Non-parametric orthogonal series identification of Hammerstein systems

WŁODZIMIERZ GREBLICKI[1]

printed in:
INTERNATIONAL JOURNAL OF SYSTEMS SCIENCE
Vol. 20, No. 12, pp. 2355–2367, 1989

The non-linearity in a discrete system governed by the Hammerstein functional is identified. The system is driven by a random white white input signal and the output is disturbed by a random white noise. No parametric a priori information concerning the non-linearity is available and non-parametric algorithms are proposed. The algorithms are derived from the trigonometric as well as Hermite orthogonal series. It is shown that the algorithms converge tothe unknown characteristic in a pointwiese manner and that the mean integrated square error converges to zero as the number of observations tends to infinity. The rate of convergence is examined. A numerical example is also given.

1. Introduction

In this work, we identify a system comprising two parts, i.e. a non-linear memoryless subsystem followed by a linear dynamic one. The system is referred to as Hammerstein one since its behaviour is governed by the Hammerstein functional. A number of works concerning identification of the system have appeared (see, e.g., Gallman 1975, Chang and Luus 1971 Tchatachar and Ramaswamy 1973, Haist *et al.* 1973, Billings and Fakhouri 1979). However, we focus our attention on recovering the non-linearity since it is undoubtedly much more interesting than identification of the linear subsystem which can be performed with standard correlation methods.

All the authors mentioned above have assumed that the non-linear characteristic can be represented in a parametric form, i.e. that the unknown function is a member of a class of functions which can be parameterized—e.g.. a family of all polynomials whose order does not exceed a known number. So, in order to apply successfully their identification algorithms we must possess information which can often be unavailable.

In contradistinction to those authors, we do not require that the class of functions the characteristic belongs to must be parameterized. It is clear that, e.g., the family of all functions belonging to L_p, $p > 1$, which is the case in the

[1]Institute of Engineering Cybernetics, Technical University of Wrocław, 50-370 Wrocław, Poland

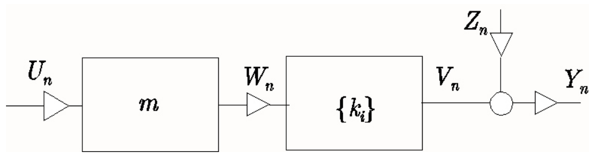

Figure 1: The identified Hammerstein system.

present work, is so ample that it cannot be parameterized. In the light of this, our problem is non-parametric.

The non-parametric approach to recovering the non-linearity in the Hammerstein system has been proposed by Greblicki and Pawlak (1986, 1987, 1989 a, b). Observing that the unknown characteristic is a regression function, they have applied a kernel regression estimate and have demonstrated the convergence of their identification algorithms.

In this work, we propose another procedure based on orthogonal expansions and examine algorithms derived from the trigonometric and Hermite orthogonal series. We show that both algorithms converge to the unknown characteristic in a pointwise manner and that the integrated square error converges to zero as the number of observations tends to infinity. Pointwise convergence has the rate $O(n^{-(2q-1)/4q})$ in probability, where q is the number of derivatives of the unknown characteristic. A numerical example is also presented.

2. The identified system

The identified Hammerstein system with input U_n and output Y_n is shown on Fig. 1. In the present work $\{U_n; n = \cdots, -1, 01, \cdots\}$ is a stationary white random process. The characteristic m of the memoryless subsystem is described by the following state-space equation:

$$X_{n+1} = AX_n + bW_n$$
$$V_n = c^T X_n,$$

where A is a matrix, b and c are vectors, and ^{T} denotes transposition. Here A, b, and c are all unknown, but it is assumed that A is asymptotically stable. Clearly $\{X_n; n = \cdots, -1, 01, \cdots\}$ is a stationary random process; V_n is not accessible for measurement and we have only Y_n, where

$$Y_n = V_n + Z_n$$

and where $\{Z_n\}$ is an additive stationary noise with zero mean and finite variance. Moreover, processes $\{U_n\}$ and $\{Z_n\}$ are mutually independent.

