Recursive Identification of MIMO Wiener Systems

Bi-Qiang Mu and Han-Fu Chen, *Fellow, IEEE*

***Abstract*—Stochastic approximation (SA) algorithms are proposed to identify a multi-input and multi-output (MIMO) Wiener system, in which the system input is taken to be a sequence of independent and identically distributed (i.i.d.) Gaussian random vectors $\boldsymbol{u}_k \in \mathcal{N}(0, \boldsymbol{I})$. The algorithm for identifying the nonlinear part is designed with multi-variable kernel functions. Under suitable conditions, we show that the estimates of the coefficients of the linear subsystem and of the values of the nonlinear function converge to the respective true values with probability one.**

***Index Terms*—MIMO Wiener system, stochastic approximation, strong consistency, α-mixing.**

I. Introduction

Wiener systems, composed of a linear subsystem followed by a static nonlinearity, are often used to model diverse practical systems; see [1]. For example, in a pH control problem [2], the linear subsystem represents the mixing dynamics of the reagent stream in a stirred vessel and the static nonlinearity describes the pH value as a function of the chemical species contained. As was shown in [3], any time-invariant system with fading memory may be approximated by a Wiener system. However, mainly single-input single-output (SISO) Wiener systems have been dealt with in the existing literature to date with only a few exceptions.

For MIMO Wiener systems, the linear subsystem is identified in [4] by using the subspace method with the help of Bussgang's theorem [5] and Gaussian input. In [6], it is assumed that the nonlinear function is invertible and the inverse function can be parameterized by using basis functions; the consistent estimates are then obtained by applying the subspace identification algorithm combined with the singular value decomposition.

In this technical brief, we adopt the nonparametric approach to identify MIMO Wiener systems by applying a stochastic approximation (SA) algorithm, without requiring invertibility of the nonlinear function. When the nonlinearity is estimated, the multi-variable kernel function is used in the SA algorithm. It is shown that the estimates for coefficients of the linear subsystem and for the static nonlinearity are strongly consistent.

This technical brief is a multidimensional extension of [7] and [8], but the extension is not straightforward. The system description, identifiability of the system, conditions imposed on the system, and probabilistic properties of some system signals are given in Section II. The estimation algorithms are proposed in Section III; their strong consistency is proved in Section IV for the linear subsystem, and in Section V for the nonlinear part. A numerical example is provided for demonstration in Section VI, and a brief conclusion is given in Section VII followed by a technique Appendix.

Manuscript received July 29, 2011; revised December 18, 2011, March 22, 2012, March 24, 2012, and June 28, 2012; accepted August 16, 2012. Date of publication August 24, 2012; date of current version February 18, 2013. This work was supported by the NSFC under Grants 61120106011, 61134013, and the National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences. Recommended by Associate Editor C. Szepesvari.

The authors are with the Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P. R. China (e-mail: bqmu@amss.ac.cn; hfchen@iss.ac.cn).

Digital Object Identifier 10.1109/TAC.2012.2215539

II. System: Identifiability and Its Probabilistic Properties

The system to be considered in the technical brief is as follows:

$$A(z)\boldsymbol{v}_k = B(z)\boldsymbol{u}_k, \quad \boldsymbol{y}_k = \boldsymbol{f}(\boldsymbol{v}_k), \quad \boldsymbol{z}_k = \boldsymbol{y}_k + \boldsymbol{\varepsilon}_k \tag{1}$$

where $A(z) = I + A_1 z + A_2 z^2 + \cdots + A_p z^p$ and $B(z) = B_1 z + B_2 z^2 + \cdots + B_q z^q$ are the $n \times n$ and $n \times m$ matrix-valued polynomials with unknown coefficients but with known orders p and q, respectively, and z is the backward-shift operator: $z\boldsymbol{y}_{k+1} = \boldsymbol{y}_k$. Moreover, $\boldsymbol{u}_k \in \mathbb{R}^m$, $\boldsymbol{y}_k \in \mathbb{R}^n$, and $\boldsymbol{z}_k \in \mathbb{R}^n$ are the system input, output, and observation, respectively; $\boldsymbol{\varepsilon}_k \in \mathbb{R}^n$ is the observation noise. The unknown nonlinear function is denoted by $\boldsymbol{f}(\cdot) = [f_1, f_2, \cdots, f_n]^T$, where $f_i : \mathbb{R}^n \to \mathbb{R}$. The problem is to recursively estimate A_i, $i = 1, \cdots, p$, B_j, $j = 1, \cdots, q$, and $\boldsymbol{f}(\boldsymbol{v})$ at any fixed $\boldsymbol{v}$ based on the available input-output data $\{\boldsymbol{u}_k, \boldsymbol{z}_k\}$.

The identifiability problem is to answer whether or not the system can be uniquely defined by the input-output data $\{\boldsymbol{u}_k, \boldsymbol{y}_k\}$. In general, the answer, is negative. To see this, let P be an $n \times n$ nonsingular matrix, and set $\tilde{A}(z) \triangleq PA(z)P^{-1}$, $\tilde{B}(z) \triangleq PB(z)$, $\tilde{\boldsymbol{f}}(\boldsymbol{x}) \triangleq \boldsymbol{f}(P^{-1}\boldsymbol{x})$, and $\tilde{\boldsymbol{v}}_k \triangleq P\boldsymbol{v}_k$. Then, the following system:

$$\tilde{A}(z)\tilde{\boldsymbol{v}}_k = \tilde{B}(z)\boldsymbol{u}_k, \quad \boldsymbol{y}_k = \tilde{\boldsymbol{f}}(P\boldsymbol{v}_k) = \boldsymbol{f}(\boldsymbol{v}_k), \quad \boldsymbol{z}_k = \boldsymbol{y}_k + \boldsymbol{\varepsilon}_k \tag{2}$$

and the systems (1) share the same input-output data$\{\boldsymbol{u}_k, \boldsymbol{y}_k\}$, but they have different linear and nonlinear parts. So, in order to have the MIMO Wiener system uniquely defined, the nonsingular matrix P should be fixed in advance.

Under what conditions, $A(z)$ and $B(z)$ are uniquely defined on the basis of $\{\boldsymbol{u}_k, \boldsymbol{v}_k\}$? The following lemma provides an answer. The proof can be found in [11].

Lemma 1: Assume $A(z)$ is stable. Under any of the following two equivalent conditions, Un1 and Un2, $A(z)$ and $B(z)$ are uniquely defined on the basis of $\{\boldsymbol{u}_k, \boldsymbol{v}_k\}$:

Un1. $A(z)$ and $B(z)$ have no common left factor, and $[A_p, B_q]$ is of row-full-rank.

Un2. There are no n-vector polynomial $d(z)$ and m-vector polynomial $c(z)$ (not both zero) with orders strictly less than p and q, respectively such that $d^T(z)H(z) + c^T(z) = 0$, where $H(z) = A^{-1}(z)B(z)$.

We now proceed to fix P. We use $\|*\|$ to denote the Euclidean norm for a vector $*$ or the Frobenius norm for a matrix $*$, and use I to denote the identity matrix. Let us first list assumptions to be imposed on the system.

H1: The input $\{\boldsymbol{u}_k, k \geq 0\}$ is a sequence of i.i.d. Gaussian random vectors $\boldsymbol{u}_k \in \mathcal{N}(0, I_{m\times m})$ and is independent of $\{\boldsymbol{\varepsilon}_k\}$.

H2: $A(z)$ and $B(z)$ have no common left factor, $[A_p, B_q]$ is of row-full-rank and $A(z)$ is stable: det $A(z) \neq 0, \forall |z| \leq 1$.

H3: $\boldsymbol{f}(\cdot)$ is a measurable vector-valued function satisfying the following condition:

$$|f_i(x_1, \cdots, x_n)| \leq L\left(M + |x_1|^{\nu_1} + \cdots + |x_n|^{\nu_n}\right), \ 1 \leq i \leq n, \tag{3}$$

where $L > 0$, $M > 0$, and $\nu_j \geq 0, 1 \leq j \leq n$ are constants.

