ON THE HERMITE SERIES APPROACH TO NONPARAMETRIC IDENTIFICATION OF HAMMERSTEIN SYSTEMS

A. Krzyżak * J. Z. Sąsiadek ** B. Kégl ***

* *Department of Computer Science and Software Engineering, Concordia University, Montreal, Quebec, Canada H3G 1M8*
** *Department of Aerospace Engineering, Carleton University, Ottawa, Ontario, Canada K1S 5B6*
*** *Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, C.P. 6128 Succ. Centre-Ville, Canada H3C 3J7*

Abstract: Nonlinear dynamic systems of Hammerstein type are identified from input and output measurements. Identification algorithms for a memoryless nonlinear part and for a linear dynamic part are proposed. Convergence and rates of convergence of the algorithms are investigated. The class of nonlinearities considered in the paper is very large and cannot be parameterized therefore nonparametric approach is used. The performance of identification algorithms is studied in simulation experiments.

Keywords: Hammerstein system identification, Hermite series, convergence

1. INTRODUCTION

Identification of nonlinear dynamic system is one of the fundamental problems of system theory. The two general identification techniques use expansions involving Volterra kernels and Wiener G-functionals. In the Volterra kernels expansion a nonlinear, dynamic system with input $x(t)$ and output $y(t)$ is represented by a generalized convolution $y(t) = \sum_{n=1}^{\infty} \int ... \int h_n(\tau_1, ..., \tau_n) x(t-\tau_1) ... \cdot x(t-\tau_n) d\tau_1 ... \tau_n$, where h_n is the Volterra kernel of order n (Sansone, 1980). The method of G-functionals, introduced by Wiener admits the description of nonlinear, dynamic system by the infinite series $y(t) = \sum_{n=0}^{\infty} G_n(k_n(\tau), x(\tau); \tau \in (-\infty, t))$, where G_n are some functionals and k_n are unknown kernels. Both methods are only applicable to smooth nonlinearities excluding such basic ones as hard and soft limiters and quantizers and their practical usefulness is limited due to excessive computational requirements.

An attractive alternative to general techniques is the block-oriented approach in which identified system is modelled by a cascade of simple functional blocks identified from the input-output observations. A simple but important block-oriented system is a sandwich system shown in Figure 1. It consists of a nonlinear, memoryless system sandwiched between two linear dynamic systems. Sandberg (Sandberg, 1992) showed that linear combinations of such systems can approximate a causal and time invariant nonlinear system. Sandwich systems have been applied among others to modelling of physiological systems (Marmarelis and Marmarelis, 1978).

[1] This work was supported by NSERC.

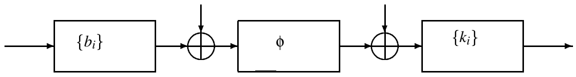


Fig. 1. Sandwich system

The most important block-oriented systems which can be derived from the sandwich system are the Hammerstein system consisting of a zero-memory nonlinearity followed by a linear dynamics (Figure 2), and the Wiener system consisting of a linear dynamic system followed by a memoryless nonlinearity (J.Bendat, 1990) (Figure 3). The Wiener system has been applied to signal detection and communication by Masry and Cambanis (Masry and Cambanis, 1980), modelling of neural network structures by Sandberg (Sandberg, 1992), and to modelling of biological systems (R. Emerson and Citron, 1992), (Hunter and Korenberg, 1986). The Hammerstein system identification has been first studied by Narendra and Gallman (Narendra and Gallman, 1982) followed by Chang and Luus (Chang and Luus, 1971), Chung and Sun (Chung and Sun, 1988) and Hsia (Hsia, 1977). It has been applied to adaptive control by Kung and Womack (Kung and Womack, 1984), to adaptive noise cancellation by Stapleton and Baas (Stapleton and Baas, 1992), to design of nonlinear predictors by McCannon et al. (T. McCannon and Wise, 1982) and to identification of biological systems by Hunter and Korenberg (Hunter and Korenberg, 1986). A blind identification of Hammerstien and Wiener systems has been studied by Bai (Bai, 2002). Comprehensive review of parametric nonlinear systems identification methods is provided in (Haber and Unbehauen, 1990) and references cited therein. All the papers mentioned above consider single-input single-output (SISO) systems with polynomial nonlinearities of fixed degree. Identification procedures for the linear and nonlinear parts are not independent and standard non-smooth nonlinearities such as dead-zone limiters, hard-limiters and quantizers are excluded. The papers also lack rigorous convergence analysis.

