CONTROL-THEORY AND ADVANCED TECHNOLOGY
Vol. 10, No. 4, Part 1, pp. 771–791, December, 1994

C92090RR

Contributed Paper

NONPARAMETRIC RECOVERING NONLINEARITIES IN BLOCK-ORIENTED SYSTEMS WITH THE HELP OF LAGUERRE POLYNOMIALS*

W. GREBLICKI[1] AND M. PAWLAK[2]

Abstract. The nonlinearity in discrete nonlinear systems of the block-oriented form is estimated. In particular, memoryless, cascade and parallel models are examined. Algorithms being in the form of the ratio of two Laguerre polynomials are proposed, and their asymptotic properties are established. The conditions for the pointwise and global consistency are found. The identification algorithms are consistent for a large class of nonlinear characteristics which cannot be parametrized.

Key Words—System identification, nonlinear systems, block-oriented models, nonparametric regression, Laguerre functions, consistency, rate of convergence.

1. Introduction

Identification of a physical process is the problem of complete determination of its characteristics from corresponding values of input and output data. A large number of techniques exist for identification of linear models (Ljung, 1987). There is no particular reason, however, to assume that the process dynamics can be described by a linear model. In fact, the linearity assumption can be regarded as a first-order approximation of stochastic discrete time phenomena, and the nonlinear behavior is the rule, rather than the exception. The identification of nonlinear stochastic systems has been an active area of research lately (see, e.g., Bendat, 1990; Priestley, 1988; Tong, 1990).

All proposed nonlinear system identification techniques strongly depend on the selected representation of the examined system. In nonparametric setting, the Volterra and Wiener representations have been traditionally used (Rugh, 1981), yielding, however, complicated identification algorithms. Some parametric models have also been extensively studied (Priestley, 1988; Tong, 1990).

Another promising approach is based on the assumption that the system can be represented by the interconnection of linear dynamic models and static nonlinear elements (Bendat, 1990; Billings, 1980), yielding the concept of block-oriented models. In particular, cascade and parallel connections have received a great deal of attention (Bendat, 1990; Billings and Fakhouri, 1982; Greblicki and

Received by the editors June 17, 1992 and in finally revised form December 25, 1993.
This work was supported by NSERC Grant A8131 and by NATO Grant CRG 900 984.
Institute of Engineering Cybernetics, Technical University of Wroclaw, Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland.
Department of Electrical and Computer Engineering, The University of Manitoba, Winnipeg, Manitoba, Canada R3T 5V6.

Pawlak, 1989). These connections can serve as basic building blocks for more complicated models. In fact, in Palm (1979), Sanberg (1991) and Stone (1985), it has been demonstrated that such models can accurately approximate a large class of nonlinear systems. See also Eskinat et al. (1991), Hunter and Korenberg (1986), Marmarelis and Marmarelis (1978) and references cited therein for a large number of applications of such models.

In this paper, we consider the problem of identification of a class of block-oriented models consisting of cascade and parallel connections. We first study the memoryless nonlinear model to establish a basic theory of our identification algorithms. Then, we extend those results to the case of cascade and parallel models. We are mostly interested in recovering a nonlinear element of the particular model. This is due to the fact that the identification of linear and nonlinear subsystems can be decoupled, and algorithms for identification of linear subsystems have been presented elsewhere (Bendat, 1990; Billings and Fakhouri, 1982; Brillinger, 1977). As for the nonlinear elements, it has been typically assumed that they are of a polynomial form; i.e., they can be parametrized by a finite number of coefficients. Not all problems are, however, parametric nor can they all be parametrized. This is the case if it is known only that, e.g., the unknown characteristic is continuous or bounded or has a finite derivative, etc.

In this paper, we propose identification techniques which are able to be consistent for a large class of nonlinear characteristics, i.e., those which cannot be parametrized. The proposed estimates are in the form of the rational function (i.e., the ratio of two polynomials) stemming from the theory of regression function estimation (see Eubank (1988) and Härdle (1990) for a detailed account of the existing theory of nonparametric curve estimation).

Specifically, we apply the Laguerre orthonormal polynomials. It is well known that they constitute an orthonormal basis of the space L_2 $([0, \infty))$.

We give conditions for the estimate consistency and rate of convergence. Both pointwise and global properties of the estimate are examined. Some small sample properties are also presented. The Laguerre functions have recently been used for the problem of identification of linear systems (see Cluett and Weng, 1992; Fu and Dumont, 1993; Mohan and Datta, 1991; Wahlberg, 1991 and references cited therein). The popularity of the Laguerre system stems from its unique properties. First, the Laplace transform of the Laguerre functions also defines the complete orthonormal system. Furthermore, the result of convoluting two Laguerre functions can be written in the additive form. This property is not shared by any other orthonormal system in L_2 $([0, \infty))$ (Dooge, 1965; Holland, 1969).

2. Preliminaries

The Laguerre polynomials are defined by the following Rodrigues formula (Sansone, 1991, p. 297)

$$L_n(x) = \frac{1}{n!} e^x \frac{d^n}{dx^n} (x^n e^{-x}), \qquad n = 0, 1, 2, 3, \cdots, \quad x \geq 0 .$$

Each $L_n(x)$ is a polynomial of degree n, and it is explicitly given by

$$L_n(x) = \sum_{k=0}^{n} (-1)^k \frac{1}{k!} \binom{n}{k} x^k, \qquad n = 0, 1, 2, \cdots.$$

Hence, $L_0(x) = 1$, $L_1(x) = -x + 1$, $L_2(x) = x^2/2 - 2x + 1$, $L_3(x) = -x^3/6 + 3x^2/2 - 3x + 1$ and so on. Note also that $L_n(0) = 1$, $n \geq 0$.

It is well known that these polynomials form an orthonormal system with respect to the weighting function e^{-x}, $x \in [0, \infty)$, i.e.,

$$\int_0^\infty l_n(x) l_m(x) dx = \begin{cases} 1 & \text{for } n = m, \\ 0 & \text{otherwise,} \end{cases}$$

where $l_n(x) = e^{-x/2} L_n(x)$ (Sansone, 1991; Szegö, 1978).

Hence, the system $\{l_n(x)\}$ forms an orthonormal basis for $L_2([0, \infty))$. Consequently, every function $h(x)$ from $L_2([0, \infty))$ has the formal representation

$$\left. \begin{aligned} h(x) &= \sum_{k=0}^{\infty} a_k l_k(x) \\ a_k &= \int_0^\infty h(v) l_k(v) dv \end{aligned} \right\}. \tag{1}$$

It is also known (Sansone, 1991) that the Nth partial sum in (1) can be written as $\int_0^\infty h(y) d_N(x, y) dy$, where $d_N(x, y)$ is the kernel of the Laguerre system defined as

$$d_N(x, y) = \sum_{k=0}^{N} l_k(x) l_k(y). \tag{2}$$

In the study of our identification algorithms, we need some properties of $d_N(x, y)$. First of all, let us recall the following inequalities:

$$\max_{\delta \leq x} |l_n(x)| \leq c_1(\delta) n^{-\frac{1}{4}} \tag{3}$$

for any $\delta > 0$, $n > 0$ (see Szegö, 1978, p. 241),

$$\max_{0 \leq x \leq \delta} |x^{\frac{1}{4}} l_n(x)| \leq c_2(\delta) n^{-\frac{1}{4}}, \tag{4}$$

for any $\delta > 0$, $n > 0$ (see Szegö, 1978, p. 178, where $c_1(\delta)$ and $c_2(\delta)$ are some constants dependent on δ but independent of n).

