Nonlinearity Recovering in Wiener System Driven with Correlated Signal

Włodzimierz Greblicki

Abstract— **The characteristic of the nonlinear part of the Wiener system is estimated. The system is driven by a Gaussian random process which may not be white. Three algorithms are presented, two semirecursive and one of the off-line type. Their pointwise convergence in probability is shown and results of numerical simulation are given.**

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

I. Introduction

In the block–oriented approach, we identify systems consisting of linear dynamic and nonlinear memoryless subsystems. In particular, in the Wiener system, a linear dynamic part is followed by a nonlinear memoryless one. The objective is to recover descriptions of the subsystems from input-output observations of the whole system. In this note we focus our attention on the nonlinear subsystem.

In the literature, the *a priori* information about both subsystems is usually parametric, see, e.g., [1], [3], [7], [8], [10]. As a result, the nonlinear characteristic is a function known up to some coefficients, [10], or just a polynomial, [3], [7], [8]. The random input signal is Gaussian, [3], [8], but not necessarily, [10], white, [3], or correlated, [8], [10]. Due to serious theoretical difficulty, properties of some algorithms are only suggested by simulation examples, [1], [7].

The characteristic can be recovered also when the *a priori* information is smaller than parametric, i.e., nonparametric. Nonparametric algorithms work properly even if we know nothing about the nonlinear characteristic which can be any Borel measurable function, [5]. Needless to say that parametric algorithms are then completely useless. Thus, for a user, nonparametric methods have a great advantage, since they can be applied when the knowledge about the system is so poor that its description can't be parameterized.

Off–line, [4], [5], and semirecursive, [6], nonparametric kernel algorithms have been examined. In all those papers, the input signal is a Gaussian white random process. The novelty of this note is the fact that we admit also a correlated input signal. To estimate the nonlinear characteristic, we propose three kernel algorithms, of which two are semirecursive and one is off-line. We show that all are pointwise consistent and present results of numerical simulation.

II. System and Algorithms

The Wiener system with input U_n and output Y_n, shown in Fig. 1 with the thick line, consists of two subsystems.

W. Greblicki is with the Institute of Engineering Cybernetics, Wrocław University of Technology, 50-370 Wrocław, Poland. E-mail: wgre@ict.pwr.wroc.pl

First is linear, dynamic while the other nonlinear, memoryless. The dynamic subsystem has an impulse response $\{\lambda_n; n = 0, 1, \ldots\}$ which means that

$$V_n = \sum_{i=-\infty}^{n} \lambda_{n-i} U_i, \tag{1}$$

where $\{U_n; n = \ldots, -1, 0, 1, \ldots\}$. The impulse response $\{\lambda_n\}$ is unknown but we assume that

$$|\lambda_n| \le \delta_1 \varepsilon_1^n \tag{2}$$

with some unknown δ_1 and ε_1, where $0 \le \varepsilon_1 < 1$. Observe that the restriction is satisfied by every stable ARMA system.

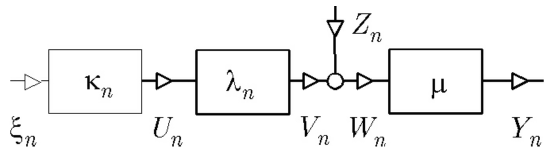

Fig. 1. Wiener system (thick line) and the system forming its input signal (thin line).

We assume that the input signal is correlated and is of the following form:

$$U_n = \sum_{i=-\infty}^{n} \kappa_{n-i} \xi_i, \tag{3}$$

where $\{\xi_n; n = \ldots, -1, 0, 1, \ldots\}$ is a stationary white random Gaussian process with zero mean and unknown variance σ_ξ^2. The signal $\{\xi_n\}$ is not measured. The weighting function $\{\kappa_n\}$ is unknown but satisfies the following inequality:

$$|\kappa_n| \le \delta_2 \varepsilon_2^n \tag{4}$$

with some unknown δ_2 and ε_2, where $0 \le \varepsilon_2 < 1$. Thus, $\{U_n\}$ is output of some system with unknown impulse response $\{\kappa_n\}$ driven by $\{\xi_n\}$, see the thin line in Fig. 1.

Output of the subsystem is disturbed by additive stationary white Gaussian noise $\{Z_n; n = \ldots, -1, 0, 1, \ldots\}$ independent of both $\{U_n\}$ and $\{V_n\}$ and having zero mean and unknown finite variance. Owing to all that, both $\{V_n; n = \ldots, -1, 0, 1, \ldots\}$ as well $\{W_n; n = \ldots, -1, 0, 1, \ldots\}$ are also stationary and Gaussian.

The unknown characteristic of the nonlinear part, denoted by μ, has a continuous derivative such that, for all $w \in R$,

$$0 < |\mu'(w)| \le c_\mu \tag{5}$$

with some unknown c_μ. Observe that, owing to that, μ is strictly monotonous and μ^{-1} exists in $R \times \mu(R)$, where $\mu(R)$ is the image of R under the mapping μ.

