Nonparametric Identification of Wiener Systems

Włodzimierz Greblicki

Abstract— **A Wiener system, i.e., a system in which a linear dynamic part is followed by a nonlinear and memoryless one, is identified. No parametric restriction is imposed on the functional form of the nonlinear characteristic of the memoryless subsystem and a nonparametric algorithm recovering the characteristic from input-output observations of the whole system is proposed. Its consistency is shown and the rate of convergence is given. An idea for identification of the impulse response of the linear subsystem is proposed. Results of numerical simulation are presented.**

Keywords— **System identification, Wiener system, nonparametric estimation, nonparametric regression, kernel regression estimation**

I. Introduction

WHEREAS theory of identification of linear dynamic systems has been already elaborated to the state of great integrity, see e.g. Norton [27], Liung [24] or Söderstöm and Stoica [30], theory of identification of nonlinear dynamic systems is not yet satisfactory. The problem of proper descriptions of nonlinear systems is still under discussion, see e.g. Sandberg [29]. In the field of nonlinear system identification, there is an obvious need for new ideas, methods, and algorithms.

One of the most promising approaches is based on the assumption that the identified system consists of relatively simple subsystems, and that the structure of the system is known. Signals interconnecting the subsystems are usually not accessible to measurement. As a consequence, the subsystems are defined on the basis of both observed input-output signals of the whole system and the *a priori* information about the system. In the literature, this approach is called block oriented. In passing, we want to mention that one can also identify the structure of the system, see Haber and Unbenhauen [20]. In this paper, however, the structure of the system under investigation is fixed and known.

The idea of block oriented identification is present in the literature in as early papers as those of Narendra and Gallman [26], Gardiner [13], Gallman [12], Webb [32], Brilinger [10], Billings and Fakhouri [4], and recently, in Hunter and Korenberg [21], as well as Korenberg and Hunter [23]. A methodical analysis of systems examined by this approach has been presented in Billings and Fakhouri [4]-[9], as well as in Bendat [2]; see also Billings [3].

Block oriented identification focuses attention mainly on two types of systems, i.e., the Hammerstein and Wiener ones. In the first, a nonlinear memoryless subsystem is followed by a linear dynamic one while the latter consists of the same subsystems connected in the reverse order. So far, much more attention has been paid in the literature to the Hammerstein system, see e.g., Narendra and Gallman [26], Gallman [12], Billings and Fakhouri [7], Hwang and Shyu [22]. While a number of identification algorithms have been proposed to identify its linear part, recovering the nonlinearity has appeared to be much more difficult problem. All the authors previously mentioned have assumed that the nonlinear characteristic is a polynomial of finite and known order. Since they estimate only a finite number of unknown parameters, i.e., polynomial coefficients, their inference problems are parametric. The accuracy of a description of a real characteristic with a finite-order polynomial has been left as an open question.

In a series of papers, Greblicki and Pawlak [15]-[19], as well as Greblicki [14], have made a step toward real situations by not imposing any parametric restriction on the functional form of the unknown characteristic. Their poor *a priori* information is closer to that encountered in real situations, when the knowledge about the system before experiment is so small that the family of all possible characteristics cannot be parameterized. They have proposed to apply nonparametric statistical methods.

In the paper, we employ the nonparametric approach in order to identify the nonlinearity in Wiener systems. We make no parametric-type restriction concerning its polynomial form; we do not assume that it is, e.g., a polynomial. In consequence, our problem is nonparametric since the family of all possible characteristics is so wide that cannot be represented in a parametric form. We observe the inverse of the characteristic can be expressed as a regression function and, making use of some ideas developed earlier in the statistical literature on nonparametric regression estimation, we present an algorithm for recovering the inverse. We show its pointwise consistency and derive the rate of convergence. Given the estimate, we easily get an estimate of the true characteristic. We also present an idea for identification of the linear subsystem, i.e., for recovering its impulse response. Results of numerical simulation are also presented.

The Wiener structure is interesting not only for a theoretical challenge but also for the fact that it is often encountered in various areas such as biology, see den Brinken [11] for a study of visual systems, industry, see Bars *et. al.* [1] for a description of a distillation plant with the help of the Wiener model, psychology, sociology, and other fields, see Korenberg and Hunter [2] and Hunter and Korenberg [2] for example.

