Pathwise Convergence of Recursive Identification Algorithms for Hammerstein Systems

Han-Fu Chen, *Fellow, IEEE*

***Abstract*—This paper gives estimates for: 1) coefficients contained in the linear part of the Hammerstein system; 2) the value of the nonlinear function $f(u)$ in the Hammerstein system at any u; 3) $Ef(u_k)$ and $Eu_kf(u_k)$ with u_k denoting the system input. No assumption is made on structure of $f(\cdot)$. The estimates given by the stochastic approximation algorithms with expanding truncations are recursive and convergent to the true values with probability one. Two numerical examples are given.**

***Index Terms*—Hammerstein system, nonparametric nonlinearity, recursive estimate, stochastic approximation, strong consistency.**

I. Introduction

THE single-input–single-output (SISO) Hammerstein system considered in the paper consists of two parts: A nonlinear memoryless element $f(\cdot)$ and a moving average type linear subsystem with disturbance as presented in Fig. 1.

By u_k and y_k we denote the system input and output, respectively, and by z_k the observation

$$z_k = y_k + \xi_k \tag{1}$$

where ξ_k is the observation noise. $\{y_k\}$ and $\{v_k\}$ are related by the linear system as follows:

$$y_{k+1} = \sum_{j=0}^{r} d_j v_{k-j} \quad d_0 = 1 \quad v_j \triangleq f(u_j). \tag{2}$$

Because of its importance in engineering applications (see, e.g., [9], [10], and [21] among others), the Hammerstein system, in particular, its identification issue has been an active research topic for many years. When identifying the system presented in Fig. 1, the only available information is the sequence $\{u_k, z_k\}$, where $\{u_k\}$ is designed by users for identification purpose. Based on $\{u_k, z_k\}$ we want to identify both the nonlinear function $f(\cdot)$ and the linear system with v_k and y_k as its input and output, respectively. This problem differs from those considered in [6], [16] for ARMAX systems. First of all, here a nonlinearity $f(\cdot)$ is involved. Even for the linear part, the identification problem is distinguished from that discussed in [6] and [16], because here the input $\{v_k\}$ for the linear system is unavailable.

Manuscript received May 21, 2003; revised November 28, 2003. Recommended by Associate Editor A. Garulli. This work was supported by the National Natural Science Foundation of China and by the Ministry of Science and Technology of China.

The author is with the Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, P. R. China (e-mail: hfchen@control.iss.ac.cn).

Digital Object Identifier 10.1109/TAC.2004.835358

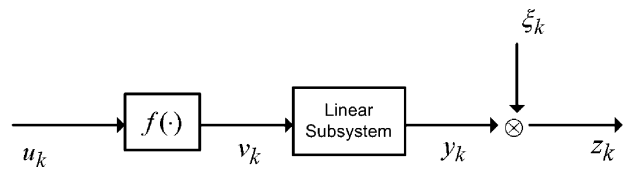

Fig. 1. Hammerstein system.

For characterizing the nonlinearity, both parametric [2]–[4], [8], [15], [18]–[20] and nonparametric [1], [12]–[14], [17] approaches are used, but almost all existing identification methods are nonrecursive and almost no result is on convergence with probability one with only a few possible exceptions to be addressed later on.

In the parametric approach, the nonlinearity is often considered as a polynomial with unknown coefficients [2], [18], [19] and, hence, the system can be written in a linear regression form with respect to coefficients of the linear subsystem and products of coefficients in both the polynomial and the linear subsystem. Therefore, identification methods developed for ARMAX systems are possible to be applied to this case. Besides the polynomial type of nonlinearity, other types of parametric nonlinearity are also discussed, e.g., linear functions with dead zone in [20], the multilayer feedforward neural network in [3], etc.

In [14], a nonparametric nonlinear function $f(\cdot)$ is considered. An approximating polynomial is constructed and shown to converge to $f(\cdot)$ in probability and in the mean square sense in a finite interval as the sample size n increases. However, for each n the whole input sequence $\{u_k, k = 1, \ldots, n\}$ has to be redefined with distribution different from that for $n-1$, and hence the approximating polynomial has to be reconstructed at each step without recursion. In [17], the nonlinear function $f(\cdot)$ multiplied by the input density is expanded to the series of Legendre polynomials and the identification problem is reduced to estimating coefficients of the first $N(n)$ terms in the series. Since the number $N(n)$ of parameters to be estimated increases with n, the identification method is nonrecursive too. In the most recent paper [1] dealing with frequency domain identification, instead of white noise, the sinusoidal inputs are applied to a continuous-time Hammerstein system, and the nonlinear function is expanded to a Fourier series. As in [17], the nonlinearity identification reduces to estimating coefficients in the Fourier expansion. The estimates are nonrecursive, and the convergence in probability is proved.

As previously mentioned, up until now there have been only a few papers on recursive and pathwise convergent algorithms for identifying the Hammerstein system; [2] and [12] are possibly among them. In [2], the nonlinearity is parameterized by a polynomial with unknown coefficients and the extended least squares (ELS) algorithm is used to estimate: 1) coefficients of the polynomial; 2) coefficients of the auto-regression part of the linear subsystem; and 3) products of coefficients contained in both the polynomial and the moving average part of the linear subsystem. In order to finally obtain estimates for coefficients of the moving average part the authors apply the LS method again. The algorithm presented there is recursive, but it seems that the authors aim at showing unbiasedness (see [2, eq. (21)]) rather than strong consistency of the estimate. For strong consistency of ELS we refer to [6], which shows that conditions listed in [2, Th. 2] are not sufficient even ignoring the second application of LS in [2].

In [12], the author based on stochastic approximation has nicely presented identification algorithms for both nonlinear and linear parts in a recursive way and has proved their convergence in the mean squares sense. However, there are some problems that remain unsolved. To be specific, instead of estimating the impulse response $k(i)(=c^T A^{i-1} b)$, and the value $m(u)$ of the unknown function $m(\cdot)$ (i.e., $f(\cdot)$ in the notation of this paper) at an arbitrary point u, the author of [12] actually estimates $\beta k(i)$ and $\alpha m(u)+\gamma$, respectively, where $\beta = E[u_0 m(u_0)]/Eu_0^2$, $\alpha = c^T b$, and $\gamma = c^T A E x_0$ with c, b, A, and x_0 given by the state representation of the linear system

$$X_{n+1} = AX_n + bm(u_n) \quad y_n = c^T X_n.$$

We see that β depends on the correlation between the input and output of the nonlinearity, and α and γ depend on the unknown state representation of the linear system. Consequently, coefficients α, β, and γ are unknown, and based on the algorithms given by [12] one can still not completely identify the Hammerstein system unless to assume some of them to be known.

