printed in IEEE Transactions on Circuits and Systems–I: Fundamental Theory and Applications, vol. 44, no. 6, pp. 538-545, June 1997 1

Nonparametric Approach to Wiener System Identification

Włodzimierz Greblicki, *Member, IEEE*

Abstract— **A Wiener system, i.e., a system consisting of a linear dynamic subsystem followed by a memoryless nonlinear one is identified. The system is driven by a stationary white Gaussian stochastic process and is disturbed by Gaussian noise. The characteristic of the nonlinear part can be of any form. The dynamic subsystem is asymptotically stable. The a priori information about both the impulse response of the dynamic part of the system and the nonlinear characteristics is nonparametric. Both subsystems are identified from observations taken at the input and output of the whole system. The kernel regression estimate is applied to estimate the invertible part of the nonlinearity. An estimate to recover the impulse response of the dynamic part is also given. Pointwise consistency of the first and consistency of the other estimate is shown. The results hold for any nonlinear characteristic, and any asymptotically dynamic subsystem. Convergence rates are also given.**

Keywords— **System identification, nonparametric identification, Wiener system, nonparametric regression.**

I. Introduction

DYNAMIC processes of various kind can be described as systems consisting of relatively simple elements called subsystems. The subsystems are usually linear dynamic or nonlinear memoryless. Such an approach called block oriented is widely used in many areas, e.g., in communication, see [10], industrial chemistry, [8], [25], automatic control, [28-29], biology, [7], [20]. Hammerstein and Wiener systems are most often investigated, see [2], [3], [5-6], [9],[20], [29], for the Hammerstein, and [2], [3], [19], [20], [22-23], [26], [28], for the Wiener one. The first consists of a memoryless nonlinear element followed by a linear dynamic subsystem while the other comprises the subsystems connected in reverse order. In the process of identification, on the basis of the *a priori* information and input-output measurements of the whole system, we construct mathematical descriptions of both subsystems. It means that the identification goal is to estimate the characteristic of the nonlinear subsystem as well as the impulse response of the dynamic part. The signal interconnecting subsystems is not measured and it is a source of difficulties impossible to overcome. No subsystem can be entirely identified since, because of the cascade structure, their descriptions can be estimated only up to some unknown constants.

As far as the dynamic subsystem is concerned, it appears that, in both Hammerstein and Wiener systems, the impulse response is proportional to some covariance functions and, therefore, can be easily estimated. The nonlinear subsystem identification is much more troublesome. Until not long ago, authors had assumed that the nonlinearity is a polynomial and thus had reduced the problem of function estimation to parameter estimation, see references cited above. In typical situations, however, the *a priori* knowledge about the nonlinear subsystem is extremely small. Therefore, we may often be able only to say that its characteristic is bounded or continuous, *etc.* So, the problem we often face in applications is nonparametric. The nonparametric approach to the estimation of the nonlinearity in Hammerstein systems has been introduced by Greblicki and Pawlak, [16]. Since then, more such estimates have been proposed and examined, see [11], [15], [17], [22], [24]. Classes of all possible nonlinearities admitted in mentioned papers are very wide and include, e.g., all bounded, or square integrable or Lipschitz functions. In turn, rudiment results on the nonparametric estimation of the nonlinear characteristic in Wiener systems have been lately given by Greblicki [12]. Orthogonal series estimates have been proposed in [13].

In this paper, we deal with a Wiener system. So far, all authors have assumed that the nonlinear characteristic is a polynomial of a known finite order, see Bendat [2], Billings and Fakhouri [3], Hunter and Korenberg [20]. Only recently, Greblicki, see [12, 13], have enlarged the class of possible characteristics to all those that are both differentiable, and invertible. In this paper the class of nonlinear characteristics is as ample as possible. We just make no restriction concerning the nonlinearity and, therefore, the characteristic can be of any form. Thus, the problem of its estimation is nonparametric. As far as the dynamic subsystem is concerned, it is just asymptotically stable and its order is unknown. Due to this, the problem of recovering its impulse response is also nonparametric. We present two algorithms. The first one estimates the invertible part of the nonlinear characteristic while the other recovers the impulse response of the dynamic subsystem. The algorithms are numerically independent of each other, i.e., are calculated separately. We estimate the invertible part of the nonlinearity with the help of the kernel regression estimate and show that the estimate is pointwise consistent. We also recover the impulse response of the dynamic subsystem and show that our identification algorithm is strongly consistent. Also convergence rates of the algorithms are presented.

In the light of this, the main contribution of the paper into the identification of the nonlinear subsystem is that we enlarge the class of all possible characteristics of the nonlinear subsystem as much as possible. The characteristic can have any form. Therefore, no further step can be made in this direction. Concerning the dynamic subsystem, the main result is that, contrary to the authors

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

mentioned above, we not only present an identification algorithm but also show its convergence. The only essential limitation of the paper is that the input signal has a Gaussian distribution. This restriction is, however, typical and have been made by all authors mentioned above.

II. Identification Problems

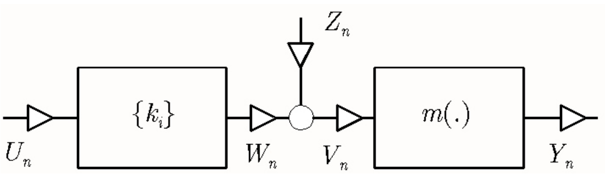

Fig. 1. The identified Wiener system.

