Continuous-Time Hammerstein System Identification

Włodzimierz Greblicki

Abstract— **A continuous-time Hammerstein system, i.e., a system consisting of a nonlinear memoryless subsystem followed by a linear dynamic one, is identified. The system is driven and disturbed by white random signals. The *a priori* information about both subsystems is nonparametric which means that functional forms of both the nonlinear characteristic and the impulse response of the dynamic subsystem are unknown. An algorithm to estimate the nonlinearity is presented and its pointwise convergence to the true characteristic is shown. The impulse response of the dynamic part is recoverd with a correlation method. The algorithms are computationally independent. Results of a simulation example are given.**

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

I. Introduction

The block oriented approach to the nonlinear system identification assumes that the system consists of nonlinear memoryless and linear dynamic subsystems, see Bendat [1]. The approach has been used in many fields, e.g., to simulate the process of eye movements, see Huebner *et al.* [12], to model visual cortex, Emerson *et al.* [2], a distillation column, a heat exchanger, Eskinat *et al.* [3], Haber and Unbehauen [10], an electrical generator, Haber and Unbehauen [10], as well as a pH-process, Kalafatis *et al.* [13]. Another applications include automotive engineering, Ralston *et al.* [19], control, Zi-Qiang [22].

Of various structures, the Hammerstein one consisting of a nonlinear part followed by a dynamic one is of great interest. Most papers devoted to its identification assume that the *a priori* information about both subsystems is parametric which usually means that the nonlinear characteristic is a polynomial of a finite known degree and that the order of the state equation of the dynamic part is finite, and known. In the consequence, finite numbers of state equation and polynomial coefficients are estimated in the process of identification. It seems, however, that in many applications our *a priori* information about the system is smaller, i.e., neither the order of the dynamic subsystem is known nor the nonlinear characteristic is a polynomial. For example, the latter is the case when the characteristic is not continuous. Problems in which the *a priori* information so small that not coefficients but whole functions are estimated are called nonparametric.

The nonparametric approach to the Hammerstein system identification has been proposed and developed by Greblicki and Pawlak [6]–[9], and examined by Pawlak [17], Greblicki [5], Krzyzak [14], [15], as well as Pawlak and Hasiewicz [18]. In all those works time is discrete. The novelty of this paper is that we identify continuous-time systems. We present a theoretical motivation and then propose a kernel algorithm to recover the characteristic of the nonlinear subsystem and show that its consistency at all points at which the characteristic is continuous as well as at almost every (with respect to the Lebesgue measure) point. The impulse response of the dynamic part of the system is recovered in a standard way via estimating a correlation function.

Finally, we want to mention that the idea of the nonparametric identification of dynamic systems has been examined earlier by Georgiev [4], but effective results have been obtained, however, when associated with the block oriented approach.

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

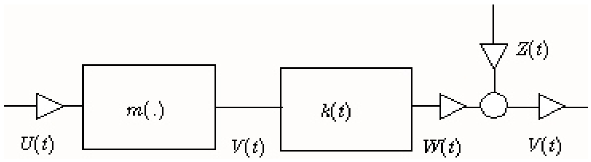

Fig. 1. The identified Hammerstein system.

II. Statement of the Problem

The continuous-time Hammerstein system, see Fig. 1, comprises two subsystems connected in a cascade. A nonlinear memoryless one is followed by a linear dynamic part. The system is driven by a stationary white random process $\{U(t); t \in (-\infty, \infty)\}$ with zero mean and autocovariance function $\sigma_U^2 \delta(.)$, where δ is the Dirac impulse. The random variable $U(t)$ has a probability density f. The first subsystem with input $U(t)$ and output $V(t)$ is nonlinear, memoryless and $V(t) = m(U(t))$, where m is the characteristic of the subsystem. Thus, $\{V(t); t \in (-\infty, \infty)\}$ is a stationary white random process with autocovariance function denoted by $\sigma_V^2 \delta(.)$. We assume that m is a Borel measurable function satisfying the following condition:

$$|m(u)| \leq c_1 + c_2|u| \tag{1}$$

with some $c_1, c_2 > 0$. The fact that σ_U^2 is finite together with (1) imply $\sigma_V^2 < \infty$. The impulse response of the linear subsystem is denoted by k which means that $W(t) = \int_{-\infty}^{t} k(t-\xi)V(\xi)d\xi$. We assume that

$$\int_0^\infty k^2(t)dt < \infty. \tag{2}$$

Owing to (2), and the fact that σ_V^2 is finite, output $W(t)$ of the dynamic part is a random variable. Therefore, the problem is well posed, i.e., can be described with probabilistic methods.