Concerning the system discussed in the paper it is noticed that considering the finite order r is a restriction, which is not imposed in [12], [13], and [17], but any stable linear system with constant coefficients can be expressed as a moving average process of infinite order, and, hence, (2) with sufficiently large r gives a good approximation. It is also noticed that the observation noise ξ_k may not be zero mean. In contrast to $E\xi_k = 0$ as required in the previous works, we allow $E\xi_k \neq 0$ but require $E\xi_k \underset{k\to\infty}{\longrightarrow} 0$.

The purpose of this paper is to identify the Hammerstein system in a recursive way and to provide almost surely convergent estimates. More precisely, $f(u)$ for any $u, d_j, i=1,\ldots,r$, as well as $Ef(u_k)$ and $Eu_k v_k$ are to be strongly consistently estimated. We share the idea proposed in [12] of using stochastic approximation with kernels for identifying Hammerstein systems, but the algorithms and the analysis method used here differ from those given in [12]. The algorithms are now truncated at expanding bounds and a sample-path based TS (trajectory-subsequence) method is applied for analyzing convergence of estimates with probability one.

The rest of the paper is arranged as follows. In Section II, the algorithms are described and conditions to be used are listed. In Section III, strong consistency of estimates for coefficients in the linear system is proved, while strong consistency of the estimate for $f(u)$ is shown in Section IV. Two numerical examples are presented in Section V. Some concluding remarks are given in Section VI.

II. Estimation Algorithms

Let the SISO Hammerstein system under consideration be presented by Fig. 1 with observation and linear part given by (1) and (2), respectively. For any fixed $u \in R$, it is required to estimate $f(u)$, $Ef(u_k)$, $Eu_k f(u_k)$, and the coefficients d_i, $i=1,\ldots,r$ of the linear system (2).

We now define system input $\{u_k\}$ and the kernel function to be used in the identification algorithms. Let $\{u_k\}$ be a sequence of bounded independent and identically distributed (iid) random variables $|u_k| < c_1$, $\forall i = 1,2,\ldots$, with $Eu_k = 0$ and with density $p(\cdot)$, where $c_1 > 0$ is a constant $c_1 \neq |u|$ and and $p(\cdot)$ is continuous at u with $p(u) > 0$. Let $\{u_k\}$ be also independent of the observation noise $\{\xi_k\}$.

The kernel function w_k to be used for estimating $f(u)$ is defined as follows:

$$w_k \triangleq \frac{1}{b_k} e^{-((u_k-u)/b_k)^2} \tag{3}$$

where $b_k = 1/k^\delta$ with $\delta \in (0, 1/2)$.

It is clear that

$$\begin{aligned} Ew_k &= \frac{1}{b_k}\int_{-c_1}^{c_1} e^{-((x-u)/b_k)^2} p(x)dx \\ &= \frac{1}{\sqrt{2}}\int_{-\sqrt{2}(c_1+u)/b_k}^{\sqrt{2}(c_1-u)/b_k} e^{-t^2/2} p\left(u+\frac{b_k t}{\sqrt{2}}\right) dt \underset{k\to\infty}{\longrightarrow} \sqrt{\pi} p(u) \end{aligned} \tag{4}$$

and

$$\begin{aligned} &E(\sqrt{b_k} w_k)^2 \\ &\quad = \frac{1}{b_k}\int_{-c_1}^{c_1} e^{-2((x-u)/b_k)^2} p(x)dx \\ &\quad = \frac{1}{2}\int_{-2(c_1+u)/b_k}^{2(c_1-u)/b_k} e^{-t^2/2} p\left(u+\frac{b_k t}{2}\right) dt \underset{k\to\infty}{\longrightarrow} \sqrt{\frac{\pi}{2}} p(u). \end{aligned} \tag{5}$$

Remark 1: In lieu of e^{-x^2} we may take any other measurable function $K(\cdot)$ satisfying the following conditions:

$$\int_{-\infty}^{\infty} |K(x)|dx < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} K^2(x)dx < \infty$$

to serve as the kernel function.

We impose the following conditions.

A1) The nonlinear function $f(\cdot)$ is measurable, locally bounded, and continuous at u.

A2) The observation noise $\{\xi_k\}$ is a sequence of mutually independent random variables with $E\xi_k \underset{k\to\infty}{\longrightarrow} 0$ and $\sup_k E\xi_k^2 < \infty$.

In what follows, denote

$$E\xi_k \triangleq r_k \quad \text{and} \quad \bar{\xi}_k \triangleq \xi_k - E\xi_k.$$

To consistently estimate $\{d_i\}$, conditions A1) and A2) are sufficient. It is worth noting that no assumption is made on the structure of $f(\cdot)$.

In order to uniquely define $f(u)$ and Ev_k, we need a condition to guarantee that the response of the linear subsystem to a nonzero constant input is nonzero.

A3) $\sum_{j=0}^{r} d_j \neq 0$.

In what follows, I_A always denotes the indicator of a set A:

$$I_A = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{otherwise} \end{cases}$$

for example, if X and Y are two random variables, then $I_{[X>Y]}$ equals 1 for those ω for which $X(\omega) > Y(\omega)$, and zero, otherwise.

For estimating $d_i, i = 1, \ldots, r$ and Eu_1v_1 we define the stochastic approximation algorithms with expanding truncations, shown in (6)–(7) at the bottom of the page, with an initial value $\theta_0(i), i = 0, 1, \ldots, r$, where $a_k = 1/k$ and $\{M_k\}$ is a sequence of increasing real numbers diverging to infinity

$$M_k > 0 \quad M_{k+1} > M_k \quad \forall k, \text{ and } M_k \underset{k\to\infty}{\longrightarrow} \infty.$$

It is worth noting that $\theta_k(0)$ is used to estimate $\rho \triangleq Eu_1v_1$, while $\theta_k(i), i = 1, \ldots, r$ are used to estimate $\rho d_i, i = 1, \ldots, r$, respectively.

We explain the algorithm (6)–(7). The truncations at expanding bounds applied in (6) are used to prevent the estimates from diverging to infinity. $\sigma_k(i)$ is the number of truncations occurred up-to time k. When calculating $\theta_{k+1}(i)$ the truncation bound is $M_{\sigma_k(i)}$. If a truncation has actually happened, then together with pulling back the algorithm to zero the truncation bound is extended from $M_{\sigma_k(i)}$ to $M_{\sigma_k(i)+1}$ for the next step. As to be pointed out in Remark 4, the truncation of the algorithm ceases in a finite time and, hence, only a finite number of $\{M_k\}$ is used in the algorithm. Therefore, the asymptotic convergence rate of $\theta_k(i)$ should not depend on the selection of $\{M_k\}$.

