Nonparametric Identification of Wiener Systems by Orthogonal Series

Włodzimierz Greblicki, *Member, IEEE*

Abstract— **A Wiener system, i.e. a system comprising a linear dynamic and a nonlinear memoryless subsystems connected in a cascade, is identified. Both the input signal and disturbance are random, white, and Gaussian. The unknown nonlinear characteristic is strictly monotonous and differentiable and, therefore, the problem of its recovering from input-output observations of the whole system is nonparametric. It is shown that the inverse of the characteristic is a regression function and, next, a class of orthogonal series nonparametric estimates recovering the regression is proposed and analyzed. The estimates apply the trigonometric, Legendre, and Hermite orthogonal functions. Pointwise consistency of all the algorithms is shown. Under some additional smoothness restrictions, the rates of their convergence are examined and compared. An algorithm to identify the impulse response of the linear subsystem is proposed.**

I. Introduction

The identification of nonlinear dynamic systems has been attracting attention of many authors for a long time. The block oriented approach is based on the assumption that the system consists of relatively simple elements which are identified from input-output observations of the whole system. The elements are usually linear dynamic or nonlinear memoryless. Mostly two types of such systems have been examined, i.e., the Hammerstein and Wiener ones. In the first, a nonlinear memoryless element is followed by a linear dynamic one, while in the other, the same subsystems are connected in the reverse order.

First results concerning the identification of Hammerstein systems can be traced back to Narendra and Gallman [32] or even earlier, see also later Billings and Fakhouri [4]-[6], Brillinger [7], Chang and Luus [9], Gallman [12], as well as Bendat [2], and Billings [3]. Wiener systems are much tougher to analyze and fewer papers on their identification can be found in literature, Bars *et al.* [1], Billings and Fakhouri [4]-[6], Hasiewicz [25], Hunter and Korenberg [28], Korenberg and Hunter [29], Westwick and Kearney [41], see also Bendat [2]. Authors mentioned above have proposed a number of algorithms, however, under the hypothesis that the nonlinearity in both the systems is a polynomial which order is known. Therefore, the parametric class of characteristics admitted by them is very narrow and contains no nontrivial bounded nor discontinuous functions, i.e. nonlinearities often encountered in applications. This restriction imposed on the nonlinear subsystem in both Hammerstein and Wiener systems is clearly an obvious and great disadvantage.

In an attempt to overcome that drawback of parametric methods, Greblicki and Pawlak [17]-[18] have proposed to apply nonparametric inference methods. In this way, they have significantly enlarged the class of nonlinearities in the Hammerstein system that can be recovered from input-output observations to all Borel measurable functions satisfying some extremely mild conditions. Their class of all possible characteristics is so wide that cannot be represented in a parametric form. So, their approach is nonparametric. The idea has next advanced in Greblicki [14], Greblicki and Pawlak [17]-[23], Krzyżak [31], Pawlak [33], Pawlak and Greblicki [34]. An effort to apply the nonparametric approach to the identification of a class of dynamic systems has been also made by Georgiev [13]. The nonparametric approach is an important proposition, since it significantly relaxes restrictions imposed on the identified system and makes the problem much closer to those encountered in real situations.

As far as the Wiener system is concerned, Greblicki [16] has recently introduced a kernel nonparametric algorithm to recover the nonlinearity and examined both its consistency and the speed of convergence. In this paper, we propose a new class of algorithms employing the trigonometric, Legendre, and Hermite orthogonal series. We show that all the algorithms converge to the unknown nonlinearity and give rates of their convergence. Compared with the kernel algorithm, they have some computational advantages. For a broad survey of nonparametric statistical methods, we refer the reader to Härdle [27] and Prakasa Rao [35].

The block oriented approach is elegant from the theoretical viewpoint. It is also an interesting proposition for the user working in such different and distant areas as biology, industry, psychology or sociology. Bars *et al.* [1] and Eskinat *et al.* [11] have applied a cascade model to identify a distillation column. The model has been also used to recover a nonlinearity in a heat exchanger, Eskinat *et al.* [11]. Parts of nervous systems are also often described by means offered by the block oriented approach. For example, den Brinker [8] has used the method to model the human transient visual system, Huebner *et al.* [26] in studies on eye movements, Emerson *et a*l. [10] have used a cascade model to identify nonlinearities in a visual cortex. Examples of other applications can be found in references given in Hunter and Korenberg [28], and Korenberg and Hunter [29]. Above examples show apparent need for algorithms able to recover nonlinearities in systems of various kinds. Many authors propose, however, algorithms without making necessary theoretical studies. It is caused by the fact that nonlinear dynamic systems are very difficult to analyze.

Responding to this demand, we give a new class of algorithms to identify nonlinear dynamic cascade systems

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

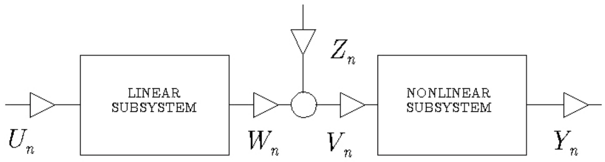

Fig. 1. The identified Wiener system.

referred to as Wiener systems. Contrary to many other authors, we prove fundamental properties of proposed algorithms. We show that our algorithms converge to unknown characteristics and, under some smoothness restrictions, examine their convergence rates. The paper is theoretical, nevertheless, results of an illustrative numerical simulation example are also presented.

II. The Identification Problem

We identify a Wiener system, i.e. a system consisting of two subsystems connected in a cascade. The linear dynamic subsystem is followed by a nonlinear memoryless one, Fig.1. The whole system is driven by a sequence $\{U_n; n = \cdots, -1, 0, 1, \cdots\}$ of independent zero-mean Gaussian random variables. Its linear part is described by state space equations

$$\left.\begin{array}{c} X_{n+1} = AX_n + bU_n \\ W_n = c^T X_n \end{array}\right\}. \tag{2.1}$$

Vectors b, c and the matrix A are all unknown. The X_n and W_n are the state vector and the output of the subsystem at time n, respectively. The subsystem is asymptotically stable, i.e.

$$\text{all eigenvalues of } A \text{ lie in the unit circle.} \tag{2.2}$$

Random variable sequences $\{X_n; n = \cdots, -1, 0, 1, \cdots\}$ and $\{W_n; n = \cdots, -1, 0, 1, \cdots\}$ are clearly stationary and Gaussian.

The output of the first subsystem is disturbed by a stationary white Gaussian random noise $\{Z_n; n = \cdots, -1, 0, 1, \cdots\}$ independent of $\{U_n\}$, i.e.,

$$V_n = W_n + Z_n.$$

The characteristic of the second subsystem is denoted by m. It means that

$$Y_n = m(V_n). \tag{2.3}$$

The m is a function defined in the entire real line R. In the paper,

$$m \text{ is differentiable and strictly monotonous} \tag{2.4}$$

and

$$\sup_{v \in R} \left| \frac{d}{dv} m(v) \right| \leq \alpha, \tag{2.5}$$

some $\alpha > 0$.

We assume that m is unknown and our goal is to estimate it from observations (U_i, Y_i)'s, $i = 0, 1, 2, \cdots$. The family of all characteristics satisfying (2.4) and (2.5) is clearly so wide that can not be represented in a parametric form. In the light of this, the problem of recovering m is nonparametric.

By M, we denote the image of R under the mapping m. Observe that M is one of the following open sets: $(-\infty, \infty)$, $(-\infty, b)$, (a, ∞), or (a, b). In the first case, m is not bounded, in the second bounded from above and $b = \sup_{v \in R} m(v)$. In the third, m is bounded from below and $a = \inf_{v \in R} m(v)$, while in the last case from both sides. Since m is a one-to-one mapping in the product $R \times M$, we can define

$$m^{-1}(y) = \begin{cases} \text{the inverse of } m \text{ at } y, & \text{for } y \in R \\ 0, & \text{otherwise.} \end{cases}$$

For convenience, we call m^{-1} the inverse of m. In the paper, we propose an algorithm to recover m^{-1} and then show how to estimate m.

