printed in IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 56, NO. 7, 3575-3580, JULY 2010

Nonparametric input density-free estimation of the nonlinearity in Wiener systems

Włodzimierz Greblicki

***Abstract*—The nonlinear characteristic of a Wiener system is estimated under nonparametric *a priori* information. The probability density of the input signal is completely unknown and can be of any form. The estimate is asymptotically biased but its bias can be made arbitrarily small. Results of computer simulations are presented.**

***Index Terms*—Nonparametric estimation, nonparametric identification, nonparametric regression, system identification, Wiener system.**

I. INTRODUCTION

RECOVERING the nonlinearity in a Wiener system, i.e., a system in which a linear dynamic part is followed by a nonlinear memoryless one, is not easy. Two types of the problem are examined in the literature. In first, the *a priori* information about both subsystems is parametric which means that the nonlinear characteristic is a function known up to some coefficients and is, e.g., a polynomial, [1], [2], or a piece-wise function, [3], [4]. In the other, the information is nonparametric. In such a situation, the inverse of the nonlinear characteristic has been estimated, provided that the probability density of the input signal is Gaussian, [5], [6], [7], [8], as well as [9]. It is worth mentioning that some authors have recently examined a problem by parts parametric and nonparametric, see [10], [11].

Our problem is entirely nonparametric. Contrary to all mentioned relevant papers, we estimate the characteristic itself, not its inverse. Another advantage of our results is that they are input density-free which means that they hold for any density of the input signal, not only for Gaussian. A drawback is that our estimate is asymptotically biased. Nevertheless, the bias can be made arbitrarily small.

II. THE PROBLEM

The Wiener system with scalar input and output, see Fig. 1, consists of a linear dynamic subsystem followed by a nonlinear memoryless one. The dynamic part is described by the following state equation:

$$\left.\begin{array}{rl} X_n = & AX_{n-1} + bU_{n-1} \\ V_n = & c^T X_n \end{array}\right\},$$

where X_n is the state vector at time n. Neither matrix A, assumed stable, nor vectors b or c are known. By $\{k_n\}$ we denote the impulse response of the subsystem. The nonlinear subsystem has a characteristic m which means that $Y_n = m(V_n) + Z_n$, where Z_n is disturbance.

The author is with the Institute of Computer Engineering, Control, and Robotics, Wrocław University of Technology, Wrocław, Poland (e-mail: wlodzimierz.greblicki@pwr.wroc.pl).

The system input is a stationary white random process $\{\ldots, U_{-1}, U_0, U_1, \ldots\}$ with finite variance. The probability density of U_n denoted by f can be of any form and is completely unknown. Disturbance $\{\ldots, Z_{-1}, Z_0, Z_1, \ldots\}$, independent of the input signal, is also a stationary process with zero mean and finite variance. Since the linear subsystem is stable, $\{\ldots, X_{-1}, X_0, X_1, \ldots\}$, $\{\ldots, V_{-1}, V_0, V_1, \ldots\}$, and $\{\ldots, Y_{-1}, Y_0, Y_1, \ldots\}$ are also stationary processes but not white.

The only assumption imposed on m is that it is a Lipschitz function, i.e., that

$$|m(u) - m(v)| \leq \alpha_m |u - v| \tag{1}$$

for all u, v, and some unknown α_m. Therefore, in particular, m is continuous. This, $EU_n^2 < \infty$, together with $\sigma_Z^2 < \infty$, implies $EY_n^2 < \infty$ needed in further parts of the paper. Our goal is to recover m from observations taken at input and output of the whole system, i.e., from (U_i, Y_i)'s.

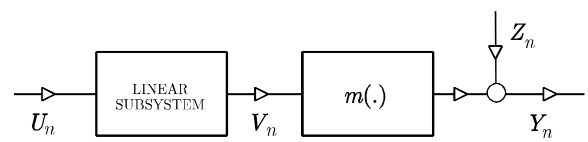

Fig. 1. The Wiener system.

III. THE ESTIMATE

A. Convergence

We fix $p \geq 1$ and apply the following algorithm:

$$\hat{m}_p(u) = \frac{\displaystyle\sum_{i=1}^{n} Y_i \prod_{j=i-p}^{i-1} K\left(\frac{u-U_j}{h_n}\right)}{\displaystyle\sum_{i=1}^{n} \prod_{j=i-p}^{i-1} K\left(\frac{u-U_j}{h_n}\right)}.$$

with $0/0$ defined as 0.