H4: $\{\boldsymbol{\varepsilon}_k\}$ is a sequence of i.i.d. random vectors with $E\boldsymbol{\varepsilon}_k = 0$ and $E[\|\boldsymbol{\varepsilon}_k\|^2] < \infty$.

From H2, it is clear that

$$H(z) \triangleq A^{-1}(z)B(z) = \sum_{i=1}^{\infty} H_i z^i \tag{4}$$

where $H_1 = B_1$, $\|H_i\| = O(e^{-ri})$, $r > 0$, $i > 1$. Assuming $\boldsymbol{u}_k = 0\ \forall k < 0$, we have

$$\boldsymbol{v}_{k+1} = \sum_{i=1}^{k+1} H_i \boldsymbol{u}_{k+1-i} \tag{5}$$

and by H1 $\boldsymbol{v}_{k+1} \in \mathcal{N}(0, \Sigma_{k+1})$ and $\Sigma_{k+1} = \sum_{i=1}^{k+1} H_i H_i^T$. It is clear that $\Sigma_{k+1} \xrightarrow[k\to\infty]{} \Sigma \triangleq \sum_{i=1}^{\infty} H_i H_i^T$. Define

$$\Gamma \triangleq \begin{pmatrix} H_q & H_{q+1} & \cdots & H_{q+np-1} \\ H_{q-1} & H_q & \cdots & H_{q+np-2} \\ \vdots & \vdots & \ddots & \vdots \\ H_{q-p+1} & H_{q-p+2} & \cdots & H_{q+(n-1)p} \end{pmatrix} \tag{6}$$

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

With the help of Lemma 1, we have the following lemma.

Lemma 2: Assume H2 holds. The matrix Γ defined by (6) is of row-full-rank.

The proof is given in the Appendix, while the corresponding assertion for SISO systems can be found in [8] and [12].

As a consequence of this lemma, the matrix composed of the first n rows of Γ is of row-full-rank, and hence Σ_k is nonsingular for all sufficiently large k.

By H3, $Q_{k+1} \triangleq E(\boldsymbol{f}(\boldsymbol{v}_{k+1})\boldsymbol{v}_{k+1}^T)$ is meaningful. Noting $\Sigma_k \xrightarrow[k\to\infty]{} \Sigma$, $Q_k \xrightarrow[k\to\infty]{} Q \triangleq E(\boldsymbol{f}(\boldsymbol{v})\boldsymbol{v}^T)$, where $\boldsymbol{v} \in \mathcal{N}(0, \Sigma)$. We need an additional condition H5.

H5: The matrix Q is nonsingular.

It is clear that H5 is a multidimensional extension of the condition $\rho \neq 0$ used in [7]–[10] for SISO Wiener systems. In order to fix coefficients of the identified system, let us choose P in (2) to be equal to $Q\Sigma^{-1}$. Then, for system (2), $\tilde{P} \triangleq \tilde{Q}\tilde{\Sigma}^{-1} = I$. Therefore, under H5, without loss of generality, we may assume that $Q\Sigma^{-1} = I$ for system (1).

The probabilistic behavior of $V_{k+1} \triangleq [\boldsymbol{v}_{k+1}^T, \boldsymbol{v}_k^T, \cdots, \boldsymbol{v}_{k+2-p}^T, \boldsymbol{u}_{k+1}^T, \boldsymbol{u}_k^T, \cdots, \boldsymbol{u}_{k+2-q}^T]^T \in \mathbb{R}^{np+mq}$ plays an important role in convergence analysis of the algorithms to be proposed below.

Introducing

$$A \triangleq \begin{bmatrix} -A_1 & \cdots & \cdots & -A_p & B_1 & \cdots & \cdots & B_q \\ I_{n\times n} & \ddots & & & 0 & & & 0 \\ & \ddots & \ddots & & \vdots & & & \vdots \\ & & I_{n\times n} & 0 & 0 & \cdots & \cdots & 0 \\ & & & & 0 & \cdots & \cdots & 0 \\ & & & & I_{m\times m} & 0 & \cdots & 0 \\ & & & & & \ddots & \ddots & \vdots \\ & & & & & & I_{m\times m} & 0 \end{bmatrix}$$

$$B \triangleq [0 \quad 0 \quad \cdots \quad 0 \quad I_{m\times m} \quad 0 \quad 0 \quad \cdots \quad 0]^T_{(np+mq)\times m},$$

$$G \triangleq [I_{n\times n} \quad 0 \quad \cdots \quad 0]_{n\times(np+mq)}$$

we then obtain the state-space representation for (1):

$$V_{k+1} = AV_k + B\boldsymbol{u}_{k+1}, \quad \boldsymbol{v}_{k+1} = GV_{k+1}. \tag{7}$$

For the process $\{X_k, k = 0, 1, \cdots\}$, we denote by $\mathcal{F}_i^j$ the σ-algebra generated by $\{X_s, 0 \leq i \leq s \leq j\}$. 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 $\alpha_k \xrightarrow[k\to\infty]{} 0$.

Remark 1: It is clear that under H1 and H2, $\{V_k\}_{k\geq 0}$ is a time-homogeneous Markov chain taking values in $(\mathbb{R}^{np+mq}, \mathcal{B}^{np+mq})$, where $\mathcal{B}^{np+mq}$ denotes the Borel σ-algebra in $\mathbb{R}^{np+mq}$, and $\{V_k\}$ is an α-mixing process with mixing coefficient α_k exponentially decaying to zero [8], [13]–[15]

$$\alpha_k \leq d\lambda^k\ \forall k \geq 1 \quad \text{for some} \quad d > 0 \quad \text{and} \quad 0 < \lambda < 1. \tag{8}$$

It is worth noting that the mixing property is hereditary [13] in the sense that the process $\{h(V_k)\}$ for any measurable function $h(\cdot)$ possesses the same mixing property as that of $\{V_k\}$. So, the processes like $\{\boldsymbol{f}(\boldsymbol{v}_k)\}$ and $\{\boldsymbol{f}(\boldsymbol{v}_k)\boldsymbol{u}_{k-i}^T, \forall i \geq 1\}$ and so on are all α-mixing sequences with mixing coefficients exponentially tending to zero.

III. Recursive Identification Algorithms

A. Estimation of $\{A_1, \cdots, A_p, B_1, \cdots, B_q\}$

We first estimate the impulse responses H_i and then the coefficients of $A(z)$ and $B(z)$.

Lemma 3: Assume H1–H5 hold. Then

$$E\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T \xrightarrow[k\to\infty]{} H_i \quad \forall i \geq 1. \tag{9}$$

Proof: From (5), it is seen that $\boldsymbol{v}_{k+1} \in \mathcal{N}(0, \Sigma_{k+1})$, $\Sigma_{k+1} \triangleq \sum_{i=1}^{k+1} H_i H_i^T$. The Gaussian vector $\boldsymbol{g}_{k,j} \triangleq [\boldsymbol{v}_{k+1}^T, \boldsymbol{u}_{k+1-j}^T]^T$, $j \geq 1$ is zero mean with covariance Matrix $G_{k,j} \triangleq \begin{pmatrix} \Sigma_{k+1} & H_j \\ H_j^T & I \end{pmatrix}$.