Now let $x > 0$ and $\delta > 0$. Then, by virtue of (2), (3) and (4)

$$\begin{aligned} \max_{\delta \leq y} |d_N(x, y)| &\leq c_1(x) c_1(\delta) \left(\sum_{k=1}^{N} k^{-\frac{1}{2}} + 1 \right) \\ &\leq c_3(x, \delta) N^{\frac{1}{2}} \end{aligned} \tag{5}$$

and

$$\begin{aligned} \max_{0 \leq y \leq \delta} |y^{\frac{1}{4}} d_N(x, y)| &\leq c_1(x) c_2(\delta) \left(\sum_{k=1}^{N} k^{-\frac{1}{2}} + 1 \right) \\ &\leq c_4(x, \delta) N^{\frac{1}{2}}, \end{aligned} \tag{6}$$

where $c_3(x, \delta)$ and $c_4(x, \delta)$ are independent of N.

3. Identification of a Memoryless System

In this section, we identify a memoryless system shown in Fig. 1. Hence,

$$Y_n = m(X_n) + \xi_n. \tag{7}$$

Input random variables $X_1, X_2, \cdots$ are independent and identically distributed. Moreover, all X_n's take only positive values. They have the probability density denoted by f. We assume that the density satisfies the following restrictions:

$$|f(x) - f(0+)| \leq Mx^{\alpha}, \quad 0 < x < \delta, \quad 0 < \alpha \leq 1, \quad M \geq 0 \tag{8}$$

and

$$f(x) = O(e^x x^{\frac{1}{6}-\varepsilon}), \quad \varepsilon > 0, \quad x \to \infty. \tag{9}$$

The condition (8) describes the behavior of $f(x)$ in the neighborhood of the origin, while (9) limits the behavior of $f(x)$ in the infinity. Clearly, (8) and (9) are satisfied for all commonly used densities on $[0, \infty)$, e.g., uniform, exponential, gamma(α, β), $\alpha \geq 1$, Weibull(α, β), $\alpha \geq 1$, lognormal, etc.

The random disturbance ξ_n is a stationary white noise independent of the input signal. Its mean is zero, and variance is denoted by σ_ξ^2.

The characteristic m of the system is a Lebesgue measurable function satisfying the following restrictions:

$$|m(x) - m(0+)| \leq Lx^{\beta}, \quad 0 < x < \delta, \quad 0 < \beta \leq 1, \quad L \geq 0, \tag{10}$$

$$\int_0^{\infty} m^2(x) f(x) dx < \infty. \tag{11}$$

It is worth nothing that the condition in (10) is satisfied if $m^{(1)}(0+)$ exists. The restriction in (11), on the other hand, describes the behavior of $m(x)$ at $x \to \infty$. It is clear that these assumptions hold for a broad class of functions, e.g., all polynomials, all rational functions on $[0, \infty)$, etc. Hence, the class of all possible functions satisfying (10) and (11) is so wide that it cannot be parametrized. Therefore our problem is nonparametric. Our goal is to recover the characteristic $m(x)$, $x \in [0, \infty)$, from the input-output observations $(X_1, Y_1), (X_2, Y_2), \cdots$.

It is clear that

$$m(x) = E\{Y_n | X_n = x\}, \tag{12}$$

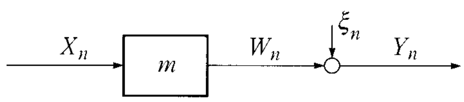

Fig. 1. Nonlinear memoryless system.

i.e., that m is a regression function of Y_n on X_n. In order to present our algorithm, we observe that

$$m(x) = \frac{g(x)}{f(x)}, \tag{13}$$

where $g(x) = m(x)f(x)$ for every x, where $f(x) \neq 0$. We shall now show how to estimate $g(x)$ and $f(x)$. Owing to (1), we have

$$g(x) = \sum_{k=0}^{\infty} a_k l_k(x), \tag{14}$$

which means that

$$a_k = \int_0^{\infty} g(x) l_k(x) dx = \int_0^{\infty} m(x) l_k(x) f(x) dx .$$

As a consequence of (12), we can write

$$a_k = E\{Y_n l_k(X_n)\} . \tag{15}$$

The coefficient a_k can be easily estimated in the following way:

$$\hat{a}_k = \frac{1}{n} \sum_{i=1}^{n} Y_i l_k(X_i) . \tag{16}$$

Taking (14), (15) and (16) into account, we get the following estimate of $g(x)$:

$$\hat{g}(x) = \sum_{k=0}^{N(n)} \hat{a}_k l_k(x), \tag{17}$$

where $\{N(n)\}$ is a sequence of integers. $N(n)$ indicates the number of terms taken in our estimate. The point at which we truncate the expansion depends on the number of observations. We shall later show that, for a suitably selected sequence $\{N(n)\}$, $\hat{g}(x)$ converges to $g(x)$ as the number of observations tends to infinity.

Similarly, $f(x) = \sum_{k=0}^{\infty} b_k l_k(x)$, which means that $b_k = \int_0^{\infty} l_k(x) f(x) dx = E\{l_k(X_n)\}$. Consequently, $\hat{f}(x) = \sum_{k=0}^{N(n)} \hat{b}_k l_k(x)$, where $\hat{b}_k = 1/n \sum_{i=1}^{n} l_k(X_i)$ is our estimate of $f(x)$. In the light of (13) and (17), our estimate of $m(x)$ is $\hat{g}(x)/\hat{f}(x)$, i.e.,

$$\hat{m}(x) = \frac{\sum_{k=0}^{N(n)} \hat{a}_k l_k(x)}{\sum_{k=0}^{N(n)} \hat{b}_k l_k(x)} . \tag{18}$$

In Greblicki and Pawlak (1985), an estimate of the similar form has been examined in the case of the Fourier and Hermite series.

Remark 1: The estimate in (18) is in the form of the ratio of two polynomials due to the identity in (13) and the fact that the density of X is unknown. In the case, when $f(x)$ is known and additionally it is exponentially distributed with a mean value 1, then, owing to (1), one can estimate $m(x)$ by

$$\bar{m}(x) = \sum_{k=0}^{N} \bar{c}_k L_k(x), \quad \bar{c}_k = n^{-1} \sum_{j=1}^{n} Y_j L_k(X_j).$$

That is, in this case, we have a purely polynomial estimate of $m(x)$. It is worth noting that if $m(x)$ is a polynomial of the order M, then $E\bar{m}(x) = m(x)$, if only $N = M$, i.e., the estimate $\bar{m}(x)$ is unbiased.

Concerning the sequence $\{N(n)\}$ appearing in (18), we assume that

$$N(n) \to \infty \qquad \text{as} \quad n \to \infty \tag{19}$$

and

$$\frac{N(n)}{n} \to 0 \qquad \text{as} \quad n \to \infty . \tag{20}$$

Our first result concerns the convergence of $\hat{m}(x)$ to $m(x)$ in the pointwise sense.

Theorem 1. Let m satisfy (10) and (11). Let f satisfy (8) and (9). Let the number sequence $\{N(n)\}$ satisfy (19) and (20). Then,

$$\hat{m}(x) \to m(x) \quad \text{as} \quad n \to \infty$$

in probability at every point $x > 0$, at which both m and f are differentiable and $f(x) > 0$.

