Identification of Discrete Hammerstein Systems Using Kernel Regression Estimate

WŁODZIMIERZ GREBLICKI AND MIROSŁAW PAWLAK

Abstract— **In this note a discrete-time Hammerstein system is identified. The weighting function of the dynamical subsystem is recovered by the correlation method. The main results concern estimation of the nonlinear memoryless subsystem. No conditions concerning functional form of the transform characteristic of the subsystem are made and an algorithm for estimation of the characteristic is presented. The algorithm is a nonparametric kernel estimate of regression functions calculated from dependent data. It is shown that the algorithm converges to the characteristic as the number of observations tends to infinity. For sufficiently smooth characteristics, the rate of convergenve is $O(n^{-2/5})$ in probability.**

I. Introduction

Various approaches which have been proposed for identification of nonlinear systems largely depend on the prior knowledge about the system, i.e., about its mathematical representation. The Volterra series and Wiener expansion [17] are quite general nonlinear representations.

Another approach is based on the assumption that the structure of the system is known. Some authors consider cascade systems described by the Hammerstein functional

$$y(t) = \int_{-\infty}^{t} k(t-\tau)\varphi(u(\tau))d\tau.$$

Such a system consists of a nonlinear memoryless element with a transform characteristic φ followed by a linear dynamical subsystem having the weighting function k. The idea of using the Hammerstein representation stems originally from Narendra and Gallman [14] and was developed by Chang and Luus [3], Thatachar and Ramaswamy [18], Haist, Chang and Luus [10], Gallman [6], [7], as well as Billings and Fakhouri [2]. All these authors presented algorithms for identification of both the subsystems, but without rigorous convergence proofs, and all of them imposed on the nonlinear element a very restrictive condition that φ is not only continuous, but also has a polynomial form, i.e.,

$$\varphi(u) = c_1u + c_2u^2 + \cdots + c_Mu^M,$$

where $c_1, \cdots, c_M$ are unknown parameters but M is a fixed and known constant. They estimate the weighting function or coefficients of the transfer functions of the linear subsystem and coefficients $c_1, \cdots, c_M$ describing the nonlinear subsystem. Moreover, if φ is not a polynomial, their algorithms do not converge to $\varphi(u)$, [6]. Algorithms for identification of the nonlinear and linear subsystems are not, however, mutually independent and the subsystems cannot be identified separately.

In this note a discrete-time system shown in Fig. 1 is identified. The first subsystem is memoryless and nonlinear and has a transform characteristic φ. Its output is disturbed by a random noise ξ_n. The second subsystem is linear and has an impulse response $\{k_n\}$ and its output depends on additive random noise η_n. We assume throughout this note that ξ_n's and η_n's are independent, identically distributed, zero mean random variables with finite variance and are independent of U_i's and W_i's, $i \le n$, respectively. The signal W_n interconnecting the subsystems is not accessible for measurement and φ as well as $\{k_n\}$ are estimated from input-output observations of the whole system. For this reason both φ and $\{k_n\}$ can be estimated generally only up to some constant multiplicative factors. Assuming that the whole system is driven by white noise, we estimate the weighting function of the linear subsystem by the standard correlation method. The main result of this note concerns, however, identification of the nonlinear subsystem.

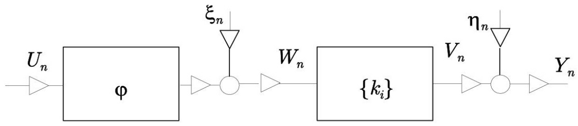

Fig. 1. The identified Hammerstein system.