We now define algorithms for estimating $f(u)$ and $Ef(u_1)$

$$\gamma_{k+1} = \begin{cases} \gamma_k - a_k(\gamma_k - z_{k+1}), & \text{if } |\gamma_k - a_k(\gamma_k - z_{k+1})| \le M_{\nu_k} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

$$\nu_k = \sum_{j=1}^{k-1} I_{[|\gamma_j - a_j(\gamma_j - z_{j+1})| > M_{\nu_j}]}, \qquad \nu_0 = 0 \tag{9}$$

with an initial value γ_0, and (10)–(11), as shown at the bottom of the page, with an initial value $\mu_0(u)$, where w_k is the kernel given by (3), $a_k = 1/k$, $b_k = 1/k^\delta$ with $\delta \in (0, 1/2)$, and $\{M_k\}$ may be any sequence of positive and increasing real numbers diverging to infinity, not necessarily to be the same as that used in (6)–(7).

Both (8)–(9) and (10)–(11) are also stochastic approximation algorithms with expanding truncations. The estimate $\{\gamma_k\}$ defined by (8)(9) is used to estimate $Ef(u_1)\sum_{j=0}^{r} d_j$, while $\{\mu_k(u)\}$ defined by (10)(11) is for estimating

$$f(u) + \sum_{j=1}^{r} d_j Ef(u_1).$$

It is worth pointing out that we cannot expect a fast rate of convergence from all algorithms (6)–(7), (8)–(9), and (10)–(11), since the asymptotic rate of stochastic approximation algorithms is not faster than $O(1/\sqrt{k})$. To be more precise, if the stochastic approximation algorithm converges to the root x^0 of a regression function $h(\cdot)$ with step-size equal to $1/k$, then the rate of convergence is $o(1/k^\delta)$ as $k \to \infty$ for any $\delta \in (0, 1/2)$ such that $H + \delta I$ remains stable, where H is the Hessian of $h(\cdot)$ at x^0. For details, we refer to [5, Secs. 3.1 and 3.2].

Remark 2: The sequences $\{a_k\}$ and $\{b_k\}$ may be more general. As a matter of fact, any sequences of positive real numbers satisfying the following conditions (12)–(13) work well:

$$a_k > 0 \quad a_k \underset{k\to\infty}{\longrightarrow} 0 \quad \sum_{k=1}^{\infty} a_k = \infty \tag{12}$$

$$b_k > 0 \quad b_k \underset{k\to\infty}{\longrightarrow} 0 \quad \frac{a_k}{b_k} \underset{k\to\infty}{\longrightarrow} 0, \text{ and } \sum_{k=1}^{\infty} \frac{a_k^2}{b_k} < \infty. \tag{13}$$

III. Strongly Consistent Estimates for $\{d_i\}$, Eu_1v_1, and Ev_1

In this section, we prove that estimates given by (6), (7), and (8), (9) are strongly consistent, i.e., we recursively derive consistent estimates for coefficients of the linear system, the cor-

$$\theta_{k+1}(i) = \begin{cases} \theta_k(i) - a_k(\theta_k(i) - u_k z_{k+1+i}), & \text{if } |\theta_k(i) - a_k(\theta_k(i) - u_k z_{k+1+i})| \le M_{\sigma_k(i)} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

$$\sigma_k(i) = \sum_{j=1}^{k-1} I_{[|\theta_j(i) - a_j(\theta_j(i) - u_j z_{j+1+i})| > M_{\sigma_j(i)}]}, \qquad \sigma_0(i) = 0 \tag{7}$$

$$\mu_{k+1}(u) = \begin{cases} \mu_k(u) - a_k w_k(\mu_k(u) - z_{k+1}), & \text{if } |\mu_k(u) - a_k w_k(\mu_k(u) - z_{k+1})| \le M_{\lambda_k(u)} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

$$\lambda_k(u) = \sum_{j=1}^{k-1} I_{[|\mu_j(u) - a_j w_j(\mu_j(u) - z_{j+1})| > M_{\lambda_j(u)}]}, \qquad \lambda_0(u) = 0 \tag{11}$$

relation between input and output of the nonlinearity, and the expectation of the output of nonlinearity.

To be precise, we have the following theorems.

Theorem 1: Assume A1) and A2) hold. Then $\theta_k(i)$, $i = 0, 1, \ldots, r$, defined by (6)–(7) are strongly consistent

$$\theta_k(0) \underset{k\to\infty}{\longrightarrow} \rho \triangleq E(u_1 f(u_1)) (= E u_1 v_1) \text{ a.s.} \tag{14}$$

and

$$\theta_k(i) \underset{k\to\infty}{\longrightarrow} \rho d_i \quad \text{a.s.}, \qquad i = 1, \ldots, r. \tag{15}$$

Remark 3: In order to obtain consistent estimates for d_i, the sequence $\{u_k\}$ should be selected such that $\rho \neq 0$. This can be done, since $f(\cdot)$ is not identically zero. Then, $\theta_k(i)/\theta_k(0) \underset{k\to\infty}{\longrightarrow} d_i$ a.s. Therefore, $\theta_k(0)$ should be monitored: If $\theta_k(0)$ approaches to zero, then the distribution of $\{u_k\}$ should be modified accordingly.

Theorem 2: Assume A1)–A3) hold and $\rho \neq 0$. Then

$$\theta_k(0)\gamma_k \left(\sum_{i=0}^{r} \theta_k(i)\right)^{-1} \underset{k\to\infty}{\longrightarrow} Ef(u_1) (= Ev_1) \text{ a.s.} \tag{16}$$

where $\theta_k(i)$ and γ_k are defined by (6)–(7) and (8)–(9), respectively.

The proof of these theorems is based on a general convergence theorem (GCT) for stochastic approximation algorithms with expanding truncations. For its proof, we refer to [5, Th. 2.2.1] or to [11, App.]. For convenience in reading this paper, let us formulate GCT for the special case where the root x^0 of $g(\cdot)$ is single: $g(x^0) = 0$.

We need the following conditions.

C1) $a_k > 0$, $a_k \underset{k\to\infty}{\longrightarrow} 0$, and $\sum_{k=1}^{\infty} a_k = \infty$.