Shown in Fig. 1 is the identified Wiener system with the input U_n and output Y_n. The system has the cascade structure and consists of two subsystems. The first is linear dynamic, has the input U_n and output W_n, and is described by the state equation:

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

$n = \cdots, -1, 0, 1, \cdots$, where X_n is the state vector. The matrix A and vectors b, c are unknown. So is the dimension of the state vector. The subsystem is asymptotically stable, i.e., all eigenvalues of A lie inside the unit circle centered at the origin. The output is disturbed by noise Z_n, which means that the input V_n of the second subsystem equals

$$V_n = W_n + Z_n.$$

The subsystem is nonlinear memoryless and has a characteristic m, which means that

$$Y_n = m(V_n).$$

The characteristic m of the second subsystem is a Borel measurable function. We want to underline that no restriction concerning its functional form is made in the paper. The whole system is driven by a stationary white stochastic process $\{U_n; n = \cdots, -1, 0, 1, \cdots\}$. Random variables U_n's have a normal distribution with zero mean and unknown variance. The noise $\{Z_n; n = \cdots, -1, 0, 1, \cdots\}$ is also a sequence of independent random variables. They have a normal distribution with zero mean and unknown variance. Moreover, the noise and the input signal are mutually independent. Owing to all that, both the state vector X_n and the output Y_n of the whole Wiener system are random variables. Stochastic processes $\{X_n; n = \cdots, -1, 0, 1, \cdots\}$ and $\{Y_n; n = \cdots, -1, 0, 1, \cdots\}$ are stationary.

Our goal is to identify both subsystems from input-output observations (U_0, Y_0), (U_1, Y_1), $\cdots$, (U_n, Y_n) of the whole Wiener system. Signals W_n and V_n are not measured. As far as the dynamic subsystem is concerned, we want to recover its impulse response $\{k_i; i = 0, 1, 2, \cdots\}$, where $k_0 = 0$, and $k_i = c^T A^{i-1} b$, for $i = 1, 2, \cdots$ .With respect to the nonlinear subsystem, our objective is to estimate its characteristic m.

Observe that the problem of the identification of the first subsystem is nonparametric since its order, i.e., the dimension of the state vector, is unknown. The same is true for the other subsystem, since the class of all possible characteristics, i.e., the family of all Borel measurable functions can't be represented in a parametric form.

Now, it should be noticed that, because of the cascade structure of the system, we can recover both the impulse response and the nonlinearity only up to some unknown constants. To observe that fact, consider two Wiener systems in which $Z_n = 0$, for all n. The first has the dynamic part having the impulse response $\{k_i; i = 0, 1, 2, \cdots\}$ and the nonlinear characteristic $m(.)$, while the other consists of the dynamic subsystem $\{\alpha k_i; i = 0, 1, 2, \cdots\}$ any $\alpha > 0$, and the nonlinearity $m(./\alpha)$. To every input signal, both systems respond identically and due to this fact can't be distinguished from each other on the basis of input-output observations.

III. Identification Algorithms

We shall now present algorithms identifying the impulse response of the dynamic subsystem, and the nonlinear characteristic of the memoryless one. By σ_V^2 and σ_U^2 we denote variance of V_n and U_n, respectively. We define the following two constants: $\alpha = (\sigma_U^2/\sigma_V^2)E\{V_0 m(V_0)\}$, and $\beta = (\sigma_U^2/\sigma_V^2)k_1$. Denote, moreover, $\mu(y) = E\{U_0 \,|Y_1 = y\}$. We shall now present a simple but unexpectedly useful lemma.

Lemma 3.1: For $i = 1, 2, \cdots$,

$$E\{U_0 \,|V_i\} = (\sigma_U^2/\sigma_V^2)k_i V_i.$$

Proof: Since $V_i = Z_i + \sum_{j=0}^{i} k_j U_{i-j}$, the pair (U_0, V_i) has a correlation coefficient $\rho_i = (\sigma_U/\sigma_V)k_i$ Because its distribution is Gaussian, we get $E\{U_0 \,|V_i\} = (\sigma_U/\sigma_V)\rho_i V_i$, which has been claimed. ∎

A. The linear subsystem

We begin with the following

Lemma 3.2: Let $E|V_0 m(V_0)| < \infty$. For $i = 1, 2, \cdots$,

$$E\{U_0 Y_i\} = \alpha k_i.$$

Proof: We have, $E\{U_0 Y_i \,|V_i\} = E\{U_0 m(V_i) \,|V_i\}$. From Lemma 3.1, it follows that the last quantity equals $(\sigma_U^2/\sigma_V^2)E\{V_i Y_i\}k_i$. The lemma follows. ∎

Having proved Lemma 3.2, we propose the following algorithm to recover the impulse response of the dynamic subsystem, i.e., to estimate αk_i:

$$\hat{\kappa}_i = \frac{1}{n}\sum_{j=1}^{n} U_j Y_{i+j}. \tag{3.1}$$

The algorithm is just a natural estimate of $E\{U_0 Y_i\}$. We shall later show that $\hat{\kappa}_i$ consistently estimates αk_i, i.e., recovers the impulse response up to an unknown constant α.

B. The nonlinear subsystem

In turn, the nonlinear subsystem is identified with the following algorithm:

$$\hat{\mu}(y) = \frac{\sum_{i=1}^{n} U_i K\left(\frac{y - Y_{i+1}}{h(n)}\right)}{\sum_{i=1}^{n} K\left(\frac{y - Y_{i+1}}{h(n)}\right)}, \tag{3.2}$$

where K and $\{h(n)\}$ are a kernel function and a number sequence, respectively. Both are suitably selected. In the definition of the estimate and throughout the paper, $0/0$ is treated as zero. We shall later show that $\hat{\mu}(y)$ is a consistent estimate of $\mu(u)$, i.e. the regression $E\{U_0\,|Y_1 = y\}$. Now we show the relation between the regression and the characteristic m of the nonlinear subsystem and, thus, explain the idea standing behind the algorithm.