II. Identification Problem

The Wiener system has a cascade structure consisting of two subsystems (see Fig. 1). The first is linear and dynamic, while the other is nonlinear and memoryless. In the paper, the system is driven by a zero-mean stationary random process $\{U_n; n = \cdots, 1, 0, 1, \ldots\}$, where the random

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

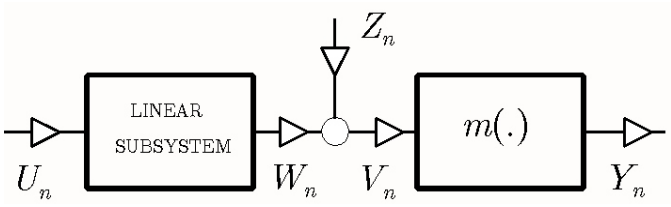

Fig. 1. Identified Wiener system.

variables U_n's are normally distributed. The behavior of the first subsystem is governed by the state space equations

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

In this equation as well as in the whole paper, ^{T} denotes the transposition.

The matrix A and the vectors b and c are unknown. Nevertheless, the system is asymptotically stable by assumption. X_n and W_n are the state vector random variable and the random variable at the output of the first subsystem, respectively. Clearly, the random processes $\{X_n; n = \cdots, 1, 0, 1, \ldots\}$ and $\{W_n; n = \cdots, 1, 0, 1, \ldots\}$ are stationary and Gaussian. The disturbance $\{Z_n; n = \cdots, 1, 0, 1, \ldots\}$ is zero-mean stationary white random noise independent of the input signal. The noise is additive, i.e.,

$$V_n = W_n + Z_n, \tag{2}$$

where the random variable V_n is the input of the second subsystem.

The second subsystem is memoryless, nonlinear. Its characteristic is denoted by m, which means that

$$Y_n = m(V_n), \tag{3}$$

where Y_n is the output random variable of the whole system. The m is a Borel measurable function defined on the real line R. We assume that m is unknown, and should be recovered from input-output observations of the system, i.e., from pairs (U_i, Y_{i+1}), $i = 1, 2, \cdots, n$.

Denote by D, $D \subseteq R$, the image of R under the mapping m. Obviously, all values of Y_n's lie in D. Assuming that m is a one-to-one mapping, we can define m^{-1} in the following way:

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

In the remainder of the paper, m^{-1} will be called the inverse of m.

The characteristic m is assumed to satisfy the Lipschitz condition of the pth order, $p > 0$, i.e.,

$$|m(v_1) - m(v_2)| \le \alpha \, |m(v_1) - m(v_2)|^p, \tag{4}$$

some $\alpha > 0$, all v_1 and v_2. Because we want Y_n to have a probability density, we assume that m is differentiable. The density is denoted by f.

One easily observe that the restrictions imposed on m are satisfied by any strictly monotonous Borel function with a bounded derivative.

In this section, we present an algorithm recovering m^{-1}, i.e., the inverse of the unknown characteristic. Given the estimate of m^{-1}, we easily define an estimate of m.

The idea of the algorithm proposed in the paper is based on the observation that m^{-1} can be represented as a regression function. In order to introduce our algorithm, we define the following regression:

$$\mu(y) = E\left[U_n \,|Y_{n+1} = y\right]. \tag{5}$$

Owing to (3), $\mu(y) = E[U_n \,|V_{n+1} = m^{-1}(y)]$. From this, (1), and (2), it follows that

$$c^T b \mu(y) = E\left[c^T b U_n \,|c^T A X_n + Z_{n+1} + c^T b U_n = m^{-1}(y)\right].$$

Clearly, $c^T b U_n$ has the normal distribution with zero-mean and variance $(c^T b)^2 \sigma_u^2$, where σ_u^2 is the variance of U_n. In turn, $c^T A X_n + Z_{n+1}$ is also Gaussian with zero-mean and variance $c^T A \Sigma A^T c + \sigma_z^2$, where $\Sigma = \mathrm{cov}(X_n, X_n)$, and where σ_z^2 is the variance of Z_n. From this, the fact that $c^T b U_n$ and $c^T A X_n + Z_{n+1}$ are independent, and Lemma 1 in Appendix, it follows that the conditional density of $c^T b U_n$, given $c^T A X_n + Z_{n+1} + c^T b U_n = d$, is also normal with mean $[(c^T b)^2 (\sigma_u^2/\sigma_v^2)]d$, where $\sigma_v^2 = (c^T b)^2 \sigma_u^2 + c^T A \Sigma A^T c + \sigma_z^2$ is the variance of V_n. Therefore

$$\mu(y) = \lambda m^{-1}(y), \tag{6}$$

where $\lambda = (c^T b)^2 \sigma_u^2 / \sigma_v^2$. Hence

$$\mu^{-1}(v) = m(v/\lambda). \tag{7}$$

In Section III, we present an estimate $\mu_n(y)$ of $\mu(y)$. This way, owing to (6), we are able to estimate $m^{-1}(y)$ up to some unknown multiplying constant λ. We then show that the estimate $\mu_n(y)$ of $\lambda m^{-1}(y)$ is consistent. By (7), we can use $\mu_n^+(y)$, a pseudoinverse of μ_n, as an estimate of $m(v/\lambda)$. We use a pseudoinverse since μ_n may not be invertible, despite the fact that μ has the inverse.

