Continuous-Time Wiener System Identification

Włodzimierz Greblicki

Abstract— **A continuous-time Wiener system is identified. The system consists of a linear dynamic subsystem and a memoryless nonlinear one connected in a cascade. The input signal is a stationary white Gaussian random process. The system is disturbed by stationary white random Gaussian noise. Both subsystems are identified from input-output observations taken at input and output of the whole system. The a priori information is very small and, therefore, resulting identification problems are nonparametric. The impulse response of the linear part is recovered by a correlation method, while the nonlinear characteristic is estimated with the help of the nonparametric kernel regression method. We prove convergence of the proposed identification algorithms and examine their convergence rates.**

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

I. Introduction

Various methods have been applied to nonlinear system identification, see, e.g., [1], [2]. The block oriented approach is based on the assumption that the identified system consists of simple elements called subsystems. In the Wiener system, a linear dynamic subsystem is followed by a memoryless nonlinearity. Such a system has been attracting attention of several authors, see, e.g., [3], [9], and [13]. Recently, the nonparametric methodology to the identification of discrete-time Wiener systems has been proposed, [5]–[7]. In this paper, we show that the approach can be successfully applied also to continuous-time Wiener systems. Our *a priori* knowledge about the system is very small which makes the problem close to those encountered in real situations. All we assume about the system is that the dynamic part is asymptotically stable, and that the nonlinear characteristic is invertible, and satisfies a Lipschitz condition. We propose algorithms to estimate the nonlinear characteristic, and the impulse response. The algorithms are numerically independent, i.e., can be calculated separately. We show that they converge to the impulse response and the characteristic, respectively. We also give convergence rates and present results of a simulation example.

II. Identification Problem

We identify a Wiener system shown in Fig. 1. The input random process $\{U(t); t \in (-\infty, \infty)\}$ is stationary, white, Gaussian, has zero mean, and an autocovariance function $\sigma_U^2 \delta(.)$. By δ, we denote the Dirac impulse. The system consists of two subsystems. The first is linear dynamic and the other memoryless nonlinear. The dynamic subsystem is described by the following state-space equation:

$$\left.\begin{aligned} \dot{X}(t) &= AX(t) + bU(t) \\ W(t) &= c^T X(t) \end{aligned}\right\}$$

in which A, b, and c are all unknown. Therefore $k(t) = c^T e^{At} b$, where k is the impulse response of the subsystem. The dynamic subsystem is asymptotically stable. Thus, $\int_0^\infty k^2(t)dt < \infty$ and, consequently, both $\{X(t); t \in (-\infty, \infty)\}$ and $\{W(t); t \in (-\infty, \infty)\}$ are stationary Gaussian processes. We also assume that

$$k(0) = 0 \tag{1}$$

which means that $c^T b = 0$. The class of systems satisfying the restriction is described in Section VI. The output of the

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

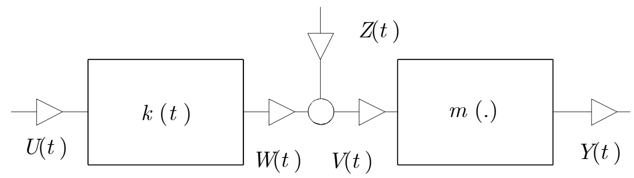

Fig. 1.

subsystem is disturbed by additive noise $\{Z(t); t \in (-\infty, \infty)\}$. Therefore, $V(t) = W(t) + Z(t)$. The noise $\{Z(t)\}$ is a zero-mean stationary white Gaussian random process with autocovariance function $\sigma_Z^2 \delta(.)$. The process is independent of $\{U(t)\}; t \in (-\infty, \infty)\}$. The other subsystem is nonlinear, memoryless and have a characteristic m. Hence $Y(t) = m(V(t))$. The characteristic is a Borel measurable function satisfying the following Lipschitz condition:

$$|m(x) - m(y)| \le c_m |x - y|, \tag{2}$$

some c_m, all x, y in R. Moreover, the characteristic m is invertible and its inverse is denoted by m^{-1}.

The aim of the paper is to identify both subsystems, i.e., to estimate the characteristic m of the nonlinear part, and the impulse response k of the linear subsystem from observations taken at input and output of the whole system, i.e., from $\{U(t), Y(t);\ t \in [0, T]\}$.

