Hammerstein System Identification by the Haar Multiresolution Approximation

(revised version)

ZYGMUNT HASIEWICZ

Institute of Engineering Cybernetics
Wrocław University of Technology
Wrocław, Poland

Affiliation address:

Institute of Engineering Cybernetics,
Wrocław University of Technology,
Janiszewskiego 11/17,
50-372 Wrocław, Poland

phone: (48 71) 320 32 77;
fax: (48 71) 21 26 77;
e-mail: zhas @ ict. pwr. wroc. pl

SUMMARY

The paper deals with recovering non-linearities in the Hammerstein systems by using multiresolution approximation - a basic concept of wavelet theory. The systems are driven by random signals and are disturbed by additive, white or coloured, random noise. The *a priori* information about system components is non-parametric and a delay in the dynamical part of systems is admitted. A non-parametric identification algorithm for estimating non-linear characteristics of static parts is proposed and investigated. The algorithm is based on the Haar multiresolution approximation. The pointwise convergence and pointwise asymptotic rate of convergence of the algorithm are established. It is shown that neither the form nor the convergence conditions of the algorithm need any modification for the noise being not white but correlated. Also the asymptotic rate of convergence is the same for white and coloured noise. The theoretical results are confirmed by computer simulations.

KEY WORDS: Hammerstein system; non-linearity recovering; non-parametric identification; multiresolution approximation

1. INTRODUCTION

In this paper, we examine the non-parametric approach to identification of non-linear characteristics of Hammerstein systems. These are cascade connections of a non-linear static part and a linear dynamic element. Our approach is suited to the case when *a priori* information about system non-linearity is poor and confined to only some qualitative features of a characteristic i.e. is much smaller than that required in the standard parametric method. We propose an identification algorithm to estimate non-linearities in the systems which are driven by random signals and are disturbed by white or correlated random noise, with possible delay in the dynamical part of systems. The identification algorithm is based on the idea of multiresolution approximation, a fundamental concept of wavelet theory, and implements the Haar multiresolution approximation, in particular. We show that our algorithm is extremely simple from the computational viewpoint and successfully recovers non-linearities in the Hammerstein systems, i.e. converges in the pointwise manner to the unknown non-linear characteristic of the system at an optimal rate.

The Hammerstein systems, located in the class of block-oriented non-linear dynamical complexes (see References 1-3 for this and other block-oriented structures), are often met in industrial (chemical) engineering[4] as well as signal processing[5], image analysis[6] and biocybernetics.[7,8] Our attention is focused on estimating non-linear part of such systems for two reasons. First, specific non-linearity is the discriminating attribute of the concrete Hammerstein system at hand. Second, ambiguity of non-linear characteristics under poor preliminary knowledge - when in particular no finite parametrization of a possible non-linearity can be reasonably motivated - makes the problem of non-linearity recovering much more pivotal than identification of a linear dynamic part - with standard form of parametrization.

The non-parametric approach to Hammerstein system identification has been originated in Reference 9 and next developed in References 10-14, where the kernel-type identification algorithms for recovering non-linear characteristic have been elaborated. In turn, in References 15-17, the orthogonal series expansion approach has been used to recover non-linearity in the Hammerstein systems. In these papers the authors applied standard kinds of estimates and assumed that the noise disturbing the system is white. In the present contribution, we admit that the system noise can be not only white but also correlated and in order to identify Hammerstein system non-linearity we apply the modern multiresolution approach

which provides a useful technique for obtaining different levels of accuracy of the characteristic estimate. Additionally, a delay in the system dynamics is allowed in our considerations.

It is well-known that in the case of conventional parametric methods, identification of systems in the presence of correlated noise is generally much more difficult than in the white noise environment, both in theoretical and practical aspects. The standard parametric identification algorithms (e.g. least squares) need then essential modifications, depending moreover on the particular correlation structure of the noise, which results in rather computationally demanding methods of local applicability.[18-20] In contradistinction to this, we shall show that our non-parametric algorithm is insensitive to the correlation of noise. It will be evidenced that no revision of the algorithm is required to ensure convergence when coloured noise instead of white noise is influencing the system, and that both the convergence conditions and asymptotic rate of convergence remain the same for white and coloured noise.