It is straightforward to verify that $G_{k,j}$ can be factorized $G_{k,j} = L_{k,j}L_{k,j}^T$ with

$$L_{k,j} \triangleq \begin{pmatrix} \Sigma_{k+1}^{\frac{1}{2}} & 0 \\ H_j^T \Sigma_{k+1}^{-\frac{1}{2}} & \left(I_{m\times m} - H_j^T \Sigma_{k+1}^{-1} H_j\right)^{\frac{1}{2}} \end{pmatrix}.$$

It then follows that $L_{k,j}^{-1} G_{k,j} L_{k,j}^{-T} = I_{(n+m)\times(n+m)}$ and $\boldsymbol{l}_{k,j} = [\boldsymbol{l}_{k,j}(1)^T, \boldsymbol{l}_{k,j}(2)^T]^T \triangleq L_{k,j}^{-1}\boldsymbol{g}_{k,j}$ is a Gaussian vector $\boldsymbol{l}_{k,j} \in \mathcal{N}(0, I_{(n+m)\times(n+m)})$, and hence the components of $\boldsymbol{g}_{k,j} = L_{k,j}\boldsymbol{l}_{k,j}$ can be expressed as follows:

$$\begin{aligned} \boldsymbol{v}_{k+1} &= \Sigma_{k+1}^{\frac{1}{2}} \boldsymbol{l}_{k,j}(1), \\ \boldsymbol{u}_{k+1-j} &= H_j^T \Sigma_{k+1}^{-\frac{1}{2}} \boldsymbol{l}_{k,j}(1) + \left(I_{m\times m} - H_j^T \Sigma_{k+1}^{-1} H_j\right)^{\frac{1}{2}} \boldsymbol{l}_{k,j}(2). \end{aligned}$$

Noting that components of $\boldsymbol{l}_{k,j}$ are orthogonal to each other, we obtain

$$\begin{aligned} &E\boldsymbol{f}(\boldsymbol{v}_{k+1})\boldsymbol{u}_{k+1-j}^T \\ &= E\boldsymbol{f}\left(\Sigma_{k+1}^{\frac{1}{2}} \boldsymbol{l}_{k,j}(1)\right) \\ &\quad \cdot \left[H_j^T \Sigma_{k+1}^{-\frac{1}{2}} \boldsymbol{l}_{k,j}(1) + \left(I_{m\times m} - H_j^T \Sigma_{k+1}^{-1} H_j\right)^{\frac{1}{2}} \boldsymbol{l}_{k,j}(2)\right]^T \\ &= E\boldsymbol{f}\left(\Sigma_{k+1}^{\frac{1}{2}} \boldsymbol{l}_{k,j}(1)\right) \left[\Sigma_{k+1}^{\frac{1}{2}} \boldsymbol{l}_{k,j}(1)\right]^T \Sigma_{k+1}^{-1} H_j \\ &= Q_{k+1}\Sigma_{k+1}^{-1} H_j \end{aligned} \tag{10}$$

where $Q_{k+1} = E(\boldsymbol{f}(\boldsymbol{v}_{k+1})\boldsymbol{v}_{k+1}^T)$.

By H1 and (10), we have

$$\begin{aligned} E\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T &= E(\boldsymbol{y}_{k+1} + \boldsymbol{\varepsilon}_{k+1})\boldsymbol{u}_{k+1-i}^T \\ &= E\boldsymbol{f}(\boldsymbol{v}_{k+1})\boldsymbol{u}_{k+1-i}^T \\ &= Q_{k+1}\Sigma_{k+1}^{-1} H_i \xrightarrow[k\to\infty]{} H_i \end{aligned}$$

since $\Sigma_k \xrightarrow[k\to\infty]{} \Sigma$, $Q_k \xrightarrow[k\to\infty]{} Q$, and $Q\Sigma^{-1} = I$. The proof is completed. ■

Based on (9), we apply the stochastic approximation algorithm with expanding truncations (SAAWET) [16] to recursively estimate H_i

$$H_{i,k+1} = \left[H_{i,k} - \frac{1}{k}\left(H_{i,k} - \boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T\right)\right] \cdot I_{\left[\left\|H_{i,k}-\frac{1}{k}\left(H_{i,k}-\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T\right)\right\|\le M_{\delta_{i,k}}\right]}, \quad (11)$$

$$\delta_{i,k} = \sum_{j=1}^{k-1} I_{\left[\left\|H_{i,j}-\frac{1}{j}\left(H_{i,j}-\boldsymbol{z}_{j+1}\boldsymbol{u}_{j+1-i}^T\right)\right\|> M_{\delta_{i,j}}\right]} \quad (12)$$

where $\{M_k\}$ is an arbitrarily chosen sequence of positive real numbers increasingly diverging to infinity, $H_{i,0}$ is an arbitrary initial value, and I_A denotes the indicator function of a set A.

Remark 2: Algorithm (11) would be the conventional Robbins-Monro (RM) algorithm if the indicator function were removed. The matrix obtained from the modified RM algorithm is compared with the sphere of radius $M_{\delta_{i,k}}$: If the matrix remains in the sphere then the algorithm continues as the RM algorithm. Otherwise, the algorithm restarts from the origin and the truncation sphere enlarges its radius from $M_{\delta_{i,k}}$ to $M_{\delta_{i,k}+1}$.

From (4), it follows that:

$$B_1 z + \cdots + B_q z^q = (I + A_1 z + \cdots + A_p z^p)(H_1 z + \cdots + H_i z^i + \cdots) \quad (13)$$

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

$$B_i = \sum_{j=0}^{i\wedge p} A_j H_{i-j}, \quad \forall 1 \le i \le q \quad (14)$$

and

$$H_i = -\sum_{j=1}^{i\wedge p} A_j H_{i-j}, \quad \forall i \ge q+1, \quad (15)$$

where $a \wedge b$ denotes $\min(a,b)$.

For $H_i, q+1 \le i \le q+np$, by (15), we obtain the following linear algebraic equation:

$$[A_1, A_2, \cdots, A_p]\Gamma = -[H_{q+1}, H_{q+2}, \cdots, H_{q+np}] \quad (16)$$

where Γ is given by (6).

Below, in Theorem 1 we show that $H_{i,k} \to H_i$ a.s. as $k \to \infty$. As a consequence, since Γ is of row-full-rank thanks to Lemma 2

$$\Gamma_k \triangleq \begin{pmatrix} H_{q,k} & H_{q+1,k} & \cdots & H_{q+np-1,k} \\ H_{q-1,k} & H_{q,k} & \cdots & H_{q+np-2,k} \\ \vdots & \vdots & \ddots & \vdots \\ H_{q-p+1,k} & H_{q-p+2,k} & \cdots & H_{q+(n-1)p,k} \end{pmatrix}$$

is also of row-full-rank when k is sufficiently large. Thus, Γ_k can serve as the kth estimate of Γ with $H_{i,k} = 0$ for $i \le 0$.

The estimates for $\{A_1, \cdots, A_p, B_1, \cdots, B_q\}$ are naturally defined as follows:

$$[A_{1,k}, A_{2,k}, \cdots, A_{p,k}] = -[H_{q+1,k}, H_{q+2,k}, \cdots, H_{q+np,k}] \times \Gamma_k^T\left(\Gamma_k\Gamma_k^T\right)^{-1} \quad (17)$$

$$B_{i,k} = \sum_{j=0}^{i\wedge p} A_{j,k} H_{i-j,k}, \ \forall 1 \le i \le q. \quad (18)$$

B. Nonparametric Estimation of $\boldsymbol{f}(\cdot)$

We now recursively estimate $\boldsymbol{f}(\boldsymbol{y})$ for any fixed $\boldsymbol{y} \in \mathbb{R}^n$. By using the estimates obtained for the coefficients of the linear subsystem we can estimate the internal signals $\boldsymbol{v}_k$ on the basis of the state-space representations of the linear subsystem. Then, applying SAAWET incorporated with a multi-variable kernel function we obtain estimates for $\boldsymbol{f}(\boldsymbol{y})$. Let us start with estimating $\boldsymbol{v}_k$.

Define

$$C \triangleq \begin{pmatrix} -A_1 & I & & \\ \vdots & & \ddots & \\ \vdots & & & I \\ -A_s & 0 & \cdots & 0 \end{pmatrix}, \ D \triangleq \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_s \end{pmatrix}, \ H \triangleq \begin{pmatrix} I \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and $s \triangleq \max(p,q)$. Then, (1) can be presented in the state-space form

$$x_{k+1} = Cx_k + D\boldsymbol{u}_k, \quad \boldsymbol{v}_{k+1} = H^T x_{k+1} \quad (19)$$

where C is an $sn \times sn$ matrix, D is an $sn \times m$ matrix, and H is an $sn \times n$ matrix, $A_k = 0$ for $k > p$, and $B_k = 0$ for $k > q$.