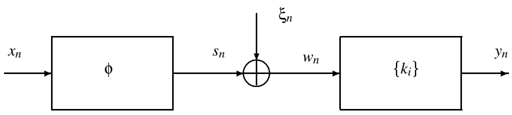


Fig. 2. Hammerstein system

The limitations imposed by the previous authors can be removed by using nonparametric approach to identification block-oriented systems. Nonparametric estimation received considerable attention in statistical literature (Eubank, 1999), (L. Györfi and Walk, 2002). The main advantage of this

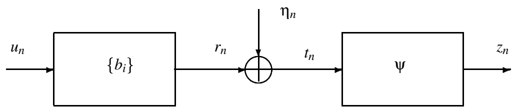


Fig. 3. Wiener system

approach is that we do not restrict nonlinearities to a class of functions described by a finite number of parameters, such as polynomials or trigonometric functions. The class of nonlinearities we are capable to recover by nonparametric estimates are all measurable L_2 functions. This class includes standard nonlinearities such as dead-zone limiters, hard-limiters and quantizers and is too large to be finitely parameterized. An elegant comparison of parametric and nonparametric techniques is provided by Hall (Hall, 1989). The most popular nonparametric techniques to date are kernel methods and methods based on the Fourier series orthogonal expansion. The former approach has been applied to memoryless block-oriented system identification by Greblicki and Krzyżak (Greblicki and Krzyżak, 1979) and Krzyżak and Partyka (Krzyżak and Partyka, 1993), and to Hammerstein system identification by Greblicki and Pawlak (Greblicki and Pawlak, 1987) and Krzyżak (Krzyżak, 1990). A recursive kernel version has been studied in (Greblicki and Pawlak, 1989), (Krzyżak, 1992). Identification algorithms based on the Fourier series expansions have been studied in Greblicki (Greblicki, 1989) and Krzyżak (Krzyżak, 1989), (Krzyżak, 1996). Kernel techniques do not work particularly well for periodic excitations whereas Fourier series approach is only applicable to inputs restricted to a bounded set. Continuous-time Hammerstein system identification has been studied by Greblicki (Greblicki, 2000).

In the present paper we identify dynamic Hammerstein systems by the algorithms based on the Hermite series expansion with the number of terms depending nonlinearly on the number of input-output measurements. We assume that Hammerstein system is driven by a stationary, white noise. We estimate the linear and nonlinear components simultaneously from the input and output observations of the whole system using the correlation method to identify the linear subsystem coefficients and nonparametric Hermite series regression estimate to recover the memoryless nonlinearity. We study convergence and convergence rates of the identification algorithms. Computer simulations of identification algorithms are presented and discussed.

The Hermite approach presented in this paper has clear advantages over kernel and Fourier series approaches. Unlike kernel approach it has modest storage requirements, that is we need to store only N coefficients instead n data measurements in case of kernel algorithms, where $N = o(n)$. The Hermite series method is applicable to the systems with input domain R, while the Fourier series method can only be used to identify systems with inputs restricted to finite intervals. Although Hermite series estimate is smooth for a finite number of measurements it can asymptotically approximate non-smooth nonlinearities such as thresholds and limiters. For a class of nonlinearities and input densities with finite Hermite series expansions (e.g of the polynomial type) the Hermite series identification approach offers better rates of convergence than the kernel and Fourier series methods.

2. HAMMERSTEIN SYSTEM

The outline of discrete Hammerstein system is given in Figure 2.

The nonlinear subsystem (I) is described by

$$W_n = \phi(X_n) + \xi_n, n = 0, \pm 1, \cdots \quad (1)$$

where X_n is R-valued stationary white noise with distribution μ and ξ_n is a stationary white noise with zero mean and finite variance σ_ξ^2. No correlation is assumed between ξ_n and X_n. Assume for simplicity that ϕ is a scalar function. The linear subsystem (II) is described by the ARMA model:

$$\begin{aligned} &Y_n + a_1 Y_{n-1} + \cdots + a_l Y_{n-l} \\ &= b_0 W_n + b_1 W_{n-1} + \cdots + b_l W_{n-l} \end{aligned}$$

where l is the order of the system (not assumed known). Coefficients $a_1, \cdots, a_l$ guarantee the asymptotic stability of the system, i.e. the roots $\mu_1, \cdots, \mu_l$ of the associated polynomial equation $z^l + a_1 z^{l-1} + \cdots + a_l = 0$ satisfy $|\mu_i| < 1, i = 1, \cdots, l$. These conditions imply that Y_n is weakly stationary as long as W_n is weakly stationary. Subsystem II can be described by state equations.

