Recursive Identification of Multi-Input Multi-Output Errors-in-Variables Hammerstein Systems

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

***Abstract*—The note considers the identification of multi-input multi-output errors-in-variables Hammerstein systems, in which both the input and output can be observed but with additive noises being ARMA processes with unknown coefficients. With the help of stochastic approximation combined with the deconvolution kernel function, the recursive algorithms are proposed for estimating coefficients of the linear subsystem and for the values of the nonlinear function. Under some reasonable conditions, all the estimates are proved to converge to the true values with probability one. These results include identification of the errors-in-variables linear systems as a special case. A simulation example is given justifying the theoretical analysis.**

***Index Terms*—α-mixing, errors-in-variables, Hammetstein systems, recursive estimation, stochastic approximation, strong consistency.**

I. Introduction

The Hammerstein system composed of a static nonlinear function followed by a dynamic linear subsystem can be thought of the simplest block-oriented nonlinear systems [1], but many practical systems, for example, the distillation column in chemical engineering [2], the power amplifier in electronic circuits [3], and solid oxide fuel cells [4] and so on, can be modeled as a Hammerstein system. It is natural that the identification of Hammerstein systems has received a considerable attraction from both the theoretical researches and engineers.

To identify the nonlinear function in a Hammerstein system there are parametric [3], [5], [6], and nonparametric approaches [7]–[9] 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, and neural networks with unknown coefficients. In this case all parameters can be estimated by the standard optimization methods. The nonparametric approach, which requires no structure information about the nonlinearity, is carried out by two steps: First, the parameters of the linear part are estimated, then the value of the nonlinear function of the Hammerstein system at any value of its argument is estimated with the help of the kernel functions. The estimation is based on the observed input-output data, which are the true input-output data corrupted by noises, i.e., we identify the multi-input multi-output (MIMO) errors-in-variables (EIV) Hammerstein systems.

Manuscript received September 24, 2012; revised March 11, 2013, November 20, 2013, April 2, 2014, and July 31, 2014; accepted August 6, 2014. Date of publication August 12, 2014; date of current version February 19, 2015. This work was supported by the National Key Basic Research Program of China (973 program) under Grant No. 2014CB845301, the National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, and the NSF of China under Grants No. 61273193, 61120106011, and 61134013. Recommended by Associate Editor L. Giarré.

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, China (e-mail: bqmu@amss.ac.cn; hfchen@iss.ac.cn).

Digital Object Identifier 10.1109/TAC.2014.2346871

The noises in Hammerstein systems may be given by the probabilistic description [5], [9], [10] or by the set-membership description [11], [12]. In the note, we adopt the nonparametric method and the probabilistic description for the noises.

If the existing estimation algorithms [9], [13] are applied to the EIV Hammerstein systems, then the resulting estimates may be inconsistent. This is seen from the simulation example given in Section VI. In this note, the consistent estimates are derived by using the stochastic approximation algorithm with expanding truncations (SAAWET) [14], modified from the well-known Robbins-Monro algorithm [15], incorporated with the deconvolution kernel functions [16], [17]. Moreover, the conditions used here on the linear part are weaker than those imposed in [9], and the corresponding estimation algorithms for the linear part are simpler in comparison with [9]. Since many parameter estimation problems can be converted to root-seeking problems by suitably choosing the regression functions, SAAWET is an appropriate tool by its simplicity of recursion and convenience of convergence analysis. The deconvolution kernel function is a commonly used technique in the statistical literature [16], [17] for EIV nonlinear systems.

Various estimation methods for identifying the linear EIV systems are well summarized in the survey paper [18], while the identifiability issue of EIV systems is discussed in [19]. It is noted that the identification methods mentioned in [18] mainly are nonrecursive. Recursive identification for linear EIV systems is considered and the strong consistency of estimates is derived in [20], [21]. There are also a few papers on identification of the nonlinear EIV systems [22], [23] among others.

In the note the MIMO EIV Hammerstein system is described as follows:

$$v_k^0 = f(u_k^0) + \omega_k \tag{1}$$

$$A(z^{-1})y_k^0 = B(z^{-1})v_k^0 + \xi_k \tag{2}$$

where $A(z^{-1}) = I + A_1z^{-1} + \cdots + A_pz^{-p}$ and $B(z) = B_1z^{-1} + B_2z^{-2} + \cdots + B_qz^{-q}$ are the $n \times n$ and $n \times m$ matrix polynomials with unknown coefficients but with known orders p, q, respectively, while z^{-1} is the backward shift operator: $z^{-1}y_k = y_{k-1}$. Further, $u_k^0 \in \mathbb{R}^m$ and $y_k^0 \in \mathbb{R}^n$ represent the noise-free system input and output, respectively, and $\omega_k \in \mathbb{R}^m$ and $\xi_k \in \mathbb{R}^n$ are the system noises. The nonlinearity $f(\cdot)$ is a vector-valued function: $f : \mathbb{R}^m \to \mathbb{R}^m$. The system input and output are observed with additive noises η_k and ε_k:

$$u_k = u_k^0 + \eta_k, y_k = y_k^0 + \varepsilon_k. \tag{3}$$

The goal of the note is to recursively estimate the unknown matrix coefficients $\{A_1, \ldots, A_p, B_1, \ldots, B_q\}$ of the linear part and the value of $f(x)$ at any given x in its domain based on the observed data $\{u_k, y_k\}$.

The rest of the note is arranged as follows. The system assumptions and the recursive algorithms are given in Section II. Some auxiliary results on α-mixing are proposed in Section III. The strong consistency of the estimates for linear subsystem and for the nonlinearity is proved in Sections IV and V, respectively. A numerical example is presented in Section VI, and a brief conclusion is given in Section VII.

II. Assumptions and Recursive Identification Algorithms

A. Assumptions

In the sequel, by $\| * \|$ denote the Euclidean norm for a vector $*$ and the Frobenius norm for a matrix $*$, and by I denote the identity matrix.

We now give the assumptions used in the note.

H1 The noise-free system input $\{u_k^0\}$ is a sequence of independent and identically distributed (iid) random vectors with zero mean and known covariance matrix $\Lambda \triangleq E u_k^0 u_k^{0T}$. Furthermore, $\{u_k^0\}$ has a density function $p(\cdot)$ in a bounded subset U of $\mathbb{R}^m$.