Obviously, $\{W(t); t \in (-\infty, \infty)\}$ is a stationary correlated random process. Output of the dynamic subsystem is disturbed by additive stationary white random noise $\{Z(t); t \in (-\infty, \infty)\}$, i.e., a processes with autocovariance function $\sigma_Z^2 \delta(.)$. We assume, moreover, that $EZ(t) = 0$ and that the process is independent of the input signal. Therefore, $Y(t) = W(t) + Z(t)$, where $Y(t)$ is output of the whole system.

We identify both subsystems, i.e., estimate the nonlinear characteristic m and the impulse response k from observations $\{U(t), Y(t); t \in [0, \infty)\}$ taken at input and output of the whole system. Since the signal $V(t)$ interconnecting the subsystems is not measured, both m and k can be estimated up to some constants only.

To simplify further considerations and formulas, we impose the following additional restrictions on the density f of $U(t)$ and the nonlinearity m:

$$f \text{ is an even function,} \tag{3}$$

$$m \text{ is an odd function.} \tag{4}$$

Restriction (3) implies $EU(t) = 0$. In turn, this, and (4) yield $EV(t) = 0$.

For the sake of simplicity, (U, V) denotes a pair of random variables distributed like $(U(t), V(t))$. For convenience, for a fixed, positive λ, we denote $\beta = k(\lambda)$. Moreover, $\alpha = E\{Um(U)\}$.

III. Identification Algorithms

The basis for recovering the impulse response is an obvious equality $\text{cov}[Y(t+\tau), U(t)] = \alpha k(\tau)$. Thus, as an estimate of

$\alpha k(\tau)$ we take

$$\hat{\kappa}(\tau;t) = \frac{1}{t}\int_0^t Y(\tau+\xi)U(\xi)d\xi. \tag{5}$$

To introduce the algorithm identifying the nonlinear subsystem, we need the following

Lemma 1: In the system,

$$E\{Y(t+\lambda)|U(t)=u\} = \beta m(u).$$

Proof: The lemma is a simple consequence of Lemma 2 in Appendix, and the fact that the zero mean disturbance $Z(t)$ is independent of the input signal. ∎

Thus, recovering $\beta m(u)$ is equivalent to estimating the regression $E\{Y(t+\lambda)\,|U(t)=u\}$. To estimate the regression, we propose the following algorithm:

$$\hat{\mu}(u;t) = \frac{\displaystyle\int_0^t Y(\lambda+\xi)K\left(\frac{u-U(\xi)}{h(t)}\right)d\xi}{\displaystyle\int_0^t K\left(\frac{u-U(\xi)}{h(t)}\right)d\xi}, \tag{6}$$

where, K and h are suitably selected kernel and bandwidth functions, respectively.

A discrete-time version of (6) has been introduced in the statistical literature independently by Nadaraya, [16], and Watson, [20], see also Härdle [11]. They have estimated $m(u)$ from independent pairs of observations (U_i, Y_i), $i = 1, 2, \cdots$, where $Y_i = m(U_i) + Z_i$, and where Z_i is a random disturbance. In other words, they have identified a nonlinear memoryless system. Next, the estimate has been applied to recover the nonlinear characteristic of a memoryless subsystem of a discrete-time Hammerstein dynamic system, see [5]–[9], [14], [15], [18]. Recovering the nonlinearity of a Hammerstein system have appeared, however, more complicated since, due to dynamics, consecutive output observations are dependent. In this paper, we also deal with a Hammerstein system, but, as we have already mentioned, time is now continuous.

The positive Borel measurable function h satisfies the following two restrictions:

$$h(t) \to 0 \text{ as } t\to\infty, \tag{7}$$

$$th(t) \to \infty \text{ as } t\to\infty, \tag{8}$$

while the Borel measurable kernel K the following ones:

$$\sup_{-\infty<v<\infty} |K(v)| < \infty, \tag{9}$$

$$\int_{-\infty}^{\infty} |K(v)|\,dv < \infty, \tag{10}$$

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

with, depending on the context, $\varepsilon = 0$ or $\varepsilon > 0$.

Observe that for $h(t) = ct^{-\gamma}$, with any positive c, (7) and (8) hold for $0 < \gamma < 1$. As the kernel, one can select, e.g., a window function equal 1 or zero according to $|u|$ does not exceed or is greater than 1. Other examples are $e^{-|u|}$, e^{-u^2}, or $1/(1+u^2)$.

