Recursive Identification of Hammerstein Systems: Convergence Rate and Asymptotic Normality

Biqiang Mu, Han-Fu Chen, *Fellow, IEEE*, Le Yi Wang, *Fellow, IEEE*, George Yin, *Fellow, IEEE*, and Wei Xing Zheng, *Fellow, IEEE*

***Abstract*—In this work, recursive identification algorithms are developed for Hammerstein systems under the conditions considerably weaker than those in the existing literature. For example, orders of linear subsystems may be unknown and no specific conditions are imposed on their moving average part. The recursive algorithms for estimating both linear and nonlinear parts are based on stochastic approximation and kernel functions. Almost sure convergence and strong convergence rates are derived for all estimates. In addition, the asymptotic normality of the estimates for the nonlinear part is also established. The nonlinearity considered in the paper is more general than those discussed in the previous papers. A numerical example verifies the theoretical analysis with simulation results.**

***Index Terms*—Asymptotic normality, Hammerstein system, kernel function, nonparametric approach, recursive estimation, stochastic approximation, strong consistency.**

I. Introduction

HAMMERSTEIN systems consisting of a static nonlinear function followed by a linear dynamic subsystem can effectively model many practical systems, such as distillation columns in chemical engineering [1], power amplifiers in electronic circuits [2] and solid oxide fuel cells [3], among others.

Manuscript received September 11, 2016; accepted November 4, 2016. Date of publication November 16, 2016; date of current version June 26, 2017. This work was supported in part by the National Key Basic Research Program of China (973 program) under Grant 2014CB845301, the National Natural Science Foundation of China under Grants 61603379, 61273193, 61120106011, and 61134013, the President Fund of Academy of Mathematics and Systems Science, CAS under Grant 2015-hwyxqnrc-mbq, the Air Force Office of Scientific Research under Grant FA9550-15-1-0131, and the Australian Research Council under Grant DP120104986. Recommended by Associate Editor M. Letizia.

B. Mu and H.-F. Chen are with the Key Laboratory of Systems and Control of CAS, 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).

L. Y. Wang is with the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202, USA (e-mail: lywang@wayne.edu).

G. Yin is with the Department of Mathematics, Wayne State University, Detroit, MI 48202, USA (e-mail: gyin@math.wayne.edu).

W. X. Zheng is with the School of Computing, Engineering and Mathematics, Western Sydney University, Sydney, NSW 2751, Australia (e-mail: w.zheng@westernsydney.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2016.2629668

Because of their effectiveness on modeling practical systems and their simple structures, identification of Hammerstein systems has received considerable attention from researchers and practitioners.

To identify the nonlinear function in a Hammerstein system, both parametric [2], [4], [5] and nonparametric [6]–[10] approaches have been proposed. In the parametric approach, the nonlinear part is approximated by a linear combination of basis functions such as polynomials [11], cubic spline functions [12], piecewise linear functions [13], neural networks [14], support vector machines [15], kernel machines [16], [17]. In this case, the nonlinear part depends only on a finite number of unknown parameters; and as a result the entire Hammerstein system is completely determined by the unknown parameters of both the nonlinear and linear parts. Therefore, parameterized Hammerstein systems are transformed into bilinear systems. The iterative algorithm proposed in [18] is a commonly used parametric method for estimating all unknown parameters of a Hammerstein system, and its convergence properties have been investigated subsequently. While the algorithm demonstrates fast convergence in some cases, it was shown to diverge and become unbounded in an example given in [11]. By normalizing model parameters, convergence of an iterative algorithm was established in [19] for Hammerstein systems with a finite impulse response (FIR) linear part if initial estimates were carefully chosen. This result was later extended to Hammerstein systems with infinite impulse responses (IIR) on their linear parts in [20] under the condition that the nonlinear functions are odd and the inputs are stochastic with symmetric probability distributions. By adding a regularization procedure, convergence of a modified algorithm was obtained in [5], removing certain restrictive conditions imposed in [20]. In addition to the iterative algorithms mentioned above, the kernel machine and space projection method [16] and the fixed point iteration for identifying bilinear models [17] are two kinds of algorithms proposed recently, where a kernel machine was employed to approximate the nonlinear part, which transforms a nonlinear relationship to a linear relationship in a higher dimensional space. It was shown that the resulting iteration in [17] was a contraction mapping on a metric space. Furthermore, two kinds of ambiguities in the identification of block-oriented systems were analyzed in [16].

When the nonlinear structure of a Hammerstein system is *a priori* known, such as its basis functions and corresponding orders, the parametric approach becomes a preferred choice. However, if such knowledge is not available or incorrect basis functions are used to represent the system, then the paramet-

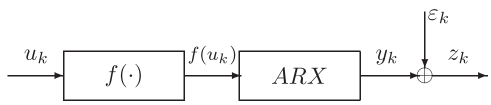


Fig. 1. Hammerstein system.

ric approach fails to generate a reliable model for modeling, prediction, and optimization. In such circumstances, the nonparametric approach is a favored alternative, since it requires no structural information about the nonlinearity of the system. The nonparametric approach is to estimate values of a nonlinear function at any points of interest by applying kernel functions.

It is well understood that for real-time implementation of identification algorithms, recursive estimation carries much lower computational complexity and uses much smaller memory space for data storage [21]–[24]. Consequently, substantial effort has been put on developing recursive algorithms for identifying Hammerstein systems. Recursive kernel estimates were used to estimate the nonlinear part of nonparametric Hammerstein systems in [6], [7], in which the assumption that the values of the nonlinear part at two points were available and different in advance is needed for completely recovering the nonlinear part. The L_1 performance index was introduced to establish the convergence and convergence rates of the estimates in [6], and the point-wise almost sure convergence was derived in [7].

A recursive algorithm for estimating a constant multiple of the impulse response sequence of the linear part and the values of a linear transform of the nonlinear part was proposed in [8]. Mean square convergence and rate of convergence for the estimates were established. The results in [8] left unaddressed some aspects on the estimates (including three undetermined constants, asymptotic normality of the nonparametric kernel estimate, among others). Recursive algorithms for estimating Hammerstein systems with FIR linear parts and general nonlinear functions were presented in [9], and point-wise almost sure convergence of the recursive algorithms was proved. Later, these recursive algorithms were extended to Hammerstein systems with ARX (autoregressive with exogenous) linear parts in [10], and point-wise almost sure convergence of the algorithms was established.

The Hammerstein systems considered in this paper are depicted in Fig. 1 and represented by

$$\begin{aligned} y_k + a_1 y_{k-1} + \cdots + a_p y_{k-p} &= b_1 f(u_{k-d}) \\ &+ b_2 f(u_{k-1-d}) + \cdots + b_q f(u_{k-(q-1)-d}) + \xi_k, \end{aligned} \quad (1)$$

where u_k, y_k, and ξ_k are the system input, the system output, and the system internal noise, respectively; (p, q) are the orders of the autoregressive (AR) part and exogenous (X) part of the linear subsystem; d is the time-delay of the system. b_1 $(b_1 \neq 0)$ is the leading coefficient. Without loss of generality, we assume $b_1 = 1$, since it is always possible to treat $\hat{f}(\cdot) \triangleq b_1 f(\cdot)$ as the nonlinear part of the system. For simplicity of presentation, we assume $d = 1$. Generalization to arbitrarily known time-delay is straightforward. It follows that the Hammerstein system to be identified can be expressed as

$$\begin{aligned} y_k + a_1 y_{k-1} + \cdots + a_p y_{k-p} &= f(u_{k-1}) \\ &+ b_2 f(u_{k-2}) + \cdots + b_q f(u_{k-q}) + \xi_k, \end{aligned} \quad (2)$$

which can be written in the compact form as

$$a(z) y_k = b(z) f(u_k) + \xi_k, \quad (3)$$

where $a(z) = 1 + a_1 z + \cdots + a_p z^p$ and $b(z) = z + b_2 z^2 + \cdots + b_q z^q$ with z being the backward shift operator: $z y_k = y_{k-1}$. The output y_k is observed with additive noise ε_k as

$$z_k = y_k + \varepsilon_k. \quad (4)$$

The goal of this paper is to recursively estimate the orders (p, q), the parameters $\{a_1, \ldots, a_p, b_2, \ldots, b_q\}$ of the linear subsystem, and the values of $f(\cdot)$ at any points of interest based on the designed inputs and observed outputs $\{u_k, z_k\}$. The paper introduces recursive algorithms and establishes their almost sure convergence, rate of convergence, and asymptotic normality. The main contributions of the paper are as follows: 1) The assumption $\sum_{j=1}^{q} b_j \neq 0$ required in [8]–[10] for recursively estimating Hammerstein systems is removed (see Remark 3 below). The minimum phase condition on the linear subsystem required in [25] is no longer assumed. Furthermore, the estimate for an important constant for completely recovering the nonlinear part is analyzed in detail (see Section II-B). 2) Instead of assuming known orders (p, q) of the linear subsystem as in [5], [9], [10], [17], the orders are estimated in this paper. The existing order estimation methods of linear systems based on the information criteria (e.g., Akaike information criterion (AIC), Bayesian information criterion (BIC), and others) are mainly applicable to batch identification, while the method proposed in this paper is built upon the rank properties of the matrices composed of impulse responses and singular value decomposition; and as such it is suitable for real-time identification. 3) In contrast to mean square convergence on the constant multiple of the impulse response of the linear part as in [8], convergence with probability one is derived here. Compared to the almost sure convergence results for parameters of the linear part and the values of the nonlinearity in [7], [9], [10], almost-sure convergence rate is obtained in this paper. 4) The estimate for a linear transform of the nonlinear part was proved to converge in mean square at rate $O(k^{-(4/5-\nu)})$ for sufficiently small $\nu > 0$ in [8] and the convergence rate of the estimate for the nonlinear part in [6] characterized by the L_1 performance index depends on the moment of the output. Here the possibly fastest rate for nonparametric kernel estimators is obtained directly for the estimate of the nonlinear part in the almost sure sense and is independent of the moment of the output. Furthermore, asymptotical normality is established. 5) The requirement in [6], [7] that the values of the nonlinear part at two points were available and different in advance for determining two constants is not needed for completely recovering the nonlinear part in this paper. 6) Three constants included in the proposed estimates were left undetermined in [8]. This problem is resolved by analyzing the identifiability of the system in this paper. 7) It is shown that a certain class of Hammerstein systems that cannot be treated by the methods in [6]–[10] can be identified in the current paper (see Remark 1).

The rest of the paper is arranged as follows. Sections II presents the recursive algorithm for estimating the parameters of the linear part and the values of the nonlinear part at any points of interest. Strong consistency, convergence rate, and asymptotic normality of the recursive estimates are investigated in Section III. A numerical example is illustrated in Section IV, and a brief conclusion is given in Section V. Some auxiliary results are placed in the Appendix.

II. Recursive Identification Algorithms

We first list the assumptions for estimating the parameters of the linear part.

Assumption 1: The designed input $\{u_k\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with zero mean, and is independent of both the internal noise $\{\xi_k\}$ and the observation noise $\{\varepsilon_k\}$. Also, $\{u_k\}$ is bounded, i.e., there exist real numbers $\underline{u}$ and $\bar{u}$ ($\underline{u} < \bar{u}$) such that $\underline{u} \le u_k \le \bar{u}$, and its density function exists and is denoted by $v(\cdot)$. The lower bound $\underline{u}$ and the upper bound $\bar{u}$ and the density $v(\cdot)$ may be unknown.

Assumption 2: $a(z)$ and $b(z)$ are coprime, and $a(z)$ is stable, i.e., $a(z) \neq 0, \forall |z| \le 1$.

Assumption 3: An upper bound n^* for $p+q$ is available.

Assumption 4: $\{\xi_k\}$ and $\{\varepsilon_k\}$ are sequences of i.i.d. random variables with zero mean and are mutually independent. Moreover, both ξ_k and ε_k have probability density, $E|\xi_k|^\Delta < \infty$ and $E|\varepsilon_k|^\Delta < \infty$ for some $\Delta > 2$.

Assumption 5: The function $f(\cdot)$ is measurable and has the left and right limits $f(x^-)$ and $f(x^+)$ at any point $x \in (\underline{u}, \bar{u})$. In addition, at least one of the constants $\tau \triangleq Ef(u_k)u_k$ and $\rho \triangleq Ef(u_k)(u_k^2 - Eu_k^2)$ is nonzero.

It is worth noting that in Assumption 4, the condition $E\varepsilon_k^2 < \infty$ usually used for proving convergence has been strengthened to $E|\varepsilon_k|^\Delta < \infty$ with $\Delta > 2$ for establishing asymptotic normality.

A. Recursive Algorithm for Estimating Linear Subsystem

Recursive identification of the linear subsystem is mainly based on the convolution relationship between its parameters and impulse response.

By stability of $a(z)$ from Assumption 2, we have

$$h(z) \triangleq \frac{b(z)}{a(z)} = \sum_{i=1}^{\infty} h_i z^i, \tag{5}$$

where $\{h_i, \, i \ge 1\}$ are impulse responses and $h_1 = 1$ since the coefficient of the power z of $b(z)$ equals 1 in (2). For any positive integers $s \ge 1, t \ge 1$, define the Toeplitz matrix

$$L(s,t) \triangleq \begin{pmatrix} h_t & h_{t-1} & \cdots & h_{t-s+1} \\ h_{t+1} & h_t & \cdots & h_{t-s+2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{t+s-1} & h_{t+s-2} & \cdots & h_t \end{pmatrix}, \tag{6}$$

where $h_i \triangleq 0$ for $i \le 0$. By [26, Theorem 4.2] it is known that the true orders of the linear subsystem are (p,q) if and only if $\operatorname{rank} L(p,q) = \operatorname{rank} L(p+1, q+1) = p$. This criterion is the basis of estimating the orders (p,q).

We now give the way of recovering the parameters by using impulse responses when the orders (p,q) are known. From (5) it follows that

$$z + b_2 z^2 \cdots + b_q z^q = (1 + a_1 z + \cdots + a_p z^p) \times (z + h_2 z^2 + \cdots + h_i z^i + \cdots).$$

Identifying coefficients for the same orders of z on both sides implies

$$b_i = \sum_{j=0}^{p} a_j h_{i-j}, \; \forall \, 1 \le i \le q, \tag{7}$$

$$h_i = -\sum_{j=1}^{p} a_j h_{i-j}, \; \forall \, i \ge q+1, \tag{8}$$

where $a_0 = 1$ and $h_i = 0$ for $i \le 0$. From (8) for the indices $q+1 \le i \le q+p$ we obtain the following linear equation:

$$L(p,q)[a_1 \, a_2 \, \cdots \, a_p]^T = -[h_{q+1} \, h_{q+2} \, \cdots \, h_{q+p}]^T.$$

Under Assumption 2, $L(p,q)$ is nonsingular by [26, Proposition 2.1], and hence the parameters of the AR-part are derived:

$$[a_1 \, a_2 \, \cdots \, a_p]^T = -L(p,q)^{-1} [h_{q+1} \, h_{q+2} \, \cdots \, h_{q+p}]^T. \tag{9}$$

The parameters of the X-part are obtained by (7) when the parameters of the AR-part are known. As a result, the $p+q$ parameters of the system can be uniquely determined via the $p+q$ values $\{h_1, \ldots, h_{p+q}\}$.

