Hammerstein system identification by non-parametric regression estimation

WŁODZIMIERZ GREBLICKI[†] and MIROSŁAW PAWLAK[‡]

A discrete-time, multiple-input non-linear Hammerstein system is identified. The dynamical subsystem is recovered using the standard correlation method, The main results concern estimation of the nonlinear memoryless subsystem. No conditions concerning the functional form of the transform characteristic of the subsystem are made and an algorithm for estimation of the characteristic is given. The algorithm is simply a non-parametric kernel estimate of the regression function calculated from the dependent data. It is shown that the algorithm converges to the characteristic of the subsystem in the pointwise as well as the global sense. For sufficiently smooth characteristics, the rate of convergence is $O(n^{-1/(2+d)})$ in probability, where d is the dimension of the input variable.

1. Introduction

In this paper we identify a discrete time Hammerstein system shown in Fig. 1. The system consists of two subsystems connected in a cascade. The first of them is multivariable non-linear and memoryless and its characteristic is denoted by m. Its output is disturbed by a random noise ξ_n. The second subsystem is linear and dynamic and has a weighting function $\{g_i\}$, $i = 0, 1, 2, \ldots$ The input of the whole system is a d dimensional stationary white noise whereas the output is disturbed by a zero mean stationary white noise η_n.

Examples of problems in which such systems are encountered are adaptive control (Kung and Womack 1984), noise cancellation models (Stapleton and Bass 1985) and image processing (Sawchuk and Strand 1982).

The problem of this paper is to identify both the subsystems from the input-output observation of the whole system.

The signal W_n interconnecting both the subsystems is inaccessible to measurements. Since the weighting function can be estimated by the standard correlation method, recovering the nonlinear characteristic appears to be the main problem of the paper. Such a problem has been investigated by a number of authors, as can be seen from the original work of Narendra and Gallman (1966) as well as later

[†]Institute of Engineering Cybernetics, Technical University of Wrocław, Wrocław, Poland.
[‡]Department of Electrical Engineering, The University of Manitoba, Winnipeg, R3T 2N2, Canada.

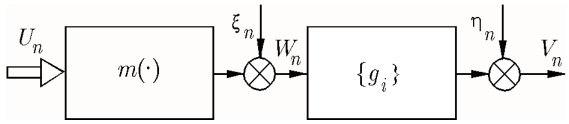

Figure 1: The identified Hammerstein system

papers of Chang and Luus (1971), Thatachar and Ramaswamy (1973), Haist *et al.* (1973), Kaminskas (1915), Gallman (1975, 1976), Hsia (1976), Billings and Fakhouri (1979) as well as Shih and Kung (1985). All of these authors present algorithms for identification of both the subsystems, but without rigorous convergence proof, and they have assumed that the non-linearity is of a polynomial form. It is clear that the class of characteristics of the non-linear subsystem considered by all these authors is not wide enough, for it does not even contain any discontinuous function. Moreover, they considered only the single-input systems and their identification algorithms of the non-linear and linear parts are computationally dependent, i.e., the parameters in one section are held constant while the parameters in the other one are determined.

Recently, observing that the non-linear characteristic can be expressed as a regression function, Greblicki and Pawlak (1986) proposed a new approach to the problem of recovering the non-linearity. Their idea, based on non-parametric estimation of the regression function, made it possible for them to introduce an algorithm which is considerably simpler than those known in previous literature. The algorithm is not only completely independent of procedures which can be used for identification of the linear subsystem, but recovers non-linear characteristics even if they are not continuous. In this paper we investigate the same procedure for multivariable inputs and additionally, we admit a broader class of dynamical subsystems, viz. asymptotically stable dynamical systems described by state equations. We examine not only pointwise, but also global properties of the algorithm and show that the integrated square error converges to zero as the number of observations increases to infinity. The rate of convergence is investigated; for the Lipschitz characteristics, it is of order $O(n^{-1/(2+d)})$ in probability, where d the dimension of the input variable. The results are valid for all possible density functions of the input random variable.