IV. Convergence of Algorithms

We shall now show that algorithms (5) and (6) converge to $\alpha k(\tau)$ and $\beta m(u)$ as t tends to infinity, respectively.

Theorem 1: Let $E\{U^2m^2(U)\} < \infty$. Then

$$E(\hat{\kappa}(\tau;t) - \alpha k(\tau))^2 \to 0 \text{ as } t\to\infty.$$

Proof: Since $E\hat{\kappa}(\tau;t) = \alpha k(\tau)$, it suffices to examine variance which equals $(1/t^2)[P_1(t)+P_2(t)]$ with

$$P_1(t) = \int_0^t\int_0^t \operatorname{cov}[W(\tau+\xi)U(\xi), W(\tau+\eta)U(\eta)]d\xi d\eta,$$

and

$$P_2(t) = \sigma_Z^2\int_0^t\int_0^t \operatorname{cov}[U(\xi),U(\eta)]d\xi d\eta,$$

respectively. Since, $EU = 0$, by virtue of Lemma 3 in Appendix, covariance under the integral in $P_1(t)$ equals

$$\delta(\xi-\eta)k^2(\tau)\operatorname{var}[Um(U)] + \delta(\xi-\eta)\sigma_V^2\sigma_U^2\textstyle\int_0^\infty k^2(\zeta)d\zeta$$
$$+k(\tau+\xi-\eta)k(\tau-\xi+\eta)E^2\{Um(U)\}.$$

Hence

$$P_1(t) = k^2(\tau)t\operatorname{var}[Um(U)] + \sigma_V^2\sigma_U^2 t\int_0^\infty k^2(\xi)d\xi$$
$$+ E^2\{Um(U)\}\int_0^t\int_0^t k(\tau+\xi-\eta)k(\tau-\xi+\eta)d\xi d\eta.$$

The double integral in the last term equals $2\int_0^t k(\tau+\lambda)k(\tau-\lambda)d\lambda$ and is bounded in absolute value by $2\int_0^\infty k^2(\xi)d\xi$. Therefore, $|P_1(t)| \le c_1 t$, some c_1 independent of t. Since a similar inequality holds for $P_2(t)$, we get $\operatorname{var}[\hat{\kappa}(\tau;t)] = O(t^{-1})$ and complete the proof. ∎

Going through the proof again, we observe that $E(\hat{\kappa}(\tau;t) - \alpha k(\tau))^2 = O(t^{-1})$. Therefore, the rate at which the estimate converges is the same as that usually encountered in parametric inference.

Our next theorem establishes convergence of the algorithm recovering the nonlinearity.

Theorem 2: Let h satisfy (7), and (8). Let the kernel K satisfy (9) and (10). If, moreover, (11) holds with $\varepsilon = 0$, then

$$\hat{\mu}(u;t) \to \beta m(u) \text{ as } t\to\infty \text{ in probability}$$

at every point u at which both m and f are continuous, and $f(u) > 0$. If (11) is satisfied with $\varepsilon > 0$, then the convergence takes place at almost every (with respect to the Lebesgue measure) point u at which $f(u) > 0$.

Proof: Suppose that (9), (10), and (11) with $\varepsilon = 0$, are satisfied. Let u be a point at which both m and f are continuous. Denote

$$\hat{\xi}(u;t) = \frac{1}{th(t)}\int_0^t Y(\lambda+\zeta)K\left(\frac{u-U(\zeta)}{h(t)}\right)d\zeta$$

$$\hat{\eta}(u;t) = \frac{1}{th(t)}\int_0^t K\left(\frac{u-U(\zeta)}{h(t)}\right)d\zeta$$

and observe $\hat{\mu}(u;t) = \hat{\xi}(u;t)/\hat{\eta}(u;t)$. From Lemma 1, we get

$$E\hat{\xi}(u;t) = \frac{1}{h(t)}E\left\{Y(\lambda)K\left(\frac{u-U(0)}{h(t)}\right)\right\}$$
$$= \frac{1}{h(t)}\beta E\left\{m(U)K\left(\frac{u-U}{h(t)}\right)\right\}.$$

Now, applying Lemma 5 in Appendix, we obtain $E\hat{\xi}(u;t) \to \beta f(u)m(u)\int_{-\infty}^{\infty}K(v)dv$ as $t\to\infty$.