Our goal is to estimate the characteristic μ of the nonlinear subsystem from observations $(U_0, Y_0), (U_1, Y_1), \ldots$ taken at input and output of the Wiener system.

The basis for our algorithms is the following lemma owing to which, to recover the inverse of $\mu(y)$, we estimate the regression $E\{U_0|Y_k = y\}$. In the lemma and further parts of the paper,

$$\alpha_k = \frac{\sigma_\xi^2}{\sigma_W^2}\sum_{i=0}^{\infty}\lambda_i\beta_{k-i}$$

with $\beta_i = \sum_{j=0}^{\infty}\kappa_j\kappa_{i+j}$.

Lemma 1: Let $\{U_n\}$ be a stationary Gaussian random process with zero mean. For $k = 0, 1, \ldots$,

$$E\{U_0|Y_k = y\} = \alpha_k\mu^{-1}(y). \tag{6}$$

Proof: Using (1) and (3), one can verify that $\mathrm{cov}\,[U_0, W_k] = \alpha_k\sigma_W^2$. Since the pair (U_0, W_k) has a normal distribution with zero marginal means, an application of Lemma 2 in Appendix A, leads to $E\{U_0|W_k = w\} = \alpha_k w$. Since $E\{U_0|\mu(W_k) = y\} = E\{U_0|W_k = \mu^{-1}(y)\}$, the proof has been completed. ■

To recover $\alpha_k\mu^{-1}(y)$, we apply the following off-line algorithm:

$$\hat{m}_k(y) = \frac{\sum_{i=1}^{n} U_i K\left(\frac{y - Y_{i+k}}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{y - Y_{i+k}}{h_n}\right)} \tag{7}$$

and two semirecursive ones, namely:

$$\tilde{m}_k(y) = \frac{\sum_{i=1}^{n} U_i \frac{1}{h_i} K\left(\frac{y - Y_{i+k}}{h_i}\right)}{\sum_{i=1}^{n} \frac{1}{h_i} K\left(\frac{y - Y_{i+k}}{h_i}\right)} \tag{8}$$

and

$$\bar{m}_k(y) = \frac{\sum_{i=1}^{n} U_i K\left(\frac{y - Y_{i+k}}{h_i}\right)}{\sum_{i=1}^{n} U_i K\left(\frac{y - Y_{i+k}}{h_i}\right)}. \tag{9}$$

In all above algorithms, K and $\{h_n\}$ are a suitably selected kernel function and a positive number sequence, respectively. On the Borel measurable kernel K, we impose the following restrictions:

$$\sup_{y\in R}|K(y)| < \infty, \tag{10}$$

$$\int K(y)dy < \infty, \tag{11}$$

$$yK(y) \to 0 \text{ as } |y| \to \infty, \tag{12}$$

$$|K(x) - K(y)| \le c_K|x - y| \tag{13}$$

with some c_K, for all $x, y \in R$. We denote $d_K = \sup_y |K(y)|$. Depending on algorithm, the positive number sequence $\{h_n\}$ satisfies some of the following conditions:

$$\{h_n\} \text{ is monotonous} \tag{14}$$

$$h_n \to 0 \text{ as } n \to \infty, \tag{15}$$

$$nh_n^2 \to 0 \text{ as } n \to \infty, \tag{16}$$

$$\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h_i^2} \to 0 \text{ as } n \to \infty, \tag{17}$$

$$h_n\sum_{i=1}^{n}h_i \to \infty \text{ as } n \to \infty. \tag{18}$$

Algorithms (7)–(9) just estimate the regression in (6). Estimates (8) and (9) are semirecursive since their numerators and denominators can be calculated recursively.

For the sake of the simplicity of the notation, W and Y stand for W_k and Y_k, respectively.

III. Convergence of algorithms

Theorems 1 and 2 establish convergence of algorithms (8) and (9). Having their proofs, the reader can easily verify convergence of (7), i.e., prove Theorem 3.

Theorem 2: Let μ satisfy (5). Let K satisfy (10)–(13). Let the number sequence satisfy (15) and (17). Then, at every $y \in \mu(R)$,

$$\tilde{m}_k(y) \to \alpha_k\mu^{-1}(y) \text{ as } n \to \infty \text{ in probability.}$$

Proof: First of all, observe that from Lemma 4 in Appendix B, it follows that the density f of Y is continuous and $0 < f(y) < \infty$ at every $y \in \mu(R)$. Let y be any point in $\mu(R)$.

