Nonlinearity Estimation in Hammerstein Systems Based on Ordered Observations

Włodzimierz Greblicki, *Member, IEEE*

Abstract—**The nonlinear subsystem of a Hammerstein system is identified, i.e., its characteristic is recovered from input-output observations of the whole system. The input and disturbance are white stochastic processes. The identified characteristic satisfies a piecewise Lipschitz condition only. Algorithms presented in the paper are calculated from ordered input-output observations, i.e., from pairs of observations arranged in a sequence in which input measurements increase in value. The mean integrated square error converges to zero as the number of observations tends to infinity. Convergence rates are insensitive to the shape of the probability density of the input signal. Results of numerical simulation are also shown.**

I. Introduction

The block oriented approach to the nonlinear system identification has been receiving growing attention in the theoretical literature, see Bendat [3], Billings [6], Billings and Fakhouri [7], Haber and Unbenhauen [28], Priestley [39], and in applications in various fields, e.g., communication, Giunta *et al.* [19], control, Lang Zi-Qiang [35], chemistry, Eskinat *et al.* [14], biology, Emerson *et al.* [13], Hunter and Korenberg [31], Korenberg and Hunter [32]. The main idea of the approach is that the identified system consists of simple subsystems such as linear, dynamic and nonlinear, memoryless. The main objective of the block-oriented identification is to recover descriptions of all subsystems from observations taken at input and output of the whole system. So far, the greatest attention has been paid to Hammerstein systems, i.e., cascade systems consisting of a nonlinear memoryless element followed by a linear dynamic one, see Narendra and Gallman [37], Chang and Luus [9], Haist *et al.* [29], Thathachar and Ramaswamy [42], Gallman [18], Brilinger [8], Fan-Chu Kung and Dong-Her Shih [16] and Lang Zi-Qiang [35]. Authors mentioned above have assumed that the nonlinearity is known up to a finite number of coefficients. The most common restriction imposed on the nonlinearity confines considerations to polynomials which coefficients are estimated. Resulting identification problems are parametric.

Greblicki and Pawlak [21]-[27], Greblicki [20], Pawlak [38], and next Krzyżak [33], [34] significantly enlarged the class of considered characteristics. Their restrictions are mild, which means that their a priori information concerning the nonlinearity is extremely poor since they assume that the characteristic is, e.g., bounded or square integrable only. Owing to that, the family of all possible characteristics admitted by them is so ample that can not be represented in a parametric form. Therefore, their nonparametric identification problems are closer to real problems encountered in applications.

In this paper, the problem of recovering the nonlinearity in a Hammerstein system is also nonparametric. The characteristic is assumed only to satisfy a piecewise Lipschitz condition. Therefore, the characteristic may be, e.g., discontinuous. We apply two algorithms to estimate the nonlinear characteristic of the memoryless subsystem. For both algorithms, the mean integrated square error converges to zero as the number of observations tends to infinity. The convergence rate depends on the nonlinearity. The smoother characteristic, the greater speed of convergence. The rate is, however, independent of the shape of the probability density of the input signal. This property is an important advantage of our algorithms over those known in the literature, since their rate gets worse for irregular densities.

We derive identification algorithms not from the original but ordered sequence of observations. Ordering means that input-output pairs of observations are rearranged with respect to input observations. In the new sequence, input observations increase in value. It seems that the computational effort caused by ordering is well compensated by the fact that the convergence rate of thus obtained algorithms is insensitive to irregularities of the probability density of the input signal.

The author is with the Institute of Engineering Cybernetics, Technical University of Wrocław, 50-370 Wrocław, Poland

II. Identification Problem

We deal with a cascade nonlinear dynamic system, i.e., a system consisting of a nonlinear memoryless subsystem followed by a linear dynamic one. The system, referred to as the *Hammerstein system*, Fig. 1, is driven by stationary white random noise $\{U_n; n = \cdots, -1, 0, 1, 2 \cdots\}$. We assume that

$$-1 \le U_n \le 1.$$

The probability density f of U_n's is unknown and satisfies the following restriction:

$$\delta < 0 \le f(u) \tag{2.1}$$

with al $u \in [-1, 1]$, some unknown δ. The nonlinear memoryless subsystem has a characteristic denoted by m, which means that

$$W_n = m(U_n).$$

The characteristic m is a Borel measurable function and satisfies a piecewise Lipschitz condition. The latter means that there exist $D_1, \cdots, D_q$ such that $D_1 \cup D_2 \cup \cdots \cup D_q = [-1, 1]$ and

$$|m(u) - m(v)| \le \alpha_i |u - v| \tag{2.2}$$

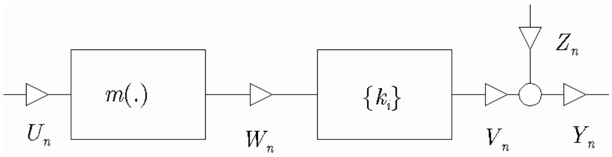

Fig. 1. Identified Hammerstein system.

with some $\alpha_i > 0$, and all u, v in D_i, $i = 1, 2, \cdots, q$. By α we denote $\max(\alpha_1, \alpha_2, \cdots, \alpha_q)$. Denote, moreover, for convenience,

$$\sup_{u\in[-1,1]} |m(u)| = M. \tag{2.3}$$

Examining convergence rates of our estimates, we shall assume additionally that m is differentiable.

The dynamic subsystem is described by a state space equation

$$\left.\begin{array}{c} X_{n+1} = AX_n + bW_n \\ V_n = c^T X_n + dW_n \end{array}\right\} \tag{2.4}$$

where X_n is the state vector at time n, and where the matrix A, vectors b, c, and the number d are all unknown. So is the dimension of the state vector. The matrix A is asymptotically stable, which means that all its eigenvalues lie in the unit circle. The output of the system is disturbed by stationary white random noise $\{Z_n; n = \cdots, -1, 0, 1, 2\cdots\}$. Therefore,

$$Y_n = V_n + Z_n.$$

The noise is independent of the input signal, has zero mean and unknown variance σ_Z^2. Owing to all that, $\{Y_n; n = \cdots, -1, 0, 1, 2\cdots\}$ is a sequence of dependent identically distributed random variables. The sequence is a stationary ARMA stochastic process.

The goal of the paper is to recover m from observations $(U_1, Y_1), (U_2, Y_2), \cdots, (U_n, Y_n)$ taken at the input and output of the whole system. Observe that

$$Y_n = \mu(U_n) + Z_n + \xi_n,$$

where

$$\mu(u) = dm(u) + c^T EX_0 \tag{2.5}$$

and where

$$\begin{aligned} \xi_n &= c^T(X_n - EX_n) \\ &= \sum_{i=1}^{\infty} c^T A^{i-1} b \left[m(U_{n-i}) - Em(U_{n-i})\right]. \end{aligned} \tag{2.6}$$

Clearly $\mu(u) = E\{Y_n | U_n = u\}$, which means that the regression μ is observed in the presence of noise $Z_n + \xi_n$. Its first component is white, while the other incurred by the dynamic subsystem is correlated and depends on both m and the input signal. Observing the input and output of the whole system, we estimate μ , i.e., the regression of Y_n on U_n. The fact that we are able to recover m only up to some unknown constants d, and is a simple consequence of the composite structure of the system. Observe that, for $EU_n = 0$ and m odd, we have $EX_0 = 0$. In such a case, our algorithms recover $dm(u)$, i.e., the nonlinear characteristic up to a multiplicative constant.