2. Problem statement and preliminaries

We assume that the input signal $\{U_n\}$, $n = \ldots -1, 0, 1, \ldots$ is d-dimensional stationary white noise and that

$$E\left\|U_n\right\|^2 < \infty \tag{1}$$

where $\|\cdot\|$ is any norm in $\mathbb{R}^d$. Throughout the paper we assume that the input signal has a probability density f. The characteristic of the first subsystem is a Borel function and satisfies the following condition:

$$|m(u)| \leq c_1 + c_2 \left\|u\right\| \tag{2}$$

c_1 and c_2 some constants. In turn, the linear dynamic subsystem is described by either the ARMA model:

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

or, more generally, by the following state equation:

$$\begin{aligned} X_{n+1} &= A_n X_n + bW_n \\ Y_n &= c^T X_n + W_n + \eta_n \end{aligned}$$

where $n = \ldots -1, 0, 1, \ldots$. In the paper the last description will be used.

X_n is a state vector of an unknown dimension, $\{\eta_n\}$ is a stationary white noise with zero mean and finite variance σ_η^2, whereas b and c are some vectors, A is a matrix assumed to be asymptotically stable. We assume, moreover, that $\{U_n\}$ and $\{\eta_n\}$ are mutually independent. Furthermore

$$W_n = m(U_n) + \xi_n \tag{4}$$

$n = \ldots -1, 0, 1 \ldots$, where $\{\xi_n\}$ is a stationary white noise with zero mean and finite variance σ_ξ^2. $\{U_n\}$ and $\{\xi_n\}$ as well as $\{\xi_n\}$ and $\{\eta_n\}$ are assumed mutually independent.

It is clear that $\sigma_\xi^2 < \infty$ and (1), (2) imply

$$EW_n^2 < \infty. \tag{5}$$

From (5), the asymptotical stability of A and Anderson (1971) it follows that both X_n and Y_n are random variables (in the mean square sense). This means that the restrictions (1) and (2) are totally independent of the identification method presented in the paper. They simply guarantee (5) to be satisfied and in this way make the problem well posed in the sense that the output of the system is a random variable. Let us note that both $\{X_n\}$ and $\{Y_n\}$, $n = \ldots -1, 0, 1 \ldots$, are strictly stationary stochastic processes; see also Anderson (1971).

The class of Borel functions satisfying (2) is so wide that it cannot be parameterized. Because of this, the problem of recovering the characteristic is nonparametric. We would like to emphasize that this is the main difference between the problem discussed here and in Greblicki and Pawlak (1986) and that considered by other authors. The case considered by the other authors is much simpler since they assumed that the characteristic is of a polynomial form of a known order. Therefore the problem investigated by them is parametric, for they estimate only a finite number of coefficients.

From (3) and (5) it follows that

$$Y_n = \sum_{i=0}^{\infty} g_i W_{n-i} + \eta_n \tag{6}$$

where $g_i = c^T A^{i-1} b$, $i = 1, 2 \ldots$, and $g_0 = 1$.

In turn from (4) and (6) it results that

$$E\{Y_n|U_n = u\} = m(u) + \gamma \tag{7}$$

where $\gamma = Em(U)\sum_{i=1}^{\infty} g_i$. We note that, by (1) and (2) and the asymptotic stability of A, $\gamma < \infty$.

In order to estimate $m(u)$ we first do the regression function in (7). For this we apply

$$\hat{\psi}_n(u) = \frac{\sum_{i=0}^{n-1} Y_i K\left(\frac{u - U_i}{h(n)}\right)}{\sum_{i=0}^{n-1} K\left(\frac{u - U_i}{h(n)}\right)} \tag{8}$$

where K is a Borel kernel function defined on $\mathbb{R}^d$ and $\{h(n)\}$ is a sequence of positive numbers. Rules for selecting K and $\{h(n)\}$ will be given later. $\{U_i Y_i\}$, $i = 0, 1 \ldots, n-1$, is a given sample of the input and output signals of the whole system.

We then make use of the only restriction of the paper strictly connected with the identification method presented by us, i.e., the assumption that

$$m(u = 0) = 0. \tag{9}$$

The restriction (3) seems to be not very restrictive and is satisfied in most applications. Owing to (9) we estimate $m(u)$ with $\hat{m}_n(u)$ which is defined in the following way

$$\hat{m}_n(u) = \hat{\psi}_n(u) - \hat{\psi}_n(u = 0) \tag{10}$$

Estimate (8) is known in the literature as non-parametric kernel regression estimate. This estimate was introduced independently by Watson (1964) and Nadaraya (1964) and since then has been studied by a number of authors, see e.g. Rosenblatt (1971), Collomb (1981), Devroye (1981) as well as Greblicki *et al.* (1984) and Greblicki and Pawlak (1986). All of these authors assumed, however, that pairs (U_i, Y_i) are mutually independent. The case considered by us is more complicated since $\{Y_n\}$, being the output of a dynamical system, is a sequence of dependent random variables.