In what follows, we begin to present the recursive algorithm for estimating the linear subsystem in terms of the input-output data $\{u_k, z_k\}$. The Lemma to follow is the basis for identifying the impulse responses sequence of the linear subsystem. Define $\bar{\xi}_k \triangleq a^{-1}(z)\xi_k$ and assume that the linear subsystem is with zero initial condition. It follows from (3) that

$$y_k = \sum_{i=1}^{k} h_i f(u_{k-i}) + a^{-1}(z)\xi_k = \sum_{i=1}^{k} h_i f(u_{k-i}) + \bar{\xi}_k. \tag{10}$$

Lemma 1: Under Assumptions 1, 2, 4, and 5, we have

$$E z_k u_{k-i} = \tau h_i, \; \forall i \ge 1, \tag{11}$$

$$E z_k (u_{k-i}^2 - E u_{k-i}^2) = \rho h_i, \; \forall i \ge 1. \tag{12}$$

Proof: Since $\{\xi_k\}, \{\varepsilon_k\}$ and $\{u_k\}$ are mutually independent, we have

$$E z_k u_{k-i} = E \left(\sum_{j=1}^{k} h_j f(u_{k-j}) + \bar{\xi}_k \right) u_{k-i}$$

$$= \sum_{j=1}^{k} h_j E f(u_{k-j}) u_{k-i} = h_i E f(u_k) u_k = \tau h_i.$$

Similarly, one obtains

$$E z_k (u_{k-i}^2 - E u_{k-i}^2) = E y_k (u_{k-i}^2 - E u_{k-i}^2)$$

$$= E \left(\sum_{j=1}^{k} h_j f(u_{k-j}) + \bar{\xi}_k \right) (u_{k-i}^2 - E u_{k-i}^2)$$

$$= \sum_{j=1}^{k} h_j E\Big(f(u_{k-j})(u_{k-i}^2 - E u_{k-i}^2) \Big)$$

$$= h_i E f(u_k)(u_k^2 - E u_k^2) = \rho h_i.$$

This finishes the proof. ∎

Remark 1: In [6]–[10], estimating the impulse responses sequence of the linear subsystem is based on (11). This naturally requires $\tau \neq 0$. However, τ may be zero for some cases, for example, in the case where the input is symmetric and $f(\cdot)$ is even. In such cases, the algorithm given in [6]–[10] does not work, but it is still possible to construct the estimation algorithm based on (12) when $\rho \neq 0$.

The idea of estimating the coefficients of the linear subsystem is as follows. We first estimate the impulse responses $\{h_i, 1 \leq i \leq n^*\}$, then estimate the orders (p, q) by the estimated impulse responses, and finally give the estimates for $\{a_1, \ldots, a_p, b_2, \ldots, b_q\}$ using the estimated impulse responses and orders.

The estimation of $\{h_i, 1 \leq i \leq n^*\}$ is motivated by (11) and (12). Note that the righthand sides of (11) and (12) are equal to τ and ρ, respectively, when $i = 1$ since $h_1 = 1$. Thus, one may take the sample average of $\{z_k u_{k-i}\}$ (or $\{z_k(u_{k-i}^2 - Eu_{k-i}^2)\}$) to serve as an estimate for τh_i (or ρh_i) and hence $\{h_i, 1 \leq i \leq n^*\}$ by (11) if $\tau \neq 0$ (or by (12) if $\rho \neq 0$). Moreover, the estimate $\frac{1}{n}\sum_{k=1}^{n} z_k u_{k-1}$ for τ can be rewritten in the recursive form as

$$\theta_k^{(1,\tau)} = \theta_{k-1}^{(1,\tau)} - \gamma_k(\theta_{k-1}^{(1,\tau)} - z_k u_{k-1}),$$

where $\gamma_k = 1/k$ is the step size, $\theta_k^{(1,\tau)}$ represents the estimate for τ at time k, and the initial value $\theta_0^{(1,\tau)} = 0$. In the below, all initial values of the following recursive algorithms are set to zero. However, by Assumption 5, it is only known that at least one of τ and ρ is nonzero, so one has no knowledge about which one is nonzero. A feasible method is to apply a switching mechanism by comparing the absolute values of the estimates for τ and ρ at each step.

Denote the estimates for τh_i and ρh_i, $1 \leq i \leq n^*$ at time k by $\theta_k^{(i,\tau)}$ and $\theta_k^{(i,\rho)}$, respectively. The recursive estimates are given by

$$\theta_k^{(i,\tau)} = \theta_{k-1}^{(i,\tau)} - \gamma_k(\theta_{k-1}^{(i,\tau)} - z_k u_{k-i}), \tag{13}$$

$$\theta_k^{(i,\rho)} = \theta_{k-1}^{(i,\rho)} - \gamma_k(\theta_{k-1}^{(i,\rho)} - z_k(u_{k-i}^2 - \theta_{k-i}^{(u)})), \tag{14}$$

where $\theta_k^{(u)} = \theta_{k-1}^{(u)} - \gamma_k(\theta_{k-1}^{(u)} - u_k^2)$ is a recursive estimate for Eu_k^2. By comparing absolute values of the estimates $\theta_k^{(1,\tau)}$ and $\theta_k^{(1,\rho)}$ for τ and ρ, the impulse responses $\{h_i, 1 \leq i \leq n^*\}$ at time k are estimated as follows:

$$h_{i,k} \triangleq \begin{cases} \frac{\theta_k^{(i,\tau)}}{\theta_k^{(1,\tau)}}, & \text{if } |\theta_k^{(1,\tau)}| \geq |\theta_k^{(1,\rho)}|, \\ \frac{\theta_k^{(i,\rho)}}{\theta_k^{(1,\rho)}}, & \text{if } |\theta_k^{(1,\tau)}| < |\theta_k^{(1,\rho)}|. \end{cases} \tag{15}$$

We are in a position to estimate the orders (p, q) using the estimated impulse responses $\{h_{i,k}, 1 \leq i \leq n^*\}$. Since the true orders of the linear subsystem (2) are (p, q) if and only if $\operatorname{rank} L(p, q) = \operatorname{rank} L(p + 1, q + 1) = p$ by [26, Theorem 4.2], the key point is to estimate the rank of the Toeplitz matrices $L(s, t)$ by the available $L_k(s, t), s \geq 1, t \geq 1$ obtained from $L(s, t)$ with h_i replaced by its estimate $h_{i,k}$, $\forall 1 \leq i \leq n^*$. Here the rank of the matrices $L(s, t), s \geq 1, t \geq 1$ is estimated by applying the singular value decomposition (SVD) method given in Appendix C to $L_k(s, t), s \geq 1, t \geq 1$. The detailed estimation algorithm is placed in Appendix C. Denote the estimated rank of the Toeplitz matrices $L_k(s, t), s \geq 1, t \geq 1$ by $r_k(s, t)$. Thus, the estimate for (p, q) is selected such that the rank condition $r_k(s, t) = r_k(s + 1, t + 1) = s$ is satisfied in the range $s \geq 1, t \geq 1, s + t \leq n^*$. Denote by (p_k, q_k) the estimated orders at step k.

At last, the estimates for the parameters of the linear subsystem are defined as follows:

$$[a_{1,k}\ a_{2,k}\ \cdots\ a_{p_k,k}]^T \triangleq -L_k^{-1}(p_k, q_k) \times [h_{q_k+1,k}\ h_{q_k+2,k}\ \cdots\ h_{q_k+p_k,k}]^T, \tag{16}$$

$$b_{i,k} \triangleq \sum_{j=1}^{p_k} a_{j,k} h_{i-j,k},\ \ i = 1, \ldots, q_k, \tag{17}$$

where

$$L_k(p_k, q_k) \triangleq \begin{pmatrix} h_{q_k,k} & h_{q_k-1,k} & \cdots & h_{q_k-p_k+1,k} \\ h_{q_k+1,k} & h_{q_k,k} & \cdots & h_{q_k-p_k+2,k} \\ \vdots & \vdots & \ddots & \vdots \\ h_{q_k+p_k-1,k} & h_{q_k+p_k-2,k} & \cdots & h_{q_k,k} \end{pmatrix} \tag{18}$$

serves as the kth estimate for $L(p, q)$ with $h_{i,k} = 0$ for $i \leq 0$.

Remark 2: It is seen that the recursive estimators of both the orders and the parameters of the linear subsystem are implemented with the help of the recursively estimated impulse responses sequence. Therefore, to some extent the proposed algorithm for estimating the orders and the parameters of the linear subsystem is semi-recursive.

B. Recursive Algorithm for Estimating $f(\cdot)$ Using Kernel Functions

In this subsection, the values of $f(\cdot)$ at any points of interest in the domain $\{\underline{u} < x < \bar{u}\}$ are recursively estimated by using the kernel functions. We first introduce the kernel function $K(\cdot)$ and the averaging kernel

$$K_{d_k}(x) = \frac{1}{d_k} K\Big(\frac{u_k - x}{d_k}\Big), \tag{19}$$

where $K(\cdot)$ and the bandwidth sequence $\{d_k\}$ are assumed to satisfy the following condition:

Assumption 6:

i) $K(\cdot)$ is bounded, symmetric, and positive such that $\int_{\mathbb{R}} K(t)\mathrm{d}t = 1$, $\lim_{|t|\to\infty} |t|K(t) = 0$, $\int_{\mathbb{R}} t^2 K(t)\mathrm{d}t < \infty$;

ii) d_k monotonically tends to 0 and $kd_k \to \infty$, and $\frac{1}{k}\sum_{i=1}^{k}(\frac{d_i}{d_k})^l \to \beta_l > 0$ for $l = -1, 2$.

The well-known kernel functions including Gaussian, Epanechnikov, uniform, triangle, biweight, etc. (see [29]) satisfy Assumption 6i). If the sequence of the bandwidth $d_k = O(1/k^c)$, $0 < c < 1/2$, then Assumption 6ii) holds.

It is worth noting that unlike [8]–[10], in the estimation algorithm of the nonlinearity below, neither the condition $\sum_{i=1}^{q} b_i \neq 0$ nor the parameter estimator of the linear subsystem is involved in the kernel functions.

The idea of using kernel functions to estimate the nonlinear part is based on the following observation. Under Assumptions 1, 5, and 6, $K_{d_k}(x)$ has the following limit properties: $EK_{d_k}(x) \xrightarrow[k\to\infty]{} v(x)$ and $EK_{d_k}(x)f(u_k) \xrightarrow[k\to\infty]{} v(x)\widetilde{f}(x)$ (see Lemma 3 below), where $v(x)$ is the density function of u_k and $\widetilde{f}(x) = \big(f(x^-) + f(x^+)\big)/2$, which equals $f(x)$ if $f(\cdot)$ is continuous at the point x. Under Assumptions 1, 2, 4, 5, and 6, it follows that

$$\begin{aligned} EK_{d_k}(x)z_{k+1} &= \sum_{j=1}^{k} h_j EK_{d_k}(x) f(u_{k+1-j}) \\ &= EK_{d_k}(x)f(u_k) + \left(\sum_{j=2}^{k} h_j\right) EK_{d_k}(x)Ef(u_k) \\ &\xrightarrow[k\to\infty]{} v(x)\left(\widetilde{f}(x) + \left(\sum_{j=2}^{\infty} h_j\right) Ef(u_k)\right) \triangleq g(x). \end{aligned} \tag{20}$$

By $EK_{d_k}(x) \xrightarrow[k\to\infty]{} v(x)$ and (20), it is possible to estimate the nonlinearity $f(\cdot)$ by using the ergodicity of $\{u_k, z_k\}$.

To derive the rate and asymptotic normality of the recursive estimate for the nonlinear part, some smooth conditions on the density function $v(\cdot)$ and the nonlinear part $f(\cdot)$ are needed since the derivation usually involves the Taylor expansion up to second order. In fact, this is a standard condition for investigating the asymptotic normality. Since this paper considers the point-wise almost sure convergence property of the estimate, the rate and asymptotic normality of the estimate still hold at the twice differentiable points in the domain even if the nonlinear part is not twice differentiable for all the points in the domain, for example, the dead-zone function, the saturation function, the quantizer function and so on.

Assumption 7: Both the density function $v(\cdot)$ and the nonlinear function $f(\cdot)$ are twice differentiable and $0 < v(x) < \infty$ in the interval $(\underline{u}, \bar{u})$.

The recursive nonparametric identification of $f(\cdot)$ is accomplished by distinguishing two cases as follows.

Case 1: $Ef(u_k) = 0$.

In this case, we have $\widetilde{f}(x) = g(x)/v(x)$ by (20). Furthermore, the case of $\sum_{i=1}^{q} b_i = 0$ falls into Case 1 ($Ef(u_k) = 0$). To see this, by setting $\bar{f}(x) \triangleq f(x) - Ef(u_k)$, the system $\{a(z)y_k = b(z)\bar{f}(u_k) + \xi_k\}$ and the original system $a(z)y_k = b(z)f(u_k) + \xi_k$ produce the same output y_k if the input u_k and the initial conditions are the same for both systems. This is because

$$\begin{aligned} & y_k + a_1 y_{k-1} + \cdots + a_p y_{k-p} \\ &= \bar{f}(u_k) + b_2\bar{f}(u_{k-1}) + \cdots + b_q\bar{f}(u_{k-q}) + \xi_k \\ &= f(u_k) + b_2 f(u_{k-1}) + \cdots + b_q f(u_{k-q}) \\ &\quad - \left(\sum_{i=1}^{q} b_i\right) Ef(u_k) + \xi_k \\ &= f(u_k) + b_2 f(u_{k-1}) + \cdots + b_q f(u_{k-q}) + \xi_k. \end{aligned} \tag{21}$$

Case 2: $Ef(u_k) \neq 0$.