For $A = 0$, the linear subsystem is just a memoryless linear amplifier and $Y_n = dm(U_n) + Z_n$. The whole system is then a nonlinear amplifier and pairs (U_i, Y_i)'s are independent. The problem of estimating the regression μ in such a situation has been already studied in the statistical literature, see, e.g., Härdle [30]. In this paper, however, A can be nonzero and, consequently, Y_i's can be correlated. Therefore, the problem examined in this paper is tougher, since we estimate a regression from a correlated stochastic process, i.e., from dependent data.

We want to mention here that recovering a regression from dependent data has been already examined mainly under mixing-type restrictions concerning the kind of dependence. Estimation of $E\{\zeta_n | \zeta_{n-1}, \cdots, \zeta_{n-p}\}$, i.e., predicting ζ_n from $\zeta_{n-1}, \cdots, \zeta_{n-p}$, where $\{\zeta_n\}$ is a uniform strong mixing stochastic process has been studied by Boente and Fraiman [5], Bierens [4] as well as Collomb and Härdle [11]. Boente and Fraiman [5] have examined the problem for strong mixing processes. Bierens [4] has weakened the strong mixing condition to the υ-stability in L_2. In turn, Yakowitz [45] have estimated the regression from a Markov process satisfying a G_2 condition satisfied by some ARMA processes. In this paper, we infer from ARMA stochastic processes generated by state space equation (2.4).

We want to stress that the class of such ARMA processes we deal with is different from the class of strong mixing processes considered by authors mentioned above. For example, the ARMA process $\{X_n\}$ in which $X_{n+1} = aX_n + W_n$, where $0 < a \le 1/2$, and where W_n's are Bernoulli random variables, is not strong mixing (see Andrews [1]). Such a process is generated by our system, if the input of the whole system has a normal density, and a nonlinear characteristic takes two different values. Strong-mixing methods are then not applicable. In the light of that, the strong mixing approach developed in the statistical literature does not apply in general to the identification of systems in which some signals are ARMA processes. In particular, it can not be used in the identification of the nonlinear part of the Hammerstein system. Our methodology aimed at ARMA processes is different from that based on the theory of mixing processes.

The problem of the identification of the linear subsystem is easier. For details, we refer the reader to papers cited in references.

III. Identification Algorithms

To identify the nonlinear part of the system, we rearrange the sequence $U_1, U_2, \cdots, U_n$ of input observations into a new one $U_{(1)}, U_{(2)}, \cdots, U_{(n)}$, in which $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$. Ties, i.e., events that $U_{(i)} = U_{(j)}$, for $i \ne j$, have zero probability since U_n's possess a density. Moreover, we define $U_{(0)} = -1$ and $U_{(n+1)} = 1$. The sequence $U_{(1)}, U_{(2)}, \cdots, U_{(n)}$ is called the order statistics of $U_1, U_2, \cdots, U_n$. We then rearrange the sequence $(U_1, Y_1), (U_2, Y_2), \cdots, (U_n, Y_n)$

of input-output observations into the following one: $(U_{(1)}, Y_{[1]}), (U_{(2)}, Y_{[2]}), \cdots, (U_{(n)}, Y_{[n]})$. Observe that $Y_{[i]}$'s are not ordered, but just paired with $U_{(i)}$'s. Similarly, also $X_{[i]}$'s are paired with $U_{(i)}$'s.

We examine the following two estimators of $\mu(u)$ based on the ordered sequence of observations:

$$\hat{\mu}(u) = \sum_{i=1}^{n} Y_{[i]} \frac{1}{h(n)} \int_{U_{(i-1)}}^{U_{(i)}} K\left(\frac{u-v}{h(n)}\right) dv \qquad (3.1)$$

and

$$\tilde{\mu}(u) = \sum_{i=1}^{n} Y_{[i]} \frac{1}{h(n)} K\left(\frac{u-v}{h(n)}\right) \left(U_{(i)} - U_{(i-1)}\right)$$
$$= \sum_{i=1}^{n} Y_{[i]} \frac{1}{h(n)} \int_{U_{(i-1)}}^{U_{(i)}} K\left(\frac{u-U_{(i)}}{h(n)}\right) dv \qquad (3.2)$$

where K is a Borel measurable kernel function on a real line R and $\{h(n)\}$ a positive number sequence. We show that, for a suitably selected kernel and number sequence, both algorithms converge to $\mu(u)$ as the number of observations increases to infinity.

For a deterministic input signal, algorithm (3.1) has been proposed by Gasser and Müller [17] and Cheng and Lin [10] while (3.2) by Priestley and Chao [40]. For the random input, estimate (3.1) has been examined by Mack and Müller [36]. In all those papers, however, $A = 0$ and, consequently, observed input-output pairs are independent.

The nonlinear characteristic in the Hammerstein system has been already recovered with the nonparametric kernel regression estimate by Greblicki and Pawlak [21], [22], and Krzyżak [34]. The orthogonal series estimate has been employed by Greblicki and Pawlak [23], [24], [27], Greblicki [20], Pawlak [38], and Krzyżak [33]. Convergence rates of their algorithms are sensitive to irregularities of the input probability density, while ours are not. Algorithms with convergence rates independent of the shape of the input signal density have been proposed by Greblicki and Pawlak [26]. Their Fourier series estimates are derived, like ours, from ordered observations.

The kernel K satisfies the following restrictions:

$$\sup_{u \in R} |K(u)| = \kappa < \infty, \qquad (3.3)$$

$$\int_{-\infty}^{\infty} K(u) du = 1.$$

For simplicity, we denote $\int_{-\infty}^{\infty} |K(u)| \, du = k$. Some results hold under an additional assumption that the kernel satisfies a piecewise Lipschitz condition, i.e., that $R = A_1 \cup A_2 \cup \cdots \cup A_r$, with some r, some $A_1, \cdots, A_r$, and

$$|K(u) - K(v)| \le \beta_i |u - v| \qquad (3.5)$$

some β_i's, all u, v in A_i, $i = 1, 2, \cdots r$. We denote $\beta = \max(\beta_1, \beta_2, \cdots, \beta_r)$. On the number sequence we impose the following restrictions:

$$h(n) \to 0 \text{ as } n \to \infty, \qquad (3.6)$$

$$nh(n) \to \infty \text{ as } n \to \infty.$$

Observe that (3.5) is not very restrictive from the practical viewpoint and that one can apply a rectangle $(1/2)I_{\{|u|\le 1\}}(u)$, triangle $(1-|u|)I_{\{|u|\le 1\}}(u)$, or parabolic $(3/4)(1-u^2)I_{\{|u|\le 1\}}(u)$ kernels, where I is the indicator function. Other examples will be given in Section IV. Kernels with unbounded support, e.g. $1/2(1+|u|)^2$, $1/2\pi(1+u^2)$, $(1/2)\exp(-|u|)$, or $(1/2\pi)\exp(-u^2)$, can be also employed. As far as the number sequence is concerned, one can choose $h(n) = cn^{-\alpha}$, any positive c, where $0 < \alpha < 1$.

IV. Ordered Observations in Dynamic Systems

In this section, we give some results concerning relations between dynamic systems and order statistics. We will now verify a lemma which plays a key role in the paper. Owing to the lemma, we can overcome analytical difficulties created by the fact that ordering takes place in a dynamic system. The lemma is derived from important Lemma B1 given in Appendix B.