The kernel function K is selected to meet the following assumptions:

$$\int K(u)du = 1, \tag{2}$$

$$\sup_u |K(u)| < \infty, \tag{3}$$

and

$$|u|K(u) \to 0 \text{ as } |u| \to \infty. \tag{4}$$

The nonnegative sequence $\{h_n\}$ is such that

$$\lim_{n\to\infty} h_n = 0 \tag{5}$$

and

$$\lim_{n\to\infty} nh_n^p = \infty. \tag{6}$$

It will be convenient to denote

$$\hat{\xi}_n(u) = \frac{1}{nh_n^p}\sum_{i=1}^{n} Y_i \prod_{j=i-1}^{i-p} K\left(\frac{u-U_j}{h_n}\right)$$

and

$$\hat{\eta}_n(u) = \frac{1}{nh_n^p}\sum_{i=1}^{n} \prod_{j=i-1}^{i-p} K\left(\frac{u-U_j}{h_n}\right),$$

since $\hat{m}_p(u) = \hat{\xi}_n(u)/\hat{\eta}_n(u)$.

The estimate being a nonparametric kernel regression estimate, see, e.g., [9, Sec. 3.1], recovers $m_p(\lambda_p u)$, where

$$\begin{aligned} m_p(\lambda_p u) &= E\left\{Y_p | U_0 = U_1 = \cdots = U_{p-1} = u\right\} \\ &= Em\left(\lambda_p u + \zeta\right) \end{aligned} \tag{7}$$

with $\lambda_p = \sum_{i=1}^{p} k_i$ and $\zeta = \sum_{i=p+1}^{\infty} k_i U_{p-i}$ which fact is established with the following theorem:

Theorem 1: Let K satisfy (2)–(4). Let h_n satisfy (5) and (6). Then

$$\hat{m}_p(u) \to m_p(\lambda_p u) \text{ as } n \to \infty$$

in probability at every point $u \in R$ at which f is continuous, and $f(u) > 0$.

Proof: From Lemmas 2 and 3 in Appendix, it follows that $\hat{\xi}_n(u) \to m_p(\lambda_p u) f^p(u)$ as $n \to \infty$ and $\hat{\eta}_n(u) \to f^p(u)$ as $n \to \infty$ in probability at every continuity point of f. To complete the proof, it suffices to take into account that $\hat{m}_p(u) = \hat{\xi}_n(u)/\hat{\eta}_n(u)$. ■

B. Asymptotic bias

Two remarks should be made. Firstly, the fact that we can recover m only up to some unknown scale factor, λ_p in our case, is caused by the cascade structure of the system and can't be avoided. Secondly, the estimate is asymptotically biased since $m_p(\lambda_p u)$ differs from $m(\lambda_p u)$. Assuming that $EU_n = 0$, we now examine the asymptotic bias, i.e., the difference between $m_p(\lambda_p u)$ and $m(\lambda_p u)$.

Using (1) and (7), we get

$$\begin{aligned} |m_p(\lambda_p u) - m(\lambda_p u)| &= |Em(\lambda_p u + \zeta) - m(\lambda_p u)| \\ &\le E\left|m(\lambda_p u + \zeta) - m(\lambda_p u)\right| \\ &\le \alpha_m E|\zeta| \le \varepsilon_p \alpha_m E\left|U_n\right| \\ &\le \varepsilon_p \alpha_m \sigma_U \end{aligned}$$

with $\varepsilon_p = \sum_{i=p+1}^{\infty} |k_i|$. Thus,

$$\hat{m}_p(u) \to m\left(\lambda_p u\right) + \varepsilon_p \alpha_m \sigma_U \phi_p(u) \text{ as } n \to \infty$$

in probability with some $\phi_p(u)$ such that $|\phi_p(u)| \le 1$. It means that we recover $m\left(\lambda_p u\right)$ up to asymptotic bias $\varepsilon_p \alpha_m \sigma_U \phi(u)$. Since $\varepsilon_p \to 0$ as $p \to \infty$, the bias can be made arbitrarily small by selecting p large enough. For example, for $k_n = \rho^{-n}$, $\varepsilon_p = \rho^{p+1}/(1-\rho)$. Thus, for $\rho = 1/2$, $\varepsilon_1 = 0.5$, $\varepsilon_3 = 0.125$, $\varepsilon_5 = 0.0312$.