As regards the multiresolution theory applied in the paper, it has found recently applications in a remarkable diversity of disciplines as, e.g., signal processing, image analysis, neural networks, approximation theory or statistics (see References 21-24). For applications in system modeling we refer the reader to the survey papers[25,26] and the papers cited therein. A comprehensive review of this methodology is given in the monographs.[27-32]

The paper is organized as follows. In Section 2, the Hammerstein system is characterized and basic assumptions are collected providing the framework for the identification task to be solved. Next, in Section 3, a brief introduction to the concept of multiresolution approximation is given and the Haar multiresolution approximation is outlined, which yields the background for the identification algorithm developed in the paper. The non-parametric algorithm for recovering non-linearity in the Hammerstein system, of the same structure for white and coloured noise, is then derived in Section 4 and convenient computer procedures for calculating the estimate from input-output measurements are presented there. Convergence and asymptotic rate of convergence of the algorithm are in turn examined in Sections 5 and 6. It is established that convergence conditions do not change when the noise becomes correlated and that the asymptotic rate of convergence is not worsen by the fact that the noise is not white but correlated. Section 7 presents the results of computer simulations to illustrate behaviour of the estimate for finite number of measurements in the case of white and coloured noise. Section 8 concludes the paper.

2. THE HAMMERSTEIN SYSTEM

We shall consider a Hammerstein system shown in Fig. 1. This is a tandem connection of a non-linear memoryless element followed by a linear output dynamics. The non-linear memoryless component has a characteristic R and the linear dynamic part is a discrete-time time-invariant

< *Fig.* 1 >

and asymptotically stable element operating in steady state, with the unknown impulse response $\{\lambda_p; p = 0, 1, \ldots\}$ where $\sum_{p=0}^{\infty} |\lambda_p| < \infty$. We assume that the system input $\{x_k; k = \ldots, -1, 0, 1, 2, \ldots\}$ is a stationary white random process with finite variance. There exists a probability density of x_k, denoted by f, and it is assumed that $\int_{-\infty}^{+\infty} f^2(x)\,dx < \infty$ and $sup_x f(x) = M_f < \infty$. The system output is disturbed by stationary random noise $\{z_k; k = \ldots, -1, 0, 1, 2, \ldots\}$. The noise is by assumption

(*a*) *white*, with zero mean, $Ez_k = 0$, and finite variance, $\operatorname{var} z_k < \infty$, (Fig. 1*a*), or

(*b*) *coloured* - obtained from a stationary white noise $\{\varepsilon_k; k = \ldots, -1, 0, 1, 2, \ldots\}$, with zero mean and finite variance, as an output of a discrete-time time-invariant and asymptotically stable linear filter, with the impulse response $\{\omega_p; p = 0, 1, \ldots\}$, operating in steady state (Fig. 1*b*), i.e.

$$z_k = \sum_{p=0}^{\infty} \omega_p \varepsilon_{k-p} \tag{2.1}$$

where $E\varepsilon_k = 0$, $\operatorname{var} \varepsilon_k < \infty$ and $\sum_{p=0}^{\infty} |\omega_p| < \infty$.

The external processes $\{x_k\}$ and $\{z_k\}$ ($\{x_k\}$ and $\{\varepsilon_k\}$) are mutually independent. We assume that one can measure the input x_k and the output y_k of the overall system however the internal signal $w_k = R(x_k)$, interconnecting both parts of the system, is not accessible for measurements. The overall system output y_k is governed by the equation

$$y_k = \sum_{p=0}^{\infty} \lambda_p R(x_{k-p}) + z_k \qquad (= v_k + z_k) \tag{2.2}$$

If $\lambda_0 = \lambda_1 = \ldots = \lambda_{d-1} = 0$ and $\lambda_d \neq 0$ then a d-step delay is present in the system.