We denote

$$\tilde{g}(y) = \frac{1}{n}\sum_{i=1}^{n}U_i\frac{1}{h_i}K\left(\frac{y - Y_{k+i}}{h_i}\right),$$

$$\tilde{f}(y) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}K\left(\frac{y - Y_{k+i}}{h_n}\right)$$

and observe $\tilde{m}_k(y) = \tilde{g}(y)/\tilde{f}(y)$. Applying Lemma1, we get $E\tilde{g}(y) = n^{-1}\sum_{i=1}^{n} m(y; h_i)$ with

$$\begin{aligned} m(y; h) &= \frac{1}{h}E\left\{U_0K\left(\frac{y - Y_k}{h}\right)\right\} \\ &= \alpha_k\frac{1}{h}E\left\{\mu^{-1}(Y)K\left(\frac{y - Y}{h}\right)\right\}. \end{aligned} \tag{19}$$

Using (15) and Lemma 3 in Appendix A, we find

$$m(y; h) \to \alpha_k\mu^{-1}(y)f(y)\int K(x)dx \text{ as } h \to 0, \tag{20}$$

since $E|\mu^{-1}(Y)| = E|W| < \infty$. Thus, $E\tilde{g}(y)$ converges to the same limit as $n \to \infty$.

Passing to variance, we find $\mathrm{var}[\tilde{g}(y)] = A_n(y) + B_n(y)$ with

$$A_n(y) = \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h_i^2}\,\mathrm{var}\left[U_0K\left(\frac{y - Y_k}{h_i}\right)\right],$$

$$\begin{aligned} B_n(y) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1, j\ne i}^{n}\frac{1}{h_ih_j} \\ \times\,\mathrm{cov}\left[U_iK\left(\frac{y - Y_{k+i}}{h_i}\right), U_jK\left(\frac{y - Y_{k+j}}{h_j}\right)\right] \end{aligned}$$

and observe that $0 \le A_n(y) \le d_K^2 \sigma_U^2 n^{-2} \sum_{i=1}^n h_i^{-2} = O(n^{-2} \sum_{i=1}^n h_i^{-2})$ and

$$B_n(y) = \frac{2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{i-1} \frac{1}{h_i h_j} \times \operatorname{cov}\left[U_i K\left(\frac{y - Y_{k+i}}{h_i}\right), U_j K\left(\frac{y - Y_{k+j}}{h_j}\right)\right].$$

By virtue of Lemma 7 in Appendix B, the double sum in the above expression is bounded in absolute value by $\varepsilon^{-k}\phi(y)\sum_{i=1}^n h_i^{-2} \sum_{j=1}^{i-1} \varepsilon^{i-j} \le \gamma\varepsilon^{-k}\phi(y)\sum_{i=1}^n h_i^{-2}$ with $\gamma = \sum_{n=0}^\infty \varepsilon^n$. Hence $B_n(y) = O(n^{-2}\sum_{i=1}^n h_i^{-2})$. Therefore, $\operatorname{var}]\tilde{g}(y)] = O(n^{-2}\sum_{i=1}^n h_i^{-2})$. In this way, we have finally shown that $E\tilde{g}(y) \to \alpha_k \mu^{-1}(y) f(y) \int K(x)dx$ as $n \to \infty$ in probability.

Using similar arguments, we can easily verify that $\tilde{f}(y) \to f(y) \int K(x)dx$ as $n \to \infty$ in probability. Since, as we have already noticed, $f(y) \neq 0$, the proof has been completed. ■

Theorem 3: Let μ satisfy (5). Let K satisfy (10)–(13). Let the number sequence satisfy (14), (15) and (18). Then, at every $y \in \mu(R)$,

$$\bar{m}_k(y) \to \alpha_k \mu^{-1}(y) \text{ as } n \to \infty \text{ in probability.}$$

Proof: Let $y \in \mu(R)$. Clearly, $\bar{m}_k(y) = \bar{g}(y)/\bar{f}(y)$ with

$$\bar{g}(y) = \frac{1}{\sum_{i=1}^n h_i} \sum_{i=1}^{n} U_i K\left(\frac{y - Y_{k+i}}{h_i}\right)$$

and

$$\bar{f}(y) = \frac{1}{\sum_{i=1}^n h_i} \sum_{i=1}^{n} K\left(\frac{y - Y_{k+i}}{h_n}\right).$$

Applying Lemma 1, we obtain

$$E\bar{g}(y) = \sum_{i=1}^{n} h_i m(y; h_i) \Bigg/ \sum_{i=1}^{n} h_i$$

with $m(y;h)$ as in (19). Since $\sum_{i=1}^\infty h_i = \infty$, see (18), using (20), we find $E\bar{g}(y) \to \alpha_k \mu^{-1}(y) f(y) \int K(x)dx$ as $n \to \infty$.