The error is much smaller for m having two bounded derivatives. In such a case

$$m\left(\lambda_p u + \zeta\right) - m\left(\lambda_p u\right) = m'\left(\lambda_p u\right)\zeta + m''\left(\lambda_p u + \theta\zeta\right)\zeta^2$$

with $|\theta| < 1$. Hence

$$\begin{aligned} m_p(\lambda_p u) - m(\lambda_p u) &= E\{m(\lambda_p u + \zeta) - m(\lambda_p u)\} \\ &= E\left\{m''\left(\lambda_p u + \theta\zeta\right)\zeta^2\right\} \end{aligned}$$

since $E\zeta = 0$. Therefore

$$\begin{aligned} |m_p(\lambda_p u) - m(\lambda_p u)| &\le E\left|m''\left(\lambda_p u + \theta\zeta\right)\zeta^2\right| \\ &\le \beta_m E\zeta^2 \le \beta_m \sigma_U^2 \sum_{i=p+1}^{\infty} k_i^2 \end{aligned}$$

with $\beta_m = \sup_v |m''(v)|$. Thus

$$\hat{m}_p(u) \to m\left(\lambda_p u\right) + \delta_p \beta_m \sigma_U^2 \psi_p(u) \text{ as } n \to \infty$$

in probability, where $\delta_p = \sum_{i=p+1}^{\infty} k_i^2$ and $|\psi_p(u)| \le 1$. For k_n as above, $\delta_p = \rho^{2p+2}/(1-\rho^2)$ and $\delta_1 = 0.0833$, $\delta_3 = 0.0052$, $\delta_5 = 0.0003$. For example, for $m(u) = (1 - e^{-\beta|u|})\text{sign}u$ with $\beta > 0$, $\beta_m = \beta^2$. For k_n as above and $p = 3$, we get

$$\hat{m}_3(u) \to m(0.875u) + 0.0052\sigma_U^2 \psi_3(u) \text{ as } n \to \infty$$

in probability.

C. Convergence rate

The analysis of convergence rate is rather standard, see, e.g., [9], but requires some modifications. We assume that both m and f and their two derivatives are bounded. Moreover, K is nonnegative, $\int vK(v)dv = 0$ and $\int v^2 K(v)dv < \infty$. Clearly,

$$\begin{aligned} &E\{\hat{\xi}_n(u)|\zeta\} \\ &= \frac{1}{h_n^p} E\left\{ m(V_p) \prod_{i=0}^{p-1} K\left(\frac{u-U_i}{h_n}\right) \middle| \zeta \right\} \\ &= \frac{1}{h_n^p} \int \cdots \int m\left(\sum_{i=0}^{p-1} k_{p-i} u_i + \zeta\right) \\ &\cdot \prod_{i=0}^{p-1} f(u_i) \prod_{j=0}^{p-1} K\left(\frac{u-u_j}{h_n}\right) du_0 \cdots du_{p-1} \\ &= \int \cdots \int m\left(\sum_{i=0}^{p-1} k_{p-i}(u - h_n v_i) + \zeta\right) \\ &\cdot \prod_{i=0}^{p-1} f(u - h_n v_i) \prod_{j=0}^{p-1} K(v_j) dv_0 \cdots dv_{p-1}, \end{aligned}$$

since $\prod_{i=0}^{p-1} f(u_i)$ is the density of $(U_0, \ldots, U_{p-1})$. Denoting

$$\begin{aligned} \phi_p(u) = \int \cdots \int m\left(\sum_{i=0}^{p-1} k_{p-i} u + \zeta\right) f^p(u) \\ \cdot \prod_{i=0}^{p-1} K(v_i) dv_0 \cdots dv_{p-1}, \end{aligned}$$

we get $E\phi_p(u) = m_p(\lambda_p u) f^p(u)$, since (2) holds. Obviously

$$E\{\hat{\xi}_n(u)|\zeta\} - \phi_p(u) = \int \cdots \int [g(u - h_n v_0, \ldots, u - h_n v_{p-1}) - g(u, \ldots, u)] \cdot \prod_{i=0}^{p-1} K(v_i) dv_0 \cdots dv_{p-1},$$

where

$$g(u_0, \ldots, u_{p-1}) = m\left(\sum_{i=0}^{p-1} k_{p-i} u_i + \zeta\right) \prod_{i=0}^{p-1} f(u_i).$$