The non-linear characteristic R of the system is unknown and it is only assumed that

$$|R(x)| \leq a_1 |x| + a_2 \tag{2.3}$$

for some positive a_1 and a_2. Our aim is to recover the non-linearity R from input-output observations $\{(x_k, y_k)\}$ of the whole system.

Let us shortly discuss some of the theoretical and practical issues associated with the above problem and the imposed assumptions. First, let us notice that inaccessibility of the internal signal w_k in Fig. 1 (generally assumed for block-oriented Hammerstein systems; see References 1-3) makes identification of non-linearity R implicit in the sense that not both the input x_k and output w_k of the static element to be identified can be directly measured. From the 'aggregated' input-output data $\{(x_k, y_k)\}$ the individual system components (i.e. memoryless non-linearity and linear dynamics) cannot be identified in a traditional fashion, i.e. each one separately - as from their own input-output measurements. Second, let us observe that output measurements y_k are correlated random variables (for coloured as well as white noise) as the outputs of a linear dynamical element (see equation (2.2)). This fact distinguishes our problem from purely memoryless system identification task (i.e. a similar task for fictitious Hammerstein system with non-linearity R, $\lambda_0 = 1$, $\lambda_p = 0$ for $p \geq 1$ and independent input and output observations) and complicates the proof of consistency of the estimate derived in Section 4 (see Appendix A).

As regards the assumptions concerning system excitations and system dynamics, they are standard and do not need comments. We want only to emphasize that (*i*) the probability density of the input signal can be in our considerations an arbitrary square integrable bounded density function, which is met by a broad class of densities, and that (*ii*) the system noise $\{z_k\}$ may be white as well as correlated, which is in contrast with earlier papers concerning Hammerstein system identification - as mentioned in Section 1. Crucial for the problem is that the underlying *a priori* knowledge about non-linearity R is poor. We merely assume that the non-linear characteristic grows (in absolute value) at most linearly (condition (2.3)). In result, the class of admissible system characteristics is very wide (including discontinuous functions) and cannot be parameterized, i.e., its members cannot be expressed by using a concise finite-dimensional parametric representation. This causes that the problem of recovering R is non-parametric and needs an adequate non-parametric approach to the solution. In most papers devoted to recovering Hammerstein system characteristics from noisy data the underlying *a priori* information is much greater than that assumed in (2.3). Typical assumption is that R is a polynomial of known order and the (parametric) identification problem is just to estimate a finite number of polynomial coefficients (see References 1, 3, 4, 8, 18, 35 and the references therein for appropriate parametric identification methods). However, smooth polynomial model of non-linearity is not always well-grounded. Assuming small *a priori* information about the system and next considering the identification problem as a non-parametric (infinite-dimensional) one is, in our belief, more realistic from the viewpoint of applications.

The non-parametric approach to recover non-linearity in the Hammerstein system has been initiated and next developed in a series of papers quoted in Section 1. For a full account of the conventional theory of non-parametric curve estimation methods from independent input-output data see, e.g., References 36-38.

3. THE HAAR MULTIRESOLUTION APPROXIMATION

In order to make the paper self-contained and to improve its readability, we present in this section a short review of the basic concepts of the multiresolution theory and the Haar multiresolution approximation, in particular. This provides the theoretical background for the identification algorithm developed in the next section. Detailed treatment of this theory can be found in References 27-31 or more recent monograph.[32]