To examine variance, observe that $\operatorname{var}[\bar{g}(y)] = A_n(y) + B_n(y)$ with

$$A_n(y) = \frac{1}{\left(\sum_{i=1}^n h_i\right)^2} \sum_{i=1}^{n} \operatorname{var}\left[U_0 K\left(\frac{y - Y_k}{h_i}\right)\right],$$

$$B_n(y) = \frac{1}{\left(\sum_{i=1}^n h_i\right)^2} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \times \operatorname{cov}\left[U_i K\left(\frac{y - Y_{k+i}}{h_i}\right), U_j K\left(\frac{y - Y_{k+j}}{h_j}\right)\right].$$

Now,

$$0 \le A_n(y) \le \frac{d_K^2 \sigma_U^2 n}{\left(\sum_{i=1}^n h_i\right)^2} \le \frac{1}{h_n \sum_{i=1}^n h_i} \frac{d_K^2 \sigma_U^2 n h_n}{\sum_{i=1}^n h_i}.$$

Since $\{h_n\}$ is monotonous, $nh_n/\sum_{i=1}^n h_i \le 1$ and $A_n(y) = O(1/h_n \sum_{i=1}^n h_i)$. In turn,

$$B_n(y) = \frac{2}{\left(\sum_{i=1}^n h_i\right)^2} \sum_{i=1}^{n} \sum_{j=1}^{i-1} \operatorname{cov}\left[U_i K\left(\frac{y - Y_{k+i}}{h_i}\right), U_j K\left(\frac{y - Y_{k+j}}{h_j}\right)\right].$$

By virtue of Lemma 7 in Appendix B, the double sum in the expression is bounded in absolute value by $\phi(y)\varepsilon^{-k} \sum_{i=1}^n \sum_{j=1}^{i-1} (h_j/h_i)\varepsilon^{i-j}$. Since $\{h_n\}$ is monotonous, the obtained double sum is not greater than $h_n^{-1} \sum_{i=1}^n \sum_{j=1}^{i-1} h_j \varepsilon^{i-j} = h_n^{-1} \sum_{j=1}^n h_j \sum_{i=j+1}^n \varepsilon^{i-j} = \gamma h_n^{-1} \sum_{j=1}^n h_j$ with $\gamma = \sum_{n=0}^\infty \varepsilon^n$. Hence, $B_n(y) = O(1/h_n \sum_{i=1}^n h_i)$ and, thus, $\operatorname{var}[\bar{g}(y)] = O(1/h_n \sum_{i=1}^n h_i)$. We have thus shown that $\bar{g}(y) \to \alpha_k \mu^{-1}(y) f(y) \int K(x)dx$ as $n \to \infty$ in probability.

Using similar arguments, we can easily verify that $\bar{f}(y) \to f(y) \int K(x)dx$ as $n \to \infty$ in probability. Since $f(y) \neq 0$, the proof has been completed. ■

Theorem 4: Let μ satisfy (5). Let K satisfy (10)–(13). Let the number sequence satisfy (15) and (16). Then, at every $y \in \mu(R)$,

$$\hat{m}_k(y) \to \alpha_k \mu^{-1}(y) \text{ as } n \to \infty \text{ in probability.}$$

IV. Simulation Example

In the simulation example, the dynamic part of the system has a transfer function $z/(z - 0.5)$. The nonlinearity is of the following form: $\mu(w) = \operatorname{sign}(w)[(|w|+1)^2 - 1]$. The input signal is produced by a dynamic system with a transfer function $z/(z-0.2)$ driven by a white Gaussian process with zero mean and variance $\sigma_\xi^2 = 1$. Disturbance has variance $\sigma_Z^2 = 0.1$. In the example, $k = 0$ and $\alpha_0 \approx 0.64$. The kernel $K(y) = \exp(-y^2)$ has been applied. The global error is measured with $\text{MISE} = \int_{-20}^{20} E(\tilde{m}_0(y) - \alpha_0\mu^{-1}(y))^2 dy$.

For $n = 500$, observations and the resulting off–line estimate obtained for h_{500} minimizing the MISE is shown in Fig. 2. In Fig. 3, for each $n; n \in [10, 1280]$, the h_n minimizing the error has been selected. For $n = 1280$, MISE ≈ 0.6 which means that the average squared point-wise error measured on the interval $[-20, 20]$ is not greater than 0.015. The error gets small rapidly as the number of observations increases, especially for n counted in dozens, and becomes quite small for n reaching a few hundreds.

V. Final Remarks

Within the nonparametric approach, new in this note is that the input signal is correlated. For input being a white process, an algorithm similar to (7) has been already examined in [4] and [5] while (8) and (9) in [6]. Applying arguments applied in [5], one can extend our results to any Borel function μ.

Finally, we want to notice that Wiener systems are used to describe processes in chemistry, [7], biology, psychology or sociology, see references in cited papers. It makes the identification of such systems important not only from the theoretical viewpoint but also for users.