By f we denote the probability density of U_n which is assumed to exist. Additionally, U_n are distributed symmetrically, i.e.,

$$f(u) = f(-u), \text{ all } u \in R. \tag{1}$$

Concerning the non-linearity, we assume that m is odd, i.e. that

$$m(u) = -m(-u), \text{ all } u \in R. \tag{2}$$

The above restrictions imposed on f and m just make it possible to present the idea of the present work in a clear way. One can easily employ the approach for any density and any non-linearity.

Our goal is to recover m from input-output observations of the whole system, i.e. from $(U_1, Y_1), \cdots, (U_n, Y_n)$. The problem of the non-parametric recovering of m has already been studied by Greblicki and Pawlak (1986, 1987, 1989a, b), who have derived their algorithms from the kernel regression estimate. In this work, we propose algorithms based on orthogonal expansions. We apply the trigonometric as well as Hermite series and show that the algorithm derived from them converge to the characteristic.

Compared with the kernel algorithm, the orthogonal series has some advantages of a computational nature since it occupies a much smaller amount of computer memory. Moreover, the orthogonal series estimate fits the unknown characteristic better in the sense that it preserves the fact that the characteristic is odd- a property which can sometimes be important.

3. The trigonometric series algorithm

The motivation for the identification algorithms examined in the paper is that the problem of recovering m is equivalent to regression function estimation. Since X_n is independent of U_n and $EX_n = 0$, which is due to (1) and (2), we have

$$E\{Y_{n+1} | U_n = u\} = \alpha m(u) \tag{3}$$

where $\alpha = c^T b$. For simplicity of notation, $\alpha = 1$. In order to recover m we just estimate the regression in (3).

In §§ 3 and 4, the support of f is a subset of the interval $[-\pi, \pi]$ which means that $f(u)$ equals zero outside the interval.

Clearly $m(u) = g(u)/f(u)$, where $g(u) = m(u)f(u)$. Since g is odd

$$g(u) \sim \sum_{k=1}^{\infty} a_k \sin ku,$$

where

$$a_k = \frac{1}{\pi} \int g(u) \sin(ku) du = \frac{1}{\pi} E\left\{Y_1 \sin kU_0\right\}.$$

The integral is taken in the set $[-\pi, \pi]$. In the remainder of the work we omit the bounds in all integrals if it is clear what they are.

Let us define the following estimate of a_k:

$$\hat{a}_k = \frac{1}{n\pi} \sum_{i=1}^{n} Y_{i+1} \sin kU_i$$

and let

$$\hat{g}(u) = \sum_{k=1}^{N} \hat{a}_k \sin ku$$

be an estimate of $g(u)$, where $N = N(n)$ and $\{N(n)\}$ is a sequence of integers.

For f even,

$$f(u) \sim \frac{1}{2\pi} + \sum_{k=1}^{\infty} b_k \cos ku$$

where

$$b_k = \frac{1}{\pi} \int f(u) \cos(ku) du = \frac{1}{\pi} E \{\cos kU_0\}.$$

Let

$$\hat{f}(u) = \frac{1}{2\pi} + \sum_{k=1}^{N} \hat{b}_k \cos ku,$$

where

$$\hat{b}_k = \frac{1}{2n\pi} \sum_{i=1}^{N} \cos kU_i.$$

We define our estimate of $m(u)$, i.e. our identification algorithm, in the following way:

$$\hat{m}(u) = \frac{\sum_{k=1}^{N} \hat{a}_k \sin ku}{\frac{1}{2\pi} + \sum_{k=1}^{N} \hat{b}_k \cos ku}. \tag{4}$$