In turn, $\operatorname{var}[\hat{\xi}(u;t)] = (1/t^2h^2(t))[R_1(u;t) + R_2(u;t)]$, where

$$R_1(u;t) = \int_0^t\int_0^t \operatorname{cov}\left[W(\lambda+\xi)K\left(\frac{u-U(\xi)}{h(t)}\right),\right.$$
$$\left.W(\lambda+\eta)K\left(\frac{u-U(\eta)}{h(t)}\right)\right]d\xi d\eta,$$

and

$$R_2(u;t)=\sigma_Z^2\int_0^t\int_0^t \mathrm{cov}\left[K\left(\frac{u-U(\xi)}{h(t)}\right),K\left(\frac{u-U(\eta)}{h(t)}\right)\right]d\xi d\eta,$$

respectively. Since (7) and (8) hold, by virtue of Lemma 4 in Appendix, $R_1(u;t)$ is bounded in absolute value by $\rho(u)(th(t)+th^2(t))$ with a finite $\rho(u)$. Since, moreover, the same holds for $R_2(u;t)$, we find, $\mathrm{var}[\hat{\xi}(u;t)]=O(1/th(t))$. Finally, $\hat{\xi}(u;t)\to\beta m(u)f(u)\int_{-\infty}^{\infty}K(v)dv$ as $t\to\infty$ in probability.

Since, using similar arguments, one can verify that $\hat{\eta}(u;t)\to f(u)\int_{-\infty}^{\infty}K(v)dv$ as $t\to\infty$ in probability at the point, the first part of the theorem has been verified. To show the other, it suffices to repeat all arguments and apply the almost everywhere version of Lemma 5. ■

Imposing some smoothness restrictions on m, we can give convergence rates for our identification algorithm recovering the nonlinearity. We fix a point u and begin with the following observation:

$$E\hat{\xi}(u;t)-\beta f(u)m(u)\int_{-\infty}^{\infty}K(v)dv$$
$$=\beta\int_{-\infty}^{\infty}\left[m\left(u-h(t)v\right)f\left(u-h(t)v\right)-m(u)f(u)\right]K(v)dv.$$

We select the kernel so that $\int_{-\infty}^{\infty}vK(v)dv=0$ and assume that both m and f have three derivatives bounded in a neighborhood of the point u. Now, expanding m and f in Taylor series at the point, we find the quantity equal to $\beta h^2(t)\rho(u)\int_{-\infty}^{\infty}v^2K(v)dv+o(h^2(t))$ with $\rho(u)=m'(u)f'(u)+(1/2)[m(u)f''(u)+m''(u)f(u)]$. Thus, $E\hat{\xi}(u;t)-\beta f(u)m(u)\int_{-\infty}^{\infty}K(v)dv=O(h^2(t))$. As variance equals $O(1/th(t))$, see the proof of Theorem 2, selecting $h(t)\sim t^{-1/5}$, we get $E(\hat{\xi}(u;t)-\beta m(u)f(u)\int_{-\infty}^{\infty}K(v)dv)^2=O(t^{-4/5})$. Since, for similar reasons, $E(\hat{\eta}(u;t)-f(u)\int_{-\infty}^{\infty}K(v)dv)^2=O(t^{-4/5})$, we finally obtain

$$\hat{\mu}(u;t)-\beta m(u)=O(t^{-2/5})\text{ as }t\to\infty\text{ in probability.}$$

Thus, the convergence rate is not very much worse than $t^{-1/2}$, i.e., the rate typical for parametric problems.

V. Simulation Example

To illustrate behavior of our algorithm recovering the nonlinearity, we present results of a numerical simulation. In the example, $m(u)$ is not continuous and equals $u+1/2$ or $u-1/2$, for $0\le u$ or $u<0$, respectively. The dynamic subsystem has a transfer function $K(s)=12/\left(s+1\right)\left(s+4\right)$ while disturbance $Z(.)$ is Gaussian with variance 0.1. Despite the fact that we have shown pointwise convergence, the mean integrated square error (MISE in short) defined as $E\int_{-2}^{2}(\hat{\mu}(u;t)-\beta m(u))^2du$ has been empirically calculated. In the algorithm, the kernel is a window function while $h(t)=ct^{-\gamma}$.

Fig. 2 shows the empirical MISE for $\gamma=-1/6$ with c varying from 0 to 8. Results suggest that the best c is close to 0.9. Observe, moreover, that one should avoid selecting the coefficient smaller than optimal since the error increases rapidly as c decreases from the optimal value. Having in mind that the estimate is consistent for γ in the interval $(0,1)$, we have shown MISE for γ varying from 0 to 4, see Fig. 3. Observe that γ close to 1/2, i.e., γ lying in the centre of the admissible interval is best, and that too small γ are highly not recommended. Next, selecting both coefficients safely, i.e. setting $c=3$ and $\gamma=4/5$, we have examined MISE against t. Results are depicted in Fig. 4 with the solid line. Observe that the error gets small very rapidly for t smaller than 200. In addition, the dashed line shows the error for $K(s)=1$, i.e. for the system with no dynamical part. We see that dynamics increases MISE. In tis context, one can observe, we omit strict reasoning, that m is recovered from observations noised by not only $Z(.)$ but also an additional disturbance incurred by the dynamic subsystem which explains why, for $K(s)=1$, MISE is smaller.