Replacing A_i and B_j in C and D with $A_{i,k}$ and $B_{j,k}$ given by (17) and (18), respectively, $i = 1, \cdots, s$, $j = 1, \cdots, s$, we obtain the estimates C_k and D_k for C and D at time k, and hence, the estimate $\widehat{\boldsymbol{v}}_{k+1}$ for $\boldsymbol{v}_{k+1}$ is given as follows:

$$\widehat{x}_{k+1} = C_{k+1}\widehat{x}_k + D_{k+1}\boldsymbol{u}_k, \quad \widehat{\boldsymbol{v}}_{k+1} = H^T\widehat{x}_{k+1} \quad (20)$$

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

To estimate $\boldsymbol{f}(\boldsymbol{y})$, let us introduce the kernel function w_k and its estimate $\widehat{w}_k$ as follows:

$$w_k(\boldsymbol{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}(b_k)^n} e^{-\frac{(\boldsymbol{y}-\boldsymbol{v}_k)^T(\boldsymbol{y}-\boldsymbol{v}_k)}{2b_k^2}}$$

$$\widehat{w}_k(\boldsymbol{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}(b_k)^n} e^{-\frac{(\boldsymbol{y}-\widehat{\boldsymbol{v}}_k)^T(\boldsymbol{y}-\widehat{\boldsymbol{v}}_k)}{2b_k^2}} \quad (21)$$

where b_k is the window width of the kernel function w_k, and we set $b_k = 1/k^a$, $a > 0$.

We apply SAAWET incorporated with the kernel function rather than the nonparametric kernel estimation method [13] and the support vector machines [17] to estimate $\boldsymbol{f}(\boldsymbol{y})$.

Let $m_k = k^b, b > 0$ such that $(n+2)a + b < (1/2)$. With $\Delta_0(\boldsymbol{y}) = 0$ and arbitrary $\mu_0(\boldsymbol{y})$ we recursively estimate $\boldsymbol{f}(\boldsymbol{y})$ as follows:

$$\mu_{k+1}(\boldsymbol{y}) = \left[\mu_k(\boldsymbol{y}) - \frac{1}{k}\widehat{w}_k\left(\mu_k(\boldsymbol{y}) - \boldsymbol{z}_k\right)\right] \cdot I_{\left[\left\|\mu_k(\boldsymbol{y})-\frac{1}{k}\widehat{w}_k(\mu_k(\boldsymbol{y})-\boldsymbol{z}_k)\right\|\le m_{\Delta_k}(\boldsymbol{y})\right]}$$

$$\Delta_k(\boldsymbol{y}) = \sum_{j=1}^{k-1} I_{\left[\left\|\mu_j(\boldsymbol{y})-\frac{1}{j}\widehat{w}_j\left(\mu_j(\boldsymbol{y})-\boldsymbol{z}_j\right)\right\|> m_{\Delta_j}(\boldsymbol{y})\right]}. \quad (22)$$

IV. Consistency of Estimates for Linear Subsystem

We now proceed to prove strong consistency of the estimates given in Section III. In the sequel to save the space, whenever the extension of a proof from SISO to MIMO systems is easy and the corresponding proof can be carried out in a similar way as that for SISO systems, we simply refer to [7] and [8] instead of giving the detailed proof.

The proof of the following lemma can be found in [8] as the proof of Lemma 4 there.

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

Lemma 5: Assume H1–H5 hold. For any $0 < \nu < (1/2)$ the following holds:

$$\sum_{k=1}^{\infty}\frac{1}{k^{1-\nu}}\left[E\left(\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T\right)-H_i\right]<\infty \quad \text{for } i\geq 1 \tag{23}$$

$$\sum_{k=1}^{\infty}\frac{1}{k^{1-\nu}}\left(E\left(\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T\right)-\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T\right)<\infty\ a.s. \quad \text{for } i\geq 1. \tag{24}$$

Proof: By noticing $E\boldsymbol{f}(\boldsymbol{v})\boldsymbol{v}^T\Sigma^{-1} = Q\Sigma^{-1} = I$, we have $E\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T - H_i = (I_{1,k} + I_{2,k})H_i$, where $I_{1,k} = E\boldsymbol{f}(\boldsymbol{v}_{k+1})\boldsymbol{v}_{k+1}^T(\Sigma_{k+1}^{-1} - \Sigma^{-1})$ and $I_{2,k} = E(\boldsymbol{f}(\boldsymbol{v}_{k+1})\boldsymbol{v}_{k+1}^T - \boldsymbol{f}(\boldsymbol{v})\boldsymbol{v}^T)\Sigma^{-1}$. Since $\boldsymbol{v}_{k+1} \in \mathcal{N}(0, \Sigma_{k+1})$ and the impulse responses H_i exponentially decay to zero as $i \to \infty$, i.e., $\|H_i\| = O(e^{-ri})$, $r > 0, i > 1$, it is shown that $I_{1,k} = O(\mu^k)$ and $I_{2,k} = O(\mu^k)$ for some $\mu \in (0,1)$, which imply (23).

For (24) it suffices to show the convergence of each element (indexed by β,γ) of the series, i.e., for $1 \leq \beta \leq n, 1 \leq \gamma \leq m, i \geq 1$

$$\begin{aligned}\sum_{k=1}^{\infty}&\frac{1}{k^{1-\nu}}\left(E\left(\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T\right)-\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T\right)_{\beta,\gamma}\\ &=\sum_{k=1}^{\infty}\frac{1}{k^{1-\nu}}\left[Ef_\beta(\boldsymbol{v}_{k+1})u_{k+1-i}(\gamma)-f_\beta(\boldsymbol{v}_{k+1})u_{k+1-i}(\gamma)\right]\\ &\quad-\sum_{k=1}^{\infty}\frac{1}{k^{1-\nu}}\varepsilon_{k+1}(\beta)u_{k+1-i}(\gamma)<\infty\ a.s.\end{aligned} \tag{25}$$

where $\boldsymbol{u}_k = [u_k(1),\cdots,u_k(m)]^T$, $\boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}),\cdots,f_n(\boldsymbol{x})]^T$.

Define $z_k^{(\beta,\gamma)} = (1/k^{1-\nu})[Ef_\beta(\boldsymbol{v}_{k+1})u_{k+1-i}(\gamma) - f_\beta(\boldsymbol{v}_{k+1})u_{k+1-i}(\gamma)]$ for $1 \leq \beta \leq n, 1 \leq \gamma \leq m$. By Remark 1, $z_k^{(\beta,\gamma)}$ is an α-mixing sequence with mixing coefficients satisfying (8). By the C_r-inequality [18] and then the Schwarz inequality, for any $\epsilon > 0$, we have $E|z_k^{(\beta,\gamma)}|^{2+\epsilon} = O(1/k^{(1-\nu)(2+\epsilon)})$, which implies $\sum_{k=1}^{\infty}(E|z_k^{(\beta,\gamma)}|^{2+\epsilon})^{(2/(2+\epsilon))} = O(\sum_{k=1}^{\infty}(1/k^{2(1-\nu)})) < \infty$. Then the first term at the right-hand side of (25) converges by Lemma 4.