Lemma 4.1: Let f satisfy (2.1). Let m satisfy (2.3). Let subsystem (2.4) be asymptotically stable. Let K satisfy (3.3). Let $\{h(n)\}$ satisfy (3.6), (3.7). Then

$$E\int_{-1}^{1} \left[\sum_{i=1}^{n+1} \xi_{[i]} \frac{1}{h(n)} \int_{U_{(i-1)}}^{U_{(i)}} K\left(\frac{u-v}{h(n)}\right) dv\right]^2 du$$
$$= O\left(\frac{1}{nh(n)}\right). \qquad (4.1)$$

If K satisfies also (3.4), then

$$E\int_{-1}^{1} \left[\sum_{i=1}^{n+1} \xi_{[i]} \frac{1}{h(n)} \int_{U_{(i-1)}}^{U_{(i)}} K\left(\frac{u-U_{(i)}}{h(n)}\right) dv\right]^2 du$$
$$= O\left(\frac{1}{nh(n)}\right).$$

Proof: To verify the first part of the lemma, observe that the quantity under the sign of integration in (4.1) equals $W_1(u) + W_2(u)$, where

$$W_1(u) = \left[\sum_{i=1}^{n+1} \xi_{[i]} \frac{1}{h(n)} \int_{U_{(i-1)}}^{U_{(i)}} K\left(\frac{u-v}{h(n)}\right) dv\right]^2$$

and

$$W_2(u) = \sum_{i=1}^{n+1} \sum_{\substack{j=1 \\ j\ne i}}^{n+1} \xi_{[i]}\xi_{[j]} \left[\frac{1}{h(n)} \int_{U_{(i-1)}}^{U_{(i)}} K\left(\frac{u-v}{h(n)}\right) dv\right]$$
$$\times \left[\frac{1}{h(n)} \int_{U_{(j-1)}}^{U_{(j)}} K\left(\frac{u-v}{h(n)}\right) dv\right].$$

Observe that $|\xi_j| \le 2M\beta$, $j = 1, 2, \cdots ..$, where $\beta = \sum_{i=0}^{\infty} \left|c^T A^i b\right|$. Since the linear subsystem is asymptotically stable, the sum exists. Therefore,

$$EW_1(u) \le 4M^2\beta^2 \sum_{i=1}^{n+1} E\left[\frac{1}{h(n)} \int_{U_{(i-1)}}^{U_{(i)}} K\left(\frac{u-v}{h(n)}\right) dv\right]^2.$$

Now, an application of Lemma A3 in Appendix A leads to $E\int_{-1}^{1} W_1(u)du = O(1/nh(n))$. In turn, applying Lemma B1, in Appendix B, we get

$$E\left|W_2(u)\right| \le 4M^2\rho\beta^2 n^{-1}\sum_{i=1}^{n+1}\sum_{\substack{j=1\\ j\neq i}}^{n+1} E \times\left\{\left[\frac{1}{h(n)}\int_{U_{(i-1)}}^{U_{(i)}}\left|K\left(\frac{u-v}{h(n)}\right)\right|dv\right] \times\left[\frac{1}{h(n)}\int_{U_{(j-1)}}^{U_{(j)}}\left|K\left(\frac{u-v}{h(n)}\right)\right|dv\right]\right\}$$

with some ρ depending only on A, b, and c. The sum is not greater than k^2. Therefore, $E\int_{-1}^{1}\left|W_2(u)\right|du = O(1/n)$ and the first part of the lemma follows. To verify the second one, it suffices to use the same arguments and apply Lemma A4 rather than Lemma A3. ■

V. Convergence of Algorithms

In this section we show that the mean integrated error for both the estimates converges to zero as the number of observations increases to infinity. We denote $\mu_h(u) = (1/h)\int_{-1}^{1}\mu(v)K((u-v)/h)dv$.

Lemma 5.1: Let f satisfy (2.1). Let subsystem (2.4) be asymptotically stable. Let m satisfy (2.2). Let the kernel K satisfy (3.3) and (3.4). Then,

$$E\int_{-1}^{1}\left[\hat{\mu}(u)-\mu_{h(n)}(u)\right]^2 du = O(1/nh(n)).$$

Proof: Since

$$\hat{\mu}(u) = \sum_{i=1}^{n}\left[Z_{[i]}+\xi_{[i]}+\mu(U_{(i)})\right] \times\frac{1}{h(n)}\int_{U_{(i-1)}}^{U_{(i)}}K\left(\frac{u-v}{h(n)}\right)dv$$

and

$$\mu_{h(n)}(u) = \sum_{i=1}^{n}\frac{1}{h(n)}\int_{U_{(i-1)}}^{U_{(i)}}\mu(v)K\left(\frac{u-v}{h(n)}\right)dv + \frac{1}{h(n)}\int_{U_{(n)}}^{U_{(n+1)}}\mu(v)K\left(\frac{u-v}{h(n)}\right)dv$$

we get $E[\hat{\mu}(u)-\mu_{h(n)}(u)]^2 \le V_1(u) + 3V_2(u)+3V_3(u)+3V_4(u)$, where

$$V_1(u) = \sigma_Z^2\sum_{i=1}^{n}E\left[\frac{1}{h(n)}\int_{U_{(i-1)}}^{U_{(i)}}K\left(\frac{u-v}{h(n)}\right)dv\right]^2,$$

$$V_2(u) = E\left[\sum_{i=1}^{n}\xi_{[i]}\frac{1}{h(n)}\int_{U_{(i-1)}}^{U_{(i)}}K\left(\frac{u-v}{h(n)}\right)dv\right]^2,$$

$$V_3(u) = E\left[\sum_{i=1}^{n}\frac{1}{h(n)}\int_{U_{(i-1)}}^{U_{(i)}}\left[\mu(U_{(i)})-\mu(v)\right] \times K\left(\frac{u-v}{h(n)}\right)dv\right]^2,$$

and

$$V_4(u) = E\left[\frac{1}{h(n)}\int_{U_{(n)}}^{U_{(n+1)}}K\left(\frac{u-v}{h(n)}\right)dv\right]^2.$$

From Lemma A3 it follows that $\int_{-1}^{1}V_1(u)du = O(1/nh(n))$, while, by virtue of Lemma 4.1, $\int_{-1}^{1}V_2(u)du = O(1/nh(n))$. Lemma A5 leads to $\int_{-1}^{1}V_3(u)du = O(1/n^2h(n))$. Since, moreover, $\int_{-1}^{1}V_4(u)du$ is of the same order, the proof has been completed. ■

From Lemma 5.1 and Lemma C2 in Appendix C, we obtain the following theorem.

Theorem 5.1: Let f satisfy (2.1). Let m satisfy (2.2). Let subsystem (2.4) be asymptotically stable. Let the kernel K satisfy (3.3) and (3.4). Let the number sequence satisfy (3.6), and (3.7). Then

$$E\int_{-1}^{1}\left[\hat{\mu}(u)-\mu(u)\right]^2 du \to 0 \text{ as } n\to\infty.$$

In order to show convergence of the second algorithm, we shall verify

Lemma 5.2: Let f satisfy (2.1). Let m satisfy (2.2). Let subsystem (2.4) be asymptotically stable. Let the kernel K satisfy (3.3), (3.4), and (3.5). Then,

$$E\int_{-1}^{1}\left[\tilde{\mu}(u)-\mu_{h(n)}(u)\right]^2 du = O(1/nh(n)).$$

Proof: We have $E\left[\tilde{\mu}(u)-\mu_{h(n)}\right]^2 \le W_1(u) + 3W_2(u)+3W_3(u)+3W_4(u)$, where

$$W_1(u) = \sigma_Z^2\sum_{i=1}^{n}E\left[\frac{1}{h(n)}\int_{U_{(i-1)}}^{U_{(i)}}K\left(\frac{u-U_{(i)}}{h(n)}\right)dv\right]^2,$$

$$W_2(u) = E\left[\sum_{i=1}^{n}\xi_{[i]}\frac{1}{h(n)}\int_{U_{(i-1)}}^{U_{(i)}}K\left(\frac{u-U_{(i)}}{h(n)}\right)dv\right]^2,$$

$$W_3(u) = E\left\{\sum_{i=1}^{n}\frac{1}{h(n)}\int_{U_{(i-1)}}^{U_{(i)}}\left[\mu(U_{(i)})K\left(\frac{u-U_{(i)}}{h(n)}\right) -\mu(v)K\left(\frac{u-v}{h(n)}\right)\right]dv\right\}^2$$

and

$$W_4(u) = E\left[\frac{1}{h(n)}\int_{U_{(n)}}^{U_{(n+1)}}K\left(\frac{u-U_{(i)}}{h(n)}\right)dv\right]^2.$$