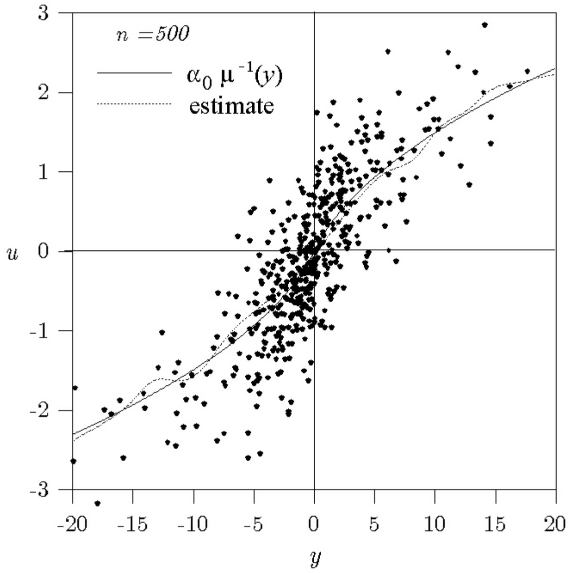

Fig. 2. Observations and estimate $m_0(y)$; $n = 500$.

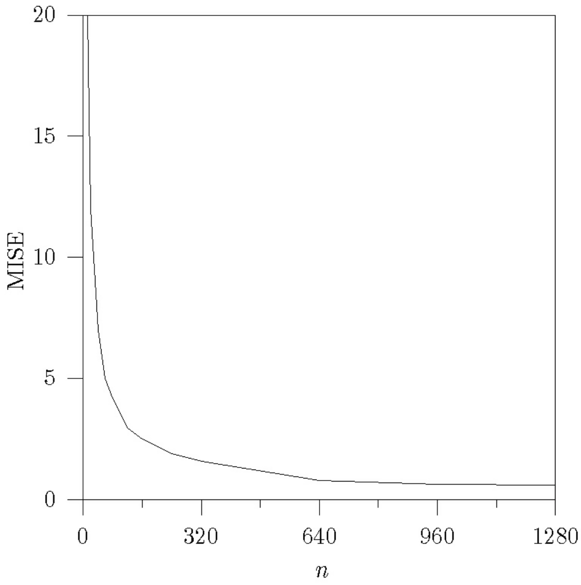

Fig. 3. MISE versus n.

Appendix A. General Results

Lemma 5: Let a pair (X, Y) have a normal distribution with zero marginal means. Then

$$E\{Y|X = x\} = \sigma_X^{-2}\operatorname{cov}[X,Y]x,$$

$$E\{Y^2|X = x\} = \sigma_X^{-4}\operatorname{cov}^2[X,Y]x^2 + \left(\sigma_Y^2 - \sigma_X^{-2}\operatorname{cov}^2[X,Y]\right).$$

Lemma 6: Let X be a random variable with a density function f and let φ be a Borel function such that $E|\varphi(X)| < \infty$. If a Borel measurable kernel K satisfies (10)–(12), then

$$\frac{1}{h}E\left\{\varphi(X)K\left(\frac{x-X}{h}\right)\right\} \to \varphi(x)f(x)\int K(y)dy$$

as $h \to 0$ and

$$\sup_{h>0}\frac{1}{h}E\left|\varphi(X)K\left(\frac{x-X}{h}\right)\right| < \infty$$

at every point $x \in R$ at which both φ and f are continuous.

The first part of the lemma we find in [9, Theorem 9.8]. The other is obvious.

Appendix B. The System

Lemma 7: The density f of Y is equal to

$$f(y) = \begin{cases} f_W(\mu^{-1}(y))\left|\frac{d}{dy}\mu^{-1}(y)\right|, & \text{for } y \in \mu(R) \\ 0, & \text{otherwise,} \end{cases}$$

where f_W is the density of W. If μ has a continuous derivative satisfying (5), then f is continuous and $0 < f(y) < \infty$ at every $y \in \mu(R)$.

Proof: The formula for $f(y)$ is not difficult to derive and can be found, e.g., in [2, Chapter 2, Theorem 1]. A μ satisfying (5) is continuous and strictly monotonous. Thus, μ^{-1} is continuous in $\mu(R)$. Since f_W is the normal density, also $f_W(\mu^{-1}(.))$ is continuous and $0 < f_W(\mu^{-1}(y)) < \infty$. As

$$\frac{d}{dy}\mu^{-1}(y) = \frac{1}{\mu'(w)|_{w=\mu^{-1}(y)}} = \frac{1}{\mu'(\mu^{-1}(y))}$$

and (5) holds, we conclude that

$$0 < \frac{d}{dy}\mu^{-1}(y) < \infty$$

which completes the proof. ■

Let $\{\eta_n; n = 0, 1, \ldots\}$ be the impulse response of the system with input ξ_n and output V_n. Owing to (2) and (4), we can write

$$|\eta_n| \le \delta_3\varepsilon_3^n \tag{21}$$

with some δ_3 and ε_3 such that $0 \le \varepsilon_3 < 1$.

Lemma 8: For $n = 0, 1, \ldots,$

$$|E\{\xi_0|W_n\}| \le \rho(W_n), E\{\xi_0^2|W_n\} \le \rho(W_n), \tag{22}$$

$$|E\{U_0|W_n\}| \le \rho(W_n), E\{U_0^2|W_n\} \le \rho(W_n) \tag{23}$$

where $\rho(W) = d_1 + d_2W^2$ with some d_1, and d_2, both independent of n.