In this case, we have $g(x)/v(x) = \widetilde{f}(x) + (\sum_{j=2}^{\infty} h_j)Ef(u_k)$ and implicitly $\sum_{i=1}^{q} b_i \neq 0$. Therefore, to obtain the estimate for $\widetilde{f}(x)$, one has to estimate the constant $\big(\sum_{j=2}^{\infty} h_j\big)Ef(u_k)$ by using the input-output data. Note that

$$\begin{aligned} \left(\sum_{j=2}^{\infty} h_j\right) Ef(u_k) &= \left(\sum_{j=1}^{\infty} h_j\right) Ef(u_k) \frac{\sum_{j=2}^{\infty} h_j}{\sum_{j=1}^{\infty} h_j} \\ &= \mu^{(z)}\left(1 - \frac{\sum_{j=0}^{p} a_j}{\sum_{j=1}^{q} b_j}\right), \end{aligned} \tag{22}$$

where

$$\mu^{(z)} \triangleq \lim_{k\to\infty} Ez_k = \left(\sum_{j=1}^{\infty} h_j\right) Ef(u_k) \tag{23}$$

since $\{z_k\}$ is asymptotically stationary due to the stability of $a(z)$, and the identities

$$\frac{\sum_{j=2}^{\infty} h_j}{\sum_{j=1}^{\infty} h_j} = 1 - \frac{1}{\sum_{j=1}^{\infty} h_j} = 1 - \frac{\sum_{j=0}^{p} a_j}{\sum_{j=1}^{q} b_j}$$

and

$$\sum_{j=1}^{\infty} h_j = \frac{\sum_{j=1}^{q} b_j}{\sum_{j=0}^{p} a_j} \tag{24}$$

obtained by setting $z = 1$ in (5) are used. Thus, in this case, one gets

$$\widetilde{f}(x) = \frac{g(x)}{v(x)} - \mu^{(z)}\left(1 - \frac{\sum_{j=0}^{p} a_j}{\sum_{j=1}^{q} b_j}\right). \tag{25}$$

To judge whether or not $Ef(u_k)$ equals zero, we show that $Ef(u_k) = 0$ if and only if $\mu^{(z)} = 0$. The necessity is directly seen from (23). Let us show the sufficiency, i.e., $\mu^{(z)} = 0$ implies $Ef(u_k) = 0$. From (23) it follows that $\mu^{(z)} = 0$ implies either $Ef(u_k) = 0$ or $\sum_{j=1}^{\infty} h_j = 0$. In the latter case, $\sum_{j=1}^{\infty} h_j = 0$ leads to $\sum_{i=1}^{q} b_i = 0$ by (24) and hence one also derives $Ef(u_k) = 0$. Thus, the recursive algorithm for estimating the nonlinear part $f(\cdot)$ is presented as follows:

1) Estimate the sample mean $\mu^{(z)}$ of z_k:

$$\mu_k^{(z)} = \mu_{k-1}^{(z)} - \gamma_k(\mu_{k-1}^{(z)} - z_k), \tag{26}$$

where $\mu_k^{(z)}$ is the estimate for $\mu^{(z)}$ at time k.

2) Define the decision number:

$$Q_k \triangleq \frac{|\mu_k^{(z)}| + \frac{1}{\log k}}{\frac{1}{\log k}} = |\mu_k^{(z)}| \log k + 1. \tag{27}$$

3) Estimate $v(x)$ and $g(x)$:

$$v_k(x) = v_{k-1}(x) - \gamma_k(v_{k-1}(x) - K_{d_k}(x)), \tag{28}$$

$$g_k(x) = g_{k-1}(x) - \gamma_k(g_{k-1}(x) - K_{d_k}(x)z_{k+1}). \tag{29}$$

4) Estimate $f(x)$:

$$f_k(x) \triangleq \begin{cases} \frac{g_k(x)}{v_k(x)} - \mu_k^{(z)}\left(1 - \frac{\sum_{j=0}^{p_k} a_{j,k}}{\sum_{j=1}^{q_k} b_{j,k}}\right), & \text{if } Q_k \geq \eta, \\ \frac{g_k(x)}{v_k(x)}, & \text{if } Q_k < \eta, \end{cases} \tag{30}$$

where $\eta > 1$ is used to judge whether $\mu^{(z)}$ is zero or not by checking $Q_k > \eta$ or not at step k.

Remark 3: It is seen from (20) that to estimate $f(\cdot)$, one has to estimate $\left(\sum_{j=2}^{\infty} h_j\right) Ef(u_k)$. Thus, from (22) it is natural to require $\sum_{i=1}^{q} b_i \neq 0$ as done in [9], [10]. However, we have just shown that for Hammerstein systems the condition $\sum_{i=1}^{q} b_i = 0$ implies $Ef(u_k) = 0$, and hence $\left(\sum_{j=2}^{\infty} h_j\right) Ef(u_k) = 0$. So, in the case $\sum_{i=1}^{q} b_i = 0$ the Hammerstein systems can still be identified. In comparison with the recursive algorithm for estimating $f(\cdot)$ in [10], the recursive procedure (28)–(29) does not use the estimates $\{a_{i,k}, 1 \le i \le p\}$ for the coefficients of the linear subsystem.

III. Convergence Analysis

In this section, we prove convergence properties of the recursive algorithms (13), (14), (26), (28), and (29), which are all stochastic approximation algorithms. For illustration convenience, the convergence results for stochastic approximation algorithms are summarized in the Appendix.

A. Convergence Analysis of the Recursive Algorithms for Estimating the Linear Subsystem

It is noted that the noise condition is satisfied when the weighted sum of the corresponding noises converges. So, we start with convergence analysis for some series.

Lemma 2: Under Assumptions 1–5, for any $0 \le \delta < 1/2$, the following series converge almost surely: $\forall\, 1 \le i \le n^*$,

$$\sum_{k=1}^{\infty} \gamma_k^{1-\delta} (z_k u_{k-i} - E z_k u_{k-i}) < \infty, \tag{31}$$

$$\sum_{k=1}^{\infty} \gamma_k^{1-\delta} \left(z_k (u_{k-i}^2 - E u_{k-i}^2) - E z_k (u_{k-i}^2 - E u_{k-i}^2) \right) < \infty. \tag{32}$$

Proof: Under Assumptions 1–5, from (10) we derive

$$z_k u_{k-i} - E z_k u_{k-i} = h_i \left[f(u_{k-i}) u_{k-i} - E f(u_{k-i}) u_{k-i} \right] + \sum_{j=1, j\neq i}^{k} h_j f(u_{k-j}) u_{k-i} + \bar{\xi}_k u_{k-i} + \varepsilon_k u_{k-i}. \tag{33}$$

Thus, we have

$$\begin{aligned} &\sum_{k=1}^{\infty} \gamma_k^{1-\delta} \left[z_k u_{k-i} - E z_k u_{k-i} \right] \\ &= h_i \sum_{k=1}^{\infty} \gamma_k^{1-\delta} \left[f(u_{k-i}) u_{k-i} - E f(u_{k-i}) u_{k-i} \right] \\ &\quad + \sum_{k=1}^{\infty} \gamma_k^{1-\delta} \left[\sum_{j=1, j\neq i}^{k} h_j f(u_{k-j}) u_{k-i} \right] \\ &\quad + \sum_{k=1}^{\infty} \gamma_k^{1-\delta} \left[\bar{\xi}_k u_{k-i} \right] + \sum_{k=1}^{\infty} \gamma_k^{1-\delta} \left[\varepsilon_k u_{k-i} \right]. \end{aligned} \tag{34}$$

Define $\psi_k^{(1)} \triangleq \gamma_k^{1-\delta} \left[f(u_{k-i}) u_{k-i} - E f(u_{k-i}) u_{k-i} \right]$ for a fixed i. It is clear that $\{\psi_k^{(1)}\}$ is a sequence of mutually independent random variables with zero mean. By the boundedness of both $\{u_k\}$ and $\{f(u_k)\}$ we have

$$\begin{aligned} \sum_{k=1}^{\infty} E\left[\psi_k^{(1)}\right]^2 &= \sum_{k=1}^{\infty} E\left[\gamma_k^{1-\delta} \left(f(u_{k-i}) u_{k-i} - E f(u_{k-i}) u_{k-i} \right) \right]^2 \\ &\le \sum_{k=1}^{\infty} \gamma_k^{2(1-\delta)} E\left[f(u_{k-i}) u_{k-i} \right]^2 < \infty, \end{aligned}$$

which implies that the first term on the righthand side of (34) converges a.s. by the Khintchine-Kolmogorov convergence theorem [30, Theorem 1 in Section 5.1].

For the second term on the right-hand side of (34), we have

$$\begin{aligned} &\sum_{k=1}^{\infty} \gamma_k^{1-\delta} \left[\sum_{j=1, j\neq i}^{k} h_j f(u_{k-j}) u_{k-i} \right] \\ &= \sum_{k=1}^{\infty} \gamma_k^{1-\delta} \left[\sum_{j=1}^{i-1} h_j f(u_{k-j}) u_{k-i} \right] \\ &\quad + \sum_{k=1}^{\infty} \gamma_k^{1-\delta} \left[\sum_{j=i+1}^{k} h_j f(u_{k-j}) u_{k-i} \right] \\ &= \sum_{j=1}^{i-1} h_j \sum_{l=0}^{i-j} \sum_{k=1}^{\infty} \frac{1}{((i-j+1)k+l)^{1-\delta}} \\ &\quad \times \left[f(u_{(i-j+1)k+l-j}) u_{(i-j+1)k+l-i} \right] \\ &\quad + \sum_{k=1}^{\infty} \gamma_k^{1-\delta} \left[\sum_{j=i+1}^{k} h_j f(u_{k-j}) u_{k-i} \right]. \end{aligned} \tag{35}$$

Define $\psi_k^{(2)} \triangleq \frac{1}{((i-j+1)k+l)^{1-\delta}} \left[f(u_{(i-j+1)k+l-j}) u_{(i-j+1)k+l-i} \right]$ for fixed i. Thus, $\{\psi_k^{(2)}\}$ is a sequence of independent random variables with zero mean. The boundedness of $\{u_k\}$ and $\{f(u_k)\}$ leads to

$$\sum_{k=1}^{\infty} E\left[\psi_k^{(2)}\right]^2 = \sum_{k=1}^{\infty} \frac{1}{((i-j+1)k+l)^{2(1-\delta)}} \times E\left[f(u_{(i-j+1)k+l-j}) u_{(i-j+1)k+l-i} \right]^2 < \infty,$$

which implies that the first term on the right-hand side of (35) converges a.s. again by the Khintchine-Kolmogorov convergence theorem [30].

Define $\psi_k^{(3)} \triangleq \sum_{j=i+1}^{k} h_j f(u_{k-j}) u_{k-i}$ and $\mathscr{F}_k \triangleq \{u_{j-i}, i \le j \le k\}$ for a fixed i. Thus, we have $E\left[\psi_k^{(3)} \middle| \mathscr{F}_{k-1}\right] = 0$, i.e., $\{\psi_k^{(3)}, \mathscr{F}_k\}$ is a martingale difference sequence (m.d.s.) and

$$\sup_k E\left(\left| \sum_{j=i+1}^{k} h_j f(u_{k-j}) u_{k-i} \right|^{\Delta} \middle| \mathscr{F}_{k-1} \right) \le O\left(\sum_{j=i+1}^{k} |h_j| \right)^{\Delta} < \infty$$

for some $\Delta > 2$ due to the boundedness of u_k and $f(u_k)$. From Theorem A2 in the Appendix it follows that

$$\sum_{l=1}^{k} \gamma_l^{1-\delta} \left[\sum_{j=i+1}^{l} h_j f(u_{l-j}) u_{l-i} \right] = O\left(W_k \left(\log W_k \right)^{\varpi} \right) < \infty,$$

where $W_k = \big(\sum_{l=1}^k \gamma_l^{2(1-\delta)}\big)^{1/2}$. With k tending to infinity, one arrives at that the second term on the righthand side of (35) converges a.s., and hence the second term on the righthand side of (34) also converges a.s.

Define $\psi_k^{(4)} \triangleq \bar{\xi}_k u_{k-i}$ and $\mathscr{F}_k \triangleq \{u_{j-i}, \bar{\xi}_{j+1}, i \le j \le k\}$ for fixed i. Thus, we have $\bar{\xi}_k \in \mathscr{F}_{k-1}$ and $E\big[\psi_k^{(4)}|\mathscr{F}_{k-1}\big] = \bar{\xi}_k E\left[u_{k-i}|\mathscr{F}_{k-1}\right] = 0$, so $\{\psi_k^{(4)}, \mathscr{F}_k\}$ is an m.d.s. and $\sup_k E\,(|\bar{\xi}_k u_{k-i}|^\Delta|\mathscr{F}_{k-1}) = \sup_k \big(E|\bar{\xi}_k|^\Delta \times E(|u_{k-i}|^\Delta|\mathscr{F}_{k-1})\big) < \infty$ for some $\Delta > 2$ due to Theorem A3 in the Appendix and the boundedness of u_k. From Theorem A2 it follows that

$$\sum_{l=1}^k \gamma_l^{1-\delta}\bar{\xi}_l u_{l-i} = O\left(W_k\big(\log W_k\big)^\varpi\right) < \infty,$$

where $W_k = \big(\sum_{l=1}^k \gamma_l^{2(1-\delta)}\big)^{1/2}$. With k tending to infinity, one derives that the third term on the righthand side of (34) converges a.s.

Finally, define $\psi_k^{(5)} \triangleq \gamma_k^{1-\delta}\varepsilon_k u_{k-i}$. It follows that $\{\psi_k^{(5)}\}$ is a sequence of mutually independent random variables with zero mean. By Assumption 4, $E\varepsilon_k^2 < \infty$, which entails

$$\sum_{k=1}^\infty E\big(\psi_k^{(5)}\big)^2 = \sum_{k=1}^\infty \frac{1}{k^{2(1-\delta)}} E\varepsilon_k^2 Eu_k^2 < \infty.$$

Hence, the last term on the right-hand side of (34) converges a.s. by the Khintchine-Kolmogorov convergence theorem [30]. Thus, we conclude (31), while (32) can be similarly proved. The proof is complete. ■

With Lemma 2, we are ready to analyze the convergence of the estimates for the linear subsystem.