Proof. First, we shall show that

$$E(\hat{g}(x) - g(x))^2 \to 0 \qquad \text{as} \quad n \to \infty \tag{21}$$

at every point $x > 0$, at which both f and m are differentiable. Obviously, $E\hat{a}_k = a_k$ and $E\hat{g}(x) = \int_0^\infty m(y)f(y)d_{N(n)}(x, y)dy = \sum_{k=0}^{N(n)} a_k l_k(x)$ is just the $N(n)$th partial sum of the expansion of g in the Laguerre system. In Appendix B, it is shown that

$$E\hat{g}(x) \to g(x) \quad \text{as} \quad N(n) \to \infty$$

at every point $x > 0$, at which both f and m are differentiable. In order to verify convergence of $\operatorname{var}\hat{g}(x)$ to zero, let us observe that

$$\hat{g}(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i d_{N(n)}(x, X_i),$$

where $d_{N(n)}$ is the Nth kernel of the Laguerre system. Thus,

$$\operatorname{var}\hat{g}(x) = \frac{1}{n}\operatorname{var}[Y_n d_{N(n)}(x, X_n)]$$

$$\leq \frac{1}{n} E\{Y_n^2 d_{N(n)}^2(x, X_n)\}.$$

Since due to (7) $E\{Y_n^2 \mid X_n = x\} = \psi(x)$, where $\psi(x) = m^2(x) + \sigma_\xi^2$, we get

$$\operatorname{var}\hat{g}(x) \leq \frac{1}{n}\int_0^\infty \psi(y) d_{N(n)}^2(x, y) f(y) dy.$$

Let $\delta > 0$. Owing to (5),

$$\int_{\delta}^{\infty}\psi(y)d_{N(n)}^2(x,y)f(y)dy \le c_3^2(x,\delta)N(n)\int_{\delta}^{\infty}\psi(y)f(y)dy = c_3^2(x,\delta)N(n)EY_n^2.$$

In turn, using (6) we have

$$\int_{0}^{\delta}\psi(y)d_{N(n)}^2(x,y)f(y)dy \le c_4^2(x,\delta)N(n)\int_{0}^{\delta}y^{-\frac{1}{2}}\psi(y)f(y)dy.$$

As a consequence of (8), $f(y) \le My^{\alpha} + f(0+)$ for $0 \le y \le \delta$ and for δ sufficiently small. For this δ, we have the above integral not greater than

$$c_4^2(x,\delta)N(n)\left[M\int_{0}^{\delta}\psi(y)y^{\alpha-\frac{1}{2}}dy + f(0+)\int_{0}^{\delta}\psi(y)y^{-\frac{1}{2}}dy\right].$$

It is plain that it suffices to consider $\int_0^{\delta}\psi(y)y^{-\frac{1}{2}}dy$. This clearly is equal to

$$\sigma_{\xi}^2\int_0^{\delta}y^{-\frac{1}{2}}dy + \int_0^{\delta}m^2(y)y^{-\frac{1}{2}}dy = 2\sigma_{\xi}^2\delta^{\frac{1}{2}} + \int_0^{\delta}m^2(y)y^{-\frac{1}{2}}dy. \qquad (22)$$

By assumption (10), we get

$$m^2(y) \le 2m^2(0+) + 2L^2y^{2\beta} \quad \text{for} \quad 0 \le y \le \delta.$$

Thus, the integral in (22) does not exceed $4m^2(0+)\delta^{1/2} + 2L^2\delta^{2\beta+1/2}/(2\beta+1/2)$. In this way, we have shown that

$$\operatorname{var}\hat{g}(x) \le \frac{c_6 N(n)}{n}, \qquad (23)$$

where c_6 is some constant dependent on x, but independent of n. Thus, (23) and (20) yield

$$\operatorname{var}\hat{g}(x) \to 0 \quad \text{as} \quad \frac{N(n)}{n} \to 0$$

for all $x > 0$. This and Appendix B give desired (21).

Using identical arguments as above, one can easily verify that

$$E(\hat{f}(x) - f(x))^2 \to 0 \quad \text{as} \quad n \to \infty \qquad (24)$$

at any point $x > 0$, at which f is differentiable. This completes the proof of Theorem 1.

Remark 2: The conditions imposed in Theorem 1 do not require that $m(x)$ and $f(x)$ are bounded. In particular, the convergence property holds for linear and polynomial characteristics on $[0,\infty)$. Furthermore, if one assumes that $m(x)$ and $f(x)$ are bounded, then the condition $N(n)/n \to 0$ can be replaced by the weaker one $N^{1/2}(n)/n \to 0$. Indeed, it is sufficient to consider $\int_0^{\infty}m(y)f(y)d_{N(n)}^2(x,y)dy$, which is bounded by $C\int_0^{\infty}d_{N(n)}^2(x,y)dy$, where $C = \sup_{x\in[0,\infty)}|m(x)f(x)|$.

Uspensky (1927, p. 614) shows that $\int_0^{\infty}d_N^2(x,y)dy = (x^{-1/2}/\pi)N^{1/2} + e^{-x}E_N$, where E_N is uniformly bounded with respect to N, as x varies in a finite interval $[x-\delta, x+\delta]$, $\delta > 0$. This yields our claim.

Remark 3: Theorem 1 gives the conditions for the estimate convergence at a point $x > 0$. At $x = 0$, our estimate need not be convergent (Szegö, 1978, p. 247). At this point, the estimate of the form $\Sigma_{k=0}^{N(n)}(1-k/N(n))\hat{a}_k l_k(x)/\Sigma_{k=0}^{N(n)}(1-k/N(n))\hat{b}_k l_k(x)$ can be considered. This is a modification of $\hat{m}(x)$ based on the Cesaro method of summation. Using Theorem 9.1.7 in Szegö (1978) and the techniques presented in the proof of Theorem 1, one can show that this estimate converges to $m(0)$, assuming that $m(x)$ and $f(x)$ are continuous at $x=0$.

Thus far we have examined only the pointwise properties of our estimate. It is also interesting to examine some global errors, as, e.g., the mean integrated squared error (MISE),

$$\mathrm{MISE}(\hat{m}) = E\int_0^\infty [\hat{m}(x) - m(x)]^2 f^2(x)dx .$$

This is a popular measure of assessing the performance of curve estimates (Härdle, 1990, Chapter 4). To evaluate $\mathrm{MISE}(\hat{m})$, let us observe that

$$(\hat{m}(x) - m(x))^2 \le \frac{2}{f^2(x)}[\hat{m}^2(x)(\hat{f}(x) - f(x))^2 + (\hat{g}(x) - g(x))^2].$$

Let us assume that $m(x)$ and ξ_n are bounded. As a consequence of this, there exists $c > 0$ such that $|\hat{m}(x)| \le c$ and $E(\hat{m}(x) - m(x))^2 \le (2/f^2(x)) \times [c^2 E(\hat{f}(x) - f(x))^2 + E(\hat{g}(x) - g(x))^2]$. Integrating the last inequality yields

$$\mathrm{MISE}(\hat{m}) \le 2c^2 E\int_0^\infty (\hat{f}(x) - f(x))^2 dx + 2E\int_0^\infty (\hat{g}(x) - g(x))^2 dx . \quad (25)$$