Contrary to authors mentioned above, we do not impose any restriction on continuity of φ or its functional form. In consequence, the estimation problem we face in this way is nonparametric in its nature. A new algorithm for recovering the transform characteristic of the nonlinear subsystem is presented, namely

$$\hat{\psi}(u) = \frac{\sum_{i=0}^{n-1} Y_{i+1} K\left(\frac{u-U_i}{h(n)}\right)}{\sum_{i=0}^{n-1} K\left(\frac{u-U_i}{h(n)}\right)} \tag{1}$$

where K is a suitably selected kernel and $\{h(n)\}$ is a sequence of positive numbers. In this definition 0/0 is understood as zero. From a computational viewpoint the algorithm is extremely simple essentially as compared to those proposed by the authors cited above. It is shown that

$$\hat{\psi}(u) \to \alpha\varphi(u) \text{ as } n \to \infty$$

in probability, where α is unknown constant factor up to which φ can be estimated, at all continuity points of φ. For φ having two derivatives and $h(n)$ selected as $n^{-1/5}$, the convergence rate equals

$$|\hat{\psi}(u) - \varphi(u)| = O(n^{-2/5})$$

in probability.

II. Identification Problem

For the first subsystem in Fig. 1 with input U_n and output W_n,

$$W_n = \varphi(U_n) + \xi_n, \tag{2}$$

$n = 0, \pm 1, \cdots$, where φ is a Borel measurable function, U_n is an input random variable, whereas ξ_n is a random disturbance. $\{U_n\}$ is a sequence of independent and identically distributed random variables having a density f. Additionally, let U_n's be distributed symmetrically, i.e., $f(u) = f(-u)$ and let variance of U_n be finite. The unknown function φ is odd, i.e., $\varphi(u) = -\varphi(-u)$, and moreover

$$|\varphi(u)| \le c_1 + c_2|u|, \tag{3}$$

where c_1, c_2 are positive constants.

It is clear that the class of all Borel functions satisfying the above restriction cannot be parameterized, and thus, the problem of estimating φ is nonparametric. Let us, moreover, observe that the above conditions imply that $\{W_n\}$ is a sequence of independent and identically distributed random variables with $EW_n = 0$ and $EW_n^2 < \infty$.

The second subsystem of an unknown order is asymptotically stable and

$$Y_n = \sum_{j=0}^{\infty} k_j W_{n-j} + \eta_n, \tag{4}$$

The authors are with the Institute of Engineering Cybernetics, Technical University of Wrocław, Poland

$n = 0, \pm 1, \cdots$, where $\{k_n; n = 0, 1, \cdots\}$ is the unknown impulse response function, and where k_0 is assumed to be zero and $k_1 \neq 0$. Since the dynamic subsystem is asymptotically stable, i.e., $\sum_{j=0}^{\infty} |k_j| < \infty$, then clearly $\sum_{j=0}^{\infty} k_j^2 < \infty$. By this and $EW_n^2 < \infty$ we observe that Y_n is a random variable [1, p. 377] and moreover $\{Y_n\}$ is strictly stationary random process (see [15, p. 143]).

The problem is to recover both φ and $\{k_n\}$ from observations (U_i, Y_i), $i = 1, 2, \cdots$, of input and output of the whole system. Since W_n's are not measured, we can estimate both φ and $\{k_n\}$ only within to some unknown factors, i.e., we can estimate $\alpha\varphi(u)$ and βk_n, $n = 1, 2, \cdots$, where α and β are unknown constants.

The problem of recovering the impulse response of the dynamical subsystem is not difficult since $\{U_n\}$ is stationary white noise. For $m \geq 1$, clearly

$$E\{Y_{n+m}U_n\} = \beta k_m, \tag{5}$$

where $\beta = E\{U\varphi(U)\}$. This suggests the following estimate of βk_m:

$$\hat{\rho}_m = n^{-1} \sum_{i=1}^{n-m} U_i Y_{i+m}$$

for $m = 1, 2, \cdots, n-1$.

Since $\{Y_n\}$ is also an ergodic random process (see [5, p. 461]) it is subject to laws of large numbers [5, p. 465] and

$$\hat{\rho}_m \to \beta k_m \text{ as } n \to \infty$$

in probability.

Remark 1: We note that if the function φ is even, then $\beta = 0$. For this case use instead (5) the fact that $E\{Y_{n+m}G(U_n)\} = \beta k_m$, $\beta = R\{G(U)\varphi(U)\}$, where G is any measurable function. If φ is even we can choose G as, e.g., $|u|$, u^2 and then clearly $\beta \neq 0$.