Let $\phi(x)$ be a real function whose translates $\{ \phi(x-n) \}$, $n \in \mathrm{Z}$, the set of integers, are orthonormal in $L^2(\mathbb{R})$ and form an orthonormal basis of a subspace V_0 of $L^2(\mathbb{R})$. Let moreover $\phi_{mn}(x) = 2^{m/2}\phi(2^m x - n)$ be scaled and translated versions of $\phi(x)$ and let $V_m = \text{span}\,\{\phi_{mn}(x), n \in \mathrm{Z}\} \subset L^2(\mathbb{R})$ be the 'dilation space' with resolution $1/2^m$ associated with V_0 (with orthonormal basis $\{\phi_{mn}(x), n \in \mathrm{Z}\}$). Suppose that for $m \in \mathrm{Z}$ the V_m's form a nested chain of subspaces of $L^2(\mathbb{R})$:

$$\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots \subset V_m \subset \cdots \subset L^2(\mathbb{R}) \tag{3.1}$$

and that

$$\bigcap_m V_m = \{0\}, \qquad \overline{\bigcup_m V_m} = L^2(\mathbb{R}) \tag{3.2}$$

or equivalently

$$\lim_{m \to -\infty} V_m = \{0\}, \qquad \overline{\lim_{m \to \infty} V_m} = L^2(\mathbb{R})$$

Such a function $\phi(x)$ is called a scaling function and the sequence of subspaces $\{ V_m \}$ constitutes a multiresolution approximation of $L^2(\mathbb{R})$.[27-34] Observe that the basis functions $\phi_{mn}(x)$ of the resolution spaces V_m are generated in an easy and automatic way from a single initial function $\phi(x)$ by only scaling (the scale factor m) and shifting (the translation factor n). We want to emphasize this fundamental and useful property of the multiresolution approximation.

Any function $F(x) \in L^2(\mathbb{R})$ can be approximated in the resolution space V_m as follows

$$F(x;m) = \sum_{n=-\infty}^{\infty} \alpha_{mn}\, \varphi_{mn}(x) \tag{3.3}$$

where the coefficients α_{mn} are given by

$$\alpha_{mn} = \int_{-\infty}^{\infty} F(x)\, \varphi_{mn}(x)\, dx. \tag{3.4}$$

The approximation $F(x;m)$ is the orthogonal projection of $F(x)$ on V_m and the appropriate L^2-approximation error (at resolution $1/2^m$) is

$$e_m^2 = \int_{-\infty}^{\infty} \big(F(x) - F(x;m)\big)^2 dx = \int_{-\infty}^{\infty} F^2(x)\, dx - \sum_{n=-\infty}^{\infty} \alpha_{mn}^2$$

Asymptotically, the error $e_m^2 \to 0$ as $m \to \infty$ by the properties of the multiresolution approximation. For finite m, the error value e_m^2 depends on the particular features of both $F(x)$ and the scaling function $\phi(x)$ which has been investigated, e.g., in References 22, 24 and 32. In this paper, we shall examine the pointwise properties of the approximations as in (3.3) in a stochastic environment (Sections 4-6).

The above requirements, in particular the basic conditions (3.1) and (3.2), hold for the scaling function of the form

$$\varphi_H(x) = I_{[0,1)}(x), \qquad x \in (-\infty, \infty) \tag{3.5}$$

where $I_{[a,b)}(x)$ is the indicator function of [a,b), i.e., we have $I_{[a,b)}(x) = 1$ if $x \in$ [a,b) and 0 otherwise. Obviously, the corresponding sequence $\{\phi_H(x-n)\}$, $n \in \mathbb{Z}$, is orthonormal in $L^2(\mathbb{R})$ and the root space V_0 consists of all piecewise constant functions in $L^2(\mathbb{R})$ with possible jumps at the integers. In this case, the resolution spaces V_m for each scale $m \in \mathbb{Z}$ are spanned by

$$\varphi_{H,mn}(x) = 2^{m/2}\, \varphi_H\big(2^m (x - \frac{n}{2^m})\big) = 2^{m/2}\, I_{[\frac{n}{2^m}, \frac{n+1}{2^m})}(x), \qquad n \in Z \tag{3.6}$$