Observe that Lemma 3.1 implies $E\{U_0\,|V_1\} = \beta V_1$. This, together with an obvious $E\{U_0\,|Y_1\} = E\{E\{U_0\,|V_1\}\,|Y_1\}$, yields $\mu(y) = \beta E\{V_0\,|Y_0 = y\}$. Therefore, (3.2) estimates $\beta E\{V_0\,|Y_0 = y\}$. We shall now demonstrate how much information about m is in $E\{V_0\,|Y_0 = y\}$. For that sake, denote by M the image of a real line R under the mapping m. Assume for a while that m is invertible in the Cartesian product $R \times M$. It means that, for every $y \in R$ there exists a unique $v \in R$ such that $m(v) = y$. Since m is invertible, $E\{V_0\,|Y_0 = y\} = m^{-1}(y)$, every $y \in M$. Consequently, $\mu(y) = \beta m^{-1}(y)$, every $y \in M$. In the assumed case, algorithm (3.2) just estimates $\beta m^{-1}(y)$. Therefore, the inverse of the characteristic is recovered up an unknown multiplicative constant β.

In general, m may be not invertible. Denote now by D the set of all y's having the property that, for every $y \in D$, there exists a unique v such that $m(v) = y$. Obviously, D is a subset of M. The part of the mapping m belonging to the product $D^{-1} \times D$, where D^{-1} is the inverse image of D, is clearly invertible in the product. Denote the part, i.e., the invertible part of m, by m_D. Hence, $E\{V_0\,|Y_0 = y\} = m_D^{-1}(y)$, all $y \in D$. Thus, $\mu(y) = \beta m_D^{-1}(y)$, all $y \in D$. In this way, we have proved

Lemma 3.3: For any m,

$$E\{U_0\,|Y_1 = y\} = \beta m_D^{-1}(y), \text{ all } y \in D.$$

For m invertible in $R \times M$,

$$E\{U_0\,|Y_1 = y\} = \beta m^{-1}(y), \text{ all } y \in M.$$

As we have already mentioned, (3.2) estimates the regression $E\{U_0\,|Y_1 = y\}$. From this and the lemma, it follows that the estimate recovers $\beta m_D^{-1}(y)$, i.e. the invertible part of m up to an unknown multiplicative constant β. For m invertible, (3.2) just estimates $\beta m^{-1}(y)$, i.e., the inversion of m.

To see what we can and what can't recover, observe that a characteristic equal to $v + 1$ or zero or v, for $v \le -1$, or $-1 < v < 1$, or $v \ge 1$, respectively, is obviously not invertible. We are, however, able to recover the inversion of its invertible part in the set $D = (-\infty, 0) \cup [1, \infty)$.

IV. Convergence of the Dynamic Subsystem Identification Algorithm

The proof of consistency of the estimate of the impulse response of the dynamic subsystem is not difficult when we apply ergodic theory arguments.

Theorem 4.1: Let $E|V_0 m(V_0)| < \infty$. Then, for $i = 1, 2, \cdots$,

$$\hat{\kappa}_i \to \alpha k_i \text{ as } n \to \infty \text{ almost surely.}$$

Proof: The theorem is a consequence of the fact that the process $\{V_n; n = \cdots, -1, 0, 1, \cdots\}$ is ergodic, see Hannan [18, Ch.4, Theorem 3]. ■

To examine the convergence rate of the algorithm we assume that m satisfies a Lipschitz condition, i.e., that

$$|m(u) - m(v)| \le \delta |u - v| \tag{4.1}$$

some δ, all $u, v \in R$. Since

$$\operatorname{var} \hat{\kappa}_i = n^{-2} \sum_{j=0}^{n-1} (n - j) \operatorname{cov}[U_0 Y_i, U_j Y_{i+j}],$$

an application of Lemma A.1 in Appendix A leads to $\operatorname{var} \hat{\kappa}_i = O(1/n)$, $i = 1, 2, \cdots$. Taking into account the fact that the estimate is unbiased, we get

$$E(\hat{\kappa}_i - \alpha k_i)^2 = O(1/n),$$

$i = 1, 2, \cdots$. The rate is not worse than n^{-1}, i.e., the one typical for the parametric inference.

V. Convergence of the Nonlinear Subsystem Identification Algorithm

As we have already mentioned, in the estimate of the nonlinear characteristic, both K and $h(n)$ should be suitably chosen. The kernel K is a nonnegative Borel measurable function satisfying the following restrictions:

$$\sup_{y \in R} K(y) = k < \infty, \tag{5.1}$$

$$c I_{\{|y| \le r\}} \le K(y), \tag{5.2}$$

some positive c and r, all $y \in R$, where I is an indicator function, i.e., where $I_A(y)$ equals 1 or zero according to $y \in A$ or $y \notin A$, respectively. It means that the kernel is bounded from below by a rectangular function of the height c, the width $2r$, located at zero. In addition, the kernel satisfies a Lipschitz condition, i.e.,

$$|K(x) - K(y)| \le \kappa |x - y|, \tag{5.3}$$

some κ, all $x, y \in R$. There are kernels satisfying above restrictions, e.g., a triangular one equal to $1 - |y|$ or zero according to $|y| \le 1$ or $|y| > 1$, respectively. Another example is a parabolic kernel, which equals $1 - y^2$ or zero for $|y| \le 1$ or $|y| > 1$, respectively.

The number sequence $\{h(n)\}$ is selected to satisfy the following conditions:

$$h(n) \to 0 \text{ as } n \to \infty, \tag{5.4}$$

$$nh^2(n) \to \infty \text{ as } n \to \infty. \tag{5.5}$$

As the number sequence, one can chose $h(n) = cn^{-\alpha}$, $c > 0$, $0 < \alpha < 1/2$.

In the paper, λ denotes the probability measure of Y_n's. Since m can be any Borel measurable function, the measure may not possess a density.