III. Nonlinear Subsystem Identification

In order to present motivation for the proposed estimate, we begin with the ARMA model of the dynamical subsystem:

$$V_n + a_1 V_{n-1} + \cdots + a_s V_{n-s} = b_1 W_{n-1} + \cdots + b_s W_{n-s}, \tag{6}$$

$$Y_n = V_n + \eta_n, \tag{7}$$

where s is the unknown order of the system and V_n is the noise-free output. By the fact that $EW_n = 0$ and (6), (7), it follows that $EY_n = 0$. It is now clear that $E\{Y_n | W_{n-1}\} = \alpha W_{n-1}$ and by stationarity

$$E\{Y_n | U_{n-1} = u\} = E\{Y_1 | U_0 = u\} = \alpha\varphi(u)$$

where $\alpha = b_1 - a_1$ is assumed to be not equal to zero. Recovering $\varphi(u)$ is thus equivalent to estimating the nonlinear regression $E\{Y_1 | U_0 = u\}$ from dependant pairs $(U_0, Y_1), (U_1, Y_2), \cdots$. In order to estimate it we apply (1) referred to as the nonparametric kernel regression estimate. This estimate has been introduced independently by Watson [19] and Nadaraya [13] and than has been studied by a number of authors (see, e.g., Rosenblatt [16], Greblicki and Krzyzak [8], and more recent works of Devroye [4], Krzyzak and Pawlak [12] as well as Greblicki, Krzyzak and Pawlak [9]). All these authors assumed that pairs (U_n, Y_{n+1})'s are independent, whereas this note deals with dependent observations.

On the bounded nonnegative Borel kernel we impose the following restrictions:

$$\int K(u)du < \infty, \tag{8}$$

$$|u|K(u) \to 0 \text{ as } |u| \to \infty. \tag{9}$$

Therefore, one can choose the following kernels:

(a) the window kernel

$$K(u) = \begin{cases} 1/2, & \text{for } |u| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

(b) the quadratic kernel

$$K(u) = \begin{cases} 3(1-u^2)/4, & \text{for } |u| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

(c) the Gauss kernel $(2\pi)^{-1/2}e^{-u^2/2}$.

In the analysis of the estimate it will be convenient to use the state-space representation of the dynamical subsystem:

$$\begin{aligned} X_{n+1} &= AX_n + bW_n \\ Y_n &= c^T X_n + \eta_n, \end{aligned} \tag{10}$$

where X_n is the s-dimensional random vector, A is unknown but asymptotically stable matrix, b and c are unknown s-vectors. A, b, and c can be easily calculated from (6) and (7).

We observe that $\{X_n\}$ is also strictly stationary random process [15, p. 797]. By this, (2) and (10) we have $E\{Y_1 | U_0 = u\} = \alpha\varphi(u)$, where $\alpha = c^T b$.

IV. Consistency of the Estimate

In this section conditions under which estimate (1) converges to $\alpha\varphi(u)$ are given. For convenience, we denote $\alpha\varphi(u) = m(u)$, $g(u) = m(u)f(u)$, where f is the input density and we shall write h for $h(n)$. In the sequel, $C(f_1, \cdots, f_m)$ denotes a set of all points at which all $f_1, \cdots, f_m$ are continuous.

Theorem 1: Let the kernel K satisfy (8) and (9). Let

$$h(n) \to 0 \text{ as } n \to \infty \tag{11}$$

and

$$nh(n) \to \infty \text{ as } n \to \infty. \tag{12}$$

Then

$$\hat{\psi}(u) \to \alpha\varphi(u) \text{ as } n \to \infty \text{ in probability} \tag{13}$$

at all points $u \in C(f, \varphi)$ at which $f(u) > 0$.

Remark 2: If it is known that $\varphi(u_0) = \varphi_0$ for some point $u_0 \in C(\varphi)$ then, by (13), $\hat{\varphi}(u) = \varphi_0\hat{\psi}(u)/\hat{\psi}(u_0)$ is a consistent estimate of $\varphi(u)$.

Next $n^{-1}\sum_{i=0}^{n-1} U_i\hat{\varphi}(U_i)$ may be used as an estimate of β in (5).