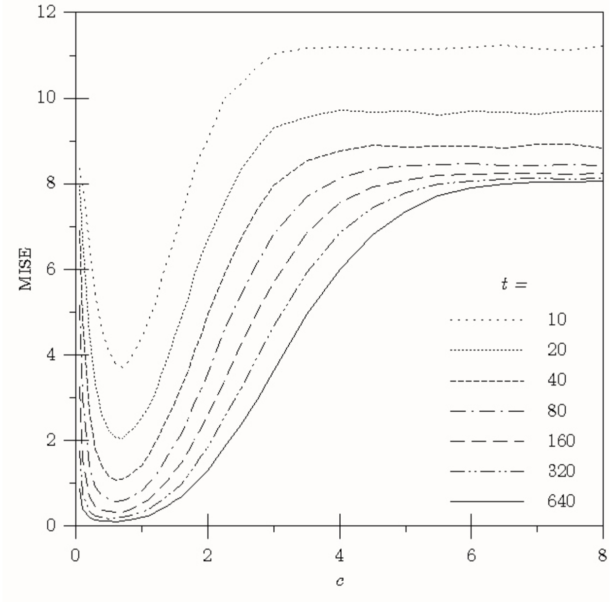

Fig. 2. MISE versus c; $h(t)=ct^{-1/6}$.

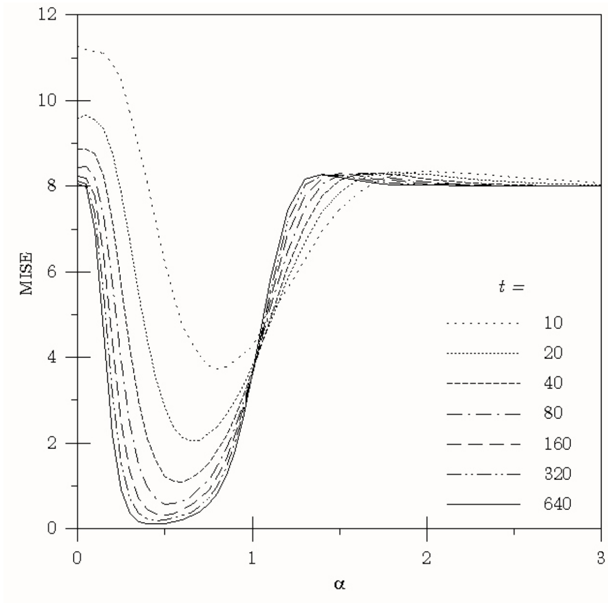

Fig. 3. MISE versus c; $h(t)=ct^{-1/6}$.

VI. Final Remarks

Identifying the continuous-time Hammerstein system, we do not impose any restriction concerning functional forms on both the nonlinear characteristic and the impulse response. Thus, the *a priori* information about the system is nonparametric. In the parametric approach, on the contrary, both forms are known and each is described by a finite number of parameters. Thus, it seems that, with this respect, nonparametric methods are more

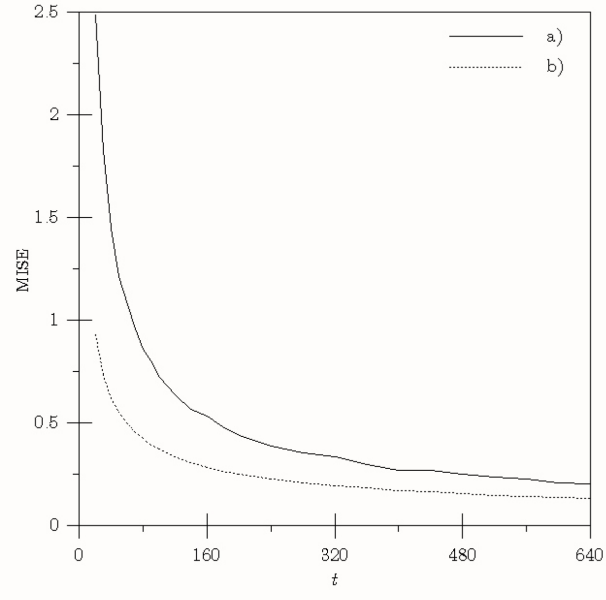

Fig. 4. MISE versus t for $h(t) = 3t^{-1/6}$; a) $K(s) = 12/(s+1)(s+4)$, b) $K(s) = 1$.

suitable to solve problems encountered in real life applications.