Since $\boldsymbol{u}_k$ and $\boldsymbol{\varepsilon}_k$ are mutually independent and $E\|\boldsymbol{\varepsilon}_k\|^2 < \infty$, we have

$$\sum_{k=1}^{\infty}E\left(\frac{1}{k^{1-\nu}}\varepsilon_{k+1}(\beta)u_{k+1-i}(\gamma)\right)^2=\sum_{k=1}^{\infty}O\left(\frac{1}{k^{2(1-\nu)}}\right)<\infty.$$

By the martingale convergence theorem [18], the last term in (25) converges too. For details, we refer to [8]. ■

Theorem 1: Assume H1–H5 hold. Then, $H_{i,k}, i \geq 1$, defined by (11)-(12) satisfy

$$\|H_{i,k}-H_i\|=o(k^{-\nu})\quad a.s.,\quad \forall\nu\in\left(0,\frac{1}{2}\right). \tag{26}$$

Proof: We rewrite (11) as

$$H_{i,k+1}=\left[H_{i,k}-\frac{1}{k}(H_{i,k}-H_i)-\frac{1}{k}\epsilon_{k+1}(i)\right]\cdot I_{\left[\left\|H_{i,k}-\frac{1}{k}(H_{i,k}-H_i)-\frac{1}{k}\epsilon_{k+1}(i)\right\|\leq M_{\delta_{i,k}}\right]}$$

where $\epsilon_{k+1}(i) = H_i - \boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T = [H_i - E\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T] + [E\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T - \boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T]$.

Since H_i is the single root of the linear function $-(\boldsymbol{x} - H_i)$, by the convergence rate theorem of SAAWET [16], it suffices to prove

$$\sum_{k=1}^{\infty}\frac{1}{k^{1-\nu}}\epsilon_{k+1}(i)<\infty\ a.s.\ i\geq 1\quad\forall\nu\in\left(0,\frac{1}{2}\right). \tag{27}$$

Write $\epsilon_{k+1}(i)$ as $\epsilon_{k+1}(i) = \epsilon'_{k+1} + \epsilon''_{k+1}$, where

$$\begin{aligned}\epsilon'_{k+1}&=E\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T-\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T,\\ \epsilon''_{k+1}&=H_i-E\boldsymbol{z}_{k+1}\boldsymbol{u}_{k+1-i}^T.\end{aligned}$$

By Lemma 5, we find that (27) is true, and hence (26) holds.

Corollary 1: From (6), (17) and (18), by Theorem 1 the following convergence rates are derived: $\|\Gamma_k - \Gamma\| = o(k^{-\nu})$, $\|A_{i,k} - A_i\| = o(k^{-\nu})$, and $\|B_{j,k} - B_j\| = o(k^{-\nu})$, $i = 1,\cdots,p$, $j = 1,\cdots,q$, for any $\nu \in (0,1/2)$.

V. Consistency of the Estimates of the Nonlinearity

Lemma 6: Assume H1–H3 hold. The following limits take place

$$\frac{\|\boldsymbol{u}_k\|}{k^c}\xrightarrow[k\to\infty]{a.s.}0,\ \frac{\|\boldsymbol{f}(\boldsymbol{v}_k)\|}{k^c}\xrightarrow[k\to\infty]{a.s.}0\ \forall c>0. \tag{28}$$

Proof:

$$P\left[\frac{\|\boldsymbol{u}_k\|}{k^c}>\varepsilon\right]=P\left[\frac{\|\boldsymbol{u}_k\|^{\frac{2}{c}}}{k^2}>\varepsilon^{\frac{2}{c}}\right]<\frac{1}{\varepsilon^{\frac{2}{c}}k^2}E\|\boldsymbol{u}_k\|^{\frac{2}{c}}$$

for any given $\varepsilon > 0$, it follows that $\sum_{k=1}^{\infty}P[(\|\boldsymbol{u}_k\|/k^c) > \varepsilon] < \infty$. Hence, by the Borel-Cantelli lemma, we derive that $(\|\boldsymbol{u}_k\|/k^c)\xrightarrow[k\to\infty]{a.s.}0$. By the growth rate restriction on $\boldsymbol{f}(\cdot)$, the second assertion of the lemma can be proved in a similar way since $\boldsymbol{v}_k$ is Gaussian with variance $\Sigma_k \xrightarrow[k\to\infty]{} \Sigma$. ■

Lemma 7: Assume H1–H4 hold. Then

$$\begin{aligned}&E[w_k]\xrightarrow[k\to\infty]{}\rho(\boldsymbol{y}),\quad E\left[w_k\boldsymbol{f}(\boldsymbol{v}_k)\right]\xrightarrow[k\to\infty]{}\rho(\boldsymbol{y})\boldsymbol{f}(\boldsymbol{y})\\ &E\left[w_k\left\|\boldsymbol{f}(\boldsymbol{v}_k)\right\|\right]\xrightarrow[k\to\infty]{}\rho(\boldsymbol{y})\left\|\boldsymbol{f}(\boldsymbol{y})\right\|\end{aligned} \tag{29}$$

$$E|w_k|^\delta=O\left(\frac{1}{b_k^{n(\delta-1)}}\right)\quad\forall\delta\geq 2 \tag{30}$$

$$E\left\|w_k\boldsymbol{f}(\boldsymbol{v}_k)\right\|^\delta=O\left(\frac{1}{b_k^{n(\delta-1)}}\right)\quad\forall\delta\geq 2 \tag{31}$$

where $\rho(\boldsymbol{y}) = (1/(2\pi)^{n/2}|\Sigma|^{1/2})e^{-(\boldsymbol{y}^T\Sigma^{-1}\boldsymbol{y}/2)}$ and $\Sigma = \sum_{i=1}^{\infty}H_iH_i^T$, and the following also holds a.s.:

$$\sum_{k=1}^{\infty}\frac{1}{k}(w_k-Ew_k)<\infty \tag{32}$$

$$\sum_{k=1}^{\infty}\frac{1}{k}\left(|w_k-Ew_k|-E|w_k-Ew_k|\right)<\infty \tag{33}$$

$$\sum_{k=1}^{\infty}\frac{1}{k}\left(w_k\|\boldsymbol{y}_k\|-E\left[w_k\|\boldsymbol{y}_k\|\right]\right)<\infty \tag{34}$$

$$\sum_{k=1}^{\infty}\frac{1}{k}(w_k\boldsymbol{y}_k-Ew_k\boldsymbol{y}_k)<\infty \tag{35}$$

$$\sum_{k=1}^{\infty}\frac{1}{k}w_k\boldsymbol{\varepsilon}_k<\infty,\sum_{k=1}^{\infty}\frac{1}{k}w_k\left(\|\boldsymbol{\varepsilon}_k\|-E\|\boldsymbol{\varepsilon}_k\|\right)<\infty. \tag{36}$$

Proof: We prove the first limit in (29). Noting $\boldsymbol{v}_k \in \mathcal{N}(0, \Sigma_k)$ with $\Sigma_k = \sum_{i=1}^{k} H_i H_i^T$, we have

$$Ew_k = \int\limits_{\mathbb{R}^n} \frac{1}{(2\pi)^n b_k^n |\Sigma_k|^{\frac{1}{2}}} e^{-\left(\frac{(\boldsymbol{y}-\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x})}{2b_k^2} + \frac{\boldsymbol{x}\Sigma_k^{-1}\boldsymbol{x}}{2}\right)} \mathrm{d}\boldsymbol{x}$$
$$= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^n |\Sigma_k|^{\frac{1}{2}}} e^{-\left(\frac{\boldsymbol{x}^T\boldsymbol{x}}{2} + \frac{(\boldsymbol{y}-b_k\boldsymbol{x})^T\Sigma_k^{-1}(\boldsymbol{y}-b_k\boldsymbol{x})}{2}\right)} \mathrm{d}\boldsymbol{x}$$
$$\xrightarrow[k\to\infty]{} \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{\boldsymbol{y}^T\Sigma^{-1}\boldsymbol{y}}{2}} \int\limits_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\left(\frac{\boldsymbol{x}^T\boldsymbol{x}}{2}\right)} \mathrm{d}\boldsymbol{x}$$
$$= \rho(\boldsymbol{y}).$$

The second limit in (29) and (30), (31) can be proved in a similar manner.