While proving Theorem 1 we shall need a lemma whose proof is in the Appendix.

Lemma: For $i \neq j$,

$$\text{cov}\left[X_{i+1}K\left(\frac{u-U_i}{h}\right), X_{j+1}K\left(\frac{u-U_j}{h}\right)\right] = A^{|i-j|}Q_h(u),$$

where

$$Q_h(u) = APA^T E^2\left\{K\left(\frac{u-U_0}{h}\right)\right\} + bb^T E\left\{\varphi^2(U_0)K\left(\frac{u-U_0}{h}\right)\right\} E\left\{K\left(\frac{u-U_0}{h}\right)\right\}$$

and where

$$P = E\left\{X_0 X_0^T\right\}.$$

Proof of Theorem 1: We begin with rewriting the estimate in the following form:

$$\hat{\psi}(u) = \hat{g}(u)/\hat{f}(u),$$

where

$$\hat{g}(u) = (nh)^{-1}\sum_{i=0}^{n-1} Y_{i+1}K_i$$

and

$$\hat{f}(u) = (nh)^{-1}\sum_{i=0}^{n-1} K_i.$$

Her and in further parts of this note K_i stands for $K((u-U_i)/h)$. Clearly,

$$E\hat{g}(u) = h^{-1}\int K\left(\frac{u-v}{h}\right)m(v)f(v)dv.$$

By this and [20, Theorem 9.9m p. 150],

$$E\hat{g}(u) \to g(u)\int K(v)dv \text{ as } h \to 0 \tag{14}$$

for $u \in C(g)$.

In turn, by virtue of the lemma,

$$\begin{aligned}
&\operatorname{var}\sum_{i=0}^{n-1} Y_{i+1}K_i \\
&= n\operatorname{var}(Y_1K_0) + \sum_{\substack{i=0\\ i\neq j}}^{n-1}\sum_{j=0}^{n-1}\operatorname{cov}(Y_{i+1}K_i, Y_{j+1}K_j) \\
&= n\operatorname{var}(Y_1K_0) + \sum_{\substack{i=0\\ i\neq j}}^{n-1}\sum_{j=0}^{n-1} c^TA^{|i-j|}cQ_h(u) \\
&= n\operatorname{var}(Y_1K_0) + 2c^T[A^n(I-A)^{-1} - nI]cQ_h(u) \\
&= n\operatorname{var}(Y_1K_0) + H_n(u), \text{ say.}
\end{aligned}$$

By the definition of $Q_h(u)$ and (14)

$$Q_h(u)/h^2 \to APA^Tf^2(u) + bb^T\varphi^2(u)f^2(u)$$

as $h \to 0$ for $u \in C(f,\varphi)$. By this and the fact that $A^n \to 0$ as $n \to \infty$,

$$H_n(u)/(nh(n))^2 \to 0 \text{ as } n \to \infty$$

for $u \in C(f,\varphi)$, provided that (11) and (12) hold. Therefore,

$$\lim_{n\to\infty} nh\operatorname{var}\hat{g}(u) = \lim_{n\to\infty} h^{-1}\operatorname{var}(Y_1K_0).$$

Obviously,

$$h^{-1}\operatorname{var}(Y_1K_0) \le kh^{-1}E\{\phi(U_0)K_0\},$$

where $k = \sup K(u)$ and $\phi(U_0) = E\{Y_1^2|U_0\}$. Using (6) and (7) one can easily get

$$\phi(u) = c_3\varphi^2(u) + c_4\varphi(u) + c_5.$$

By this and (14)

$$\limsup_{n\to\infty} h^{-1}\operatorname{var}(Y_1K_0) \text{ is finite}$$

for $u \in C(f,\varphi)$. By this and (14)

$$\hat{g}(u) \to g(u)\int K(v)dv \text{ as } n \to \infty$$

in probability for $u \in C(f,\varphi)$.