III. Identification Algorithms

Obviously, $V(t)$ has zero mean and variance $\sigma_V^2 = \sigma_U^2 \int_0^\infty k^2(t)dt + \sigma_Z^2$. Observe now that the pair $(U(t-\tau), V(t))$ has a Gaussian distribution with marginal variances σ_U^2, σ_V^2, and the correlation coefficient $(\sigma_U/\sigma_V)k(\tau)$. Therefore, $E\{U(t-\tau)|V(t)\} = (\sigma_U^2/\sigma_V^2)k(\tau)V(t)$. In this way, we have verified

Lemma 1: In the system,

$$E\{U(t-\tau)|Y(t) = y\} = \alpha m^{-1}(y),$$

where $\alpha = (\sigma_U^2/\sigma_V^2)k(\tau)$.

Next, observe $E\{U(t-\tau)Y(t)\} = E\{Y(t)E\{U(t-\tau)|Y(t)\}\}$. Applying Lemma 1, we find the quantity equal to $\alpha E\{Y(t)V(t)\}$. In this way, we have verified our next lemma.

Lemma 2: In the system,

$$E\{U(t-\tau)Y(t)\} = \beta k(\tau),$$

where $\beta = (\sigma_U/\sigma_V)E\{V(t)Y(t)\}$.

Having Lemma 2, we propose the following algorithm to estimate $\beta k(t)$:

$$\hat{\kappa}_T(t) = \frac{1}{T}\int_0^T U(\lambda - t)Y(\lambda)d\lambda. \tag{3}$$

In turn, Lemma 1 suggests the following estimate of $\alpha m^{-1}(y)$:

$$\hat{\mu}_T(y) = \frac{\displaystyle\int_0^T U(t-\tau)K\left(\frac{y - Y(t)}{h(T)}\right)dt}{\displaystyle\int_0^T K\left(\frac{y - Y(t)}{h(T)}\right)dt} \tag{4}$$

where K is a suitably selected Borel measurable kernel function and h is a bandwidth function, respectively. From the statistical viewpoint, the algorithm estimates the regression in Lemma 1 in a nonparametric way, see e.g., [10] or [11].

On the nonnegative Borel measurable kernel K, we impose the following constraints:

$$c_1 H(y) \le K(y) \le c_2 H(y), \tag{5}$$

some positive c_1, c_2, where a Borel measurable function H is such that

$$\sup_{-\infty<y<\infty} |H(y)| < \infty, \tag{6}$$

$$|y|H(y) \to 0 \quad \text{as} \quad |y| \to \infty, \tag{7}$$

$$cI_{\{|y|\le r\}}(y) \le H(y), \tag{8}$$

some positive c, r, where I is the indicator function, i.e., where $I_A(y)$ equals 1 or 0 for $y \in A$ or $y \notin A$, respectively. Moreover

$$|K(x) - K(y)| \le c_K |x - y|, \tag{9}$$

all x, y in R, some c_K. The positive function h satisfies the following conditions:

$$h(t) \to 0 \quad \text{as} \quad t \to \infty, \tag{10}$$

$$th^2(t) \to \infty \quad \text{as} \quad t \to \infty. \tag{11}$$

To explain the idea standing behind (4), rewrite the estimate as $\hat{\mu}_T(y) = \hat{g}(y)/\hat{f}(y)$, where

$$\hat{g}(y) = \frac{1}{Th(T)} \int_0^T U(t-\tau) K\left(\frac{y - Y(t)}{h(T)}\right) dt, \tag{12}$$

and

$$\hat{f}(y) = \frac{1}{Th(T)} \int_0^T K\left(\frac{y - Y(t)}{h(T)}\right) dt. \tag{13}$$

One can expect that, for large T, $\hat{g}(y)$ is close to

$$E\{U(-\tau) \frac{1}{h(T)} K\left(\frac{y - Y(0)}{h(T)}\right)$$

which, by virtue of our crucial Lemma 1, equals

$$E\left\{E\{U(-\tau)\,|Y(0)\}\frac{1}{h(T)} K\left(\frac{y - Y(0)}{h(T)}\right)\right\}$$
$$= E\left\{\alpha m^{-1}(Y(0))\frac{1}{h(T)} K\left(\frac{y - Y(0)}{h(T)}\right)\right\}.$$