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

H3 The measurement noises η_k and ε_k are both ARMA processes:

$$P(z^{-1})\eta_k = Q(z^{-1})\zeta_k \tag{4}$$

$$R(z^{-1})\varepsilon_k = S(z^{-1})\varsigma_k \tag{5}$$

where

$$P(z^{-1}) = I + P_1 z^{-1} + P_2 z^{-2} + \cdots + P_{n_p} z^{-n_p} \tag{6}$$

$$Q(z^{-1}) = I + Q_1 z^{-1} + Q_2 z^{-2} + \cdots + Q_{n_q} z^{-n_q} \tag{7}$$

$$R(z^{-1}) = I + R_1 z^{-1} + R_2 z^{-2} + \cdots + R_{n_r} z^{-n_r} \tag{8}$$

$$S(z^{-1}) = I + S_1 z^{-1} + S_2 z^{-2} + \cdots + S_{n_s} z^{-n_s} \tag{9}$$

where $P(z^{-1})$ and $R(z^{-1})$ are both stable. The driven noises $\{\zeta_k\}$, $\{\varsigma_k\}$ and the internal noises $\{\xi_k\}$, $\{\omega_k\}$ are all mutually independent sequences of iid zero mean random vectors having probability densities. Moreover, ζ_k, ς_k, ξ_k and ω_k are independent of u_k^0, and $E(\|\zeta_k\|^{\Delta-1}) < \infty$, $E(\|\xi_k\|^{\Delta}) < \infty$, $E(\|\omega_k\|^{\Delta}) < \infty$, and $E(\|\varsigma_k\|^{\Delta}) < \infty$ for some $\Delta > 3$.

H4 The function $f(\cdot)$ is measurable, locally bounded and continuous at x where $f(x)$ is estimated. Further, the input-output-correlation-matrix of the nonlinearity $\Upsilon \triangleq E(f(u_k^0)u_k^{0T})$ is nonsingular.

H5 $zB(z^{-1})$ is of column-full-rank for $z^{-1} = 0,1$.

H6 The driven noise $\{\zeta_k\}$ in (4) is a sequence of iid zero mean Gaussian random vectors.

Before proceeding further, let us explain these assumptions. To identify the linear system it is sufficient to require H1-H4, while to estimate $f(\cdot)$ we have to additionally impose H5-H6. For a practical system, the iid assumption on the input in H1 may be a restriction, but which is used to derive the simple relationship connecting the observed input and output with the impulse responses of the linear part. Assumption H2 is the necessary and sufficient condition for identifiability of the multivariable linear subsystem. In H3, it is assumed that the input and output measurement noises $\{\eta_k\}$ and $\{\varepsilon_k\}$ are ARMA processes, which can characterize most noises, but the internal noises $\{\xi_k\}$ and $\{\omega_k\}$ may or may not exist. Assumption H5 makes it possible to estimate the nonlinearity $\{f(\cdot)\}$ by the observed $\{y_k\}$. To impose the Gaussian assumption on $\{\zeta_k\}$ in H6 it is just to make the deconvolution kernel functions $w_k(x)$ (see (20)) simpler. The deconvolution kernel function $w_k(x)$ requires the covariance of the input noise $\{\eta_k\}$ be known, but which is impossible to achieve only by using the observed input $\{u_k\}$ if no additional condition is imposed. Because of this, we require the covariance Λ of $\{u_k^0\}$ be available. It is worth noting that the availability of Λ required in H1 is needed for identifying $f(\cdot)$ only, and it is not required when identifying the linear part as to be seen later on.

Remark 1: As pointed out in [9], $A^{-1}(z^{-1})B(z^{-1})f(u_k^0)$ and $A^{-1}(z^{-1})(B(z^{-1})P)(P^{-1}f(u_k^0))$ for any nonsingular matrix P produce the same output under the same input $\{u_k^0\}$. So, to determine the system, the indeterminate matrix P has to be fixed. Let us choose $P \triangleq \Upsilon$. Then the correlation matrix between the input and output of the nonlinearity for the new system $\widetilde{\Upsilon} = E(\Upsilon^{-1} f(u_k^0)u_k^{0T}) = \Upsilon^{-1}\Upsilon = I$ under H4. Therefore, without loss of generality, we may assume that the correlation matrix between the input and output of the nonlinearity, $\Upsilon = I$ under H4.

B. Estimation of the Linear Part

Since $A(z^{-1})$ is stable by H2, we have

$$H(z^{-1}) \triangleq A^{-1}(z^{-1})B(z^{-1}) = \sum_{i=1}^{\infty} H_i z^{-i} \tag{10}$$

where $\|H_i\| = O(e^{-ri})$, $r > 0$, $i > 1$ and $H_1 = B_1$.

Defining $\bar{\xi}_k \triangleq A^{-1}(z^{-1})\xi_k$ and assuming $u_k^0 = 0 \ \forall\, k < 0$, we have

$$y_k^0 = \sum_{i=1}^{k} H_i f(u_{k-i}^0) + \sum_{i=1}^{k} H_i \omega_{k-i} + \bar{\xi}_k. \tag{11}$$

The following lemma may not be new but is very essential for identifying the linear part.

Lemma 1: Assume H1–H4 hold but without need for availability of Λ in H1. The following formulas take place

$$E y_k u_{k-i}^T = H_i, \quad for\ i \geq 1. \tag{12}$$

Proof: Since ξ_k, ζ_k, ς_k, ω_k, and u_k^0 are zero mean and mutually independent, it follows that:

$$\begin{aligned} E\left(y_k u_{k-i}^T\right) &= E\left(y_k^0 + \varepsilon_k\right)\left(u_{k-i}^0 + \eta_{k-i}\right)^T = E y_k^0 u_{k-i}^{0T} \\ &= E\left(\sum_{j=1}^{k} H_j v_{k-j}^0 + \bar{\xi}_k\right) u_{k-i}^{0T} = \sum_{j=1}^{k} H_j E\left(f\left(u_{k-j}^0\right) u_{k-i}^{0T}\right) \\ &= H_i E\left(f\left(u_{k-i}^0\right) u_{k-i}^{0T}\right) = H_i \Upsilon = H_i,\ \forall\, i \geq 1. \qquad \blacksquare \end{aligned}$$

Similar to [24], [25], we first estimate the impulse responses $\{H_i, i \geq 1\}$ and then obtain the estimates for the coefficients $\{A_1, \ldots, A_p, B_1, \ldots, B_q\}$ by the deconvolution relationship between them. Motivated by (12), the impulse responses $\{H_i, i \geq 1\}$ are recursively estimated by applying SAAWET:

$$\begin{aligned} H_{i,k} = &\left[H_{i,k-1} - \frac{1}{k}\left(H_{i,k-1} - y_k u_{k-i}^T\right)\right] \\ &\cdot I_{\left[\left\|H_{i,k-1} - \frac{1}{k}\left(H_{i,k-1} - y_k u_{k-i}^T\right))\right\| \leq M_{\delta_{i,k}}\right]} \end{aligned} \tag{13}$$

$$\delta_{i,k} = \sum_{j=1}^{k-1} I_{\left[\left\|H_{i,j-1} - \frac{1}{j}\left(H_{i,j-1} - y_j u_{j-i}^T\right))\right\| > M_{\delta_{i,j}}\right]} \tag{14}$$