The fact that we are not able to determine λ nor to estimate it is a simple consequence of the cascade structure of the system. Having input-output observations of the whole system, it is impossible to distinguish a Wiener system with an impulse response $\{k(i)\}$ and a nonlinear characteristic $m(v)$ from a Wiener system with an impulse response $\{\lambda k(i)\}$ and a nonlinearity $m(v/\lambda)$. This property holds for any identification algorithm using our *a priori* information and measurement data.

III. Nonlinear Subsystem Identification

We now introduce an algorithm for recovering the unknown characteristic of the nonlinear part of the Wiener system. Using (5), we suggest to estimate $\mu(y)$ by the nonparametric kernel regression estimate

$$\mu_n(y) = \frac{\sum_{i=1}^{n} U_i K\left(\dfrac{y - Y_{i+1}}{h(n)}\right)}{\sum_{i=1}^{n} K\left(\dfrac{y - Y_{i+1}}{h(n)}\right)}, \tag{8}$$

where K is a kernel function and $\{h(n)\}$ a sequence of positive numbers. Both the kernel and the sequence are selected by the user.

The motivation for the algorithm can be easily explained by employing the rectangular kernel. For this kernel,

$$\mu_n(y)=\left((1/2nh(n)\sum_{i\in I_y}U_i\right)\Bigg/\left((1/2nh(n)\sum_{i\in I_y}1\right),$$

where $I_y=\{i:|y-Y_{i+1}|\le h(n)\}$. The numerator and denominator clearly estimate

$$\int_{y-h(n)}^{y+h(n)}\mu(\eta)f(\eta)d\eta \text{ and } \int_{y-h(n)}^{y+h(n)}f(\eta)d\eta,$$

respectively. Under some regularity conditions, these quantities converge to $\mu(y)f(y)$ and $f(y)$ as $h(n)$ tends to zero, respectively. In view of this, one can expect the estimate to converge to $\mu(y)$ in the process of identification.

We assume that K is a nonnegative Borel measurable function such that

$$\sup_{y\in R}|K(y)|=k<\infty, \tag{9}$$

$$\int K(y)dy=1, \tag{10}$$

and

$$K(y)y^{1+\varepsilon}\to 0 \text{ as } |y|\to\infty \tag{11}$$

for some $\varepsilon>0$. In (10), as well as in the remainder of the paper, all integrals are taken on the entire real line.

Restrictions (9)-(10) are typical in nonparametric regression. Since our observations are dependent, we assume additionally that the kernel is Lipschitz, i.e., that

$$|K(y_1)-K(y_2)|\le\gamma|y_1-y_2|, \tag{12}$$

some $\gamma>0$, all y_1, and y_2. This assumption plays an important role in the proof of our Lemma 4 in the Appendix.

The sequence $h(n)$ satisfies the following restrictions:

$$h(n)\to 0 \text{ as } n\to\infty, \tag{13}$$

$$nh^2(n)\to\infty \text{ as } n\to\infty. \tag{14}$$

There are many functions that can be used as kernels, a few of which are

a) $K(y)=\begin{cases}1-|y|, & \text{for } |y|\le 1,\\ 0, & \text{otherwise,}\end{cases}$

b) $K(y)=\begin{cases}(3/4)(1-y)^2, & \text{for } |y|\le 1,\\ 0, & \text{otherwise,}\end{cases}$

c) $K(y)=(1/2\pi)^{1/2}\exp(-y^2/2)$,

d) $K(y)=(1/2)\exp(-|y|)$,

e) $K(y)=(1/\pi)(1/(1+y^2))$.

An example of $h(n)$ is given by $h(n)=cn^{-\rho}$, where $c>0$, and $0<\rho<1/2$.

The kernel nonparametric regression has been introduced independently by Nadaraya [25] and Watson [31] and examined by a number of authors. We refer the reader to Rao [28] for an extensive list of papers on nonparametric regression estimation. We also remark that the statistical literature previously mentioned deals mainly with estimating regression function from independent pairs of observations. Our problem is more difficult since our pairs (U_i,Y_{i+1})'s are dependent because the Wiener system is dynamic.

In the paper, we use the following notation:$\|x\|=\left|x^{(1)}\right|+\cdots+\left|x^{(d)}\right|$, where d is the dimension of a vector x, and $x^{(i)}$ its ith coordinate. By $\|P\|$ we denote the appropriate norm of a matrix P. We write "almost every" for "almost every with respect to the Lebesgue measure". For a sequence $\{X_n;n=1,2,\cdots\}$ of random variables, $X_n=O(a_n)$ in probability means that $(\gamma_n/a_n)X_n\to 0$ as $n\to\infty$ in probability for every sequence $\{\gamma_n\}$ convergent to zero.

