Continuous-time Hammerstein System Identification From Sampled Data

Włodzimierz Greblicki

Abstract— **A continuous-time Hammerstein system driven by a random signal is identified from observations sampled in time. The sampling may be uniform or not. The *a priori* information about the system is nonparametric, functional forms of both the nonlinear characteristic and the impulse response are completely unknown. Three kernel algorithms, one off-line and two semirecursive are presented. Their convergence to the true characteristic of the nonlinear subsystem is shown. The distance between consecutive sampling times must not decrease too fast for the algorithms to converge.**

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

I. Introduction

In a Hammerstein system a memoryless nonlinear subsystem is followed by a linear dynamic one. The identification of the nonlinear subsystem has been the subject of a significant number of works, see [1] and survey papers [2], [3]. This paper belongs to those in which the *a priori* information about the system is nonparametric. Classes of all possible nonlinear characteristics and equations of the dynamic subsystem are so wide that they cannot be represented in any parametric form. Moreover, the system is driven by a random signal and the output is disturbed by a random noise.

Kernel nonparametric algorithms were applied in [4]-[8] to recover the nonlinearity in discrete-time systems. Recently kernel algorithms have been applied to recover the nonlinear characteristics in continuous-time systems, [9], [10]. The novelty of the paper is that we infer from sampled input-output signals, i.e., not from functions of time but from number sequences being a result of sampling. The distance between consecutive sampling instants may change in time, sampling can be uniform or not. We show that our kernel algorithms, one off-line and two semirecursive, converge to the unknown characteristic and give convergence rates. We also examine the relation between various ways of sampling and convergence and demonstrate which methods of sampling are good and which are not. The time interval the sampling is performed on must increase to infinity sufficiently quickly.

II. Statement of the problem

The identified continuous-time Hammerstein system is shown in Fig. 1. The input signal $\{U(t); t \in (-\infty, \infty)\}$ is a stationary white random process with autocovariance function $\sigma_U^2\delta(t)$, where $\delta(t)$ is the Dirac impulse and $\sigma_U^2 < \infty$. The random variable $U(t)$ has a probability density f. The first subsystem has a characteristic m which means that $V(t) = m(U(t))$. We assume that m is a Borel measurable function such that $EV^2(t) < \infty$. This is, e.g., the case for $|m(u)| \leq c_1 + c_2|u|$ with any $c_1, c_2 > 0$. Thus $\{V(t); t \in (-\infty, \infty)\}$ is a stationary white random process with autocovariance function $\sigma_V^2\delta(t)$, where $\sigma_V^2 < \infty$.

W. Greblicki is with the Institute of Computer Engineering, Control, and Robotics, Wrocław University of Technology, Wyspiańskiego 27, 50-370 Wrocław, Poland, E-mail: wlodzimierz.greblicki@pwr.wroc.pl

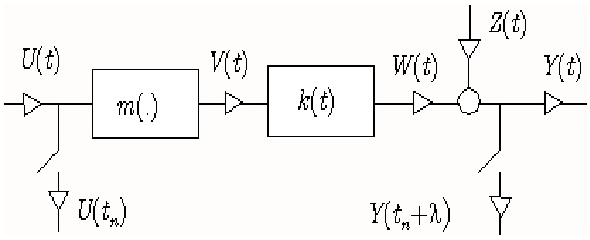

Fig. 1. Hammerstein system.

The unknown impulse response $k(t)$ of the dynamic subsystem satisfies the following inequality:

$$|k(t)| \leq \alpha e^{-\beta t}, \qquad (1)$$

with some unknown positive α and β, which is always the case for a stable dynamic subsystem described by a state equation. Clearly, $W(t) = \int_{-\infty}^{t} k(t-\tau)V(\tau)d\tau$. Due to the fact that $\sigma_V^2 < \infty$ and (1), $\{W(t); t \in (-\infty, \infty)\}$ is a stationary correlated random process such that $\sigma_W^2 < \infty$. Additive stationary random noise $\{Z(t); t \in (-\infty, \infty)\}$ has zero mean, finite variance and is independent of input signal. Therefore, $Y(t) = W(t) + Z(t)$, where $Y(t)$ is output of the whole system.

We estimate the nonlinear characteristic m from sampled observations, i.e., from pairs $(U(t_1), Y(t_1+\lambda))$, $(U(t_2), Y(t_2+\lambda))$, $(U(t_3), Y(t_3+\lambda))$,... with $\lambda \geq 0$ and $0 \leq t_1 < t_2 < \ldots$. The inner signal of the system is not measured. In the paper, λ is fixed and known.