$$\begin{aligned} \hat{X}_{n+1} &= A\hat{X}_n + \mathbf{b}W_n \\ Y_n &= c^T \hat{X}_n + d_1 W_n \end{aligned} \quad (2)$$

where

$$A = \begin{bmatrix} \mathbf{0} & I \\ & a^T \end{bmatrix} \mathbf{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} a = \begin{bmatrix} a_l \\ \vdots \\ a_1 \end{bmatrix}$$

$$b = \begin{bmatrix} h_1 \\ \vdots \\ h_l \end{bmatrix} c = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} d_1 = h_0$$

$$\begin{aligned} &h_0 = b_0, h_i = b_i - \sum_{j=1}^{i} a_j h_{l-j}, \\ &i = 2, \cdots, l, n = 0, \pm 1, \cdots \end{aligned}$$

$\hat{X}_n$ is an l-dimensional state vector, while A is assumed to be asymptotically stable (i.e. eigenvalues of $\mathbf{A}$ lie inside a unit circle). These conditions imply that $\hat{X}_n$ and Y_n are weakly stationary as long as W_n is weakly stationary. By (2)

$$E\{Y_n|X_n\} = d_1\phi(X_n) + \alpha = m(X_n) \quad (3)$$

where $\alpha = E\phi(X)c^T(I-A)^{-1}b$.

From equation (2) we obtain an IIR representation

$$Y_n = \sum_{j=0}^{\infty} k_j W_{n-j} \quad (4)$$

where $k_0 = d_1 \neq 0, k_i = c^T A^{i-1} b, i = 1, 2, \cdots$ and $\sum_{i=0}^{\infty} |k_i| < \infty$ guarantees asymptotic stability of the linear subsystem. It follows from (4) that

$$E\{Y_n|X_n\} = k_0\phi(X_n) + \beta$$

where $\beta = E\phi(X)\sum_{i=1}^{\infty} k_i$.

The estimation problem is well defined if Y_n and X_n are random variables in the L_p sense, $p \geq 1$. This is the case if II is asymptotically stable and

$$EW_n^p < \infty. \quad (5)$$

Equations (1), (5) and assumptions on X_n imply that W_n is weakly stationary. Condition (5) holds if $E\xi_n^p < \infty$ and either

$$\phi \in L_p(\mu) \ \ (\text{i.e. } E|\phi(X)|^p < \infty) \quad (6)$$

or

$$\phi(x) < P(|x|) \ \text{ and } \ E|X|^{ps} < \infty \quad (7)$$

where P is a polynomial of order s. Observe that condition (6) is weaker than (7) and that it is satisfied for a random variable X with arbitrary distribution. Asymptotic stability of II and (6) imply that α and β exist. Restrictions (6) and (7) on nonlinearity are totally independent of the estimation algorithms discussed in this paper. It is obvious that the class of Borel functions satisfying (6) or (7) is so large that it cannot be parameterized. It contains nonlinearities with jump discontinuities such as dead-zone limiters, hard-limiters, smooth-limiters and quantizers. This is the reason we use the nonparametric approach to estimate ϕ.

3. IDENTIFICATION ALGORITHMS AND CONVERGENCE

To estimate both ϕ and the linear subsystem parameters the Hammerstein system we will use the sequence $\{(X_i, Y_i)\}, i = 0, 1, \cdots, n-1$ of n dependent observations of the input X and the output Y of the whole system. In order to estimate the impulse response function $\{k_j\}, j = 0, 1, \cdots$ we will use the correlation approach. Asymptotic stability of the system implies that $k_j \to 0$ as n $\to \infty$. Therefore we can assume that for j greater than some threshold N the $k_j' s$ are negligible and we can estimate the finite set of N parameters. We obtain from (4)

$$\rho_{i,n} = cov\{Y_{n+i}, X_n\} = k_i/k_0 cov\{Y_n, X_n\} \quad (8)$$

where $cov\{Y_n, X_n\} = E\phi(X_0)X_0$.