From Lemma A.4, it follows that $\int_{-1}^{1}W_1(u)du = O(nh(n))$. By virtue of Lemma 4.1, $\int_{-1}^{1}W_2(u)du$ is of the same order. An application of Lemma A.5 yields $\int_{-1}^{1}W_3(u)du =$

$O(1/n^2h(n))$. Since $\int_{-1}^{1} W_2(u)du$ is of the same order, the proof has been completed. ■

Combining Lemma 5.2 with Lemma C2 in Appendix C, we obtain

Theorem 5.2: Let f satisfy (1). Let m satisfy (2.2). Let subsystem (2.4) be asymptotically stable. Let the kernel K satisfy (3.3), (3.4), and (3.5). Let the number sequence satisfy (3.6) and (3.7). Then

$$E\int_{-1}^{1}\left[\tilde{\mu}(u)-\mu(u)\right]^2du \to 0 \text{ as } n\to\infty.$$

VI. Convergence Rate

In the previous section, we have shown that our algorithms converge for a wide class of nonlinear characteristics. To examine their convergence rate, we assume that the unknown nonlinearity is a smooth function, i.e., differentiable. The rate will be expressed in terms of the number of existing derivatives of the characteristic. We need

Lemma 6.1: Let m have p derivatives and let its pth derivative be bounded in the interval $[-1,1]$. Let K have bounded support and let

$$\left.\begin{aligned}&\int_{-1}^{1}u^qK(u)du=0 \qquad , \text{ for } q=1,2,\cdots,p-1\\ &\left|\int_{-1}^{1}u^pK(u)du\right|<\infty\end{aligned}\right\}. \tag{6.1}$$

Then, for any ϵ, such that $0<\epsilon<1$, we have

$$\begin{aligned}&\int_{|u|<1-\epsilon}\left|\mu_h(u)-\mu(u)\right|^2du\\ &=(d/p!)^2h^{2p}\left[\int_{-1}^{1}u^pK(u)du\right]^2\sup_{|u|\le 1}\left[m^{(p)}(u)\right]^2\\ &+o(h^{2p}).\end{aligned}$$

Proof: With no loss of generality we can assume that $[-1,1]$ is support of K. Observing,

$$\mu_h(u)=\int_{(-1-u)/h}^{(1-u)/h}\mu(u+hv)K(-v)dv$$

and expanding $\mu(u+hv)$ in a Taylor series, we get

$$\begin{aligned}&\mu_h(u)\\ &=\mu(u)+\sum_{k=1}^{p-1}\mu^{(k)}(u)\int_{(-1-u)/h}^{(1-u)/h}((hv)^k/k!)K(-v)dv\\ &+\int_{(-1-u)/h}^{(1-u)/h}((hv)^p/p!)\mu^{(p)}(u+\theta_uhv)K(-v)dv\end{aligned}$$

where $|\theta_u|<1$. Let $0<\epsilon<1$. For h sufficiently small and every $u\in[-1+\epsilon,-\epsilon]$, we have $[(-1-u)/h,(1-u)/h]\subset[-1,1]$, and an application of (6.1) completes the proof. ■

Suppose now that pth derivative of m exists and is square integrable. Suppose, moreover, that the kernel has compact support and satisfies (3.3), (3.4). Theorem 1 combined with Lemma 6.1 imply that, for $h(n)\sim n^{-1/(2p+1)}$, we get

$$\int_{|u|\le 1-\epsilon}(\hat{\mu}(u)-\mu(u))^2du=O\left(n^{-2p/(2p+1)}\right)$$

with any ϵ, such that $0<\epsilon<1$. If, moreover, the kernel satisfies (3.5), we have

$$\int_{|u|\le 1-\epsilon}(\tilde{\mu}(u)-\mu(u))^2du=O\left(n^{-2p/(2p+1)}\right).$$

There exist kernels satisfying (3.3)-(3.5), (6.1). For example, one can select $[(-3/4)u^2+3/4]I_{(|u|\le 1)}(u)$ or $[(-63/32)u^4-(45/16)u^2+27/32]I_{(|u|\le 1)}(u)$, for p equal to 1 or 3, respectively.

Observe that the convergence rate is the same for both algorithms. Contrary to other algorithms examined in the literature, the rate is independent of the shape of the probability density f of the input signal, provided that the density is bounded from zero. Irregularities of f do not worsen the rate. It depends however on the number p of existing derivatives of m. The larger p, i.e., the smoother characteristic, the greater speed of convergence. For large p, the rate gets very close to n^{-1}, i.e., the rate typical in parametric inference. For twice differentiable m, the rate is $O(n^{-4/5})$, which is quite encouraging.

It is interesting to compare our rate with other relevant results. First of all, it is known that, for $A=0$ and normal Z_n's, the optimal, i.e., best possible rate of any nonparametric regression estimate is $n^{-2p/(2p+1)}$ (see Härdle [30, Theorem 4.1.2]). Conditions, we identify here are more complicated since A is usually nonzero and noise has an arbitrary distribution. Despite these adverse circumstances, our estimators have the optimal convergence rate.

As far as other nonparametric estimates of the nonlinearity in Hammerstein systems are concerned, Greblicki and Pawlak [24] have shown that the Hermite series estimate has the rate $O(n^{-6p/(5+6p)})$, provided that m has p derivatives and f is Gaussian. The rate is worse than ours. The Legendre series estimate converges as fast as $O(n^{-2/3})$, for both m and f differentiable (see Greblicki and Pawlak [23]). The rate is the same as ours, for $p=1$, but requires f to be differentiable. To pass to the kernel estimate, notice that our integrated square error converges to zero in probability at the speed $O(n^{-2p/(2p+1)})$. Recall that the rate holds for m having p derivatives and that f may not be differentiable. The pointwise square error for the mentioned kernel estimate has the same in-probability convergence rate, provided that both m and f have p derivatives, Greblicki and Pawlak [25]. Paying no attention to the fact that our error is global while their pointwise, observe that they have obtained the rate assuming that not only m but also f is differentiable.

It is interesting to compare our rate with that derived by Greblicki and Pawlak [26] for the Fourier series estimate based, like our ones, on ordered observations. For m satisfying a Lipschitz condition, their estimates converge as fast as $O(n^{-2/3})$. Since their restriction imposed on the nonlinearity is close to differentiability of m, their result is very similar to ours with $p=1$.

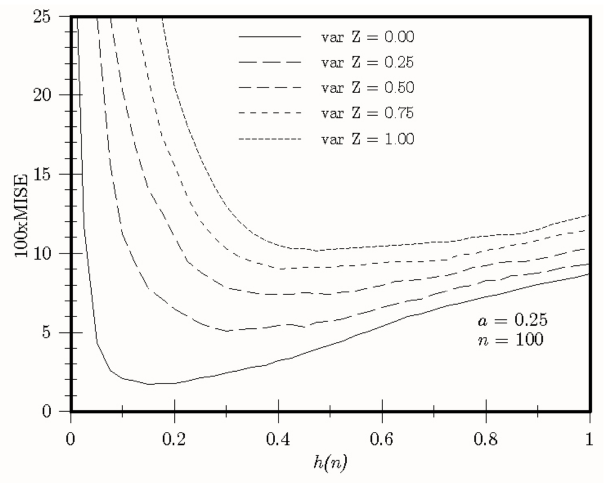

Fig. 3. MISE versus variance of noise.