This inequality allows us to establish the global convergence of $\hat{m}$. Indeed, using Parseval's formula we have

$$E\int_0^\infty (\hat{g}(x) - g(x))^2 dx = \sum_{k=0}^{N(n)} E(\hat{a}_k - a_k)^2 + \sum_{k=N(n)+1}^{\infty} a_k^2 . \quad (26)$$

The second term tends to zero as $N(n) \to \infty$. To evaluate the first term in (26), let us note that $E(\hat{a}_k - a_k)^2$ is bounded by

$$\frac{1}{n}\left[\int_0^\infty m^2(x) f(x) l_k^2(x)dx + \sigma_\xi^2 \int_0^\infty f(x) l_k^2(x)dx\right].$$

Employing inequalities in (3), (4), conditions (8), (10), (11) and arguments like in the proof of Theorem 1, one can easily show that both $\int_0^\infty m^2(x)f(x)l_k^2(x)dx$ and $\int_0^\infty f(x)l_k^2(x)dx$ are of order $(k+1)^{-1/2}$. Hence, the first term on the right-hand-side of (26) does not exceed $cN^{1/2}(n)/n$ for some $c > 0$. Clearly, the first term in (25) can be evaluated in an analogous way. This proves the following theorem.

Theorem 2. Let all the conditions of Theorem 1 concerning m and f be satisfied. Let, additionally, m and ξ_n be bounded. If (19) and $N^{1/2}(n)/n \to 0$ hold, then

$$\mathrm{MISE}(\hat{m}) \to 0 \quad \text{as} \quad n \to \infty .$$

Some involved analysis (see Greblicki and Pawlak (1985) and Hall (1980) for a related discussion) allows us also to establish the rate at which the terms $\sum_{k=N+1}^{\infty} a_k^2$ and $\sum_{k=N+1}^{\infty} b_k^2$ tend to zero as $N \to \infty$. In fact, if

$$\int_0^{\infty} x^p (m^{(p)}(x))^2 e^{-x} dx < \infty, \quad \int_0^{\infty} x^p (f^{(p)}(x))^2 e^{-x} dx, \quad p \geq 1, \tag{27}$$

then $\sum_{k=N+1}^{\infty} a_k^2 + \sum_{k=N+1}^{\infty} b_k^2 = O(N^{-p})$.

As a result,

$$\text{MISE}(\hat{m}) \leq \frac{1}{n}(c_1 N^{\frac{1}{2}}(n)) + c_2 N^{-p}(n)$$

for some positive constants c_1, c_2 which are independent of N and n. Hence, under the conditions of Theorem 2 and (27), the rate of convergence of $\text{MISE}(\hat{m})$ is not slower than $O(n^{-2p/(2p+1)})$ because $N(n)$ is selected as $n^{2/(2p+1)}$. It is worth noting that the identification algorithms based on Hermite polynomials studied in Greblicki and Pawlak (1992) can reach (under similar assumptions on $m(x)$) the global rate $O(n^{-6p/(6p+5)})$, which is slower than ours, which is $O(n^{-2p/(2p+1)})$. The latter rate also holds for identification techniques based on the Legendre polynomials (Pawlak, 1991). Here, however, one has to assume that the characteristic is defined on a finite interval.

4. Recovering the Nonlinearity in Cascade Systems

In this section, we identify the nonlinear characteristic of the cascade system shown in Fig. 2.

The system comprises two subsystems. The first is nonlinear and memoryless, and its characteristic is denoted by m. It means that

$$W_n = m(X_n). \tag{28}$$

The other subsystem is linear dynamic and is described by the discrete convolution

$$Y_n = \sum_{i=0}^{\infty} k_i W_{n-i} + Z_n, \tag{29}$$

where $\{k_n; n = 0, 1, 2, \cdots\}$ is the impulse response of the subsystem. The system of this form is often referred to as the Hammerstein model (Greblicki and Pawlak, 1986; 1989; 1991; 1992; Narendra and Gallman, 1966; Pawlak, 1991). $Z_0, Z_1, Z_2, \cdots$ is a sequence of uncorrelated random variables with zero mean and variance σ_Z^2.

The input signals $X_0, X_1, X_2, \cdots$ form a sequence of independent and identically distributed random variables which are independent of $\{Z_n\}$. Their probability density is denoted by f, and it is supported on $[0, \infty)$. We assume that it

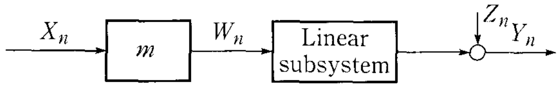

Fig. 2. Cascade nonlinear system.

satisfies restrictions (8) and (9). In turn, the nonlinear characteristic satisfies (10), (11), i.e., $Em^2(X_0) < \infty$. Note that now the output process $\{Y_n\}$ is no longer white.

We assume, moreover, that

$$\sum_{n=0}^{\infty} k_n^2 < \infty. \tag{30}$$

This and the fact that W_0 has the second moment imply that Y_n is a random variable.

In further parts of this section, we shall need the following restrictions:

$$\sum_{n=0}^{\infty} |k_n| < \infty \tag{31}$$

and

$$\sum_{n=0}^{\infty} \sum_{j=0}^{\infty} |k_{n-j}|\,|k_j| < \infty. \tag{32}$$

Let us also assume, without loss of generality, that $k_0 = 1$. Let us now shortly discuss the problem of identification of the linear subsystem. First, let us observe that

$$\operatorname{cov}(Y_n, X_n) = \gamma \quad \text{and} \quad \operatorname{cov}(Y_{n+s}, X_n) = \gamma k_s, \quad s = 1, 2, \cdots, \tag{33}$$

where $\gamma = \operatorname{cov}(m(X_n), X_n)$. Let us assume that $\gamma \neq 0$. Thus, clearly

$$\hat{k}_s = \frac{\hat{\theta}_s}{\hat{\theta}_0},$$

where

$$\hat{\theta}_s = \frac{1}{n} \sum_{j=1}^{n-s} (Y_{j+s} - \bar{Y})(X_j - \bar{X}), \quad s \geq 0, \tag{34}$$

$$\bar{Y} = \frac{1}{n} \sum_{j=1}^{n} Y_j, \quad \bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j$$

can serve as an estimate of k_s for $s = 1, 2, \cdots$. This result allows us to carry out the identification in the frequency domain. Indeed, formation of the Fourier transform of the relationship in (33) yields

$$h_{XY}(\omega) = \gamma K(\omega), \quad |\omega| \leq \pi,$$

where $h_{XY}(\omega) = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \theta_s \cos(s\omega)$ is the cross-spectral density function of processes $\{X_n\}$ and $\{Y_n\}$ and $\theta_s = \operatorname{cov}(Y_{n+s}, X_n)$.

Moreover,

$$K(\omega) = (2\pi)^{-1} \sum_{s=0}^{\infty} k_s \exp(-is\omega)$$

is the transfer function of the linear subsystem. Thus, the standard spectral estimation theory (Priestley, 1981) allows us to estimate $K(\omega)$ by

$$(2\pi\hat{\theta}_0)^{-1} \sum_{|s| \leq n} \chi(s) \hat{\theta}_s \cos(s\omega),$$

where $\chi(s)$ is the lag window (Priestley, 1981). See Brillinger (1977) for the consistency result for this estimate.