Let $y \in D$ be such that m satisfies a Lipschitz condition in a neighborhood of v, where $v = m_D^{-1}(y)$. It means that there exists a neighborhood of v such that $|m(v) - m(u)| \le \rho_v |v-u|$, some ρ_v depending on v, all u belonging to the neighborhood. Let D_L denote the set of all such y's. We shall now show that, for almost every $y \in D_L$, algorithm (3.2) consistently estimates the inversion of the invertible part of m up to a multiplicative constant β, i.e., estimates $\beta m_D^{-1}(y)$.

Lemma 5.1: Let the nonnegative kernel K have bounded support and satisfy (5.1)-(5.3). Let the number sequence $\{h(n)\}$ satisfy (5.4) and (5.5). Then, for any m,

$$\hat{\mu}(y) \to \mu(y) \text{ as } n \to \infty \text{ in probability,}$$

almost every (λ) $y \in D_L$.

Proof: We have $\hat{\mu}(y) = \hat{g}(y)/\hat{f}(y)$, where

$$\hat{g}(y) = \sum_{i=1}^{n} U_i K\left(\frac{y - Y_{i+1}}{h(n)}\right) \Big/ nEK\left(\frac{y - Y_{i+1}}{h(n)}\right) \tag{5.6}$$

and

$$\hat{f}(y) = \sum_{i=1}^{n} K\left(\frac{y - Y_{i+1}}{h(n)}\right) \Big/ nEK\left(\frac{y - Y_{i+1}}{h(n)}\right),$$

respectively. Observing,

$$E\hat{g}(y) = \sum_{i=1}^{n} E\left\{\mu(Y_0) K\left(\frac{y - Y_0}{h(n)}\right)\right\} \Big/ EK\left(\frac{y - Y_0}{h(n)}\right)$$

and applying Lemma B1 given in the Appendix B, we get

$$E\hat{g}(y) \to \mu(y) \text{ as } n \to \infty, \tag{5.7}$$

almost every (λ) $y \in R$.

In turn, $\operatorname{var}[\hat{g}(y)] = Q_1(y) + Q_2(y)$, where

$$Q_1(y) = \frac{1}{nE^2K\left(\dfrac{y - Y_0}{h(n)}\right)} \operatorname{var}\left[U_0 K\left(\frac{y - Y_1}{h(n)}\right)\right],$$

and where $Q_2(y)$ equals

$$\frac{1}{n^2E^2K\left(\dfrac{y - Y_0}{h(n)}\right)} \times \sum_{i=1}^{n-1} (n-i) \operatorname{cov}\left[U_0 K\left(\frac{y - Y_1}{h(n)}\right), U_i K\left(\frac{y - Y_{i+1}}{h(n)}\right)\right].$$

We easily find

$$|Q_1(y)| \le \frac{k}{nh(n)} \frac{h(n)}{EK\left(\dfrac{y - Y_0}{h(n)}\right)} \frac{E\left\{\psi(Y_0) K\left(\dfrac{y - Y_0}{h(n)}\right)\right\}}{EK\left(\dfrac{y - Y_0}{h(n)}\right)},$$

where $\psi(y) = E\{U_0^2 | Y_1 = y\}$. Applying Lemmas B1 and B2 in Appendix B, we find $Q_1(y) = O(1/nh(n))$, almost every (λ) $y \in R$.

In order to examine Q_2, fix y and suppose that $y \in D_L$. By virtue of Lemma A.2 in Appendix A, for h small enough, its absolute value of is not greater than

$$\frac{k\delta_y}{n^2h^2(n)} \frac{h(n)}{EK\left(\dfrac{y - Y_0}{h(n)}\right)} \times \frac{E\left\{\phi(Y_0) K\left(\dfrac{y - Y_0}{h(n)}\right)\right\}}{EK\left(\dfrac{y - Y_0}{h(n)}\right)} \sum_{i=1}^{n-1} (n-i) \left\| c^T A^i \right\|,$$

some δ_y, some integrable ϕ. Recalling that A is asymptotically stable and applying Lemmas B1 and B2 in Appendix B again, we come to the conclusion that $Q_2(y) = O(1/nh^2(n))$, almost every (λ) point y. Therefore,

$$\operatorname{var} \hat{g}(y) = O(1/nh^2(n)) \tag{5.8}$$

at almost every (λ) assumed y. Recalling (5.7), we find

$$\hat{g}(y) \to \mu(y) \text{ as } n \to \infty \text{ in probability}$$

at almost every (λ) $y \in D_L$. Since similar convergence of $\hat{f}(y)$ to 1 can be now easily verified, the proof has been completed. ∎

Combining the lemma with Lemma 3.3, we get

Theorem 5.1: Let the nonnegative kernel K have bounded support and satisfy (5.1)-(5.3). Let the number sequence $\{h(n)\}$ satisfy (5.4) and (5.5). Then, for any m,

$$\hat{\mu}(y) \to \beta m_D^{-1}(y) \text{ as } n \to \infty \text{ in probablity}$$

at almost every (λ) $y \in D_L$. In particular, for m invertible in $R \times M$,

$$\hat{\mu}(y) \to \beta m^{-1}(y) \text{ as } n \to \infty \text{ in probablity}$$

at almost every (λ) $y \in D_L$.