Proof: Since $W_n = Z_n + \sum_{i=-\infty}^{n}\eta_{n-i}\xi_i$, $\operatorname{cov}[\xi_0, W_n] = \sigma_\xi^2\eta_n$, using Lemma 2, we get $E\{\xi_0|W_n\} = (\sigma_\xi^2/\sigma_W^2)\eta_nW_n$. Thus $|E\{\xi_0|W_n\}| \le c_1|W_n|$ with c_1 independent of n. For similar reasons, $E\{\xi_0^2|W_n\} \le c_1^2W_n^2 + c_2$ with c_2 independent of n. Therefore (22) holds. The proof of (23) is similar. ■

Our next lemma is crucial for this note.

Lemma 9: Let φ be a Borel function such that

$$|\varphi(v) - \varphi(w)| \le c_\varphi\,|v - w| \tag{24}$$

with some c_φ, for all v, $w \in R$. Let $d_\varphi = \sup_{-\infty<v<\infty}|\varphi(v)| < \infty$. Then

$$|\operatorname{cov}[U_n\varphi(W_{n+k}), U_0\psi(W_k)]| \le \begin{cases} \beta_1 d_\varphi E\{\rho(W)|\psi(W)|\}, & \text{for } 0 < n \le k, \\ (\beta_2 d_\varphi + \beta_3 c_\varphi)\,\varepsilon^{n-k}E\{\rho(W)|\psi(W)|\}, & \text{for } 0 \le k < n \end{cases}$$

with some β_1, β_2, β_3 independent of k, n, φ, and ψ. Moreover, $0 \le \varepsilon < 1$, and $\rho(w) = c_1 + c_2 w^2$ with some c_1, and c_2, both independent of k, n, φ, and ψ.

Proof: For the sake of simplicity, let $\delta = \max(\delta_1, \delta_2, \delta_3)$ and $\varepsilon = \max(\varepsilon_1, \varepsilon_2, \varepsilon_3)$. To avoid complicated notation, we assume that ψ is nonnegative.

Suppose that $0 < n \le k$. The covariance in the assertion equals $A_n - B_n$ with $A_n = E\{U_n U_0 \varphi(W_{n+k})\psi(W_k)\}$, and $B_n = E\{U_n\varphi(W_{n+k})\}E\{U_0\psi(W_k)\}$. Observe

$$\begin{aligned} |A_n| &\le d_\varphi E|U_n U_0 \psi(W_k)| \\ &\le d_\varphi E^{1/2}\{U_0^2\psi(W_k)\}E^{1/2}\{U_n^2\psi(W_k)\} \end{aligned}$$

which, by virtue of Lemma 5, is bounded by $d_\varphi E\{\rho(W)\psi(W)\}$. In turn,

$$|B_n| \le d_\varphi E|U_n| E\{\rho(W)\psi(W)\}$$

and the inequality in the assertion follows.

Let now $0 \le k < n$. Since

$$U_n = \sum_{i=-\infty}^{k} \kappa_{n-i}\xi_i + \sum_{i=k+1}^{n} \kappa_{n-i}\xi_i,$$

we get

$$\operatorname{cov}\left[U_n\varphi(W_{n+k}), U_0\psi(W_k)\right] = T_n + S_n$$

with

$$\begin{aligned} T_n &= \sum_{i=-\infty}^{k} \kappa_{n-i} \operatorname{cov}\left[\xi_i\varphi(W_{n+k}), U_0\psi(W_k)\right] \\ &= \sum_{i=0}^{\infty} \kappa_{n+i-k} \operatorname{cov}\left[\xi_{k-i}\varphi(W_{n+k}), U_0\psi(W_k)\right], \\ S_n &= \sum_{i=k+1}^{n} \kappa_{n-i} \operatorname{cov}\left[\xi_i\varphi(W_{n+k}), U_0\psi(W_k)\right]. \end{aligned}$$

Using (4), we obtain

$$|T_n| \le \delta\varepsilon^{n-k} \sum_{i=0}^{\infty} \varepsilon^i \left|\operatorname{cov}\left[\xi_{k-i}\varphi(W_{n+k}), U_0\psi(W_k)\right]\right|. \quad (25)$$

The covariance in (25) equals $T_{1n} - T_{2n}$ with $T_{1n} = E\{\xi_{k-i}\varphi(W_{n+k})U_0\psi(W_k)\}$ and $T_{2n} = E\{\xi_{k-i}\varphi(W_{n+k})\}$ $E\{U_0\psi(W_k)\}$. Observe that

$$\begin{aligned} |T_{1n}| &\le d_\varphi E|\xi_{k-i}U_0\psi(W_k)| \\ &= d_\varphi E\{E\{|\xi_{k-i}U_0|\,|W_k\}\psi(W_k)\}. \end{aligned}$$