Let us now return to the problem of estimating $m(x)$. First, let us observe that

$$Y_n = \mu(X_n) + \xi_n,$$

where

$$\mu(X_n) = m(X_n) + Em(X_0)\sum_{i=1}^{\infty} k_i$$

and

$$\xi_n = \sum_{i=1}^{\infty} k_i[m(X_{n-i}) - Em(X_{n-i})] + Z_n,$$

which means that we observe the regression noised by ξ_n. Hence, the cascade model can be represented in the form like in (7). Contrary to Sec. 3, however, the noise is now not white since ξ_i and ξ_j are correlated for $i \neq j$.

In order to recover $\mu(x)$, we apply estimate (18). Symbols $g, \hat{g}$ and $f, \hat{f}$ have the same meaning as in Sec. 3.

Arguing as in Sec. 3, we can verify that

$$E\hat{g}(x) \to \mu(x)f(x) \quad \text{as} \quad N(n) \to \infty$$

at every point $x > 0$, at which both f and m are differentiable. In turn,

$$\begin{aligned} \operatorname{var}\hat{g}(x) &= \frac{1}{n}\operatorname{var}[Y_n d_{N(n)}(x, X_n)] \\ &\quad + \frac{2}{n}\sum_{s=1}^{n-1}\left(1 - \frac{s}{n}\right)\operatorname{cov}[Y_s d_{N(n)}(x, X_s), Y_0 d_{N(n)}(x, X_0)] \\ &= V_1(x) + V_2(x). \end{aligned} \tag{35}$$

Using arguments similar to those in Sec. 3, one can show that

$$V_1(x) \to 0 \quad \text{as} \quad \frac{N(n)}{n} \to 0.$$

We shall now verify that

$$V_2(x) \to 0 \quad \text{as} \quad N(n) \to \infty \ \text{ and } \ \frac{N(n)}{n} \to 0. \tag{36}$$

Let us observe that due to (29), for $s \geq 1$,

$$\begin{aligned} &\operatorname{cov}[Y_s d_{N(n)}(x, X_s), Y_0 d_{N(n)}(x, X_0)] \\ &= \sum_{i=-\infty}^{s}\sum_{j=-\infty}^{0} k_{s-i}k_{-j}\operatorname{cov}[W_i d_{N(n)}(x, X_s), W_j d_{N(n)}(x, X_0)] \\ &= k_s\sum_{j=-\infty}^{0} k_{-j}\operatorname{cov}[W_0 d_{N(n)}(x, X_s), W_j d_{N(n)}(x, X_0)] \\ &\quad + \sum_{j=-\infty}^{-1} k_{s-j}k_{-j}\operatorname{cov}[W_j d_{N(n)}(x, X_s), W_j d_{N(n)}(x, X_0)], \end{aligned} \tag{37}$$

where W_i is defined in (28).

We used above the fact that the input is a white noise. The covariance term in the first sum in (37) is equal to $EW_0 Ed_{N(n)}(x, X_0)\operatorname{cov}(W_0, d_{N(n)}(x, X_0))$ for $j < 0$ and $Ed_{N(n)}(x, X_0)\operatorname{cov}(W_0, W_0 d_{N(n)}(x, X_0))$ for $j = 0$. Since $Ed_{N(n)}(x, X_0) = E\hat{f}(x)$ converges to $f(x)$ and $E\{W_0^2 d_{N(n)}(x, X_0)\} \to m^2(x)f(x)$ as $N \to \infty$, the absolute value of the first term in (37) is bounded by $c|k_s|\sum_{j=0}^{\infty}|k_j|$ at every $x > 0$, at which both f and m have derivatives, where c is a positive constant independent of N.

In turn, the absolute value of the covariance in the second term in (37) is not greater than

$$(\operatorname{var} W_0 \operatorname{var} d_{N(n)}(x, X_0))^{\frac{1}{2}} \leq (EW_0^2 Ed_{N(n)}^2(x, X_0))^{\frac{1}{2}},$$

which is bounded at every $x > 0$, at which f is differentiable.

In light of this, at every point $x > 0$, at which both f and m are differentiable, the absolute value of $V_2(x)$ is not greater than

$$\frac{c}{n}\left[\left(\sum_{j=0}^{\infty}|k_j|\right)^2 + \sum_{s=0}^{\infty}\sum_{j=0}^{\infty}|k_{s+j}||k_j|\right],$$

where c is some constant dependent on x but independent of n.

Thus, (36) holds at every $x > 0$, at which both f and m are differentiable.

Since (24) holds, we have proved Theorem 3.

Theorem 3. Let f and m satisfy the assumptions of Theorem 1. Let the pulse response satisfy (30), (31) and (32). Let the number sequence $\{N(n)\}$ satisfy (19) and (20). Then,

$$\hat{m}(x) \to m(x) + \alpha \quad \text{as} \quad n \to \infty$$

in probability at every point $x > 0$, at which both f and m are differentiable.

Remark 4: Note that $\alpha = Em(X_0)\sum_{i=1}^{\infty} k_i$. Thus, the nonlinearity $m(x)$ can be recovered only up to the additive constant. If, however, one knows the value of the characteristic at one point, say $x = \theta$, then $\hat{m}(x) - \hat{m}(\theta) + m(\theta)$ converges to $m(x)$, $x > 0$.

Conditions (30), (31) and (32) restrict the class of admissible linear subsystems. One can, however, verify that every asymptotically stable system described by the state space equations fulfills all three restrictions.

It is clear that, using the results of Theorems 2 and 3, we can also obtain the global convergence of the estimate $\hat{m}(x)$ applied to the cascade model. The resulting rate of convergence (under the condition (27)) is the same as for the memoryless system, i.e., $O(n^{-2p/(2p+1)})$. Nevertheless, this is an asymptotic result, and a finite sample behavior of the estimate in those two cases can be drastically different.

In order to illustrate the aforementioned results, let us consider a simple simulation example.

Example 1. Let

$$\eta_n = q\eta_{n-1} + m(X_n),$$

$$Y_n = \eta_n + Z_n, \quad n = 0, \pm 1, \pm 2, \cdots$$

be the cascade model with the first-order stable autoregressive dynamic subsystem, i.e., $|q| < 1$.

Note that here, $k_n = q^n$, $n \geq 0$. Let Z_n be uniformly distributed over $[-0.1, 0.1]$ and X_n be uniform on $[0, 10]$. It is assumed that $m(x) = (5x^3 - 2x^2 + x)e^{-x}$, $x \geq 0$ and $q = 0.1$. The estimate performance is measured by the discrete version of our global criterion, i.e.,

$$\text{Error} = n^{-1} E\left\{ \sum_{j=1}^{n} |\tilde{m}(X_j) - m(X_j)|^2 \right\},$$

where $\tilde{m}(x) = \hat{m}(x) - \alpha$, and α is defined in Theorem 3. A simple algebra shows that $\alpha = q(1-q)^{-1} E\{m(X)\}$.