Theorem 1: Assume that Assumptions 1–5 hold. Then the following assertions take place: i) The estimates $\{h_{i,k}, 1 \le i \le n^*\}$ for the impulse responses given by (15) converge to $\{h_i, 1 \le i \le n^*\}$ almost surely with the rate:

$$|h_{i,k} - h_i| = o(k^{-\delta})\, a.s., \,\forall\, \delta \in (0, 1/2). \quad (36)$$

ii) The order estimates are strongly consistent:

$$p_k \xrightarrow[k\to\infty]{} p\, a.s. \text{ and } q_k \xrightarrow[k\to\infty]{} q\, a.s. \quad (37)$$

iii) The estimates for the parameters of the linear subsystem given by (16)–(17) converge to the true values with the rate:

$$|a_{i,k} - a_i| = o(k^{-\delta})\, a.s., \,\forall\, \delta \in (0, 1/2),\, 1 \le i \le p, \quad (38)$$

$$|b_{i,k} - b_i| = o(k^{-\delta})\, a.s., \,\forall\, \delta \in (0, 1/2),\, 1 \le i \le q. \quad (39)$$

Proof: We first show that $\theta_k^{(i,\tau)}$ and $\theta_k^{(i,\rho)}$ given by (13) and (14) converge to τh_i and ρh_i with the rate:

$$\left|\theta_k^{(i,\tau)} - \tau h_i\right| = o(k^{-\delta})\, a.s. \forall\delta \in (0, 1/2), \quad (40)$$

$$\left|\theta_k^{(i,\rho)} - \rho h_i\right| = o(k^{-\delta})\, a.s. \forall\delta \in (0, 1/2) \quad (41)$$

for $1 \le i \le n^*$, respectively. Rewrite the recursive algorithm (13) as $\theta_k^{(i,\tau)} = \theta_{k-1}^{(i,\tau)} + \gamma_k\big(-(\theta_{k-1}^{(i,\tau)} - \tau h_i) + e_k^{(i,\tau)}\big)$, where $e_k^{(i,\tau)} = z_k u_{k-i} - \tau h_i = z_k u_{k-i} - Ez_k u_{k-i}$. Thus, the corresponding regression function of the algorithm (13) is $-(x - \tau h_i)$ and the noise is $e_k^{(i,\tau)}$. By Theorem A1 in the Appendix, for proving (40), it suffices to prove $\sum_{k=1}^\infty \gamma_k^{1-\delta}(z_k u_{k-i} - Ez_k u_{k-i}) < \infty\, a.s., \forall\delta \in [0, 1/2)$. This is guaranteed by (31) in Lemma 2, and hence (40) holds. By noticing $Ez_k(u_{k-i}^2 - Eu_{k-i}^2) = Ef(u_k)\big(u_{k-i}^2 - Eu_{k-i}^2\big) = \rho h_i$, the algorithm (14) can be rewritten as $\theta_k^{(i,\rho)} = \theta_{k-1}^{(i,\rho)} + \gamma_k\big(-(\theta_{k-1}^{(i,\rho)} - \rho h_i) + e_k^{(i,\rho)}\big)$, where

$$\begin{aligned} e_k^{(i,\rho)} &= z_k(u_{k-i}^2 - \theta_{k-i}^{(u)}) - \rho h_i \\ &= \big(z_k(u_{k-i}^2 - Eu_{k-i}^2) - Ez_k(u_{k-i}^2 - Eu_{k-i}^2)\big) \\ &\quad + (Eu_{k-i}^2 - \theta_{k-i}^{(u)})\sum_{j=1}^\infty h_j f(u_{k-j}) \\ &\quad + (Eu_{k-i}^2 - \theta_{k-i}^{(u)})\bar{\xi}_k + (Eu_{k-i}^2 - \theta_{k-i}^{(u)})\varepsilon_k. \end{aligned} \quad (42)$$

Thus, the corresponding regression function of the algorithm (14) is $-(x - \rho h_i)$ and the noise is $e_k^{(i,\rho)}$. Similarly, by Theorem A1 in the Appendix, for proving (41) it suffices to show that

$$\sum_{k=1}^\infty \gamma_k^{1-\delta} e_k^{(i,\rho)} < \infty\, a.s. \forall\delta \in [0, 1/2). \quad (43)$$

Clearly, (43) holds with $e_k^{(i,\rho)}$ replaced by the first term on the righthand side of (42) by (32).

Since $\{u_k\}$ is a sequence of i.i.d. random variables with zero mean, by Theorem A2, $\sum_{i=1}^k (u_i^2 - Eu_i^2) = O(\sqrt{k}(\log\sqrt{k})^\varpi)$ for any $\varpi > 1/2$. This means $|\theta_k^{(u)} - Eu_k^2| = |\frac{1}{k}\sum_{i=1}^k (u_i^2 - Eu_k^2)| = O((\log\sqrt{k})^\varpi/\sqrt{k}) \le O(k^{-(1/2-\nu)})$ for any sufficiently small $\nu > 0$. Thus, (43) with $e_k^{(i,\rho)}$ replaced by the second term on the righthand side of (42) takes place, since $\sum_{j=1}^\infty h_j f(u_{k-j})$ is bounded.

For the third term on the right-hand side of (42), we have

$$\begin{aligned} \left|\sum_{k=1}^\infty \gamma_k^{1-\delta}(\theta_k^{(u)} - Eu_k^2)\bar{\xi}_k\right| &\le \sum_{k=1}^\infty \gamma_k^{3/2-\delta-\nu}|\bar{\xi}_k| \\ &= \sum_{k=1}^\infty \gamma_k^{3/2-\delta-\nu}(|\bar{\xi}_k| - E|\bar{\xi}_k|) + \sum_{k=1}^\infty \gamma_k^{3/2-\delta-\nu}E|\bar{\xi}_k|. \end{aligned}$$

For a fixed δ, we can always choose an appropriate $\nu > 0$ such that $3/2 - \delta - \nu > 1$, so $\sum_{k=1}^\infty \gamma_k^{3/2-\delta-\nu}E|\bar{\xi}_k| < \infty$. By Theorem A3, $\{|\bar{\xi}_k| - E|\bar{\xi}_k|\}$ is a zero mean α-mixing with mixing coefficient exponentially decaying to zero and $E|\bar{\xi}_k|^\Delta < \infty$ for some $\Delta > 2$ under Assumption 4. Note that

$$\begin{aligned} &\sum_{k=1}^\infty \left(E\Big(\gamma_k^{3/2-\delta-\nu}(|\bar{\xi}_k| - E|\bar{\xi}_k|)\Big)^\Delta\right)^{2/\Delta} \\ &\quad \le \sum_{k=1}^\infty \gamma_k^{3-2\delta-2\nu}(E|\bar{\xi}_k|^\Delta)^{2/\Delta} < \infty. \end{aligned}$$

Thus, we have $\sum_{k=1}^\infty \gamma_k^{3/2-\delta-\nu}(|\bar{\xi}_k| - E|\bar{\xi}_k|) < \infty\, a.s.$ by Theorem A6. This yields that (43) with $e_k^{(i,\rho)}$ replaced by the third term on the righthand side of (42) takes place.

Similarly, (43) also holds with $e_k^{(i,\rho)}$ replaced by the last term of (42). Therefore, (41) holds.

By Assumption 5 at least one of τ and ρ is nonzero, and the convergence of $\theta_k^{(1,\tau)}$ and $\theta_k^{(1,\rho)}$ implies that 1) the switching in (15) will cease for $|\tau| \neq |\rho|$ after a finite number of steps; 2) the switching in (15) often happens for $|\tau| = |\rho|$. Note that (15) is either $\theta_k^{(i,\tau)}/\theta_k^{(1,\tau)}$ or $\theta_k^{(i,\rho)}/\theta_k^{(1,\rho)}$ at each step for $|\tau| = |\rho|$ and both of them have the desired rate due to (40) and (41). Thus, (36) holds for both cases thanks to (40) and (41). Accordingly, (37) follows from [31], [26] under (36), while (38) and (39) straightforwardly follow from (36) together with (37). Hence, the proof is complete. ■

Remark 4: The rates (38)–(39) in Theorem 1 will not be influenced by the switchings in the algorithm due to the fact that the estimates (13)–(14) are calculated at each recursive step and the estimates $\theta_k^{(i,\tau)}, \theta_k^{(i,\rho)}$ always have the rates (40)–(41) independent of the switchings; and the rates (38)–(39) are on the asymptotic property of the estimates. Since the order estimation is also asymptotically convergent by (37), the rates (38)–(39) hold in the asymptotic case.

Remark 5: Actually, the estimates (15), (16), and (17) are asymptotically normal. This can be proved by using the similar procedure as that carried out in Theorem 2, but their asymptotic variances are difficult to derive explicitly due to the complicated relationship. This means that the rates in Theorem 1 can be improved and become $O_p(1/k^{1/2})$.

B. Convergence Analysis of the Recursive Algorithms for Estimating $f(\cdot)$

Prior to proving the convergence of the recursive algorithms (26)–(30), we first show some properties of the averaging kernel $K_{d_k}(x)$.

Lemma 3: Under Assumptions 1, 5, and 6, for the averaging kernel $K_{d_k}(x)$ defined by (19), the following limits take place

$$EK_{d_k}(x) \xrightarrow[k\to\infty]{} v(x), EK_{d_k}(x)f(u_k) \xrightarrow[k\to\infty]{} v(x)\widetilde{f}(x), \quad (44)$$

where $\widetilde{f}(x) = \big(f(x^-) + f(x^+)\big)/2$, which equals $f(x)$ if $f(\cdot)$ is continuous at x. In addition, if Assumption 7 also holds, then

$$EK_{d_k}(x) - v(x) = \frac{d_k^2 v''(x)}{2}\int_{\mathbb{R}} t^2 K(t)\mathrm{d}t + o(d_k^2), \quad (45)$$

$$\begin{aligned} &EK_{d_k}(x)f(u_k) - v(x)f(x) \\ &= d_k^2[v'(x)f'(x) + v''(x)f(x)/2 + v(x)f''(x)/2]\int_{\mathbb{R}} t^2 K(t)\mathrm{d}t \\ &\quad + o(d_k^2), \end{aligned} \quad (46)$$

and if $d_k = O(1/k^c)$ with $0 < c < 1/2$, then

$$\sum_{k=1}^{\infty}\frac{1}{k^{1-\zeta}}\big(K_{d_k}(x) - EK_{d_k}(x)\big) < \infty\, a.s., \quad (47)$$

$$\sum_{k=1}^{\infty}\frac{1}{k^{1-\zeta}}\big(K_{d_k}(x)f(u_k) - EK_{d_k}(x)f(u_k)\big) < \infty\, a.s., \quad (48)$$

where $0 \le \zeta < (1-c)/2$.

Proof: By the definition of $K_{d_k}(x)$ we have

$$\begin{aligned} &EK_{d_k}(x)f(u_k) = \int_{\underline{u}}^{\bar{u}} \frac{1}{d_k}K\Big(\frac{y-x}{d_k}\Big)f(y)v(y)\mathrm{d}y \\ &= \int_{\underline{u}}^{x}\frac{1}{d_k}K\Big(\frac{y-x}{d_k}\Big)f(y)v(y)\mathrm{d}y + \int_{x}^{\bar{u}}\frac{1}{d_k}K\Big(\frac{y-x}{d_k}\Big)f(y)v(y)\mathrm{d}y \\ &= \int_{\frac{\underline{u}-x}{d_k}}^{0} K(t)f(x+d_k t)v(x+d_k t)\mathrm{d}t \\ &\quad + \int_{0}^{\frac{\bar{u}-x}{d_k}} K(t)f(x+d_k t)v(x+d_k t)\mathrm{d}t \\ &\xrightarrow[k\to\infty]{} v(x)\big(f(x^-)+f(x^+)\big)/2. \end{aligned} \quad (49)$$

Moreover, if Assumption 7 holds, then by the Taylor expansion we have

$$\begin{aligned} v(x+d_k t)f(x+d_k t) - v(x)f(x) &= [v'(x)f(x) + v(x)f'(x)]d_k t \\ &\quad + [v''(x)f(x) + 2v'(x)f'(x) + v(x)f''(x)]d_k^2 t^2/2 + o(d_k^2), \end{aligned}$$

which implies

$$\begin{aligned} &EK_{d_k}(x)f(u_k) - v(x)f(x) \\ &= \int_{\mathbb{R}} K(t)[v(x+d_k t)f(x+d_k t) - v(x)f(x)]\mathrm{d}t \\ &= d_k^2[v'(x)f'(x) + v''(x)f(x)/2 + v(x)f''(x)/2]\int_{\mathbb{R}} t^2K(t)\mathrm{d}t \\ &\quad + o(d_k^2) \xrightarrow[k\to\infty]{} 0, \end{aligned} \quad (50)$$

where $\int_{\mathbb{R}} tK(t)\mathrm{d}t = 0$ is used. The first assertion of (44) and (45) can be similarly proved. Notice that $\{u_k\}$ is a sequence of i.i.d. random variables and so does the sequence $\{K_{d_k}(x) - EK_{d_k}(x)\}$. Utilizing the derivation similar to that used in (49), we can prove $E[K_{d_k}(x)]^2 = O(1/d_k) = O(k^c)$. When $0 \le \zeta < (1-c)/2$, i.e., $2(1-\zeta) - c > 1$, we obtain

$$\begin{aligned} &\sum_{k=1}^{\infty} E\Big(\frac{1}{k^{1-\zeta}}(K_{d_k}(x) - EK_{d_k}(x))\Big)^2 \\ &= \sum_{k=1}^{\infty}\frac{1}{k^{2(1-\zeta)}}E\Big(K_{d_k}(x) - EK_{d_k}(x)\Big)^2 \\ &\le \sum_{k=1}^{\infty}\frac{1}{k^{2(1-\zeta)}}E[K_{d_k}(x)]^2 \\ &= \sum_{k=1}^{\infty}\frac{1}{k^{2(1-\zeta)}}O(k^c) = \sum_{k=1}^{\infty}\frac{1}{k^{2(1-\zeta)-c}} < \infty, \end{aligned}$$

which implies (47) by the Khintchine-Kolmogorov convergence theorem [30]. Similarly, we obtain (48). ■

Lemma 4: Under Assumptions 1, 2, 4, and 5, it holds that

$$Q_k \xrightarrow[k\to\infty]{} \begin{cases} \infty & \text{if } \mu^{(z)} \neq 0, \\ 1 & \text{if } \mu^{(z)} = 0, \end{cases} \quad (51)$$

where Q_k is the decision number defined by (27).

Proof: First, one concludes that under Assumptions 1, 2, 4, and 5, the estimate $\mu_k^{(z)}$ given by (26) converges to $\mu^{(z)}$ a.s. with the rate: $|\mu_k^{(z)} - \mu^{(z)}| = o(k^{-\delta})\, \forall\, 0 < \delta < 1/2$. This can be proved similarly as that done in Theorem 1. Therefore, $|\mu_k^{(z)}| = |\mu^{(z)}| + o(k^{-\delta})$.

By the definition of Q_k, we have $Q_k = |\mu_k^{(z)}| \log k + 1$. If $\mu^{(z)} \neq 0$, then $Q_k = |\mu^{(z)}| \log k + o(k^{-\delta}) \log k + 1 \xrightarrow[k\to\infty]{} \infty$. On the other hand, if $\mu^{(z)} = 0$, then $|\mu_k^{(z)}| = o(k^{-\delta})$, and hence $Q_k = o(k^{-\delta}) \log k + 1 \xrightarrow[k\to\infty]{} 1$. ■

The next theorem makes an analysis of convergence of the nonlinear part estimate, including its asymptotic normality.