C2) There is a continuously differentiable function $v(\cdot) : I\!R^l \to R$ such that

$$\sup_{\delta \le \|x - x^0\| \le \Delta} g^T(x) v_x(x) < 0$$

for any $\Delta > \delta > 0$, where $v_x(x)$ denotes the gradient of $v(\cdot)$. Further, x^* used in (18) is such that

$$v(x^*) < \inf_{\|x\|=c_0} v(x) \text{ for some } c_0 > 0 \text{ and } \|x^*\| < c_0.$$

C3) For the sample path ω under consideration

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left\| \sum_{i=n_k}^{m(n_k, T_k)} a_i \epsilon_{i+1} \right\| = 0 \qquad \forall T_k \in [0, T] \tag{17}$$

for any n_k such that x_{n_k} converges, where

$$m(k, T) = \max\left\{ m : \sum_{j=k}^{m} a_j \le T \right\}.$$

C4) $g(\cdot)$ is measurable and locally bounded.

We explain these conditions. C1) is an ordinary selection of the step-size. The first part of C2) means that the product $g^T(x) v_x(x)$ should be negative in-between two spheres centered at x^0 with any positive radiuses. For C3) to hold, either of the following conditions $\epsilon_k \underset{k\to\infty}{\longrightarrow} 0$ or $\sum_{k=1}^{\infty} a_k \epsilon_{k+1}$ is sufficient. In this case, (17) is satisfied along the whole sequence $\{x_n\}$, but in many applications before establishing the convergence of $\{x_n\}$, (17) may be verified along any convergent subsequences of $\{x_n\}$ rather than along the whole sequence. This can be seen from the proof of Theorem 3 to be given later on. C4) is very general without any growth rate restriction.

General Convergence Theorem (GCT for the Special Case of Single Root): Assume C1), C2), and C4) hold. Let $\{x_k\}$ be given by the following algorithm:

$$x_{k+1} = \begin{cases} x_k + a_k y_{k+1}, & \text{if } \|x_k + a_k y_{k+1}\| \le M_{\sigma_k} \\ x^*, & \text{otherwise} \end{cases} \tag{18}$$

$$\sigma_k = \sum_{i=1}^{k-1} I_{[\|x_i + a_i y_{i+1}\| > M_{\sigma_i}]} \quad \sigma_0 = 0 \tag{19}$$

$$y_{k+1} = g(x_k) + \epsilon_{k+1}. \tag{20}$$

Then $x_k \underset{k\to\infty}{\longrightarrow} x^0$ for those ω where C3 holds.

Remark 4: In the proof of GCT [5] it is shown that the number of truncations for each sample path ω where C3) holds is finite and, hence, only a finite number (which may depend on ω) of $\{M_k\}$ is used in the algorithm. Therefore, the selection of $\{M_k\}$ asymptotically should not effect the convergence rate of $\{x_k\}$.

Proof of Theorem 1: We rewrite (6) as shown in (21) at the bottom of the page, where

$$\epsilon_{k+1}(i) = -u_k z_{k+1+i} + d_i \rho, \qquad i = 0, 1, \ldots, r. \tag{22}$$

The linear function $g^{(i)}(x) \triangleq -(x - d_i\rho)$ in (21) corresponds to the regression function $g(x)$ in GCT. The root $x^0(i)$ of $g^{(i)}(\cdot)$ is unique $x^0(i) = \rho d_i$, $i = 0, 1, \ldots, r$. It is clear that $\|x - x^0(i)\|^2$ may serve as the Lyapunov function required in C2). The fixed point x^* in (18) is now equal to zero in (21). Thus, conditions C1), C2), and C4) are satisfied. Therefore, the assertion of Theorem 1 follows from GCT if we can show that

$$\lim_{T\to 0} \limsup_{n\to\infty} \frac{1}{T} \left| \sum_{j=n}^{m(n,t)} a_j \epsilon_{j+1}(i) \right| = 0 \qquad \forall t \in [0, T], \quad i = 0, 1, \ldots, r \tag{23}$$

For (23), it suffices to show that

$$\lim_{T\to 0} \limsup_{n\to\infty} \frac{1}{T} \sum_{j=n}^{m(n,t)} a_j u_j \left(\sum_{\substack{s=0 \\ s\neq i}}^{r} d_s v_{j+i-s} + r_{j+1+i} + \bar{\xi}_{j+i+1} \right) = 0 \tag{24}$$

$$\theta_{k+1}(i) = \begin{cases} \theta_k(i) - a_k(\theta_k(i) - d_i\rho) - a_k\epsilon_{k+1}(i), & \text{if } |\theta_k(i) - a_k(\theta_k(i) - d_i\rho) - a_k\epsilon_{k+1}(i)| \le M_{\sigma_k}(i) \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

and

$$\lim_{T\to 0}\limsup_{n\to\infty}\frac{1}{T}\sum_{j=n}^{m(n,t)} a_j(u_j d_i v_j - d_i\rho) = 0 \qquad \forall t\in[0,T]. \tag{25}$$

Notice that $\{u_j r_{j+i+1}, j = 1,2,\ldots\}$, $\{u_j\bar{\xi}_{j+i+1}, j = 1,2,\ldots\}$, and $\{u_j v_j - \rho\}$ are sequences of mutually independent random variables with zero mean and bounded second moments. Therefore, [7]

$$\sum_{j=1}^{\infty} a_j u_j r_{j+i+1} < \infty \text{ a.s.}, \ \sum_{j=1}^{\infty} a_j u_j \bar{\xi}_{j+i+1} < \infty \text{ a.s.}$$
$$\text{and } \sum_{j=1}^{\infty} a_j(u_j v_j - \rho) < \infty \text{ a.s.}$$

and for (24) and (25), it suffices to show

$$\lim_{T\to 0}\limsup_{n\to\infty}\frac{1}{T}\sum_{j=n}^{m(n,t)} a_j u_j \left(\sum_{\substack{s=0\\ s\neq i}}^{r} d_s v_{j+i-s}\right) = 0 \ \forall t\in[0,T]. \tag{26}$$

Rewriting $\sum_{j=n}^{m(n,t)} a_j u_j \sum_{\substack{s=0\\ s\neq i}}^{r} d_s v_{j+i-s}$ as the sum of two terms

$$\sum_{j=n}^{m(n,t)} a_j u_j \sum_{\substack{s=0\\ s\neq i}}^{r} d_s(v_{j+i-s} - Ev_1) + \sum_{j=n}^{m(n,t)} a_j u_j \sum_{\substack{s=0\\ s\neq i}}^{r} d_s Ev_1 \tag{27}$$

and noticing that $\{u_j(v_{j+i-s} - Ev_1)\}$ is a martingale difference sequence whatever $i > s$ or $i < s$ and both $\{u_j\}$ and $\{v_j\}$ are bounded, we see that [7]

$$\sum_{j=1}^{\infty} a_j u_j \sum_{\substack{s=0\\ s\neq i}}^{r} d_s(v_{j+i-s} - Ev_1) < \infty \text{ a.s..}$$