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, while the estimates for $\{A_1, \ldots, A_p, B_1, \ldots, B_q\}$ are obtained as follows:

$$\begin{aligned} [A_{1,k}, A_{2,k}, \ldots, A_{p,k}] = &-[H_{q+1,k}, H_{q+2,k}, \cdots, H_{q+np,k}] \\ &\times \Gamma_k^T \left(\Gamma_k \Gamma_k^T\right)^{-1} \end{aligned} \tag{15}$$

$$B_{i,k} = \sum_{j=0}^{i \wedge p} A_{j,k} H_{i-j,k},\ \forall\, 1 \leq i \leq q \tag{16}$$

where

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

with $H_{i,k} = 0$ for $i \leq 0$ is of row-full-rank when k is sufficiently large since $H_{i,k} \xrightarrow[k\to\infty]{} H_i$ a.s. as to be shown by Theorem 1 and the limit of Γ_k is of row-full-rank under H2 [24], [25].

C. Nonparametric Estimation of $f(\cdot)$

We now recursively estimate $f(x)$ by the observed input $\{u_k\}$ and output $\{y_k\}$, where x is an arbitrary point in the domain U. Since $\{u_k^0\}$ is corrupted by the noise $\{\eta_k\}$, the conventional kernel function leads to a biased estimate. To remove the influence of the input noise $\{\eta_k\}$, we use the deconvolution kernel functions [16], [17] to be defined in the sequel.

A conventional kernel function $K(x)$ is formally related with its Fourier transformation $\Phi_K(t) \triangleq \int_{\mathbb{R}^m} e^{it^T x} K(x)\mathrm{d}x$ by

$$K(x) = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} e^{-it^T x} \Phi_K(t)\mathrm{d}t \tag{17}$$

where $x = [x_1, \ldots, x_m]^T$, $t = [t_1, \ldots, t_m]^T$, and i stands for the imaginary unit satisfying $i^2 = -1$.

Dividing the integrand of (17) by the characteristic function of the input noise $\{\eta_k\}$ leads to the following function

$$K_k(x) \triangleq \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} e^{-it^T x} \frac{\Phi_K(t)}{\Phi_{\eta_k}\left(\frac{t}{b_k}\right)} \mathrm{d}t \tag{18}$$

where $\Phi_{\eta_k}(t) = \int_{\mathbb{R}^m} e^{it^T x}(1/((2\pi)^{m/2}|\Sigma|^{1/2}))e^{-x^T\Sigma^{-1}x/2}\mathrm{d}t = e^{-(t^T\Sigma t/2)}$ is the characteristic function of $\{\eta_k\}$, $\Sigma \triangleq E(\eta_k\eta_k^T)$, and $b_k = (b\lambda_{\max}(\Sigma)/\log k)^{1/2}$ is the bandwidth with $b \geq 3$ m being a constant, and $\lambda_{\max}(\Sigma)$ denotes the maximum eigenvalue of Σ.

Finally, the deconvolution kernel function is defined as follows:

$$w_k(x) \triangleq \frac{1}{b_k^m} K_k\left(\frac{u_k - x}{b_k}\right). \tag{19}$$

The Sinc function [16] (i.e., $K(x) = \prod_{j=1}^m (\sin(x_j)/\pi x_j)$) is chosen to serve as the function $K(x)$ because its Fourier transformation is very simple: $\Phi_K(t) \triangleq \int_{\mathbb{R}^m} e^{it^T x} K(x)\mathrm{d}x = \prod_{j=1}^m I_{[|t_j|\leq 1]}$. Thus, the resulting deconvolution kernel function $w_k(x)$ is given by

$$\begin{aligned} w_k(x) &= \frac{1}{b_k^m} K_k\left(\frac{u_k - x}{b_k}\right) \\ &= \frac{1}{(2\pi b_k)^m} \int_G e^{\left[-it^T(u_k - x)/b_k\right]} e^{\frac{t^T\Sigma t}{2b_k^2}} \mathrm{d}t \\ &= \frac{1}{(2\pi b_k)^m} \int_G \cos\left(\frac{t^T(u_k - x)}{b_k}\right) e^{\frac{t^T\Sigma t}{2b_k^2}} \mathrm{d}t \end{aligned} \tag{20}$$

where $G \triangleq [-1, 1]^m$.

Remark 2: Based on the limits to be given in (41), we can estimate the density function $p(x)$ of $\{u_k^0\}$ and $p(x)f(x)$, respectively, and then obtain the estimate for the nonlinear function $f(\cdot)$. The sinc function is used to guarantee the finiteness of the integrals (18) and (20), since the support of its Fourier transformation is the bounded set $[-1, 1]^m$. Actually, any function whose Fourier transformation has a bounded support can be used to replace the sinc function. Since the characteristic function of zero equals 1, $K_k(x)$ turns to be the sinc function if the input noise $\{\eta_k\}$ vanishes. In this case the deconvolution kernel function becomes the conventional kernel function. Therefore, the deconvolution kernel functions are the extension of the conventional kernel functions.

Since the covariance Σ of η_k in (20) is unknown, we apply the following SAAWET to recursively estimate Σ:

$$\Sigma_k = \left[\Sigma_{k-1} - \frac{1}{k}\left(\Sigma_{k-1} + \Lambda - u_k u_k^T\right)\right] \cdot I_{\left[\left|\Sigma_{k-1} - \frac{1}{k}\left(\Sigma_{k-1} + \Lambda - u_k u_k^T\right)\right| \leq M_{\phi_k}\right]} \tag{21}$$

$$\phi_k = \sum_{j=1}^{k-1} I_{\left[\left|\Sigma_{j-1} - \frac{1}{j}\left(\Sigma_{j-1} + \Lambda - u_j u_j^T\right)\right| > M_{\phi_j}\right]} \tag{22}$$

where Σ_k denotes the estimate for Σ at time k.

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

$$\widehat{w}_k(x) \triangleq \frac{1}{(2\pi\widehat{b}_k)^m} \int_G \cos\left(\frac{t^T(u_k - x)}{\widehat{b}_k}\right) e^{\frac{t^T\Sigma_k t}{2\widehat{b}_k^2}} \mathrm{d}t \tag{23}$$

where $\widehat{b}_k = (b\lambda_{\max}(\Sigma_k)/\log k)^{1/2}$.