Figure 3 depicts the Error (calculated from 30 repetitions of the input-output data) versus n. For each n, the value of N minimizing the Error has been chosen. The selection of the optimal truncation point is an important issue for our estimates. Figure 4 shows the Error as a function of N for $n = 100$. The optimal N

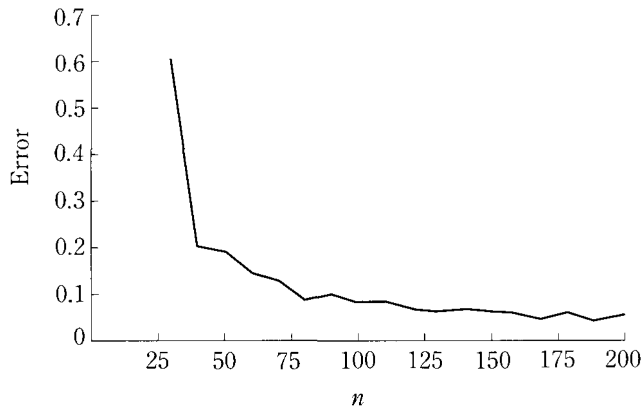

Fig. 3. Error versus n for the cascade system.

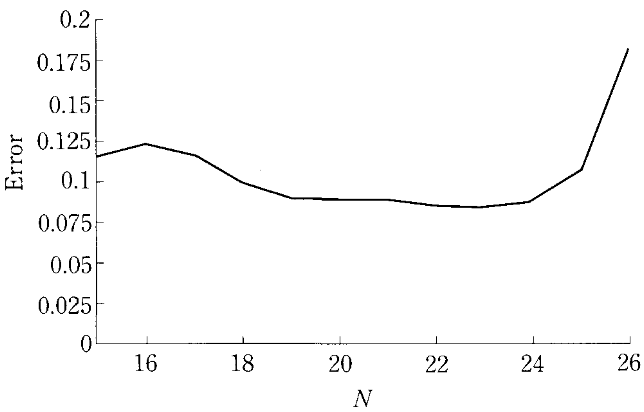

Fig. 4. Error versus N for the cascade system, $n = 100$, optimal $N = 23$.

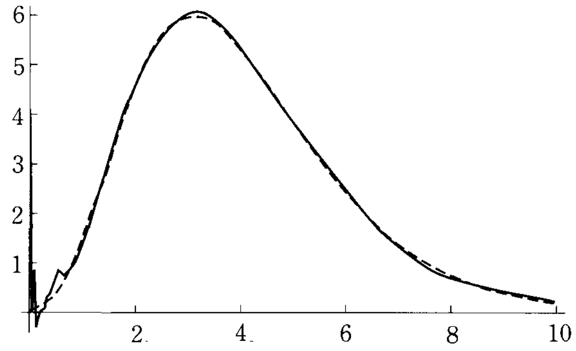

Fig. 5. The characteristic $m(x) = (5x^3 - 2x^2 + x)e^{-x}$, $x \geq 0$ (dashed line) and its estimate $\tilde{m}(x)$ (solid line) for the cascade system, $n = 100$, $N = 23$.

is equal to 23 with the corresponding Error $= 0.083$.

Figure 5, moreover, shows the plot of the characteristic $m(x)$ and its estimate $\tilde{m}(x)$ for $n = 100$, $N = 23$. The behavior of the estimate at the boundary is clearly revealed. The relatively large bias of the estimate for $x \to 0$ is due to the fact that the Laguerre series does not converge at $x = 0$ (see Remark 3 in Sec. 3). This problem could be easily fixed by using $\hat{m}(x) - \hat{m}(0) + m(0+)$ instead of $\hat{m}(x)$.

5. Recovering the Nonlinearity in Parallel Systems

A model of some physical interest is pictured in Fig. 6, where a nonlinear element $m(x)$ is connected in parallel with a linear dynamic system $\{k_i\}$ (Bendat, 1990, Sec. 5). That is,

$$Y_n = m(X_n) + \sum_{i=0} k_i X_{n-1} + Z_n . \tag{38}$$

We assume that $\{k_i\}$ satisfies the assumption (30), while for $m(x)$, the condition (11) is in force. $\{Z_n\}$ is a white noise with zero mean and finite variance. All these assumptions imply that $EY_n^2 < \infty$. The estimation of the linear subsystem can be carried out in a way similar to that in Sec. 4. In fact, one can note that

$$\operatorname{cov}(Y_{n+s}, X_n) = \tau^2 k_s, \quad s = 1, 2, \cdots,$$

where $\tau^2 = \operatorname{var}(X_n)$. Now k_s can be estimated by $\hat{\theta}_s / \hat{\tau}^2$, where $\hat{\theta}_s$ is defined (34) and $\hat{\tau}^2$ is an ordinary estimate of the variance.

Clearly, (38) can also be represented in a like that in (7),

$$Y_n = \mu(X_n) + \xi_n ,$$

where

$$\mu(X_n) = m(X_n) + X_n + \sum_{i=1} k_i E(X_0)$$

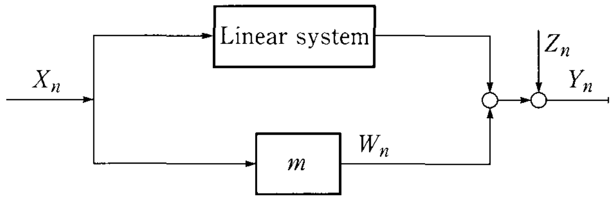

Fig. 6. Parallel nonlinear system.

and

$$\xi_n = \sum_{i=1}^{\infty} k_i[X_{n-i} - E(X_{n-i})] + Z_n.$$

The noise ξ_n, like in Sec. 4, is not white since ξ_i and ξ_j are correlated for $i \neq j$.

In order to recover $\mu(x)$, we apply estimate (18). Notation is the same as in Secs. 3 and 4. It is plain that

$$E\hat{g}(x) \to \mu(x)f(x) \quad \text{as} \quad N(n) \to \infty$$

at every point $x > 0$, at which both f and m are differentiable. In turn, $\operatorname{var}\hat{g}(x)$ can be written as in (35), where now Y_n is defined in (38).

Following the proof of Theorem 3, we have $V_1(x) \to 0$ as $N(n)/n \to 0$, whereas the covariance in $V_2(x)$ can be written in the following form:

$$\begin{aligned}
&\operatorname{cov}[Y_s d_{N(n)}(x, X_s), Y_0 d_{N(n)}(x, X_0)] \\
&\quad = k_s E d_{N(n)}(x, X_0) \operatorname{cov}[X_0, W_0 d_{N(n)}(x, X_0)] \\
&\quad + \sum^{s} \sum^{0} k_{s-i} k_{-j} \operatorname{cov}[X_i d_{N(n)}(x, X_s), X_j d_{N(n)}(x, X_0)],
\end{aligned}$$

where $W_0 = m(X_0)$. The first term in the above formula tends to $k_s f^2(x) m(x)[x - E(X_0)]$ as $N \to \infty$. The second term, in turn, can be analyzed in the same fashion as the term in (37) (just replace W_i with X_i). It should be noted that $E(X_0^2) < \infty$ is required here. Also, conditions (31) and (32) are needed. All these considerations yield the following result.

Theorem 4. Let all the assumptions of Theorem 3 be satisfied. Let, additionally, $\int_0^\infty x^2 f(x)dx < \infty$. Then,

$$\hat{m}(x) \to m(x) + x + \alpha \quad \text{as} \quad n \to \infty$$

in probability at every point $x > 0$, at which both f and m are differentiable.