Assume, for a while, that $Y(t)$ has a probability density f, and observing that the quantity equals

$$\int_{-\infty}^{\infty} \alpha m^{-1}(\eta) \frac{1}{h(T)} K\left(\frac{y-\eta}{h(T)}\right) f(\eta) d\eta.$$

Since, in view of (7) and (10),

$$\frac{1}{h(T)} K\left(\frac{y-\eta}{h(T)}\right)$$

converges to $\delta(y-\eta)\int_{-\infty}^{\infty} K(\xi)d\xi$, one can expect $\hat{g}(y)$ to converge to

$$\alpha m^{-1}(y) f(y) \int_{-\infty}^{\infty} K(\xi) d\xi$$

in the process of identification. We have assumed here that $\int_{-\infty}^{\infty} K(\xi)d\xi < \infty$. For similar reasons, $\hat{f}(y)$ can be expected to converge to $f(y)\int_{-\infty}^{\infty} K(\xi)d\xi$. Eventually, the above reasoning suggests that the estimate converges to $\alpha m^{-1}(y)$. We want, however, to stress that the assumption that $Y(t)$ has a probability density has been only temporary, and that, unless explicitly stated, $Y(t)$ may not possess a probability density.

Algorithms presented in the paper recover the impulse response of the dynamic subsystem and the characteristic of the nonlinear part only up to some multiplicative constants. This is caused by the cascade structure of the system, and the fact that the signal connecting the subsystems is not measured. Given the *a priori* information, no identification method can estimate these unknown constants. Observe, moreover, that, from the computational viewpoint, algorithms (3) and (4) are independent, i.e., can be calculated separately.

IV. Convergence of the algorithms

In this section, we show that our algorithms converge to the impulse response, and the nonlinear characteristic, respectively.

A. Dynamic Subsystem Identification

We shall now show that estimate (4) recovering the impulse response of the dynamic subsystem is consistent.

Theorem 1: Let the impulse response satisfy (1) and let m satisfy (2). For any $t \in (0, \infty)$,

$$E(\hat{\kappa}_T(t) - \beta k(t))^2 \to 0 \quad \text{as} \quad T \to \infty.$$

Proof: We have, $E\{\hat{\kappa}_T(t)\} = E\{U(0)Y(t)\}$ which, by virtue of Lemma 2, equals $\beta k(t)$. Thus $\hat{\kappa}_T(t)$ is an unbiased estimate of $\beta k(t)$. Its variance equals

$$\frac{1}{T^2} \int_0^T \int_0^T \operatorname{cov}[U(\lambda-\tau)Y(\lambda), U(\eta-\tau)Y(\eta)] d\lambda d\eta,$$

which, by virtue of Lemma 8 in Appendix, is bounded from above by c/T, some c independent of T. The proof has been completed. ∎

The mean square error converges to zero as fast as $O(1/T)$. Despite the fact that our *a priori* knowledge about the subsystem is nonparametric, the rate is the same as that shared by parametric algorithms.

B. Nonlinear Subsystem Identification

We shall now examine estimate (4) recovering the nonlinear characteristic. By μ, we denote the probability measure of $Y(t)$.

Theorem 2: Let the impulse response satisfy (2) and let m satisfy (2). Let the nonnegative Borel kernel K satisfy (5)-(9). Let the positive bandwidth function h satisfy (10)-(11). Then,

$$\hat{\mu}_T(y) \to \alpha m^{-1}(y) \quad \text{as} \quad T \to \infty \quad \text{in probability},$$

almost every (μ) $y \in R$, where $\alpha = (\sigma_U^2/\sigma_V^2)k(\tau)$.