and contain piecewise constant functions with jumps at dyadic grid points $n/2^m$, $n \in \mathbb{Z}$. They represent in the ladder (3.1) successive 'resolution levels', associated with different scales m, and (3.2) says in particular that any function in $L^2(\mathbb{R})$ can be approximated with arbitrary precision by using the step functions $\phi_{H,mn}(x)$ in (3.6). The V_m's defined above (associated with the scaling function (3.5)) form the so-called *Haar multiresolution approximation* of $L^2(\mathbb{R})$ which will be fundamental for our paper. We note that $\phi_{H,mn}(x)$ of (3.6) are supported in the intervals $[n/2^m, (n+1)/2^m)$ and that for each scale m and each point $x \in [n/2^m, (n+1)/2^m)$ the Haar approximation in V_m of a square integrable function $F(x)$ takes the form (see (3.3)-(3.4))

$$F(x;m) = 2^{m/2}\alpha_{H,mn} \tag{3.7}$$

where

$$\alpha_{H,mn} = 2^{m/2}\int_{n/2^m}^{(n+1)/2^m} F(x)dx \tag{3.8}$$

(since two Haar basis functions $\phi_{H,mn}(x)$ of the same scale m do not overlap), i.e. $F(x;m)$ is simply a *mean value* of $F(x)$ in the interval $[n/2^m, (n+1)/2^m)$ of the length $1/2^m$.

4. IDENTIFICATION ALGORITHM

We are now in a position to develop the identification algorithm for recovering non-linearity of the Hammerstein system and discuss some of the practical aspects associated with its application. We begin our considerations with the observation that, taking account of equation (2.2), the Hammerstein system output y_k can be expressed as follows

$$y_k = \lambda_d R(x_{k-d}) + \sum_{\substack{p=0 \\ p\neq d}}^{\infty} \lambda_p R(x_{k-p}) + z_k$$

or equivalently

$$y_k = [cR(x_{k-d}) + e] + \sum_{\substack{p=0 \\ p\neq d}}^{\infty} \lambda_p [R(x_{k-p}) - ER(x_{k-p})] + z_k$$

where $c = \lambda_d$ and $e = ER(x_k)\sum_{p\neq d}\lambda_p$ and where we have used stationarity of the process $\{x_k\}$ (see Section 2).

The key for derivation of the identification algorithm is the recognition that:

1) under assumptions of Section 2, for whichever white or coloured noise, it holds

$$E\{y_k \mid x_{k-d} = x\} = cR(x) + e, \tag{4.1}$$

2) for each point x such that $f(x) > 0$ the right hand side of (4.1) can be factorized as follows

$$cR(x) + e = g(x)/f(x), \tag{4.2}$$

where $g(x) = E\{y_k \mid x_{k-d} = x\} f(x)$, and where f is the probability density of x_k (see Section 2).

The equation (4.1) reveals that the unknown non-linearity R (up to a scale and translation constants c and e) is - for white and coloured noise - a regression function (a conditional expectation of output y on shifted input x) and hence estimating a regression in (4.1) is equivalent to estimating (scaled and

translated) non-linearity of the Hammerstein system. Such a property, for $d = 0$, was originally observed by Greblicki and Pawlak in Reference 9. At present, d can be an arbitrary shift factor such that $\lambda_d \neq 0$. In particular, it is admitted that in equation (2.2) we have $\lambda_p = 0$ for $p < d$, i.e., a d-step delay is allowed in the system, without any revision of the formulas (4.1) and (4.2). Moreover, we let in Section 2 the noise $\{z_k\}$ be correlated. The ratio $g(x)/f(x)$ as in (4.2) has been previously applied in statistical as well as engineering literature.[15,16,17,43,44]

Since in the case under consideration $g(x)$ $(= (cR(x) + e)f(x))$ and $f(x)$ are square integrable functions, $\int_{-\infty}^{+\infty} g^2(x)\,dx < \infty$ and $\int_{-\infty}^{+\infty} f^2(x)\,dx < \infty$ (see assumptions in Section 2), the numerator $g(x)$ and denominator $f(x)$ of (4.2) may be approximated in the Haar resolution space V_m by the series (see (3.3)-(3.4) in Section 3)

$$g(x;m) = \sum_{n=-\infty}^{\infty} a_{mn}\varphi_{H,mn}(x) \quad \text{and} \quad f(x;m) = \sum_{n=-\infty}^{\infty} b_{mn}\varphi_{H,mn}(x),$$

where

$$a_{mn} = E\{y_d\,\varphi_{H,mn}(x_0)\} \quad \text{and} \quad b_{mn} = E\,\varphi_{H,mn}(x_0).$$