Bierens (1983), Georgiev (1984) Collomb (1984) and Collomb and Härdle (1986) investigated the behaviour of the kernel regression estimate under dependence conditions. They assumed, however, that pairs (U_i, Y_i), $i = 1, 2, \ldots$, create a strictly stationary stochastic process which, additionally, satisfies some mixing conditions.

In this paper we show that $\hat{m}_n$ converges to m as n tends to infinity, in probability. We examine a pointwise well as global L_2 error. We give the rate of convergence, and furthermore, we examine the dynamical subsystem identification by the correlation method.

3. Linear subsystem identification

For identification of the linear subsystem, we shall employ the standard correlation approach.

In order to estimate $\{g_i\}$, the weighting sequence, we make use of the following equalities:

$$\mathrm{cov}[Y_n, \|U_n\|] = 0 \tag{11}$$

and

$$\mathrm{cov}[Y_{n+i}, \|U_n\|] = \alpha g_i,\ \ i = 1, 2, \ldots \tag{12}$$

where $\alpha = \mathrm{cov}[m(U_0), \|U_0\|]$. Let us observe that (1) and (2) assure that α exists.

Owing to (11) and (12), if only $\alpha \neq 0$ we can define

$$\hat{g}_{i,n} = \frac{\hat{\rho}_{i,n}}{\hat{\rho}_{0,n}},$$

where

$$\hat{\rho}_{i,n} = \frac{1}{n} \sum_{j=0}^{n-1-i} (Y_{j+1} - \bar{Y})(\|U_j\| - \overline{\|U\|})$$

and

$$\bar{Y} = \frac{1}{n} \sum_{j=1}^{n-1-i} Y_{j+i},\ \ \overline{\|U\|} = \frac{1}{n} \sum_{j=0}^{n-1} \|U_j\|\,,$$

as an estimate of g_i.

Recalling that $\{Y_n\}$ given by (6) is a strictly stationary and ergodic process, Hannan (1970, p. 210), one can easily show that

$$\hat{\rho}_{i,n} \overset{n}{\to} \alpha g_i \text{ in probability, } i = 1, 2, \ldots$$

and

$$\hat{\rho}_{0,n} \overset{n}{\to} \alpha \text{ in probability.}$$

Consequently we have

$$\hat{g}_{i,n} \overset{n}{\to} g_i \text{ in probability.}$$

In this way we have proved the following theorem:

Theorem 1
Let matrix A be asymptotically stable. Let (1) and (2) hold. If $\alpha \neq 0$, then

$$\hat{g}_{i,n} \overset{n}{\to} g_i \text{ in probability.}$$

4. Non-linear subsystem identification

In this section we shall show that estimate $\hat{m}(u)$ converges to $m(u)$ regardless of the density of U_0. The bounded and non-negative Borel kernel K satisfies the following two conditions:

$$\int K(u)du < \infty \tag{13}$$

and

$$K(u) = O(\|u\|^{-(d+\lambda)}) \text{ as } \|u\| \to \infty \tag{14}$$

for some $\lambda > 0$.

There exists a wide class of kernels satisfying (13) and (14), some of them listed below:

(a) the window kernel

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

(b) the triangular kernel

$$(u) = \begin{cases} 1 - \|u\|, & \text{for } \|u\| \le 1 \\ 0, & \text{otherwise,} \end{cases}$$

(c) the Gauss-Weierstrass kernel

$$K(u) = e^{-\|u\|^2},$$

(d) the Poisson kernel

$$K(u) = \frac{1}{1 + \|u\|^{2d}},$$

(e) the product kernel

$$K(u) = \prod_{j=1}^{d} K_j(u^{(j)}),$$

where K_j's are any kernels satisfying (13) and (14) for $d = 1$ and $u = (u^{(1)}, \ldots, u^{(d)})^T$.