Convergence of the series (32)-(35) can be proved by a treatment similar to that used in [8] consisting of verifying the conditions required in Lemma 4, while for (36) the convergence theorem for martingale difference sequences (mds) is applied by noticing that $\boldsymbol{u}_k$ and $\boldsymbol{\varepsilon}_k$ are mutually independent and $E\|\boldsymbol{\varepsilon}_k\|^2 < \infty$, hence $\sum_{k=1}^{\infty}(1/k^2)E(w_k^2\|\boldsymbol{\varepsilon}_k\|^2) = O(\sum_{k=1}^{\infty}(1/k^{2-na})) < \infty$. ■

Lemma 8: Assume that H1–H5 hold. There exists a constant $c > 0$ with $(1/2) - (n+2)a - b - 3c > 0$ such that

$$\|\boldsymbol{v}_k - \widehat{\boldsymbol{v}}_k\| = o\left(\frac{1}{k^{1/2-2c}}\right), \ |w_k - \widehat{w}_k| = o\left(\frac{1}{k^{1/2-(n+2)a-2c}}\right). \quad (37)$$

Proof: From (19) and (20), we have

$$\begin{aligned}\hat{x}_{k+1} - x_{k+1} &= C_{k+1}\hat{x}_k - Cx_k + (D_{k+1} - D)\boldsymbol{u}_k \\ &= C_{k+1}(\hat{x}_k - x_k) + (C_{k+1} - C)x_k + (D_{k+1} - D)\boldsymbol{u}_k.\end{aligned}$$

Since C is stable and $C_k \to C$, there exists a $\lambda \in (0,1)$ such that

$$\|\hat{x}_{k+1} - x_{k+1}\| \le N_1\lambda^{k+1}\|\hat{x}_0 - x_0\| + S(\lambda, k) \quad (38)$$

where

$$S(\lambda, k) = N_2 \sum_{j=1}^{k+1} \lambda^{k-j+1} \left(\|C_j - C\| \cdot \|x_{j-1}\| \right. \\ \left. \|D_j - D\| \cdot \|\boldsymbol{u}_{j-1}\|\right)$$

with $N_1 > 0$ and $N_2 > 0$ being constants.

By Corollary 1 and Lemma 6, we have $S(\lambda, k) = o(1/k^{1/2-2c})$, which incorporating with (38) implies $\|\hat{x}_k - x_k\| = o(1/k^{1/2-2c})$.

Since $\|\boldsymbol{v}_k - \hat{\boldsymbol{v}}_k\|$ and $\|\hat{x}_k - x_k\|$ are of the same order, we have

$$\begin{aligned}|\hat{w}_k - w_k| &= o\left(\frac{1}{b_k^{(n+2)}}\|\hat{\boldsymbol{v}}_k - \boldsymbol{v}_k\|\right) = o\left(\frac{1}{b_k^{(n+2)}}\|\hat{x}_k - x_k\|\right) \\ &= o\left(\frac{1}{k^{1/2-(n+2)a-2c}}\right).\end{aligned}$$

■

Theorem 2: Assume that H1–H5 hold. Then $\mu_k(\boldsymbol{y})$ defined by (22) is strongly consistent:

$$\mu_k(\boldsymbol{y}) \xrightarrow[k\to\infty]{} \boldsymbol{f}(\boldsymbol{y}) \quad a.s., \quad \boldsymbol{y} \in \mathbb{R}^n. \quad (39)$$

Proof: The algorithm (22) can be rewritten as

$$\mu_{k+1}(\boldsymbol{y}) = \left[\mu_k(\boldsymbol{y}) - \frac{1}{k}\rho(\boldsymbol{y})\left(\mu_k(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{y})\right) - \frac{1}{k}\bar{\epsilon}_{k+1}(\boldsymbol{y})\right] \\ \cdot I_{\left[\left\|\mu_k(\boldsymbol{y}) - \frac{1}{k}\rho(\boldsymbol{y})(\mu_k(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{y})) - \frac{1}{k}\bar{\epsilon}_{k+1}(\boldsymbol{y})\right\| \le m_{\Delta_k}(\boldsymbol{y})\right]}$$

where $\bar{\epsilon}_{k+1}(\boldsymbol{y}) = \widehat{w}_k(\mu_k(\boldsymbol{y}) - \boldsymbol{z}_k) - \rho(\boldsymbol{y})(\mu_k(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{y}))$.

Since $\boldsymbol{f}(\boldsymbol{y})$ is the unique root of $-\rho(\boldsymbol{y})(\boldsymbol{x} - \boldsymbol{f}(\boldsymbol{y}))$, by Theorem 2.2.1 in [16], for (39) it suffices to prove

$$\lim_{T\to 0}\limsup_{k\to\infty} \frac{1}{T}\left\|\sum_{j=n_k}^{m(n_k,T_k)} \frac{1}{j}\bar{\epsilon}_{j+1}(\boldsymbol{y})\right\| = 0 \quad \forall T_k \in [0, T] \quad (40)$$

for any convergent subsequence $\mu_{n_k}(\boldsymbol{y})$, where $m(k,T) \triangleq \max\{m : \sum_{j=k}^{m}(1/j) \le T\}$.

Write $\bar{\epsilon}_{k+1}(\boldsymbol{y})$ as $\bar{\epsilon}_{k+1}(\boldsymbol{y}) = \sum_{i=1}^{4} \bar{\epsilon}_{k+1}^{(i)}(\boldsymbol{y})$, where

$$\bar{\epsilon}_{k+1}^{(1)}(\boldsymbol{y}) = (\widehat{w}_k - w_k)\left(\mu_k(\boldsymbol{y}) - \boldsymbol{z}_k\right), \quad \bar{\epsilon}_{k+1}^{(4)}(\boldsymbol{y}) = -w_k\boldsymbol{\varepsilon}_k$$
$$\bar{\epsilon}_{k+1}^{(3)}(\boldsymbol{y}) = \rho(\boldsymbol{y})\boldsymbol{f}(\boldsymbol{y}) - w_k\boldsymbol{f}(\boldsymbol{v}_k),$$
$$\bar{\epsilon}_{k+1}^{(2)}(\boldsymbol{y}) = (w_k - \rho(\boldsymbol{y}))\,\mu_k(\boldsymbol{y}).$$

Noticing $\|\mu_k(\boldsymbol{y})\| \le k^b$ by Lemmas 6 and 8 it follows that:

$$\sum_{k=1}^{\infty} \frac{1}{k}(\widehat{w}_k - w_k)\left(\mu_k(\boldsymbol{y}) - \boldsymbol{z}_k\right) < \infty. \quad (41)$$

Hence, (40) holds for $i = 1$.

For $i = 3$, we have

$$\begin{aligned}\bar{\epsilon}_{k+1}^{(3)}(\boldsymbol{y}) &= \rho(\boldsymbol{y})\boldsymbol{f}(\boldsymbol{y}) - w_k\boldsymbol{f}(\boldsymbol{v}_k) \\ &= [\rho(\boldsymbol{y})\boldsymbol{f}(\boldsymbol{y}) - Ew_k\boldsymbol{f}(\boldsymbol{v}_k)] \\ &\quad + [Ew_k\boldsymbol{f}(\boldsymbol{v}_k) - w_k\boldsymbol{f}(\boldsymbol{v}_k)]. \quad (42)\end{aligned}$$

By the second limit in (29) of Lemma 7 and by convergence of the series in (35) it follows that (40) holds for $i = 3$.

Convergence of the first series in (36) assures that (40) holds for $i = 4$.