By this and by noticing $\sum_{j=1}^{\infty} a_j u_j < \infty$ a.s., the sum given in (26) tends to zero as $n\to\infty$. Thus, we have shown that both (24) and (25) are true, which in turn implies (23), and the theorem follows from GCT. □

Proof of Theorem 2: We first show that

$$\gamma_k \underset{k\to\infty}{\longrightarrow} \sum_{j=0}^{r} d_j Ef(u_1). \tag{28}$$

We write (18) as (29), as shown at the bottom of the page, where

$$\delta_{k+1} = z_{k+1} - \sum_{j=0}^{r} d_j Ef(u_1). \tag{30}$$

Since in (29) the regression function is linear, similar to Theorem 1, for (28) we need only to prove

$$\lim_{T\to 0}\limsup_{n\to\infty}\frac{1}{T}\sum_{j=n}^{m(n,t)} a_j\delta_{j+1} = 0 \qquad \forall t\in[0,T]. \tag{31}$$

Since $\sum_{s=1}^{\infty} a_s \sum_{j=0}^{r} d_j(v_{s-j} - Ef(u_1)) < \infty$ a.s. and $\sum_{s=1}^{\infty} a_s\bar{\xi}_{s+1} < \infty$ a.s., (31) follows from $E\xi_k \underset{k\to\infty}{\longrightarrow} 0$ by A2). Thus, by GCT we have shown (28).

Since $\sum_{j=0}^{r} d_j \neq 0$ by A3), we have

$$\gamma_k \left(\sum_{j=0}^{r} d_j\right)^{-1} \underset{k\to\infty}{\longrightarrow} Ef(u_1) \text{ a.s..} \tag{32}$$

By Theorem 1, we have $(\theta_k(0))^{-1}\sum_{i=0}^{r}\theta_k(i) \underset{k\to\infty}{\longrightarrow} \sum_{j=0}^{r} d_j$ a.s., which incorporating with (32) yields the assertion of the theorem. □

IV. Strongly Consistent Estimate for $f(u)$

We now proceed to estimate the value of the nonlinear function at any fixed point u. We have the following theorem.

Theorem 3: Assume A1) and A2) hold. Then

$$\mu_k(u) \underset{k\to\infty}{\longrightarrow} \mu(u) \triangleq f(u) + \sum_{j=1}^{r} d_j Ef(u_1) \text{ a.s..} \tag{33}$$

Further, if in addition, A3) holds, then

$$\mu_k(u) - \frac{\sum_{j=1}^{r}\theta_k(j)\gamma_k}{\sum_{j=0}^{r}\theta_k(j)} \underset{k\to\infty}{\longrightarrow} f(u) \text{ a.s.} \tag{34}$$

where $\theta_k(i)$, γ_k, and $\mu_k(u)$ are defined by (6)–(7), (8)–(9), and (10)–(11), respectively.

To prove the theorem, we start with a lemma.

Lemma 1: Assume A1) and A2) hold. Then, there is Ω_0 with $P\Omega_0 = 1$ such that for any fixed sample path $\omega\in\Omega_0$ if $\mu_{n_k}(u)$ is a convergent subsequence of $\{\mu_k(u)\}$: $\mu_{n_k}(u) \underset{k\to\infty}{\longrightarrow} \bar{\mu}(u)$, then for all large enough k and sufficiently small $T > 0$

$$\mu_{s+1}(u) = \mu_s(u) - a_s w_s(\mu_s(u) - z_{s+1}) \tag{35}$$

and

$$\|\mu_{s+1}(u) - \mu_{n_k}(u)\| \leq cT, \ s = n_k, n_k+1, \ldots, m(n_k, T) \tag{36}$$

where c is a constant independent of k but may depend on sample path ω.

Proof: Define

$$\Phi_{k,j} \triangleq (1-a_k w_k)\ldots(1-a_j w_j), \qquad k\geq j, \qquad \Phi_{j,j+1} = 1. \tag{37}$$

$$\gamma_{k+1} = \begin{cases} \gamma_k - a_k\left(\gamma_k - \sum_{j=0}^{r} d_j Ef(u_1)\right) + a_k\delta_{k+1}, & \text{if } \left|\gamma_k - a_k\left(\gamma_k - \sum_{j=0}^{r} d_j Ef(u_1) - \delta_{k+1}\right)\right| \leq M_{\nu_k} \\ 0, & \text{otherwise} \end{cases} \tag{29}$$

By (4), (5), and the fact that $\sum_{k=1}^{\infty} a_k^2/b_k < \infty$, we have

$$\sum_{j=1}^{\infty} \frac{a_j}{\sqrt{b_j}}(\sqrt{b_j}w_j - \sqrt{b_j}Ew_j) < \infty \text{ a.s.} \tag{38}$$

since $\{\sqrt{b_j}w_j\}$ is a sequence of mutually independent random variables with bounded second moments. Consequently

$$\sum_{j=n_k}^{s} a_j w_j = \sum_{j=n_k}^{s} \frac{a_j}{\sqrt{b_j}}(\sqrt{b_j}w_j - \sqrt{b_j}Ew_j) + \sum_{j=n_k}^{s} a_j Ew_j = O(T) \quad \forall s \in [n_k, \ldots, m(n_k T)] \tag{39}$$

as $k \to \infty$ and $T \to 0$. This implies that

$$\log \Phi_{s,n_k} = O\Big(\sum_{j=n_k}^{s} a_j w_j\Big)$$

and

$$\Phi_{s,n_k} = 1 + O(T) \quad \forall s \in [n_k, \ldots, m(n_k, T)] \tag{40}$$

as $k \to \infty$ and $T \to 0$.

Denoting $E|\xi_j| \triangleq c_j$, we have

$$\sum_{j=1}^{\infty} a_j(w_j|\xi_{j+1}| - Ew_j c_{j+1}) < \infty \text{ a.s.} \tag{41}$$

and

$$\begin{aligned}\left|\sum_{j=n_k}^{s} \Phi_{s,j+1} a_j w_j \xi_{j+1}\right| &\le \sum_{j=n_k}^{s} a_j w_j |\xi_{j+1}| \\ &= \sum_{j=n_k}^{s} a_j(w_j|\xi_{j+1}| - Ew_j c_{j+1}) \\ &\quad + \sum_{j=n_k}^{s} a_j Ew_j c_{j+1} = O(T) \\ &\quad \forall s \in [n_k, \ldots, m(n_k, T)]\end{aligned} \tag{42}$$

as $k \to \infty$ and $T \to 0$.