We now estimate $\chi_k \triangleq B_1^\dagger A(z^{-1})y_{k+1}$, where $B_1^\dagger \triangleq (B_1^T B_1)^{-1}B_1^T$ denotes the pseudo-inverse of B_1. Then by (1)–(3), we have

$$\begin{aligned} \chi_k = f\left(u_k^0\right) + B_1^\dagger \sum_{j=2}^q B_j f\left(u_{k-j+1}^0\right) + B_1^\dagger B(z^{-1})\omega_{k+1} \\ + B_1^\dagger \xi_{k+1} + B_1^\dagger A(z^{-1})\varepsilon_{k+1}. \end{aligned} \tag{24}$$

By (15) and (16), it is natural to estimate χ_k by

$$\widehat{\chi}_k = B_{1,k}^\dagger \sum_{j=0}^p A_{j,k} y_{k+1-j}. \tag{25}$$

Since u_k is independent of $\{u_j^0, j < k\}$, we have

$$E(w_k(x)\chi_k) = E\left(w_k(x)f\left(u_k^0\right)\right) + (Ew_k(x))B_1^\dagger \sum_{j=2}^q B_j\left(Ef\left(u_k^0\right)\right).$$

Then, by (41) to be given in Lemma 6, it follows that:

$$E(w_k(x)\chi_k) \xrightarrow[k\to\infty]{} p(x)\left(f(x) + B_1^\dagger \sum_{j=2}^q B_j Ef\left(u_k^0\right)\right) \triangleq h(x). \tag{26}$$

Based on the limits $Ew_k(x) \xrightarrow[k\to\infty]{} p(x)$ and (26), we use the following algorithms to estimate $p(x)$ and $h(x)$, respectively:

$$\tau_k(x) = \left[\tau_{k-1}(x) - \frac{1}{k}\left(\tau_{k-1}(x) - \widehat{w}_k(x)\right)\right] \cdot I_{\left[\left|\tau_{k-1}(x) - \frac{1}{k}\left(\tau_{k-1}(x) - \widehat{w}_k(x)\right)\right| \leq M_{\delta_k^{(\tau)}(x)}\right]} \tag{27}$$

$$\delta_k^{(\tau)}(x) = \sum_{j=1}^{k-1} I_{\left[\left|\tau_{j-1}(x) - \frac{1}{j}\left(\tau_{j-1}(x) - \widehat{w}_j(x)\right)\right| > M_{\delta_j^{(\tau)}(x)}\right]} \tag{28}$$

$$\beta_k(x) = \left[\beta_{k-1}(x) - \frac{1}{k}\left(\beta_{k-1}(x) - \widehat{w}_k(x)\widehat{\chi}_k\right)\right] \cdot I_{\left[\left\|\beta_{k-1}(x) - \frac{1}{k}\left(\beta_{k-1}(x) - \widehat{w}_k(x)\widehat{\chi}_k\right)\right\| \leq M_{\delta_k^{(\beta)}(x)}\right]} \tag{29}$$

$$\delta_k^{(\beta)}(x) = \sum_{j=1}^{k-1} I_{\left[\left\|\beta_{j-1}(x) - \frac{1}{j}\left(\beta_{j-1}(x) - \widehat{w}_j(x)\widehat{\chi}_j\right)\right\| > M_{\delta_j^{(\beta)}(x)}\right]}. \tag{30}$$

Since $h(x)/p(x)=f(x)+B_1^\dagger\sum_{j=2}^q B_jEf(u_k^0)$ and $\{B_j, j=1\cdots,m\}$ are estimated by (16), to estimate $f(x)$ it suffices to obtain an estimate for $Ef(u_k^0)$. By stability of $A(z^{-1})$, $\{y_k^0\}$ is asymptotically stationary, and we have $\mu^{(y)}\triangleq\lim_{k\to\infty}Ey_k$. Set $\mu^{(f)}\triangleq Ef(u_k^0)$. By (2) it follows that $\sum_{j=0}^p A_j\mu^{(y)}=\sum_{i=1}^q B_i\mu^{(f)}$, which yields $\mu^{(f)}=\left(\sum_{i=1}^q B_i\right)^\dagger\sum_{j=0}^p A_j\mu^{(y)}$ by H5. Therefore, we can estimate $\mu^{(f)}$ at time k by

$$\mu_k^{(f)}=\left(\sum_{i=1}^q B_{i,k}\right)^\dagger\sum_{j=0}^p A_{j,k}\mu_k^{(y)} \tag{31}$$

where $\mu_k^{(y)}$ is recursively estimated by

$$\mu_k^{(y)}=\left[\mu_{k-1}^{(y)}-\frac{1}{k}\left(\mu_{k-1}^{(y)}-y_k\right)\right]\cdot I_{\left[\left\|\mu_{k-1}^{(y)}-\frac{1}{k}\left(\mu_{k-1}^{(y)}-y_k\right)\right\|\le M_{\delta_k^{(y)}}\right]} \tag{32}$$

$$\delta_k^{(y)}=\sum_{j=1}^{k-1}I_{\left[\left\|\mu_{j-1}^{(y)}-\frac{1}{j}\left(\mu_{j-1}^{(y)}-y_j\right)\right\|>M_{\delta_j^{(y)}}\right]}. \tag{33}$$

Finally, the estimate for $f(x)$ is given by

$$f_k(x)\triangleq\frac{\beta_k(x)}{\tau_k(x)}-B_{1,k}^\dagger\sum_{j=2}^q B_{j,k}\mu_k^{(f)}. \tag{34}$$

III. Auxiliary Results on Weak Dependence

In order to establish the strong consistency of the estimates given in Section II, one has to verify the conditions for convergence of SAAWET. For this the basic step often consists in verifying convergence of some series of random variables forming such as the martingale difference sequence (mds), α-mixing, and others.

We first define the α-mixing [26] and then introduce the convergence results for α-mixing [24]. For a process $\{X_k, k=0,1,\cdots\}$, we denote by $\mathcal{F}_i^j$ the σ-algebra generated by $\{X_s, 0\le i\le s\le j\}$. Define

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

The process $\{X_k\}$ is called the α-mixing if $\alpha_k\xrightarrow[k\to\infty]{}0$, and the numbers α_k are called the mixing coefficients of $\{X_k\}$. It is worth noting that the mixing property is hereditary [26] in the sense that the process $\{g(X_k)\}$ for any measurable function $g(\cdot)$ possesses the mixing property of $\{X_k\}$. The next lemma indicates that an asymptotically stable ARMA process is an α-mixing.