In the parametric identification, considerations are usually confined to nonlinear characteristics being polynomials of a finite known degree. Nevertheless, the approximation theorem of Weierstrass is an argument often used in favor of the polynomial approach. According to the theorem, any continuous function can be uniformly approximated in any interval to any degree of accuracy by power polynomials. It suggests that applying a polynomial of a high degree, one could expect good results. This argument gets, however, weak when the estimated nonlinearity is not continuous. In this context, we point out once again that our algorithm successfully identifies the nonlinear subsystem even when its characteristic is not continuous.

Concluding, we emphasize that parametric and nonparametric approaches do not compete with each other since they are designed to work under different *a priori* information. When the information is large enough, the former is applied, when isn't, only the latter can recover unknown characteristics.

Finally, we want also to stress that, from the computational viewpoint, algorithms (5) and (6) are independent, i.e., that estimates $\hat{\kappa}(\tau;t)$ and $\hat{\mu}(u;t)$ are calculated separately.

Appendix

A. The System

Lemma 2: In the system,

$$E\{W(t+\lambda)|U(t)\} = k(\lambda)m(U(t)).$$

Proof: To prove the lemma, it suffices to verify $E\{W(t+\lambda)|V(t)\} = k(\lambda)V(t)$. To do this, denote $\rho(V(t)) = E\{W(t+\lambda)|V(t)\}$, and $\psi(V(t)) = k(\lambda)V(t)$. Since, due to (3) and (4), $EV = 0$, for any Borel function ϕ, we get $E\{W(t+\lambda)\phi(V(t))\} = E\{\rho(V)\phi(V)\}$. On the other hand, for any ϕ, $E\{W(t+\lambda)(V(t))\} = k(\lambda)E\{V\phi(V)\}$. Thus, for any ϕ, $E\{\rho(V)\phi(V)\} = k(\lambda)E\{V\phi(V)\}$. Observe, moreover, that, for any ϕ, $E\{\psi(V)\phi(V))\} = k(\lambda)E\{V\phi(V)\}$. Hence, $\rho(V) = \psi(V)$ with probability 1. This implies the desired equality and completes the proof. ■

Remark 1: Observe that if restrictions (3) and (4) are not satisfied,

$$E\{W(t+\lambda)|U(t) = u\} = k(\lambda)m(u) + \gamma,$$

with $\gamma = E\{m(U)\}\int_0^\infty k(\xi)d\xi$.

Lemma 3: For any Borel function φ,

$$\begin{aligned}
&\operatorname{cov}\left[W(t+\lambda)\varphi(U(t)), W(\tau+\lambda)\varphi(U(\tau))\right] \\
&= \delta(t-\tau)k^2(\lambda)\operatorname{var}\left[m(U)\varphi(U)\right] \\
&+ \delta(t-\tau)\sigma_V^2 E\{\varphi^2(U)\}\int_0^\infty k^2(v)dv \\
&+ k(\lambda)k(\tau-t+\lambda)E\{\varphi(U)\}E\{m^2(U)\varphi(U)\} \\
&+ k(\lambda)k(t-\tau+\lambda)E\{\varphi(U)\}E\{m^2(U)\varphi(U)\} \\
&+ k(t-\tau+\lambda)k(\tau-t+\lambda)E^2\{m(U)\varphi(U)\}.
\end{aligned}$$

Proof: The covariance in the assertion equals

$$\begin{aligned}
&\int_{-\infty}^{t+\lambda}\int_{-\infty}^{\tau+\lambda} k(t+\lambda-\xi)k(\tau+\lambda-\eta) \\
&\times \operatorname{cov}\left[m(U(\xi))\varphi(U(t)), m(U(\eta))\varphi(U(\tau))\right]d\xi d\eta.
\end{aligned}$$

Since $\{U(t)\}$ is white noise, and $Em(U) = 0$, we get

$$\begin{aligned}
&\operatorname{cov}\left[m(U(\xi))\varphi(U(t)), m(U(\eta))\varphi(U(\tau))\right] \\
&= \delta(t-\tau)\delta(\tau-\xi)\delta(\tau-\eta)\operatorname{var}\left[m(U)\varphi(U)\right] \\
&+ \delta(t-\tau)\delta(\xi-\eta)E\{\varphi^2(U)\}\operatorname{var}\left[m(U)\right] \\
&+ \delta(t-\xi)\delta(\eta-\xi)E\{\varphi(U)\}E\{m^2(U)\varphi(U)\} \\
&+ \delta(\tau-\xi)\delta(\tau-\eta)E\{\varphi(U)E\{m^2(U)\varphi(U)\} \\
&+ \delta(t-\eta)\delta(\tau-\xi)E^2\{m(U)\varphi(U)\}
\end{aligned}$$

and complete the proof. ■

We are now in a position to verify our next lemma.