If $E\phi(X_0)X_0 \neq 0$ then (8) yields

$$k_i = k_0 \rho_{i,n} \rho_{0,n}/|\rho_{0,n}|^2, \quad i = 1, \cdots, N. \quad (9)$$

We define the following estimate of k_i motivated by (9):

$$\hat{k}_{i,n} = k_0 \hat{\rho}_{i,n} \hat{\rho}_{0,n}/|\hat{\rho}_{0,n}|^2$$

where

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

$$\bar{Y} = \frac{1}{n} \sum_{i=0}^{n-1} Y_i, \quad \bar{X} = \frac{1}{n} \sum_{i=0}^{n-1} X_i.$$

Since Y_n is a weakly stationary process then $\hat{\rho}_{i,n} \to cov\{Y_{n+i}, X_n\}$ in probability as $n \to \infty$. Hence we have

Lemma 1. If (6) holds, the linear, dynamic subsystem is asymptotically stable and $E\phi(X)X \neq 0$ then

$$\hat{k}_{i,n} \to k_i \text{ almost surely}$$

as $n \to \infty, i = 1, 2, \cdots, N$.

In order to estimate the cross-correlation function γ_i of the linear subsystem, notice

$$\gamma_i = cov\{S_{n+i}, W_n\}/cov\{S_n, W_n\}$$
$$= k_i/k_0, i = 1, \cdots, N.$$

We can estimate γ_i by

$$\hat{\gamma}_{i,n} = \hat{\rho}_{i,n} \hat{\rho}_{0,n}/|\hat{\rho}_{0,n}|^2 i = 1, \cdots, N.$$

In order to recover ϕ we first estimate the regression function m defined in (3) by the kernel regression estimate. The presence of undetermined coefficients d_1 and α in (3) is a consequence of the fact that the signal W_n is not accessible for the measurements. It would be possible to determine d_1 and α if $\phi(x_1)$ and $\phi(x_2)$ were known for some x_1 and x_2 at which $m_n(x)$ were consistent and $\phi(x_1) \neq \phi(x_2)$. Then we could estimate d_1 by:

$$d_{1,n} = \frac{m_n(x_1) - m_n(x_2)}{\phi(x_1) - \phi(x_2)}$$

and α by:

$$\alpha_n = m_n(x_1) - \frac{\phi(x_1)}{\phi(x_1) - \phi(x_2)} (m_n(x_1) - m_n(x_2))$$

Using (3) and the formulas above we define an estimate of ϕ by

$$\phi_n(x) = (m_n(x) - \alpha_n)/d_{n,1}.$$

Similarly to estimating regression function Φ in section 3 we can estimate m by the Hermite series estimate

$$m_n(x) = \frac{\sum_{i=1}^n Y_i D_N(x, X_i)}{\sum_{i=1}^n D_N(x, X_i)} \quad (10)$$

where $(X_1, Y_1), \ldots, (X_n, Y_n)$ be observations of an $R^d \times R$-valued random vector (X,Y) and $E|Y| < \infty$. In the theorem below we will consider the convergence of algorithm (10). Let $g(x) = \Phi(x) f(x)$.

Theorem 1. Let $EY^2 < \infty$, A be asymptotically stable, and $f, g \in L_2$. If N satisfies

$$N \to \infty \quad (11)$$

$$N/n \to 0 \quad (12)$$

then

$$m_n(x) \to m(x) \text{ in probability}$$

as $n \to \infty$ for almost all x.

If, in addition, $t_r(\cdot) = (x - d/dx)^r(\cdot) \in L_2$ for some integer $r > 0$ for both f, g, then

$$|m_n(x) - m(x)| = O\left(n^{-\frac{2r-1}{2(2r+1)}}\right) \text{ in probability}$$

for almost all x for optimal $N = n^{2/(2r+1)}$.

Remark 1. If condition (12) in Theorem 1 is strengthened to

$$N/n \log n \to 0 \text{ as n} \to \infty \quad (13)$$

then

$$m_n(x) \to m(x) \text{ almost surely}$$

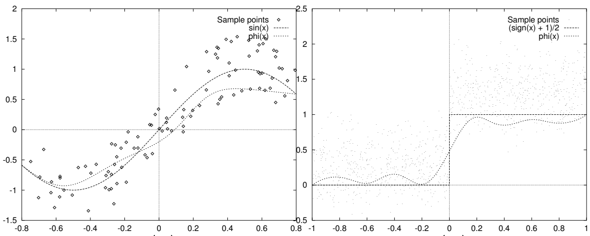
as $n \to \infty$ for almost all x. If f and g satisfy the smoothness assumptions of Theorem 1 then

(13) implies $|m_n(u) - m(u)| = O\left(n^{\frac{-2r+1}{4r+2}} / \log n\right)$ almost surely for almost all x for optimal $N(n) = (n/\log n)^{1/(r+1/2)}$.