Remark 5: Theorem 4 reveals that $\hat{m}(x) - x$ converges to $m(x) + \alpha$, where $\alpha = E(X_0)\sum_{i=1}^{\infty} k_i$. If, however, $m(\theta), \theta \geq 0$ is known, then $\hat{m}(x) + m(\theta) - \hat{m}(\theta) + \theta - x$ converges to $m(x)$. In particular, the knowledge of $m(\theta)$ for $\theta = 0$ can enforce our estimate to be convergent at the boundary point.

It is also clear that the global convergence can be established along with the rate as it has been done in Secs. 3 and 4. In order to complete our studies on the parallel model, let us consider the following example.

Example 2. Let us consider the parallel model,

$$\eta_n = q\eta_{n-1} + X_n ,$$

$$Y_n = \eta_n + m(X_n) + Z_n, \quad n = 0, \pm 1, \pm 2,$$

Here, Z_n, X_n, $m(x)$ and q are the same as in Example 1. Furthermore, let $\tilde{m}(x) = \hat{m}(x) - x - \alpha$, where a simple algebra shows that $\alpha = q(1-q)^{-1}E\{X\}$.

Figure 7 depicts the Error versus n, while Fig. 8 shows the Error versus N for $n = 100$. The optimal $N = 19$ with the corresponding Error $= 0.16$. Figure 9

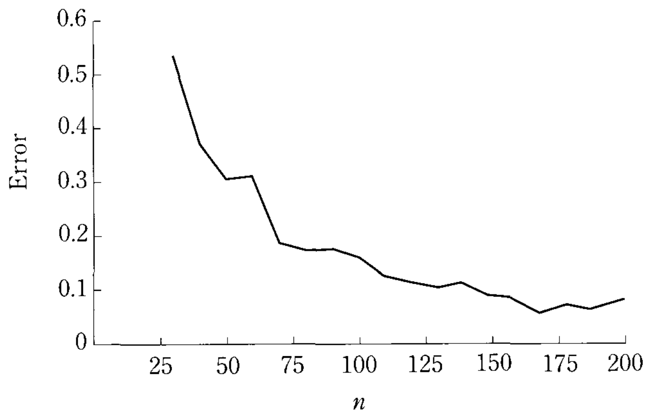

Fig. 7. Error versus n for the parallel system.

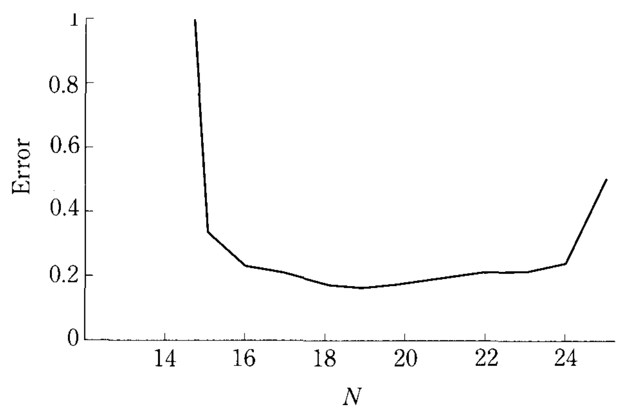

Fig. 8. Error versus N for the parallel system, $n = 100$, optimal $N = 19$.

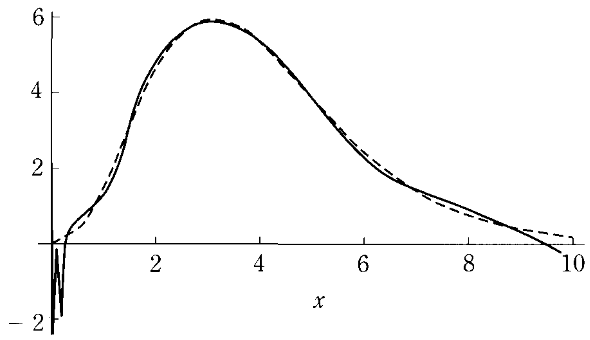

Fig. 9. The characteristic $m(x) = (5x^3 - 2x^2 + x)e^{-x}$, $x \geq 0$ (dashed line) and its estimate $\tilde{m}(x)$ (solid line) for the parallel system, $n = 100$, $N = 19$.

displays the plot of $m(x)$ and $\tilde{m}(x)$ for $n = 100$, $N = 19$. It should be noted that the Error for the parallel model is larger than that one for the cascade connection (see Example 1). This phenomenon can be explained by noting that $Y_n = m(X_n) + \zeta_n$ with $\zeta_n = \sum_{j=-\infty}^{n-1} q^{n-j} m(X_j) + Z_n$ for the cascade model and $\zeta_n = \sum_{j=-\infty}^{n-1} q^{n-j} X_j + Z_n$ for the parallel one. Then, a simple algebra shows that the variance of ζ_n for the parallel model is greater than the var ζ_n for the cascade one. Hence, the parallel connection has a smaller signal to noise ratio than the cascade one. It is worth noting that this need not always be the case, and the inverse situation can occur.

6. Concluding Remarks

It is clear that some other nonlinear models can be identified in a fashion similar to that used in Secs. 4 and 5. A simple extension could be carried out for models which are a combination of the cascade and parallel connections, e.g., a model with parallel linear and nonlinear cascade systems (see Bendat (1990), Sec. 7 for some examples of such models). Furthermore, a reverse regression $E\{X_n | Y_n = y\}$ can be employed for estimating a nonlinearity in a Wiener system, i.e., a system in which a linear dynamic part is followed by a nonlinear memoryless subsystem. We refer the reader to Greblicki (1992) for a detailed discussion of nonparametric identification algorithms of this system.

Yet, another interesting extension is to replace the memoryless nonlinear system in the aforementioned models by the non-instantaneous element. This is illustrated in Fig. 10, where the cascade model with a nonlinear element possessing a memory of the size 1 is used. Here, the input-output relation is given by

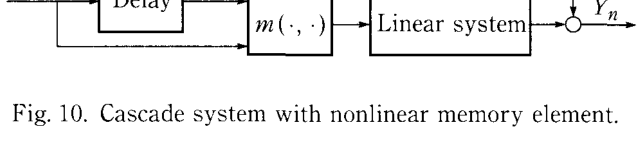

Fig. 10. Cascade system with nonlinear memory element.

Greblicki, W. and M. Pawlak (1985). Fourier and Hermite series estimates of regression functions. *Annals of the Inst. Statist. Math.*, **37**, 3, 443–459.

Greblicki, W. and M. Pawlak (1986). Identification of discrete Hammerstein systems using kernel regression estimates. *IEEE Trans. Automatic Control*, **AC-31**, 1, 74–77.

Greblicki, W. and M. Pawlak (1989). Nonparametric identification of Hammerstein systems. *IEEE Trans. Information Theory*, **IT-35**, 2, 409–418.

Greblicki, W. and M. Pawlak (1991). Nonparametric identification of a cascade nonlinear time series system. *Signal Processing*, **22**, 1, 61–75.

Greblicki, W. and M. Pawlak (1992). Nonparametric identification of a particular nonlinear time series system. *IEEE Trans. Signal Processing*, **40**, 4, 985–989.

Hall, P. (1980). Estimating a density on the positive half line by the method of orthogonal series. *Annals of the Inst. Satist. Math.*, **32**, 2, 351–362.

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.

Holland, F. (1969). An algebraic characterization of the Laguerre functions. *Quart. J. Math.*, **20**, 2, 193–194.