Theorem 2: Under Assumptions 1–6, the estimate $f_k(x)$ defined by (30) converges:

$$f_k(x) \xrightarrow[k\to\infty]{} \widetilde{f}(x) \;\; a.s. \tag{52}$$

If, in addition, Assumption 7 holds and $d_k = O(1/k^c), 0 < c < 1/2$, then $f_k(x)$ converges with the rate

$$|f_k(x) - f(x)| = o(k^{-\varsigma}) \;\; a.s. \tag{53}$$

where $0 < \varsigma = \min(2c, (1-c)/2 - \nu) < 2/5$ with sufficiently small $\nu > 0$. Moreover, if $d_k = O(1/k^{1/5})$, then $f_k(x)$ is asymptotically normal:

$$\sqrt{kd_k}(f_k(x) - f(x) - B_k) \xrightarrow[k\to\infty]{} \mathscr{N}(0, \chi^2(x)), \tag{54}$$

where

$$B_k = d_k^2 \beta_2 \Big(f''(x)/2 + f'(x)v'(x)/v(x)\Big) \int t^2 K(t) \mathrm{d}t,$$

$$\chi^2(x) = \beta_{-1}\sigma^2 \int K(t)^2 \mathrm{d}t / v(x),$$

with $\beta_2 = 5/3$, $\beta_{-1} = 5/6$, and $\sigma^2 = \mathrm{Var}(\bar{\xi}_k) + \mathrm{Var}(\varepsilon_k) + \sum_{j=2}^{\infty} h_j^2 \mathrm{Var}(f(u_k))$. This means that $|f_k(x) - f(x)| = O_p(1/k^{2/5})$. Furthermore, if the bandwidth $d_k = O(1/k^c)$, $1/5 < c < 1/2$, then (54) still holds with $B_k = 0$ and $\beta_{-1} = 1/(1+c)$.

Proof: By ergodicity of z_k and by (38), (39), it is seen from (30) that for proving (52) it suffices to show that $v_k(x) \xrightarrow[k\to\infty]{} v(x)\, a.s.$ and $g_k(x) \xrightarrow[k\to\infty]{} g(x)\, a.s.$ We first prove $v_k(x) \xrightarrow[k\to\infty]{} v(x)\, a.s.$ The recursive algorithm (28) for estimating the density $v(x)$ of u_k can be rewritten as $v_k(x) = v_{k-1}(x) + \gamma_k\big(-(v_{k-1}(x) - v(x)) + e_k^{(v)}(x)\big)$, where $e_k^{(v)}(x) = K_{d_k}(x) - v(x) = \big(K_{d_k}(x) - EK_{d_k}(x)\big) + \big(EK_{d_k}(x) - v(x)\big)$. Clearly, the regression function of the algorithm (28) is $f(y) = -(y - v(x))$ and $e_k^{(v)}(x)$ can be regarded as the noise. By Theorem A1 in the Appendix for the convergence of $v_k(x)$ it suffices to show that 1) $\sum_{k=1}^{\infty} \gamma_k \big(K_{d_k}(x) - EK_{d_k}(x)\big) < \infty$ a.s. and 2) $EK_{d_k}(x) \xrightarrow[k\to\infty]{} v(x)$.

The convergence stated in 1) holds by (47) with $\zeta = 0$, while the convergence 2) is the first assertion of (44). Furthermore, the assertion $g_k(x) \xrightarrow[k\to\infty]{} g(x)\, a.s.$ can be proved in a similar way, and hence (52) holds.

For proving (53), one needs to show that $|v_k(x) - v(x)| = o(k^{-\varsigma})\, a.s.$ and $|g_k(x) - g(x)| = o(k^{-\varsigma})\, a.s.$ We first show the first assertion. By Theorem A1 in the Appendix it suffices to show that $\sum_{k=1}^{\infty} (1/k^{1-\varsigma})\big(v(x) - EK_{d_k}(x)\big) < \infty\, a.s.$ and $EK_{d_k}(x) - v(x) = O(k^{-\varsigma})$. The assertions (45) and (47) imply $EK_{d_k}(x) - v(x) = O(d_k^2) = O(k^{-2c})$ and $\sum_{k=1}^{\infty} \frac{1}{k^{1-\zeta}} \big(K_{d_k}(x) - EK_{d_k}(x)\big) < \infty\, a.s.$ As a result, the rate of the convergence for the term $v(x) - EK_{d_k}(x)$ is $O(k^{-2c})$, while for the term $EK_{d_k}(x) - K_{d_k}(x)$ it is $O(k^{-\zeta})$. Since $0 \le \zeta < (1-c)/2$, we can choose a sufficiently small $\nu > 0$ such that $\vartheta \triangleq (1-c)/2 - \nu > 0$ and hence

$$\sum_{k=1}^{\infty} \frac{1}{k^{1-\vartheta}} \big(EK_{d_k}(x) - K_{d_k}(x)\big) < \infty\, a.s.$$

So, the convergence rate is $O(k^{-\varsigma})$ with $\varsigma = \min(2c, \vartheta)$ for the bandwidth $d_k = O(1/k^c)$ with $0 < c < 1/2$. In particular, when we choose $2c = (1-c)/2 - \nu$, i.e., $c = 1/5 - 2\nu/5$, the two terms achieve the same convergence rate $O(1/k^{2/5 - 4\nu/5})$. This makes the algorithm (28) achieve the fastest rate of convergence $O(1/k^{2/5-4\nu/5})$ when $c = 1/5 - 2\nu/5$. Similarly, one can show $|g_k(x) - g(x)| = O(k^{-\varsigma})\, a.s.$ By (30), these rates incorporating with (38) and (39) yield the assertion (53).

We now proceed to show the asymptotical normality (54). For simplicity of notation, let us denote the constant $\mu^{(z)}(1 - \frac{\sum_{j=0}^{p} a_j}{\sum_{j=1}^{q} b_j})$ and its estimate $\mu_k^{(z)}(1 - \frac{\sum_{j=1}^{p_k} a_{j,k}}{\sum_{j=1}^{q_k} b_{j,k}})$ by ϕ and ϕ_k, respectively, and set $\omega_k \triangleq \sum_{j=2}^{k} h_j(f(u_{k-j}) - Ef(u_{k-j})) + \bar{\xi}_k + \varepsilon_k$. Thus, we have $g(x) = v(x)(f(x) + \phi)$ and $z_k = f(u_{k-1}) + \omega_k + \sum_{j=2}^{k} h_j Ef(u_k) = f(u_{k-1}) + \omega_k + \phi + O(\lambda^k)$ for some $0 < \lambda < 1$. Noting that

$$\sqrt{kd_k}\,(\phi_k - \phi) = \sqrt{kd_k} \times o(k^{-\delta}) = o(1),$$

by Theorem 1 and the relation $|\mu_k^{(z)} - \mu^{(z)}| = o(k^{-\delta}), \forall\, 0 < \delta < 1/2$, we see from (30) that for proving (54) it suffices to show

$$\sqrt{kd_k}\Big(\frac{g_k(x)}{v_k(x)} - \frac{g(x)}{v(x)} - B_k\Big) \xrightarrow[k\to\infty]{} \mathscr{N}(0, \chi^2(x)). \tag{55}$$

First, we have

$$\begin{aligned} \frac{g_k(x)}{v_k(x)} - \frac{g(x)}{v(x)} &= \frac{g_k(x)v(x) - g(x)v_k(x)}{v_k(x)v(x)} \\ &= \frac{\big(g_k(x) - g(x)\big) - \big(f(x) + \phi\big)\big(v_k(x) - v(x)\big)}{v_k(x)} \\ &= J_1(x) + J_2(x), \end{aligned}$$

where

$$J_1(x) \triangleq \frac{\big(g_k(x) - Eg_k(x)\big) - \big(f(x) + \phi\big)\big(v_k(x) - Ev_k(x)\big)}{v_k(x)},$$

which is asymptotically normally distributed, and

$$J_2(x) \triangleq \frac{\big(Eg_k(x) - g(x)\big) - \big(f(x) + \phi\big)\big(Ev_k(x) - v(x)\big)}{v_k(x)},$$

is the bias term.

We first consider the bias term $J_2(x)$. Simple calculation indicates that the recursive forms (28) and (29) can be expressed by $v_k(x) = \frac{1}{k}\sum_{i=1}^{k} K_{d_i}(x)$ and $g_k(x) = \frac{1}{k}\sum_{i=1}^{k} K_{d_i}(x) z_{i+1}$, respectively. Further, by Assumption 6ii) it follows from (45) and (46) that

$$Ev_k(x) - v(x) = \frac{1}{k}\sum_{i=1}^{k}(EK_{d_i}(x) - v(x))$$
$$= \frac{1}{k}\sum_{i=1}^{k}\Big(\frac{v''(x)d_i^2}{2}\int_{\mathbb{R}} t^2K(t)\mathrm{d}t + o(d_i^2)\Big)$$
$$= \frac{d_k^2 v''(x)}{2}\int_{\mathbb{R}} t^2K(t)\mathrm{d}t\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{d_i}{d_k}\right)^2\right) + o(d_k^2)$$
$$= \frac{d_k^2\beta_2 v''(x)}{2}\int_{\mathbb{R}} t^2K(t)\mathrm{d}t + o(d_k^2)$$

and

$$Eg_k(x) - g(x) = \frac{1}{k}\sum_{i=1}^{k}(EK_{d_i}(x)z_{i+1} - g(x))$$
$$= \frac{1}{k}\sum_{i=1}^{k}(EK_{d_i}(x)f(u_i) - v(x)f(x))$$
$$+ \frac{1}{k}\sum_{i=1}^{k}\Big(EK_{d_i}(x)\sum_{j=2}^{i}h_jEf(u_k) - v(x)\phi\Big)$$
$$= \frac{1}{k}\sum_{i=1}^{k}(EK_{d_i}(x)f(u_i) - v(x)f(x))$$
$$+ \frac{1}{k}\sum_{i=1}^{k}(EK_{d_i}(x) - v(x))\phi + O(1/k)$$
$$= d_k^2\beta_2[v'(x)f'(x) + v''(x)f(x)/2$$
$$+ v(x)f''(x)/2]\int_{\mathbb{R}} t^2K(t)\mathrm{d}t$$
$$+ (Ev_k(x) - v(x))\phi + o(d_k^2),$$

where the condition $\frac{1}{k}\sum_{i=1}^{k}(\frac{d_i}{d_k})^2 \xrightarrow[k\to\infty]{} \beta_2$ in Assumption 6ii) is used. Since $v_k(x) \xrightarrow[k\to\infty]{} v(x)\ a.s.$, one derives

$$J_2(x) = d_k^2\beta_2[f''(x)/2 + v'(x)f'(x)/v(x)]\int_{\mathbb{R}} t^2K(t)\mathrm{d}t$$
$$+ o(d_k^2) = B_k + o(d_k^2),$$

which implies

$$\sqrt{kd_k}\Big(\frac{g_k(x)}{v_k(x)} - \frac{g(x)}{v(x)} - B_k\Big) = \sqrt{kd_k}J_1(x) + o\Big(\sqrt{kd_k^5}\Big)$$
$$= \sqrt{kd_k}J_1(x) + o(1)$$

due to $d_k = O(1/k^{1/5})$. Thus for asymptotic normality it remains to show

$$\sqrt{kd_k}J_1(x) \xrightarrow[k\to\infty]{} \mathscr{N}(0,\chi^2(x)). \tag{56}$$

The term $g_k(x) - Eg_k(x)$ in $J_1(x)$ can be decomposed into

$$g_k(x) - Eg_k(x) = \frac{1}{k}\sum_{i=1}^{k}\Big(K_{d_i}(x)z_{i+1} - EK_{d_i}(x)z_{i+1}\Big)$$
$$= \frac{1}{k}\sum_{i=1}^{k}\Big(K_{d_i}(x)f(u_i) - EK_{d_i}(x)f(u_i)\Big) + \frac{1}{k}\sum_{i=1}^{k}K_{d_i}(x)\omega_{i+1}$$
$$+ \frac{1}{k}\sum_{i=1}^{k}\Big(K_{d_i}(x) - EK_{d_i}(x)\Big)\sum_{j=2}^{i}h_jEf(u_k)$$
$$= \frac{1}{k}\sum_{i=1}^{k}\Big(K_{d_i}(x)f(u_i) - EK_{d_i}(x)f(u_i)\Big) + \frac{1}{k}\sum_{i=1}^{k}K_{d_i}(x)\omega_{i+1}$$
$$+ (v_k(x) - Ev_k(x))\phi + O(1/k).$$

Since $v_k(x) \xrightarrow[k\to\infty]{} v(x)\ a.s.$, one obtains

$$J_1(x) = J_{11}(x) + J_{12}(x) + J_{13}(x) + o(1),$$

where

$$J_{11}(x) = \frac{1}{k}\sum_{i=1}^{k}\Big(K_{d_i}(x)f(u_i) - EK_{d_i}(x)f(u_i)\Big)/v(x),$$
$$J_{12}(x) = -f(x)\left(\frac{1}{k}\sum_{i=1}^{k}\big(K_{d_i}(x) - EK_{d_i}(x)\big)\right)/v(x),$$
$$J_{13}(x) = \frac{1}{k}\sum_{i=1}^{k}K_{d_i}(x)\omega_{i+1}/v(x).$$

Let us first show that $\sqrt{kd_k}\big(J_{11}(x) + J_{12}(x)\big)$ converges to zero in probability. Clearly, we have

$$\mathrm{Var}(J_{11}(x))$$
$$= \frac{1}{v(x)^2k^2}\sum_{i=1}^{k}\Big(E\big[K_{d_i}(x)f(u_i)\big]^2 - \big[EK_{d_i}(x)f(u_i)\big]^2\Big)$$

due to the mutual independence of $\{u_k\}$. A treatment similar to that used in (49) leads to $E\big[K_{d_i}(x)f(u_i)\big]^2 = d_i^{-1}v(x)f(x)^2 \int K(t)^2\mathrm{d}t + O(d_i)$ and $\big[EK_{d_i}(x)f(u_i)\big]^2 = v(x)^2f(x)^2 + O(d_i^2)$. Thus,

$$\mathrm{Var}(J_{11}(x)) = \frac{f(x)^2\int K(t)^2\mathrm{d}t}{v(x)kd_k}\left(\frac{1}{k}\sum_{i=1}^{k}\frac{d_k}{d_i}\right) + \frac{f(x)^2}{k}$$
$$+ O\Big(\frac{d_k}{k}\Big) + O\Big(\frac{d_k^2}{k}\Big).$$

This entails

$$kd_k\mathrm{Var}(J_{11}(x)) = \frac{f(x)^2\int K(t)^2\mathrm{d}t}{v(x)}\left(\frac{1}{k}\sum_{i=1}^{k}\frac{d_k}{d_i}\right) + O(d_k)$$
$$= \frac{\beta_{-1}f(x)^2\int K(t)^2\mathrm{d}t}{v(x)} + o(1), \tag{57}$$

where the condition $\left(\frac{1}{k}\sum_{i=1}^{k}\frac{d_k}{d_i}\right) \xrightarrow[k\to\infty]{} \beta_{-1}$ in Assumption 6 is used. Similarly, we have

$$kd_k \mathrm{Var}(J_{12}(x)) = \frac{\beta_{-1} f(x)^2 \int K(t)^2 \mathrm{d}t}{v(x)} + o(1),$$

$$kd_k \mathrm{Cov}(J_{11}(x), J_{12}(x)) = -\frac{\beta_{-1} f(x)^2 \int K(t)^2 \mathrm{d}t}{v(x)} + o(1).$$

It follows that

$$\mathrm{Var}\big(\sqrt{kd_k}\,[J_{11}(x) + J_{12}(x)]\big) = kd_k \mathrm{Var}(J_{11}(x)) + kd_k \mathrm{Var}(J_{12}(x)) + 2kd_k \mathrm{Cov}(J_{11}(x), J_{12}(x)) = o(1),$$

which implies $\sqrt{kd_k}\,(J_{11}(x) + J_{12}(x)) \xrightarrow[k\to\infty]{} 0$ in probability. Therefore, to show the asymptotic normality (56), it suffices to prove

$$\sqrt{kd_k}\,J_{13}(x) \xrightarrow[k\to\infty]{} \mathcal{N}(0, \chi^2(x)), \tag{58}$$

which will be given in Appendix D. ■

IV. Illustrative Example

Consider the Hammerstein system

$$y_k + a_1 y_{k-1} + a_2 y_{k-2} = f(u_{k-1}) + b_2 f(u_{k-2}) + b_3 f(u_{k-3}) + \xi_k, \quad f(u_k) = u_k^3,$$

where $a_1 = 0.3$, $a_2 = 0.6$, $b_2 = 0.8$, and $b_3 = -1.8$, and the true orders (2, 3) of the linear subsystem are unknown. It is seen that $\sum_{j=1}^{3} b_i = 0$, and hence the nonlinear part cannot be identified by the algorithms given in [8]–[10], while it can be estimated by the algorithm proposed in the paper.