The number sequence $\{N(n)\}$ satisfies the following two restrictions:

$$N(n) \xrightarrow{n} \infty \tag{5}$$

and

$$\frac{N(n)}{n} \xrightarrow{n} 0. \tag{6}$$

The estimate can be rewritten in an alternative form. Observe that

$$2 \sum_{k=1}^{N} \sin kx \sin ky = D_N(x-y) - D_N(x+y), \tag{7}$$

where

$$D_N(x) = \frac{\sin\left(N + \frac{1}{2}\right) x}{2 \sin \frac{x}{2}}$$

is the Nth Dirichlet kernel and denote $w_N(x,y) = D_N(x-y) - D_N(x+y)$. Moreover,

$$1 + 2\sum_{k=1}^{N} \cos kx \cos ky = v_N(x,y), \tag{8}$$

where $w_N(x,y) = D_N(x-y) + D_N(x+y)$. Proofs of (7) and (8) are given in the Appendix. Hence

$$\hat{m}(u) = \frac{\displaystyle\sum_{i=1}^{n} Y_{i+1} w_N(U_i, u)}{\displaystyle\frac{1}{2\pi} + \sum_{k=1}^{N} \hat{b}_k \cos ku}. \tag{9}$$

4. Convergence of the trigonometric series algorithm

We shall now show that our algorithm converges to the unknown characteristic in a pointwise manner. First of all, we shall verify the convergence of $\hat{g}(u)$ to $g(u)$.

Since $E\hat{a}_k = a_k$, we have

$$E\hat{g}(u) = \sum_{k=1}^{N} a_k \sin ku = \int D_N(u-v) g(v) f(v) dv. \tag{10}$$

The above expression is just the Nth partial sum of the trigonometric expansion of g. Now, assuming that $f \in L_p$, $p > 1$, and that m is bounded we get $g \in L_p$ and then using Theorem A.1 in the Appendix and (5) we have

$$E\hat{g}(u) \xrightarrow{n} g(u), \text{ almost every } u \in [-\pi, \pi]. \tag{11}$$

Here and throughout the work "almost everywhere" means "almost everywhere with respect to the Lebesgue measure". In order to show that

$$E(\hat{g}(u) - g(u))^2 \xrightarrow{n} 0, \text{ almost every } u \in [-\pi, \pi] \tag{12}$$

it now suffices merely to verify that

$$\operatorname{var} \hat{g}(u) \xrightarrow{n} 0, \text{ almost every } u \in [-\pi, \pi]. \tag{13}$$

Obviously,

$$\begin{aligned} 4\pi^2 \operatorname{var} \hat{g}(u) &= \frac{1}{n^2} \operatorname{var}\left(\sum_{i=1}^{n} Y_{i+1} w_N(U_i, u) \right) \\ &\le \frac{2}{n^2} \operatorname{var}\left(\sum_{i=1}^{n} Y_{i+1} D_N(U_i - u) \right) \\ &+ \frac{2}{n^2} \operatorname{var}\left(\sum_{i=1}^{n} Y_{i+1} D_N(U_i + u) \right). \end{aligned} \tag{14}$$

We shall now show that

$$\frac{1}{n^2}\operatorname{var}\left(\sum_{i=1}^{n} Y_{i+1} D_N(U_i - u)\right) = O\left(\frac{N}{n}\right) \tag{15}$$

for almost every $u \in [-\pi, \pi]$. The quantity in (15) equals

$$\begin{aligned}
&\frac{1}{n}\operatorname{var}\left[Y_1 D_N(U_0 - u)\right] \\
&+ \frac{1}{n^2}\sum_{i=1}^{n}(n-i)\operatorname{cov}\left(Y_{i+1} D_N(U_i - u), Y_1 D_N(U_0 - u)\right) \\
&= V_1(u) + V_2(u)
\end{aligned}$$

say. Clearly

$$V_1(u) = \frac{1}{n} E\left\{r(U_0) D_N^2(U_0 - u)\right\} - \frac{1}{n} E^2\left\{m(U_0) D_N(U_0 - u)\right\}$$

where $r(u) = R\{Y_1^2 | U_0 = u\} = EY_0^2 + m^2(u)$. Since

$$D_N^2(x) = \frac{N+1}{2} F_N(x),$$

where F_N is the Nth Fejér kernel, we have

$$V_1(u) = \frac{N+1}{2n}\int r(v) f(v) F_N(u - v) dv - \frac{1}{n}\left[\int m(v) f(v) D_N(u - v) dv\right]^2.$$