In order to introduce the algorithm, we begin with the observation that the pair (U_{n-1}, V_n) has a zero-mean Gaussian distribution with marginal variance σ_U^2 and σ_V^2, and the correlation coefficient $\rho = \lambda_1 \sigma_U / \sigma_V$ where $\sigma_V^2 = \sigma_V^2 \sum_{i=1}^{\infty} \lambda_i^2 + \sigma_Z^2$. By σ_U^2 σ_V^2 and σ_Z^2 we denote variance of U_n, V_n, and Z_n, respectively, while $\lambda_i = c^T A^{i-1} b$. Obviously $\{\lambda_n; n = 0, 1, 2, \cdots\}$, where $\lambda_0 = 0$, and where $\lambda_i, i = 1, 2, \cdots$, is defined above, is the impulse response of the dynamic subsystem. Therefore, $E\{U_{n-1} | V_n = v\} = \lambda v$, where $\lambda = \lambda_1 \sigma_U^2 / \sigma_V^2$. Recalling (2.3), we get finally

$$E\{U_{n-1} | Y_n = y\} = \lambda m^{-1}(y), \tag{2.6}$$

where λ is unknown. Having observed that $\lambda m^{-1}(y)$ can be expressed as a regression function, we propose to estimate the regression, i.e., $E\{U_{n-1} | Y_n = y\}$ from input-output observations. Denoted by $\mu_n(y)$, our estimate of $\lambda m^{-1}(y)$ employs the trigonometric, Legendre, and Hermite orthogonal series. In this way, we can recover the inverse of the characteristic up to some unknown multiplicative constant λ.

The fact that the constant cannot be estimated is an obvious consequence of the cascade structure of the system. Observe that, having input-output observations, we cannot distinguish a system consisting of a linear subsystem with an impulse response $\{\lambda_n\}$, and nonlinearity m from a system having a linear part $\{\alpha \lambda_n\}$, and a characteristic $(1/\alpha)m$, any $\alpha \neq 0$.

Next, using the notion of a pseudoinversion μ_n^+ of defined in the following way:

$$\mu_n^+(v) = \begin{cases} \text{any } y & \text{, for which } \mu_n(y) = v \\ 0 & \text{, otherwise} \end{cases}$$

one can take $\mu_n^+(v)$ as an estimate of the inverse of $\lambda m^{-1}(y)$, i.e., of $m(v/\lambda)$. The usage of a pseudoinversion is caused by the fact that the estimate itself may not be invertible. Therefore, because of the cascade structure of the system, we can recover m^{-1} only up to the multiplicative factor λ and, as a consequence, m up to the dilation constant $1/\lambda$.

Since m is a differentiable one-to-one mapping, a probability density f of Y_n exists and equals

$$f(y) = \begin{cases} f_V(m^{-1}(y)) \left| \left[m^{-1}(y)\right]' \right| & \text{, for } y \in M \\ 0 & \text{, otherwise,} \end{cases}$$

where f_V is the density of V_n, i.e., a normal density with zero mean and variance σ_V^2. Observing that $\left[m^{-1}(y)\right]' = 1/m'(y)$ at $v = m^{-1}(y)$, we verify

$$f(y) = \begin{cases} f_V(v)/\left|m'(y)\right| \text{ at } v = m^{-1}(y) & \text{, for } y \in M \\ 0 & \text{, otherwise.} \end{cases} \tag{2.7}$$

This, along with (2.5), yield

Lemma 1: Let (2.2) hold. Let m satisfy (2.4), and (2.5). Then $f(y) > 0$ at every $y \in M$.

Throughout the paper, we denote $g(y) = \lambda m^{-1}(y) f(y)$. Observing that $[1/m'(v)]' = \left[m^{-1}(y)\right]''$ at $v = m^{-1}(y)$, we obtain the following.

Lemma 2: Let (2.2) hold. Let m satisfy (2.4), and (2.5). Then both $f'(y)$ and $g'(y)$ exist at every $y \in M$ at which $\left[m^{-1}(y)\right]''$ exists, i.e., is finite.

III. The Algorithm

The algorithm proposed in the paper applies orthogonal series. Suppose that a complex-valued series $\{\varphi_k; k = 0, 1, 2, \cdots\}$ is orthonormal in a set D, i.e. that

$$\int_D \varphi_i(y)\varphi_j^*(y)dy = \begin{cases} 1, & \text{for } i = j \\ 0, & \text{otherwise.} \end{cases}$$

The idea of the algorithm is based on the fact that, owing to (2.6), $\lambda m^{-1}(y) = g(y)/f(y)$, where $g(y) = E\{U_{n-1} | Y_n = y\} f(y)$. Defining

$$\begin{aligned} a_k &= \int_D \varphi_k^*(y) E\left\{U_{n-1} | Y_n = y\right\} f(y)dy \\ &= E\left\{U_{n-1}\varphi_k^*(y) I_D(Y_n)\right\} \end{aligned}$$

and

$$b_k = \int_D \varphi_k^*(y) f(y)dy = E\left\{\varphi_k^*(y) I_D(Y_n)\right\}$$

where I_D is a characteristic function of $_D$, we can expand g and f in the orthogonal series, i.e.,

$$g(y) \sim \sum_{k=0}^{\infty} a_k \varphi_k(y) \tag{3.1}$$

and

$$f(y) \sim \sum_{k=0}^{\infty} b_k \varphi_k(y),$$

respectively. Coefficients a_k's and b_k's of the above expansions can be easily estimated in the following way

$$a_{kn} = n^{-1} \sum_{i=1}^{n} U_{n-1} \varphi_k^*(y) I_D(Y_i)$$

and

$$b_{kn} = n^{-1} \sum_{i=1}^{n} \varphi_k^*(y) I_D(Y_i)$$

respectively. As an estimate of $\lambda m^{-1}(y)$, we propose

$$\mu_n(y) = \frac{\sum_{k=0}^{N(n)} a_{kn}\varphi_k(y)}{\sum_{k=0}^{N(n)} b_{kn}\varphi_k(y)},$$

where $\{N(n); n = 1, 2, 3, \cdots\}$ is a sequence of integers. In the definition, $0/0$ is treated as an arbitrary constant. We show that, for a suitably selected $\{N(n)\}$,

$$\mu_n(y) \xrightarrow{n} \lambda m^{-1}(y) \text{ in probability} \tag{3.3}$$

at some points $y \in R$ which will be specified later. Denoting

$$K_N(x, y) = \sum_{k=0}^{N} \varphi_k^*(x)\varphi_k(y), \tag{3.4}$$

where K is the kernel of the series, we can rewrite the estimate in the following alternative form:

$$\mu_n(y) = \frac{n^{-1}\sum_{i=1}^{n} U_{i-1} K_{N(n)}(Y_i, y) I_D(Y_i)}{n^{-1}\sum_{i=1}^{n} K_{N(n)}(Y_i, y) I_D(Y_i)}. \tag{3.2a}$$

We denote the numerator and the denumerator of the estimate by $g_n(y)$ and $f_n(y)$, respectively.

Various orthogonal systems lead to various versions of both (3.2) and (3.2a). In the paper, we apply the trigonometric, Legendre, and Hermite series. For the consecutive series, D is $(-\pi, \pi)$, $(-1, 1)$, and R, respectively.