Proof: We have $\hat{\mu}_T(y) = \hat{\xi}_T(y)/\hat{\eta}_T(y)$, where

$$\hat{\xi}_T(y) = \frac{\dfrac{1}{T}\displaystyle\int_0^T U(t-\tau) K\left(\frac{y-Y(t)}{h(T)}\right) dt}{EK\left(\dfrac{y-Y(t)}{h(T)}\right)},$$

and

$$\hat{\eta}_T(y) = \frac{\dfrac{1}{T}\displaystyle\int_0^T K\left(\frac{y-Y(t)}{h(T)}\right) dt}{EK\left(\dfrac{y-Y(t)}{h(T)}\right)}.$$

Observing

$$E\hat{\xi}_T(y) = \frac{E\left\{\alpha m^{-1}(Y(t)) K\left(\dfrac{y-Y(t)}{h(T)}\right)\right\}}{EK\left(\dfrac{y-Y(t)}{h(T)}\right)}$$

recalling (10) and applying Lemma 9 in Appendix to find

$$E\hat{\xi}_T(y) \to \alpha m^{-1}(y) \text{ as } T \to \infty, \tag{14}$$

almost every (μ) $y \in R$. In turn, $\operatorname{var}[\hat{\xi}_T(y)]$ equals

$$\frac{1}{T^2 E^2 K\left(\dfrac{y-Y(t)}{h(T)}\right)} \int_0^T \int_0^T \operatorname{cov}\left[U(\xi-\tau)K\left(\frac{y-Y(t)}{h(T)}\right),\right.$$
$$\left. U(\varsigma-\tau)K\left(\frac{y-Y(t)}{h(T)}\right)\right] d\xi d\varsigma.$$

Making use of Lemma 7 in the Appendix, we find the above expression not greater than

$$\frac{4\kappa^2\sigma_U^2\tau}{Th^2(T)} \frac{h^2(T)}{E^2K\left(\dfrac{y-Y(t)}{h(T)}\right)} + \frac{2c_K}{Th^2(T)}$$
$$\cdot \frac{h(T)}{EK\left(\dfrac{y-Y(t)}{h(T)}\right)} \frac{E\left\{\phi_\tau(Y(t))K\left(\dfrac{y-Y(t)}{h(T)}\right)\right\}}{EK\left(\dfrac{y-Y(t)}{h(T)}\right)},$$

where ϕ_τ is some function independent of T, such that $E\{|\phi_\tau(Y(t)|\} < \infty$. Invoking now Lemma 9 in Appendix and using (10), we find

$$\frac{E\left\{\phi_\tau(Y(t))K\left(\dfrac{y-Y(\varsigma)}{h(T)}\right)\right\}}{EK\left(\dfrac{y-Y(t)}{h(T)}\right)}$$

converging to $\phi_\tau(y)$ as $T \to \infty$, almost every (μ) $x \in R$. In turn, be virtue of Lemma 10 in Appendix

$$\sup_{T\in(0,\infty)} \frac{h(T)}{EK\left(\dfrac{y-Y(t)}{h(T)}\right)}$$

is finite for almost every (μ) $x \in R$. Thus

$$\operatorname{var}[\hat{\xi}_T(y)] = O\left(\frac{1}{Th^2(T)}\right), \tag{15}$$

almost every (μ) $y \in R$. Recalling now (14), we get

$$\hat{\xi}_T(y) \to \alpha m^{-1}(y) \text{ as } T \to \infty \text{ in probability,}$$

almost every (μ) $y \in R$. Since, using similar arguments, we can verify $\hat{\eta}_T(y) \to 1$ as $\to \infty$ in probability, almost every (μ) $y \in R$, the proof has been completed. ■

V. Convergence Rate and Simulation Example

As the kernel we can apply, e.g., $1/(1+y^2)$, $\exp(-|y|)$, $\exp(-y^2)$. Kernels with bounded support can be also employed. Examples of such are, e.g., a triangle kernel equal $1-|y|$ or 0 for $|y| \le 1$ or $|y| > 1$, respectively, and a parabolic kernel equal $1-y^2$ for $|y| \le 1$ and 0 otherwise, respectively. As far as the bandwidth function h is concerned, we can use, e.g., $h(t) = ct^{-\alpha}$, any positive c. In such a case, (10)-(11) is satisfied for $0 < \alpha < 1/2$.