Using Theorem A.1 in the Appendix and (5) we find the integrals in the above expression converging to $r(u)f(u)$ and $m(u)f(u)$ for almost every $u \in [-\pi, \pi]$, respectively. Thus

$$\operatorname{var} V_1(u) = O\left(\frac{N}{n}\right), \text{ almost every } u \in [-\pi, \pi]. \tag{16}$$

In turn, by virtue of Lemma A.1 in the Appendix,

$$V_2(u) = \frac{1}{n^2} c^T \left[d_N(u) \sum_{i=1}^{n} (n-i) A^i\right] c$$

where

$$\begin{aligned}
d_N(u) = {} & A \operatorname{cov}(X_0, X_0) A^T E^2\{D_N(U_0 - u)\} \\
& + bb^T E\{D_N(U_0 - u)\} \operatorname{cov}\left[W_0, W_0 D_N(U_0 - u)\right].
\end{aligned}$$

Applying again Theorem A.1 we find that

$$\operatorname{var} V_2(u) = O\left(\frac{N}{n}\right), \text{ almost every } u \in [-\pi, \pi].$$

This and (16) imply (15).

Since similar property can be shown for the second term in (4), we have finally verified (13).

Using similar arguments we can easily show that

$$E(\hat{f}(u) - f(u))^2 \overset{n}{\to} 0, \text{ almost every } u \in [-\pi, \pi]. \tag{17}$$

From this and (12) we get the following theorem.

Theorem 1

Let $f \in L_p$, $p > 1$, and let m be bounded. If (5) and (6) hold, then we have

$$\hat{m}(u) \overset{n}{\to} m(u) \text{ in probability}$$

for almost every $u \in [-\pi, \pi]$ at which $f(u) > 0$.

Theorem 1 says that algorithm (4) converges to m in a pointwise manner.. However, one can be interested in a global error of the estimate. The integrated square error is dealt with the next one.

Theorem 2

Let $f \in L_p$, $p > 1$, and let m be bounded. Let, moreover,

$$|Z_n| \le c \text{ almost surely}$$

for some $c > 0$. If (5) and (6) hold, then

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

and

$$E \int_D (\hat{m}(u) - m(u))^2 du \overset{n}{\to} 0,$$

where D is the support of f.

Theorem 2 is an immediate consequence of Theorem 1 and the Lebesgue dominated convergence theorem for random functions (Glick 1974).

Let us observe that in order to memorize estimate (4) it suffices to store only $\hat{a}_k$ and $\hat{b}_k$, i.e. $2N$ numbers. The kernel algorithm examined by Greblicki and Pawlak (1986, 1987, 1989 a, b) require that all n pairs (U_i, Y_{i+1}), i.e., $2n$ numbers be kept in the memory. Having in mind that $N/n \to 0$ as $n \to \infty$, we conclude that the orthogonal series algorithm is much more convenient from the viewpoint of the amount of occupied computer memory.

Observe, moreover, that our estimate of an odd characteristic is also odd, which is not the case with the kernel estimate.

Imposing some smoothness conditions on f and m we can obtain results on the asymptotic rate of convergence of $\hat{m}(u)$ to $m(u)$ and give some recommendations concerning the choice of N.