Concerning the positive number of sequence, we assume that

$$h(n) \xrightarrow{n} 0 \tag{15}$$

and

$$nh^d(n) \xrightarrow{n} \infty. \tag{16}$$

If for example $h(n) = n^{-\tau}$, (15) and (16) hold for $0 < \tau < 1/d$. The following theorem present the conditions for the estimate consistency.

Theorem 2
Let matrix A be asymptotically stable and let (1) and (2) hold. Suppose that a bounded non-negative Borel kernel satisfies (13) and (14). If the non-negative number sequence $\{h(n)\}$ satisfies (15) and (16), then

$$\hat{m}_n(u) \xrightarrow{n} m(u) \text{ in probability}$$

for almost all (with respect to the Lebesgue measure in $\mathbb{R}^d$) $u \in \mathbb{R}^d$ at which $f(u) > 0$.

The proof of the theorem is deferred to the next section.

As a simple consequence of Theorem 2 and Lebesgue's dominated convergence theorem on product spaces (see for example Glick 1974), we obtain a result with global estimation error.

Theorem 3
Let the density of U_0 have a compact support. Under the conditions of Theorem 3,

$$E\int(\hat{m}_n(u)-m(u))^2 f(u)du \overset{n}{\to} 0.$$

It is worth mentioning that (5) is also met if we assume that

$$\sup_u |m(u)| < \infty. \tag{17}$$

In this case equation (5) is satisfied for all input densities. Thus, replacing (1) and (2) by (17) we obtain a density-free version of Theorem 3, i.e. a version true for all densities of the input random variable.

5. Proof of Theorem 2

In order to prove Theorem 2, we need a few lemmas. For convenience, by K_i, we denote $K((u-U_i)/h(n))$, $i=0,1,\ldots$.

Lemma 1
Let the bounded non-negative kernel satisfy (13) and (14). Then

$$\lim_{n\to\infty}\frac{1}{h^d}E\left\{Y_0 K\left(\frac{u-U_0}{h}\right)\right\} = (m(u)+\gamma)f(u)\int K(v)dv$$

for almost all $u\in\mathbb{R}^d$.

The lemma is a consequence of Theorem 9.13 in Wheeden and Zygmund (1977).

Lemma 2
Let K be a Borel integrable function. Then, for $0\le i\le j$,

$$\mathrm{cov}[X_iK_i, X_jK_j] = Q_{ij}(u)EK_0,$$

where

$$\begin{aligned} Q_{ij}(u) &= EK_0A^jP(A^i)^T + \mathrm{cov}[K_0, m(U_0)]A^i\bar{X}_0(A^{j-1-i}b)^T \\ &\quad + \mathrm{var}(m(U_0))EK_0\sum_{r=1}^{i}A^{i-r}b(A^{i-r}b)^T \\ &\quad + \sigma_Z^2 EK_0\sum_{r=1}^{i}A^{i-r}b(A^{j-r}b)^T \\ &\quad + Em(U_0)\mathrm{cov}[K_0, m(U_0)]A^{j-i-1}b\sum_{r=1}^{i-1}(A^rb)^T, \end{aligned}$$

with $P = \operatorname{cov}[X_0, X_0]$, $\bar{X} = EX_0$.

Proof.
Let $\theta_n = m(U_n) - Em(U_n)$. Using (3) and (4), we have

$$\begin{aligned} X_i &= A^i X_0 + \sum_{r=1}^{i-1} A^{i-1-r} b\theta_r + Em(U_0) \sum_{r=0}^{i-1} A^{i-1-r} b + \sum_{r=1}^{i-1} A^{i-1-r} \xi_r \\ &= G_{i1} + G_{i2} + G_{i3} + G_{i4}, \end{aligned} \tag{18}$$

say. Clearly

$$\operatorname{cov}[K_i G_{i4}, K_j G_{jr}] = 0 \text{ for } r = 1, 2, 3$$

and

$$\begin{aligned} \operatorname{cov}[K_i G_{i4}, K_j G_{j4}] &= \sum_{r=0}^{i-1} \sum_{s=0}^{j-1} A^{i-1-r} b (A^{j-1-s} b)^T \operatorname{cov}[K_i \xi_r, K_j \xi_s] \\ &= \sigma_\xi^2 E^2 K_0 \sum_{r=0}^{i-1} A^{i-1-r} b (A^{j-1-r} b)^T. \end{aligned}$$