Notice that Theorem 5.1 given above holds for any m and any asymptotically stable dynamic subsystem. Convergence, however, takes place at some specified points, i.e., at almost every (λ) $y \in D_L$, i.e., at almost every (λ) $y \in D$ having the property that m is a Lipschitz function in a neighborhood of a point v, where $v = m_D^{-1}(y)$. Assume now, moreover, that m has a nonzero derivative at the point v, i.e., that $m'(v) \neq 0$. Since $(d/dy)m_D^{-1}(y) = 1/m'(v)$ at $v = m_D^{-1}(y)$, we have $|m_D^{-1}(y) - m_D^{-1}(x)| \le c_y |y - x|$,

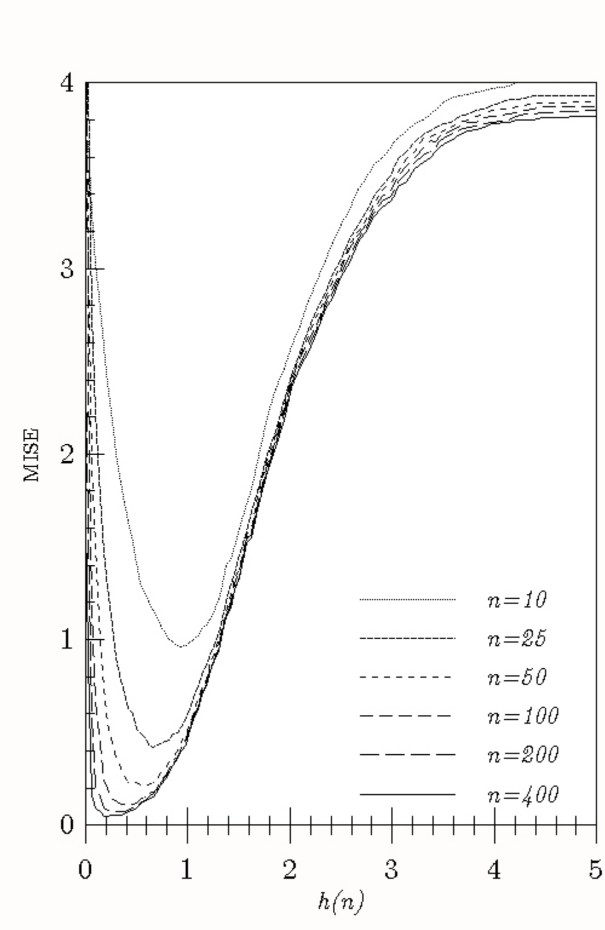

Fig. 2. MISE versus *h(n)*.

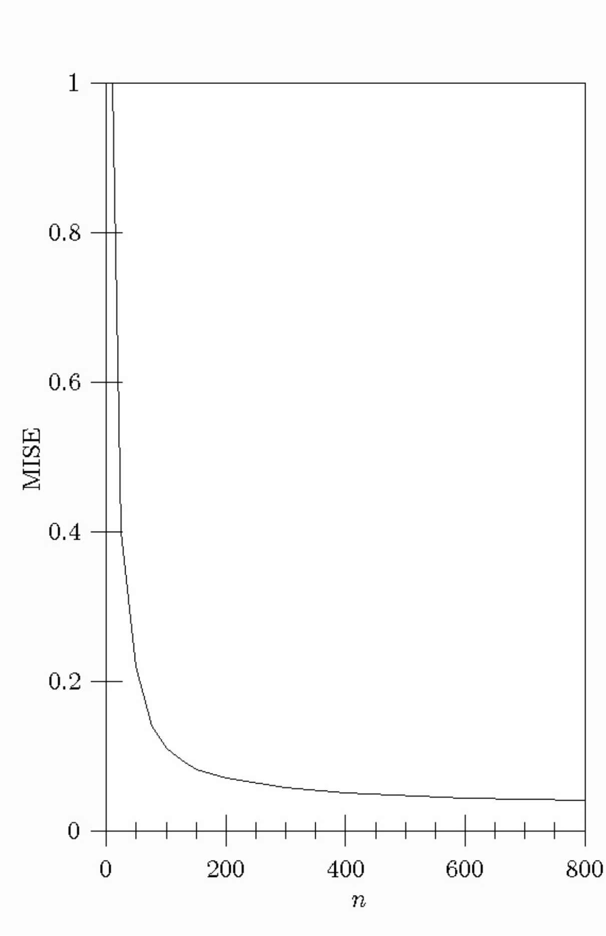

Fig. 3. Optimal MISE versus n.

some c_y, for all x in a neighborhood of the y. Therefore, $|\mu(y) - \mu(x)| \leq c_y \beta |y - x|$ in the neighborhood. Consequently

$$|E\hat{g}(y) - \mu(y)| \leq \beta c_y h(n) \frac{E\left\{\dfrac{|y - Y_0|}{h(n)} K\left(\dfrac{y - Y_0}{h(n)}\right)\right\}}{EK\left(\dfrac{y - Y_0}{h(n)}\right)}$$

at the point y, where $\hat{g}$ is the function defined in the proof of Theorem 5.1. Since K has bounded support, $|x|K(x) \leq \alpha K(x)$, some α, all x, and we can bound the above quantity by $\alpha\beta c_y h(n)$. Hence $|E\hat{g}(y) - \mu(y)| = O(h(n))$ at the point. From this and (5.8), it follows that

$$E(\hat{g}(y) - \mu(y))^2 = O(h^2(n)) + O(1/nh^2(n))$$

at almost every (λ) such a point. Thus, selecting $h(n) \sim n^{-1/4}$, we get $E(\hat{g}(y) - \mu(y))^2 = O(n^{-1/2})$ at almost every such (λ) a point. Since the rate is achieved also by $\hat{f}$, recalling (3.3), we get finally

$$\hat{\mu}(y) - \beta m_D^{-1}(y) = O(n^{-1/4}) \text{ in probability}$$

at almost every (λ) such a point. Observe that the result holds for any m. The rate is somewhat worse than $n^{-1/2}$, i.e., the rate usually encountered in parametric inference problems.

VI. Simulation Example

To illustrate the behavior of algorithm (3.2), we present results of a simulation example. In the example, the system is described by the following equations: $W_{n+1} = (1/2)W_n + U_n$, $V_n = W_n + Z_n$, $Y_n = m(V_n)$, where $m(v)$ equals $v - 1/2$, or $v + 1/2$, for $v \leq 0$ or $v > 0$, respectively. The input signal has zero mean and variance 1. The noise has also zero mean but its variance equals 0.1. The triangular kernel with support [-1,1], and the height 1 has been applied. The mean integrated square error (MISE in brief) has been numerically calculated in the interval [-1.5,1.5] to which fall as much as 90% of observations.