For convenience, we shall write N for $N(n)$. The integer sequence satisfies the following obvious condition:

$$N \to \infty \text{ as } n \to \infty. \tag{3.5}$$

Additional restrictions imposed on the sequence depend on the applied orthogonal series and will be given later.

We shall say that $\mu_n(y) \to \lambda m^{-1}(y)$ as $n \to \infty$ in probability at a rate $O(a_n)$, i.e., that in-probability convergence (3.3) is of order $O(a_n)$, if

$$(\gamma_n/a_n) \left|\mu_n(y) - \lambda m^{-1}(y)\right| \xrightarrow{n} 0 \text{ in probability,}$$

every number sequence $\{\gamma_n\}$ convergent to zero.

From the statistical viewpoint, our algorithm recovering the nonlinearity just estimates a regression function and makes it with the help of various orthogonal series. We want to mention that the idea of the nonparametric regression estimation by orthogonal series has been present in literature, mainly statistical, for some time, e.g. Greblicki [14], Greblicki and Pawlak [17], Greblicki *et al.* [24], Rafajłowicz [36]. Rudiment results concerning the application of such series in nonparametric inference have been, however, presented earlier by e.g. Kronmal and Tarter [30], Schwartz [38], Walter [40] in the context of the probability density estimation.

Apart from two lemmas given in Section II, the others are gathered in two appendixes. Those in Appendix A deal with the Wiener system, while Appendix B is of general character.

In the paper we estimate a regression $E\{U_{n-1}|Y_n = y\}$ and next, assuming that λ_1 is different from zero, recover $\lambda_1 m^{-1}(y)$. In general, it can be easily verified that estimating a regression $E\{U_{n-k}|Y_n = y\}$, $k \geq 1$, we can recover $\lambda(k)m^{-1}(y)$, where $\lambda(k) = \lambda_k \sigma_U^2/\sigma_V^2$.

IV. The Trigonometric Series Algorithm

In this section, we estimate λm^{-1} in the interval $(-\pi,\pi)$. It is very well known that the trigonometric system $\{(2\pi)^{-1/2}e^{iky}; k = 0, \pm 1, \pm 2, \cdots\}$ is orthonormal in the interval. Therefore, in this section, $D = (-\pi,\pi)$ and $\varphi_k(y) = (2\pi)^{-1/2}e^{iky}$, where $k = 0, \pm 1, \pm 2, \cdots$. Contrary to the previous section, now the index k takes both nonnegative as well as negative integer values. We shall not rewrite the estimate but we want, however, to stress that, for the estimate with this particular series, the index k in the sums in both the numerator and denumerator in definition (3.2) of the estimate do not vary from 0 till k but from $-k$ through k. In other words, $|k|$ varies from zero till N.

Clearly,

$$K_N(x,y) = (1/2\pi)\sum_{k=0}^{N} \cos k(x-y)$$

which obviously equals $(1/\pi)D_N(x-y)$, where $D_N(t) = [\sin(N+1/2)t]/2\sin(t/2)$ is the Nth Dirichlet kernel.

Theorem 1: Let (2.2) hold. Let m satisfy (2.4), and (2.5). Let, in addition to (3.5),

$$N^3/n \to \infty \text{ as } n \to \infty. \tag{4.1}$$

Then, for the estimate with the trigonometric series, (3.3) holds at every $y \in (-\pi,\pi) \cap M$ at which $[m^{-1}(y)]''$ exists.

The set of points at which the algorithm converges can be commented on in the following way. For $(-\pi,\pi) \subseteq M$, (3.3) holds for every $y \in (-\pi,\pi)$ at which $[m^{-1}(y)]''$ exists. For a bounded m and $(a,b) \subseteq (-\pi,\pi)$, (3.3) holds at every $y \in (a,b)$ at which $[m^{-1}(y)]''$ exists.

There obviously exist number sequences satisfying appropriate conditions of Theorem 1. For example, we can select $N \sim n^{\alpha}$, where $0 < \alpha < 1/3$.

Proof of Theorem 1: We have

$$Eg_n(y) = \sum_{|k|\leq N} a_k \varphi_k(y) I_D(y)$$

which is simply the Nth partial sum of the expansion of g on $(-\pi,\pi)$ in the trigonometric series which, due to Dini's theorem, [37, pp. 65-66], converges to $g(y)$ at every point at which g is differentiable. Hence,

$$Eg_n(y) \to g(y) \text{ as } n \to \infty$$

at every $y \in (-\pi,\pi)$ at which g is differentiable.

In turn, using (3.2a), we get

$$\begin{aligned} \operatorname{var}[g_n(y)] &= n^{-1}\operatorname{var}[U_0 K_N(Y_1,y) I_D(Y_1)] \\ &+ n^{-2}\sum_{j=1}^{n-1}(n-j)\operatorname{cov}[U_j K_N(Y_{j+1},y) I_D(Y_{j+1}), \\ &\qquad U_j K_N(Y_1,y) I_D(Y_1)] \end{aligned}$$

The first term in the above expression does not exceed

$$\begin{aligned} & n^{-1}E\left\{U_0^2 K_N^2(Y_1,y) I_D(Y_1)\right\} \\ & \leq n^{-1}E\{U_0^2\} \max_{x,y\in D} K_N^2(x,y) = O(N^2/n). \end{aligned}$$

Recalling (2.5) and using Lemmas A3 and B2 in Appendixes A and B, respectively, we find the covariance under the sum in the second term is not greater than $c(y)\left\|A^j\right\| N^3$, where c is some function independent of j, n, and N. This and the fact that all eigenvalues of A are in the unit circle imply that the term is of order $O(N^3/n)$. Therefore,

$$\operatorname{var}[g_n(y)] = O(N^3/n) \tag{4.2}$$

every $y \in (-\pi,\pi)$. In this way, we have shown that $g_n(y)$ converges to $g(y)$ in probability at every point $y \in (-\pi,\pi)$ at which g is differentiable.

Using similar arguments, one can verify that $f_n(y)$ converges to $f(y)$ as n tends to infinity in probability at every point $y \in (-\pi,\pi)$ at which f has a derivative. Thus, (3.3) holds at every $y \in (-\pi,\pi) \cap M$, at which both g and f are differentiable and $f(y) > 0$. Applying now Lemmas 1 and 2, we complete the proof of the theorem. ■

Imposing some additional smoothness restrictions on f and m, we shall now examine the rate of convergence in (3.3). Let m be bounded and let $(a,b) \subseteq (-\pi,\pi)$. For convenience, denote $\varphi(v) = \exp(-v^2/2\sigma_V^2)$ and then suppose that functions $\varphi(v)/m'(v)$, $v\varphi(v)/m'(v)$, and $\varphi(v)m''(v)/[m'(v)]^3$ converge to zero as $|v|$ tends to infinity. Owing to that, by virtue of Lemma A1 in Appendix A, $f(y)$ and $f'(y)$ converge to zero as y tends to a or b. The same clearly holds as y tends to $-\pi$ or π. Suppose, moreover, that functions $v^2\varphi(v)/m'(v)$, $v\varphi(v)m''(v)/[m'(v)]^3$, and $\varphi(v)/[m'(v)]^2$, also converge to zero as $|v|$ tends to infinity. Applying now Lemma A2 in Appendix A, we find both $g(y)$ and $g'(y)$ converging to zero as y tends to $-\pi$ or π. In passing, observe that all the above conditions concerning m, however somewhat complicated, are not very restrictive and are satisfied, e.g., by $m(v) = v/(|v|+\alpha)$, any $\alpha > 0$. Integrating $\int_{-\pi}^{\pi} e^{iky}g(y)dy$ and $\int_{-\pi}^{\pi} e^{iky}f(y)dy$ by parts, we find $a_k = -i\alpha_k/k$ and $b_k = -i\beta_k/k$, where α_k and β_k are Fourier coefficients of g' and f', respectively. In general, assuming that $g^{(s)}$ and $f^{(s)}$, $s = 1, 2, \cdots, p$, exist and equal zero at points a and b, we can verify that $a_k = (-i)^p\alpha_{kp}/k^p$ and $b_k = (-i)^p\beta_{kp}/k^p$, where α_{kp} and