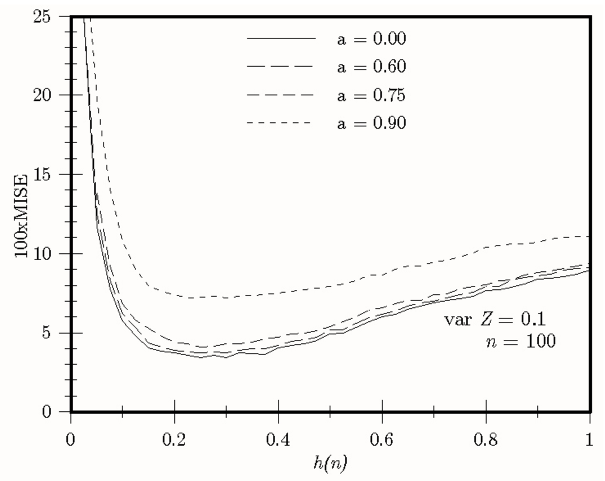

Fig. 4. MISE and the linear subsystem.

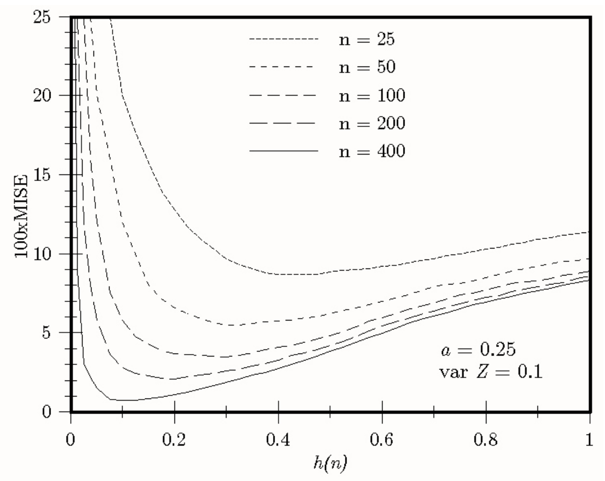

Fig. 2. MISE versus the bandwidth parameter $h(n)$.

All rates given above, are in an asymptotic sense. It means that, e.g., the error of our estimates is not greater than $c_1 n^{-\alpha}$, with some c_1. As we have mentioned, owing to the suitable choice of $\{h(n)\}$, α is optimal. The constant c_1 depends mainly on the kernel. Achieving the optimal c_1 is obviously less important. Numerical simulation shows that different kernels change the constant by only a few percents (see Fan [15]).

VII. Simulation Example

All results given in the previous section hold for large n. To illustrate the behavior of our estimates for a moderate and small number of observations, we present results of an numerical experiment. In the example, the system is described by the equation:

$$\left.\begin{aligned} X_{n+1} &= aX_n + m(U_n) \\ Y_n &= 0.5X_n + m(U_n) + Z_n, \end{aligned}\right\}$$

where the state is a scalar. The input density is not continuous and equals 2/3, 1/3 or zero according to $|u| \le 1/2$, $1/2 < |u| \le 1$, or $1 < |u|$, respectively. Moreover, $m(u)$ is defined in the following way:
a)

$$m(u) = \begin{cases} -1/2, & \text{for } u \le -2/3 \\ -1/4, & \text{for } -2/3 < u \le 1/3 \\ 0, & \text{for } 1/2 < u \le 1/3 \\ 1/4, & \text{for } 1/3 < u \le 2/3 \\ 1/2, & \text{for } 2/3 < u. \end{cases}$$

The characteristic is not continuous. Observe that $\mu(u) = m(u)$, all u such that $|u| \le 1$. In estimates (3.1) and (3.2), the kernel is rectangular with support $[-1, 1]$. The mean integrated square error (MISE for short) has been numerically calculated in the interval $[-0.9, 0.9]$. For $n = 25, 50, 100, 200$, and 400, the MISE is plotted in Fig. 2. In the figure and in the others, each curve represents both estimates since numerical results obtained for them are practically identical. Observe that selecting $h(n)$ smaller than optimal can cause large errors, particularly for large n. On the other hand, for a large number of observations, choosing the optimal bandwidth is more important and gives apparent results. The error is sensitive to variance of noise (see Fig. 3). For large variance, the problem of selecting the optimal value of $h(n)$ gets less vital. In Fig. 4, each plot represents different value of the parameter a of the linear subsystem. The MISE is relatively insensitive to the parameter and increases only for a close to 1. For the optimal bandwidth, the error decreases at the greatest possible speed, see the solid curve in Fig. 5. In a real situation, the characteristic would be, however, unknown. Therefore the bandwidth parameter could not be optimally selected, and the error would decrease at a slower rate. Fig. 5 shows also the error for next two characteristics:
b) $m(u) = u^2 \operatorname{sign}(u)$,
c) $m(u) = u$.

For discontinuous characteristic a), the error is surprisingly smaller than for the linear one. It seems that examined algorithms behave particularly well for piecewise constant nonlinearities. Results obtained for other kernels differs by a few percents only and are not presented.

The error, i.e., the accuracy of algorithms, depends more or less heavily on the choice of the bandwidth parameter

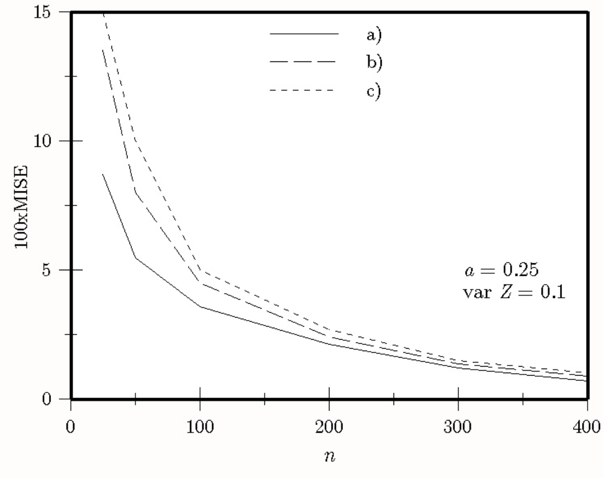

Fig. 5. Optimal MISE versus the number of observatios.

$h(n)$. In the simulation example, the parameter has been selected optimally since m has been known. The problem of proper selecting the parameter is far from a satisfactory solution. Reasonable is the idea of making the parameter a function of not only n but also observations; see Härdle [30] for inferring from independent pairs of observations. For such a $h(n)$, the parameter becomes a random variable and the theoretical analysis of the estimate gets much more difficult. As regards to the Hammerstein system identification, the problem is particularly tough since consecutive pairs of observations are mutually dependent.

VIII. Final Remarks

An important advantage of our estimates over those known in literature is the fact that their convergence rate is independent of the shape of the density of the input signal, i.e., is insensitive to its irregularities. The rate depends only on the smoothness of m and holds for any f bounded from zero. Convergence rates of other algorithms given in literature depend on the smoothness of both m and f and get worse for rough f. We want to mention another difference between algorithms examined in the paper and those studied in literature. The latter have a quotient form, while ours are of a simpler form. For example, the kernel algorithm

$$\sum_{i=1}^{n} Y_i K\left(\frac{u-U_i}{h(n)}\right) \Big/ \sum_{i=1}^{n} K\left(\frac{u-U_i}{h(n)}\right)$$

has been used by Greblicki and Pawlak [21], [22], and Krzyżak [34]. Rounding up can cause large errors when the denominator is near zero. Therefore our algorithms have not only a simpler but also a safer form. In the light of this, it seems that an extra computational effort caused by ordering is well compensated by described properties of algorithms.