The error has been calculated for $n = 10$, 25, 50, 100, 200, 400, and 800. From results presented in Fig.2, we conclude that the error is sensitive to the bandwidth $h(n)$ and that its suitable selection is very important. In particular, $h(n)$ smaller than optimal should be avoided at every cost. The problem of the optimal choice of $h(n)$ gets more important for large n since the error becomes more sensitive to $h(n)$ when the number of observations increases. For the bandwidth parameter selected optimally, the MISE decreases at the highest possible speed, see Fig.3. In the example, finding the best value of $h(n)$ has not been difficult.

If, however, the situation examined in the example were real, the bandwidth parameter would be more or less distant from the optimum and the error would decrease slower.

VII. Discussion

Examining nonparametric regression estimate (3.2), we use the state vector description of the dynamic subsystem. An alternative way for the analysis of nonparametric regression estimates derived from dependent data is based on the assumption that the underlying stochastic process is mixing in some way, e.g., uniformly, see [1] and [4]. Authors of mentioned papers have been interested in $E\{Y_n|Y_{n-1},\cdots,Y_{n-k}\}$, some k, i.e., in prediction rather than system identification. Nevertheless, we shall briefly compare consequences of their methodology with those caused by ours. When applied to the analysis of dynamic systems, their approach has some important disadvantage since it covers a considerably smaller class of linear dynamic subsystems and consequently a smaller class of Wiener systems.

To demonstrate this fact, observe that we infer from the stochastic process $\{(U_n,Y_n); n = \cdots,-1,0,1\cdots\}$, where $\{U_n; n=\cdots,-1,0,1\cdots\}$ is white, Gaussian, and where $\{Y_n; n = \cdots,-1,0,1\cdots\}$ is the result of a nonlinear transformation of a Gaussian process $\{V_n; n = \cdots,-1,0,1\cdots\}$ having the autocorrelation function $R_{VV}(n)=\sum_{i=0}^{\infty}k_ik_{i+n}$, $n=0,\pm1,\pm2,\cdots$. For simplicity, we have assumed here that the noise Z_n just equals zero. It is known that a Gaussian process satisfies the uniform mixing condition mentioned above if and only if its autocorrelation function is identically zero for all $|n|$ larger than some n_0, see Ibragimov and Linnik [21, Theorem 17.3.2]. Therefore, $\{V_n; n=\cdots,-1,0,1\cdots\}$ is a uniformly mixing process if and only if $R_{VV}(n)$ equals zero, for $|n|\geq n_0$, some n_0, i.e., if and only if the impulse response $\{k_n\}$ of the dynamic subsystem is zero for all n larger than n_0. This is the case if and only if the dynamic subsystem has a finite memory. The class of all such subsystems includes, e.g., all MA but only a small amount of ARMA processes. For example, in a system with the dynamic part described by a differential equation $W_{n+1}-(1/2)W_n=U_n$, the process $\{V_n; n=\cdots,-1,0,1\cdots\}$ is not uniformly mixing since k_n equals 0 or $n^{-1/2}$, for $n=0$ or $n=1,2,\cdots$, respectively. Thus, the class of linear dynamic systems considered by them is rather narrow while our results hold for any asymptotically stable dynamic subsystem described by a state equation.

Having Lemma 3.2, we have proposed algorithm (3.1) to identify the dynamic subsystem. Using the correlation function notation, we can rewrite the equality in the lemma in the following alternative form:

$$R_{UY}(n)=(\sigma_U^2/\sigma_V^2)E\{V_0m(V_0)\}k_n, \tag{6.1}$$

where R_{UY} is the appropriate correlation function. This formula is not new and has been already reported for continuous-time systems, see Billings and Fakhouri [3]. Their result has been given, however, for m being a polynomial only, while our Lemma 3.2 holds for any m.

Lemma 3.2 can be derived also from Brillinger's observation, see [5]. Due to his result, for any two Gaussian random variables U, V and any Borel function m, we have $\operatorname{cov}[U,m(V)] = (1/\sigma_U^2)\operatorname{cov}[V,m(V)]\operatorname{cov}[U,V]$, provided that $E|Vm(V)|<\infty$. To obtain the lemma it suffices to substitute U and V for U_0 and V_n, respectively.

Finally, we want to mention that the lemma has a bit in common with Busgang's theorem, see Bendat [2, Ch.2, Theorem 3], due to which

$$R_{VY}(n)=(1/\sigma_V^2)R_{VV}(n)E\{V_0m(V_0)\}$$

or equivalently

$$\sum_{i=0}^{\infty}k_iR_{UY}(n+i) = (\sigma_U^2/\sigma_V^2)E\{V_0m(V_0)\}\sum_{i=0}^{\infty}k_ik_{n+i}. \tag{6.2}$$

Observe that our (6.1) implies Busgang (6.2) but not conversely. In [2], moreover, Busgang's theorem has been proved only for a particular case in which Y_n possesses a density, i.e., for a sufficiently smooth m.

VIII. Conclusion

In the paper, we have identified both the linear dynamic and nonlinear memoryless subsystems. The problem of recovering the impulse response of the linear part is nonparametric since the order of the part is unknown. The fact that the characteristic of the nonlinear subsystem can be of any form makes the problem of its estimation also nonparametric. Observe that our restrictions imposed on the subsystem can't be weakened in any respect. Our Theorem 4.1 shows that estimate (3.1) recovers the impulse response of the dynamic subsystem. No similar result is known to the author. In turn, Theorem 5.1 improves results given in [12], and [13] since imposes no restriction on the nonlinear characteristic. In the mentioned papers, m has been assumed both differentiable, and invertible.