β_{kp} are Fourier coefficients of $g^{(p)}$ and $f^{(p)}$, respectively. Hence, assuming that $g^{(p)}$ is square integrable, we get

$$|Eg_n(y) - g(y)| = \left|\sum_{|k|>N} a_k e^{-iky}\right| \le \sum_{|k|>N} |a_k|$$

$$\le \left|\sum_{|k|>N} \alpha_{kp}^2\right|^{1/2} \left|\sum_{|k|>N} k^{-2p}\right|^{1/2}$$

which, for $p > 1$, is of order $O(N^{-p+1/2})$. This and (4.2) yield $E(g_n(y) - g(y))^2 = O(N^{-2p+1}) + O(N^3/n)$, which obviously equals $O(n^{-(2p-1)/(2p+2)})$, for $N \sim n^{1/(2p+2)}$. Since the error concerning fn vanishes at the same rate, an application of Lemma B1 in Appendix B gives finally

$$P\left\{\mu_n(y) - \lambda m^{-1}(y) > \varepsilon\lambda m^{-1}(y)\right\}$$
$$= O(n^{-(2p-1)/(2p+2)})$$

any $\varepsilon > 0$, and

$$\left|\mu_n(y) - \lambda m^{-1}(y)\right| = O(n^{-(2p-1)/(4p+4)})$$

in probability, every $y \in (-\pi, \pi)$. For example, for $p = 2$, the in-probability convergence rate equals $O(n^{-1/4})$.

V. The Legendre Series Algorithm

In this section, D is $(-1, 1)$ and we apply the Legendre system of orthogonal polynomials. The kth Legendre polynomial P_k is defined by the following Rodrigues formula: $P_k(y) = (1/2^k k!)(d^k/dy^k)(y^2 - 1)^k$, $k = 0, 1, 2, \cdots$. One can easily verify that $P_0(y) = 1$, $P_1(y) = y$, $P_2(y) = 3y^2/2 - 1/2$, $P_3(y) = 5y^3/2 - 3y/2$, and so on. The system $\{p_k; k = 0, 1, 2, \cdots\}$, where $p_k(y) = (k + 1/2)^{1/2} P_k(y)$, is orthonormal in $(-1, 1)$, [37, p. 190]. Thus, in this section $\varphi_k = p_k$, $k = 0, 1, 2, \cdots$.

Theorem 2: Let (2.2) hold. Let m satisfy (2.4), and (2.5). Let the algorithm apply the Legendre orthogonal series. If, in addition to (3.5),

$$N^{9/2}/n \to 0 \text{ as } n \to \infty \tag{5.1}$$

then (3.3) holds at every $y \in (-1, 1) \cap M$ at which $[m^{-1}(y)]''$ exists. Let, additionally, m be bounded, i.e., let $M = (a, b)$, some a and b. Let, moreover, $[a, b] \subset (-1, 1)$. If (3.5) is satisfied and

$$N^3/n \to 0 \text{ as } n \to \infty$$

then (3.3) holds at every $y \in M$ at which $[m^{-1}(y)]''$ exists.

In the first part of the theorem m can be bounded or not. For m bounded and $[a, b]$ being a subinterval of $(-1, 1)$, the second part of Theorem 2 applies and significantly relaxes the condition imposed on the number sequence $\{N(n)\}$. As we later prove, the convergence rate is better in that case.

Proof of Theorem 2: First of all, observe that due to Lemma 2, g is differentiable at every point y at which m^{-1} has a derivative. Applying now Hobson's theorem on pointwise convergence of Legendre expansions, [37, pp. 234-235], we find the partial sum of the expansion of g in the Legendre series to converge to $g(y)$ at every $y \in (-1, 1)$ at which m^{-1} is differentiable. Next, using Lemmas 2 and B3, and proceeding like in the proof of- Theorem 1, we verify that $\text{var}[g_n(y)]$ equals $O(N^{9/2}/n)$ or $O(N^3/n)$, for the first or the second case, respectively. Since, similar property can be shown for f, the proof has been completed. ■

We shall now examine the convergence rate of the algorithm. Suppose that both f and g have two derivatives. Invoking Jackson's theorem, [37, p. 206], we obtain

$$|Eg_n(y) - g(y)| = \left|\sum_{k=N+1}^{\infty} a_k p_k(y)\right|$$

which is of order $O(N^{-1/2})$, every $y \in (-1, 1)$. Recalling that variance is of order $O(N^{9/2}/n)$, we find the estimate to converge in probability at the rate $O(n^{-1/11})$, provided that $N \sim n^{2/11}$. In turn, assuming that m is bounded and that $[a, b] \subset (-1, 1)$ and applying again Jackson's theorem, [37, p. 20], we find $|Eg_n(y) - g(y)| = O(N^{-1})$, every $y \in [a, b]$. Now, since variance vanishes as fast as $O(N^4/n)$, the in-probability convergence rate is $O(n^{-1/5})$, for $N \sim n^{1/5}$. The rate is very close to $O(n^{-1/4})$, i.e., that at which the trigonometric series estimate converges for twice differentiable m.

VI. The Hermite Series Algorithm

In this section, we estimate λm^{-1} in the entire real line, i.e,. we have $D = (-\infty, \infty)$, and do it with the help of the Hermite orthogonal series, i.e., a series $\{h_k; k = 0, 1, 2, \cdots\}$ in which

$$h_k(y) = (2^k k! \pi^{1/2})^{-1/2} H_k(y) e^{-y^2/2},$$

where

$$H_k(y) = e^{y^2}(d^k/dy^k)e^{-y^2}$$

is the kth Hermite polynomial. One can easily verify that $H_0(y) = 1$, $H_1(y) = -2y$, $H_2(y) = 4y^2 - 2$, $H_3(y) = -8y^3 + 12y$, and so on. The series is orthonormal in the entire real line, Sansone [37] or Szegö [39]. In other words, in this section $\varphi k = h_k$, $k = 0, 1, 2, \cdots$.

Theorem 3: Let (2.2) hold. Let m satisfy (2.4), and (2.5). Let the algorithm use the Hermite orthogonal series. If (3.5) is satisfied and

$$N^{5/3}/n \to 0 \text{ as } n \to \infty \tag{6.1}$$

then (3.3) holds at every $y \in M$ at which $[m^{-1}(y)]''$ exists. Let, moreover, m be bounded, i.e. let M be finite. If (3.5) is satisfied and

$$N^{3/2}/n \to 0 \text{ as } n \to \infty$$

then (3.3) holds at every $y \in M$ at which $[m^{-1}(y)]''$ exists.