Consider $\operatorname{cov}[K_i G_{i2}, K_j G_{j2}]$. Obviously,

$$\begin{aligned} \operatorname{cov}[K_i G_{i2}, K_j G_{j2}] &= \sum_{r=0}^{i-1} \sum_{s=0}^{j-1} A^{i-1-r} b (A^{j-1-s} b)^T \operatorname{cov}[K_i \theta_r, K_j \theta_s] \\ &= \sum_{r=0}^{i-1} \sum_{s=0}^{i-1} A^{i-1-r} b (A^{j-1-s} b)^T \operatorname{cov}[K_i \theta_r, K_j \theta_s] \\ &= \sum_{r=0}^{i-1} A^{i-1-r} b (A^{j-1-s} b)^T E^2 K_0 \operatorname{var}(m(U_0)). \end{aligned}$$

Since the remaining terms can be evaluated in the same way, the proof is completed. □

Lemma 3
Let K be a Borel function. Then, for $0 \le i \le j$,

$$\begin{aligned} \operatorname{cov}[Y_i K_i, Y_j K_j] &= c^T \operatorname{cov}[X_i K_i, X_j K_j] c \\ &\quad + g_{j-i}(EK_0 \operatorname{cov}[K_0 m(U_0), m(U_0)] + \sigma_\xi^2 E^2 K_0). \end{aligned}$$

Proof.
Using (1) and (4), we have

$$\begin{aligned} \operatorname{cov}[Y_i K_i, Y_j K_j] &= c^T \operatorname{cov}[X_i K_i, X_j K_j] c + c^T \operatorname{cov}[X_i K_i, W_j K_j] \\ &\quad + c^T \operatorname{cov}[W_i K_i, X_j K_j] + c^T \operatorname{cov}[W_i K_i, W_j K_j] \\ &= F_1 + F_2 + F_3 + F_4, \end{aligned}$$

say. Clearly, $F_2 = F_4 = 0$. In turn, by (18)

$$\begin{aligned} F_3 &= g_{j-i}\ \mathrm{cov}[W_i K_i, W_j K_j] \\ &= g_{j-i}(EK_0\ \mathrm{cov}[m(U_0), m(U_0)K_0] + \sigma_\xi^2 E^2 K_0). \end{aligned}$$

Thus, the lemma has been proved. □

Lemma 4
Let the conditions of Theorem 2 be satisfied. Then, for any i,

$$\begin{aligned} \mathrm{var}(Y_i K_i) &= \mathrm{var}(K_0 m(U_0)) + EK_0^2(\sigma_\xi^2 + \sigma_\eta^2) \\ &\quad + EK_0^2(\mathrm{var}(m(U_0)) + \sigma_\xi^2) \sum_{i=1}^{\infty} g_i^2 \\ &\quad + E^2 m(U_0) \mathrm{var}(K_0) \left(\sum_{i=1}^{\infty} g_i \right)^2 \\ &\quad + 2Em(U_0)\,\mathrm{cov}[m(U_0)K_0, K_0] \sum_{i=1}^{\infty} g_i. \end{aligned}$$

Proof.
Using (6) and the stationarity assumption, we get

$$\begin{aligned} \mathrm{var}(Y_i K_i) &= \mathrm{var}(Y_0 K_0) \\ &= \mathrm{var}\left(\left[m(U_0) + \xi_0 + \eta_0 + \sum_{i=1}^{\infty} g_i W_{-i} \right] K_0 \right). \end{aligned}$$

After some algebra it is easy to show that the above expression is equal to

$$\begin{aligned} &\mathrm{var}(K_0 m(U_0)) + EK_0^2(\sigma_\xi^2 + \sigma_\eta^2) + EK_0^2 EW_0^2 \sum_{i=1}^{\infty} g_i^2 \\ &+ EK_0^2 E^2 W_0 \sum_{j=1}^{\infty} \sum_{i=1, i\neq j}^{\infty} g_i g_j - E^2 K_0 E^2 W_0 \left(\sum_{i=1}^{\infty} g_i \right)^2 \\ &+ 2Em(U_0)\,\mathrm{cov}[m(U_0)K_0, K_0] \sum_{i=1}^{\infty} g_i. \end{aligned}$$

Because $EW_0 = Em(U_0)$ and $EW_0^2 = Em^2(U_0) + \sigma_\xi^2$, the lemma has been proved.
□

Proof of Theorem 2.
It is clear that it suffices to verify that

$$\hat{\psi}_n(u) \stackrel{n}{\to} m(u) + \gamma \text{ in probability}$$

for almost all $u \in \mathbb{R}^d$.