Lemma 2. ([27]): Let $\{X_k\}$ be an asymptotically stable vector-valued ARMA process whose driven noise is a sequence of iid random vectors with density and has a bounded moment of order >2. Then $\{X_k\}$ is an α-mixing with mixing coefficients decaying exponentially to zero.

Lemma 3. ([28]): Let $\{\varpi_k\}$ be a sequence of random vectors with $\sup_k E\|\varpi_k\|^\delta<\infty$ for some $\delta\ge 2$, and let L_i with $\|L_i\|=O(e^{-ri})$, $r>0$ be a sequence of real matrices. 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. ([24]): Let $\{X_k,\mathcal{F}_k\}$ be a zero mean α-mixing with mixing coefficient α_k exponentially decaying to zero and $\sum_{k=1}^\infty(E|X_k|^{2+\epsilon})^{(2/2+\epsilon)}<\infty$ for some $\epsilon>0$. Then $\sum_{k=1}^\infty X_k<\infty$ *a.s.*

Remark 3: Since both the iid sequence and the mds are the α-mixings with mixing coefficients exponentially decaying to zero, the convergence of their sums can also be established by Lemma 4. According to Lemma 2 and the heredity of the mixing processes, we see that all signals appearing in the proof presented below are the α-mixings with mixing coefficients exponentially decaying to zero. Therefore, to assure the convergence of the corresponding random series, one only needs to verify the convergence condition moments in Lemma 4, while this can be derived by Lemma 3. Since the similar convergence analysis for SAAWET has been elaborately described in the papers [9], [24], [28], here we only outline the proof.

IV. Strong Consistency of Estimates for Linear Part

Lemma 5: Assume H1-H4 hold but without need for availability of Λ in H1. Then, for any $0\le\nu<1/2$, the following series converge:

$$\sum_{k=1}^\infty\frac{1}{k^{1-\nu}}\left(E\left(y_ku_{k-i}^T\right)-y_ku_{k-i}^T\right)<\infty\ a.s.\ for\ i\ge 1. \tag{35}$$

Proof: Noticing u_{k-i}^0, η_{k-i}, $\bar{\xi}_k$, ω_k, and ε_k are mutually independent, by (11) we have

$$\begin{aligned}\sum_{k=1}^\infty\frac{1}{k^{1-\nu}}\left(E\left(y_ku_{k-i}^T\right)-y_ku_{k-i}^T\right)&=-\sum_{k=1}^\infty\frac{1}{k^{1-\nu}}\left(\bar{\xi}_ku_{k-i}^{0T}\right)\\ &-\sum_{k=1}^\infty\frac{1}{k^{1-\nu}}\left(\bar{\xi}_k\eta_{k-i}^T\right)-\sum_{k=1}^\infty\frac{1}{k^{1-\nu}}\left(\varepsilon_ku_{k-i}^{0T}\right)\\ &-\sum_{k=1}^\infty\frac{1}{k^{1-\nu}}\left(\varepsilon_k\eta_{k-i}^T\right)-\sum_{k=1}^\infty\frac{1}{k^{1-\nu}}\left(\sum_{\substack{j=1\\ j\ne i}}^k H_jf\left(u_{k-j}^0\right)u_{k-i}^{0T}\right)\\ &-\sum_{k=1}^\infty\frac{1}{k^{1-\nu}}\left(\sum_{j=1}^k H_j\omega_{k-j}\eta_{k-i}^T\right)\\ &-\sum_{k=1}^\infty\frac{1}{k^{1-\nu}}\left(\sum_{j=1}^k H_j\omega_{k-j}u_{k-i}^{0T}\right)\\ &-\sum_{k=1}^\infty\frac{1}{k^{1-\nu}}\left(\sum_{j=1}^k H_jf\left(u_{k-j}^0\right)\eta_{k-i}^T\right)\\ &+H_i\sum_{k=1}^\infty\frac{1}{k^{1-\nu}}\left(Ef\left(u_{k-i}^0\right)u_{k-i}^{0T}-f\left(u_{k-i}^0\right)u_{k-i}^{0T}\right).\end{aligned} \tag{36}$$

Since each term on the right-hand side of (36) is an α-mixing with mixing coefficients exponentially decaying to zero, we see that (36) converges by Lemma 4 and Remark 3, and hence (35) follows. □

Theorem 1: Assume H1-H4 hold but without need for availability of Λ in H1. Then, $H_{i,k}$, $i\ge 1$, defined by (13), (14) have the convergence rates

$$\|H_{i,k}-H_i\|=o(k^{-\nu})\quad a.s.\quad i\ge 1\quad\forall\nu\in\left(0,\frac{1}{2}\right). \tag{37}$$

As consequences, the following convergence rates also take place:

$$\|A_{i,k}-A_i\|=o(k^{-\nu})\ a.s.\ 1\le i\le p\ \ \forall\nu\in\left(0,\frac{1}{2}\right) \tag{38}$$

$$\|B_{j,k}-B_j\|=o(k^{-\nu})\ a.s.\ 1\le j\le q\ \ \forall\nu\in\left(0,\frac{1}{2}\right). \tag{39}$$

Proof: We rewrite (13) as

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

where $e_{i,k} = H_i - y_k u_{k-i}^T = [Ey_k u_{k-i}^T - y_k u_{k-i}^T]$.

Since H_i is the single root of the linear function $-(y - H_i)$, by Theorem 3.1.1 in [14], the convergence rate theorem of SAAWET, it suffices to prove

$$\sum_{k=1}^{\infty} \frac{1}{k^{1-\nu}} e_{i,k} < \infty \quad a.s. \quad i \ge 1 \quad \forall \nu \in \left(0, \frac{1}{2}\right). \tag{40}$$

By Lemma 5, we find that (40) is true, and hence (37) holds, while (38) and (39) are derived directly from (37) by (15), (16). □

V. Strong Consistency of Estimates for $f(\cdot)$

The limits (41) to be given in Lemma 6 reveal the relationship between the deconvolution kernel functions $w_k(x)$ and the nonlinear function $f(\cdot)$, and they indicate that the deconvolution kernel functions can really depress the effect of the input noise $\{\eta_k\}$. Namely based on them the SAAWET (27)–(34) are designed to estimate the nonlinear function $f(\cdot)$.

Lemma 6: Under H1, H3, H4, and H6, for $w_k(x)$ defined by (20) the following assertions take place:

$$E[w_k(x)] \underset{k\to\infty}{\longrightarrow} p(x), E\left[w_k(x) f\left(u_k^0\right)\right] \xrightarrow[k\to\infty]{} p(x)f(x) \tag{41}$$

$$\sum_{k=1}^{\infty} \frac{1}{k}\left(Ew_k(x) - w_k(x)\right) < \infty \tag{42}$$

where $p(x)$ is the density function of $\{u_k^0\}$.