4. SIMULATION EXPERIMENTS

In this section we present simulation results for identification of Hammerstein systems. We focus our attention on recovering static nonlinearity.

We simulated Hammerstein system from Figure 2 using two nonlinear functions, $f_1(x) = \sin x$ and $f_2(x) = (sign\, x + 1)/2$ in the intervals $[a_1, b_1] = [-0.8, 0.8]$ and $[a_2, b_2] = [-1, 1]$, respectively. The input x_n was generated uniformly in $[a_i, b_i], i = 1, 2$. The impulse response of the linear system was set to $k_i = e^{-i}$. We tested both systems for six different sample sizes, $n = 30, 50, 100, 200, 500, 1000$, and three different noise levels $\sigma_\xi = 0.01, 0.1, 0.2$, where ξ was zero mean Gaussian. Figure 4 shows two examples of individual experiments.



(a) (b)

Fig. 4. Results of individual experiments. (a) $n = 100$, $N = 50$, $\sigma_\xi = 0.01$, $f_1(x) = \sin x$, (b) $n = 1000$, $N = 120$, $\sigma_\xi = 0.2$, $f_2(x) = sign\, x$.

The experiments show that the identification algorithm is rather robust in terms of the choice of order of the Hermite kernel. This made it possible to choose $N(n)$ uniformly over different nonlinearities and noise levels, only as a function of n so that the corresponding error terms $L(n, N(n))$ were relatively small in each experiment (see Table 1). Although in real applications the nonlinear function is obviously unknown, our experiments suggest that practitioners have a relatively free choice of the order of the Hermite kernel. As a guideline, we fitted a logarithmic function $(32 \log n - 100)$ to the empirical function $N(n)$.

Table 1. The chosen order of the Hermite kernel as a function of the data size.

n	30	50	100	200	500	1000
N	16	20	50	70	100	120

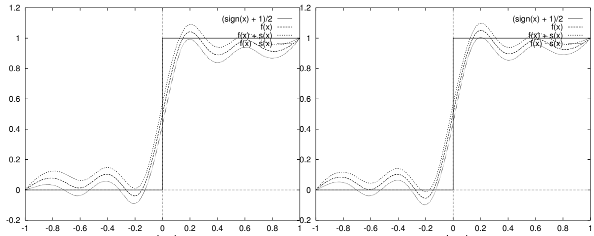
Our experiments show the average mean square error does not depend strongly on the random noise ξ_n between the nonlinear and the linear components of the Hammerstein system. For each of the 36 combinations of parameters (two nonlinearities, six data sizes, and three noise levels), we carried out 100 experiments, and we computed the pointwise averages

$$f(x) = \frac{1}{100} \sum_{j=1}^{100} \phi_n^{(j)}(x),$$

and pointwise standard deviations

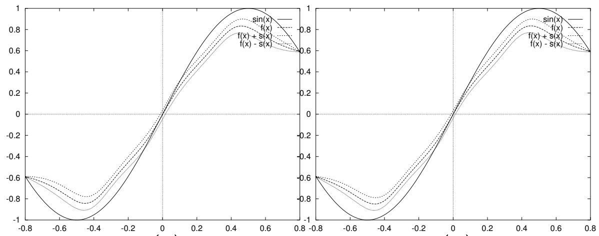
$$s(x) = \left(\frac{1}{100} \sum_{j=1}^{100} (\phi_n^{(j)}(x) - f(x))^2 \right)^{1/2}.$$

The results of the individual experiments are shown in Figure 5 and Figure 6.



(a) (b)

Fig. 5. $n = 1000$ (a) $\sigma_\xi = 0.2$. (b) $\sigma_\xi = 0.01$.



(a) (b)

Fig. 6. $n = 1000$ (a) $\sigma_\xi = 0.2$. (b) $\sigma_\xi = 0.01$.

5. COMMENTS AND CONCLUSIONS

In the paper we considered nonparametric identification of Hammerstein systems by the nonparametric Hermite series regression estimate. Identification algorithms have been proposed and their convergence and the rates investigated under very mild restrictions on the measurements and parameters. One important problem not considered in the paper is data-dependent selection of algorithms parameters. There are many powerful statistical techniques for optimizing parameters. They include cross-validation and bootstrap (see (Efron and Tibshirani, 1993) for a comprehensive discussion). Application of these techniques to adaptive selection of parameters in our algorithms is left for the future work.

REFERENCES

Bai, E. W. (2002). A blind approach to the hammerstein-wiener model identification. *Automatica* **38**, 967–979.