Since, by virtue of Lemma 5, the conditional expectation in the expression is not greater than $E^{1/2}\{\xi_{k-i}^2|W_k\}$ $E^{1/2}\{U_0^2|W_k\} \le \rho(W_k)$, we conclude that $|T_{1n}| \le d_\varphi E\left\{\rho(W)\psi(W)\right\}$. In turn, $|T_{2n}| \le d_\varphi E|\xi_0|E|U_0\psi(W_k)|$ which does not exceed $d_\varphi\sigma_\xi E\{\rho(W)\psi(W)\}$. Therefore, the absolute value of the covariance in (25) is bounded by $d_\varphi(1+\sigma_\xi)E\{\rho(W)\psi(W)\}$. Finally, denoting $\gamma = \sum_{i=0}^{\infty}\varepsilon^i$, we get

$$|T_n| \le d_\varphi\gamma\delta(1+\sigma_\xi)\varepsilon^{n-k}E\{\rho(W)\psi(W)\}.$$

To examine S_n, observe that $W_{n+k} = P_k + Q_{n+k}$ with $P_k = \sum_{j=-\infty}^{k}\eta_{n+k-j}\xi_j$, $Q_{n+k} = \sum_{j=k+1}^{n+k}\eta_{n+k-j}\xi_j + Z_{n+k}$. Notice also that, for $i = k+1,\dots,k+n$, $\operatorname{cov}[\xi_i\varphi(Q_{n+k}), U_0\psi(W_k)] = 0$, since pairs (ξ_i, Q_{n+k}) and (U_0, W_k) are independent. Thus,

$$\begin{aligned} S_n &= \sum_{i=k+1}^{n} \kappa_{n-i} \\ &\quad\times \operatorname{cov}\left[\xi_i\left(\varphi(W_{n+k}) - \varphi(Q_{n+k})\right), U_0\psi(W_k)\right] \\ &= \sum_{i=k+1}^{n} \kappa_{n-i}\left(A_{ni} - B_{ni}\right) \end{aligned}$$

with

$$A_{ni} = E\{\xi_i[\varphi(W_{n+k}) - \varphi(Q_{n+k})]U_0\varphi(W_k)\},$$

$$B_{ni} = E\{\xi_i[\varphi(W_{n+k}) - \varphi(Q_{n+k})]\}E\{U_0\psi(W_k)\}.$$

Owing to (24), $|\varphi(W_{n+k}) - \varphi(Q_{n+k-1})| \le c_\varphi|W_{n+k} - Q_{n+k}| \le c_\varphi|P_k|$. This and the fact that ξ_i is independent of U_0, W_k, and P_k, imply

$$\begin{aligned} |A_{ni}| &\le c_\varphi E|\xi_i|E|P_k U_0\psi(W_k)| \\ &\le c_\varphi\sigma_\xi E|P_k U_0\psi(W_k)|. \end{aligned}$$

Using (4) and (21), we find

$$|P_k| \le \delta\varepsilon^n \sum_{j=-\infty}^{k} \varepsilon^{k-j}|\xi_j|,$$

and

$$\begin{aligned} |U_0| &= \Big|\sum_{i=-\infty}^{0} \kappa_{-i}\xi_i\Big| \le \delta\sum_{i=-\infty}^{0}\varepsilon^{-i}|\xi_i| \\ &= \delta\varepsilon^{-k}\sum_{i=-\infty}^{k}\varepsilon^{k-i}|\xi_i|. \end{aligned}$$

Owing to that, we can write

$$\begin{aligned} |P_k U_0| &\le \delta^2\varepsilon^{n-k}\Big(\sum_{i=-\infty}^{k}\varepsilon^{k-i}|\xi_i|\Big)^2 \\ &\le \delta^2\varepsilon^{n-k}\sum_{i=-\infty}^{k}\varepsilon^{k-i}\sum_{i=-\infty}^{k}\varepsilon^{k-i}\xi_i^2 \\ &= \gamma\delta^2\varepsilon^{n-k}\sum_{i=-\infty}^{k}\varepsilon^{k-i}\xi_i^2. \end{aligned}$$

Using Lemma 5, we thus find

$$\begin{aligned} E|P_k U_0\psi(W_k)| &\le \gamma\delta^2\varepsilon^{n-k}\sum_{i=-\infty}^{k}\varepsilon^{k-i}E\{\xi_i^2\psi(W_k)\} \\ &\le \gamma^2\delta^2\varepsilon^{n-k}E\{\rho(W)\psi(W)\}. \end{aligned}$$

Finally

$$|A_{ni}| \le \gamma^2\delta^2 c_\varphi\sigma_\xi\varepsilon^{n-k}E\{\rho(W)\psi(W)\}. \quad (26)$$