It will be convenient to denote $\Delta_n = \min_{i=1,2,\ldots,n-1} |t_{i+1} - t_i|$. Observe that Δ_n is monotonous and converges to either a positive number or zero as $n \to \infty$. We shall say that the sampling is uniform if $t_{n+1} - t_n = \Delta$, some $\Delta > 0$, for all $n = 1, 2, \ldots$. If a_n/b_n has a nonzero limit as $n \to \infty$, we write $a_n \sim b_n$.

For the sake of simplicity, U denotes a random variable distributed like $U(t)$.

III. Algorithms

The basis for our identification algorithms is the fact that $E\{Y(t+\lambda)|U(t) = u\} = k(\lambda)m(u) + \int_0^\infty k(\tau)d\tau Em(U)$, see [9]. To make our considerations easier, we assume that f and m are even and odd functions, respectively. Owing to that

$$E\{Y(t+\lambda)|U(t) = u\} = k(\lambda)m(u), \qquad (2)$$

since $Em(U) = 0$. We thus recover m up to some unknown multiplicative constant $k(\lambda)$ which is caused by the fact that the inner signal in the system of a cascade structure is not measured.

Recovering $k(\lambda)m(u)$ is thus equivalent to estimating a regression function which will be achieved with the following off-line algorithm:

$$\hat{\mu}_n(u) = \frac{\sum_{i=1}^{n} Y(t_i+\lambda) K\left(\frac{u-U(t_i)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{u-U(t_i)}{h_n}\right)}, \quad (3)$$

where K is a Borel measurable kernel while $\{h_n\}$ a sequence of nonnegative numbers, both suitably selected. The kernel is such that

$$\int K(v)dv = 1,\ \sup_v |K(v)| < \infty,\ \lim_{|v|\to\infty} vK(v) = 0. \quad (4)$$

The nonnegative number sequence satisfies the following restrictions:

$$h_n \to 0 \text{ as } n \to \infty, \quad (5)$$

$$nh_n \to \infty \text{ as } n \to \infty. \quad (6)$$

We also examine the following two semirecursive versions:

$$\tilde{\mu}_n(u) = \frac{\sum_{i=1}^{n} Y(t_i+\lambda) \frac{1}{h_i} K\left(\frac{u-U(t_i)}{h_i}\right)}{\sum_{i=1}^{n} \frac{1}{h_i} K\left(\frac{u-U(t_i)}{h_i}\right)}, \quad (7)$$

$$\bar{\mu}_n(u) = \frac{\sum_{i=1}^{n} Y(t_i+\lambda) K\left(\frac{u-U(t_i)}{h_i}\right)}{\sum_{i=1}^{n} K\left(\frac{u-U(t_i)}{h_i}\right)}. \quad (8)$$

In (7) and (8), in addition to (5), $\{h_n\}$ satisfies restrictions

$$\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h_i} \to 0 \text{ as } n \to \infty, \quad (9)$$

$$\sum_{n=1}^{\infty} h_n = \infty, \quad (10)$$

respectively.

IV. Convergence of the off-line algorithm

We shall now show convergence of estimate (3). Clearly, $\hat{\mu}_n(u) = \hat{\xi}_n(u)/\hat{\eta}_n(u)$ with

$$\hat{\xi}_n(u) = \frac{1}{nh_n}\sum_{i=1}^{n} Y(t_i+\lambda) K\left(\frac{u-U(t_i)}{h_n}\right),$$

$$\hat{\eta}_n(u) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{u-U(t_i)}{h_n}\right).$$

We now study asymptotic properties of $\hat{\xi}_n(u)$ and $\hat{\eta}_n(u)$. In Lemmas 1 and 2, bias and variance are examined.

Lemma 1: Let the kernel K satisfy (4). If (5) holds, then

$$E\hat{\xi}_n(u) \to k(\lambda)m(u)f(u) \text{ as } n \to \infty$$

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

Proof: From (2), we get

$$\begin{aligned} E\hat{\xi}_n(u) &= \frac{1}{h_n} E\left\{Y(\lambda+t)K\left(\frac{u-U(t)}{h_n}\right)\right\} \\ &= k(\lambda)\frac{1}{h_n}E\left\{m(U)K\left(\frac{u-U}{h_n}\right)\right\} \end{aligned}$$

and using (20) in the appendix verify the lemma. ■

In the next lemma and the remaining of the paper

$$\varphi_n(\lambda) = \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \left[k(\lambda+|t_i-t_j|) + k(\lambda-|t_i-t_j|)\right].$$

Lemma 2: Let the kernel K satisfy (4). If (5) and (6) hold, then

$$\begin{aligned} \mathrm{var}[\hat{\xi}_n(u)] &= \frac{1}{nh_n}\left(\sigma_Z^2 + m^2(u)\right) f(u) \\ &\times \int K^2(v)dv + o\left(\frac{1}{nh_n}\right) + k(\lambda) \\ &\times m^2(u) f^2(u)\left(1+o(1)\right)\frac{1}{n^2}\varphi_n(\lambda) \end{aligned} \quad (11)$$

and

$$\mathrm{var}[\hat{\xi}_n(u)] = O\left(\frac{1}{nh_n}\right) + O\left(\frac{1}{n\Delta_n}\right) \quad (12)$$