Clearly, $g(u, \ldots, u) = m(\lambda_p u + \zeta) f^p(u)$. Therefore, expanding $g(u - h_n v_0, \ldots, u - h_n v_{p-1})$ in a Taylor series at a point $(u, \ldots, u)$, we find the quantity in the square brackets equal

$$-h_n \sum_{i=0}^{p-1} v_i \frac{\partial}{\partial u_i} g(u_0, \ldots, u_{p-1})\Big|_{(u,\ldots,u)} + R(u),$$

where

$$R(u) = \frac{1}{2} h_n^2 \sum_{i=0}^{p-1} \sum_{j=0}^{p-1} v_i v_j \cdot \frac{\partial^2}{\partial u_i u_j} g(u_0, \ldots, u_{p-1})\Big|_{(u-\theta_0 h_n v_0, \ldots, u - \theta p-1 h_n v_{p-1})}$$

with all $|\theta_i| \leq 1$. Since m and f and their derivatives are bounded, so are $(\partial^2/\partial u_i \partial u_j) g(u_0, \ldots, u_{p-1})$. Hence

$$|R(u)| \leq \frac{1}{2} h_n^2 M \sum_{i=0}^{p-1} \sum_{j=0}^{p-1} |v_i v_j|$$

with some M. Recalling that $\int vK(v)dv = 0$, we thus get

$$\begin{aligned}
&\left|E\{\hat{\xi}_n(u)|\zeta\} - \phi_p(u)\right| \\
&= \int \cdots \int |R(u)| \prod_{i=0}^{p-1} K(v_i) dv_0 \cdots dv_{p-1} \\
&\leq \frac{1}{2} h_n^2 M \sum_{i=0}^{p-1} \sum_{j=0}^{p-1} \int \cdots \int |v_i v_j| \\
&\cdot \prod_{m=0}^{p-1} K(v_m) dv_0 \cdots dv_{p-1} \\
&= \frac{1}{2} h_n^2 M \left[p\alpha + (p^2 - p)\beta^2\right]
\end{aligned}$$

with $\alpha = \int v^2 K(v) dv$ and $\beta = \int |v| K(v) dv$. In this way, we have shown that $E\{\hat{\xi}_n(u)|\zeta\} - \phi_p(u) = O(h_n^2)$ which implies $E\hat{\xi}_n(u) - m_p(\lambda_p u) f^p(u) = O(h_n^2)$. From this and Lemma 3, we conclude that

$$\begin{aligned}
&E\left(\hat{\xi}_n(u) - m_p(\lambda_p u) f^p(u)\right)^2 \\
&= \left[E\hat{\xi}_n(u) - m_p(\lambda_p u) f^p(u)\right]^2 + \text{var}\left[\hat{\xi}_n(u)\right] \\
&= O(h_n^4) + O\left(\frac{1}{n h_n^p}\right)
\end{aligned}$$

which equals $O(n^{-4/(p+4)})$, provided that $h_n \sim n^{-1/(4+p)}$. Thus

$$\hat{\xi}_n(u) - m_p(\lambda_p u) f^p(u) = O\left(n^{-2/(p+4)}\right) \text{ as } n \to \infty$$

in probability and, for similar reasons, $\hat{\eta}_n(u) - f^p(u) = O(n^{-2/(p+4)})$ as $n \to \infty$ in probability. Hence, see, e.g., [9, Lemma C.8],

$$\hat{m}_p(u) - m_p(\lambda_p u) = O\left(n^{-2/(p+4)}\right) \text{ as } n \to \infty$$

in probability.

IV. Numerical Example

In the numerical example, the linear subsystem system is described by the following equation: $X_{n+1} = aX_n + U_n$ with $a = 0.5$ while

$$m(u) = \begin{cases} -1 & \text{for } u < -\pi/2, \\ \sin u & -\pi/2 \leq u \leq \pi/2, \\ 1 & \text{for } 1 < \pi/2. \end{cases}$$

The input signal is distributed uniformly over the interval $[-2, 2]$. Gaussian disturbance has variance 0.1.

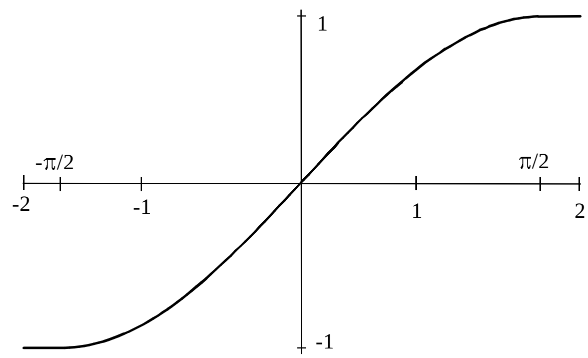

Fig. 2. The nonlinear characteristic m.