Proof: The proof follows the lines similar to that of Theorem 1. As far as convergence of the partial expansion of g in the Hermite series, according to the equiconvergence theorem in Szegö [39, p. 247], the Hermite and trigonometric expansions of g, taken in an arbitrary interval containing y

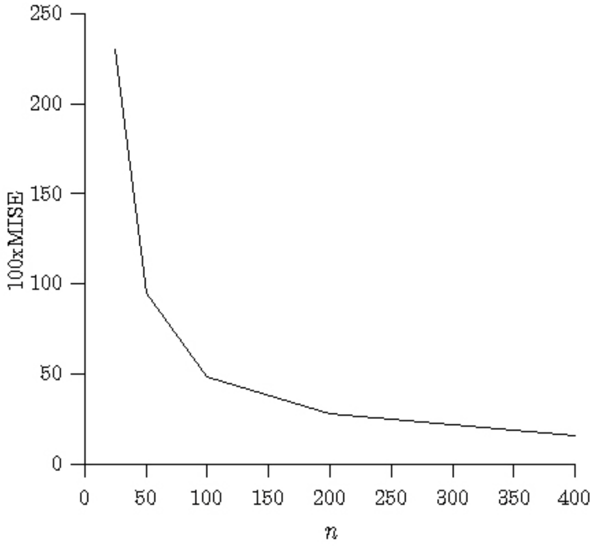

Fig. 2. MISE versus the number of observations.

in its interior, have the same limit. Therefore, the Hermite expansion of g converges to $g(y)$ at every $y \in M$ at which m^{-1} has a derivative. Now, in order to complete the proof, it suffices to apply Lemma B4 in Appendix B. ■

In the analysis of the convergence rate of $\mu_n(y)$, we impose, as usual in the paper, some regularity conditions on g and f. Assuming that m has p derivatives and denoting $\tau_p(y;g) = (y - d^p/dy^p)g(y)$, we find in Walter [40] that $|a_k| \leq |\alpha_{kp}|\,(k+1)^{-p/2}$, where α_{kp} is the kth coefficient of the expansion of $\tau(y;g)$ in the Hermite series. Thus,

$$|Eg_n(y) - g(y)| = \left|\sum_{k>N} a_k h_k(y)\right|$$

$$\leq \left|\sum_{|k|>N} \alpha_{kp}^2\right|^{1/2} \left|\sum_{|k|>N} (k+1)^p h_k^2(y)\right|^{1/2}$$

which, owing to (B.5) in Appendix B, is not greater than

$$c(y)\,\|\tau_p\| \left[\sum_{k>N} (k+1)^{-p-1/2}\right]^{1/2} = O(N^{-p/2+1/4}),$$

some function c, where $\|\cdot\|$ is the L_2 norm. Because, moreover, $\text{var}[g_n(y)] = O(N^{5/3}/n)$, $E[g_n(y) - g(y)]^2$ is of order $O(N^{-(6p-3)/(6p+7)})$, provided that $N \sim n^{1/(p+7/6)}$. Since $f_n(y)$ converges to $f(y)$ at the same rate, the in-probability convergence rate of the estimate is $O(n^{-(6p-3)/(12p+14)})$. In particular, it equals $O(n^{-9/38})$, for $p = 2$. For m bounded, variance is of order $O(N^{3/2}/n)$ and, for $N \sim n^{1/(p+1)}$ the estimate rate becomes $O(n^{-(2p-1)/(4p+4)})$. For $p = 2$, the rate is $O(n^{-1/4})$ and is slightly better.

At last, observe that the estimate can be also rewritten in a form more convenient from the computational viewpoint:

$$\mu_n(y) = \sum_{k=0}^{N} \alpha_{kn} H_k(y) \Bigg/ \sum_{k=0}^{N} \beta_{kn} H_k(y)\ .$$

where $\alpha_{kn} = \sum_{i=1}^{n} U_{i-1} H_k(Y_i)$ and $\beta_{kn} = \sum_{i=1}^{n} H_k(Y_i)$. Now, we do not calculate the time consuming operation $\exp(\cdot)$ but deal only with polynomials H_k's.

VII. The Modified Algorithm

In previous sections we have shown that (3.3) holds at every point $y \in D \cap M$ at which m^{-1} is differentiable. Obviously, D is $(-\pi,\pi)$, $(-1,1)$ or $(-\infty,\infty)$, for the trigonometric, Legendre or Hermite orthogonal series version of the estimate, respectively. Since f equals zero outside M, proofs of theorems given in the paper do not apply to $D \cap \bar{M}$, where $\bar{M}$ is the complement of M. In other words, convergence of the estimate at points in $D \cap \bar{M}$ has not been examined. It can be, however, sometimes desirable for our algorithms to converge to zero for y's in the set. In order to obtain that property, we propose the following very easy modification of our estimate:

$$\bar{\mu}_n(y) = \begin{cases} \mu_n(y), & \text{for } Y_{(1)} \leq y \leq Y_{(n)} \\ 0, & \text{otherwise,} \end{cases}$$

where we denote $Y_{(1)} = \min(Y_1, Y_2, \cdots, Y_n)$ and $Y_{(n)} = \max(Y_1, Y_2, \cdots, Y_n)$.

Theorem 4: Let (3.3) be satisfied. In addition, let (4.1), (5.1) or (6.1) hold, for the trigonometric, Legendre or Hermite version of the algorithm, respectively. Then

$$\bar{\mu}_n(y) \xrightarrow{n} \lambda m^{-1}(y) \text{ in probability}$$

at every $y \in D \cap M$ at which m^{-1} is differentiable and at every $y \in D \cap \bar{M}$. The set D is $(-\pi,\pi)$, $(-1,1)$ or $(-\infty,\infty)$, for the trigonometric, Legendre or Hermite series version of the estimate, respectively.

Proof: Observe that $\int_{\inf m(v)}^{Y_{(1)}} f(y)dy$, where $inf m(v)$ can be finite of infinite, has a beta distribution with parameters 1 and n, Wilks [42, p. 236], i.e., mean $1/(n+1)$ and variance $n/(n+1)2(n+2)$. This and the fact that the set $M = (\inf m(v), \sup m(v))$ is support of f imply in-probability convergence of $Y_{(1)}$ to $\inf m(v)$ as n tends to infinity. For similar reasons, $Y_{(n)}$ converges to $\sup m(v)$, where $\sup m(v)$ can be finite or infinite. Therefore, for every $y \in \bar{M}$, $\bar{\mu}_n(y)$ converges to zero as n tends to infinity in probability. Recalling the definition of m^{-1} and Theorems 1, 2, and 3, we complete the proof of the theorem. ■

VIII. Simulation Example

The algorithm has been numerically examined. In the example, the linear subsystem is governed by the equation: $X_{n+1} = 0.25X_n + U_n$, $W_n = X_n$. In turn, $Y_n = W_n + Z_n$, where Z_n has a normal distribution with zero mean and variance 0.1. The nonlinearity has been defined in the following way: $m(v) = v^2\text{sign}(v)$. The trigonometric series algorithm has been used to recover λm^{-1} in the interval $(-4,4)$. Since trigonometric expansions do not behave well at the interval ends, the MISE has been taken in the interval $(-3.17, 3.17)$. Notice that about 95% of Y_i's falls in the interval. The MISE has been numerically determined, for $n = 25, 50, 100, 200,$ and 400. For each of those n's, the

$N(n)$ minimizing the MISE has been found and the optimal MISE calculated. The result, i.e., the optimal MISE versus n, is presented in Fig. 2. Observe that the MISE is rather great for small n and decreases fast for n getting larger. Results of simulating the other algorithms, i.e, those derived from the Legendre and Hermite series, are very much like those presented above.

Compared with parametric algorithms, nonparametric require more observations to give satisfactory results. It is obviously caused by poor *a priori* information. In this paper, an additional reason, i.e., the composite structure of the system, plays a significant role. It is very well known that in Wiener systems, statistical dependence between their input and output signals is complicated. As a consequence of all those, the adequate number of observations, as suggested by the example, may be counted in hundreds rather than dozens. On the other hand, our algorithms are very easy and fast from the computational viewpoint. They are obviously much faster than those offered by parametric methods, since the latter usually invert matrices, i.e., perform time consuming operations.

At last, theoretical properties of our algorithms are rigorously proved, they are shown to converge to the unknown nonlinear characteristic. Behavior of parametric algorithms proposed by other authors to identify the nonlinearity in Wiener systems has not been theoretically verified so far and is just vague. Arguments in their favor are only that they do not contradict common sense and that have passed some numeric simulation tests. In the light of this, a good deal of optimism is necessary to apply them.