Suppose that both f and m have q derivatives, and that $q-1$ first derivatives of both f and m equal zero for $u=-1$ and $u=1$, and that qth derivatives of f and g are square integrable. Thus, integrating $\int g(u)\sin kudu$ by parts one can find $a_k = |\alpha_k|\,k^{-q}$, where α_k equals $\int g^{(q)}(u)\sin kudu$ or $\int g^{(q)}(u)\cos kudu$ for q odd or even, respectively. Hence

$$\begin{aligned}|E\hat{g}(u)-g(u)| &\le \sum_{k=N+1}^{\infty}|a_k| \le \sum_{k=N+1}^{\infty}k^{-q}|\alpha_k| \\ &\le \left(\sum_{k=N+1}^{\infty}\alpha_k^2\right)^{1/2}\left(\sum_{k=N+1}^{\infty}k^{-2q}\right)^{1/2} = O(N^{-(q-1/2)}).\end{aligned}$$

Recalling (16) and selecting

$$N(n) \sim n^{1/2q}$$

we get

$$E(\hat{g}(u)-g(u))^2 = O(n^{-(2q-1)/2q}).$$

For similar reasons

$$E(\hat{f}(u)-f(u))^2 = O(n^{-(2q-1)/2q}).$$

Making use of Lemma 2 given by Greblicki and Pawlak (1985), we find

$$P\{|\hat{m}(u)-m(u)| > \varepsilon m(u)\} = O(n^{-(2q-1)/4q}) \tag{18}$$

for any $\varepsilon > 0$. Similarly, we have

$$|\hat{m}(u)-m(u)| = O(n^{-(2q-1)/4q}) \text{ in probability.}$$

Observe that for large q, i.e. for smooth f and m, the rate in (18) gets close to $1/n$, i.e. the rate that is typical for parametric inference.

5. The Hermite series algorithm

In §§ 3 and 4, we assumed that the absolute value of the input signal does not exceed π, i.e. that $|U_0| \le \pi$. If, however, $a \le U_0 \le b$ and both a and b are known we can simply replace the system $\{\sin ku, \cos ku; k=1,2,\cdots\}$ by

$$\left\{[(b-a)\pi]^{1/2}\sin\frac{2u-(a+b)}{b-a},\right.$$
$$\left.[(b-a)\pi]^{1/2}\cos\frac{2u-(a+b)}{b-a}; k=1,2,\cdots\right\}.$$

A problem arises when a or b or both are unknown or U_n are not bounded. In order to overcome this difficulty one can apply the Hermite orthogonal series, i.e. the system orthogonal on the real line R.

The Hermite system $\{h_k; k = 0, 1, 2, \cdots\}$ is a complete system of orthonormal functions defined in the following way:

$$h_k(u) = \frac{1}{\sqrt{2^k k! \sqrt{\pi}}} H_k(u) \exp\left(-\frac{u^2}{2}\right)$$

where

$$H_k(u) = e^{u^2} \frac{d^k}{du^k} e^{-u^2}$$

is the $k$the Hermite polynomial. For example, $H_0(u) = 1$, $H_1(u) = -2u$, $H_2(u) = 4u^2 - 2$, $H_3(u) = -8u^3 + 12u$, and so on. Clearly H_k and consequently h_k is even or odd foe even or odd k, respectively. Therefore

$$g(u) \sim \sum_{k=0}^{\infty} a_{2k+1} h_{2k+1}(u)$$

and

$$f(u) \sim \sum_{k=0}^{\infty} b_{2k} h_{2k}(u)$$

where

$$a_k = E\{Y_1 h_k(U_0)\}$$

and

$$b_k = E\{h_k(U_0)\}.$$

Let

$$\hat{a}_k = \frac{1}{n} \sum_{i=1}^{n} Y_{i+1} h_k(U_i)$$

and

$$\hat{b}_k = \frac{1}{n} \sum_{i=1}^{n} h_k(U_i)$$

be estimates of $\hat{a}_k$ and $\hat{b}_k$. Define

$$\hat{g}(u) = \sum_{k=0}^{N} \hat{a}_{2k+1} h_{2k+1}(u)$$

and

$$\hat{f}(u) = \sum_{k=0}^{N} \hat{b}_{2k} h_{2k}(u).$$

Let

$$\hat{m}(u) = \frac{\hat{g}(u)}{\hat{f}(u)} = \frac{\sum_{k=0}^{N} \hat{a}_{2k+1} h_{2k+1}(u)}{\sum_{k=0}^{N} \hat{b}_{2k} h_{2k}(u)}$$

be an estimate of $m(u)$.