Since w_j is positive, from (39) and the boundedness of $\{v_j\}$, it follows that

$$\begin{aligned}\sum_{j=n_k}^{s} \Phi_{s,j+1} a_j w_j y_{j+1} &= \sum_{j=n_k}^{s} \Phi_{s,j+1} a_j w_j \sum_{l=0}^{r} d_l v_{j-l} \\ &= O\left(\sum_{j=n_k}^{s} a_j w_j\right) = O(T).\end{aligned} \tag{43}$$

Combining (40), (42), and (43), we arrive at

$$\Phi_{s,n_k}\mu_{n_k}(u) + \sum_{j=n_k}^{s} \Phi_{s,j+1} a_j w_j z_{j+1} = \mu_{n_k}(u) + O(T) \quad s \in [n_k, \ldots, m(n_k, T)]. \tag{44}$$

This means that $\mu_s(u)$ with s varying in $[n_k, \ldots, m(n_k, T)]$ is iterated according to

$$\mu_{s+1}(u) = \mu_s(u) - a_s w_s(\mu_s(u) - z_{s+1})$$

and the left-hand side of (44) is nothing else but $\mu_{s+1}(u)$.

Let $\mathcal{N}$ denote the exceptional set where at least one of (38) and (41) does not hold. Then, we may take $\Omega_0 = \Omega \backslash \mathcal{N}$. It is clear $P\Omega_0 = 1$. The lemma is proved. $\square$

Proof of Theorem 3: We first prove (33). For this, we rewrite (10) as (45), as shown at the bottom of the page, where

$$e_{k+1}(u) = w_k(\mu_k(u) - z_{k+1}) - \sqrt{\pi}p(u)\Big(\mu_k(u) - f(u) - \sum_{j=1}^{r} d_j Ef(u_1)\Big). \tag{46}$$

The linear function

$$\sqrt{\pi}p(u)\Big(x - f(u) - \sum_{j=1}^{r} d_j Ef(u_1)\Big) \tag{47}$$

corresponds to the regression function $g(x)$ in GCT. The root of the linear function (47) is $f(u) + \sum_{j=1}^{r} d_j Ef(u_1)$. Therefore, the conclusion (33) follows from GCT if we can show that there is Ω_1 with $P\Omega_1 = 1$ such that for any sample path $\omega \in \Omega_1$ we have

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \sum_{j=n_k}^{m(n_k,T_k)} a_j e_{j+1}(u) = 0 \quad \forall T_k \in [0, T] \tag{48}$$

for any $\{n_k\}$ such that $\mu_{n_k}(u)$ converges: $\mu_{n_k}(u) \underset{k\to\infty}{\longrightarrow} \bar{\mu}(u)$.

Let us first define the exceptional set. By the convergence theorem for martingale difference sequences [7] all series listed here are convergent, a.s.

$$\sum_{j=1}^{\infty} a_j(w_j - Ew_j) < \infty \text{ a.s.} \tag{49}$$

$$\sum_{j=1}^{\infty} a_j(|w_j - Ew_j| - E|w_j - Ew_j|) < \infty \text{ a.s.} \tag{50}$$

$$\sum_{j=1}^{\infty} a_j w_j \bar{\xi}_{j+1} < \infty \text{ a.s.} \tag{51}$$

$$\sum_{j=1}^{\infty} a_j(w_j v_j - Ew_j v_j) < \infty \text{ a.s.} \tag{52}$$

$$\mu_{k+1}(u) = \begin{cases} \mu_k(u) - a_k\sqrt{\pi}p(u)\Big(\mu_k(u) - f(u) - \sum_{j=1}^{r} d_j Ef(u_1)\Big) - a_k e_{k+1}(u)), & \text{if } \left|\mu_k(u) - a_k\sqrt{\pi}p(u)\left(\mu_k(u) - f(u) - \sum_{j=1}^{r} d_j Ef(u_1) - a_k e_{k+1}(u)\right)\right| \le M_{\lambda_k(u)} \\ 0, & \text{otherwise} \end{cases} \tag{45}$$

and

$$\sum_{j=1}^{\infty} a_j(w_j v_{j-s} - Ew_j Ev_1) < \infty \text{ a.s.}, \qquad s = 1, 2, \ldots, r. \tag{53}$$

Noticing that $\{r_j\}$ is bounded and deterministic and, hence, $\sum_{j=1}^{\infty} a_j^2 r_j^2 / b_j < \infty$. Thus, we have

$$\sum_{j=1}^{\infty} a_j r_{j+1}(w_j - Ew_j) < \infty \text{ a.s.}. \tag{54}$$

Denote by Ω_1 the set where (41) and (49)–(54) are convergent. Then, $P\Omega_1 = 1$ and $\Omega_1 \subset \Omega_0$ where Ω_0 is defined in Lemma 1.

Let $\omega \in \Omega_1$ be fixed, and let $(\mu_{n_k}(u))$ be a convergent subsequence: $\mu_{n_k}(u) \underset{k\to\infty}{\longrightarrow} \bar{\mu}(u)$.

We write $e_{k+1}(u)$ given by (46) as a sum

$$e_{k+1}(u) = e_{k+1}^{(1)}(u) + e_{k+1}^{(2)}(u) + e_{k+1}^{(3)}(u) + e_{k+1}^{(4)} \tag{55}$$

where

$$e_{k+1}^{(1)}(u) = (w_k - \sqrt{\pi} p(u))\mu_k(u) \tag{56}$$

$$e_{k+1}^{(2)}(u) = -(w_k v_k - \sqrt{\pi} p(u) f(u)) \tag{57}$$

$$e_{k+1}^{(3)}(u) = -\left(w_k \sum_{l=1}^{r} d_l v_{k-l} - \sqrt{\pi} p(u) \sum_{l=1}^{r} d_l E v_1\right) \tag{58}$$

$$e_{k+1}^{(4)} = -w_k \xi_{k+1}. \tag{59}$$

We now show

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \sum_{j=n_k}^{m(n_k,T_k)} a_j e_{j+1}^{(i)}(u) = 0, \qquad i = 1, 2, 3, 4 \qquad \forall T_k \in [0, T] \tag{60}$$

where $e_{j+1}^{(4)}(u) \equiv e_{j+1}^{(4)}$. If this is done, then the desired (48) follows immediately.