Since using similar arguments one can easily show that

$$\hat{f}(u) \to f(u)\int K(v)dv \text{ as } n \to \infty$$

in probability for $u \in C(f,\varphi)$, the theorem has been proved. ■

In [20, p. 152] we find that, under additional assumptions on K, (14) holds also for almost all points $u \in R$, R is the real line. Thus, (13) holds at almost all $u \in R$ for which $f(u) > 0$. Proceeding as in Greblicki, Krzyzak and Pawlak [9] and using the lemma one can verify that (13) takes place for almost all (μ) $u \in R$, where μ is the probability measure of U_n's. The measure has not to be absolutely continuous and, consequently, cannot posses the density. All the kernels given in (a)–(c) satisfy their additional conditions on K.

V. The Rate of the Convergence

Imposing some regularity conditions on the transform characteristic of the nonlinear subsystem and on the input density f, we show how to select $\{h(n)\}$ and give the rate of convergence.

Writing $a_n \sim b_n$ we mean that a_n/b_n has a nonzero limit as $n \to \infty$. For a sequence $\{\theta_n\}$ of random variables, $\theta_n = O(a_n)$ in probability says that $\lambda_n\theta_n/a_n \to 0$ as $n \to \infty$ in probability for any number sequence $\{\lambda_n\}$ convergent to zero..

Theorem 2: Let both φ and f have bounded derivatives up to the second order in some neighborhood of $u \in R$. Let the bounded Borel kernel satisfy (8) and (9) and let

$$\int vK(v)dv = 0, \tag{16}$$

$$\int v^2K(v)dv < \infty. \tag{17}$$

If

$$h(n) \sim n^{-1/5}, \tag{18}$$

then

$$|\hat{\psi}(u) - \alpha\varphi(u)| = O(n^{-2/5})$$

in probability.

Remark 3: Estimation accuracy is very sensitive to selection of parameter h. In general, the optimal choice of h depends on the unknown input density, the nonlinear transformation φ, as well the transfer function of the dynamical subsystem. For large n, the optimal h becomes, however, independent of the last factor. For independent observations Hall [11] showed that the asymptotically optimal h is of order $n^{-1/5}$. In the light of this, it seems that h given in (18) is also asymptotically optimal. At last we would like to mention that the choice of the kernel is much less critical

Proof of Theorem 2: Without any loss of generality we assume that $\int K(v)dv = 1$. Let $\hat{g}(u)$ and $\hat{f}(u)$ have the same meaning as in the proof of Theorem 1. Obviously

$$E\hat{g}(u) = \alpha\int K(v)\varphi(u-hv)f(u-hv)dv.$$

Expanding both φ and f in the Taylor series and using (16), (17), we get

$$|E\hat{g}(u) - \alpha\varphi(u)f(u)| = O(h^2).$$

From this, (16) and (18), it follows that

$$E(\hat{g}(u) - \alpha\varphi(u)f(u))^2 = O(n^{-4/5}).$$

i.e., $|\hat{g}(u) - \alpha\varphi(u)f(u)| = O(n^{-2/5})$ in probability.

In the same way one can verify that

$$|\hat{f}(u) - f(u)| = O(n^{-2/5})$$

in probability.

Observing

$$\left|\frac{\hat{g}(u)}{\hat{f}(u)} - \alpha\varphi(u)\right| \leq \left|\frac{\hat{g}(u) - \alpha\varphi(u)f(u)}{\hat{f}(u)}\right| + \left|\frac{\hat{g}(u)(\hat{f}(u) - f(u))}{\hat{f}(u)f(u)}\right|$$

we complete the proof. ■

VI. Final Remarks

We have assumed that the transfer characteristic of the nonlinear subsystem is odd and the input distribution is symmetrical. One can easily verify that if these conditions are not satisfied,

$$E\{Y_1|U_0 = u\} = \alpha\varphi(u) + \delta,$$

where α and δ are some fixed but unknown constants.

Now δ can be estimated if it is known that, e.g., $\varphi(u_1) = \varphi_1$, where $u_1 \in C(\varphi)$. Then one can take $\hat{\psi}(u) - \hat{\psi}(u_1) + \alpha\varphi(u_1)$ as an estimate of $\alpha\varphi(u)$. Consequently, by Remark 2, if we know the values of φ in two points, what is often the case, we can estimate the constants α and δ and finally $\varphi(u)$. We observe that condition (3) may be relaxed to $|\varphi(u)| \leq c_1 + c_2|u|^\tau, \tau > 0$. Then in order to have $EW_n^2 < \infty$ (what is necessary on existence of the random variable Y_n) we assume $E|U|^{2\tau} < \infty$. Moreover, if the density of U is known we may assumed the most general $\int \varphi^2(u)f(u)du < \infty$.