Theorem 2 establishes convergence of algorithm (4). To examine its convergence rate assume that m^{-1} is differentiable. Therefore, $Y(t)$, has a density $f(y) = f_V(|m^{-1}(y)|)[m^{-1}(y)]'$,

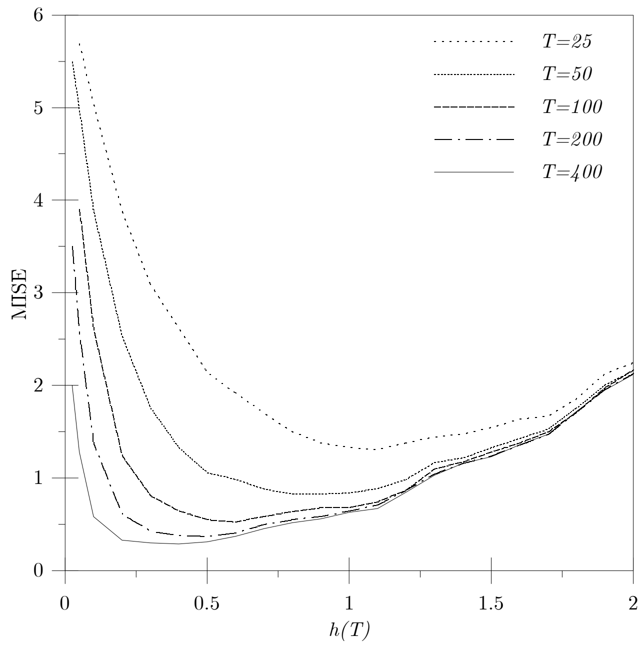

Fig. 2.

where f_V is the density of $V(t)$. Observe that f exists at each point at which the derivative of m^{-1} exists. Suppose, moreover, that $\int_{-\infty}^{\infty} K(x)dx < \infty$, $\int_{-\infty}^{\infty} xK(x)dx = 0$, and $\int_{-\infty}^{\infty} x^2K(x)dx < \infty$ and rewrite the estimate in the following form: $\hat{\mu}_T(y) = \hat{g}(y)/\hat{f}(y)$, where $\hat{g}(y)$, and $\hat{f}(y)$ are defined in (12), and (13), respectively. Applying now Lemma 11 in Appendix, we find $E\hat{g}_T(y) - \alpha m^{-1}(y)f(y)\int_{-\infty}^{\infty} K(x)dx = O(h^2(T))$. Recalling (15), and selecting $h(T) \sim T^{-1/6}$, we get

$$E(\hat{g}_T(y) - \alpha m^{-1}(y)f(y)\int_{-\infty}^{\infty} K(x)dx)^2 = O\left(\frac{1}{T^{2/3}}\right).$$

Since, for similar reasons, $\hat{f}_T(y)$ converges to $f(y)\int_{-\infty}^{\infty} K(x)dx$ at the same rate, we obtain, see, e.g., [5, Lemma B1],

$$\hat{\mu}_T(y) - \alpha m^{-1}(y) = O\left(\frac{1}{T^{1/3}}\right) \quad \text{as} \quad T \to \infty \quad \text{in probability.}$$

Comparing it with $1/T^{1/2}$, i.e., with the rate usually obtained for parametric inference, we find the rate derived by us quite good.

All results given above are asymptotic, i.e., hold for large T. As far as the behavior of estimate (4) for small or moderate T, all we can do now is to present results of a simulation example. In the example, the transfer function of the linear part is $6/(s+1)(s+4)$ while $m(v) = (1/2)v + v^2\operatorname{sign}(v)$. In the example, $\sigma_U^2 = 1$, and $\sigma_Z^2 = .1$. As the kernel, a parabolic one equal to 0 or $1-v^2$ for $|v| > 1$ or $|v| \le 1$, respectively, has been selected. In the paper, we have shown pointwise convergence of estimate (4), nevertheless, in the example, we have examined the mean integrated square error (MISE for short) defined as $\int_{-2}^{2}(\hat{\mu}(y) - \alpha m^{-1}(y))^2 dy$. The MISE has been measured for T varying from 25 to 800, and for various $h(T)$. Results are shown in Fig. 2. Observe that the problem of a proper selection of $h(T)$ is particularly important for large T, and that the error is especially sensitive to $h(T)$ for $h(T)$ close to the optimum.