For $i = 1$, we have

$$\begin{aligned}
&\sum_{j=n_k}^{m(n_k,T_k)} a_j \mu_j(u)(w_j - \sqrt{\pi} p(u)) \\
&= \sum_{j=n_k}^{m(n_k,T_k)} a_j \mu_j(u)(w_j - Ew_j) + \sum_{j=n_k}^{m(n_k,T_k)} a_j \mu_j(u)(Ew_j - \sqrt{\pi} p(u)) \\
&= \sum_{j=n_k}^{m(n_k,T_k)} a_j(\mu_j(u) - \bar{\mu}(u))(w_j - Ew_j) + \bar{\mu}(u) \sum_{j=n_k}^{m(n_k,T_k)} a_j(w_j - Ew_j) \\
&\quad + \sum_{j=n_k}^{m(n_k,T_k)} a_j \mu_j(u)(Ew_j - \sqrt{\pi} p(u)).
\end{aligned} \tag{61}$$

On the right-hand side of (61), the second term tends to zero as $k \to \infty$ by (49), the last term tends to zero as $k \to \infty$ by (4) and (36), while the first term can be estimated as follows.

By (36) and $\mu_{n_k}(u) \underset{k\to\infty}{\longrightarrow} \bar{\mu}(u)$, we have

$$\left| \sum_{j=n_k}^{m(n_k,T_k)} a_j(\mu_j(u) - \bar{\mu}(u))(w_j - Ew_j) \right| = O(T) \sum_{j=n_k}^{m(n_k,T_k)} a_j(|w_j - Ew_j| - E|w_j - Ew_j| + E|w_j - Ew_j|)$$

and by (50) and (4), it follows that

$$\begin{aligned}
&\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \sum_{j=n_k}^{m(n_k,T_k)} a_j(\mu_j(u) - \bar{\mu}(u))(w_j - Ew_j) \\
&= \lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} O(T) \sum_{j=n_k}^{m(n_k,T_k)} a_j E|w_j - Ew_j| \\
&= \lim_{T\to 0} \limsup_{k\to\infty} O(1) \sum_{j=n_k}^{m(n_k,T_k)} a_j Ew_j = 0.
\end{aligned} \tag{62}$$

Thus, we have shown (60) for $i = 1$. We now show (60) for $i = 2$.

We have

$$\begin{aligned}
&\frac{1}{T} \sum_{j=n_k}^{m(n_k,T_k)} a_j(w_j v_j - \sqrt{\pi} p(u) f(u)) \\
&= \frac{1}{T} \sum_{j=n_k}^{m(n_k,T_k)} a_j(w_j v_j - Ew_j v_j) \\
&\quad + \frac{1}{T} \sum_{j=n_k}^{m(n_k,T_k)} a_j \Big(\int_{-c_1}^{c_1} \frac{1}{b_j} e^{-((x-u)/b_j)^2} p(x) f(x) dx \\
&\quad - \sqrt{\pi} p(u) f(u) \Big).
\end{aligned} \tag{63}$$

Similar to (4), it is shown that

$$\int_{-c_1}^{c_1} \frac{1}{b_j} e^{-((x-u)/b_j)^2} p(x) f(x) dx \underset{j\to\infty}{\longrightarrow} \sqrt{\pi} p(u) f(u).$$

Consequently, the last term in (63) tends to zero as $k \to \infty$, while the first term on the right-hand side of (63) also tends to zero by (52). This proves (60) for $i = 2$.

We now show (60) for $i = 3$. By (4), it suffices to show

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \sum_{j=n_k}^{m(n_k,T_k)} a_j \left(w_j \sum_{l=1}^{r} d_l v_{j-l} - Ew_j \sum_{l=1}^{r} d_l E v_1 \right) = 0 \tag{64}$$

but which is a consequence of (53).

Finally, for (60) for $i = 4$ by (51), we need only to show

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \sum_{j=n_k}^{m(n_k,T_k)} a_j w_j r_{j+1} = 0. \tag{65}$$

By (54), we have

$$\begin{aligned}
&\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \sum_{j=n_k}^{m(n_k,T_k)} a_j w_j r_{j+1} \\
&= \lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \sum_{j=n_k}^{m(n_k,T_k)} a_j(w_j - Ew_j + Ew_j) r_{j+1} \\
&= \lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \sum_{j=n_k}^{m(n_k,T_k)} a_j Ew_j r_{j+1}
\end{aligned}$$

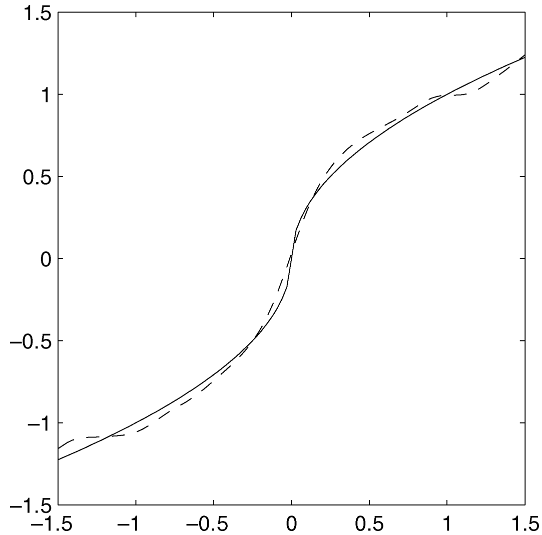

Fig. 2. Example 1.

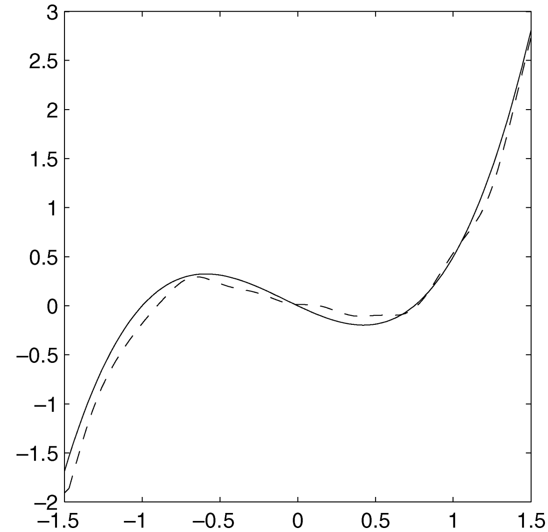

Fig. 3. Example 2.

which tends to zero by (4) and $r_j \xrightarrow[j\to\infty]{} 0$ by A2.

Thus, we have shown (48), and at same time completed the proof of (33).

Putting consistent estimates obtained in Theorems 1 and 2 into (33) leads to (34).