We have also presented convergence rates of the estimates. As far as the impulse response estimate is concerned, its rate is not worse than that n^{-1}, i.e., the rate typical in parametric inference. The rate obtained for the nonlinear characteristic estimate is somewhat worse, which is not a disappointment since the *a priori* information about the memoryless subsystem in extremely poor.

The essential feature of the approach presented in the paper is that it works for $\{W_n; n=0,1,2,\cdots\}$ being any ARMA stochastic process. Owing to that, our convergence results hold for any asymptotically stable dynamic subsystem. An alternative approach examined mainly in the statistical literature is based on the assumption that the process is mixing. As we have shown in section VII, the class of Wiener systems which can be treated in this way is very narrow, and not very interesting. The approach presented in the paper covers all asymptotically stable subsystems, and, therefore, is much more interesting from the practical viewpoint. Thus, it is worth further studies.

Appendix A. The Wiener System

All results given in this appendix deal with the Wiener system examined in the paper.

Lemma A.1: Let m satisfy (4.1). Then, for $0 < p < n$,

$$|\mathrm{cov}[U_0 Y_p, U_n Y_{p+n}]| \leq \gamma \|A^n\|,$$

some γ independent of both n and p.

Proof: Let $0 < p < n$. As $V_{p+n} = c^T A^n X_p + \zeta_{p+n}$, where

$$\zeta_{p+n} = \sum_{i=0}^{n-1} c^T A^{n-i-1} b U_{p+i} + Z_{p+n},$$

we observe $\mathrm{cov}[U_0 m(V_p), U_p m(\zeta_{p+n})] = 0$. Hence, the examined variance equals

$$\mathrm{cov}\{U_0 m(V_p), U_n [m(V_{p+n}) - m(\zeta_{p+n})]\}.$$

This, in turn, equals

$$\begin{aligned} &E\{U_0 U_n m(V_p)[m(V_{p+n}) - m(\zeta_{p+n})]\} \\ &- E\{U_0 m(V_p)\} E\{U_p [m(V_{p+n}) - m(\zeta_{p+n})]\} \\ &= Q_1 - Q_2, \end{aligned}$$

say. Since m satisfies inequality (4.1),

$$|m(V_{p+n}) - m(\zeta_{p+n})| \leq \delta |V_{p+n} + \zeta_{p+n}|,$$

which is not greater than $\delta \|c^T A^n X_p\|$. It is now easy to verify that the absolute value of Q_1 does not exceed $\alpha_p \|A^n\|$, where

$$\alpha_p = \delta \|c\| E\{|U_0 U_n m(V_{p+n})| \|X_p\|\}.$$

Applying Cauchy's inequality, we can easily bound α_p by a constant independent of p. Since similar reasoning holds also for Q_2, we complete the proof. ■

Lemma A.2: Let m be any Borel measurable function. Let a nonnegative Borel measurable kernel K have bounded support and satisfy (5.3). Then, for every $y \in D_L$, there exists h_0 such that, for $0 < h < h_0$,

$$\begin{aligned} &\left| \mathrm{cov}\left[U_0 K\left(\frac{y - Y_1}{h}\right), U_n K\left(\frac{y - Y_{n+1}}{h}\right)\right]\right| \\ &\leq (\kappa \delta_y / h) \left\|c^T A^n\right\| E\left\{\phi(Y_0) K\left(\frac{y - Y_0}{h}\right)\right\}, \end{aligned}$$

some δ_y depending on y, some function ϕ independent of n. Moreover, ϕ is nonnegative and $E\phi(Y_0) < \infty$.

Proof: Let $y \in D_L$ and let $v = m_D^{-1}(y)$. Therefore, there exists a positive ε such that $|m(\xi) - m(\zeta)| \leq \delta_y |\xi - \zeta|$, some δ_y depending on y, all ξ, ζ, belonging to $[v - \varepsilon, v + \varepsilon]$. Define

$$\bar{m}(\xi) = \begin{cases} m(v - \varepsilon), & \text{for} \quad \xi \leq v - \varepsilon \\ m(\xi), & \text{for} \quad v - \varepsilon < \xi < v + \varepsilon \\ m(v + \varepsilon), & \text{for} \quad v + \varepsilon \leq \xi, \end{cases}$$

and observe that $\bar{m}$ satisfies a Lipschitz condition in the whole real line, i.e., that $|\bar{m}(\xi) - \bar{m}(\zeta)| \leq \delta_y |\xi - \zeta|$, all ξ, ζ, in R. Obviously, all values of $\bar{m}$ lie in the interval $A = [y - \varepsilon\delta_y, y + \varepsilon\delta_y]$. Since K has bounded support, there exists a positive h_0 such that, for $h < h_0$, support of $K(./h)$ is a subset of $(-\varepsilon\delta_y, \varepsilon\delta_y)$. Hence, support of $K((y - .)/h)$ is a subset of A. Consequently,

$$K\left(\frac{y - m(\xi)}{h}\right) = K\left(\frac{y - \bar{m}(\xi)}{h}\right),$$

all ξ in R. Thus, for $0 < h < h_0$,

$$\begin{aligned} &\left| K\left(\frac{y - m(\xi)}{h}\right) - K\left(\frac{y - m(\zeta)}{h}\right)\right| \\ &= \left| K\left(\frac{y - \bar{m}(\xi)}{h}\right) - K\left(\frac{y - \bar{m}(\zeta)}{h}\right)\right|, \end{aligned}$$

which is bounded by $(\kappa/h)|\bar{m}(\xi) - \bar{m}(\zeta)| \leq (\kappa\delta_y/h)|\xi - \zeta|$, all ξ, ζ, in R. Finally, we have shown that, there exists a positive h_0 such that, for any positive $h < h_0$,

$$\begin{aligned} &\left| K\left(\frac{y - m(\xi)}{h}\right) - K\left(\frac{y - m(\zeta)}{h}\right)\right| \\ &\leq (\kappa\delta_y/h) |\xi - \zeta|, \end{aligned} \tag{A.1}$$

some δ_y depending on y, all, ξ, ζ in R, every $y \in D_L$.