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

Proof: Obviously, $\mathrm{var}[\hat{\xi}_n(u)] = P_n(u) + Q_n(u) + R_n(u)$ with

$$\begin{aligned} P_n(u) &= \frac{\sigma_Z^2}{n^2h_n^2}\sum_{i=1}^{n}\mathrm{var}\left[K\left(\frac{u-U(t_i)}{h_n}\right)\right] \\ &= \frac{\sigma_Z^2}{nh_n^2}\mathrm{var}\left[K\left(\frac{u-U}{h_n}\right)\right], \end{aligned}$$

$$\begin{aligned} Q_n(u) &= \frac{1}{n^2h_n^2}\sum_{i=1}^{n}\mathrm{var}\left[W(t_i+\lambda)K\left(\frac{u-U(t_i)}{h_n}\right)\right] \\ &= \frac{1}{nh_n^2}\mathrm{var}\left[W(t+\lambda)K\left(\frac{u-U(t)}{h_n}\right)\right], \end{aligned}$$

$$\begin{aligned} R_n(u) &= \frac{1}{n^2h_n^2} \\ &\times \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\mathrm{cov}\left[W(t_i+\lambda)K\left(\frac{u-U(t_i)}{h_n}\right),\right. \\ &\left. W(t_j+\lambda)K\left(\frac{u-U(t_j)}{h_n}\right)\right]. \end{aligned}$$

Applying (21) in Appendix, we find $P_n(u) = (1/nh_n)\sigma_Z^2 f(u) \int K^2(v)dv + o(1/nh_n)$. From Lemma 8 in the appendix we obtain $Q_n(u) = (1/nh_n)k^2(\lambda)m^2(u)f(u)\int K^2(v)dv +$

$o(1/nh_n)$. Applying Lemma 8 again, we get $R_n(u) = k(\lambda)m^2(u)f^2(u)\,(1+o(1))\,n^{-2}\varphi_n(\lambda)$ which completes the proof of (11). To verify (12) it suffices to use Lemma 9 in the appendix. ∎

From Lemmas 1 and 2 we easily obtain

Theorem 1: Let the kernel K satisfy (4). Let h_n satisfy (5) and (6). If

$$n\Delta_n \to \infty \text{ as } n \to \infty, \tag{13}$$

then

$$\hat{\mu}_n(u) \to k(\lambda)m(u) \text{ as } n \to \infty \text{ in probability.}$$

at every point u at which both m and f are continuous and $f(u) > 0$.

Proof: Convergence $\hat{\xi}_n(u) \to k(\lambda)m(u)$ in probability is implied by Lemmas 1 and 2. Since, for similar reasons $\hat{\eta}_n(u) \to f(u)$ as $n \to \infty$ in probability, the theorem follows. ∎

V. Convergence of semirecursive algorithms

We now pass to semirecursive algorithms. Rewriting (7) as $\tilde{\mu}_n(u) = \tilde{\xi}_n(u)/\tilde{\eta}_n(u)$ with

$$\tilde{\xi}_n(u) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}Y(t_i+\lambda)K\left(\frac{u-U(t_i)}{h_i}\right),$$

$$\tilde{\eta}_n(u) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}K\left(\frac{u-U(t_i)}{h_i}\right).$$

Both $\tilde{\xi}_n(u)$ and $\tilde{\eta}_n(u)$ can be calculated on-line. For example,

$$\tilde{\xi}_n(u) = \tilde{\xi}_{n-1}(u) - \frac{1}{n}\left(\tilde{\xi}_{n-1}(u) - \frac{1}{h_n}Y(t_n+\lambda)K\left(\frac{u-U(t_n)}{h_n}\right)\right).$$

Lemma 3: Let the kernel K satisfy (4). If (5) holds, then

$$E\tilde{\xi}_n(u) \to k(\lambda)m(u)f(u) \text{ as } n \to \infty$$

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

Proof: From (2), we get

$$E\tilde{\xi}_n(u) = k(\lambda)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}E\left\{m(U)K\left(\frac{u-U}{h_i}\right)\right\}$$

which, in view of (20), converges to $k(\lambda)m(u)f(u)$ as $n \to \infty$. ∎

Lemma 4: Let the kernel K satisfy (4). If (5) holds, then

$$\begin{aligned}\operatorname{var}[\tilde{\xi}_n(u)] &= \left(\sigma_Z^2 + m^2(u)\right)f(u)\\ &\times \int K^2(v)dv\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h_i} + o\left(\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h_i}\right)\\ &+ k(\lambda)m^2(u)f^2(u)\,(1+o(1))\,\frac{1}{n^2}\varphi_n(\lambda)\end{aligned} \tag{15}$$

at every point u at which both m and f are continuous. If, moreover, (9) is satisfied, then

$$\operatorname{var}[\tilde{\xi}_n(u)] = O\left(\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h_i}\right) + O\left(\frac{1}{n\Delta_n}\right) \tag{16}$$

at the same points.