IX. Final Remarks and Conclusion

The paper is mainly devoted to the identification of the nonlinear part of the Wiener system. Nevertheless, observing that $\sigma_U^2\lambda_i = E\{V_nU_{n-i}\} = E\{m^{-1}(Y_n)U_{n-i}\}$, one can take $n^{-1}\sum_{j=1}^{n}\mu_n(Y_{j+i})U_j$ as an estimate of $\chi\lambda_i$. Clearly $\chi = (\sigma_U^2/\lambda) = \sigma_V^2/\lambda_1$ is unknown and can not be estimated because of the cascade structure of the system. Statistical dependence between μ_n, U_j's and Y_j's is, however, very complicated and an analysis of asymptotic behavior of the algorithm proposed to recover the impulse response of the linear subsystem is left to future works.

In the paper we have shown how to estimate the nonlinearity in the Wiener system. The m can be unbounded, or bounded from above or below, or from both sides. For m unbounded or bounded with unknown a or b, in order to recover the entire m, we use the Hermite series. For m bounded with known a and b, we can apply also the trigonometric and Legendre series. At any case, using the trigonometric or Legendre series, we recover a part of m, i.e., a part which values are in the interval $(-\pi, \pi)$ or $(-1, 1)$, respectively.

It is interesting to compare our asymptotic convergence rates obtained for particular versions of the algorithm. Suppose for this sake that m is bounded, twice differentiable and that $[a, b]$ is a subset of both $(-\pi, \pi)$ and $(-1, 1)$. Thus, roughly speaking, taking $p = 2$ in our analysis of the speed of convergence, we obtain the in-probability convergence rate $O(n^{-1/4})$, for both the trigonometric and Hermite series estimates. Therefore, as far as the asymptotic convergence rate is concerned, all three versions of the estimate behave very similarly.

Not less interesting it is to compare our rates of convergence with those known for other algorithms. For twice differentiable m, the nonparametric kernel estimate presented in Greblicki [16] converges at a rate $O(n^{-1/3})$, i.e., somewhat faster. Unfortunately, we are not able to compare rates given in this paper with those attained by parametric algorithms since no such result has been presented in literature, at least to the best author's knowledge. Nevertheless, our $O(n^{-1/4})$ and $O(n^{-1/5})$ discussed above do not look bad at $n^{-1/2}$, i.e. the rate typical for many kinds of parametric inference. Observe, moreover, that, for smooth m, i.e., for large p, our rates derived for both the trigonometric and Hermite versions of the algorithm become very close to the parametric rate $n^{-1/2}$.

Above considerations concerning convergence rates have an asymptotic sense, i.e., hold for n tending to infinity. The user can be interested in behavior of the algorithms for small and moderate n. The numerical example suggests that, for small number of observations, the error is rather large and that we may likely be forced to make more than a hundred input-output observations to achieve satisfactory accuracy.

Observe that in order to memorize our estimates, it obviously suffices to store all a_k's and b_k's, i.e., $2(N + 1)$ numbers, where N increases much slower than n. Therefore, having calculated a_k's and b_k's from n observed pairs, we can with little computational effort determine the value of the estimate at a point. The kernel estimate presented in Greblicki [16] requires all $2n$ observations, i.e., a larger amount of data, to be kept.

In the paper, m is assumed differentiable, which makes the density f of the output signal smooth and the analysis easier. Nevertheless, one can slightly relax restriction (2.5) by assuming that m is only piecewise differentiable and then examine consistency of the estimate at points at which the orthogonal expansions of both g and f converge, i.e., e.g., at points at which m^{-1} is differentiable. In turn, the hypothesis that the nonlinear characteristic has a bounded derivative is essential in the paper and plays a crucial role in the proof of fundamental Lemma A3.

We want to underline that the parametric and nonparametric approaches do not compete with each other, at least from the theoretical point of view. Just the contrary, the first can be successfully applied only when the *a priori* knowledge about the system is sufficiently large, whereas the latter does not require so much information. This difference makes nonparametric algorithms interesting in applications since in many real situations the *a priori* knowledge is very small and nonparametric in nature. On the other hand, we can apply nonparametric algorithms also when the *a priori* information is parametric, i.e., when the functional form of the characteristic is known up to a finite number of parameters. If the number is large, computational effort can become so great that we can be forced to resort to the nonparametric approach.

Appendix A

Lemma A1: Suppose that (2.2), (2.3), and (2.4) hold. Let $M = (a, b)$. If $\varphi(v)/m'(v) \to 0$ as $|v| \to \infty$, where $\varphi(v) = \exp(-v^2/2\sigma_V^2)$, then

$$f(y) \to 0 \text{ as } y \to a, \text{ and as } y \to b.$$

If both $v\varphi(v)/m'(v)$ and $\varphi(v)m''(v)/[m'(v)]^3$ converge to zero as $|v|$ tends to infinity, then

$$f'(y) \to 0 \text{ as } y \to a, \text{ and as } y \to b.$$

Proof: With no loss of generality, suppose that $m'(v)$ is positive in R. The first part of the lemma is a simple consequence of (2.7). In turn, taking into account that $[m^{-1}(y)]'' = -m''(v)/[m'(v)]^3$ at $v = m^{-1}(y)$, we obtain

$$f'(y) = f_V'(v)/m'(v) - f_V(v)m''(v)/[m'(v)]^3$$

at $v = m^{-1}(y)$, every $y \in M$. Hence, for some c_1 and c_2 independent of y

$$f'(y) = c_1 v\varphi(v)/m'(v) + c_2\varphi(v)m''(v)/[m'(v)]^3$$

at $v = m^{-1}(y)$. Since $y \to a$ or $y \to b$ is equivalent to $m^{-1}(y) \to -\infty$ or $m^{-1}(y) \to \infty$, respectively, the proof has been completed. ■

Lemma A2: Let (2.2), (2.3), and (2.4) hold. Let $M = (a, b)$. If $v\varphi(v)/m'(v) \to 0$ as $|v| \to \infty$, where $\varphi(v) = \exp(-v^2/2\sigma_V^2)$, then

$$f(y)m^{-1}(y) \to 0 \text{ as } y \to a, \text{ and as } y \to b.$$

If functions $v^2\varphi(v)/m'(v)$, $v\varphi(v)m''(v)/[m'(v)]^3$, and $\varphi(v)/[m'(v)]^2$, then

$$\left[f(y)m^{-1}(y)\right]' \to 0 \text{ as } y \to a, \text{ and as } y \to b.$$

Proof: The first part of the lemma is a simple consequence of (7.2). In order to verify the second, observe $g'(y) = f'(y)m^{-1}(y) + f(y)[m^{-1}(y)]'$ and proceed as in the proof of Lemma A1. ■

Lemma A3: Suppose that (2.2), (2.3), and (2.5) hold. Let t, a function defined in R, equal zero outside D. Moreover, let t be differentiable in D. Then, for $n = 1, 2, \cdots$,

$$|\text{cov}\,[U_n t(Y_{n+1}), U_0 t(Y_1)]| \le \alpha\beta\gamma\delta \left\|A^n\right\|, \tag{A.1}$$

where α is defined in (2.5), $\beta = \sup_{y\in M\cap D} |t(y)|$, and $\gamma = \sup_{y\in M\cap D} |t'(y)|$. The δ is a constant independent of both n, and t.