Appendix

Proof of the Lemma

For simplicity of notation, let φ_i denote $\varphi(U_i)$. Since $EX_0 = 0$,

$$\begin{aligned}
&E\{X_{i+1}X_1^T K_i K_0\} \\
&= E\{[A^{i+1}X_0 + (A^i b\varphi_0 + \cdots + b\varphi_i)](X_0^T A^T + b^T\varphi_0)K_i K_0\} \\
&= A^{i+1}PA^T EK_0^2 + E\{(A^i b\varphi_0 + \cdots + b\varphi_i)b^T\varphi_0 K_i K_0\} \\
&= A^{i+1}PA^T EK_0^2 + A^i bb^T E\{\varphi_0^2 K_0\}EK_0 + bb^T E^2\{\varphi_0 K_0\}.
\end{aligned}$$

We used here the following equality:

$$(A^{i-2}b + \cdots + Ab + b)E\varphi_0 = (I - A^{i-1})EX_0.$$

Now one can easily get the desired assertion.

Acknowledgment

The authors wish to express their thanks to the referees for suggestions concerning some improvements.

References

[1] T.W. Anderson, *The Statistical Analysis of Time Series*, New York: Wiley, 1971.

[2] S.A. Billings and S.Y. Fakhouri, "Non-linear system identification using the Hammerstein model", *Int. J. Syst. Sci.*, vol. 10, pp. 567-578, no. 5, 1979.

[3] F.H.I. Chang and R. Luus, "A noniterative method for identification using the Hammerstein model", *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 464-468, 1971.

[4] L. Devroye, "On the almost everywhere convergence of nonparametric regression function estimate", *Ann. Statist.*, vol. 9, pp. 1310-1319, 1981.

[5] T.J. Doob, *Stochastic Processes*, New York: Wiley, 1953.

[6] P.G. Gallman, "An iterative method for the identification of nonlinear systems using a Uryson model", *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 71-775, 1975.

[7] P.G. Gallman, "A comparison of two Hammerstein model identification algorithms", *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 124-126, 1976.

[8] W. Greblicki and A. Krzyżak, "Asymptotic properties of kernel estimates of a regression function", *J. Statist. Plann. Inference*, vol. 4, pp. 81-90, 980.

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

[10] N.D. Haist, F.H.I. Chang and R. Luus, "Nonlinear identification in the presence of correlated noise using a Hammerstein model", *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 552-555, 1973.

[11] P. Hall, "Integrated square error properties of kernel estimators of regression functions", *Ann. Statist.*, vol. 12, pp. 241-260, 1984.

[12] A. Krzyzak and M. Pawlak, "Distribution free consistency of a nonparametric kernel regression estimate and classification", *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 78-81, 1984.

[13] E.A. Nadaraya, "On estimating regression", *Theor. Probability Appl.*, vol. 9, pp. 141-142, 1964.

[14] K.S. Narendra and P.G. Gallman, "An iterative method for the identification of nonlinear systems using the Hammerstein model", *IEEE Trans. Automat. Contr.*, vol. AC-11, pp. 546-550, 1966.

[15] M. B. Priestley, *Spectral Analysis and Time Series.* New York: Wiley, 1981.

[16] M. Rosenblatt, "Curve estimation", *Ann. Math. Statist.*, vol. 42, pp. 1815-1842, 1971.

[17] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems.* New York: Wiley, 1980.

[18] M.A.L. Thatachar and S. Ramaswamy, "Identification of a class of nonlinear systems", *Int. J. Contr.*, vol. 18, pp. 741-752, 173.

[19] G. S. Watson, "Smooth regression analysis,", *Shankhyā*, Series A, vol. 26, pp. 355-372, 164.

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