Proof: By the Fubini theorem changing the order of taking expectation and integral, and noticing that the density function of η_k is even, we have

$$Ew_k(x) = \frac{1}{(2\pi b_k)^m} \int_{\mathbb{R}^m} E e^{\left[-it^T(u_k - x)/b_k\right]} \frac{\Phi_K(t)}{\Phi_{\eta_k}\left(\frac{t}{b_k}\right)} \mathrm{d}t$$

$$= \frac{1}{(2\pi b_k)^m} \int_{\mathbb{R}^m} E e^{\left[-it^T\left(u_k^0 - x\right)/b_k\right]} \Phi_K(t) \mathrm{d}t$$

$$= \frac{1}{b_k^m} \int_U \left(\frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} e^{\left[-it^T(y-x)/b_k\right]} \Phi_K(t)\mathrm{d}t\right) p(y)\mathrm{d}y$$

$$= \frac{1}{b_k^m} \int_U K\left(\frac{y-x}{b_k}\right) p(y)\mathrm{d}y = \int_{U_k} K(t)p(x + b_k t)\mathrm{d}t \xrightarrow[k\to\infty]{} p(x)$$

where U_k tends to $\mathbb{R}^m$ as $k \to \infty$. Likewise, we can prove the second limit of (41), while (42) can be shown by Lemma 4 since $w_k(x)$ is α-mixing with mixing coefficients exponentially decaying to zero. □

Lemma 7: Assume that H1, H3, H4, and H6 hold. Then Σ_k given by (21) and (22) and $\widehat{w}_k(x)$ given by (23) have the convergence rates

$$\|\Sigma_k - \Sigma\| = o\left(\frac{1}{k^{1/2-c}}\right) \quad \forall c > 0 \tag{43}$$

$$|w_k(x) - \widehat{w}_k(x)| = o\left(\frac{(\log)^{\frac{m}{2}+1}}{k^{\frac{1}{2}-\frac{m}{2b}-c}}\right) \quad \forall c > 0. \tag{44}$$

Proof: The proof of (43) is similar to that of Theorem 1, so we omit it here. According to (20) and (23), we have $w_k(x) - \widehat{w}_k(x) = I_1 + I_2 + I_3$, where

$$I_1 = \frac{1}{(2\pi)^m}\left(\frac{1}{b_k^m} - \frac{1}{\widehat{b}_k^m}\right) \int_G \cos\left(\frac{t^T(u_k - x)}{b_k}\right) e^{\frac{t^T\Sigma t}{2b_k^2}} \mathrm{d}t,$$

$$I_2 = \frac{1}{(2\pi\widehat{b}_k)^m} \int_G \cos\left(\frac{t^T(u_k - x)}{b_k}\right)\left(e^{\frac{t^T\Sigma t}{2b_k^2}} - e^{\frac{t^T\Sigma_k t}{2\widehat{b}_k^2}}\right) \mathrm{d}t,$$

$$I_3 = \frac{1}{(2\pi\widehat{b}_k)^m} \int_G \left(\cos\left(\frac{t^T(u_k - x)}{b_k}\right) - \cos\left(\frac{t^T(u_k - x)}{\widehat{b}_k}\right)\right) e^{\frac{t^T\Sigma_k t}{2\widehat{b}_k^2}} \mathrm{d}t.$$

Since $\|\Sigma_k - \Sigma\| = o(1/k^{1/2-c})$ for any $c > 0$, we have

$$\left|\frac{1}{b_k^m} - \frac{1}{\widehat{b}_k^m}\right| = \left(\frac{\log k}{b}\right)^{\frac{m}{2}} \frac{\left|(\lambda_{\max}(\Sigma))^{\frac{m}{2}} - (\lambda_{\max}(\Sigma_k))^{\frac{m}{2}}\right|}{(\lambda_{\max}(\Sigma_k))^{\frac{m}{2}}(\lambda_{\max}(\Sigma))^{\frac{m}{2}}} = o\left(\frac{(\log k)^{\frac{m}{2}}}{k^{\frac{1}{2}-c}}\right). \tag{45}$$

Consequently, it follows that

$$|I_1| \le \frac{1}{(2\pi)^m}\left|\frac{1}{b_k^m} - \frac{1}{\widehat{b}_k^m}\right|\left|\int_G e^{\frac{t^T\Sigma t}{2b_k^2}} \mathrm{d}t\right| \le o\left(\frac{(\log k)^{\frac{m}{2}}}{k^{\frac{1}{2}-c}}\right) O\left(k^{\frac{m}{2b}}\right) = o\left(\frac{(\log k)^{\frac{m}{2}}}{k^{\frac{1}{2}} - \frac{m}{2b} - c}\right). \tag{46}$$

Again by $\|\Sigma_k - \Sigma\| = o(1/k^{1/2-c})$ for any $c > 0$ and (45), we have

$$\left|e^{\frac{t^T\Sigma t}{2b_k^2}} - e^{\frac{t^T\Sigma_k t}{2\widehat{b}_k^2}}\right| = e^{\frac{t^T\Sigma t}{2b_k^2}}\left|1 - e^{t^T\left(\frac{\Sigma_k}{2\widehat{b}_k^2} - \frac{\Sigma}{2b_k^2}\right)t}\right| \le e^{\frac{t^T\Sigma t}{2b_k^2}} o\left(\frac{\log k}{k^{\frac{1}{2}-c}}\right).$$

Thus, it follows that

$$|I_2| \le \frac{1}{(2\pi\widehat{b}_k)^m} \int_G \left|e^{\frac{t^T\Sigma t}{2b_k^2}} - e^{\frac{t^T\Sigma_k t}{2\widehat{b}_k^2}}\right| \mathrm{d}t \le \frac{1}{(2\pi\widehat{b}_k)^m} o\left(\frac{\log k}{k^{\frac{1}{2}-c}}\right) \int_G e^{\frac{t^T\Sigma t}{2b_k^2}} \mathrm{d}t = o\left(\frac{(\log k)^{\frac{m}{2}+1}}{k^{\frac{1}{2}-\frac{m}{2b}-c}}\right). \tag{47}$$

Finally, by (45) we have

$$|I_3| \le \frac{1}{(2\pi\widehat{b}_k)^m} \int_G \left|-2\sin\left(t^T(u_k - x)\frac{1}{2}\left(\frac{1}{b_k} + \frac{1}{\widehat{b}_k}\right)\right) \cdot \sin\left(t^T(u_k - x)\frac{1}{2}\left(\frac{1}{b_k} - \frac{1}{\widehat{b}_k}\right)\right)\right| e^{\frac{t^T\Sigma_k t}{2\widehat{b}_k^2}} \mathrm{d}t$$

$$\le \frac{1}{(2\pi\widehat{b}_k)^m}\left(\frac{1}{b_k} - \frac{1}{\widehat{b}_k}\right) \int_G e^{\frac{t^T\Sigma_k t}{2\widehat{b}_k^2}} \mathrm{d}t = o\left(\frac{(\log k)^{\frac{m+1}{2}}}{k^{\frac{1}{2}-\frac{m}{2b}-c}}\right). \tag{48}$$

Therefore, together with (46)–(48), we have

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

According to the convergence rates of the estimates for the linear part in Theorem 1 and Remark 3, the estimates appearing in Section II-C have the following convergence rates.