Chang, F.H. and R. Luus (1971). A noniterative method for identification using Hammerstein system. *IEEE Trans. Automat. Contr* **AC-16**, 464–468.

Chung, H. and Y.Y. Sun (1988). Analysis and parameter estimation of nonlinear systems with Hammerstein model using Taylor series approach. *IEEE Trans. Circuits Syst.* **CAS-35**, 1533–1544.

Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap.* Chapman and Hall. New York.

Eubank, R. (1999). *Nonparametric Regression and Spline Smoothing.* Marcel Dekker. New York.

Greblicki, W. (1989). Non-parametric orthogonal series identification of Hammerstein systems. *Int. J. Syst. Science* **20**, 2355–2367.

Greblicki, W. (2000). Continuos time Hammerstein system identification. *IEEE Trans. Automat. Contr.* **AC-45**, 1232–1236.

Greblicki, W. and A. Krzyżak (1979). Nonparametric identification of a memoryless system with cascade structure. *Int. J. Syst. Science* **10**, 1301–1310.

Greblicki, W. and M. Pawlak (1987). Hammerstein system identification by nonparametric regression estimation. *Int. J. Contr.* **45**, 343–354.

Greblicki, W. and M. Pawlak (1989). Recursive nonparametric identification of Hammerstein systems. *J. of the Franklin Institute* **326**, 461–481.

L. Györfi, M. Kohler, A. Krzyzżak and H. Walk (2002). *Distribution-free Theory on Nonparametric Regression.* Springer-Verlag. New York.

Haber, R. and H. Unbehauen (1990). Structure identification of nonlinear dynamic systems - a survey on input/output approaches. *Automatica* **26**, 651–677.

Hall, P. (1989). On convergence rates in nonparametric problems. *Int. Statist. Review* **57**, 45–58.

Hsia, T. (1977). *System Identification.* D.C. Heath and Company. Lexington, Mass.

Hunter, I. and M. Korenberg (1986). The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics* **55**, 135–144.

J.Bendat (1990). *Nonlinear System Analysis and Identification from Random Data.* Wiley. New York.

Krzyżak, A. (1989). Identification of discrete Hammerstein systems by the Fourier series regression estimate. *Int. J. Syst. Science* **20**, 1729–1744.

Krzyżak, A. (1990). On estimation of a class of nonlinear systems by the kernel regression estimate. *IEEE Trans. Inform. Theory* **IT-36**, 141–152.

Krzyżak, A. (1992). Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation. *IEEE Trans. Inform. Theory* **IT-38**, 1323–1338.

Krzyżak, A. (1996). On nonparametric estimation of nonlinear dynamic systems by the Fourier series estimate. *Signal Processing* **52**, 299–321.

Krzyżak, A. and M. Partyka (1993). Identification of block oriented systems by nonparametric techniques. *Int. J. Syst. Science* **24**, 1049–1066.

Kung, M. and B. Womack (1984). Discrete-time adaptive control of linear dynamic systems with a two-segment piecewise-linear asymmetric nonlinearity. *IEEE Trans. Automat. Contr* . **AC-29**, 170–172.

Marmarelis, P. and V. Marmarelis (1978). *Analysis of Physiological Systems: the White Noise Approach.* Plenum Press. New York.

Masry, E. and S. Cambanis (1980). Signal identification after noisy, nonlinear transformations. *IEEE Trans. Inform. Theory* **IT-26**, 50–58.

Narendra, K. and P. Gallman (1982). An iterative method for the identification of nonlinear systems using the Hammerstein model. *IEEE Trans. Automat. Contr.* **AC-11**, 546–550.

R. Emerson, M. Korenberg and M. Citron (1992). Identification of complex-cell intensive nonlinearities in a cascade model of cat visual cortex. *Biological Cybernetics* **66**, 291–300.

Sandberg, I. (1992). Approximations of nonlinear functionals. *IEEE Trans. Circuits and Systems* **CAS-39**, 65–67.

Sansone, G. (1980). *The Volterra and Wiener Theories of Nonlinear Systems.* Wiley-Interscience. New York.

Stapleton, J. and S. Baas (1992). Adaptive noise cancelation for a class of nonlinear, dynamic reference channels. *IEEE Trans. Circuits and Systems* **CAS-32**, 65–67.

T. McCannon, W. Gallagher, D. Minoo-Hamedani and G. Wise (1982). On the design of nonlinear discrete-time predictors. *IEEE Trans. Inform. Theory* **IT-28**, 366–371.