Proof: Obviously, $\operatorname{var}[\tilde{\xi}_n(u)] = P_{1n}(u) + Q_{1n}(u) + R_{1n}(u)$ with

$$\begin{aligned}P_{1n}(u) &= \frac{\sigma_Z^2}{n^2}\sum_{i=1}^{n}\frac{1}{h_i^2}\operatorname{var}\left[K\left(\frac{u-U(t_i)}{h_i}\right)\right]\\ &= \frac{\sigma_Z^2}{n^2}f(u)\int K^2(v)dv\sum_{i=1}^{n}\frac{1}{h_i}\,(1+o_i(1))\\ &= \sigma_Z^2 f(u)\int K^2(v)dv + o\left(\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h_i}\right),\end{aligned}$$

$$\begin{aligned}Q_{1n}(u) &= \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h_i^2}\operatorname{var}\left[W(t_i+\lambda)K\left(\frac{u-U(t_i)}{h_i}\right)\right]\\ &= k^2(\lambda)m^2(u)f(u)\int K^2(v)dv\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h_i}\\ &+ o\left(\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h_i}\right),\end{aligned}$$

$$\begin{aligned}R_{1n}(u) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{1}{h_ih_j}\\ &\times \operatorname{cov}\left[W(t_i+\lambda)K\left(\frac{u-U(t_i)}{h_i}\right),\right.\\ &\left. W(t_j+\lambda)K\left(\frac{u-U(t_j)}{h_j}\right)\right]\\ &= k(\lambda)m^2(u)f^2(u)\,(1+o(1))\,\frac{1}{n^2}\varphi_n(\lambda).\end{aligned}$$

We have applied (21) in the appendix and used Lemma 8. The proof of (15) is completed. Arguing as in the proof of Lemma 2, we verify (16) and complete the proof. ∎

Our next theorem is now obvious.

Theorem 2: Let the kernel K satisfy (4). Let (5) and (9) hold. If, moreover, (14) holds, then

$$\tilde{\mu}_n(u) \to k(\lambda)m(u) \text{ as } n \to \infty$$

at every point u at which m and f are continuous and $f(u) > 0$.

To examine algorithm (8), we denote

$$\bar{\xi}_n(u) = \frac{1}{\sum_{i=1}^{n}h_i}\sum_{i=1}^{n}Y(t_i+\lambda)K\left(\frac{u-U(t_i)}{h_i}\right),$$

$$\bar{\eta}_n(u) = \frac{1}{\sum_{i=1}^{n}h_i}\sum_{i=1}^{n}K\left(\frac{u-U(t_i)}{h_i}\right)$$

and observe that $\bar{\mu}_n(u) = \bar{\xi}_n(u)/\bar{\eta}_n(u)$. Both $\bar{\xi}_n(u)$ and $\bar{\eta}_n(u)$ can be calculated recursively. For example

$$\bar{\xi}_n(u) = \bar{\xi}_{n-1}(u) - \frac{h_n}{\sum_{i=1}^{n}h_i}\times\left(\bar{\xi}_{n-1}(u) - \frac{1}{h_n}Y(t_n+\lambda)K\left(\frac{u-U(t_n)}{h_n}\right)\right).$$

The next lemma concerning the estimate is given without proof.

Lemma 5: Let the kernel K satisfy (4). If (5) and (10) hold, then

$$E\bar{\xi}_n(u) \to k(\lambda)m(u)f(u) \text{ as } n\to\infty,$$

$$\begin{aligned}
&\operatorname{var}[\bar{\xi}_n(u)]\\
&= \left(\sigma_Z^2 + m^2(u)\right) f(u) \int K^2(v)dv \frac{1}{\sum_{i=1}^n h_i} + o\left(\frac{1}{\sum_{i=1}^n h_i}\right)\\
&+ k(\lambda)m^2(u)f^2(u)\left(1+o(1)\right)\frac{1}{\left(\sum_{i=1}^n h_i\right)^2}\varphi_n(\lambda)\\
&= O\left(\frac{1}{\sum_{i=1}^n h_i}\right) + O\left(\frac{1}{\Delta_n\sum_{i=1}^n h_i}\right) \qquad (17)
\end{aligned}$$

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

From Lemma 5 we obtain

Theorem 3: Let the kernel K satisfy (4). Let (5) and (10) hold. If

$$\Delta_n \sum_{i=1}^n h_i \to \infty \text{ as } n\to\infty, \qquad (18)$$

then

$$\bar{\mu}_n(u) \to k(\lambda)m(u) \text{ as } n\to\infty$$

at every point u at which m and f are continuous, and $f(u) > 0$.