A rectangular kernel equal $1/2$ or 0 for $|u| \leq 1$ or $|u| > 1$, respectively, has been applied. Since $m(u)$ is estimated by $\hat{m}_p(u/\lambda_p)$, a global error defined as MISE $= \int_{-1}^{1} E(\hat{m}_p(u/\lambda_p) - m(u))^2 du$ has been empirically calculated. For $p = 2$ results are shown in Fig. 3 (h plays the role of h_n). They are typical for the kernel estimate and suggest that too small h should be avoided. Nevertheless, the error isn't very sensitive to unoptimal choice of h. For h selected in the optimal way suggested by Fig. 3, MISE versus n is shown in Fig. 4. For $p = 1$, the error gets small a little faster than for $p = 2$ and 3, but asymptotic error is greater. For $n \geq 300$, no noticeable difference is, however, between $p = 2$ and 3.

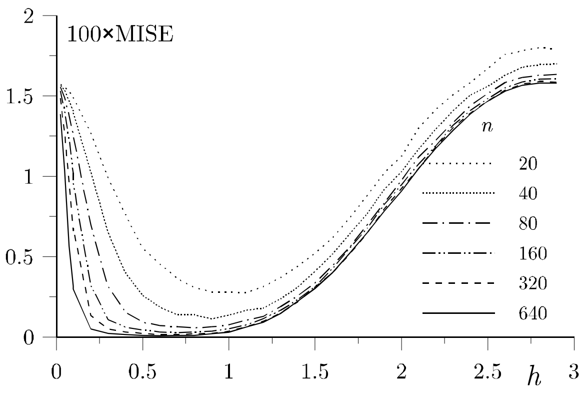

Fig. 3. MISE versus h, $p = 2$.

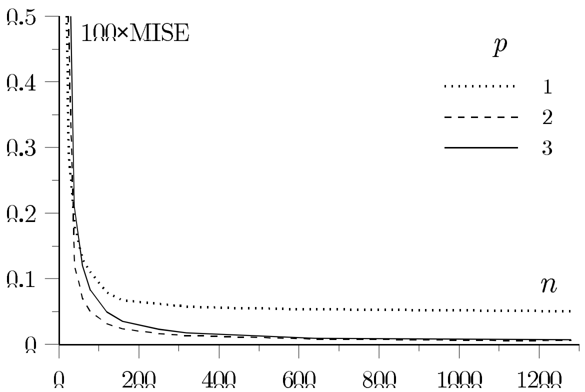

Fig. 4. Minimal MISE versus n.

V. Final remarks

Our estimate recovers the nonlinear characteristic regardless the shape of the input density. It can, therefore, be Gaussian, which is typical in the literature devoted to the nonparametric problem, or not Gaussian. In other words, the density can be of any form which means that our results are input density-free. Moreover, contrary to all mentioned papers dealing with the nonparametric problem, we estimate m, but not its inverse.

The drawback of our estimate is its asymptotic biasedness. The bias, however, can be made arbitrarily small by selecting p large enough. On the other hand, the price paid for that is lower convergence rate. Nevertheless, results of the numerical example look quite encouraging since the number of required observations is counted in hundreds rather than thousands.

VI. Appendix

A. Lemma

The following lemma is a multivariate extension of a result in [12, Theorem 9.8]:

Lemma 1: If K satisfies restrictions (3) and (4), then

$$\lim_{h\to 0}\frac{1}{h^p}\int\cdots\int\varphi(v_1,\ldots,v_p)\prod_{i=1}^{p}K\left(\frac{u-v_i}{h}\right)dv_1\cdots dv_p = \varphi(u,\ldots,u)\left(\int K(v)dv\right)^p$$

at every point $(u,\ldots,u)$ at which φ is continuous.

Since $f^p(u)$ is the density of $(U_1,\ldots,U_p)$, we easily obtain

Corollary 1: Let $\varphi(v_1,\ldots,v_p)$ be continuous. If K satisfies restrictions (3) and (4), then

$$\lim_{h\to 0}\frac{1}{h^p}E\left\{\varphi(U_1,\ldots,U_p)\prod_{i=1}^{p}K\left(\frac{u-U_i}{h}\right)\right\} = \varphi(u,\ldots,u)f^p(u)\left(\int K(v)dv\right)^p$$

and

$$\sup_{h>0}\frac{1}{h^p}\left|E\left\{\varphi(U_1,\ldots,U_p)\prod_{i=1}^{p}K\left(\frac{u-U_i}{h}\right)\right\}\right|<\infty$$

at every point u at which f is continuous.