Obviously,

$$\hat{g}(u) = \frac{1}{n}\sum_{i=1}^{n} Y_{i+1} t_{2N+1}(U_i, u)$$

and

$$\hat{f}(u) = \frac{1}{n}\sum_{i=1}^{n} s_{2N}(U_i, u)$$

where

$$t_{2N+1}(v,u) = \sum_{k=0}^{N} h_{2k+1}(v) h_{2k+1}(u)$$

and

$$s_{2N}(v,u) = \sum_{k=0}^{N} h_{2k}(v) h_{2k}(u).$$

It is well known that by virtue of Christoffel's formula

$$d_N(v,u) = \sqrt{\frac{N+1}{2}} \frac{h_{N+1}(v)h_N(u) - h_{N+1}(u)h_N(u)}{(u-v)}$$

where

$$d_N(v,u) = \sum_{k=0}^{N} h_k(v) h_k(u).$$

Clearly

$$t_{2N+1}(v,u) + s_{2N}(v,u) = d_{2N+1}(v,u).$$

Since h_k is ever or odd for k even or odd, respectively, then

$$-t_{2N+1}(v,u) + s_{2N}(v,u) = d_{2N+1}(v,-u).$$

Therefore

$$2t_{2N+1}(v,u) = d_{2N+1}(v,u) + d_{2N+1}(v,-u)$$

and

$$2s_{2N}(v,u) = d_{2N+1}(v,u) - d_{2N+1}(v,-u).$$

We are now in a position to rewrite our estimate in a kernel-type form

$$\hat{m}(u) = \frac{\sum_{i=1}^{n} Y_{i+1} t_{2N+1}(U_i, u)}{\sum_{i=1}^{n} s_{2N}(U_i, u)}. \tag{20}$$

Arguing as in §3 one can verity the following theorem:

Theorem 3

Let $f \in L_p$, $p > 1$, and let $g \in L_p$. If (5) holds and

$$\frac{N^{1/2}}{n} \overset{n}{\to} 0 \tag{21}$$

then we have

$$\hat{m}(u) \overset{n}{\to} m(u) \text{ in probability}$$

for almost every $u \in R$ at which $f(u) > 0$.

While proving Theorem 3 one should use the equiconvergence theorem (Szegö 1959, p. 243) which states that at every point $u \in R$ both the Hermite expansion of any integrable function and the trigonometric expansion of the function taken at any interval containing the point in its interior have the same limit. This and Theorem A.1 in the Appendix yield

$$E\hat{g}(u) \overset{n}{\to} g(u) \text{ almost every } u \in R.$$

In the analysis of $\operatorname{var} \hat{g}(u)$ one can apply Lema 4 given in Greblicki and Pawlak (1985) which states that

$$\int d_N^2(v,u)t(v)dv \leq c(u)(N+1)^{1/2},$$

where t is any integrable Borel function and c is finite for almost every $u \in R$.

One can also easily show that if f and g , $t_q(\cdot;f)$ and $t_q(\cdot;g)$, where

$$t_q(\cdot;f) = \left(u - \frac{d}{du}\right)^q f(u)$$

are square integrable and if, moreover, $N(n) \sim n^{1/q}$, we have

$$P\{|\hat{m}(u) - m(u)| > \varepsilon m(u)\} = O(n^{-(2q-1)/2q})$$

and

$$|\hat{m}(u) - m(u)| = O(n^{-(2q-1)/4q}) \text{ in probability.}$$

In order to verify the above, it just suffices to argue as in § 4 and use a result given be Walter (1977) which states that

$$|a_k| \leq (k+1)^{-q/2}|\alpha_k| \text{ and } |b_k| \leq (k+1)^{-q/2}|\beta_k|,$$

where α_k and β_k are kth coefficients of expansions of $t_q(\cdot;f)$ and $t_q(\cdot;g)$ in the Hermite series, respectively.

The final theorem is on the global error.