Let the input signal $\{u_k\}$ be a sequence of i.i.d. random variables uniformly distributed over $[-1, 1]$. Assume that the internal noise $\{\xi_k\}$ and the observation noise $\{\varepsilon_k\}$ are sequences of mutually independent Gaussian random variables: $\xi_k \in \mathcal{N}(0, 0.3^2)$ and $\varepsilon_k \in \mathcal{N}(0, 0.3^2)$. The resulting signal-to-noise ratio (SNR) is 8.4285 dB. The sample size at each Monte-Carlo experiment is $N = 3000$. The simulation results below are based on 101 Monte-Carlo experiments. The implementation of each experiment is summarized as follows: 1) Estimate the impulse responses by the algorithms (13)–(15); 2) Estimate the orders of the linear subsystem by the estimated impulse responses with the help of the SVD method introduced in the Appendix; 3) Estimate the parameters of the linear subsystem by (16)–(18); 4) Estimate the nonlinear part by (26)–(30). In one implementation, the needed recursive steps for correctly finding the true orders of the linear system is an important index to evaluate the order estimation algorithm, which is defined as the minimum number such that the estimated orders are correct when the recursive steps are greater than or equal to the number.

Fig. 2 illustrates the distribution of the needed recursive steps for correctly finding the true orders of the AR-part, the X-part, and the linear subsystem by box plots, respectively. It is seen that the first quantile (the 25th percentile), the second quantile (median), and the third quantile (the 75th percentile) of the needed steps for correctly finding the true orders of the linear subsystem are 117, 194, and 256, respectively. This means that the steps for correctly finding the true orders (2, 3) are less than 194 in half simulations. In the following plots, the solid lines, dashed lines, dotted lines represent the true values, the estimates based on the average of 101 experiments, and the one unit of standard deviation of the corresponding estimates, respectively. The recursive estimates for the parameters of the linear part are presented in Figs. 3 and 4, while Fig. 5 gives the estimates for the nonlinear part for its arguments taking values in the interval $[-1, 1]$. Figs. 3 and 4 show that the estimates have a large fluctuation before the true orders are correctly found for all the experiments. To some extent, this is caused by the estimation algorithm (16)–(18) since the obtained parameter estimation is incorrect when the estimated orders are wrong. Meanwhile, it is also seen from Fig. 5 that the nonparametric estimate for the nonlinear part based on kernel functions are subjected to so-called boundary effects, a phenomenon in which the bias of an estimator increases near the endpoints of the estimation interval [32]. To reduce the impact of these boundary effects, boundary kernels can be applied by modifying kernel estimators near boundaries (see [33] for details). From the simulation results it

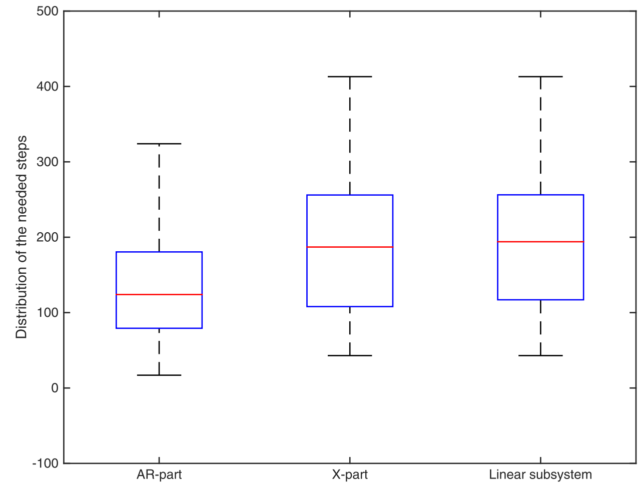


Fig. 2. Boxplot of the needed steps for correctly finding the true orders of the AR-part, the X-part, and the linear subsystem, respectively.

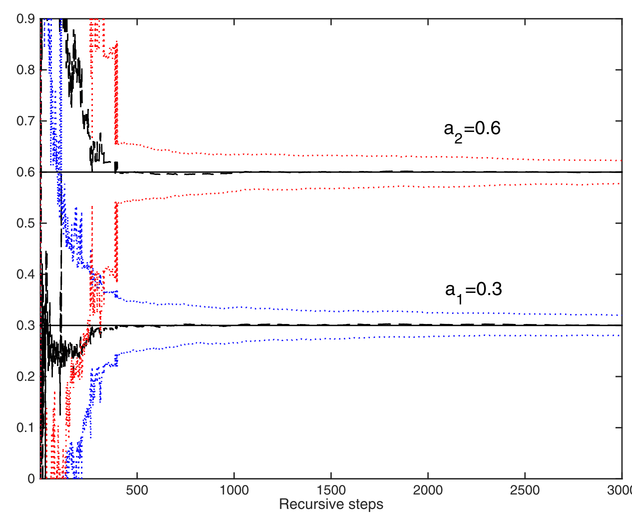


Fig. 3. Recursive estimates for AR-part. The black solid lines, black dashed lines, dotted lines (blue and red) represent the true values, the estimates based on the average of 101 experiments, and the one unit of standard deviation of the corresponding estimates, respectively.

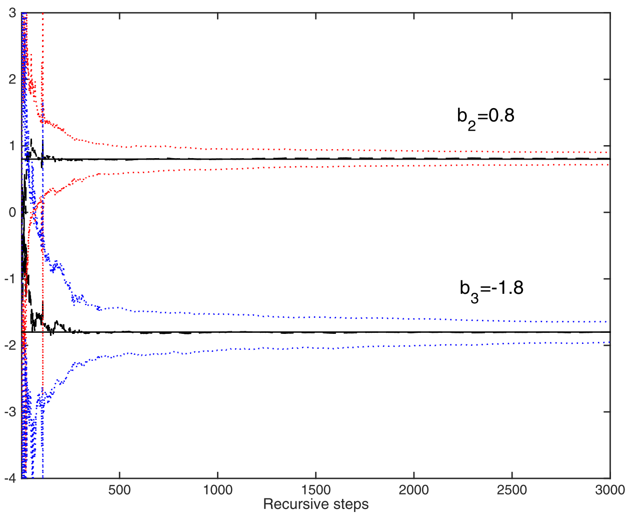


Fig. 4. Recursive estimates for X-part. The black solid lines, black dashed lines, dotted lines (blue and red) represent the true values, the estimates based on the average of 101 experiments, and the one unit of standard deviation of the corresponding estimates, respectively.

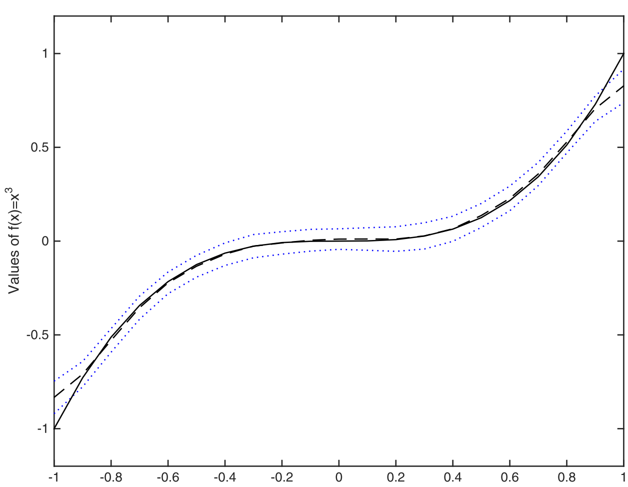


Fig. 5. Nonparametric estimate for the nonlinear part in the interval $[-1, 1]$. The black solid line, black dashed line, blue dotted lines represent the true values, the estimates based on the average of 101 experiments, and the one unit of standard deviation of the estimates, respectively.

TABLE I
STATISTICAL RESULTS ON THE TIME SPENT BY THE ALGORITHM FOR 101 RUNS: UNIT (SECONDS)

Quantile	5%	25%	50%	75%	95%	Avg
Time	4.4073	4.4146	4.4187	4.4243	4.4360	4.4197

is clearly seen that the proposed recursive algorithms perform very well as predicted in the preceding theoretical analysis.

In order to illustrate the computational complexity of the algorithm proposed, the quantiles and the average of the time spent by the algorithm for 101 runs are reported in Table I. This result shows that the algorithm can be operated quickly and thus is suitable for real-time applications. The hardware used for this computation includes a 3.5 GHz Intel Core i5 CPU and an 8 GB RAM while the software platform is Matlab 2014b running under OS X 10.10 operation system.

V. CONCLUSION

Recursive identification algorithms for Hammerstein systems have been proposed based on stochastic approximation incorporated with kernel functions. The new findings include the following key aspects. 1) Some restrictive conditions used in the literature have been removed; 2) The orders of the linear subsystem may be unknown and are consistently estimated recursively. 3) Almost sure convergence rate of the estimate for the parameters of the linear subsystem has been established; 4) The rate of point-wise convergence and asymptotic normality of the estimate for the nonlinearity have been derived.

Note that the recursive estimates for the orders and parameters of the linear subsystem are implemented with the help of recursively estimated impulse response sequences. For further research it is of interest to derive recursive estimates for both the orders and parameters of the linear subsystem directly based on the input-output data rather than on some intermediate estimates.

APPENDIX

A. Stochastic Approximation With Linear Functions

The stochastic approximation is a recursive method used for estimating roots of an unknown function $f(\cdot)$ (regression function) from the observation that may be corrupted by errors and noises. It updates the estimate as follows:

$$x_k = x_{k-1} + \gamma_k O_k, \tag{59}$$

where γ_k is the step size and it may be taken as $\gamma_k = 1/k$, and O_k is the observation of $f(\cdot)$ at time k. The observation O_k can always be decomposed as $O_k = f(x_{k-1}) + \varepsilon_k$, where $f(x_{k-1})$ represents the value of $f(\cdot)$ at x_{k-1} and ε_k is the resulting observation error.

Since all regression functions of the recursive algorithms involved in the paper are linear, i.e., $f(x) = -(x - x^*)$, where x^* is the parameter that needs to be estimated, we only introduce the convergence results of stochastic approximation with linear regression functions.

Theorem A1: ([34, Theorem 2.5.1, Remark 2.5.2, and Theorem 2.6.1])

Let the estimation sequence $\{x_k\}$ be produced by the stochastic approximation algorithm (59) with linear regression function $f(x) = -(x - x^*)$. Then the recursive estimate x_k converges to the true value x^* if and only if the observation noise ε_k can be decomposed into two parts $\varepsilon_k = \varepsilon_k' + \varepsilon_k''$ such that

$$\sum_{k=1}^{\infty} \gamma_k \varepsilon_k' < \infty \, a.s. \text{ and } \varepsilon_k'' \xrightarrow[k\to\infty]{} 0 \, a.s.$$

Further, x_k converges to x^* with the rate $|x_k - x^*| = o(\gamma_k^\delta)$ if the observation noise ε_k can be decomposed into two parts $\varepsilon_k = \varepsilon_k' + \varepsilon_k''$ such that

$$\sum_{k=1}^{\infty} \gamma_k^{1-\delta} \varepsilon_k' < \infty \, a.s. \text{ and } \varepsilon_k'' = O(\gamma_k^\delta) \, a.s.$$

for some $\delta \in (0, 1]$.

B. Convergence for Series of Random Variables

Theorem A2: ([35, Lemma 2]) Let $\{X_k, \mathscr{F}_k\}$ be a martingale difference sequence satisfying $\sup_k E(|X_k|^\Delta|\mathscr{F}_{k-1}) < \infty\, a.s.$ for some $\Delta > 2$. Let $\{M_k\}$ be a sequence of random variables such that M_k is $\mathscr{F}_{k-1}$ measurable. Then

$$\sum_{i=1}^{k} M_i X_{i+1} = O\Big(W_k\Big(\log W_k\Big)^{\varpi}\Big)\; a.s., \forall \varpi > 1/2$$

with $W_k = (\sum_{i=1}^k M_i^2)^{1/2}$.

For the process $\{X_k,\, k = 1, 2, \ldots\}$, denote by $\mathscr{F}_i^j$ the σ-algebra generated by $\{X_s,\, 1 \le i \le s \le j\}$. For simplicity, $\mathscr{F}_1^k$ is abbreviated as $\mathscr{F}_k$. Define

$$\alpha(k) \overset{\Delta}{=} \sup_{n, A \in \mathscr{F}_n, B \in \mathscr{F}_{n+k}^{\infty}} |P(A)P(B) - P(AB)|.$$

The process $\{X_k\}$ is called α-mixing if $\alpha(k) \xrightarrow[k\to\infty]{} 0$, and the numbers $\alpha(k)$ are called the mixing coefficients of the random process $\{X_k\}$.