Proof: Clearly,

$$V_{n+1} = c^T A^n X_1 + \xi_n, \tag{A.2}$$

where $\xi_n = \sum_{i=1}^n c^T A^{n-i} bU_i + Z_{n+1}$. Since obviously $\text{cov}[U_n t(m(\xi_n)), U_0 t(Y_1)] = 0$, the covariance in (A.1) equals

$$\begin{aligned}
&\text{cov}\left\{U_n\left[t(Y_{n+1}) - t(m(\xi_n))\right], U_0 t(Y_1)\right\}\\
&= E\left\{U_n\left[t(Y_{n+1}) - t(m(\xi_n))\right], U_0 t(Y_1)\right\}\\
&- E\left\{U_n\left[t(Y_{n+1}) - t(m(\xi_n))\right]\right\} E\left\{U_0 t(Y_1)\right\}\\
&= S + T
\end{aligned}$$

say. From the fact that all arguments of t are in $M \cap D$ and from the inequality $|t(y_1) - t(y_2)| \le \gamma\,|y_1 - y_2|$, for all y_1 and y_2 in $M \cap D$, it follows that $|t(Y_{n+1}) - t(m(\xi_n))| \le \gamma\,|Y_{n+1} - m(\xi_n)|$, which, by virtue of (2.3), equals $\gamma\,|m(V_{n+1}) - m(\xi_n)|$. Recalling (2.5), we find the quantity bounded by $\alpha\gamma\,|V_{n+1} - \xi_n|$ which, by virtue of (A.2), is not greater than $\alpha\gamma\,\|c\|\,\|A^n\|\,\|X_1\|$. Therefore, the absolute value of S is not greater than $\alpha\beta\gamma\,\|c\|\,\|A^n\|\,E\{|U_0|\}E\{\|X_1\|\,|U_0|\}$ and $|T|$ does not exceed $\alpha\beta\gamma\,\|A^n\|\,E^2\{|U_0|\}E\,\|X_0\|$, which completes the proof. ■

Appendix B

Lemma B1: Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables convergent to a and b, respectively. Let $a_n = E(X_n - a)^2$, $b_n = E(Y_n - b)^2$ and let $b \neq 0$. Then,

$$P\left\{|X_n/Y_n - a/b| > \varepsilon\right\} = O(c_n)$$

any $\varepsilon > 0$ and

$$|X_n/Y_n - a/b| = O(c_n^{1/2}) \text{ in probability,}$$

where $c_n = \max(a_n, b_n)$.

Proof: Suppose that both a and b are positive. Obviously,

$$|X_n/Y_n - a/b| \le |X_n/Y_n|\,/b + |X_n - a|\,/b.$$

Therefore, the following two inequalities $|X_n - a| < a\varepsilon/(2+\varepsilon)$, and $|Y_n - b| < b\varepsilon/(2+\varepsilon)$ imply

$$\begin{aligned}
&P\left\{|X_n/Y_n - a/b| > \varepsilon a/b\right\}\\
&\le P\left\{|X_n - a| > a\varepsilon/(2+\varepsilon)\right\}\\
&+ P\left\{|Y_n - b| > b\varepsilon/(2+\varepsilon)\right\}
\end{aligned}$$

any $\varepsilon > 0$. An application of Chebyshev's inequality ends the proof. ■

Lemma B2: For the trigonometric system and any $y \in (-\pi, \pi)$,

$$\max_{-\pi\le x\le\pi} |K_N(x, y)| = O(N)$$

and

$$\max_{-\pi\le x\le\pi} \left|\frac{\partial}{\partial x} K_N(x, y)\right| = O(N^2).$$

Proof: Clearly $\max_t |D_N(t)| = O(N)$, where $t \in [\pi, \pi]$, and the first part of the theorem follows. Since, moreover, $D_N(t) = 1/2 + \sum_{k=1}^N \cos kt$, the proof has been completed. ■

Lemma B3: For the Legendre system and any $y \in (-1, 1)$,

$$\max_{-1\le x\le 1} |K_N(x, y)| = O(N),$$

$$\max_{-1\le x\le 1}\left|\frac{\partial}{\partial x}K_N(x,y)\right| = O(N^{7/2})$$

and

$$\max_{-1+\delta\le x\le 1-\delta}\left|\frac{\partial}{\partial x}K_N(x,y)\right| = O(N^{2})$$

any $\delta > 0$.

Proof: Fix $y \in (-1,1)$. From Christoffel-Darboux's formula

$$K_N(x,y) = [(N+1)/2] \times [P_N(x)P_{N+1}(y) - P_{N+1}(x)P_N(y)]/(y-x)$$

[37, p. 179], and the inequality $|P_N(x)| \le 1$, also [37, p. 181], we easily get

$$\max_{|x-y|\ge\rho}|K_N(x,y)| = O(N), \quad \text{(B.1)}$$

where both x and y are in $[-1,1]$, any $\rho > 0$.

The case $|x-y| < \rho$ requires much more attention. By virtue of Darboux's formula, [39, p. 196],

$$P_k(\cos\xi) = (2/\pi k\sin\xi)^{1/2}\cos(\kappa\xi+\pi/4)+O(k^{-3/2}) \quad \text{(B.2)}$$

where $\kappa = k+1/2$, and where $\xi \in (\varepsilon, \pi-\varepsilon)$, any $\varepsilon > 0$. The bound for the error, i.e., for the second term, is independent of ξ. Thus, in the interval, we have

$$P_k(\cos\xi) = O(k^{-1/2})\cos(\kappa\xi+\pi/4) + O(k^{-3/2})$$

where both bounds are independent of ξ. Hence

$$\begin{aligned} &K_N(\cos\xi,\cos\zeta)\\ &= O(1)\left[\frac{\cos[M\xi+\pi/4]\cos[(M+1)\zeta+\pi/4]}{\cos\xi-\cos\zeta}\right.\\ &\left.- \frac{\cos[M\zeta+\pi/4]\cos[(M+1)\xi+\pi/4]}{\cos\xi-\cos\zeta}\right]\\ &+ O(N^{-1/2}) \end{aligned}$$

where $M = N+1/2$. The quantity equals

$$\begin{aligned} &O(1)\left[\frac{\cos[M(\xi+\zeta)+\xi+\pi/2]\cos[M(\xi-\zeta)+\xi]}{\cos\xi-\cos\zeta}\right.\\ &\left.- \frac{\cos[M(\xi+\zeta)+\zeta+\pi/2]\cos[M(\xi-\zeta)-\zeta]}{\cos\xi-\cos\zeta}\right]\\ &+ O(N^{-1/2})\\ &= O(1)\left[\frac{\sin[(M+1/2)(\xi-\zeta)]}{\sin[(\xi-\zeta)/2]}\right.\\ &\left.+ \frac{\sin[(M+1/2)(\xi+\zeta)+\pi/2]}{\sin[(\xi+\zeta)/2]}\right] + O(N^{-1/2}) \end{aligned}$$

for ξ and ζ in $(\varepsilon, \pi-\varepsilon)$. Since ε, can be arbitrarily small, we easily obtain $\max_{|x-y|\ge\rho}|K_N(x,y)| = O(N)$, any $\rho < 0$. This and (B.1) yield finally the first part of the assertion.

To verify the second part, we begin with the following inequality: $\max_{|x|\le 1}|P'_k(y)| \le k$, [37, p. 250]. In turn, (B.2) leads to the following inequality $\max_{|y|\le\delta}|P_k(y)| = O(k^{1/2})$, any δ such that $0 < \delta < 1$. Since we, moreover, easily observe that $(\partial/\partial x)K_N(x,y) = \sum_{k=0}^{N}[(2k+1)/2]\,P'_k(x)P_k(y)$, the proof of this part of the assertion has been completed.