Theorem 4

Let $f, g \in L_p$, $p > 1$. Let, moreover,

$$|Z_n| \leq c \text{ almost surely}$$

for some $c > 0$. If (5) and (21) hold, then

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

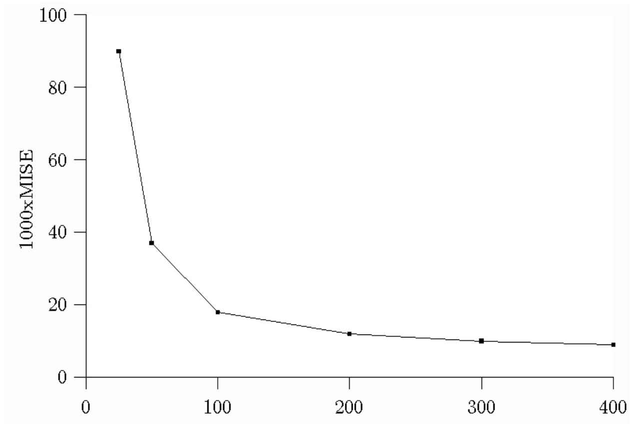

Figure 2: Convergence of the MISE simulation results.

6. Numerical example

In the numerical example illustrating the method, $m(u)$ equals -0.5, u or 0.5 according to $n < -0.5$, $-0.5 \le u \le 0.5$ or $u > 0.5$, respectively. The state vector simply a scalar and $X_{n+1} = aX_n + W_n$, $a = 0.25$, $Y_n = X_n + Z_n$. The input signal and output noise are distributed uniformly in the intervals $(-1, 1)$ and $(-0.5, 0.5)$, respectively. We apply the trigonometric series $1/2^{1/2}$, $\sin \pi u$, $\cos \pi u$, $\sin 2\pi u$, $\cos 2\pi u$, $\cdots$ which is orthonormal in the interval $[-1, 1]$. It is well known that at $u = -1$ and $u = 1$ the expansion of g converges to $[g(-1) + g(1)]/2$, i.e. to zero. For this reason $\hat{m}(u)$ does not behave well at the ends of the interval. In order to avoid this boundary effect in our measure of the quality of the algorithm, we define MISE in the following way:

$$\text{MISE} = \int_{-0.9}^{0.9} (\hat{m}(u) - m(u))^2 du.$$

For $N(n) = 0.7n^{-2/5}$, MISE has been estimated numerically for $n = 25$, 50, 100, 200, 300, and 400 and the results are shown in Fig. 2. We can observe that, for small n, MISE decreases rapidly and good fit of the estimate is achieved for quite moderate n.

7. Final remarks

In this work, f is even and g odd and consequently (3) holds. Hence, if α is not known we can estimate only $\alpha m(u)$, where α is unknown. This fact is not surprising since, whatever the method, we can identify characteristics of both the subsystems only up to some multiplicative factors. If, however, f is not even or g not odd, we have

$$E\{Y_{n+1}|U_n = u\} = \alpha m(u) + \beta,$$

where $\beta = c^T AEX_n$. In this case, we can expand f and g in a full trigonometric series and then take $\hat{m}_1(u) = \hat{m}(u) - \hat{m}(0)$ as the estimate of $m(u)$ and still to able to recover $\alpha m(u)$.

The orthogonal series algorithm presented here is just a non-parametric orthogonal series estimate of regression functions studied in the statistical literature, see, e.g., Greblicki and Pawlak (1985). The problem examined by is nevertheless more complicated since here pairs (U_n, Y_{n+1}) are dependent while in the work mentioned above, like in the most statistical literature, the pairs are independent.

Finally, let us observe that our algorithm estimates $m(u)$ consistently only if f and g belong to L_p, $p > 1$. From the practical viewpoint the restriction is of no importance and we are not far from the truth saying the we do not need to possess any *a priori* information in order to apply the algorithm and recover the unknown characteristic. This is not necessarily the case when applying algorithms offered by the parametric approach. The algorithm converge to the characteristic if only underlying *a priori* information is known. If not, the usage of any parametric procedure with expectation of its consistency is evidence of ungrounded optimism.