VI. Sampling and convergence

To guarantee that the algorithms converge, we make their numerators and denominators converge, see proofs of Theorems 1, 2, and 3. We achieve this goal by appropriate choice of K and $\{h_n\}$. Additional conditions concern Δ_n, i.e., the way sampling is performed, see (14) for algorithms (3) and (7), and (18) for algorithm (8).

We shall now examine (14) and (18). Focusing on (14) we observe that it is satisfied for both uniform sampling and that in which the distance $t_{n+1} - t_n$ between consecutive observations increases. In both cases the horizon of observations $T = \lim_{n\to\infty} t_n$ is infinite. It is interesting that (3) and (7) can be consistent also when $t_{n+1} - t_n$ decreases to zero. For example, for $t_n = n^{1/2}$ we have $t_i - t_{i-1} \sim i^{-1/2}$ which implies $\Delta_n \sim n^{-1/2}$. Therefore $n\Delta_n \to \infty$ as $n\to\infty$ and, as a consequence, (14) is satisfied. The horizon of observations is, however, also infinite. As we later show, compared with uniform, such sampling may worsen the convergence rate.

The horizon T must be infinite anyway. If T is finite, neither $\operatorname{var}[\hat{\xi}_n(u)]$ nor $\operatorname{var}[\tilde{\xi}_n(u)]$ converge to zero as $n\to\infty$. To verify that, recall (11) and (15), suppose that $k(t) = e^{-t}$ for $t \ge 0$, and observe that, since $|t_i - t_j| \le T$,

$$\begin{aligned}
\frac{1}{n^2}\varphi_n(\lambda) &\ge \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1, j\ne i}^n k(\lambda + |t_i - t_j|)\\
&\ge e^{-\lambda}\frac{1}{n^2}\sum_{i=1}^n \sum_{j=1, j\ne i}^n e^{-|t_i - t_j|}\\
&\ge e^{-\lambda}e^{-T}\frac{n-1}{n}
\end{aligned}$$

which converges to a positive limit as $n\to\infty$. All that happens despite the fact that input and output processes are sampled infinitely many times. Thus, $T=\infty$ is necessary for $\operatorname{var}[\hat{\xi}_n(u)]$ and $\operatorname{var}[\tilde{\xi}_n(u)]$ to converge to zero as $n\to\infty$.

It is interesting that even the infinite horizon of observations, i.e., $T=\infty$, does not guarantee convergence. For example, for $t_n = \ln n$, T is also infinite, but $\hat{\xi}_n(u)$ as well as $\tilde{\xi}_n(u)$ may be not consistent for any $\{h_n\}$. Since $t_j - t_i = \ln(j/i)$, assuming that $k(t) = e^{-t}$ for $t\ge 0$, we again recall (11) and (15) to find

$$\begin{aligned}
\frac{1}{n^2}\varphi_n(\lambda) &\ge e^{-\lambda}\frac{2}{n^2}\sum_{i=1}^n \sum_{j=1, j\ne i}^{i-1} e^{t_j - t_i}\\
&= e^{-\lambda}\frac{2}{n^2}\sum_{i=1}^n \frac{1}{i}\sum_{j=1}^{i-1} j = e^{-\lambda}\frac{n-1}{2n}.
\end{aligned}$$

Hence, neither $\operatorname{var}[\hat{\xi}_n(u)]$ not $\operatorname{var}[\tilde{\xi}_n(u)]$ vanish as $n\to\infty$. Notice in passing that, for such sampling, (14) is not satisfied since $n\Delta_n \sim 1$. To explain this fact, we recall that (13) should converge to zero, which means that covariance terms under the sum incurred by consecutive observations must be small enough. Observing that from Lemma 8 it follows that, for, e.g., $\lambda = 0$,

$$\begin{aligned}
&\frac{1}{h_n^2}\operatorname{cov}\left[W(t_i+\lambda)K\left(\frac{u-U(t_i)}{h_n}\right),\right.\\
&\left.W(t_i+\lambda)K\left(\frac{u-U(t_j)}{h_n}\right)\right]\\
&= e^{-|t_i - t_j|}\left(m^2(u)f^2(u) + o_{t_i}(1) + o_{t_j}(1)\right),
\end{aligned}$$

we conclude that $t_n - t_i$, $i = 1, 2, \ldots, n-1$, must all be large enough. In particular, $t_n - t_{n-1}$ must be large enough, too. As we have shown, $t_n - t_{n-1} \sim n^{-1}$, i.e., $t_n = \ln n$, does not suffice, despite the fact that the horizon of observations is infinite.