Theorem A3: ([36, Theorem 1] [25, Lemma 4.2]) Let $\{X_k\}$ be a stable autoregressive moving average (ARMA) process driven by a white noise sequence $\{e_k\}$ with a continuous density function. Then $\{X_k\}$ is α-mixing with mixing coefficients (or mixing rates) decaying exponentially to zero. Moreover, if $E|e_k|^\nu < \infty$ for some $\nu > 0$, then $E|X_k|^\nu < \infty$.

Theorem A4: ([37, Lemma 1]) Let $\{X_k\}$ be an α-mixing with the mixing coefficients $\alpha(k)$. Let r_1, r_2, r_3 be positive numbers such that $r_1^{-1} + r_2^{-1} + r_3^{-1} = 1$. Suppose that Y and Z are random variables measurable with respects to the σ-algebras $\mathscr{F}_l$ and $\mathscr{F}_{l+k}$, respectively. Then

$$|E(YZ) - EYEZ| \le 10(\alpha(k))^{\frac{1}{r_3}} (E|Y|^{r_1})^{\frac{1}{r_1}} (E|Z|^{r_2})^{\frac{1}{r_2}}.$$

Theorem A5: ([38, Lemma 1.1]) Let $\{X_k\}$ be an α-mixing with mixing coefficients $\alpha(k)$. Let $m_k, t_k, k = 1, \ldots, n$ be integers such that $1 = m_1 < t_1 < \cdots < m_n < t_n$ with $m_{k+1} - t_k \ge l$, $k = 1, 2, \ldots, n-1$. Suppose that $Y_1, Y_2, \ldots, Y_n$ are random variables with $|Y_k| \le 1$ and Y_k is measurable with respect to $\mathscr{F}_{m_k}^{t_k}$. Then

$$|E(Y_1 \cdots Y_n) - EY_1 \cdots EY_n| \le 16(n-1)\alpha(l).$$

Theorem A6: ([39, Lemma 4]) Let $\{X_k, \mathscr{F}_k\}$ be a zero mean α-mixing with the mixing coefficients $\alpha(k)$ exponentially decaying to zero and $\sum_{k=1}^\infty (E|X_k|^\Delta)^{\frac{2}{\Delta}} < \infty$ for some $\Delta > 2$. Then $\sum\limits_{k=1}^{\infty} X_k < \infty\, a.s.$

C. Estimating the Effective Rank of a Matrix

The method of estimating the effective rank of a matrix corrupted by disturbance is based on the following theorem. Let $A = [a_{ij}]$ be an $m \times n$ matrix of complex valued elements. One now seeks for an $m \times n$ matrix $B = [b_{ij}]$ of rank r minimizing the criterion

$$\|A - B\|_F = \left[\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij} - b_{ij}|^2\right]^{1/2},$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. Let the singular value decomposition (SVD) of A be

$$A = U\Sigma V, \tag{60}$$

where U and V are $m \times m$ and $n \times n$ unitary matrices, respectively, and $\Sigma = [\sigma_{jj}]$ is an $m \times n$ nonnegative diagonal matrix whose elements are ordered such that $\sigma_{11} \ge \sigma_{22} \ge \cdots \sigma_{ll} \ge 0$, where $l = \min(m, n)$. The diagonal elements are called the singular values of A.

Theorem A7: ([27, Theorem 7.2]) The unique $m \times n$ matrix of rank $r \le \operatorname{rank}(A)$ which best approximates the $m \times n$ matrix A in the Frobenius norm sense is given by $A^{(r)} = U\Sigma_r V$, where U and V are as given in (60) while Σ_r is obtained from Σ by keeping its r largest singular values and setting the rest to zero. This optimal approximation provides the minimum of the criterion:

$$\|A - A^{(r)}\|_F = \left[\sum_{j=r+1}^{l} \sigma_{jj}^2\right]^{1/2}. \tag{61}$$

As r approaches to l, this sum in (61) is decreasing and eventually becomes zero at $r = l$. To provide a convenient measure for this approximation independent of the size of matrix A, consider the normalized ratio:

$$\nu(r) = \frac{\|A^{(r)}\|_F}{\|A\|_F} = \sqrt{\frac{\sigma_{11}^2 + \sigma_{22}^2 + \cdots + \sigma_{rr}^2}{\sigma_{11}^2 + \sigma_{22}^2 + \cdots + \sigma_{ll}^2}},\ 1 \le r \le l.$$

Clearly, this normalized ratio approaches its maximum 1 as r tends to l. If the quantity $\nu(r)$ is close to one for some r significantly smaller than l, then the matrix A is of low effective rank. On the other hand, if in order for $\nu(r)$ to be close to one, the corresponding r must take values close to l (i.e., $r \approx l$), then A is said to be of high effective rank.

Based on the explanation given above, the estimate of the effective rank of the matrix $L_k(s,t), s \ge 1, t \ge 1$ is given by

$$r_k(s,t) = \min\{r \mid \nu(r) \ge \varrho,\ 1 \le r \le s\},$$

where $\nu(r)$, $1 \le r \le s$ is the normalized ratio of the matrix $L_k(s,t), s \ge 1, t \ge 1$ and ϱ is a fixed threshold close to but less than one, e.g., $\varrho = 0.999$ [27], [28].

Remark 6: Note that a method for estimating the rank of the matrices $L(s,t)$ by using $L_k(s,t)$ was given in [26] based on the characteristic polynomial of the matrix $L_k(s,t)L_k^T(s,t)$. Our practical experience shows that the SVD method is more robust and performs better than the method used in [26] and hence the SVD method is adopted in this paper. It is also seen from the illustrative example in Section IV that the SVD method works well since the true orders are almost correctly found for all 101 runs when the data size is greater than 500.

D. Proof of Asymptotic Normality for Theorem 2

Theorem A8: ([30, Corollary 1 in Section 9.1]) If $\{X_k\}$ are mutually independent with $EX_k = 0$ and $\sum_{j=1}^k E|X_j|^\Delta = o(s_k^\Delta)$ for some $\Delta > 2$, where $s_k^2 = \sum_{j=1}^k EX_j^2$, then $\sum_{j=1}^k X_j/s_k \to \mathscr{N}(0,1)$.

Proof of Asymptotic Normality for Theorem 2 Let us continue to show the asymptotic normality (54) with the help of Theorem A8, where we apply the technique of big blocks separated by small blocks [40, Theorem 7.5, pages 228–231] and [41].

Define the σ-algebra $\mathscr{F}_k \triangleq \{u_j, \xi_{j+1}, \varepsilon_{j+1}, j \leq k\}$ and $Z_k \triangleq K_{d_k}(x)\omega_{k+1}$. Thus, $\{Z_k, \mathscr{F}_k\}$ is adapted and the sequence $\{Z_k\}$ is an α-mixing with exponentially decaying mixing coefficients $\alpha(k)$ by Theorem A3, i.e., $\alpha(k) = O(\lambda^k)$ for some $0 < \lambda < 1$. Let the integer sequences $\{\widetilde{p}_k\}$, $\{\widetilde{q}_k\}$, and $\{\widetilde{r}_k\}$ be defined as follows:

$$\widetilde{p}_k \triangleq \lfloor (kd_k)^\beta \rfloor,\ \widetilde{q}_k \triangleq \lfloor \sqrt{\widetilde{p}_k} \rfloor,\ \widetilde{r}_k \triangleq \left\lfloor \frac{k}{\widetilde{p}_k + \widetilde{q}_k} \right\rfloor,$$

where $\beta(\Delta - 1) < \Delta/2 - 1$ for some $\Delta > 2$, and $\lfloor x \rfloor$ denotes the integer part of a real number x. Clearly, $0 < \beta < 1/2$. Further, define

$$\varphi_m = \sum_{j=k_m+1}^{l_m} \sqrt{\frac{d_k}{k}} Z_j,\ \varphi'_m = \sum_{j=l_m+1}^{l_m+\widetilde{q}_k} \sqrt{\frac{d_k}{k}} Z_j,$$

$$\varphi''_{\widetilde{r}+1} = \sum_{j=\widetilde{k}+1}^{k} \sqrt{\frac{d_k}{k}} Z_j,$$

where $\widetilde{k} \triangleq \widetilde{r}_k(\widetilde{p}_k + \widetilde{q}_k)$, and $k_m \triangleq (m-1)(\widetilde{p}_k + \widetilde{q}_k)$, $l_m \triangleq (m-1)(\widetilde{p}_k + \widetilde{q}_k) + \widetilde{p}_k$ for $m = 1, \ldots, \widetilde{r}_k$. Define also the partial sums

$$S_k = \sum_{m=1}^{\widetilde{r}_k} \varphi_m,\ S'_k = \sum_{m=1}^{\widetilde{r}_k} \varphi'_m,\ S''_k = \varphi''_{\widetilde{r}+1}.$$

Then we have $\sqrt{kd_k} J_{13}(x) v(x) = S_k + S'_k + S''_k$. Let us first show that ES'^2_k and ES''^2_k converge to zero. Clearly,

$$d_k \mathrm{Var}(K_{d_i}(x)) = \frac{d_k}{d_i} v(x) \int K(t)^2 \mathrm{d}t - d_k v(x)^2 = O(1)$$

for all $1 \leq i \leq k$ since d_k monotonically decreases and $\mathrm{Cov}(Z_i, Z_j) = O(1)$ for $1 \leq i < j \leq k$.

Observe that

$$ES'^2_k = \sum_{m=1}^{\widetilde{r}_k} \mathrm{Var}(\varphi'_m) + 2 \sum_{1 \leq i < j \leq \widetilde{r}_k} \mathrm{Cov}(\varphi'_i, \varphi'_j),$$

where at the righthand side the first term is estimated as

$$\sum_{m=1}^{\widetilde{r}_k} \mathrm{Var}(\varphi'_m) = \frac{d_k}{k} \sum_{m=1}^{\widetilde{r}_k} \sum_{j=l_m+1}^{l_m+\widetilde{q}_k} \mathrm{Var}(Z_j) + 2\frac{d_k}{k} \sum_{m=1}^{\widetilde{r}_k} \sum_{l_m+1 \leq i < j \leq l_m+\widetilde{q}_k} \mathrm{Cov}(Z_i, Z_j) \leq O\Big(\frac{\widetilde{q}_k \widetilde{r}_k}{k}\Big) + O\Big(\frac{d_k \widetilde{r}_k \widetilde{q}^2_k}{k}\Big) \to 0,$$

while the second term is estimated as

$$\sum_{1 \leq i < j \leq \widetilde{r}_k} \mathrm{Cov}(\varphi'_i, \varphi'_j) = \frac{d_k}{k} \sum_{1 \leq i < j \leq \widetilde{r}_k} \sum_{s=l_i+1}^{l_i+\widetilde{q}_k} \sum_{t=l_j+1}^{l_j+\widetilde{q}_k} \mathrm{Cov}(Z_s, Z_t)$$

$$= \frac{d_k}{k} \sum_{i=1}^{\widetilde{r}_k-1} \sum_{j=i+1}^{\widetilde{r}_k} \sum_{s=l_i+1}^{l_i+\widetilde{q}_k} \sum_{t=l_j+1}^{l_j+\widetilde{q}_k} \frac{\big(\alpha(t-s)\big)^{\frac{\Delta-2}{\Delta}}}{(d_s d_t)^{(\Delta-1)/\Delta}}$$

$$\leq O\left(\frac{d_k \widetilde{q}^2_k \widetilde{r}_k}{k d_k^{2(\Delta-1)/\Delta}} \sum_{l=1}^{\widetilde{r}_k-1} \big(\alpha(l\widetilde{p}_k)\big)^{\frac{\Delta-2}{\Delta}} \right)$$

$$= O\left(\frac{\widetilde{q}^2_k \widetilde{r}_k}{k d_k^{1-2/\Delta}} \frac{\lambda^{\widetilde{p}_k(\Delta-2)/\Delta}(1 - \lambda^{\widetilde{p}_k(\Delta-2)(\widetilde{r}_k-1)/\Delta})}{1 - \lambda^{\widetilde{p}_k(\Delta-2)/\Delta}} \right)$$

$$= O\left(\frac{\widetilde{q}^2_k \widetilde{r}_k \lambda^{\widetilde{p}_k(\Delta-2)/\Delta}}{k d_k^{1-2/\Delta}} \right) \to 0,$$

where Theorem A4 is used. Analogously, one derives

$$ES''^2_k = \frac{d_k}{k} \sum_{j=\widetilde{k}+1}^{k} \mathrm{Var}(Z_j) + 2\frac{d_k}{k} \sum_{\widetilde{k}+1 \leq i < j \leq k} \mathrm{Cov}(Z_i, Z_j) \leq O\Big(\frac{\widetilde{p}_k + \widetilde{q}_k}{k}\Big) + O\Big(\frac{d_k(\widetilde{p}_k + \widetilde{q}_k)^2}{k}\Big) \to 0.$$

Therefore, one needs only to show that S_k in distribution converges to $\mathscr{N}(0, \chi^2(x) v^2(x))$. By Theorem A5 we have

$$\left| E\left[\prod_{m=1}^{\widetilde{r}_k} \exp(jt\varphi_m) \right] - \prod_{m=1}^{\widetilde{r}_k} E\left[\exp(jt\varphi_m)\right] \right| \leq 16(\widetilde{r}_k - 1)\alpha(\widetilde{q}_k) \to 0,$$

since $\alpha(k)$ exponentially tends to zero, where j is the imaginary unit. This means that S_k and the random variable $\sum_{m=1}^{\widetilde{r}_k} Y_m$ asymptotically are identically distributed as $k \to \infty$, where $\{Y_m, m = 1, \ldots, \widetilde{r}_k\}$ are mutually independent with $EY_m = 0$ and Y_m and φ_m have the same distribution. So, to establish (54), it remains to show that the distribution of $\sum_{m=1}^{\widetilde{r}_k} Y_m$ converges to $\mathscr{N}(0, \chi^2(x) v^2(x))$. Clearly,

$$\sum_{m=1}^{\widetilde{r}_k} EY^2_m = \sum_{m=1}^{\widetilde{r}_k} E\varphi^2_m = \frac{d_k}{k} \left(\sum_{m=1}^{\widetilde{r}_k} \sum_{j=k_m+1}^{l_m} EZ^2_j + 2 \sum_{m=1}^{\widetilde{r}_k} \sum_{k_m+1 \leq i < j \leq l_m} \mathrm{Cov}(Z_i, Z_j) \right).$$

For the first term at its righthand side,

$$\frac{d_k}{k} \sum_{m=1}^{\widetilde{r}_k} \sum_{j=k_m+1}^{l_m} EK^2_{d_j}(x) Ew^2_{j+1} = \left(\frac{1}{k} \sum_{m=1}^{\widetilde{r}_k} \sum_{j=k_m+1}^{l_m} \frac{d_k}{d_j} \right) v(x) \int K^2(t) \mathrm{d}t Ew^2_k + O(d^2_k) \to \chi^2(x) v^2(x).$$