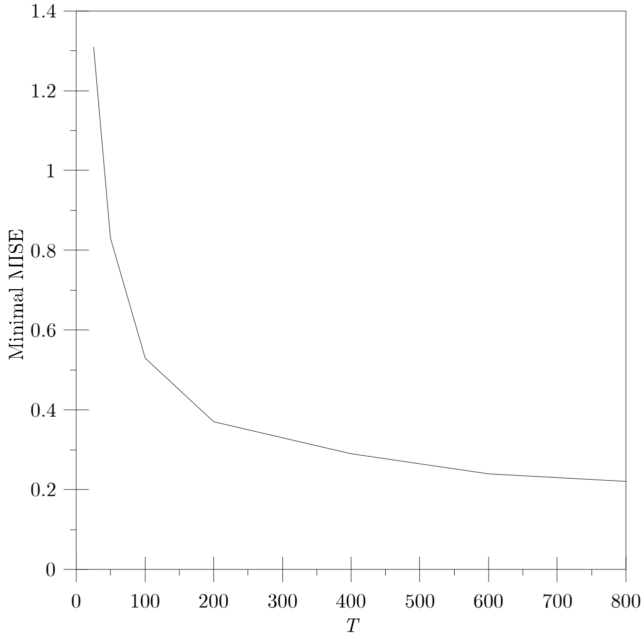

Fig. 3.

Observe, moreover, that $h(T)$ smaller than optimal should be avoided. In Fig. 3, the optimal MISE is shown.

VI. Final Remarks

We want to remark here that condition (1) is satisfied by a wide class of systems, i.e., by those with a transfer function $L(s)/M(s)$, where L, and M are polynomials such that the degree of M is greater than that of L by 2.

We should mention here that, for discrete-time systems, the result given in Lemma 2 has been already known, [4]. The same result, but for continuous-time systems has been presented in [3], however, for the nonlinear characteristic of a polynomial of a finite order, only. As far as the assumption that the input signal is Gaussian, it is crucial in their papers as well as in ours.

Appendix

A. The Wiener System

Lemma 3: Let (1), and (2) hold. Let a nonnegative Borel function φ satisfy the following Lipschitz condition:

$$|\varphi(x) - \varphi(y)| \le c_\varphi |x - y|, \tag{16}$$

some c_φ, all x, y in R. Let $0 < \tau$. Then,

$$\begin{aligned} &|\operatorname{cov}[U(t-\tau)\varphi(Y(t)), U(t-\tau+\lambda)\varphi(Y(t+\lambda))]| \\ &\le \begin{cases} 2d_\varphi^2\sigma_U^2(1+\delta(\lambda)), & \text{for } 0 \le \lambda \\ c_\varphi \|e^{A\lambda}\| E\{\rho_\tau(Y(t))\varphi(Y(t))\}, & \text{for } 0 < \tau < \lambda, \end{cases} \end{aligned}$$

where $d_\varphi = \sup_y \varphi(y)$, some function ρ_τ independent of φ, t, and λ. Moreover, $E|\rho_\tau(Y(t))| < \infty$.

Proof: First of all, observe that the covariance in the assertion equals

$$\operatorname{cov}[U(-\tau)\varphi(Y(0)), U(-\tau+\lambda)\varphi(Y(\lambda))]. \tag{17}$$

Suppose $0 \le \lambda$ and observe that the above quantity equals $S_1 - S_2$, where

$$S_1 = E\{U(-\tau)U(-\tau+\lambda)\varphi(Y(0))\varphi(Y(\lambda))\},$$

and

$$S_2 = E^2\{U(-\tau)\varphi(Y(0))\}.$$

Observing

$$|S_1| \le d_\varphi^2 E\{U|(-\tau))U(-\tau+\lambda)|\} \le d_\varphi^2\sigma_U^2(1+\delta(\lambda))$$

and $|S_2| \le d_\varphi^2\sigma_U^2$, we complete the proof of the first part of the inequality in the assertion.

To verify the other, suppose $0 < \tau < \lambda$. Denote

$$\xi(\lambda) = \int_0^\lambda k(\lambda-\eta)U(\eta)d\eta + Z(\lambda)$$

and observe that $V(\lambda) = c^T e^{A\lambda}X(0) + \xi(\lambda)$. Thus, by virtue of (2) and (16)

$$|\varphi(m(V(\lambda))) - \varphi(m(\xi(\lambda)))| \le c_\varphi c_m |c^T e^{A\lambda}X(0)|. \tag{18}$$