Hunter, I.W. and M.J. Korenberg (1986). The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biol. Cybern.*, **55**, 2, 135–144.

Ljung, L. (1987). *System Identification Theory for the User*. Prentice-Hall, Englewood Cliffs, N.J.

Marmarelis, P.Z. and V.Z. Marmarelis (1978). *Analysis of Physiological Systems: The White Noise Approach*. Plenum Press, N.Y.

Mohan, B.M. and K.B. Datta (1991). Identification of linear time-invariant single-input single-output continuous-time lumped parameter system via orthogonal functions. *Control–Theory and Advanced Technology* (*C-TAT*), **7**, 3, 387–406.

Narendra, K.S. and P.G. Gallman (1966). An iterative method for the identification of nonlinear systems using the Hammerstein model. *IEEE Trans. Automatic Control*, **AC-11**, 7, 546–550.

Palm, G. (1979). On representation and approximation of nonlinear systems. *Biol. Cybernetics*, **34**, 1, 49–52.

Pawlak, M. (1991). On the series expansion approach to the identification of Hammerstein systems. *IEEE Trans. Automatic Control*, **AC-36**, 6, 763–767.

Priestley, M.B. (1981). *Spectral Analysis and Time Series*. Academic Press, N.Y.

Priestley, M.B. (1988). *Non-linear and Non-stationary Time Series Analysis*. Academic Press, London.

Rugh, W.J. (1981). *Nonlinear System Theory — The Volterra/Wiener Approach*. John Hopkins University Press, Baltimore.

Sanberg, I.W. (1991). Approximation theorems for discrete-time systems. *IEEE Trans. Circuits and Systems*, **CAS-38**, 5, 564–566.

Sansone, G. (1991). *Orthogonal Functions*. Dover, N.Y.

Schuster, E.F. and S. Yakowitz (1985). Parametric/nonparametric mixture density estimation with application to flood frequency analysis. *Water Resources Bull.*, **21**, 5, 797–804.

Stone, C.T. (1985). Additive regression and other nonparametric models. *Ann. Statist.*, **13**, 2, 689–705.

Szegö, G. (1978). *Orthogonal Polynomials*. Amer. Math. Soc., Providence, Fourth Edition.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.

Uspensky, J.V. (1927). On the development of arbitrary functions in series of Hermite's and Laguerre's polynomials. *Annals. of Math.*, **28**, 4, 593–619.

Wahlberg, B. (1991). System identification using Laguerre models. *IEEE Trans. Automatic Control*, **AC-36**, 5, 551–562.

Appendix A

In Szegö (1978, p. 246), we have:

Theorem 1A. Let φ be a Lebesgue measurable function in $[0,\infty)$ and let the following integrals exist:

$$\int_0^1 |\varphi(x)|dx, \quad \int_0^1 x^{-\frac{1}{4}} |\varphi(x)|dx. \tag{A.1}$$

Let, moreover,

$$n^{\frac{1}{2}} \int_n^\infty e^{-\frac{x}{2}} x^{-\frac{13}{12}} |\varphi(x)|dx \to 0 \quad \text{as} \quad n \to \infty. \tag{A.2}$$

Denote by $S_n(x)$ the nth partial sum of the expansion of φ in the Laguerre series. Let $\tilde{S}_n(x)$ denote the nth partial sum of the expansion of $\varphi(x^2)$ in the trigonometric series in interval $[\sqrt{x}-\delta, \sqrt{x}+\delta]$, $\delta > 0$. Then,

$$\lim_{n\to\infty}[S_n(x) - \tilde{S}_n(\sqrt{x})] = 0.$$

As a consequence of Theorem 1A, the asymptotic behavior of $S_n(x)$ at point x is the same as the behavior of the trigonometric series of $\varphi(y^2)$ at the point $y = \sqrt{x}$. Since the trigonometric expansion of any integrable function converges to the function at every point at which the function is differentiable, we get:

Corollary. Let φ be a Lebesgue measurable function in $[0,\infty)$, and let all the integrals in (A.1) exist. Let (A.2) hold. Then,

$$\lim_{n\to\infty} S_n(x) = \varphi(x)$$

at every differentiability point of φ, where $S_n(x)$ is the nth partial sum of the expansion of φ in the Laguerre series.

Appendix B

Let us now verify that g satisfies restrictions (A.1) and (A.2) of Theorem 1A in Appendix A. The conditions (8), (9) and (11) will be employed.

First, observe that

$$\int_0^1 |m(x)| f(x)dx \le \left(\int_0^\infty m^2(x) f(x)dx\right)^{\frac{1}{2}},$$

which is finite due to (11).

Furthermore,

$$\int_0^1 x^{-\frac{1}{4}} |m(x)| f(x)dx \le \left(\int_0^1 x^{-\frac{1}{2}} f(x)dx\right)^{\frac{1}{2}} \left(\int_0^1 m^2(x) f(x)dx\right)^{\frac{1}{2}}.$$

Clearly, using (8), we have

$$\begin{aligned}\int_0^1 x^{-\frac{1}{2}} f(x)dx &\le M\int_0^1 x^{\alpha-\frac{1}{2}}dx + f(0)\int_0^1 x^{-\frac{1}{2}}dx \\ &\le \frac{M}{\alpha+1/2} + 2f(0) < \infty.\end{aligned}$$

The integral in (A.2) does not exceed

$$\left(\int_n^\infty e^{-x}x^{-\frac{13}{6}}f(x)dx\right)^{\frac{1}{2}}\left(\int_0^\infty m^2(x)f(x)dx\right)^{\frac{1}{2}}.$$

It is clear that due to (9) we have $n\int_n^\infty e^{-x}x^{-13/6}f(x)dx \to 0$ as $n \to \infty$.

Therefore,

$$E\hat{g}(x) \to g(x) \quad \text{as} \quad n \to \infty$$

at every point $x > 0$, at which both f and m are differentiable.

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

Since 1966, he has been with the Technical University of Wrocław, Wrocław, Poland. He currently holds the position of Professor in the Institute of Engineering Cybernetics, Technical University of Wrocław. In the academic year 1986–1987, he was a Visiting Professor at the University of Manitoba, Winnipeg, Manitoba, Canada.

His research interests include nonparametric methods of mathematical statistics and their applications to system identification and classification. He teaches courses in control theory and system identification.

Miroslaw Pawlak was born in Poland on July 21, 1954. He received the M.Sc. and Ph.D. degrees in computer engineering from the Technical University of Wrocław, Poland, in 1978 and 1982, respectively.

In 1982, he joined the Department of Technical Cybernetics, Technical University of Wrocław, as an Assistant Professor. From 1984 to 1985, he was a visiting Research Associate at the Department of Computer Science, Concordia University, Montreal. Since 1985, he has been with the University of Manitoba, Winnipeg, Canada, where he is currently an Associate Professor in the Department of Electrical and Computer Engineering.

During 1992, he was at the Technical University of Wrocław, Poland and at the University of Ulm, Germany as an Alexander von Humboldt Foundation fellow. He has also held visiting appointments at the Center of Multivariate Analysis, University of Pittsburgh, and at the Center for Stochastic Processes, University of North Carolina, Chapel Hill.

His research interests are in pattern recognition, statistical signal and image processing, non-linear system identification and nonparametric curve estimation.

mita

PUBLISHED BY MITA PRESS

TOKYO, JAPAN

1 9 9 4