Denoting

$$\hat{g}_n(u) = \frac{1}{nh_n^d} \sum_{i=0}^{n-1} Y_i K_i \tag{19}$$

and

$$\hat{f}_n(u) = \frac{1}{nh_n^d} \sum_{i=0}^{n-1} K_i, \tag{20}$$

we have

$$\hat{\psi}_n(u) = \frac{\hat{g}_n(u)}{\hat{f}_n(u)}. \tag{21}$$

Using Lemma 1 and (15), we obtain

$$E\hat{g}_n(u) = \frac{1}{h^d(n)} EY_0K_0 \xrightarrow{n} (m(u)+\gamma)f(u)\int K(v)dv \tag{22}$$

for almost all $u \in \mathbb{R}^d$.

We shall now show that $\mathrm{var}(\hat{g}_n(u)) \xrightarrow{n} 0$, for almost all $u \in \mathbb{R}^d$. Due to the stationarity, we get

$$\begin{aligned} \mathrm{var}(\hat{g}_n(u)) &= \frac{1}{n^2h^{2d}(n)} \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \mathrm{cov}[Y_iK_i, Y_jK_j] \\ &= \frac{1}{nh^d(n)} \frac{1}{h^d(n)} \mathrm{var}(Y_0K_0) \\ &+ \frac{2}{n^2h^{2d}(n)} \sum_{j=1}^{n-1} (n-j)\mathrm{cov}[Y_0K_0, Y_jK_j]. \end{aligned} \tag{23}$$

Let us take the second term in (23) into account, i.e., let us consider $\mathrm{cov}[Y_0K_0, Y_jK_j]$. By Lemmas 2 and 3,

$$\begin{aligned} &\mathrm{cov}[Y_0K_0, Y_jK_j] \\ &= g_j\left(\sigma_\xi^2 E^2K_0 + EK_0\mathrm{cov}[K_0m(U_0), m(U_0)]\right) + EK_0c^TQ_{0,j}(u)c \\ &= g_j\left(\sigma_\xi^2 E^2K_0 + EK_0\mathrm{cov}[K_0m(U_0), m(U_0)]\right) \\ &+ E^2K_0c^TA^jPc + g_jEK_0\mathrm{cov}[K_0, m(U_0)]c^T\bar{X}_0. \end{aligned}$$

Because of the asymptotic stability of A, $\sum_{i=1}^{\infty}|g_i| < \infty$. Consequently, using Toeplitz's lemma (see e.g. Loéve 1977),

$$\sum_{j=1}^{\infty}\left(1-\frac{j}{n}\right)g_j$$

has a finite limit as $n \to \infty$.

In turn, by virtue of Lemma 1,

$$\frac{1}{h^d(n)} EK_0 \overset{n}{\to} f(u) \int K(v)dv,$$

$$\frac{1}{h^d(n)} \operatorname{cov}[K_0 m(U_0), m(U_0)] \overset{n}{\to} m^2(u) f(u) \int K(v)dv$$

and

$$\frac{1}{h^d(n)} \operatorname{cov}[K_0, m(U_0)] \overset{n}{\to} m(u) f(u) \int K(v)dv$$

for almost all $u \in \mathbb{R}^d$. Therefore the second term in (23) is of order $O(1/n)$.

Considering the first term in (23), using Lemmas 4 and 1, we get

$$\begin{aligned}\frac{1}{h^d(n)} \operatorname{var}(Y_0 K_0) &\overset{n}{\to} \Bigg[m^2(u) + \sigma_\xi^2 + \sigma_\eta^2 + \big(\operatorname{var}(m(U_0)) + \sigma_\xi^2\big) \sum_{i=1}^{\infty} g_i^2 \\ &+ E^2 m(U_0) \left(\sum_{i=1}^{\infty} g_i \right)^2 + 2Em(U_0) m(u) \sum_{i=1}^{\infty} g_i \Bigg] f(u) \int K^2(v)dv \\ &= \left[(m(u) + \gamma)^2 + \operatorname{var}(m(U_0) \sum_{i=1}^{\infty} g_i^2 + \sigma_\xi^2 \sum_{i=1}^{\infty} g_i^2 + \sigma_\eta^2 \right] f(u) \int K^2(v)dv\end{aligned}$$

for almost all $u \in \mathbb{R}^d$.