Obviously, $V_{n+1} = c^T A^n X_1 + \xi_{n+1}$, where

$$\xi_{n+1} = \sum_{i=1}^{n} c^T A^{n-i} b U_i + Z_{n+1}.$$

Since

$$\mathrm{cov}\left[U_0 K\left(\frac{y - Y_1}{h}\right), U_n K\left(\frac{y - m(\xi_{n+1})}{h}\right)\right] = 0,$$

we find the examined covariance equal to

$$\mathrm{cov}\left[U_0 K\left(\frac{y - Y_1)}{h}\right), U_n \Phi_h(y)\right],$$

where

$$\Phi_h(y) = K\left(\frac{y - m(V_{n+1})}{h}\right) - K\left(\frac{y - m(\xi_{n+1})}{h}\right).$$

In turn, this quantity equals

$$\begin{aligned} &E\left\{U_0 U_n \Phi_h(y) K\left(\frac{y - Y_1}{h}\right)\right\} \\ &- E\{U_n \Phi_h(y)\} E\left\{U_0 K\left(\frac{y - Y_1}{h}\right)\right\}. \end{aligned} \tag{A.2}$$

Since, by virtue of (A.1), $|\Phi_h(y)| \leq (\delta_y/h) \left|c^T A^n X_1\right|$, we find the absolute value of the first term in (A.2) not greater than

$$(\kappa\delta_y/h) \left\|c^T A^n\right\| E\{|U_0|\} E\left\{\|X_1\| |U_0| K\left(\frac{y - Y_1}{h}\right)\right\},$$

which equals

$$(\kappa\delta_y/h) \left\|c^T A^n\right\| E\{|U_0|\} E\left\{\phi_1(Y_1) K\left(\frac{y - Y_1}{h}\right)\right\},$$

where $\phi_1(y) = E\{|U_0| \|X_1\| |Y_1 = y\}$. For similar reasons the absolute value of the second term does not exceed

$$\begin{aligned} &(\kappa\delta_y/h) \left\|c^T A^n\right\| E\{|U_0|\} E\{\|X_0\|\} \\ &\times E\left\{\phi_2(Y_1) K\left(\frac{y - Y_1}{h}\right)\right\}, \end{aligned}$$

where $\phi_2(y) = E\{|U_0| |Y_1 = y\}$. The proof has been completed. ■

Appendix B. General Results

Results given in the appendix are of general character. By λ, we denote the probability measure of an arbitrary random variable X.

Lemma B.1: Let a nonnegative Borel measurable kernel K have bounded support and satisfy (5.1)-(5.3). Let ϕ be a Borel measurable function such that $E|\phi(X)| < \infty$. Then

$$\lim_{h\to 0} \frac{E\left\{\phi(X)K\left(\dfrac{x-X}{h}\right)\right\}}{EK\left(\dfrac{x-X}{h}\right)} = \phi(x),$$

for almost all (λ) $x \in R$.

Proof: The lemma is a consequence of Lemma 1 in Greblicki *et al.* [14]. ■

Lemma B.2: Let a nonnegative Borel measurable kernel K satisfy (5.2). Then

$$\limsup_{h\to 0} \frac{h}{EK\left(\dfrac{x-X}{h}\right)}$$

is finite for almost all (λ) $x \in R$.

Proof: By virtue of (5.2), $h/E\{K((x-X)/h)\}$ does not exceed $h/c\lambda(S_{rh}(x))$. This quantity equals $\upsilon(S_{rh}(y))/2c\lambda(S_{rh}(x))$, where $S_r(x)$ is a sphere of radius r centered at x, and where υ is the Lebesgue measure. Invoking the fact that $\upsilon(S_h(x))/\lambda(S_h(x))$ has a finite limit as h tends to zero, almost all (λ) $x \in R$, see Wheeden and Zygmund [27, Corollary 10.50], we complete the proof. ■

References

[1] J. Baek, and T. E. Wehrly, "Kernel estimation for additive models under dependence", *Stochastic Processes and their Applications*, vol. 47, pp. 95-112, 1993.

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

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

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

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

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

[7] 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.

[8] 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.

[9] P. G. Gallman, "A comparison of two Hammerstein model identification algorithms", *IEEE Transactions on Automatic Control*, vol. 21, pp. 124-126, 1966.

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

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

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

[13] W. Greblicki, "Nonparametric identification of Wiener systems by othogonal series", *IEEE Transactions on Automatic Control*, vol. 39, pp. 2077-2086, 1994.

[14] W. Greblicki, A. Krzyżak, and M. Pawlak, "Distribution-free pointwise consistency of kernel regression estimate", *Annals of Statistics*, vol. 12, pp. 1570-1575, 1984.

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

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

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

[18] E. J. Hannan, *Multiple Time Series*, New York: Wiley, 1970.

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

[20] 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.

[21] I. A. Ibragimov, and Yu. V. Linnik, *Independent and Stationary Sequences of Random Variables*, The Netherlands: Wolters-Noordhoff Publishing Groningen, 1971.

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

[23] H. E. Liao, and W. A. Sethares, "Suboptimal identification of nonlinear ARMA models using an orthogonality approach", *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, vol. 42, no. 1, pp. 14-22, 1995.

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

[25] R. K. Pearson, "Nonlinear input/output modeling", *Control of Chemical Processes, Preprints IFAC Symposium*, Kyoto, Japan, May 25-27, 1994, pp. 1-15, 1994.

[26] 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.

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

[28] T. Wigren, "Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model", *IEEE Transactions on Automatic Control*, vol. 39, pp. 2191-2206, 1994.

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

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

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