It remains to show (40) for $i = 2$. For this it is first shown that if $\mu_{n_k}(\boldsymbol{y})$ is a convergent subsequence of $\{\mu_k(\boldsymbol{y})\}$: $\mu_{n_k}(\boldsymbol{y}) \xrightarrow[k\to\infty]{} \bar{\mu}(\boldsymbol{y})$, then for all large enough k and sufficiently small $T > 0$

$$\mu_{i+1}(\boldsymbol{y}) = \mu_i(\boldsymbol{y}) - \frac{1}{i}\widehat{w}_i\left(\mu_i(\boldsymbol{y}) - \boldsymbol{z}_i\right) \quad (43)$$

and

$$\|\mu_{i+1}(\boldsymbol{y}) - \mu_{n_k}(\boldsymbol{y})\| \le cT, \quad i = n_k, n_k+1, \cdots, m(n_k, T), \quad (44)$$

where $c > 0$ is a constant, which is independent of k but may depend on sample path ω. We then have

$$\begin{aligned}\sum_{j=n_k}^{m(n_k,T_k)} \frac{1}{j}\bar{\epsilon}_{j+1}^{(2)}(\boldsymbol{y}) &= \sum_{j=n_k}^{m(n_k,T_k)} \frac{1}{j}\left(\mu_j(\boldsymbol{y}) - \bar{\mu}(\boldsymbol{y})\right)(w_j - Ew_j) \\ &\quad + \bar{\mu}(\boldsymbol{y})\sum_{j=n_k}^{m(n_k,T_k)} \frac{1}{j}(w_j - Ew_j) \\ &\quad + \sum_{j=n_k}^{m(n_k,T_k)} \frac{1}{j}\left(Ew_j - \rho(\boldsymbol{y})\right)\mu_j(\boldsymbol{y}). \quad (45)\end{aligned}$$

On the right-hand side of the equality in (45), the second term tends to zero as $k \to \infty$ by (32), the last term tends to zero as $k \to \infty$ by the first limit in (29) and (44), while the first term is analyzed as follows. By (29), (33), (44), and $\mu_{n_k}(\boldsymbol{y}) \xrightarrow[k\to\infty]{} \bar{\mu}(\boldsymbol{y})$, we have

$$\begin{aligned}&\lim_{T\to 0}\limsup_{k\to\infty} \frac{1}{T}\sum_{j=n_k}^{m(n_k,T_k)} \frac{1}{j}\left(\mu_j(\boldsymbol{y}) - \bar{\mu}(\boldsymbol{y})\right)(w_j - Ew_j) \\ &= \lim_{T\to 0}\limsup_{k\to\infty} \frac{1}{T}O(T)\sum_{j=n_k}^{m(n_k,T_k)} \frac{1}{j}E|w_j - Ew_j| \\ &= \lim_{T\to 0}\limsup_{k\to\infty} O(1)\sum_{j=n_k}^{m(n_k,T_k)} \frac{1}{j}Ew_j = 0. \quad (46)\end{aligned}$$

Thus, (40) is also valid for $i = 2$. ■

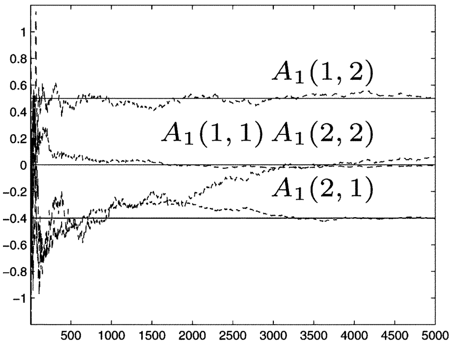

Fig. 1. Estimates for A_1.

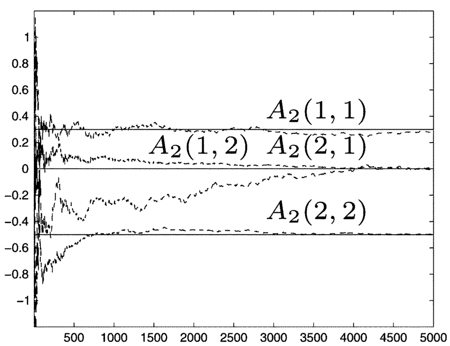

Fig. 2. Estimates for A_2.

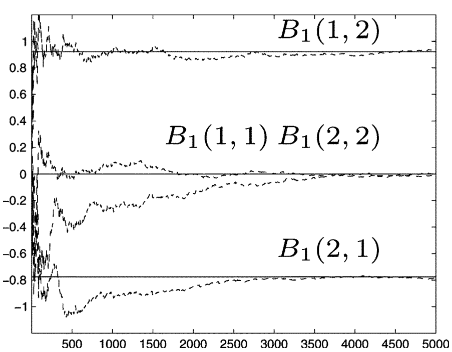

Fig. 3. Estimates for B_1.

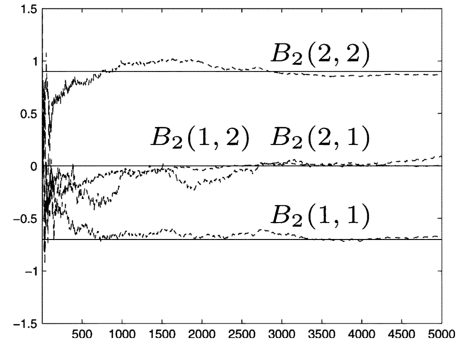

Fig. 4. Estimates for B_2.

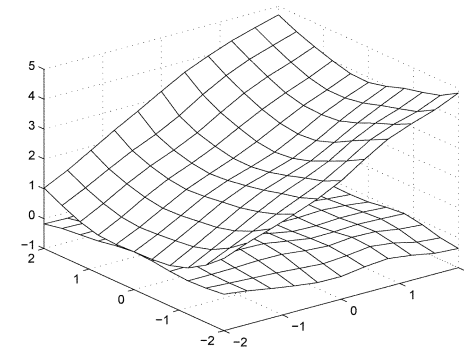

Fig. 5. Estimates for f_1 and estimation errors.

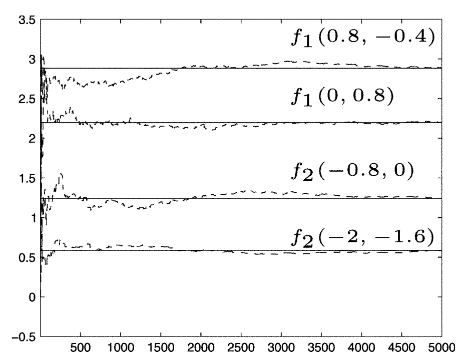

Fig. 6. Estimates for $\boldsymbol{f}$ at some particular points.

VI. Numerical Example

Let the linear subsystem be given as follows:

$$\boldsymbol{v}_k + A_1 \boldsymbol{v}_{k-1} + A_2 \boldsymbol{v}_{k-2} = B_1 \boldsymbol{u}_{k-1} + B_2 \boldsymbol{u}_{k-2}, \tag{47}$$

where $A_1 = \begin{pmatrix} 0 & 0.5 \\ -0.4 & 0 \end{pmatrix}$, $A_2 = \begin{pmatrix} 0.3 & 0 \\ 0 & -0.5 \end{pmatrix}$, $B_1 = \begin{pmatrix} 0 & \sqrt{0.85} \\ -\sqrt{0.6} & 0 \end{pmatrix}$, and $B_2 = \begin{pmatrix} -0.7 & 0 \\ 0 & 0.9 \end{pmatrix}$, and let the nonlinear function be given by $\boldsymbol{y}_k = \boldsymbol{f}(\boldsymbol{v}_k)$ with $\boldsymbol{f}(x,y) = \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix} = \begin{pmatrix} x + 0.3y^{2+2} \\ 0.3x^2 + y + 1 \end{pmatrix}$. Let the observation noise $\varepsilon_k \in \mathcal{N}(0, I_2)$ be i.i.d.. By H1 $\boldsymbol{u}_k \in \mathcal{N}(0, I_2)$ is i.i.d. and independent of $\{\varepsilon_k\}$.

It is noticed that P corresponding to the system (47) equals I, so the coefficients of the linear subsystem and the nonlinear function to be estimated coincide with those listed above.

The estimates of A_1, A_2, B_1, B_2 are given by Fig. 1–4, respectively, the estimated curve for $f_1(\cdot)$ is given by Fig. 5, where the lower surface represents the estimation errors, while the estimates for $f_1(\cdot)$ and $f_2(\cdot)$ at some particular points, namely, the estimates for $f_1(0, 0.8)$, $f_1(0.8, -0.4)$, and $f_2(-0.8, 0)$, $f_2(-2, -1.6)$ are presented in Fig. 6.