Therefore the first term in (23) is of order

$$O\left(\frac{1}{nh^d(n)} \right). \tag{24}$$

In turn, along with (15), (16), (22) and (23), this implies

$$\hat{g}_n(u) \overset{n}{\to} (m(u) + \gamma) f(u) \int K(v)dv$$

in probability for almost all $u \in \mathbb{R}^d$.

Using similar arguments, we can easily show that

$$\hat{f}_n(u) \overset{n}{\to} f(u) \int K(v)dv$$

in probability for almost all $u \in \mathbb{R}^d$.

Finally, recalling (19), (20) and (21), we obtain the desired convergence of $\hat{\psi}_n(u)$. The proof of the theorem has been completed. □

6. The rate of convergence

Imposing some mild regularity conditions on the transform characteristic of the non-linear subsystem, we show how to select $\{h(n)\}$ and give the rate of convergence.

Writing $X_n = O(a_n)$ in probability, we mean that $d_n X_n / a_n \overset{n}{\to} 0$ in probability for any number sequence $\{d_n\}$ convergent to zero.

Theorem 4
Let all the assumptions of Theorem 2 be satisfied. Suppose that K has a compact support. Let m satisfy a Lipschitz condition of order 1 in a neighbourhood of u and let $f(u) > 0$. If

$$h(n) \sim n^{-1/(2+d)}, \tag{25}$$

then

$$|\hat{m}_n(u) - m(u)| = O(n^{1/(2+d)}) \text{ in probability.} \tag{26}$$

Proof.
Let $\hat{g}_n(u)$ and $\hat{f}_n(u)$ have the same meaning as in the proof of Theorem 2. Clearly,

$$\begin{aligned} P\{|\hat{\psi}_n(u) - (m(u)+\gamma)| > \varepsilon\} \le P\{|\hat{a}_n(u) - (m(u)+\gamma)| > t\} \\ + P\{|\hat{b}_n(u) - E\hat{b}_n(u)| > t\}, \end{aligned} \tag{27}$$

where

$$\hat{a}_n(u) = \frac{\hat{g}_n(u)}{E\hat{f}_n(u)},\ \hat{b}_n(u) = \frac{\hat{f}_n(u)}{E\hat{f}_n(u)}$$

and

$$t = \frac{\varepsilon}{\varepsilon + |m(u)+\gamma| + 1}.$$

By Chebyshev's inequality, the first probability on the right-hand side of (27) is bounded from above by

$$\frac{\operatorname{var}(\hat{a}_n(u)) + (E\hat{a}_n(u) - (m(u)+\gamma))^2}{t^2}.$$

In turn, by (24) and Lemma 1, we have

$$\operatorname{var}(\hat{a}_n(u)) = O\left(\frac{1}{nh^d(n)f(u)}\right).$$

Moreover, by (6) and the Lipschitz condition, we obtain

$$\begin{aligned} |E\hat{a}_n(u) - (m(u)+\gamma))| &= \left|\frac{E\{m(U_0)K_0\}}{EK_0} - m(u)\right| \\ &= \left|\frac{\int |m(v) - m(u)| K\left(\frac{u-v}{h(n)}\right) f(v)dv}{\int K\left(\frac{u-v}{h(n)}\right) f(v)dv}\right| \\ &\le \sup_{v \in u + h(n)A} |m(v) - m(u)| = O(h(n)). \end{aligned}$$

Here A is the support of the kernel function. Since $\operatorname{var}(\hat{b}_n(u)) = O(1/nh^d(n))$ and (25) holds, we have completed the proof. □