To examine B_{ni}, observe that

$$|E\{\xi_i[\varphi(W_{n+k}) - \varphi(Q_{n+k})]\}| \le c_\varphi E\{|\xi_i P_k|\} = c_\varphi E|\xi_i| E|P_k|$$

which is bounded by $\delta c_\varphi \sigma_\xi \varepsilon^n \sum_{j=-\infty}^{k} \varepsilon^{k-j} E|\xi_j| \le \gamma\delta c_\varphi \sigma_\xi^2 \varepsilon^n$. Hence, using Lemma 5, we get $|B_{ni}| \le c_\varphi \sigma_\xi^2 \gamma\delta\varepsilon^{n-k} E\{\rho(W)\psi(W)\}$. This and (26) give

$$|A_{ni} - B_{ni}| \le c_\varphi \varepsilon^{n-k}(\sigma_\xi \gamma^2\delta^2 + \sigma_\xi^2\gamma\delta)E\{\rho(W)\psi(W)\}.$$

Using (4) again, we finally obtain

$$|S_n| \le c_\varphi \varepsilon^{n-k}\delta\gamma(\sigma_\xi\gamma^2\delta^2 + \sigma_\xi^2\gamma\delta)E\{\rho(W)\psi(W)\}$$

and complete the proof. ■

Having Lemma 6, we obtain

Lemma 10: Let μ satisfy (5). Let K satisfy (10)–(13). Let $\{h_n\}$ be a sequence of positive numbers. Let $0 \le k$. Then, for any $n \ge 1$,

$$\left|\operatorname{cov}\left[U_n K\left(\frac{y - Y_{n+k}}{h_n}\right), U_0 K\left(\frac{y - Y_k}{h_0}\right)\right]\right| \le \frac{h_0}{h_n}\varepsilon^{n-k}\phi(y)$$

with some $\phi(y)$ finite for every $y \in \mu(R)$ and independent of both k, and n.

Proof: Since (5) and (13) hold, for any nonnegative h,

$$\left|K\left(\frac{y-\mu(v)}{h}\right) - K\left(\frac{y-\mu(w)}{h}\right)\right| \le \frac{1}{h}c_K\left|\mu(v) - \mu(w)\right| \le \frac{1}{h}c_K c_\mu \left|v - w\right|.$$

Observe, moreover, that the examined covariance equals

$$\operatorname{cov}\left[U_n K\left(\frac{y - \mu(W_{n+k})}{h_n}\right), U_0 K\left(\frac{y - \mu(W_k)}{h_0}\right)\right].$$

Therefore, from Lemma 6, it follows that the covariance in the assertion is bounded in absolute value by

$$\begin{cases} \gamma_1 d_K h_0 \phi(y; h_0), & \text{for } 0 < n \le k, \\ \left(\gamma_2 d_K + \gamma_3 \dfrac{1}{h_n} c_K c_\mu\right) h_0 \varepsilon^{n-k}\varphi(y; h_0), & \text{for } 0 \le k < n, \end{cases}$$

with some γ_1, γ_2, γ_3, and

$$\varphi(y; h) = \frac{1}{h}E\left\{\rho(\mu^{-1}(Y))\left|K\left(\frac{y - Y}{h}\right)\right|\right\},$$

where $\rho(w) = c_1 + c_2 w^2$. Constants γ_1, γ_2, γ_3, as well as ρ are all independent of k, n, and K. Observe that $E|\rho(\mu^{-1}(Y))| \le E|\rho(W)| < \infty$. Thus, denoting $\varphi(y) = \sup_{h>0}\varphi(y; h)$, and applying Lemma 3 in Appendix A, we find $\varphi(y)$ finite for every $y \in \mu(R)$. In view of this and the fact that $0 \le \varepsilon < 1$, we can bound the examined quantity by

$$\begin{cases} \gamma_1 d_K h_0 \varepsilon^{n-k}\varphi(y), & \text{for } 0 < n \le k, \\ \left(\gamma_2 d_K + \gamma_3 \dfrac{1}{h_n} c_K c_\mu\right) h_0 \varepsilon^{n-k}\varphi(y), & \text{for } 0 \le k < n. \end{cases}$$

Since $\sup_n |h_n| < \infty$, the proof has been completed. ■

References

[1] H. Al-Duwaish, M. N. Karim, and V. Chandrasekar, Use of multilayer feedforward neural networks in identification and control of Wiener model, *IEE Proc.–Control Theory Appl.*, vol. 143, no. 3, pp. 255-258, May 1996.

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

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

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

[5] W. Greblicki, Nonparametric approach to Wiener system identification. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 44, no. 6, pp. 538–545, June 1997.

[6] W. Greblicki, Recursive identification of Wiener systems. *International Journal of Applied Mathematics and Computer Science*, vol. 11, no. 4, pp. 977-991, 2002.

[7] A. D. Kalafatis, L. Wang, and W. R. Cluett, Identification of Wiener-type nonlinear systems in a noisy environment. *International Journal of Control*, vol. 66, no. 6, pp. 923–941, 1997.

[8] D. Westwick, and M Verhaegen, Identifying MIMO Wiener systems using subspace model identification methods. *Signal Processing*, vol. 52, pp. 235–258, 1996.

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

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