Lemma 4: Let K satisfy (9), (10), and (11). Then,

$$\begin{aligned}
&\int_0^t\int_0^t \left|\operatorname{cov}\left[W(\xi+\lambda)K\left(\frac{u-U(\xi)}{h(t)}\right),\right.\right. \\
&\left.\left. W(\eta+\lambda)K\left(\frac{u-U(\eta)}{h(t)}\right)\right]\right| d\xi d\eta \\
&\le \rho(u)\left[th(t) + th^2(t)\right]
\end{aligned}$$

with some function ρ. For $\varepsilon = 0$, ρ is finite at every point u at which both m and f are continuous. For $\varepsilon > 0$, ρ is finite at almost every (the Lebesgue measure) point u.

Proof: From Lemma 3, it follows that the double integral in the assertion is bounded in absolute value by $Q_1(u;t) + Q_2(u;t) + 2Q_3(u;t) + Q_4(u;t)$ with

$$Q_1(u;t) = k^2(\lambda)t\operatorname{var}\left[m(U)K\left(\frac{u-U}{h(t)}\right)\right],$$

$$Q_2(u;t) = \sigma_V^2 tE\left\{K^2\left(\frac{u-U}{h(t)}\right)\right\}\int_0^\infty k^2(\xi)d\xi,$$

$$\begin{aligned}
Q_3(u;t) &= k(\lambda)E\left\{K\left(\frac{u-U}{h(t)}\right)\right\}E\left\{m^2(U)K\left(\frac{u-U}{h(t)}\right)\right\} \\
&\times \int_0^t\int_0^t k(\lambda+\xi-\eta)d\xi d\eta,
\end{aligned}$$

and

$$\begin{aligned}
Q_4(u;t) &= E^2\left\{m(U)K\left(\frac{u-U}{h(t)}\right)\right\} \\
&\int_0^t\int_0^t k(\lambda-\xi+\eta)k(\lambda+\xi-\eta)d\xi d\eta,
\end{aligned}$$

respectively.

Observe that $|Q_1(u;t)| \leq \kappa k^2(\lambda)\rho_1(u)th(t)$, where

$$\rho_1(u) = \sup_{0<h} \frac{1}{h} E \left| m^2(U) K\left(\frac{u-U}{h}\right)\right|,$$

and where $\kappa = \sup_u |K(u)|$. In turn, $|Q_2(u;t)| \leq \kappa\sigma_V^2\rho_2(u)th(t)\int_0^\infty k^2(\xi)d\xi$, where

$$\rho_2(u) = \sup_{0<h} \frac{1}{h} E \left| K\left(\frac{u-U}{h}\right)\right|$$

while

$$|Q_3(u;t)| \leq \rho_1(u)\rho_2(u)k(\lambda)h^2(t)\left|\int_0^t\int_0^t k(\lambda+\xi-\eta)d\xi d\eta\right|.$$

Since the integral in the obtained expression equals $2\int_0^t(t-\tau)k(\lambda-\tau)d\tau$ and is bounded in absolute value by $2t\int_0^\lambda |k(\xi)|d\xi$. Denoting the quantity by ε_1, we find $|Q_3(u;t)| \leq \varepsilon_1\rho_1(u)\rho_2(u)k(\lambda)th^2(t)$. Further, the absolute value of $Q_4(u;t)$ is bounded by

$$\rho_3^2(u)h^2(t)\left|\int_0^t\int_0^t k(\eta-\xi+\lambda)k(\lambda-\eta+\xi)d\xi d\eta\right|,$$

where

$$\rho_3(u) = \sup_{0<h} \frac{1}{h} E \left| m(U) K\left(\frac{u-U}{h}\right)\right|.$$

The double integral in the above expression equals $2\int_0^t(t-\tau)k(\lambda-\tau)k(\lambda+\tau)d\tau$ which absolute value is not greater than $2t\int_0^\infty k^2(\xi)d\xi$. Denoting the quantity by ε_2, we get $|Q_4(u;t)| \leq \varepsilon_2\rho_1(u)\rho_2(u)k(\lambda)th^2(t)$.