Since (1) holds, $Y(0)$ is independent of $U(t)$, $t \in [0,\infty)$. Therefore, $\xi(\lambda)$ and $Y(0)$ are independent. Because this implies mutual independence of pairs $(U(-\tau), Y(0))$ and $(U(-\tau+\lambda), \xi(\lambda))$,

$$\operatorname{cov}[U(-\tau)\varphi(Y(0)), U(-\tau+\lambda)\varphi(m(\xi(\lambda)))] = 0. \tag{19}$$

Having (18) and (19), we begin the main part of the proof. Recalling (17) and applying (19) to find the above quantity equal to

$$\begin{aligned} &\operatorname{cov}\{U(-\tau)\varphi(Y(0)), U(-\tau+\lambda)[\varphi(m(V(\lambda))) - \varphi(m(\xi(\lambda)))]\} \\ &= Q_1 - Q_2, \end{aligned}$$

where

$$\begin{aligned} Q_1 = E\{&U(-\tau)U(-\tau+\lambda)\varphi(Y(0)) \\ &\cdot [\varphi(m(V(\lambda))) - \varphi(m(\xi(\lambda)))]\}, \end{aligned}$$

and

$$\begin{aligned} Q_2 = {}&E\{U(-\tau)\varphi(Y(0))\} \\ &\cdot E\{U(-\tau+\lambda)[\varphi(m(V(\lambda))) - \varphi(m(\xi(\lambda)))]\}. \end{aligned}$$

Using (19), we find $|Q_1|$ not greater than

$$c_\varphi c_m |c| \|e^{A\lambda}\| E\{|U(-\tau)|U(-\tau+\lambda)|\varphi(Y(0))\|X(0)\|\}.$$

Since $< -\tau+\lambda$, the quantity equals

$$c_\varphi c_m |c| \|e^{A\lambda}\| E\{|U(-\tau+\lambda)|\} E\{\rho_1(Y(0))\varphi(Y(0))\},$$

where $\rho_1(Y(0)) = E\{|U(-\tau)|\|X(0)\|\,|U(0)\}$. For similar reasons $|Q_2|$ is bounded by

$$\begin{aligned} &c_\varphi c_m |c| \|e^{A\lambda}\| E\{U(-\tau+\lambda)\|X(0)\|\} E\{U(-\tau)(Y(0))\} \\ &\le c_\varphi c_m |c| \|e^{A\lambda}\| E\{|U(0)|\} E\{\|X(0)\|\} E\{\rho_2(Y(0))\varphi(Y(0))\}, \end{aligned}$$

where $\rho_2(Y(0)) = E\{|U(-\tau)|\,|Y(0)\}$. Thus, for $0 < \tau < \lambda$, the examined covariance does not exceed

$$c_\varphi c_m |c| \|e^{A\lambda}\| E\{\rho_\tau(Y(0))\varphi(Y(0))\},$$

some function ρ_τ independent of both φ, and λ. We have, thus, verified the first inequality in the assertion and, therefore, completed the proof. ■

We can now verify our next lemma.

Lemma 4: Let (1), and (2) hold. Let a nonnegative Borel function φ satisfy (16). Then

$$\left|\int_0^T\int_0^T \operatorname{cov}[U(\xi-\tau)\varphi(Y(\xi)),U(\varsigma-\tau)\varphi(Y(\varsigma))]d\xi d\varsigma\right| \le 4d_\varphi^2\sigma_U^2T(\tau+1)+2c_\varphi dTE\{\rho_\tau(Y(t))\varphi(Y(t)\},$$

where ρ_τ is some Borel function independent of T, and φ, such that $E|\rho_\tau(Y(t))|<\infty$, and where $d=\int_0^\infty \left\|e^{A\xi}\right\|d\xi$.

Proof: Noticing that the covariance in the assertion is a function of $\xi-\varsigma$ and taking into account the fact that $\int_0^T\int_0^T g(x-y)dxdy = 2\int_0^T(T-x)g(x)dx$, any g, we find the integral in the lemma equal to

$$2\int_0^T (T-\xi)\operatorname{cov}[U(\xi-\tau)\varphi(Y(\xi)),U(-\tau)\varphi(Y(0)]d\xi.$$

By virtue of Lemma 5, $|\int_0^\tau| \le 2d_\varphi^2\sigma_U^2T\tau$, and $|\int_\tau^T| \le c_\varphi dTE\{\rho_\tau(Y(t))\varphi(Y(t)\}$. The lemma has been verified. ■

The lemma leads to next.