The third part can be now easily verified by using the inequality $|P'_k(y)| \le 2(k/\pi)^{1/2}/(1-x^2)$, $k = 1,2,\cdots$, [37, p. 201], which completes the proof. ∎

Lemma B.4: For the Hermite series and any $y \in (-\infty,\infty)$,

$$\max_{-\infty\le x\le\infty}|K_N(x,y)| = O(N^{1/2}),$$

$$\max_{-\infty\le x\le\infty}\left|\frac{\partial}{\partial x}K_N(x,y)\right| = O(N^{7/6})$$

and

$$\max_{|x|\le d}\left|\frac{\partial}{\partial x}K_N(x,y)\right| = O(N)$$

any $d > 0$.

Proof: Fix $y \in R$. From Christoffel's formula

$$K_N(x,y) = [(N+1)/2]^{1/2} \times [h_{N+1}(x)h_N(y) - h_{N+1}(y)h_N(x)]/(y-x)$$

[37, p. 307], and the inequality

$$\max_{-\infty\le x\le\infty}|h_k(x)| \le c(k+1)^{-1/12} \le c \quad \text{(B.3)}$$

where c is some constant independent of k, [39, p. 242], we get

$$\max_{|x-y|>\rho}|K_N(x,y)| = O(N^{1/2}) \quad \text{(B.4)}$$

any $\rho > 0$. In turn, $\pi K_N(x,y) = [\sin(v(x-y))]/(x-y) + T_N(x,y)/v$, where $2v < (2N+1)^{1/2} + (2N+3)^{1/2}$, and where T_N is bounded by a constant independent of N, for x and y varying in finite intervals, [37, p. 377]. Thus, for any $\rho > 0$, $\max_{|x-y|\le\rho}|K_N(x,y)| = O(N^{1/2})$ and, in the light of (B.4), the first part of the assertion follows.

From the definition of the Hermite series and the fact that $H'_{k+1}(x) = -2(k+1)H_k(x)$, [37, p. 304], we easily obtain

$$h'_k(x) = (k/2)^{1/2}h_{k-1}(x) - [(k+1)/2)]^{1/2}h_{k+1}(x).$$

Thus, using (B.3) and the inequality

$$\max_{|x|<t}|h_k(x)| \le c(k+1)^{-1/4} \quad \text{(B.5)}$$

any $t > 0$, where c depends on t but is independent of k, [39, p. 242], we get finally

$$\max_{-\infty<x<\infty}|h'_k(x)| \le O(k^{-5/12})$$

and

$$\max_{|x|\le d}|h'_k(x)| \le O(k^{1/4})$$

which completes the proof. ∎

Acknowledgement

The author thanks a reviewer for exceptionally careful reading the manuscript and helpful suggestions.

References

[1] R. Bars, I. Bézi, B. Pilipàr, and B. Ojhelyi, "Nonlinear long-range control of a distillation pilot plant", *Identification and System Parameter Estimation, Preprints of the 9th IFAC/IFORS Symposium*, Budapest, Hungary, 8-12 July 1990, vol. 2, pp. 848-853, 1990.

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

[3] S. A. Billings, "Identification of nonlinear systems -a survey", *IEE Proceedings*, vol 127, pp. 272-285, 1980.

[4] S. A. Billings and S. Y. Fakhouri, "Identification of nonlinear systems using the Wiener model", *Electronics Letters*, vol. 17, pp. 502-504, 1977.

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

[6] S. A. Billings and S. Y. Fakhouri, "Theory of separable processes with applications to the identification of nonlinear systems", P*roceedings of the IEE*, vol. 125, pp. 1051-1058, 1978.

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

[8] A. C. den Brinker, "A comparison of results from parameter estimation of impulse responses of the transient visual systems", *Biological Cybernetics*, vol. 61, pp. 139-151, 1989.

[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] 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.

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

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

[13] A. Georgiev, "Nonparametric kernel algorithm for recovering of functions from noisy measurements with applications", *IEEE Transactions on Automatic Control*, vol. AC-30, pp. 782-784, 1985.

[14] W. Greblicki, "Nonparametric system identification by orthogonal series", *Problems of Control and Information Theory*, vol. 8, pp. 67-73, 1979.

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

[16] W. Greblicki, "Nonparametric identification of Wiener systems", *IEEE Transactions on Information Theory*, vol. IT-38, pp. 1487-1493, Sept. 1992.

[17] W. Greblicki and M. Pawlak, "Fourier and Hermite series estimates of regression functions", *Annals of the Institute of Statistical Mathematics*, vol. 37, pp. 443-459, 1985.

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

[19] W. Greblicki and M. Pawlak, "Hammerstein system identification by non-parametric regression estimation", *International Journal of Control*, vol. 45, pp. 345-354, 1987.

[20] W. Greblicki and M. Pawlak, "Nonparametric identification of Hammerstein systems", *IEEE Transactions on Information Theory*, vol. IT-35, pp. 409-412, March 1989.

[21] W. Greblicki and M. Pawlak, "Recursive nonparametric identification of Hammerstein systems", *Journal of the Franklin Institute*, vol. 326, pp. 461-481, 1989.

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

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

[24] W. Greblicki, D. Rutkowska, and L. Rutkowski, "An orthogonal series estimate of time-varying regression", *Annals of the Institute of Statistical Mathematics*, vol. 35, pp. 215-228, 1983.

[26] Z. Hasiewicz, "Identification of a linear system observed through zero-memory non-linearity", *International Journal of Systems Science*, vol. 18, pp. 1595-1607, 1987.

[26] W. P. Huebner, G. M. Saidel, and R. L. Leigh, "Nonlinear parameter estimation applied to a model of smooth pursuit eye movements", *Biological Cybernetics*, vol. 62, pp. 265-273, 1990.

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

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

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

[30] R. Kronmal and M. Tarter, "The estimation of probability densities and cumulatives by Fourier series methods", *Journal of the American Statistical Association*, vol. 63, pp. 925-952, 1968.

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

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

[33] M. Pawlak, "On the series expansion approach to the identification of Hammerstein systems", *IEEE Transactions on Automatic Control*, vol. AC-36, pp. 763-767, 1991.

[34] M. Pawlak and W. Greblicki, "Nonparametric estimation of a class of nonlinear time series", *Nonparametric Estimation and Related Topics*, ed. G. Roussas. Kluver Academic Publishers, pp. 541-552, 1991.

[35] B. L. S. Prakasa Rao, *Nonparametric Functional Estimation.* Orlando: Academic Press, 1983.

[36] E. Rafajłowicz, "Nonparametric orthogonal series estimators of regression: a class attaining the optimal rate in L2", *Statistics and Probability Letters*, vol. 5, pp. 219-224, 1987.

[37] G. Sansone, O*rthogonal Functions.* Interscience Publishers Inc., 1959.

[38] S. C. Schwartz, "Estimation of probability density by an orthogonal series, *Annals of Mathematical Statistics*, vol. 38, pp. 1261-1265, 1967.

[39] G. Szegö, *Orthogonal Polynomials.* Providence: Amer. Math. Soc. Coll. Publ., 1966.

[40] G. G. Walter, "Properties of Hermite series estimation of probability density", *Annals of Statistics*, vol. 5, pp. 1258-1264, 1977.

[41] D. T. Westwick and R. E. Kearney, "A new algorithm for the identification of multiple input Wiener systems", *Biological Cybernetics*, vol. 68, pp. 75-85, 1992.

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

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.

Since 1966, Dr. Greblicki has been with the Technical University of Wrocław, Wrocław, Poland. He currently holds a Professor position in the Institute of Engineering Cybernetics, Technical University of Wrocław. In the academic year 1986-1987 he held an appointment as a visiting professor with the Department of Electrical Engineering, University of Manitoba, Winnipeg, Manitoba, Canada. His research interests include nonparametric mathematical statistics methods and their application to system identification and classification. He teaches courses in control theory and system identification.