The formulas on the weighting coefficients a_{mn} and b_{mn}, being now simple expectations, follow immediately from the definition (3.4), definitions of $g(x)$ and $f(x)$ and stationarity of the processes $\{x_k\}$ and $\{y_k\}$, and are the same for white and coloured noise $\{z_k\}$.

The above two expressions and (4.2) lead together to the following natural m-scale estimate $R_N(x;m)$ of $cR(x) + e$:

$$R_N(x;m) = \frac{g_N(x;m)}{f_N(x;m)} = \frac{\sum_{n=-\infty}^{\infty} a_{mn,N}\,\varphi_{H,mn}(x)}{\sum_{n=-\infty}^{\infty} b_{mn,N}\,\varphi_{H,mn}(x)} \tag{4.3}$$

where $a_{mn,N}$ and $b_{mn,N}$ (estimates of a_{mn}'s and b_{mn}'s) are in (4.3) computed from N (random) observations $\{(x_k, y_{k+d});\, k = 1, 2, \ldots, N\}$ of the whole system input and output as usual sample means:

$$a_{mn,N} = \frac{1}{N}\sum_{k=1}^{N} y_{k+d}\,\varphi_{H,mn}(x_k)\;; \qquad b_{mn,N} = \frac{1}{N}\sum_{k=1}^{N} \varphi_{H,mn}(x_k) \tag{4.4}$$

Owing to the specific form of the Haar basis functions $\phi_{H,mn}(x)$ (see (3.6)), for each point $x \in [n/2^m, (n+1)/2^m)$ the algorithm (4.3)-(4.4) yields

$$R_N(x;m) = \frac{a_{mn,N}}{b_{mn,N}} \tag{4.3a}$$

where (we omit below the common factor $2^{m/2}/N$)

$$a_{mn,N} = \sum_{\{k:\, x_k \in [\frac{n}{2^m}, \frac{n+1}{2^m})\}} y_{k+d}\ ; \qquad b_{mn,N} = \#\{\, x_k \in [\frac{n}{2^m}, \frac{n+1}{2^m})\,\} \tag{4.4a}$$

or equivalently

$$a_{mn,N} = \sum_{\{k:\, u_k \in [0,1)\}} y_{k+d}\ ; \qquad b_{mn,N} = \#\{\, u_k \in [0,1)\,\} \tag{4.4b}$$

where $u_k = 2^m x_k - n$ and # denotes the cardinality of a collection. The denominator of the estimate (4.3a) counts the number of measurements x_k in the interval $[n/2^m, (n+1)/2^m)$ containing the point x at which the estimation is carried out (changing during the recovery process). The numerator selects and sums up the corresponding, d steps ahead shifted, output measurements y_{k+d} including thereby the possible d-step delay in the system. The ratio (4.3a) gives eventually the *sample mean* of (shifted) output measurements y_{k+d} for x_k in the respective interval of the length $1/2^m$ as the estimate of $cR(x) + e$ at x (compare (3.7)-(3.8) in Section 3). The scale factor m (precisely: the resolution level $1/2^m$) controls the size of the neighbourhood around the point x in which averaging is made and changes sensitivity of the estimate to the details in the run of $cR(x) + e$.

For the use of the above algorithm, practically only a finite number of coefficients $a_{mn,N}$ and $b_{mn,N}$ must be calculated from experimental data, and stored in a computer memory, for each scale m and all computations can be done automatically and in short time. If, for instance, all input observations x_k are contained in the interval $[x_{\min}, x_{\max}]$ then for $x \in [x_{\min}, x_{\max}]$ and a given scale m the translation factor n varies in the estimate (4.3)-(4.4) over the range of integers $[2^m x_{\min}] \le n \le [2^m x_{\max}]$ and the total number of pairs $(a_{mn,N}, b_{mn,N})$ which should be computed and kept in a memory does not exceed $2^m[x_{\max} - x_{\min}] + 2$, where $[v]$ stands for the integer part of v.