B. Bias and variance

Since m satisfies (1), both

$$\begin{aligned}&\mu_p(u_0,\ldots,u_{p-1})\\ &\quad= E\left\{m(V_p)|U_0=u_0,\ldots,U_{p-1}=u_{p-1}\right\}\\ &\quad= Em\left(\sum_{i=0}^{p-1}k_{p-i}u_i+\zeta\right)\end{aligned}$$

as well as

$$\begin{aligned}&\nu_p(u_0,\ldots,u_{p-1})\\ &\quad= E\left\{m^2\left(V_p\right)|U_0=u_0,\ldots,U_{p-1}=u_{p-1}\right\}\end{aligned}$$

are Lipschitz function and, as a consequence, continuous.

Lemma 2: Let K satisfy (2)–(4). Let h_n satisfy (5). Then

$$\lim_{n\to\infty}E\hat{\xi}_n(u)=m_p(\lambda_p u)f^p(u)$$

and

$$\lim_{n\to\infty}E\hat{\eta}_n(u)=f^p(u)$$

at every point $u\in R$ at which f is continuous.

Proof: Since $\mu_p(u_0,\ldots,u_{p-1})$ is continuous and

$$E\hat{\xi}_n(u)=\frac{1}{h_n^p}E\left\{\mu_p(U_0,\ldots,U_{p-1})\prod_{i=0}^{p-1}K\left(\frac{u-U_i}{h_n}\right)\right\},$$

applying Corollary 1, we find the quantity converging to $\mu_p(u,\ldots,u)f^p(u)$ as $h_n\to\infty$ at every continuity point u of f. Observing that $\mu_p(u,\ldots,u)=m_p(\lambda_p u)$, we verify the first part of the lemma. Convergence of $E\hat{\eta}_n(u)$ is now obvious. ■

Lemma 3: Let K satisfy (2)–(4). Let h_n satisfy (5) and (6). Then

$$\operatorname{var}\left[\hat{\xi}_n(u)\right]=O\left(\frac{1}{nh_n^p}\right) \quad (8)$$

and

$$\operatorname{var}\left[\hat{\eta}_n(u)\right]=O\left(\frac{1}{nh_n^p}\right) \quad (9)$$

at every point $u\in R$ at which f is continuous.

Proof: Let u be a continuity point of f. Obviously

$$\begin{aligned} & nh_n^p \text{var}\left[\hat{\xi}_n(u)\right] \\ &= \frac{1}{nh_n^p}\text{var}\left[\sum_{i=1}^n Z_i \prod_{j=i-1}^{i-p} K\left(\frac{u-U_j}{h_n}\right)\right] \\ &\quad + \frac{1}{nh_n^p}\text{var}\left[\sum_{i=1}^n m(V_i) \prod_{j=i-1}^{i-p} K\left(\frac{u-U_j}{h_n}\right)\right] \\ &= P_n(u)+Q_n(u)+R_n(u) \end{aligned} \tag{10}$$

with

$$P_n(u)=\sigma_Z^2 \frac{1}{h_n^p}\prod_{i=1}^p EK^2\left(\frac{u-U_i}{h_n}\right),$$

$$Q_n(u)=\frac{1}{h_n^p}\text{var}\left[m(V_p)\prod_{i=0}^{p-1} K\left(\frac{u-U_i}{h_n}\right)\right],$$

and

$$R_n(u)=\frac{2}{nh_n^p}\sum_{q=1}^{n-1}(n-q)\,C_q(u),$$

where

$$C_q(u)=\text{cov}\left[m(V_q)\prod_{i=q-p}^{q-1} K\left(\frac{u-U_i}{h_n}\right), \; m(V_0)\prod_{j=-p}^{-1} K\left(\frac{u-U_j}{h_n}\right)\right].$$

By virtue of Corollary 1, $P_n(u) \to \sigma_Z^2 f^p(u)\left(\int K^2(v)dv\right)^p$ as $n\to\infty$. Thus

$$\sup_{h_n>0}|P_n(u)| \le \alpha(u) \tag{11}$$

with some finite $\alpha(u)$. In turn,

$$\begin{aligned} & h_n^p Q_n(u) \\ &\le E\left\{m^2(V_p)\prod_{i=0}^{p-1} K^2\left(\frac{u-U_i}{h_n}\right)\right\} \\ &= E\left\{E\left\{m^2(V_p)|U_0,\ldots,U_{p-1}\right\}\prod_{i=0}^{p-1} K^2\left(\frac{u-U_i}{h_n}\right)\right\} \\ &= E\left\{\nu_p(U_0,\ldots,U_{p-1})\prod_{i=0}^{p-1} K^2\left(\frac{u-U_i}{h_n}\right)\right\}. \end{aligned}$$

