^{\infty} \frac{1}{k}\left((\mathrm{E}w_k(x) - w_k(x)) + (w_k(x) - \widehat{w}_k(x))\right) < \infty$ *a.s.* by (102) and (95), and $\mathrm{E}w_k(x) \xrightarrow[k\to\infty]{} p(x)$ by (87), we conclude $\mu_k(x) \xrightarrow[k\to\infty]{} p(x)$ *a.s.* by Theorem 2.1.1 in Chen (2002).

The proof of (106) can similarly be carried out, if we rewrite the algorithm (58) as follows:

$$\beta_k(x) = \left[\beta_{k-1}(x) - \frac{1}{k}(\beta_{k-1}(x) - p(x)\widetilde{f}(x)) - \frac{1}{k}\tilde{e}_k(x)\right] I_{A_k},$$

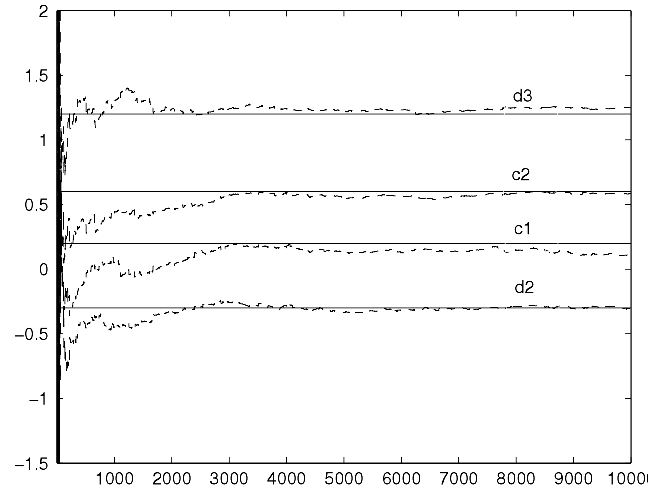

Fig. 2. Estimates for c_1, c_2, d_2, d_3.

where

$$\begin{aligned}\tilde{e}_k(x) &= \left(p(x)\widetilde{f}(x) - \mathrm{E}w_k(x)f(v_k)\right) \\ &\quad + (\mathrm{E}w_k(x)f(v_k) - w_k(x)f(v_k)) \\ &\quad - (\widehat{w}_k(x) - w_k(x))(f(v_k) + \varepsilon_k) - w_k(x)\varepsilon_k. \end{aligned} \tag{109}$$

Each term on the right-hand side of (109) satisfies the convergence condition of SAAWET by noticing (88), (103), (95) and (104). So, the estimate (60) is strongly consistent. $\square$

6. Example

Let the nonlinear function and the linear subsystem be such that

$$f(x) = x^2 - 0.5x - 1$$
$$v_k + c_1 v_{k-1} + c_2 v_{k-2} = u_{k-1}^0 + d_2 u_{k-2}^0 + d_3 u_{k-3}^0,$$

where $c_1 = 0.2$, $c_2 = 0.6$, $d_2 = -0.3$ and $d_3 = 1.2$.

Let the input $\{u_k^0\}$, the driven noises $\{\zeta_k\}$ and $\{\varsigma_k\}$ be mutually independent and Gaussian: $u_k^0 \in \mathcal{N}(0, 1)$, $\zeta_k \in \mathcal{N}(0, 0.3^2)$, and $\varsigma_k \in \mathcal{N}(0, 0.3^2)$. The measurement noises η_k and ε_k are ARMA processes:

$$\eta_k - 0.7\eta_{k-1} = \zeta_k + 0.5\zeta_{k-1},$$
$$\varepsilon_k + 0.4\varepsilon_{k-1} = \varsigma_k - 0.6\varsigma_{k-1}.$$

The parameters used in the algorithms are as follows: $b = 4$ and $M_k = 2^k + 10$.

For parameter estimation, the solid lines are the true values, while the dashed lines denote the corresponding estimates. Fig. 2 demonstrates the estimates for coefficients of the linear subsystem, while Fig. 3 gives the performance of the estimate for $\sigma_k^2(e)$. In Fig. 4 the true nonlinear function is denoted by the solid curve, and its estimates at 31 points equally chosen from the interval $[-3, 3]$ are shown by symbols $+$. The behavior of the estimates at points $\{-2.4, -2, -0.2, 1.8\}$ versus time is shown by Fig. 5.