Lemma 5: Let (1) and (2) hold. Let the nonnegative Borel kernel K satisfy (6), and (9). Let $0<\tau$. Then, for any positive h,

$$\left|\int_0^T\int_0^T \operatorname{cov}\left[U(\xi-\tau)K\left(\frac{y-Y(\xi)}{h}\right),\right.\right. \left.\left. U(\varsigma-\tau)K\left(\frac{y-Y(\varsigma)}{h}\right)\right]d\xi d\varsigma\right| \le 4\kappa^2\sigma_U^2T(\tau+1)+2c_Kc_mTd \cdot E\left\{\psi_\tau(Y(t))\frac{1}{h}K\left(\frac{y-Y(t)}{h}\right)\right\},$$

where $\kappa=\sup_y K(y)$ and where ψ_τ is a Borel function independent of both T and K, such that $E|\psi_\tau(Y(t))|<\infty$, and where d is as in the previous Lemma.

Arguing in a similar way we can verify

Lemma 6: Let (1) and (2) hold. Then,

$$\left|\int_0^T\int_0^T \operatorname{cov}[U(\xi-\tau)Y(\xi),U(\varsigma-\tau+)Y(\varsigma)]d\xi d\varsigma\right| \le c_1T(\tau+1),$$

some c_1 independent of T.

B. General results

Results given below are of general character.

Lemma 7: Let X be a random variable with the probability measure μ. Let ρ be a Borel measurable function such that $E|\rho(X)|\le\infty$. Let K be a nonnegative Borel measurable function satisfying (5)-(8). Then

$$\frac{E\left\{\rho(X)K\left(\dfrac{x-X}{h}\right)\right\}}{EK\left(\dfrac{x-X}{h}\right)} \to \rho(x) \quad \text{as} \quad h\to 0,$$

almost every (μ) $x\in R$.

The lemma can be found in [8, Lemma 1].

Lemma 8: Let a nonnegative Borel kernel K satisfy (5), and (8). Then, for any positive γ,

$$\sup_{h\in[0,\gamma]}\frac{h}{EK\left(\dfrac{x-X}{h}\right)}<\infty,$$

almost every (μ) $x\in R$.

Proof: From (5) and (8), it follows that the examined quantity is not greater than $h/c\mu(S_{rh}(x))$, which equals $(1/2c)\nu(S_{rh}(x))/\mu(S_{rh}(x))$, where $S_r(x)$ is a sphere of radius r centered at x, and where ν is the Lebesgue measure. Since $\nu(S_h(x))/\mu(S_h(x))$ has a finite limit as h tends to zero, almost all $(\mu\ )$ $x\in R$, see [12, p. 189], the proof has been completed. ■

Lemma 9: Let a random variable X have a probability density f, twice differentiable a point x. Let a Borel function ρ be also twice differentiable at the point. Let a Borel measurable kernel K be such that $\int_{-\infty}^\infty K(x)dx<\infty$, $\int_{-\infty}^\infty xK(x)dx=0$. Then

$$\frac{1}{h}E\left\{\rho(X)K(\frac{x-X}{h})\right\}-\rho(x)\int_{-\infty}^\infty K(x)dx=O(h^2).$$

Proof: Since the quantity in the assertion equals

$$\int_{-\infty}^\infty[\rho(x-hy)f(x-hy)-\rho(x)f(x)]K(y)dy,$$

to verify the lemma it suffices to expand $\rho(x-hy)$, and $f(x-hy)$ in a Taylor series. ■

References

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

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

[3] S.A. Billings, and S.Y. Fakhouri, "Identification of a class of nonlinear systems using correlation analysis", *IEE Proceedings*, vol. 125, pp. 691-697, 1978.

[4] D. Brillinger, "The identification of a particular nonlinear time series system", *Biometrica*, vol. 61, pp. 509-515.

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

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

[7] W. Greblicki, "Nonparametric approach to Wiener system identification", *IEEE Transactions on Circuits and Systems - I: Fundamental Theory and Applications*, in print.

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

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

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

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

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

[13] T. Wigren, "Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model", *IEEE Transactions on Automatic Control*, vol. 39, pp. 2191-2106, 1994.