V. Examples

We now give two numerical examples demonstrating how the nonparameterized unknown nonlinear function $f(u)$ is approximated by the estimate given by the left-hand side of (34). In both examples, we take

$$y_k = v_k - 0.5 v_{k-1} \quad v_k = f(u_k) \quad z_k = y_k + \xi_k$$
$$a_k = k^{-3/4} \quad b_k = k^{-1/4}$$

$\{u_k\}$ to be a sequence of iid random variables uniformly distributed over $[-2, 2]$, and assume $\xi_k \in \mathcal{N}(0, 0.1)$.

Example 1: $f(u) = sign(u)\sqrt{|u|}$

Example 2: $f(u) = u^3 + (1/4)u^2 - 3/4$

In both Figs. 2 and 3, the solid line is the true function $f(u)$ and the dashed line represents the estimate of $f(u)$. The estimate is derived in the following way: the interval $[-1.5, 1.5]$ where the function is defined on is equally divided into 100 subintervals, and at each endpoint u of subintervals $f(u)$ is estimated by the left-hand side of (34). The dashed line corresponds to the estimate for $f(u)$ at $k = 1000$.

For the linear part, there is only one parameter $d_1 = -0.5$ to be estimated. At $k = 1000$, the estimates for d_1 are -0.4962 and -0.4578 for examples 1 and 2, respectively.

The numerical simulation justifies the strong consistency theoretically proved in Sections III and IV.

VI. Concluding Remarks

In this paper, new algorithms are proposed for identifying Hammerstein systems. We note the following properties of the obtained results, which are justified by numerical simulation.

i) The estimates are recursive and, thus, they are updated after receiving new data u_k and z_{k+1} for each k.

ii) The estimates are convergent to the true values with probability one.

iii) No structure assumption is made on the nonlinearity $f(\cdot)$.

iv) Not only $f(u)$ and coefficients of the linear system but also $Ef(u_k)$ and the correlation $Eu_k f(u_k)$ between the input and output of the nonlinear part are strongly consistently estimated.

For further research it is natural to consider more general linear subsystems, e.g., general ARMAX systems. The multidimensional systems and correlated observation noises are also of interest to consider.

Acknowledgment

The author would like to thank Mr. X. Hu for his help in performing the numerical simulation.

References

[1] E. W. Bai, "Frequency domain identification of Hammerstein models," *IEEE Trans. Automat. Contr.*, vol. 48, pp. 530–542, Apr. 2003.

[2] M. Boutayeb, H. Rafaralahy, and M. Darouach, "A robust and recursive identification method for the Hammerstein model," in *Proc. 13th IFAC World Congr.*, vol. I, San Francisco, CA, 1996, pp. 447–452.

[3] H. Al-Duwaish and M. N. Karim, "A new method for the identification of Hammerstein model," *Automatica*, vol. 33, no. 10, pp. 1871–1875, 1997.

[4] F. H. I. Chang and R. Luus, "A noniterative method for identification using Hammerstein model," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 465–468, Oct. 1971.

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

[6] H. F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*. Boston, MA: Birkhäuser, 1991.

[7] Y. S. Chow and H. Teicher, *Probability Theory*. New York: Springer-Verlag, 1978.

[8] P. Crama and J. Schoukens, "Initial estimates of Wiener and Hammerstein systems using multisine excitation," *IEEE Trans. Instrum. Measure.*, vol. 50, pp. 1791–1795, Dec. 2001.

[9] R. C. Emerson, M. J. Korenberg, and M. C. Citron, "Identification of complex-cell intensive nonlinearities in a cascade model of cat visual cortex," *Bio. Cybern.*, vol. 66, pp. 291–300, 1992.

[10] E. Eskinat and S. H. Johnson, "Use of Hammerstein models in identification of nonlinear systems," *Amer. Inst. Chem. Eng.*, vol. 37, no. 2, pp. 255–268, 1991.

[11] H. T. Fang and H. F. Chen, "Stability and instability of limit points for stochastic approximation algorithms," *IEEE Trans. Automat. Contr.*, vol. 45, pp. 413–420, Mar. 2000.
[12] W. Greblicki, "Stochastic approximation in nonparametric identification of Hammerstein systems," *IEEE Trans. Automat. Contr.*, vol. 47, pp. 1800–1810, Nov. 2002.
[13] W. Greblicki and M. Pawlak, "Nonparametric identification of Hammertein systems," *IEEE Trans. Inform. Theory*, vol. 35, pp. 409–418, Mar. 1989.
[14] Z. Q. Lang, "A nonparametric polynomial identification algorithm for the Hammerstein system," *IEEE Trans. Automat. Contr.*, vol. 42, pp. 1435–1441, Oct. 1997.
[15] H. E. Liao and W. A. Sethares, "Suboptimal identification of nonlinear ARMA models using an orthogonality approach," *IEEE Trans. Circuits Syst. I*, vol. 42, pp. 14–22, Jan. 1995.
[16] L. Ljung, *System Identification*. Upper Saddle River, NJ: Prentice-Hall, 1987.
[17] M. Pawlak, "On the series expansion appraoch to the identification of Hammertein system," *IEEE Trans. Automat. Contr.*, vol. 36, pp. 763–767, June 1991.
[18] P. Stoica and T. Söderström, "Instrumental-variable methods for identification of Hammerstein systems," *Int. J. Control*, vol. 35, no. 3, pp. 459–476, 1982.
[19] M. Verhaegen and D. Westwick, "Identifying MIMO Hammerstein systems in the context of subspace model identification methods," *Int. J. Control*, vol. 63, no. 2, pp. 331–349, 1996.
[20] J. Vörös, "Parameter identification of discontinuous Hammerstein systems," *Automatica*, vol. 33, no. 6, pp. 1141–1146, 1997.
[21] W. Wang and R. Henriksen, "Generalized predictive control of nonlinear systems of the Hammerstein form," *Model. Ident. Control*, vol. 15, no. 4, pp. 253–262, 1994.

Han-Fu Chen (SM'94–F'97) graduated from Leningrad (St. Petersburg) University, Leningrad, Russia, in 1961.

He joined the Institute of Mathematics, Chinese Academy of Sciences, Beijing, China, in 1961. Since 1979, he has been with the Institute of Systems Science, which is now a part of the Academy of Mathematics and Systems Science, Chinese Academy of Sciences. He is a Professor of the Laboratory of Systems and Control of the Institute. His research interests are mainly in stochastic systems, including system identification, adaptive control, and stochastic approximation and its applications to systems, control, and signal processing. He was elected a Member of the Chinese Academy of Sciences in 1993. He now serves as a Council Member of the International Federation of Automatic Control (IFAC), and as Editor of both *Systems Science and Mathematical Sciences* and *Control Theory and Applications*.