IV. Convergence of the Algorithm

In this section, we prove that our algorithm converges to the inverse of the unknown characteristic and examine the rate of convergence. We show how the rate is affected by the number sequence and the kernel. Next, we find $h(h)$ that optimizes the rate. Finally, we show how to estimate the true characteristic.

Theorem 1: Let the nonnegative Borel kernel measurable kernel satisfy (9)-(12). For the number sequence, let (13) and (14) hold. Then, for the Wiener system,

$$\mu_n(y)\to\lambda m^{-1}(y), \text{ as } n\to\infty \text{ in probability}$$

at almost every $y\in R$ at which $f(y)>0$, i.e., at almost every $y\in R$ belonging to support of f.

Proof: The estimate can be rewritten in the following form:

$$\mu_n(y)=g_n(y)/f_n(y), \tag{15}$$

where

$$g_n(y)=\frac{1}{nh(n)}\sum_{i=1}^{n}U_iK\left(\frac{y-Y_{i+1}}{h(n)}\right)$$

and

$$f_n(y)=\frac{1}{nh(n)}\sum_{i=1}^{n}K\left(\frac{y-Y_{i+1}}{h(n)}\right).$$

We have

$$\begin{aligned}Eg_n(y)&=\frac{1}{h(n)}E\left\{E[U_0|Y_1]K\left(\frac{y-Y_1}{h(n)}\right)\right\}\\&=\frac{1}{h(n)}\int\mu(\eta)f(\eta)K\left(\frac{y-\eta}{h(n)}\right)d\eta.\end{aligned}$$

By virtue of (13), the fact that $\int\mu(\eta)f(\eta)d\eta=\lambda EV_n<\infty$, and Lemma 2 in the Appendix, this quantity converges to $g(y)=\mu(y)/f(y)$ at almost every point $y\in R$. In turn,

$$\begin{aligned}&\operatorname{var}(g_n(y))\\&=\frac{1}{nh^2(n)}\operatorname{var}\left(U_0K\left(\frac{y-Y_1}{h(n)}\right)\right)\\&+\frac{1}{n^2h^2(n)}\sum_{i=1}^{n-1}(n-i)\\&\cdot\operatorname{cov}\left(U_iK\left(\frac{y-Y_{i+1}}{h(n)}\right),U_0K\left(\frac{y-Y_1}{h(n)}\right)\right).\end{aligned}\tag{16}$$

Using (9), we see that the first term in this expression does not exceed $(1/nh^2(n))k^2EU_0^2 = c_1/nh^2(n)$, where c_1 is clearly independent of both n and $h(n)$.

The second term in (16) requires more attention. Observing that $|K(y_1/h) - K(y_2/h)| \le (\gamma/h)\,|y_1 - y_2|$, $h > 0$, and then applying Lemma 4, we find the absolute value of the covariance under the sum in (16) not greater than $(1/h(n))\left\|A^i\right\|^p \omega_{h(n)}(y)$, where $\omega_h(y)(1/h)E[\psi(Y_1)K((y - Y_1)/h)]$, and where ψ is some nonnegative function such that $E[\psi(Y_1)]$ exists. Moreover, ψ is independent of both i and $h(n)$. Thus, with the help of Lemma 2, the second term in (16) can be bounded in absolute value by

$$\begin{aligned}&(1/n^2h^2(n))\omega(y)\sum_{i=1}^{n-1}(n-i)\left\|A^i\right\|^p\\ &\le (1/nh^2(n))c_2(y),\end{aligned} \tag{17}$$

where $\omega(y) = \sup_{0<h<\infty}\omega_h(y)$ and $c_2(y) = \sum_{i=1}^{\infty}\left\|A^i\right\|^p$. Since all eigenvalues of A are inside the unit circle, the sum in the definition of $c_2(y)$ is finite. Obviously, $c_2(y)$ is finite at almost every $y \in R$, and independent of both n, as well as $h(n)$.

In this way we have shown that

$$\mathrm{var}((g_n(y)) \le c_3(y)/nh^2(n), \tag{18}$$

where $c_3(y)$ is finite at almost every $y \in R$, and independent of both n and $h(n)$. Therefore, we have shown that $g_n(y) \to g(y)$ as $n \to \infty$ at almost every $y \in R$.