Therefore, due to Corollary 1,

$$Q_n(u)\to \nu_p(u,\ldots,u)f^p(u)\left(\int K^2(v)dv\right)^p \text{ as } n\to\infty$$

which implies

$$\sup_{h_n>0}|Q_n(u)| \le \beta(u) \tag{12}$$

with some finite $\beta(u)$. Passing to $R_n(u)$, we observe

$$\begin{aligned} R_n(u) &= \frac{1}{nh_n^p}\sum_{q=1}^{p}(n-q)\,C_q(u) \\ &\quad + \frac{1}{nh_n^p}\sum_{q=p+1}^{n-1}(n-q)\,C_q(u). \end{aligned} \tag{13}$$

For any $1\le q$, in particular for $1\le q<p$,

$$\begin{aligned} \frac{1}{h_n^p}|C_q(u)| &\le \frac{1}{h_n^p}E\left\{m^2(V_p)\prod_{i=0}^{p-1}K^2\left(\frac{u-U_i}{h_n}\right)\right\} \\ &\quad -\frac{1}{h_n^p}E^2\left\{m(V_p)\prod_{i=0}^{p-1}K\left(\frac{u-U_i}{h_n}\right)\right\} \\ &\le \frac{1}{h_n^p}E\left\{m^2(V_p)\prod_{i=0}^{p-1}K^2\left(\frac{u-U_i}{h_n}\right)\right\} \end{aligned}$$

which, by virtue of Corollary 1, converges to

$$\nu_p(u,\ldots,u)f^p(u)\left(\int K^2(v)dv\right)^p$$

as $n\to\infty$. In such a case

$$\sup_{h_n>0}\frac{1}{h_n^p}|C_q(u)| \le \gamma(u) \tag{14}$$

with a finite $\gamma(u)$. Let now $p\le q$. Defining

$$c_q(u)=\text{cov}\left[m(\theta_q)\prod_{i=q-p}^{q-1}K\left(\frac{u-U_j}{h_n}\right), \; m(V_0)\prod_{j=-p}^{-1}K\left(\frac{u-U_j}{h_n}\right)\right]$$

with $\theta_q=\sum_{k=0}^{q-1}c^TA^{q-k-1}bU_k$, we find $c_q(u)=0$, since $U_{q-p},\ldots,U_{q-1}$ are independent of $U_{-p},\ldots,U_{-1}$, and V_0. Thus, denoting $\rho_q=m(V_q)-m(\theta_q)$, we get

$$\begin{aligned} C_q(u) &= C_q(u)-c_q(u) \\ &= \text{cov}\left[\rho_q\prod_{i=q-p}^{q-1}K\left(\frac{u-U_j}{h_n}\right), \; m(V_0)\prod_{j=-p}^{-1}K\left(\frac{u-U_j}{h_n}\right)\right] \\ &= S_q(u)-T_q(u), \end{aligned}$$

where

$$S_q(u) = E\left\{\rho_q m(V_0)\prod_{i=q-p}^{q-1}K\left(\frac{u-U_i}{h_n}\right) \cdot \prod_{j=-p}^{-1}K\left(\frac{u-U_j}{h_n}\right)\right\}$$

and

$$T_q(u) = E\left\{\rho_q\prod_{i=q-p}^{q-1}K\left(\frac{u-U_i}{h_n}\right)\right\} \cdot E\left\{m(V_0)\prod_{i=-p}^{-1}K\left(\frac{u-U_i}{h_n}\right)\right\}.$$

Since $V_q = c^T A^q X_0 + \theta_q$, due to (1), $|\rho_q| \leq \alpha_m |c^T A^q X_0| \leq \alpha_m \|c\| \, \|X_0\| \, \|A^q\|$. Therefore $|S_q(u)|$ is not greater than

$$\begin{aligned} & \alpha_m \|c\| \, \|A^q\| \, E \left\{ |m(V_0)| \, \|X_0\| \prod_{i=q-p}^{q-1} \left| K\left(\frac{u-U_i}{h_n}\right)\right| \right. \\ & \left. \cdot \prod_{j=-p}^{-1} \left| K\left(\frac{u-U_j}{h_n}\right)\right| \right\} \\ \leq \; & \alpha_m \kappa^p \|c\| \, \|A^q\| \\ & \cdot E \left\{ |m(V_0)| \, \|X_0\| \prod_{i=q-p}^{q-1} \left| K\left(\frac{u-U_i}{h_n}\right)\right| \right\} \end{aligned}$$