We now pass to algorithm (8) and restriction (18). For the algorithm, T must be infinite too. To show that, suppose to the contrary, i.e., that T is finite. Assuming that $k(t) = e^{-t}$ for $t\ge 0$, we recall (17) and observe that

$$\frac{1}{\left(\sum_{i=1}^n h_i\right)^2}\varphi_n(\lambda) \ge e^{-\lambda}e^{-T} - \frac{\sum_{i=1}^n h_i^2}{\left(\sum_{i=1}^n h_i\right)^2}e^{-\lambda}e^{-T}$$

which converges to a positive limit $e^{-\lambda}e^{-T}$ as $n\to\infty$.

VII. Convergence rate

Imposing some smoothness restrictions on f and m, we give convergence rate for our identification algorithms recovering the nonlinearity. Assuming that f has three bounded derivatives and expanding $f(u - h_n v)$ in a Taylor series, we get

$$\begin{aligned}
E\hat{\eta}_n(u) - f(u) &= \int \left[f(u - h_n v) - f(u)\right] K(v)dv\\
&= -h_n f'(u)\int vK(v)dv
\end{aligned}$$

$$+\frac{1}{2}h_n^2 f''(u)\int v^2K(v)dv + o(h_n^2).$$

Selecting the kernel such that $\int vK(v)dv = 0$, we thus obtain $|E\hat{\eta}_n(u) - f(u)| = O(h_n^2)$. Recalling (12), we thus find

$$E(\hat{\eta}_n(u)-f(u))^2 = O(h_n^4)+O\left(\frac{1}{nh_n}\right)+O\left(\frac{1}{n\Delta_n}\right). \quad (19)$$

Let $t_n = n^\alpha$, $\alpha > 0$. For $1 < \alpha$ the distance between consecutive sampling instants increases, for $\alpha = 1$ sampling is uniform. Since in both cases $\Delta_n =$ const, the second term in (19) dominates the third and

$$E(\hat{\eta}_n(u) - f(u))^2 = O(h_n^4) + O\left(\frac{1}{nh_n}\right).$$

Selecting $h_n \sim n^{-1/5}$, we thus find $E(\hat{\eta}_n(u) - f(u))^2 = O(n^{-4/5})$, i.e., $|\hat{\eta}_n(u) - f(u)| = O(n^{-2/5})$ in probability. Since, for m having three derivatives, the same rate holds for $\hat{\xi}_n(u)$,

$$|\hat{\mu}_n(u) - k(\lambda)m(u)| = O(n^{-2/5}) \text{ in probability.}$$

For $0 < \alpha < 1$ the distance between consecutive sampling instants decreases. Selecting the same h_n, we observe that, for $4/5 \le \alpha < 1$, the second term is not dominated by third which means that the rate still holds. For $0 < \alpha < 4/5$, $1/n\Delta_n = O(n^{-\alpha})$ and the third term dominates the others. The whole sum is of order $O(n^{-\alpha})$. Hence,

$$|\hat{\mu}_n(u) - k(\lambda)m(u)| = O(n^{-\alpha/2}) \text{ in probability.}$$

Therefore, for the distance $t_n - t_{n-1} \sim n^{\alpha-1}$ between sampling instants decreasing quickly, i.e., for $0 < \alpha < 4/5$, the rate gets worse than $O(n^{-2/5})$, i.e., that obtained for uniform sampling.

Observe that for $t_n - t_{n-1} \sim n^{-1+\varepsilon}$ with arbitrarily small $\varepsilon > 0$, both $\hat{\xi}_n(u)$ and $\hat{\eta}_n(u)$ converge. It is interesting that, as we have shown in section VI, for $t_n - t_{n-1} \sim n^{-1}$, i.e., for $t_n - t_{n-1}$ decreasing a little bit faster, neither $\hat{\xi}_n(u)$ nor $\hat{\eta}_n(u)$ converge. In such a case $t_n = \ln n$.

Comparing Lemmas 1 and 2 with 3 and 4, we conclude that all those remarks apply also to algorithm (7). One can verify that they hold for (8), too.

VIII. Simulation example

In the simulation example, the scalar dynamic subsystem is described by the equation $\dot{X} = aX + m(U)$ with $a = 0.8$. Moreover, $m(u) = (1/4 + u^2)\text{sign}(u)$. The system is driven by Gaussian noise with variance 1. Since we set $\lambda = 0$, $\mu(u) = m(u)$. Outer disturbance $Z(t)$ is zero and, therefore, $Y(t) = m(U(t)) + \xi(t)$ with $\xi(t) = Y(t) - m(U(t))$ being noise incurred by the dynamic subsystem.