7. Conclusion

The recursive estimation for identifying EIV Wiener systems is proposed in the paper. The estimation is carried out by the SAAWET incorporated with the deconvolution kernel. The estimates for the linear subsystem 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 systems or more complicated EIV Wiener–Hammerstein systems.

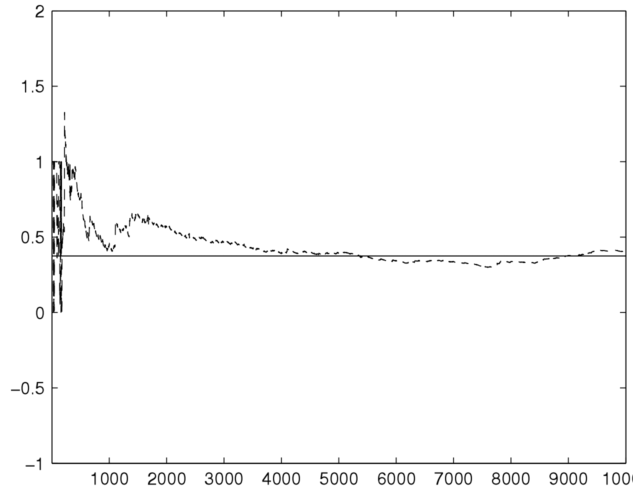

Fig. 3. Estimates for $\sigma_k^2(e)$.

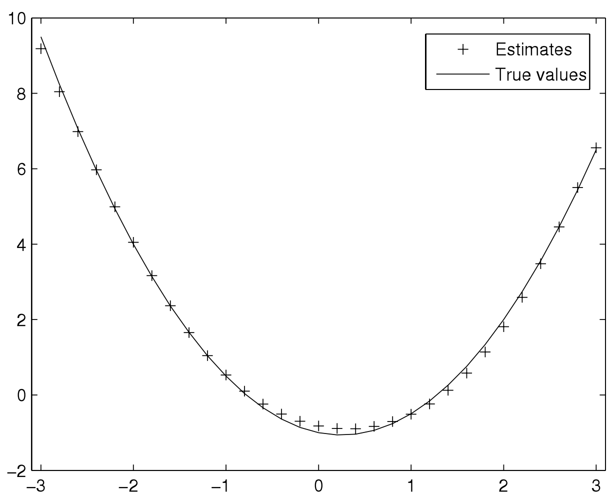

Fig. 4. Estimates for $f(x) = x^2 - 0.5x - 1$.

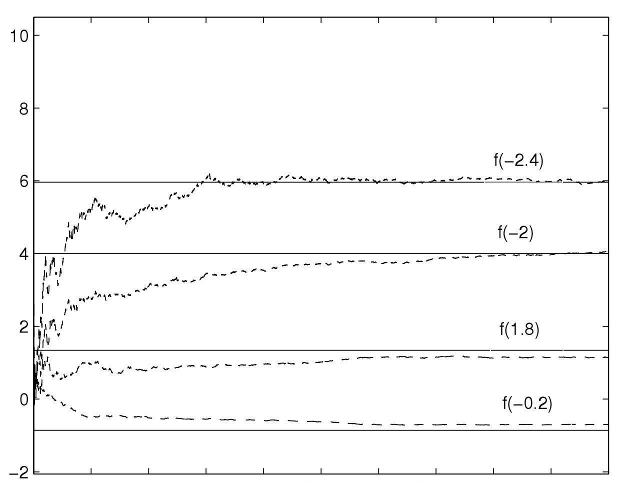

Fig. 5. Estimates for $f(x)$ at some fixed points.

References

Agüero, J. C., & Goodwin, G. C. (2008). Identifiability of errors in variables dynamic systems. *Automatica*, *44*, 371–382.

Bai, E. W. (2003). Frequency domain identification of Wiener models. *Automatica*, *39*, 1521–1530.

Boyd, S., & Chua, L. O. (1985). Fading memory and the problem of approximating nonlinear operators with volterra series. *IEEE Transactions on Circuits and Systems*, *32*, 1150–1161.

Chen, H. F. (2002). *Stochastic approximation and its applications*. Dordrecht, The Netherlands: Kluwer.