which equals

$$d \, \|A^q\| \, E \left\{ \prod_{i=1}^{p} \left| K\left(\frac{u-U_i}{h_n}\right)\right| \right\},$$

where $d = \alpha_m \kappa^p \|c\| \, E\{|m(V_0)|\} \, E\{\|X_0\|\}$, and $\kappa = \sup_u |K(u)|$. Therefore, recalling Corollary 1, we find $\sup_{h_n>0} h_n^{-p} |S_q(u)| = \|A^q\| \, \delta_1(u)$ with a finite $\delta_1(u)$. For similar reasons $\sup_{h_n>0} h_n^{-p} |T_q(u)| = \|A^q\| \, \delta_2(u)$ with a finite $\delta_2(u)$. Thus, for such p and q,

$$\sup_{h_n>0} \frac{1}{h_n^p} |C_q(u)| = \|A^q\| \, \delta(u)$$

with a finite $\delta(u)$. Therefore, recalling (13), and (14), get

$$\begin{aligned} \sup_{h_n>0} |R(u)| &\leq p\gamma(u) + \delta(u) \frac{1}{n} \sum_{q=p+1}^{n-1} (n-q) \, \|A^q\| \\ &= O(1) \end{aligned} \tag{15}$$

since, owing to the stability of A, $n^{-1} \sum_{q=1}^{n} (n-q) \, \|A^q\|$ converges as $n \to \infty$. Finally, from (10), (11), (12), and (15), it follows that $n h_n^p \mathrm{var}[\hat{\xi}_n(u)] = O(1)$. In this way we have verified (8). Since (9) is now obvious, the proof has been completed. ■

Acknowledgement

The author would like to thank to reviewers for their comments and suggestions.

References

[1] S.A. Billings and S.Y. Fakhouri, "Theory of separable processes with applications to the identification of nonlinear systems," *IEE Proceedings*, vol. 125, no. 9, pp. 1051–1058, October 1978.

[2] 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.

[3] T. Wigren, "Recursive prediction error identification using the nonlinear Wiener model," *Automatica*, vol. 29, no. 4, pp. 1011–1025, 1993.

[4] H.F. Chen, "Recursive identification for Wiener model with discontinuous piece-wise linear function," *IEEE Transactions on Automatic Control*, vol. 51, pp. 390–400, March 2006.

[5] W. Greblicki, "Nonparametric identification of Wiener systems," *IEEE Transactions on Information Theory*, vol. 38, no. 5, pp. 1487–1493, September 1992.

[6] ——, "Nonparametric identification of Wiener systems by orthogonal series," *IEEE Transactions on Automatic Control*, vol. 39, no. 10, pp. 2077–2086, October 1994.

[7] M. Pawlak and Z. Hasiewicz, "Nonlinear system identification by the Haar multiresolution analysis," *IEEE Transactions on Circuits and Systems-I: Fundamental Theory And Applications*, vol. 45, pp. 945–961, September 1998.

[8] W. Greblicki, "Nonlinearity recovering in Wiener system driven with correlated signal," *IEEE Transactions on Automatic Control*, vol. 49, no. 10, pp. 1805–1810, October 2004.

[9] W. Greblicki and M. Pawlak, *Nonparametric System Identification*. Cambridge: Cambridge University Press, 2008.

[10] M. Pawlak, Z. Hasiewicz, and P. Wachel, "On nonparametric identification of Wiener systems," *IEEE Transactions on Signal Processing*, vol. 55, pp. 482–492, February 2007.

[11] G. Mzyk, "A censored sample mean approach to nonparametric identification of nonlinearities in Wiener systems," *IEEE Transactions on Circuit and Systems-II: Express Briefs*, vol. 54, pp. 897–901, October 2007.

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

Włodzimierz Greblicki was born in 1943 in Poland. He received the M.Eng., Ph.D., and D.Sc. degrees from the Wrocław University of Technology, Wrocław, Poland, in 1966, 1971, and 1975, respectively. He is now Full Professor at the University and teaches courses in dynamic systems and automatic control. His current research interest is in system identification, in particular nonparametric. He is a coauthor (with M. Pawlak) of the book *Nonparametric System Identification* (Cambridge University Press, 2008).