For dealing with the last term, let us divide the index set $\{(i,j) \mid k_m+1 \le i < j \le l_m\}$ into two disjoint subsets $Q_1 \triangleq \{(i,j) \mid i,j \in \{k_m+1,\ldots,l_m\}, 1 \le j-i \le t_k\}$ and $Q_2 \triangleq \{(i,j) \mid i,j \in \{k_m+1,\ldots,l_m\}, t_k < j-i < \widetilde{p}_k\}$, where t_k is such that $t_k \to \infty$ and $d_k t_k \to 0$. Thus, we have

$$\frac{2d_k}{k}\sum_{m=1}^{\widetilde{r}_k}\sum_{k_m+1\le i<j\le l_m}\mathrm{Cov}(Z_i,Z_j)$$
$$=\frac{2d_k}{k}\sum_{m=1}^{\widetilde{r}_k}\sum_{Q_1}\mathrm{Cov}(Z_i,Z_j)+\frac{2d_k}{k}\sum_{m=1}^{\widetilde{r}_k}\sum_{Q_2}\mathrm{Cov}(Z_i,Z_j),$$

where the first term at the righthand side is

$$\frac{2d_k}{k}\sum_{m=1}^{\widetilde{r}_k}\sum_{Q_1}\mathrm{Cov}(Z_i,Z_j)=O\left(\frac{d_k t_k \widetilde{p}_k \widetilde{r}_k}{k}\right)=o(1),$$

while the last term item is estimated by

$$\frac{2d_k}{k}\sum_{m=1}^{\widetilde{r}_k}\sum_{Q_2}\mathrm{Cov}(Z_i,Z_j)$$
$$=O\left(\sum_{m=1}^{\widetilde{r}_k}\sum_{Q_2}\frac{d_k}{k}\left(\frac{1}{d_i d_j}\right)^{\frac{\Delta-1}{\Delta}}(\alpha(j-i))^{\frac{\Delta-2}{\Delta}}\right)$$
$$=O\Big(\Big[d_k^{\frac{2}{\Delta}-1}\sum_{l=t_k+1}^{\widetilde{p}_k-1}\left(\lambda^{\frac{\Delta-2}{\Delta}}\right)^l\Big]\Big[\frac{1}{k}\sum_{m=1}^{\widetilde{r}_k}\sum_{i=1}^{\widetilde{p}_k-1}\left(\frac{d_k}{d_i}\right)^{\frac{2(\Delta-1)}{\Delta}}\Big]\Big)$$
$$=O\Big(d_k^{\frac{2}{\Delta}-1}\left(\lambda^{\frac{\Delta-2}{\Delta}}\right)^{t_k}\Big)=o(1),$$

where the covariance inequality given in Theorem A4 is used.

Therefore, we have $\sum_{m=1}^{\widetilde{r}_k} EY_m^2 \to \chi^2(x)v^2(x)$. Using the C_r-inequality [34, Page 6] leads to

$$E|Y_m|^\Delta = E|\varphi_m|^\Delta \le \left(\frac{d_k}{k}\right)^{\Delta/2}\widetilde{p}_k^{\Delta-1}\sum_{j=k_m+1}^{l_m}E|Z_j|^\Delta$$
$$=O\left(\left(\frac{d_k}{k}\right)^{\Delta/2}\widetilde{p}_k^{\Delta-1}\sum_{j=k_m+1}^{l_m}\frac{1}{d_j^{\Delta-1}}\right),$$

which implies

$$\sum_{m=1}^{\widetilde{r}_k}E|Y_m|^\Delta=O\left(\frac{\widetilde{p}_k^{\Delta-1}}{(kd_k)^{\Delta/2-1}}\left(\frac{1}{k}\sum_{m=1}^{\widetilde{r}_k}\sum_{j=k_m+1}^{l_m}\left(\frac{d_k}{d_j}\right)^{\Delta-1}\right)\right)$$
$$=O\left(\frac{\widetilde{p}_k^{\Delta-1}}{(kd_k)^{\Delta/2-1}}\right)=o(1).$$

Thus, we have shown that

$$\sum_{m=1}^{\widetilde{r}_k}E|Y_m|^\Delta\Big/\left(\sum_{m=1}^{\widetilde{r}_k}EY_m^2\right)^{\Delta/2}=o(1).$$

By Theorem A8 we conclude that $\sum_{m=1}^{\widetilde{r}_k}Y_m$ converges to $\mathcal{N}(0,\chi^2(x)v^2(x))$ in distribution, and the asymptotic normality (54) has now been established. ■

References

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

[2] 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.

[3] 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.

[4] B. Ninness and S. Gibson, "Quantifying the accuracy of Hammerstein model estimation," *Automatica*, vol. 38, no. 12, pp. 2037–2051, 2002.

[5] 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.

[6] A. Krzyżak, "Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1323–1338, Aug. 1992.

[7] A. Krzyżak, "Identification of nonlinear block-oriented systems by the recursive kernel estimate," *J. Franklin Inst.*, vol. 330, no. 3, pp. 605–627, 1993.

[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] 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.

[10] W. 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.

[11] P. Stoica, "On the convergence of an iterative algorithm used for Hammerstein system identification," *IEEE Trans. Autom. Control*, vol. 26, no. 4, pp. 967–969, Apr. 1981.

[12] E. J. Dempsey and D. T. Westwick, "Identification of Hammerstein models with cubic spline nonlinearities," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 2, pp. 237–245, Feb. 2004.

[13] H.-F. Chen, "Strong consistency of recursive identification for Hammerstein systems with discontinuous piecewise-linear memoryless block," *IEEE Trans. Autom. Control*, vol. 50, no. 10, pp. 1612–1617, Oct. 2005.

[14] X. Hong, S. Chen, Y. Gong, and C. J. Harris, "Nonlinear equalization of Hammerstein OFDM systems," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5629–5639, Nov. 2014.

[15] 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.

[16] G. Li, C. Wen, W. X. Zheng, and Y. Chen, "Identification of a class of nonlinear autoregressive models with exogenous inputs based on kernel machines," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2146–2159, May 2011.

[17] G. Li, C. Wen, and A. Zhang, "Fixed point iteration in identifying bilinear models," *Syst. Control Lett.*, vol. 83, pp. 28–37, 2015.

[18] K. S. Narendra and P. G. Gallman, "An iterative method for the identification of nonlinear systems using a Hammerstein model," *IEEE Trans. Autom. Control*, vol. 11, no. 3, pp. 546–550, Mar. 1966.

[19] 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.

[20] Y. Liu and E. W. Bai, "Iterative identification of Hammerstein systems," *Automatica*, vol. 43, no. 2, pp. 346–354, Feb. 2007.

[21] E. Masry, "Recursive probability density estimation for weakly dependent stationary processes," *IEEE Trans. Inf. Theory*, vol. 32, no. 2, pp. 254–267, Apr. 1986.

[22] W. 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, June 2010.

[23] W. Zhao, W. X. Zheng, and E.-W. Bai, "A recursive local linear estimator for identification of nonlinear ARX systems: Asymptotical convergence and applications," *IEEE Trans. Autom. Control*, vol. 58, no. 12, pp. 3054–3069, Dec. 2013.

[24] Y. Huang, X. Chen, and W. B. Wu, "Recursive nonparametric estimation for time series," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1301–1312, Apr. 2014.

[25] B.-Q. Mu and H.-F. Chen, "Recursive identification of Wiener-Hammerstein systems," *SIAM J. Control Optim.*, vol. 50, no. 5, pp. 2621–2658, 2012.

[26] B.-Q. Mu, H.-F. Chen, L. Y. Wang, and G. Yin, "Characterization and identification of matrix fraction descriptions for LTI systems," *SIAM J. Control Optim.*, vol. 52, no. 6, pp. 3694–3721, 2014.

[27] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, no. 9, pp. 907–939, Sept. 1982.

[28] B. Aksasse, L. Badidi, and L. Radouane, "A rank test based approach to order estimation. I. 2-D AR models application," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 2069–2072, July 1999.

[29] W. Greblicki and M. Pawlak, *Nonparametric System Identification*. New York: Cambridge University Press, 2008.

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

[31] H.-F. Chen and W. Zhao, "New method of order estimation for ARMA/ARMAX processes," *SIAM J. Control Optim.*, vol. 48, no. 6, pp. 4157–4176, 2010.

[32] J. D. Hart and T. E. Wehrly, "Kernel regression when the boundary region is large, with an application to testing the adequacy of polynomial models," *J. Amer. Stat. Assoc.*, vol. 87, no. 420, pp. 1018–1024, 1992.

[33] H.-G. Müller, "Smooth optimum kernel estimators near endpoints," *Biometrika*, vol. 78, no. 3, pp. 521–530, 1991.

[34] H.-F. Chen and W. Zhao, *Recursive Identification and Parameter Estimation*. Boca Raton, FL: CRC Press, 2014.

[35] C.-Z. Wei, "Asymptotic properties of least-squares estimates in stochastic regression models," *Ann. Stat.*, vol. 13, no. 4, pp. 1498–1508, 1985.

[36] A. Mokkadem, "Mixing properties of ARMA processes," *Stoch. Process. Their Appl.*, vol. 29, no. 2, pp. 309–315, 1988.

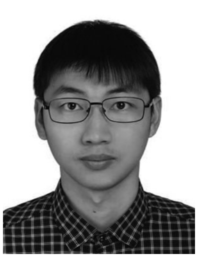
[37] C. M. Deo, "A note on empirical processes of strong-mixing sequences," *Ann. Probab.*, vol. 1, no. 5, pp. 870–875, 1973.

[38] V. Volkonskii and Y. A. Rozanov, "Some limit theorems for random functions. I," *Theory Probab. Appl.*, vol. 4, no. 2, pp. 178–197, 1959.

[39] 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.

[40] J. L. Doob, *Stochastic Processes*. New York: Wiley, 1953.

[41] G. G. Roussas and L. T. Tran, "Asymptotic normality of the recursive kernel regression estimate under dependence conditions," *Ann. Stat.*, vol. 20, no. 1, pp. 98–120, 1992.



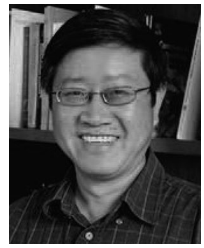
Biqiang Mu was born in Sichuan, China in 1986. He received the B.Eng. degree in material formation and control engineering from Sichuan University, China, in 2008 and the Ph.D. degree in operations research and cybernetics from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, in 2013.

He was a Postdoctoral Fellow at the Wayne State University from 2013 to 2014 and also at the Western Sydney University from 2015 to 2016. He is currently an Assistant Professor at the Academy of Mathematics and Systems Science, Chinese Academy of Sciences. His research interests include system identification and applications.

Han-Fu Chen (SM'94–F'97) received the Ph.D. degree from the Leningrad (St. Petersburg) State University, Leningrad, 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. He authored and coauthored more than 200 journal papers and seven books. His research interests are mainly in stochastic systems, including system identification, adaptive control, and stochastic approximation and its applications.

Dr. Chen is an IFAC Fellow, a Member of TWAS, and a Member of Chinese Academy of Sciences. 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).



Le Yi Wang (S'85–M'89–SM'01–F'12) received the Ph.D. degree in electrical engineering from McGill University, Montreal, QC, Canada, in 1990.

Since 1990, he has been with Wayne State University, Detroit, Michigan, where he is currently a professor in the Department of Electrical and Computer Engineering. His research interests are in the areas of complexity and information, system identification, robust control, H-infinity optimization, time-varying systems, adaptive systems, hybrid and nonlinear systems, information processing and learning, as well as medical, automotive, communications, power systems, and computer applications of control methodologies. He was a keynote speaker in several international conferences.

Dr. Wang serves on the IFAC Technical Committee on Modeling, Identification and Signal Processing. He was an Associate Editor of the IEEE Transactions on Automatic Control and several other journals, and currently is an Associate Editor of Journal of Control Theory and Applications. He was a Visiting Faculty at University of Michigan in 1996 and Visiting Faculty Fellow at University of Western Sydney in 2009 and 2013. He is a member of a Foreign Expert Team in Beijing Jiao Tong University and a member of the Core International Expert Group at Academy of Mathematics and Systems Science, Chinese Academy of Sciences.

George Yin (S'87–M'87–SM'96–F'02) received the B.S. degree in mathematics from the University of Delaware in 1983, M.S. degree in electrical engineering, and Ph.D. in applied mathematics from Brown University in 1987.

He joined Wayne State University in 1987, and became a Professor in 1996. His research interests include stochastic systems and applications.

Dr. Yin is a Fellow of SIAM and a Fellow of IFAC. He served on the IFAC Technical Committee on Modeling, Identification and Signal Processing, and many conference program committees; he was Co-Chair of SIAM Conference on Control & Its Application, 2011, and Co-Chair of a couple AMS-IMS-SIAM Summer Research Conferences; he also chaired a number SIAM prize selection committees. He is Chair SIAM Activity Group on Control and Systems Theory, and serves on the Board of Directors of American Automatic Control Council. He is an Associate Editor of SIAM Journal on Control and Optimization, and on the editorial board of a number of other journals. He was an Associate Editor of Automatica (2005–2011) and IEEE Transactions on Automatic Control (1994–1998).

Wei Xing Zheng (M'93–SM'98–F'14) received the B.Sc. degree in applied mathematics, the M.Sc. degree in electrical engineering, and the Ph.D. degree in electrical engineering, all from Southeast University, Nanjing, China, in 1982, 1984, and 1989, respectively.

He is currently a Professor at Western Sydney University, Sydney, Australia. Over the years he has also held various faculty/research/visiting positions at Southeast University, China, Imperial College of Science, Technology and Medicine, UK, University of Western Australia, Curtin University of Technology, Australia, Munich University of Technology, Germany, University of Virginia, USA, and University of California-Davis, USA. His research interests are in the areas of systems and controls, signal processing, and communications.

Dr. Zheng is a Fellow of IEEE. He served as an Associate Editor for IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications, IEEE Transactions on Automatic Control, IEEE Signal Processing Letters, IEEE Transactions on Circuits and Systems-II: Express Briefs, and IEEE Transactions on Fuzzy Systems, and as a Guest Editor for IEEE Transactions on Circuits and Systems-I: Regular Papers. Currently, he is an Associate Editor for Automatica, IEEE Transactions on Automatic Control (the second term), IEEE Transactions on Cybernetics, IEEE Transactions on Neural Networks and Learning Systems, and other scholarly journals. He is also an Associate Editor of IEEE Control Systems Society's Conference Editorial Board.