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

Chen, H. F., & Zhao, W. X. (2010). New method of order estimation for arma/armax processes. *SIAM Journal on Control and Optimization*, *48*, 4157–4176.

Davis, K. B. (1975). Mean square error properties of density estimates. *The Annals of Statistics*, *3*, 1025–1030.

Doukhan, P. (1994). *Mixing: properties and examples*. New York: Springer-Verlag.

Fan, J., & Truong, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics*, *21*, 1900–1925.

Fan, J., & Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric approach*. New York: Springer-Verlag.

Greblicki, W., & Pawlak, M. (2008). *Nonparametric system identification*. New York: Cambridge University Press.

Hagenblad, A., Ljung, L., & Wills, A. (2008). Maximum likelihood identification of Wiener models. *Automatica*, *44*, 2697–2705.

Hu, X. L., & Chen, H. F. (2006). Identification for Wiener systems with rtf subsystems. *European Journal of Control*, *12*, 581–594.

Hunter, I. W., & Korenberg, M. J. (1986). The identification of nonlinear biological system: Wiener and Hammerstein cascade models. *Biological Cybernetics*, *55*, 135–144.

Kalafatis, A., Arifin, N., Wang, L., & Cluett, W. R. (1995). A new approach to identification of ph process based on the Wiener model. *Chemical Engineering Science*, *50*, 3693–3701.

Kang, H. W., Cho, Y. S., & Youn, D. H. (1999). On compensating nonlinear distortions of an ofdm system using an efficient adaptive predistorter. *IEEE Transactions on Communications*, *47*, 522–526.

Mu, B. Q., & Chen, H. F. (2012). Recursive identification of Wiener–Hammerstein systems. *SIAM Journal on Control and Optimization*, *50*, 2621–2658.

Mu, B. Q., & Chen, H. F. (2013). Recursive identification of mimo Wiener systems. *IEEE Transactions on Automatic Control*, *58*, 802–808.

Söderström, T. (2007). Errors-in-variables methods in system identification. *Automatica*, *43*, 939–958.

Song, Q. J., & Chen, H. F. (2008). Identification of errors-in-variables systems with arma observation noise. *Systems & Control Letters*, *57*, 420–424.

Stefanski, L., & Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics*, *21*, 169–184.

Stoica, P. (1983). Generalized Yule–Walker equations and testing the orders of multivariate time series. *International Journal of Control*, *37*, 1159–1166.

Wigren, T. (1994). Convergence analysis of recursive identification algorithm based on the nonlinear Wiener model. *IEEE Transactions on Automatic Control*, *39*, 2191–2206.

Zhao, W. X., & Chen, H. F. (2012a). Markov chain approach to identifying Wiener systems. *Science China: Information Sciences*, *55*, 1201–1217.

Zhao, W.X., & Chen, H.F. (2012b). Stochastic approximation based pca and its application to identification of eiv systems. In *Proceeding of the 10-th world congress on intelligent control and automation*, Beijing (pp. 3276–3280).

Zhu, Y. (1999). Distillation column identification for control using Wiener model. In *Proceedings of American control conference*, San Diego, California (pp. 3462–3466).

Bi-Qiang Mu was born in Sichuan, China, 1986. He received the B.E. degree from Sichuan University in 2008, and defended his Ph.D. thesis in May 2013 at the Institute of Systems Science, AMSS, Chinese Academy of Sciences. His research interests are in the recursive identification for nonlinear systems such as Wiener systems, Hammerstein systems, Wiener–Hammerstein systems, and the errors-in-variables problems for these systems.

Han-Fu Chen graduated from Leningrad (St. Petersburg) State University, Russia. Since 1979 he has been with the Institute of Systems Science, which now is a part of the Academy of Mathematics and Systems Science, Chinese Academy of Sciences (CAS). He is a Professor of the Key Laboratory of Systems and Control of CAS. His research interests are mainly in stochastic systems, including system identification, adaptive control, and stochastic approximation and its applications. He has authored and coauthored more than 200 journal papers and seven books.

He served as an IFAC Council Member (2002–2005), President of the Chinese Association of Automation (1993–2002), and a Permanent member of the Council of the Chinese Mathematics Society (1991–1999).

He is an IEEE Fellow, IFAC Fellow, a Member of TWAS, and a Member of the Chinese Academy of Sciences.