Applying similar arguments, one can verify $Ef_n(y) \to f(y)$ as $n \to \infty$ and $\mathrm{var}(f(y)) \le c_4(y)/nh^2(n)$, where $c_4(y)$ is finite at almost every $y \in R$ and independent of both n and $h(n)$. Recalling (15), we complete the proof. ■

We shall now examine the rate of convergence of the algorithm and show how to select $h(n)$ to obtain the optimal value of the rate. Taking a kernel for which $\int yK(y)dy = 0$ and $\int y^2K(y)dy < \infty$, which is the case with examples a)-d) but not e), and expanding both $\mu(y - h(n)\eta)$ and $f(y - h(n)\eta)$ in a Taylor series, we get

$$\begin{aligned}Eg_n(y) - g(y) &= \int (g(\eta) - g(y - h(n)\eta))K(\eta)d\eta\\ &= (h^2(n)/2)g(y)\int y^2K(y)dy\\ &\quad + o(h^2(n)),\end{aligned}$$

where $g(y) = \mu(y)f(y)$. We have assumed that all derivatives in this expression exist. Therefore,

$$(Eg_n(y) - g(y))^2 \le c_5(y)h^4(n) + o(h^4(n)),$$

where c_5 is independent of both n and $h(n)$. This together with (1) yields,

$$\begin{aligned}&E(g_n(y) - g(y))^2\\ &\le c_5(y)h^4(n) + c_3(y)/nh^2(n) + o(h^4(n)).\end{aligned}$$

Clearly, the sum of the two most significant terms on the right-hand side of the previous expression is minimized for

$$h(n) = cn^{-\rho}, \tag{19}$$

where $\rho = 1/6$, and where c can be easily expressed in terms of c_3 and c_4. For this number sequence,

$$\begin{aligned}&E(g_n(y) - g(y))^2\\ &\le (c^4c_5(y) + c^{-2}c_3(y))n^{-2/3} + o(n^{2/3})\\ &= O(n^{-2/3}).\end{aligned} \tag{20}$$

This rate holds also for $h(n)$ selected according to (19) with $\rho = 1/6$ and any $c > 0$.

For similar reasons,

$$E(f_n(y) - f(y))^2 = O(n^{-2/3}) \tag{21}$$

for the same number sequence.

From (20), (21) and Lemma 3 in the Appendix, we finally obtain

$$\mu_n(y) - \mu(y) = O(n^{-1/3}) \text{ in probability,} \tag{22}$$

for almost every $y \in R$ at which $f(y) > 0$. Equation (22) means that the estimate converges to $\mu(y)$ in probability at the rate equal $O(n^{-1/3})$, i.e., not slower than $n^{-1/3}$, provided that $h(n) = cn^{-1/6}$, any $c > 0$.

Repeating the previous arguments, one can easily observe that for $\{h(n)\}$ selected according to (19), any $\rho > 0$ and any $c > 0$, the rate at which the algorithm converges to $\mu(y)$ in probability is $O(n^{-\delta})$, where δ depends on ρ but is independent of both c and the kernel. The rate is optimized for $\rho = -1/6$ and its optimal value equals $O(n^{-1/3})$. In the light of this, the sequence $h(n) = cn^{-1/6}$ can be recommended for a differentiable m. In general, selecting the sequence according to (19), one should pay more attention to ρ than c, at least for large n.

All we have said about the choice of the number sequence is true only in the asymptotic sense. Analyzing the estimate for moderate and small n appears to be much more difficult for even the system in which the dynamic subsystem is just a delay, i.e., in which $A = 0$. The problem of selecting $h(n)$, in this case, has been already studied in the statistical literature and has not yet been solved despite the fact that the inference is then much easier because is made from independent pairs.

Given the estimate $\mu_n(y)$ of $\lambda m^{-1}(y)$, one can recover $m(v/\lambda)$, where λ is some unknown number. The problem, however, is that μ_n may not have an inverse. In order to calculate an estimate of $m(v/\lambda)$, we use μ_n^+, a pseudoinverse of μ_n, defined in the following way:

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

where $I_v = \{y : \mu_n(y) = v\}$. We can now apply $\mu_n^+(v)$ as an algorithm recovering $m(v/\lambda)$.

In the paper, m is assumed differentiable, so the output signal has a density. This restriction can be relaxed, however. It suffices to use Lemma 4 and then proceed as in Greblicki and Pawlak [17] in the proof of their Theorem 2. Although, in the paper, the nonlinearity is globally Lipschitz, it is possible to extend our results to locally Lipschitz characteristics.. All those improvements and details are left to future works.

V. Identification of the Dynamic Subsystem

In this section, we present an idea for identification of the linear dynamic subsystem. The suggested algorithm recovers the impulse response $\{k(i); i = 1, 2, \cdots\}$ of the subsystem. We shall need the following Corollary obtainable fro Theorem 1.

Corollary: Under the conditions of Theorem 1,

$$\mu_n(Y) \to \mu(Y) \text{ as } n \to \infty \text{ in probability,}$$

where Y is a random variable independent of both $\{U_i; i = 1, 2, \cdots\}$ and $\{Y_i; i = 1, 2, \cdots\}$, and having the probability density f.