Concluding, we want to state again that nonparametric methods offered by the statistical literature to recover nonlinear characteristics from dependent data assume that the stochastic process $\{Y_n; n = 0, 1, 2 \cdots\}$ is mixing. Unfortunately, processes encountered in applications are often not mixing. Therefore, nonparametric system identification algorithms require an independent treatment focused on ARMA processes.

Appendix A. Spacings and Related Results

In this section we give some lemmas concerning spacings and related problems. For any f, define $b_i = \int_{U_{(i-1)}}^{U_{(i)}} f(v)dv$. Random variables b_i's are called spacings. It is known, see Wilks [44, Ch. 8], David [12, Ch. 2], Balakrishnan and Cohen [2, Ch. 2], or Pyke [41], that b_i's have a beta distribution with parameters one and n while, for $i \neq j$, (b_i, b_j) has a Dirichlet distribution. Therefore, we have

Lemma A1: For any f, $Eb_i^p = 1 \cdot 2 \cdot \cdots \cdot n/(p+1)(p+2) \cdot \cdots \cdot (p+n)$, any $p > 0$, any $i = 1, 2, \cdots, n+1$. Therefore, in particular, $Eb_i = 1/(n+1)$, $Eb_i^2 = 2/(n+1)(n+2)$, $Eb_i^3 = 6/(n+1)(n+2)(n+3)$, Moreover, $Eb_ib_j = 1/(n+1)(n+2)$, any $i, j = 1, 2, \cdots, n+1$, such that $n \neq j$.

Observations $U_1, U_2, \cdots, U_n$ have randomly split the interval $[-1, 1]$ into $n+1$ subintervals which lengths equal $U_{(1)} - U_{(0)}$, $U_{(2)} - U_{(1)}$, $\cdots, U_{(n+1)} - U_{(n)}$, respectively. Observing that (2.1) implies $U_{(i)} - U_{(i-1)} = \int_{U_{(i-1)}}^{U_{(i)}} dv \leq \delta^{-1} \int_{U_{(i-1)}}^{U_{(i)}} f(v)dv$, and using Lemma A1, we easily obtain

Lemma A2: Let f satisfy (2.1). Then, $E(U_{(i)} - U_{(i-1)}) \leq 1/\delta(n+1)$, $E(U_{(i)} - U_{(i-1)})^2 \leq 2/\delta^2(n+1)(n+2)$, $E(U_{(i)} - U_{(i-1)})^3 \leq 6/\delta^3(n+1)(n+2)(n+3)$, $E(U_{(i)} - U_{(i-1)})(U_{(j)} - U_{(j-1)}) \leq 1/\delta^2(n+1)(n+2)$, any $i, j = 1, 2, \cdots, n+1$, such that $i \neq j$.

Lemma A3: Let f satisfy (2.1). Let K satisfy (3.3) and (3.4). Let $h > 0$. Then

$$\sum_{i=1}^{n+1} E \int_{-1}^{1} \left[\frac{1}{h} \int_{U_{(i-1)}}^{U_{(i)}} K\left(\frac{u-v}{h}\right) dv\right]^2 du \leq 2\delta^{-2}\kappa k(nh)^{-1} \tag{A.1}$$

and

$$E \int_{-1}^{1} \left[\sum_{i=1}^{n+1} \frac{1}{h} \int_{U_{(i-1)}}^{U_{(i)}} K\left(\frac{u-v}{h}\right) dv\right]^2 du \leq \delta^{-2}\kappa k h^{-1}.$$

Proof: The quantity under the sign of expectation in (A.1) is not greater than $(\kappa k/h)(U_{(i)} - U_{(i-1)})^2$ and, by virtue of Lemma A2, the first part of the lemma follows. To verify the other, observe that, for any $i, j = 1, 2, \cdots, n+1$, such that $i \neq j$,

$$\begin{aligned} &\int_{-1}^{1} \left[\int_{U_{(i-1)}}^{U_{(i)}} \frac{1}{h} K\left(\frac{u-v}{h}\right) f(v)dv \right. \\ &\quad \times \left. \int_{U_{(j-1)}}^{U_{(j)}} \frac{1}{h} K\left(\frac{u-v}{h}\right) f(v)dv\right] du \\ &\leq (\kappa k/h)(U_{(i)} - U_{(i-1)})(U_{(j)} - U_{(j-1)}). \end{aligned}$$

An application of Lemma A2 completes the proof. ∎

Lemma A4: Let f satisfy (2.1). Let K satisfy (3.3) and (3.4). Let $h > 0$. Then, Then

$$\sum_{i=1}^{n+1} E\int_{-\infty}^{\infty}\left[\frac{1}{h}\int_{U_{(i-1)}}^{U_{(i)}} K\left(\frac{u-U_{(i)}}{h}\right)f(v)dv\right]^2 du \le 2\kappa k(nh)^{-1}$$

and

$$E\int_{-\infty}^{\infty}\left[\sum_{i=1}^{n+1}\frac{1}{h}\int_{U_{(i-1)}}^{U_{(i)}} K\left(\frac{u-U_{(i)}}{h}\right)f(v)dv\right]^2 du \le \kappa k h^{-1}.$$

To verify the lemma, it suffices to repeat arguments used in the proof of Lemma A3.

Lemma A5: Let f satisfy (2.1) and let m satisfy (2.2). Let K satisfy (3.3) and (3.4). Let $h > 0$. Then

$$E\int_{-1}^{1}\left|\sum_{i=1}^{n+1}\frac{1}{h}\int_{U_{(i-1)}}^{U_{(i)}} \left|(\mu(U_{(i)})-\mu(v))K\left(\frac{u-U_{(i)}}{h}\right)\right| dv\right]^2 du \le c_1(n^2h)^{-1} \tag{A.2}$$

and

$$E\int_{-1}^{1}\left[\sum_{i=1}^{n+1}\frac{1}{h}\int_{U_{(i-1)}}^{U_{(i)}} \left|(\mu(U_{(i)})-\mu(v))K\left(\frac{u-v}{h}\right)\right| dv\right]^2 du \le c_2(n^2h)^{-1},$$

some c_1 and c_2 independent of both n and h.

Proof: Suppose at first that m satisfies a Lipschitz condition in the whole interval $[-1,1]$, i.e., that $|m(u)-m(v)| \le \alpha|u-v|$, some α, all u, v in the interval. Hence, for all i, all $v \in [-1,1]$,

$$\left|\mu(U_{(i)})-\mu(v)\right| \le \alpha d\left|U_{(i)}-v\right|. \tag{A.3}$$

Thus, the amount in (A.2) does not exceed $(\alpha d)^2\kappa k h^{-1}$ $E[\sum_{i=1}^{n+1}\int_{U_{(i-1)}}^{U_{(i)}}(U_{(i)}-v)dv]^2$. The expectation is bounded by $(1/4)E[\sum_{i=1}^{n+1}(U_{(i)}-U_{(i-1)})^2]^2$. Since the sum under the sign of expectation is not greater than $2\sum_{i=1}^{n+1}(U_{(i)}-U_{(i-1)})^3$, by virtue of Lemma A2, the amount in (A.2) is bounded by

$$3\delta^{-3}(\alpha d)^2\kappa k(n^2h)^{-1} \tag{A.4}$$

and inequality (A.2) follows. For m satisfying (2.2), (A.3) may not be satisfied for at most $q-1$ indexes i. For every such an i, $\left|\mu(U_{(i)})-\mu(v)\right| \le 2M$. Thus, the quantity in the first part of the assertion is not greater than twice that in (A.4) plus $2\kappa k(2M)^2h^{-1}E[\sum_{i=1}^{q-1}(U_{(i)}-U_{(i-1)})]^2$. Recalling Lemma A2, we complete the proof of the first part of the lemma.