VII. Conclusion

In this technical brief, recursive estimation algorithms are designed for MIMO Wiener systems, which estimate both the matrix coefficients of the linear subsystems and the values of the nonlinear functions at any fixed points of interest. Under a set of reasonable conditions, all estimates are proved to be strongly consistent. It is noted that for the nonlinear function $\boldsymbol{f}(\boldsymbol{x})$, there is only a growth rate restriction as $\|\boldsymbol{x}\|$ tends to infinity. No other conditions like invertibility and expansion to a linear combination of basis functions are needed. The numerical simulation complements the theoretical analysis given in the technical brief. For further research it is of interest to consider identification of MIMO Wiener–Hammerstein systems and Hammerstein–Wiener systems.

Appendix

Proof of Lemma 2: Notice that $H(z) = A^{-1}(z)B(z) = (A^*(z)B(z)/a(z)) = (B^*(z)/a(z))$, where $a(z) = \det(A(z)) =$

$\sum_{i=0}^{np} a_i z^i$, $B^*(z) \triangleq A^*(z)B(z) = \sum_{j=1}^{(n-1)p+q} B_j^* z^j$, and $A^*(z)$ is the adjoint matrix of $A(z)$.

The linear subsystem (1) can be expressed as

$$\begin{aligned}\boldsymbol{v}_{k+1} + a_1\boldsymbol{v}_k + \cdots + a_{np}\boldsymbol{v}_{k-np+1} \\ = B_1^*\boldsymbol{u}_k + B_2^*\boldsymbol{u}_{k-1} + \cdots + B_{(n-1)p+q}^*\boldsymbol{u}_{k+1-[(n-1)p+q]}\end{aligned}$$

and hence

$$\begin{aligned}&E(\boldsymbol{v}_{k+1} + a_1\boldsymbol{v}_k + \cdots + a_{np}\boldsymbol{v}_{k-np+1})\boldsymbol{u}_{k+1-t}^T \\ &= E\bigg(B_1^*\boldsymbol{u}_k + B_2^*\boldsymbol{u}_{k-1} + \cdots \\ &\quad + B_{(n-1)p+q}^*\boldsymbol{u}_{k+1-[(n-1)p+q]}\bigg)\boldsymbol{u}_{k+1-t}^T \\ &= 0, \quad for \quad t > q + (n-1)p.\end{aligned}$$

Consequently, we have

$$H_t = -\sum_{i=1}^{np} a_i H_{t-i}, \quad for \quad t > q + (n-1)p. \tag{48}$$

If the matrix Γ were not of row-full-rank, then there would exist a vector $x = (x_1^T, \cdots, x_p^T)^T \neq 0$ with $x_i \in \mathbb{R}^n$ such that $x^T\Gamma = 0$, i.e.,

$$\sum_{j=1}^{p} x_j^T H_{q-j+l} = 0, 1 \leq l \leq np. \tag{49}$$

In this case we show that (49) holds $\forall l \geq 1$. Noticing (48) and (49), for $l = np + 1$, we have

$$\begin{aligned}\sum_{j=1}^{p} x_j^T H_{q-j+np+1} &= -\sum_{j=1}^{p} x_j^T \sum_{i=1}^{np} a_i H_{q-j+np+1-i} \\ &= -\sum_{i=1}^{np} a_i \sum_{j=1}^{p} x_j^T H_{q-j+np+1-i} = 0.\end{aligned} \tag{50}$$

Hence, (49) holds for $i = np+1$. Carrying out the similar treatment as that done in (50), we find

$$\sum_{j=1}^{p} x_j^T H_{q-j+l} = 0 \quad \forall l \geq 1. \tag{51}$$

Defining $d(z) \triangleq (\sum_{i=1}^{p} x_i z^{i-1})$, we have

$$\begin{aligned}d^T(z)H(z) &= \left(\sum_{i=1}^{p} x_i^T z^{i-1}\right) \cdot \left(\sum_{j=1}^{\infty} H_j z^j\right) \\ &= \sum_{i=1}^{p}\left(x_i^T z^{i-1}\left(\sum_{j=1}^{q-i} H_j z^j + \sum_{j=q-i+1}^{\infty} H_j z^j\right)\right) \\ &= \sum_{i=1}^{p}\sum_{j=1}^{q-i} x_i^T H_j z^{i+j-1} \triangleq -c^T(z).\end{aligned}$$

Consequently, $d^T(z)H(z) + c^T(z) = 0$ and the orders of $d(z)$ and $c(z)$ are strictly less than p and q, respectively. This, however, contradicts H2 by Lemma 1.

Therefore, the matrix Γ defined by (6) is of row-full-rank when H2 holds. ∎

References

[1] F. Giri and E. W. Bai, "Block-oriented Nonlinear System Identification," in *Lecture Notes in Control and Information Sciences*, M. Thoma, F. Allgöower, and M. Morari, Eds. Berlin, Germany: Springer, 2010.

[2] A. D. Kalafatis, L. Wang, and W. R. Cluett, "Identification of time-varying pH processes using sinusoidal signals," *Automatica*, vol. 41, no. 4, pp. 685–691, 2005.

[3] S. Boyd and L. O. Chua, "Fading memory and the problem of approximating nonlinear operators with volterra series," *IEEE Trans. Circuits Syst.*, vol. 32, no. 11, pp. 1150–1161, Nov. 1985.

[4] D. Westwick and M. Verhaegen, "Identifying MIMO Wiener systems using subspace model identification methods," *Signal Process.*, vol. 52, no. 2, pp. 235–258, 1996.

[5] J. J. Bussgang, Crosscorrelation functions of amplitude-distorted Gaussian signals. MIT Electrical Research Lab, Tech. Rep. 216, 1952.

[6] J. C. Gömez and E. Baeyens, "Subspace identification of multivariable Hammerstein and Wiener models," *Eur. J. Control*, vol. 11, no. 2, pp. 1–10, 2005.

[7] X. L. Hu and H. F. Chen, "Identification for Wiener systems with RTF subsystems," *Eur. J. Control*, vol. 12, no. 6, pp. 581–594, 2006.

[8] W. X. Zhao and H. F. Chen, "Markov chain approach to identifying Wiener systems," *Sci. China: Inform. Sci.*, vol. 55, no. 5, pp. 1201–1217, 2012.

[9] X. L. Hu and H. F. Chen, "Strong consistency of recursive identification for Wiener systems," *Automatica*, vol. 41, no. 11, pp. 1905–1916, 2005.

[10] Q. J. Song and H. F. Chen, "Identification for wiener systems with internal noise," *J. Syst. Sci. Complex.*, vol. 21, no. 3, pp. 378–393, 2008.

[11] P. O. Arambel and G. Tadmor, "Identifiability and persistent excitation in full matrix fraction parameter estimation," *Automatica*, vol. 33, no. 4, pp. 689–692, 1997.

[12] A. M. Walker, "Large-sample estimation of parameters for autoregressive process with moving-average residuals," *Biometrika*, vol. 49, no. 1, pp. 117–131, 1962.

[13] J. Q. Fan and Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Approach*. New York: Springer–Verlag, 2003.

[14] T. D. Pham and L. T. Tran, "Some mixing properties of time series models," *Stochastic Process. Appl.*, vol. 19, no. 2, pp. 297–303, 1985.

[15] W. X. Zhao, H. F. Chen, and W. X. Zheng, "Recursive identification for nonlinear ARX systems based on stochastic approximation algorithm," *IEEE Trans. Autom. Control*, vol. 55, no. 6, pp. 1287–1299, Jun. 2010.

[16] H. F. Chen, *Stochastic Approximation and Its Applications*. Dordrecht, The Netherlands: Kluwer, 2002.

[17] T. Falck, P. Dreesen, K. D. Brabanter, K. Pelckmans, B. D. Moor, and J. A. K. Suykens, "Least-squares support vector machines for the identification of Wiener–Hammerstein systems,," *Control Eng. Practice*, vol. 20, no. 11, pp. 1165–1174, 2012.

[18] Y. S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*. New York: Springer–Verlag, 1978.