We begin with the presentation of the algorithm with the following observation:$\lambda k(i) = \lambda E[U_0 V_i] = E[Y_0 \mu(Y_i)]$. The difficulty is that we do not know the characteristic m nor, in consequence, μ. However, we can use μ_n instead of μ. Suppose that he impulse response is recovered fro input-output observations independent of those from which μ has been estimated. Let us denote those observations by $(U_1', Y_2'), (U_2', Y_3'), \cdots$. We propose

$$\chi_{n,N}(i) = (1/N) \sum_{j=1}^{N} U_j' \mu_n(Y_{i+j}'), \tag{23}$$

some $N > 0$, as our estimate of $\lambda k(i)$.

We have already shown, see the corollary,

$$\mu_n(Y_{i+j}') \to \mu(Y_{i+j}') \text{ as } n \to \infty \text{ in probability.}$$

Hence, for any fixed N,

$$\mu_n(Y_{i+j}') \to (1/N) \sum_{j=1}^{N} U_j' \mu_n(Y_{i+j}')$$
$$\text{as } n \to \infty \text{ in probability.}$$

Since the process $(U_1', Y_2'), (U_2', Y_3'), \cdots$ is strictly stationary, by virtue of the ergodic theorem,

$$(1/N) \sum_{j=1}^{N} U_j' \mu_n(Y_{i+j}') \to E[U_0 \mu(Y_i)], \text{ as } N \to \infty,$$

almost surely. As the limit equals $\lambda k(i)$, we can use (23) in order to estimate the impulse response. The previous discussion shows that the larger both n and N, the better the result. Thus, one can estimate μ from the first sequence of input-output observations and then, after a sufficiently long time interval, observe the second one. The latter, practically independent of the first, can be used to estimate the impulse response. What we have shown here is merely a suggestion, rather than a formal proof of convergence of the algorithm.

VI. Simulation Results

In this section, we present results of numerical simulation of the identification algorithm. In the example, X_n is a scalare and, consequently, the matrix A is reduced to a scalar denoted by a. We have taken $m(v) = v^2 \operatorname{sign}(v)$ which leads to $\mu(y) = \lambda y^{1/2} \operatorname{sign}(y)$. It is not difficult to verify $\lambda = \sigma_u^2/(b\sigma_u^2 + \sigma_z^2)$, where $b = (2 - a^2)/(1 - a^2)$. We have set $a = 0.2$, $\sigma_u^2 = 1$, $\sigma_z^2 = 0.1$, which gives $\lambda = 0.47$. The parabolic kernel defined in example b) in Section III has been used.

We have already shown that our algorithm converges to $\mu(y)$ pointwise. Since we want to have a global measure of the quality of the algorithm, the following error has been empirically evaluated:

$$\text{ERROR}(h(n)) = \frac{1}{101} \sum_{j=0}^{100} E(\mu_n(-3.18 + j\delta) - E\mu(-3.18 + j\delta))^2,$$

where $\delta = 2 \cdot 3.18/100$. In other words, we have calculated errors at 101 points, namely, 3.18, $-3.18 + \delta, \cdots$, 3.18, and defined the ERROR$(h(n))$ as their average. The probability that Y_n lies in the interval $(-3.18, 3.8)$ equals 0.95 and therefore, the ERROR$(h(n))$ seems to be a good global measure of the quality of our algorithm. The error has been empirically determined for $n =, 25, 100, 200$, and 400. For each n, we have found $h^*(n)$, i.e., $h(n)$ minimizing the ERROR$(h(n))$. Fig. 2 shows ERROR$(h^*(n))$, i.e., the minimal possible ERROR$(h(n))$ as a function of n. It is seen that the error decreases rapidly for small n and relatively slowly for n larger than 200.

VII. Final Remarks and Conclusion

The idea of applying the nonparametric approach to identify a class of nonlinear dynamic cascade systems has been proposed and developed by Greblicki and Pawlak in a series of papers mentioned earlier. They have suggested and examined a number of algorithm recovering the nonlinearity in Hammerstein systems under both stationary as well as quasi-stationary conditions. In this paper, we have shown that the nonparametric approach can be also applied to recover the nonlinearity in Wiener systems. Compared to those identifying Hammerstein systems, the algorithm examined in this paper converges at a smaller rate and restrictions imposed on the number sequence are sharper.. It is caused by the fact that the environment in which the algorithm works is much more complicated from a statistical point of view since the type of the statistical dependence is more inconvenient. In this paper, regressors, i.e., Y_n's are dependent while in Hammerstein system identification, the situation is simpler because U_n's are regressors and are independent. Despite those adverse circumstances, the rate of convergence given in (22) is encouraging since it is not much worse than $n^{-/1/2}$, which is a typical rate in the parametric inference.