In order to verify the second part is suffices to observe that

$$\left|\mu(U_{(i)})K\left(\frac{u-U_{(i)}}{h}\right)-\mu(u)K\left(\frac{u-v}{h}\right)\right| \le \left|\left[\mu(U_{(i)})-\mu(u)\right]K\left(\frac{u-U_{(i)}}{h}\right)\right| + \left|\mu(u)\left[K\left(\frac{u-U_{(i)}}{h}\right)-K\left(\frac{u-v}{h}\right)\right]\right|.$$

Take (3.5) into account and proceed in a similar manner. ■

Appendix B. Order Statistics and Dynamic Systems

The lemma given below is fundamental for the paper. It deals with the linear dynamic subsystem. For any matrix $P=[p_{ij}]$, $|P|$ denotes a matrix $[|p_{ij}|]$. For two matrices P and Q, $P \le Q$ means that the inequality holds for each pair of corresponding elements of the matrices, respectively.

Lemma B1: Let subsystem (2.4) be asymptotically stable. Let $\{U_i; i=\cdots-1,0,1,\cdots\}$ be a stationary white random stochastic process. Let m be a Borel function satisfying (2.3). Then, for any nonnegative Borel function g and any $i \ne j$,

$$E\left\{\left|\xi_{[i]}\xi_{[j]}\right| g(U_{(i)},U_{(j)},U_{(k)},U_{(m)})\right\} \le \rho M^2 n^{-1}E\{g(U_{(i)},U_{(j)},U_{(k)},U_{(m)})\}$$

some ρ dependent on A, b, and c only .

The lemma is an immediate consequence of the following one:

Lemma B2: Let subsystem (2.4) be asymptotically stable. Let $\{W_i; i=\cdots-1,0,1,\cdots\}$ be a stationary white random stochastic process such that $|W_i| \le \omega$, some ω. Then, for i, j, k, m, all different,

$$E\left\{\left|\left(X_{[i]}-EX_{[i]}\right)\left(X_{[j]}-EX_{[j]}\right)^T\right| \big| W_{(i)},W_{(j)},W_{(k)},W_{(m)}\right\} \le n^{-1}\omega^2 Q,$$

where the matrix Q depends only on A and b.

Proof: For convenience, denote $\eta_i = X_i - EX_i$, $\lambda_i = W_i - EW_i$ and suppose that $p > q$. We have $\eta_q = \sum_{m=-\infty}^{q-1}A^{q-m-1}b\lambda_m$ and $\eta_p = A^{p-q-1}\sum_{k=-\infty}^{q}A^{q-k}b\lambda_m + \sum_{k=q+1}^{p-1}A^{p-k-1}b\lambda_k$. Therefore, $E\{\eta_p\eta_q^T\,|\lambda_p,\lambda_q,\lambda_r,\lambda_s\} = E\{V_1\,|\lambda_p,\lambda_q,\lambda_r,\lambda_s\} + E\{V_2\,|\lambda_p,\lambda_q,\lambda_r,\lambda_s\}$, where

$$V_1 = A^{p-q-1}\sum_{k=-\infty}^{q}A^{q-k}b\lambda_k\left[\sum_{m=-\infty}^{q-1}A^{q-m-1}b\lambda_m\right]^T$$

and

$$V_2 = \sum_{k=q+1}^{p-1}A^{p-k-1}b\lambda_k\left[\sum_{m=-\infty}^{q-1}A^{q-m-1}b\lambda_m\right]^T.$$

Since $|\lambda_i| \le \omega$, then

$$E\{|V_1|\,|\lambda_p,\lambda_q,\lambda_r,\lambda_s\} \le \omega^2\left|A^{|p-q|-1}\right||b|\left|SS^T\right|\left|b^T\right|, \tag{B.1}$$

where $S = \sum_{n=0}^{\infty} |A^n|$. Observe that the above inequality holds for any different p and q.

Let again $p > q$. Since $E\lambda_k = 0$, for p, q, r, s all different, we find

$$E\{V_2\,|\lambda_p, \lambda_q, \lambda_r, \lambda_s\} = \begin{cases} A^{p-r-1}b\left[A^{q-s-1}b\right]^T \lambda_r\lambda_s, & \text{for } s < q < r < p \\ A^{p-s-1}b\left[A^{q-r-1}b\right]^T \lambda_r\lambda_s, & \text{for } r < q < s < p \\ 0, & \text{otherwise.} \end{cases}$$

From this and (B.1), it follows that , for any different p, q, r, s, we have finally

$$E\{\left|\eta_p\eta_q^T\right|\,|\lambda_p, \lambda_q, \lambda_r, \lambda_s\} \le \Gamma_{pqrs}, \tag{B.2}$$

some matrix Γ_{pqrs} such that $\sum_{p,q,r,s=1}^{n} \Gamma_{pqrs} \le n^3\Gamma$, where p, q, r, s under the sum are all different, and where some matrix Γ depends only on A and b.

Define the following event: $A_{pqrs} = \{\lambda_{(i)}$ came pth, $\lambda_{(j)}$ came qth, $\lambda_{(k)}$ came rth, $\lambda_{(m)}$ came sth$\}$ and observe that (B.2) implies

$$E\{\left|\eta_{[i]}\eta_{[j]}^T\right|\,|A_{pqrs}, \lambda_{(i)}, \lambda_{(j)}, \lambda_{(k)}, \lambda_{(m)}\} \le \Gamma_{pqrs},$$

any i, j, such that $i \ne j$. This, the equality

$$E\{\left|\eta_{[i]}\eta_{[j]}^T\right|\,|\lambda_{(i)}, \lambda_{(j)}, \lambda_{(k)}, \lambda_{(m)}\} = \sum_{\substack{p,q,r,s=1 \\ p,q,r,s \text{ all different}}}^{n} E\{\left|\eta_{[i]}\eta_{[j]}^T\right|\,|A_{pqrs}, \lambda_{(i)}, \lambda_{(j)}, \lambda_{(k)}, \lambda_{(m)}\} \times P\{A_{pqrs}\}$$

and the fact that $P\{A_{pqrs}\} = 1/n(n-1)(n-2)(n-3)$, complete the proof. ■

Appendix C

Denote $\bar{\mu}_h(u) = (1/h)\int_{-\infty}^{\infty} \mu(v)K((u-v)/h)dv$. In Wheeden and Zygmund [43, p. 148] we find

Lemma C1: Let μ be a Borel square integrable function on a real line. Let a Borel measurable kernel K satisfy (3.4) and be square integrable in a real line. Then

$$\int_{-\infty}^{\infty} (\bar{\mu}_h(u) - \mu(u))^2 du \to 0 \text{ as } h \to 0.$$

From Lemma C1, we easily obtain

Lemma C2: Let μ be bounded Borel integrable on $[-1, 1]$. Let K satisfy (3.3) and (3.4). Then

$$\int_{-\infty}^{\infty} (\mu_h(u) - \mu(u))^2 du \to 0 \text{ as } h \to 0.$$

References

[1] D. W. K. Andrews, "Non-strong mixing autoregressive processes", *Journal of Applied Probability*, vol. 21, pp. 930-934, 1984.

[2] N. Balakrishnan, and A. C. Cohen, *Order Statistics and Inference*, Academic Press: Boston, 1991,

[3] J. S. Bendat, *Nonlinear System Analysis and Identification*, Wiley: New York, 1990.

[4] H. J. Bierens, "Uniform consistency of kernel estimators of a regression function under generalized conditions," *Journal of the American Statistical Association*, vol. 78, pp. 699-707, 1983.

[5] G. Boente, and R. Fraiman, "Robust nonparametric regression estimation for dependent observations", *The Annals of Statistics*, vol. 17, pp. 1242-1256, 1989.