Lemma 8: Assume H1–H6 hold. Then $\widehat{\chi}_k$ given by (25), $\mu_k^{(y)}$ given by (32) and (33), and $\mu_k^{(f)}$ given by (31) have the following convergence rates:

$$\|\widehat{\chi}_k - \chi_k\| = o\left(\frac{1}{k^{\frac{1}{6}-c}}\right) \ \forall c > 0 \tag{49}$$

$$\left\|\mu_k^{(y)} - \mu^{(y)}\right\| = o\left(\frac{1}{k^{\frac{1}{2}-c}}\right) \ \forall c > 0 \tag{50}$$

$$\left\|\mu_k^{(f)} - \mu^{(f)}\right\| = o\left(\frac{1}{k^{\frac{1}{2}-c}}\right) \ \forall c > 0. \tag{51}$$

Theorem 2: Assume H1-H6 hold. Then $\tau_k(x)$ defined by (27) and (28) and $\beta_k(x)$ defined by (29) and (30) are convergent

$$\tau_k(x) \xrightarrow[k\to\infty]{} p(x) \ a.s. \tag{52}$$

$$\beta_k(x) \xrightarrow[k\to\infty]{} h(x) \ a.s. \tag{53}$$

As a consequence, $f_k(x)$ defined by (34) is strongly consistent

$$f_k(x) \xrightarrow[k\to\infty]{} f(x) \ a.s. \tag{54}$$

Proof: The algorithm (27) can be rewritten as

$$\tau_k(x) = \left[\tau_{k-1}(x) - \frac{1}{k}\left(\tau_{k-1}(x) - p(x)\right) - \frac{1}{k}\bar{e}_k(x)\right] \times I_{\left[\left|\tau_{k-1}(x) - \frac{1}{k}\left(\tau_{k-1}(x)-p(x)\right) - \frac{1}{k}\bar{e}_k(x)\right| \le M_{\delta_k^{(\tau)}(x)}\right]} \tag{55}$$

where

$$\bar{e}_k(x) = p(x) - \widehat{w}_k(x) = (p(x) - Ew_k(x)) + (Ew_k(x) - w_k(x)) + (w_k(x) - \widehat{w}_k(x)). \tag{56}$$

Since $p(x)$ is the single root of the linear function $-(y - p(x))$, by Theorem 2.2.1 in [14], for convergence of $\tau_k(x)$ it suffices to show

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left|\sum_{j=n_k}^{m(n_k,T_k)}\frac{1}{j}\bar{e}_j(x)\right| = 0 \quad \forall T_k \in [0, T] \tag{57}$$

for any convergent subsequence $\tau_{n_k}(x)$, where $m(k,T) \triangleq \max\{m : \sum_{j=k}^{m}(1/j) \le T\}$. By the first assertion in (41) together with (42) and (44), it follows that (57) holds for $\{\bar{e}_k(x)\}$. The proof of (53) can similarly be carried out. Therefore, the estimate (34) is strongly consistent. $\blacksquare$

VI. Example

In order to justify the effectiveness of the algorithm proposed in the note, we compare our method with the identification algorithms given in [9], [13] for Hammerstein systems without input noise. We adopt the simulation example considered in [13], which is as follows:

$$y_k^0 - 0.3y_{k-1}^0 - 0.2y_{k-2}^0 + 0.3y_{k-3}^0 = 2v_{k-1}^0 + v_{k-2}^0 + v_{k-3}^0 + \xi_k,$$
$$v_k^0 = 0.9545g_1\left(u_k^0\right) + 0.2983g_2\left(u_k^0\right), \quad g_1(u) = u, g_2(u) = 2u^2$$

TABLE I
Estimation of the Linear Part

True value	$(-0.3, -0.2, 0.3)$	$(2, 1, 1)$	SSEE
Bai & Li	(-0.4694 -0.2351 0.3969)	(1.6447 0.4955 0.5109)	0.6593
Chen & Chen	(-0.2745 -0.1949 0.2993)	(2.0159 1.0532 1.0808)	0.0103
This note	(-0.3040 -0.1956 0.2903)	(2.0184 0.9937 1.0278)	0.0013

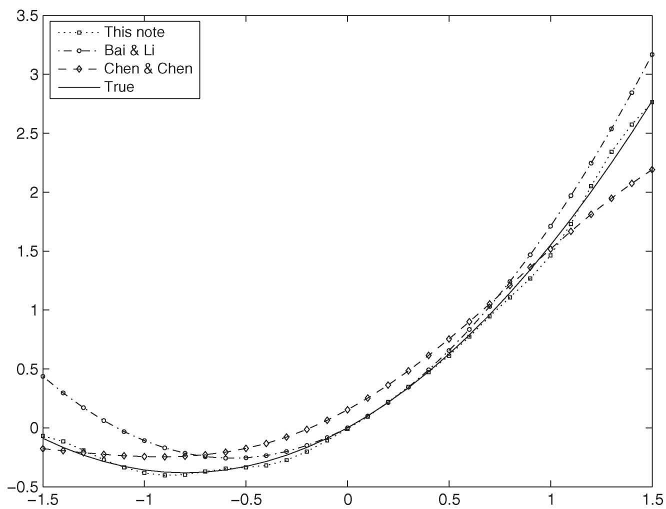

Fig. 1. Estimation of the nonlinearity by three algorithms.

where the input signal $\{u_k^0\}$ is a sequence of random variables uniformly distributed over the interval $[-1.7729, 1.7729]$, all the noises $\{\zeta_k\}$, $\{\omega_k\}$, $\{\xi_k\}$, and $\{\varsigma_k\}$ are mutually independent Gaussian random variables: $\mathcal{N}(0, 0.1^2)$, and the measurement noises η_k and ε_k are as follows:

$$\eta_k - 0.7\eta_{k-1} = \zeta_k + 0.5\zeta_{k-1}$$

$$\varepsilon_k + 0.4\varepsilon_{k-1} = \varsigma_k - 0.6\varsigma_{k-1}.$$

It is clear that the Frobenius norm of the coefficients [0.9545 0.2983] of the basis functions equals 1, which is required in [13], and the cross-correlation between the input and output of the nonlinearity is $\Upsilon \triangleq E(f(u_k^0)u_k^0) = 1$.