The algorithm presented in the paper is convenient fro the computational viewpoint since no time consuming operations are involved. Calculations are extremely quick for the kernel defined in example a) of Section III. On the other hand, due to the smoothness of kernel d), the estimate using it is not rough as that the kernel a). Taking all that into account, the parabolic kernel b) seem to be good for both its numerical simplicity and smoothness..

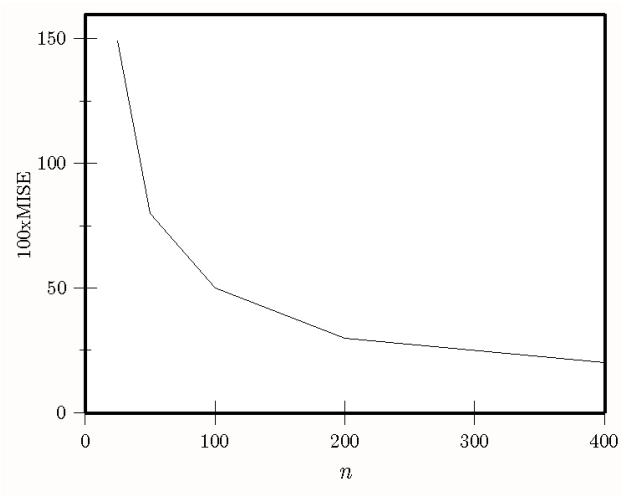

Fig. 2. ERROR ($h^*(n)$) versus the number of observations.

Appendix

Lemma 1: Let $Z = X + Y$, where X and Y are two independent Gaussian random variables with zero-mean and variance σ_x^2 and σ_y^2, respectively. Then, the conditional density of X conditioned on $Z = z$ is also normal with mean $[\sigma_y^{-2}/(\sigma_x^{-2} + \sigma_y^{-2})]z$.

The proof is straightforward and can be found in many text-books on probability and statistics.

Lemma 2: Let

$$u_h(x) = \frac{1}{h}\int \varphi(y) K\left(\frac{x-y}{h}\right) dy,$$

where φ is an integrable Borel measurable function and where K is a Borel measurable kernel satisfying (9)-(1). Then

$$u_h(x) \to \varphi(x) \text{ as } h \to 0,$$

almost every $x \in R$. Moreover,

$$\sup_{0<h<\infty} |u_h(x)|$$

is finite for almost every $x \in R$.

The first part of the lemma is in Wheeden and Zygmund [33, p.152] while the second is an immediate consequence of the first.

Lemma 3: For random variables X_n and Y_n, let us denote $a_n = E(X_n - a)^2$ and $b_n = E(X_n - b)^2$, $b > 0$. Then, $X_n/Y_n = O(c_n^{1/2})$ in probability, where $c_n = \max(a_n, b_n)$.

Proof: Clearly $X_n = O(a^{1/2})$ and $X_n = O(b^{1/2})$ in probability. This and the inequality

$$\left|\frac{X_n}{Y_n} - \frac{a}{b}\right| \le \left|\frac{X_n}{Y_n}\right|\left|\frac{Y_n - b}{b}\right| + \left|\frac{X_n - b}{b}\right|$$

complete the proof. ■

Lemma 4: Let t be a Borel measurable function such that $|t(y_1) - t(y_2)| \le \beta |y_1 - y_2|$ for some $\beta > 0$ and all $y_1$and y_2. Then, in the Wiener system,

$$|\mathrm{cov}[U_n t(Y_{n+1}), U_0 t(Y_1)]| \le \beta \|A^n\|^p E[\varphi(Y_1)|t(Y_1)|], \quad (24)$$

where φ is some nonnegative function such that $E[\varphi(Y_1)]$ exists. Moreover, φ is independent of n, t, and β.

Proof: Obviously,

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

where $\xi_n = \sum_{i=1}^n c^T A^{n-1} b U_i$. Since

$$\mathrm{cov}\{U_n t[m(\xi_n + Z_{n+1})], U_0 t(Y_1)\} = 0,$$

we find the covariance in (24) equal

$$\begin{aligned}&\mathrm{cov}\{U_n(t[m(V_{n+1})] - t[m(\xi_n + Z_{n+1})]), U_0 t(Y_1)\}\\ &= E\{U_n U_0 t(Y_1)(t[m(V_{n+1})] - t[m(\xi_n + Z_{n+1})])\}\\ &-E\{U_n(t[m(V_{n+1})] - t[m(\xi_n + Z_{n+1})])\}\\ &\qquad \cdot E\{U_0 t(Y_1)\} = S + T, \text{ say.}\end{aligned}$$