[6] [6] S. A. Billings, "Identification of nonlinear systems - a survey", *Proceedings of the IEE*, vol. 127, 1990 pp. 272-285.

[7] S. A. Billings, and S. Y. Fakhouri, "Identification of systems containing linear dynamic and static nonlinear elements," *Automatica*, vol. 18, pp. 15-26, 1982.

[8] D. R. Brilinger, "The identification of a particular nonlinear time series system," *Biometrica*, vol. 64, pp. 509-515, 1977.

[9] F. H. I. Chang, and R. Luus, "A noniterative method for identification using Hammerstein model," *IEEE Transactions on Automatic Control*, vol. AC-16, pp. 464-468, 1971.

[10] K. F. Cheng, and P. E. Lin, "Nonparametric estimation of a regression function," *Z. Wahrscheinlichtkeitstheorie verw. Gebiete*, vol. 57, pp. 223-233, 1989.

[11] G. Collomb, and W. Härdle, "Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations," *Stochastic Processes and their Applications*, vol. 23, pp. 77-89, 1986.

[12] H. A. David, *Order Statistics*. Wiley: New York, 1981.

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

[14] E. Eskinat, S. H. Johnson, and W. L. Luyben, "Use of Hammerstein models in identification of nonlinear systems," *AICh Journal*, vol. 37, pp. 255-268, 1991.

[15] J. Fan, "Design-adaptive nonparametric regression," *Journal of the American Statistical Association*, vol. 87, pp. 998-1004, 1992.

[16] Fan-Chu Kung, and Dong-Her Shih, "Analysis and identification of Hammerstein model non-linear systems using block-pulse function expansions," *International Journal of Control*, vol. 43, pp. 139-147, 1986.

[17] T. Gasser, and H. G. Müller, "Kernel estimation of regression functions," in *Smoothing Techniques for Curve Estimation. Lecture Notes in Mathematics 757*, eds. T. Gasser and M. Rosenblatt. Heidelberg: Springer-Verlag, pp. 23-68, 1979.

[18] P. G. Gallman, "A comparison of two Hammerstein model identification algorithms," *IEEE Transactions on Automatic Control*, vol.AC-21, pp. 124-126, 1976.

[19] G. Giunta, G. Jacovitti, and A. Neri, "Bandpass nonlinear system identification by higher cross correlation", *IEEE Transactions on Signal Processing*, vol. 39, pp. 2092-2095, 1991.

[20] W. Greblicki, "Non-parametric orthogonal series identification of Hammerstein systems," *International Journal of Systems Science*, vol. 20, pp. 2355-2367, 1989.

[21] W. Greblicki, and W. Pawlak, "Identification of discrete Hammerstein systems using kernel regression estimates," *IEEE Transactions on Automatic Control*, vol. AC-31, pp. 74-77, 1986.

[22] W. Greblicki, and W. Pawlak, "Nonparametric identification of Hammerstein systems," *IEEE Transactions on Information Theory*, vol. IT-35, pp. 409-418, 1989.

[23] W. Greblicki, and W. Pawlak, "Nonparametric identification of a cascade nonlinear time series system," *Signal Processing*, vol. 22, pp. 61-75, 1991.

[24] W. Greblicki, and W. Pawlak, "Nonparametric identification of a particular nonlinear time series system," *IEEE Transactions on Signal Processing*, vol. SP-40, pp. 985-989, 1992.

[25] W. Greblicki, and W. Pawlak, "Cascade non-linear system identification by a non-parametric method", *International Journal of Systems Science*, vol. 25, pp. 129-153, 1994.

[26] W. Greblicki, and W. Pawlak, "Dynamic system identification with order statistics", *IEEE Transactions on Information Theory*, vol. 40, pp. 1474-1489, 1994.

[27] W. Greblicki, and W. Pawlak, "Nonparametric recovering nonlinearities in block-oriented systems with the help of Laguerre polynomials", *Control - Theory and Advanced Technology*, vol. 10, Part 1, pp. 771-791, 1994.

[28] R. Haber, and H. Unbenhauen, "Structure identification of nonlinear dynamic systems - A survey on input/output approaches," *Automatica*, vol. 26, pp. 651-677, 1990.

[29] N. D. Haist, F. H. I. Chang, and R. Luus, "Nonlinear identification in the presence of correlated noise using a Hammerstein

model,“ *IEEE Transactions on Automatic Control*, vol. AC-18, pp. 552-555, 1973.

[30] W. Härdle, *Applied Nonparametric Regression.* Cambridge University Press: Cambridge, 1990.

[31] I. W. Hunter, and M. J. Korenberg, “The identification of nonlinear biological systems: Wiener and Hammerstein cascade systems,“ *Biological Cybernetics*, vol. 55, pp. 135-144, 1986.

[32] M. J. Korenberg, and I. W. Hunter, “The identification of nonlinear biological system: LNL cascade models,“ *Biological Cybernetics*, vol. 55, pp. 125-134, 1986.

[33] A. Krzyżak, “Identification of discrete Hammerstein systems by the Fourier series regression estimate,“ *International Journal of Systems Science*, vol. 20, pp. 1729-1744, 1989.

[34] A. Krzyżak, “On estimation of a class of nonlinear systems by the kernel regression estimate,“ *IEEE Transactions on Information Theory*, vol. IT-36, pp. 141-152, 1990.

[35] L. Zi-Qiang, “Controller design oriented model identification method for Hammerstein system,“ *Automatica*, vol. 29, pp. 767-771, 1993.

[36] Y. P. Mack, and H. G. Müller, “Convolution type estimators for nonparametric regression,“ *Statistics and Probability Letters*, vol. 7, pp. 229-239, 1989.

[37] K. S. Narendra, and P. G. Gallman, “An iterative model for the identification of nonlinear systems using a Hammerstein model,“ *IEEE Transactions on Automatic Control*, vol. AC-11, pp. 546-550, 1966.

[38] M. Pawlak, “On the series expansion approach to the identification of Hammerstein systems,“ *IEEE Transactions on Automatic Control*, vol. AC- 36, pp. 736-767, 1991.

[39] M. B. Priestley, *Non-linear and Non-stationary Time Series Analysis*, London: Academic Press, 1991.

[40] M. B. Priestley, and M. T. Chao, “Nonparametric function fitting,“ *Journal of the Royal Statistical Society*, vol. B34, pp. 385-392, 1972.

[41] R. Pyke, “Spacings,“ *Journal of the Royal Statistical Society*, Ser. B., vol. 27, pp. 395-436, 1965.

[42] M. A. L. Thathachar, and S. Ramaswamy, “Identification of a class of nonlinear systems,“ *International Journal of Control*, vol. 18, pp. 741-752, 1973.

[43] R. L. Wheeden, and A. Zygmund, *Measure and Integral*, New York: Dekker, 1977.

[44] S. S. Wilks, *Mathematical Statistics*, Wiley: New York, 1962.

[45] S. J. Yakowitz, “Nonparametric density estimation, prediction, and regression for Markov sequences,“ *Journal of the American Statistical Society*, vol. 80, pp. 215-221, 1985.

Włodzimierz Greblicki (M’92) was born in Poland in 1943. He received the M.Eng., Ph.D., and D.Sc. degrees in electronics from Technical University of Wrocław, Wrocław, Poland, in 1966, 1971, and 1975, respectively.

In 1966, he joined the Technical University of Wrocław and is now a professor at the Institute of Engineering Cybernetics at the University. During the 19861987 academic year, he held an appointment as a visiting professor with the Department of Electrical Engineering, University of Manitoba, Winnipeg, Manitoba, Canada. He teaches courses in control theory. His research interests include classification and system identification as well as mathematical statistics, in particular, nonparametric estimation.