All results are the averages of 10 Monte Carlo simulations where the date size of each simulation is $N = 5000$. The estimates and the resulting square sum of estimation errors (SSEE) for the linear part generated by the three different algorithms are listed in Table I, respectively. The true nonlinear function together with the estimation results given by the three algorithms for the nonlinearity are illustrated in Fig. 1. We see that the algorithms given here work well, while the algorithms proposed in [9] and [13] do not lead to the satisfactory results.

VII. Conclusion

The recursive estimation algorithms for identifying MIMO EIV Hammerstein systems are proposed in the note. The estimation is carried out by the stochastic approximation algorithms, and additionally incorporated with a deconvolution kernel when estimating the nonlinear part. All the estimates are proved to converge to the true values with probability one.

For further research it is of interest to consider identification of other EIV nonlinear systems, for example, EIV Wiener systems or more complicated EIV Wiener-Hammerstein systems.

REFERENCES

[1] F. Giri and E. W. Bai, *Block-Oriented Nonlinear System Identification*, vol. 404, T. Manfred, A. Frank, and M. Manfred, Eds. London, U.K.: Springer, 2010, ser. Lecture Notes in Control and Information Sciences.

[2] E. Eskinat, S. Johnson, and W. L. Luyben, "Use of hammerstein models in identification of nonlinear systems," *AIChE Journal*, vol. 37, no. 2, pp. 255–268, 1991.

[3] J. Kim and K. Konstantnou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electron. Lett.*, vol. 37, no. 23, pp. 1417–1418, 2001.

[4] F. Jurado, "A method for the identification of solid oxide fuel cells using a hammerstein model," *J. Power Sources*, vol. 154, no. 1, pp. 145–152, 2006.

[5] E. W. Bai and D. Li, "Convergence of the iterative hammerstein system identification algorithm," *IEEE Trans. Autom. Control*, vol. 49, no. 11, pp. 1929–1940, Nov. 2004.

[6] I. Goethals, K. Pelckmans, J. A. K. Suykens, and B. D. Moor, "Identification of mimo hammerstein models using least squares support vector machines," *Automatica*, vol. 41, no. 7, pp. 1263–1272, 2005.

[7] W. X. Zhao and H. F. Chen, "Recursive identification for hammerstein systems with arx subsystems," *IEEE Trans. Autom. Control*, vol. 51, no. 12, pp. 1966–1974, Dec. 2006.

[8] W. Greblicki, "Stochastic approximation in nonparametric identification of hammerstein systems," *IEEE Trans. Autom. Control*, vol. 47, no. 11, pp. 1800–1810, Nov. 2002.

[9] X. M. Chen and H. F. Chen, "Recursive identification for mimo hammerstein systems," *IEEE Trans. Autom. Control*, vol. 56, no. 4, pp. 895–902, Apr. 2011.

[10] H. F. Chen, "Pathwise convergence of recursive identification algorithms for hammerstein systems," *IEEE Trans. Autom. Control*, vol. 49, no. 10, pp. 1641–1649, Oct. 2004.

[11] A. Garulli, L. Giarrè, and G. Zappa, "Identification of approximated hammerstein models in a worst-case setting," *IEEE Trans. Autom. Control*, vol. 47, no. 12, pp. 2046–2050, Dec. 2002.

[12] V. Cerone, D. Piga, and D. Regruto, "Bounded error identification of hammerstein systems through sparse polynomial optimization," *Automatica*, vol. 48, no. 10, pp. 2693–2698, 2012.

[13] E. W. Bai and K. Li, "Convergence of the iterative algorithm for a general hammerstein system identification," *Automatica*, vol. 46, no. 11, pp. 1891–1896, 2010.

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

[15] H. Robbins and S. Monro, "A stochastic approximation method," *Annal. Mathemat. Statist.*, vol. 22, no. 3, pp. 400–407, 1951.

[16] L. Stefanski and R. J. Carroll, "Deconvoluting kernel density estimators," *Statistics*, vol. 21, no. 2, pp. 169–184, 1990.

[17] J. Fan and Y. K. Truong, "Nonparametric regression with errors in variables," *Annal. Statist.*, vol. 21, no. 4, pp. 1900–1925, 1993.

[18] T. Söderström, "Errors-in-variables methods in system identification," *Automatica*, vol. 43, no. 6, pp. 939–958, 2007.

[19] J. C. Agüero and G. C. Goodwin, "Identifiability of errors in variables dynamic systems," *Automatica*, vol. 44, no. 2, pp. 371–382, 2008.

[20] H. F. Chen, "Recursive identification for multivariate errors-in-variables systems," *Automatica*, vol. 43, no. 7, pp. 1234–1242, 2007.

[21] Q. J. Song and H. F. Chen, "Identification of errors-in-variables systems with arma observation noise," *Syst. & Control Lett.*, vol. 57, no. 5, pp. 420–424, 2008.

[22] I. Vajk and J. Hetthéssy, "Identification of nonlinear errors-in-variables models," *Automatica*, vol. 39, no. 12, pp. 2099–2107, 2003.

[23] B. E. Jun and D. S. Bernstein, "Extended least-correlation estimates for errors-in-variables non-linear models," *Int. J. Control*, vol. 80, no. 2, pp. 256–267, 2007.

[24] B.-Q. Mu and H.-F. Chen, "Recursive identification of mimo wiener systems," *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 802–808, Mar. 2013.

[25] B.-Q. Mu and H.-F. Chen, "Hankel matrices for system identification," *J. Mathemat. Anal. and Applic.*, vol. 409, no. 1, pp. 494–508, 2014.

[26] J. Fan and Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Approach*. New York, NY, USA: Springer-Verlag, 2003.

[27] A. Mokkadem, "Mixing properties of arma processes," *Stochastic Processes and Their Applic.*, vol. 29, no. 2, pp. 309–315, 1988.

[28] B.-Q. Mu and H.-F. Chen, "Recursive identification of wiener-hammerstein systems," *SIAM J. Control and Optimiz.*, vol. 50, no. 5, pp. 2621–2658, 2012.