Recalling that both t and m are Lipschitz, and using (25), we get

$$\begin{aligned}&|t[m(V_{n+1})] - t[m(\xi_n + Z_{n+1})]|\\ &\le \alpha\beta |c^T A^n X_1|^p \le \alpha\beta \left\|c^T\right\| \|A^n\| \|X_1\|^p.\end{aligned}$$

Thus

$$|S| \le a\beta \left\|c^T\right\| \|A^n\| E\{\|X_1\|^p |U_0 t(Y_1)|\}$$

and

$$|T| \le a\beta \left\|c^T\right\| \|A^n\| E\{\|X_1\|^p\} E\{|U_0 t(Y_1)|\}$$

where $a = \alpha\beta \left\|c^T\right\| E|U_0|$. Denoting

$$\begin{aligned}\varphi(y) = &aE\{\|X_1\|^p |U_0| \,|Y_1 = y\}\\ &+ aE\{\|X_1\|^p\} E\{|U_0| \,|Y_1 = y\},\end{aligned}$$

we complete the proof. ■

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 nonlinear systems using correlation analysis", *Proceedings of the IEEE*, vol. 125, pp. 691-697, 1978.

[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] S. A. Billings and S. Y. Fakhouri, "Identification of nonlinear systems using correlation analysis and pseudorandom inputs", *International Journal of Systems Science*, vol. 11, pp. 261-279, 1980.

[8] S. A. Billings and S. Y. Fakhouri, "Analysis of estimation errors in the identification of nonlinear systems", *International Journal of Systems Science*, vol. 12, pp. 205-225, 1981.

[9] S. A. Billings and S. Y. Fakhouri, "Identification of systems containing linear dynamic and static nonlinear elements", *Automatica*, vol. 18, pp. 15-26, 1982.
[10] D. Brilinger, "The identification of a particular nonlinear time series system", *Biometrica*, vol. 64, pp. 509-515, 1977.
[11] 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.
[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. B. Gardiner, "Identification of processes containing single-valued non-linearities", *International Journal of Control*, vol. 18, pp. 1029-1039, 1973.
[14] W. Greblicki, "Non-parametric orthogonal series identification of Hammerstein systems", *International Journal of Systems Science*, vol. 20, pp. 2355-2367, 1989.
[15] 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.
[16] W. Greblicki and M. Pawlak, "Hammerstein system identification by non-parametric regression estimation", *International Journal of Control*, vol. 45, pp. 345-354, 1987.
[17] W. Greblicki and M. Pawlak, "Nonparametric identification of Hammerstein systems", *IEEE Transactions on Information Theory*, vol. IT-35, pp. 409-412, March 1989.
[18] W. Greblicki and M. Pawlak, "Recursive nonparametric identification of Hammerstein systems", *Journal of the Franklin Institute*, vol. 326, pp. 461-481, 1989.
[19] W. Greblicki and M. Pawlak, "Nonparametric identification of a cascade nonlinear time series", *Signal Processing*, vol. 22, pp. 61-75, 1991.
[20] R. Haber and H. Unbenhauen, "Structure identification of nonlinear systems: A survey of input-output approaches", *Automatica*, vol. 26, pp. 651-677, 1990.
[21] 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.
[22] C. Hwang and K. K. Shyu, "Series approach to the analysis and identification of discrete Hammesrtein systems", *International Journal of Control,* vol. 47, pp. 1961-1972, 1988.
[23] M. J. Korenberg and W. Hunter, "The identification of nonlinear biological systems: LNL cascade models", *Biological Cybernetics*, vol. 55, pp. 125-134, 1986.
[24] L. Liung, *System Identification: Theory for the User*, Englewood Cliffs, NJ: Prentice Hall, 1987.
[25] E. Nadaraya, "On regression estimators", *Theory Probability and its Applications*, vol. 9, pp. 157-159, 1964.
[26] 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.
[27] J. P. Norton, *An Introduction to Identification*, London: Academic Press, 1986.
[28] B. L. S. Rao, *Nonparametric Functional Estimation.* Orlando: Academic Press, 1983.
[29] I. W. Sandberg, "Expansions for nonlinear systems", *Bell Syst. Tech. J.*, vol. 61, pp. 159-199, 1982.
[30] T. Söderström and P. Stoica, *System Identification*, New York: Prentice Hall, 1989.
[31] G. S. Watson, "Smooth regression analysis", *Sankhyā Ser. A*, vol. 26, pp. 359-372, 1964.
[32] R. V. Webb, "Identification of the Volterra kernels of a process containing single-valued nonlinearities" *Electronics Letters*, vol. 110, pp. 344-346, 1974.
[33] R. L. Wheeden and A. Zygmund, *Measure and Integral*, New York: Dekker, 1977.