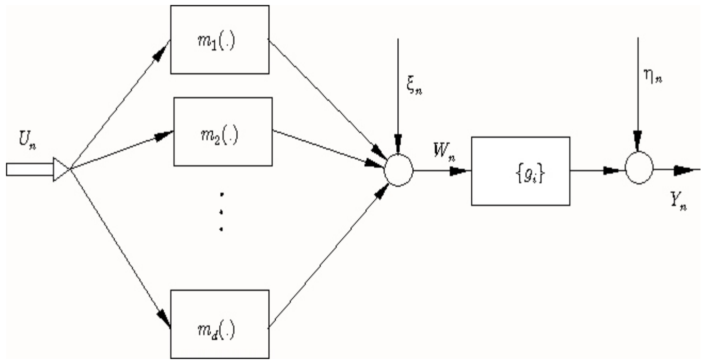

Figure 2: The additive Hammerstein system.

The rate obtained in (26) depends essentially on the dimension of the input signal, i.e. the bigger the dimension, the slower the rate of convergence. This "curse of dimensionality" can be partially overcome if either the non-linear characteristic is smoother than that in Theorem 4 or it has a special form. The latter can have the additive form of $m(u) = m_1(u^{(1)}) + \cdots + m_d(u^{(d)})$, where m_i, $i = 1, \ldots, d$ are real variable functions (Fig. 2). It is natural to conjecture that for such a case, $\{\hat{m}_n\}$ achieves the same rate as in (26) with $d = 1$, i.e., $O(n^{-1/3})$. The solution to this problem, however, is not obvious even if pairs (U_i, Y_i) are mutually independent, i.e. if the dynamic subsystem degenerates to a memoryless linear amplifier (Stone 1985).

7. Concluding remarks

In this paper, description (3) of the dynamical subsystem has been assumed. If the second equation in (3) has the form $Y_n = c^T X_n + g_0 W_n + \eta_n$, then $E\{Y_n | U_n = u\} = g_0 m(u) + \gamma$ and $\mathrm{cov}[Y_n, \|U_n\|] = g_0\alpha$. Therefore both $m(u)$ and g_i can be estimated generally only up to some constant factors. However, if the values of the non-linear characteristics are known at two points, which is often the case, then we can estimate $m(u)$ as well as g_i.

Acknowledgement

This work was supported by the University of Manitoba Research Committee of the Senate under contract 431-2730-63.

References

Anderson, T. W., 1971, *The Statistical Analysis of Time Series* (New York: Wiley).

Bierens, H. T., 1983, *J. Am. statist. Ass.*, 78, 699.

Billings, S. A., and Fakhouri, S. Y., 1979, *Int. J. Systems Sci.*, 10, 567.

Chang, F. H. I., and Luus, R., 1971, *I.E.E.E. Trans. autom. Control*, 16, 464.

Collomb, G., 1981, *Int. statist. Rev.*, 49, 75; 1984, *Z. Wahr. Verw. Geb.*, 66, 441.

Collomb, G., and Härdle, W., 1986, *Stoch. Process. Appl.*, to be published.

Devroye, L., 1981, *I.E.E.E. Conf on Decision Control*, 934.

Kaminskas, V. A., 1975, *Automn remote Control,* 36, 1107.

Kung, M. C., and Womack, B. F., 1984, *I.E.E.E. Trans. autom. Control*, 29, 170.

Loève, M., 1977, *Probability Theory I* (New York: Springer-Verlag).

Nadaraya, E. A., 1964, *Theor. Probability Appl.*, 9, 141.

Narendra, K. S., and Gallman, P. G., 1966, *I.E.E.E. Trans. autom. Control*, 11, 546.

Rosenblatt, M., 1971, *Ann. Math. Statist.*, 42, 1815.

Sawchuk, A. A., and Strand, T. C., 1982, *Applications of Optical Fourier Transforms*, edited by H. Stark (New York: Academic Press), p. 371.

Shih, D. H., and Kung, F. C., 1985, *Int. J. Control*, 42, 1399.

Stapleton, J. C., and Bass, S. C., 1985, *I.E.E.E. Trans. Circuits Syst.*, 32, 143.

Stone, C., 1985, *Ann. Statist.*, 13, 689.

Thatachar, M. A. L., and Ramaswamy, S., 1973, *Int. J. Control*, 18, 741.

Watson, G. S., 1964, *Sankhyā Ser. A*, 26, 355.

Wheeden, R. L., and Zygmund, A., 1977, *Measure and Integral* (New York: Dekker).