Applying now the continuous version of Lemma 5, we find that for K satisfying (9), (10), and (11) with $\varepsilon = 0$, all functions ρ_1, ρ_2, and ρ_3 are finite for every point u at which m, and f are continuous. Since the almost everywhere version can be verified in a similar way, the proof has been completed. ∎

B. General result

Lemma 5: Let f be a density function of a random variable U. Let m be a Borel function such that $E|m(U)| < \infty$. Let a Borel measurable kernel K satisfy (9), (10). If, moreover, (11) holds with $\varepsilon = 0$, then

$$\frac{1}{h}E\left\{m(U)K\left(\frac{u-U}{h}\right)\right\} \to m(u)f(u)\int_{-\infty}^{\infty}K(v)dv \text{ as } h \to 0 \tag{12}$$

at every point u at which both f and m are continuous and

$$\sup_{0<h} \frac{1}{h} E \left| m(U) K\left(\frac{u-U}{h}\right)\right| < \infty \tag{13}$$

at the same points. If, in addition, (11) holds with $\varepsilon > 0$, then (12) and (13) are satisfied at almost every (the Lebesgue measure) point u (in particular, at every point u at which both f and m are continuous).

Proof: Continuous and almost everywhere versions of (12) can be found in [21, Theorems 9.9 and 9.13]. (13) follows immediately. ∎

References

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

[2] R. Emerson, M. Korenberg, and M. Citron, "Identification of complex-cell intensive nonlinearities in a cascade model of cat visual cortex," *Biological Cybernetics*, vol. 66, pp. 291–300, 1992.

[3] E. Eskinat, S.H. Johnson, and L. Luyben, "Use of Hammerstein models in identification of nonlinear systems," *American Institute of Chemical Engineers Journal*, vol. 37, pp. 255–268, 1991.

[4] A. Georgiev, "Nonparametric kernel algorithm for recovering of functions from noisy measurements with applications," *IEEE Transactions on Automatic Control*, vol. AC-30, pp. 782-784, 1985.

[5] W. Greblicki, "Nonlinearity estimation in Hammerstein systems based on ordered observations," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1224–1233, May 1996.

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

[7] W. Greblicki and M. Pawlak, "Hammerstein system identification by nonparametric regression estimation," *International Journal of Control*, vol. 45, pp. 343–354, 1987.

[8] W. Greblicki and M. Pawlak, "Nonparametric identification of Hammerstein systems," *IEEE Transactions on Information Theory*, vol. IT-35, pp. 409–418, March 1989.

[9] W. Greblicki and M. Pawlak, "Cascade non-linear system identification by a non-parametric method," *International Journal of Control*, vol. 25, pp. 129-153, 1994.

[10] R. Haber and H. Unbehauen, "Structure identification of nonlinear dynamic systems - a survey of input/output approaches," *Automatica*, vol. 24, pp. 651–677, 1990.

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

[12] W.P. Huebner, G.M. Saidel, and R.J. Leigh, "Nonlinear parameter estimation applied to a model of smooth pursuit eye movements," *Biological Cybernetics*, vol. 62, pp. 265–273, 1990.

[13] A. Kalafatis, L. Wang, and W. Cluett, "Identification of Wiener-type nonlinear systems in a noisy environment," *International Journal of Control*, vol. 66,pp. 923–941, 1997.

[14] A. Krzyzak, "On estimation of a class of nonlinear systems by the kernel regression estimate," *IEEE Transactions on Information Theory*, vol. 36, pp. 141–152, February 1990.

[15] A. Krzyzak, "Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation," *IEEE Transactions on Information Theory*, vol. 38, pp. 1323–1338, July 1992.

[16] E. Nadaraya, "On regression estimators" *Theory Probab. Appl.*, vol. 9, pp. 157–159, 1964.

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

[18] 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 1989.

[19] J.C. Ralston, A.M. Zoubir, and B. Boashash, "Identification of a class of nonlinear systems under stationary non-Gaussian excitation," *IEEE Transactions on Signal Processing*, vol. 45, pp. 719–735, March 1997.

[20] G.S. Watson, "Smooth regression analysis," *Sankhyā, Ser. A*, vol. 26, pp. 359–372, 1964.

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

[22] L. Zi-Qiang, "Controller design oriented model identification method for Hammerstein system," *Automatica*, vol. 39, pp. 767–771, 1993.