Appendix

Proof of (7) and (8)

Equality (7) is a consequence of the following two facts:

$$2 \sin x \sin y = \cos(x - y) - \cos(x + y)$$

and

$$D_N(x) = \frac{1}{2} + \sum_{k=1}^{N} \cos kx$$

(see, e.g., Sansone 1959). In order to verify (8) it suffices to recall that $\cos x \cos y = \cos(x + y) + \cos(x - y)$.

Theorem A.1

Let t be a Borel measurable function defined in the interval $[-\pi, \pi]$. Let

$$S_n(x) = \int D_n(x - y)t(y)dy \text{ and } s_n(x) = \int F_n(x - y)t(y)dy,$$

where D_n and F_n are the nth Dirichlet and Fejér kernel, respectively. If $t \in L_p$, $p > 1$, then

$$S_n(x) \xrightarrow{n} t(x), \text{ almost every } x \in [-\pi, \pi]. \tag{A.1}$$

If $t \in L_1$, then

$$s_n(x) \xrightarrow{n} t(x), \text{ almost every } x \in [-\pi, \pi]. \tag{A.2}$$

Remark A.1

For $p = 2$, result (22) has been shown by Carleson (1966) and then extended by Hunt (1968) for $p > 1$. Unfortunately, (A.1) does not hold for $p = 1$ and it is known that there exist integrable functions whose trigonometric expansions do not converge for almost every $x \in [-\pi, \pi]$ (see, e.g., Zygmund 1959, § 8.4).

Result (A.2) is simply the Lebesgue theorem on the Cesaro summability of the trigonometric series (see, e.g., Sansone 1959). The lemma below holds for any f and any m.

Lemma A.1

Let t be a Borel measurable function. Then in the Hammerstein system

$$\begin{aligned}
&\operatorname{cov}[X_{n+1}t(U_n), X_1 t(U_0)] \\
&= A^{n+1} \operatorname{cov}[X_0, X_0]\, A^T E^2\{t(U_0)\} \\
&+ A^n b E\{X^T\} A^T \operatorname{cov}[m(U_0), t(U_0)]\, E\{t(U_0)\} \\
&+ A^n b b^T E\{t(U_0)\} A^T \operatorname{cov}[m(U_0), m(U_0) t(U_0)]
\end{aligned}$$

for $n = 1, 2, 3, \cdots$.

References

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

[2] Carleson, L., 1966, *Acta Math.*, **116**, 135.

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

[4] Gallman, P.G., 1975, *I.E.E.E. Trans. autom. Control*, **20**, 771.

[5] Greblicki, W., and Pawlak, W., 1985, *Ann. Inst. statist. Math.*, **37**, 443; 1986, *I.E.E.E. Trans. autom. Control*, **31**, 74; 1987, *Int. J. Control*, **45**, 343; 1989 a, *I.E.E.E. Trans. Inform. Theory*, **35**, 409; 1989 b, *J. Franklin Inst.*, to be published.

[6] Haist, N.D., Chang, F.H.I., and Luus, R., 1973, *I.E.E.E. Trans. autom. Control*, **18**, 552.

[7] Hunt, R.A., 1968, On the convergence of Fourier series. *Orthogonal Expansions and their Continuous Analogues* (Carbondale: Southern Illinois University Press), pp. 235-255.

[8] Sansone, G., 1959, *Orthogonal Functions* (New York: Wiley-Interscience).

[9] Szegö, G., 1959, *Orthogonal Polynomials* (Providence, RI: American Mathematical Society).

[10] Tchatachar, M.A.L., and Ramaswamy, S., 1973, *Int. J. Control*, **18**, 741.

[11] Walter, G.G., 1977, *Ann. Statist.*, **5**, 1258.

[12] Zygmund, A., 1959, *Trigonometric Series* (Cambridge, UK; Cambridge University Press).