Algorithm (3) with the rectangular kernel, i.e., a kernel equal 1 or zero according to whether $|u|$ does not exceed or is greater than 1, and $h_n = h$ is examined. MISE is defined as $\int_{-3}^{3}(\hat{\mu}(u) - m(u))^2du$ and examined for $n = 10$, 40, 160, 640, and 2560. In Fig. 2, $t_n = n$, i.e., $t_{n+1} - t_n = 1$, i.e., sampling is uniform. In Fig. 3, $t_n = n^{4/5}$, the distance $t_{n+1} - t_n \sim n^{-1/5}$ between sampling instants decreases. The error is greater, which is caused by the fact that covariance between two output observations is greater when the time distance between them is smaller.

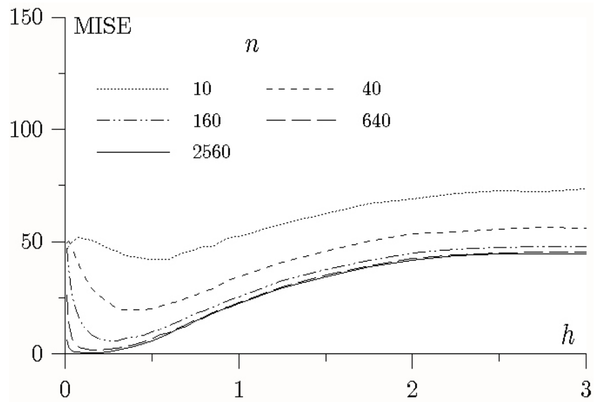

Fig. 2. MISE versus h; $t_n = n$.

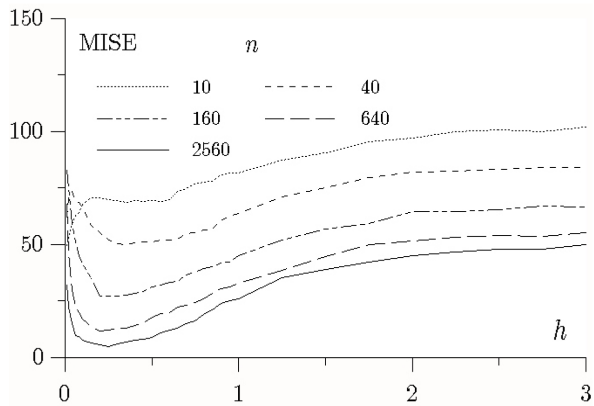

Fig. 3. MISE versus h; $t_n = n^{4/5}$.

IX. Final remarks

Our algorithms converge for both uniform and nonuniform sampling. In the other case, the distance between consecutive sampling instants must not decrease too quickly. For $t_n = \ln n$, e.g., the algorithms may not converge. Moreover, compared to uniform sampling, nonuniform can worsen convergence rate. This is the case if the distance between consecutive sampling instants decreases too quickly.

Appendix

The next lemma is an immediate consequence of Theorem 9.9 in [11].

Lemma 6: Let f be a density function of a random variable U. Let a Borel measurable kernel K be such that $\int |K(v)|dv < \infty$ and satisfy the second and third assumptions in (4). If, moreover, m is a Borel function such that

$E|m(U)| < \infty$, then

$$\frac{1}{h}E\left\{m(U)K\left(\frac{u-U}{h}\right)\right\} \to m(u)f(u)\int K(v)dv, \quad (20)$$

$$\frac{1}{h}\operatorname{var}\left[m(U)K\left(\frac{u-U}{h}\right)\right] \to m^2(u)f(u)\int K^2(v)dv. \quad (21)$$

as $h \to 0$, at every point u at which both f and m are continuous. If, moreover, $Em(U) = 0$, then

$$\frac{1}{h}\operatorname{cov}\left[m(U)K\left(\frac{u-U}{h_n}\right), m(U)\right] \to m^2(u)f(u)\int K(v)dv \quad (22)$$

at the same points.

Lemma 7: Let $U(t)$ and $U(\tau)$ be samples from the process $\{U(.)\}$. For any Borel functions φ and ψ,

$$\operatorname{cov}\left[W(\lambda+t)\varphi(U(t)), W(\lambda+\tau)\psi(U(\tau))\right] = \begin{cases} k^2(\lambda)\operatorname{cov}\left[m(U)\varphi(U), m(U)\psi(U)\right], & \text{for } t=\tau \\ k(\lambda+t-\tau)k(\lambda)E\{\varphi(U)\} \times \operatorname{cov}\left[m(U)\varphi(U), m(U)\right] + k(\lambda+\tau-t)k(\lambda)E\{\psi(U)\} \times \operatorname{cov}\left[m(U)\psi(U), m(U)\right], & \text{for } t \neq \tau. \end{cases}$$

From Lemmas 6 and 7, we obtain the following one:

Lemma 8: Let $U(t)$ and $U(\tau)$ be samples from the process $\{U(.)\}$. Let the kernel K satisfy (4). Let $h_n \to 0$ as $n \to \infty$. Then, at every point at which both m and f are continuous,

$$\operatorname{cov}\left[W(\lambda+t)K\left(\frac{u-U(t)}{h(t)}\right), W(\lambda+\tau)K\left(\frac{u-U(\tau)}{h(\tau)}\right)\right] = \begin{cases} h(t)k^2(\lambda)m^2(u)f(u)\int K^2(v)dv + h(t)o_t(1), & \text{for } t=\tau \\ h(t)h(\tau)k(\lambda)\left(k(\lambda+|t-\tau|)+k(\lambda-|t-\tau|)\right) \times \left(m^2(u)f^2(u)+o_t(1)+o_\tau(1)\right), & \text{for } t\neq\tau \end{cases}$$

with $o_t(1)$ and $o_\tau(1)$ independent of k, t, and τ.

Lemma 9: If (1) holds, then

$$\varphi_n(\lambda) = O\left(\frac{n}{\Delta_n}\right).$$

Proof: From (1) and $|i-j|\Delta_n \le |t_i - t_j|$, we get $|k(\lambda + |t_i - t_j|)| \le \alpha e^{-\beta\lambda}e^{-\beta|t_i-t_j|} \le \alpha e^{-\beta\lambda}e^{-|i-j|\beta\Delta_n}$. Hence,

$$\left|\sum_{i=1}^{n}\sum_{j=1}^{n}k(\lambda+|t_i-t_j|)\right| \le \alpha e^{-\beta\lambda}\sum_{i=1}^{n}\sum_{j=1}^{n}e^{-|i-j|\beta\Delta_n} = \alpha e^{-\beta\lambda}\sum_{i=1}^{n}(n-i)e^{-\beta i\Delta_n}.$$

Since, for $|q| < 1$, $\sum_{i=1}^{n}(n-i)q^i \le nq/(1-q)$, we obtain $\sum_{i=1}^{n}(n-i)e^{-\beta i\Delta_n} \le n\gamma(\Delta_n)$, where $\gamma(\Delta) = e^{-\beta\Delta}/(1-e^{-\beta\Delta})$. As $0 < \gamma(\Delta) \le 1/\beta\Delta$, we find that the quantity is of order $O(n/\Delta_n)$. Moreover, because $|k(t)| \le \alpha$,

$$\left|\sum_{i=1}^{n}\sum_{j=1}^{n}k(\lambda-|t_i-t_j|)\right| \le \sum_{i=1}^{n}\sum_{j=1}^{n}k(\lambda-|t_i-t_j|)I_{\{j;|t_i-t_j|\le\lambda\}}(j) \le \alpha\sum_{i=1}^{n}\sum_{j=1}^{n}I_{\{j;|t_i-t_j|\le\lambda\}}(j),$$

where $I_A(j) = 1$ for $j \in A$ and $I_A(j) = 0$ otherwise. As

$$\sum_{j=1}^{n}I_{\{j;|t_i-t_j|\le\lambda\}}(j) \le \text{the number of } t_j\text{'s in the interval } [t_i-\lambda, t_i+\lambda] \le \frac{2\lambda}{\Delta_n}+1,$$

the quantity is also of order $O(n/\Delta_n)$ which completes the proof. ■

References

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

[2] S. A. Billings, "Identification of nonlinear systems - a survey," *IEE Proceedings*, vol. 127, no. 6, Part D, pp. 272–285, November 1980.

[3] G. B. Giannakis and E. Serpendin, "A bibliography on nonlinear system identification," *Signal Processing*, vol. 81, pp. 533–580, 2001.

[4] W. Greblicki and M. Pawlak, "Identification of discrete Hammerstein systems using kernel regression estimates," *IEEE Transactions on Automatic Control*, vol. AC-31, no. 1, pp. 74–77, January 1986.

[5] W. Greblicki and M. Pawlak, "Recursive identification of Hammerstein systems," *Journal of the Franklin Institute*, vol. 326, no. 4, pp. 461–481, 1989.

[6] A. Krzyżak, "On estimation of a class of nonlinear systems by the kernel regression estimate," *IEEE Transactions on Information Theory*, vol. 36, no. 1, pp. 141–152, February 1990.

[7] A. Krzyżak, "Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation," *IEEE Transactions on Information Theory*, vol. 38, no. 4, pp. 1323–1338, July 1992.

[8] A. Krzyżak and M. A. Partyka, "On identification of block orientated systems by non-parametric techniques," *International Journal of Systems Science*, vol. 24, no. 6, pp. 1049–1066, 1993.

[9] W. Greblicki, "Continuous-time Hammerstein system identification," *IEEE Transactions on Automatic Control*, vol. 45, no. 6, pp. 1232–1236, June 2000.

[10] W. Greblicki, "Recursive identification of continuous-time Hammerstein systems," *International Journal of Systems Science*, vol. 33, no. 12, pp. 969–977, 2002.

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