In the *on-line* application, convenient recursive procedures for updating the coefficients $a_{mn,N}$ and $b_{mn,N}$ can be used:

$$a_{mn,N} = a_{mn,N-1} + y_{N+d}\, I_{[0,1)}(u_N),$$

$$b_{mn,N} = b_{mn,N-1} + I_{[0,1)}(u_N)$$

where the initial conditions are $a_{mn,0} = 0$, $b_{mn,0} = 0$ and $u_N = 2^m x_N - n$.

We point out that the form of the algorithm is the same for white and each coloured noise. The computations needed by the algorithm do not thus depend on any specific correlation structure of the disturbances corrupting the system. This is in contrast to parametric methods, where correlation of the noise requires generally a substantial modification of the estimation routine and results in far more demanding computer procedures than for the white noise case.[18,20] We accentuate this property as an important advantage of the algorithm extending the range of its applicability (see also Theorem 1 and Theorem 2 in Sections 5 and 6). Some weaker point is that we can only estimate the Hammerstein system non-linearity with accuracy to the (unknown) scaling and translation constants c and e (i.e. we may factually discover only the 'sketchy' shape $cR(x) + e$ of non-linearity $R(x)$). Moreover, the computations as in (4.4a), (4.4b) must be repeated for each individual sub-interval $[n/2^m, (n+1)/2^m)$ of the resolution grid, contained in the estimation region. The first weakness is however not a drawback of the algorithm in itself but an inherent feature of the problem under study, independent of the applied identification method. It follows as an unavoidable consequence of a cascade complex structure of the system and assembling character of the data $\{(x_k, y_{k+d})\}$ used for identification (inaccessibility of the internal signal w_k and linear filtering of the non-linearity output before measurement; see Section 2). Recalculation of the estimate is, in turn, not an excessive computational problem because of simplicity of computations required by the algorithm.

In the following, for the ease of analysis, we shall assume that $ER(x_k) = 0$. Then $e = 0$ and $cR(x) + e$ is reduced to $cR(x)$. Consequently, the algorithm (4.3)-(4.4) ((4.3a)-(4.4a)) provides the estimate of a scaled non-linearity $cR(x)$, where $c = \lambda_d$ is an unknown constant. Such assumption is for instance fulfilled if the non-linearity R is an odd function and the input probability density f is symmetric, i.e.

$$R(x) = -R(-x), \qquad x \in (-\infty, \infty)$$

and

$$f(x) = f(-x), \qquad x \in (-\infty, \infty)$$

In such a case, if the value of a non-linear characteristc R is *a priori* known in an arbitrary point (x_0 say) such that $R(x_0) \neq 0$, the ratio

$$c_N = \lambda_{d,N} = R_N(x_0; m) \,/\, R(x_0)$$

may be used as an estimate of the constant c and further

$$\big(R(x_0) \,/\, R_N(x_0; m)\big) R_N(x; m)$$

can be applied as the estimate of the 'exact' non-linearity $R(x)$, provided that $R_N(x_0;m) \neq 0$. Obviously, $R_N(x;m)$ should be then computed according to (4.3)-(4.4).

In Sections 5 and 6 we shall show that in addition to easiness of implementation, the algorithm has good convergence properties and satisfactory rate of convergence which compete with those of the classical non-parametric orthogonal series algorithms worked out earlier.

5. CONVERGENCE OF THE ALGORITHM

Let us consider convergence of the algorithm (4.3)-(4.4). Supposing that $ER(x) = 0$, we shall show that the estimate $R_N(x;m)$ tends, in a pointwise manner and under the same conditions for white and coloured noise, to the unknown non-linearity $cR(x)$ of the Hammerstein system as the number N of measurement data grows large. Such a property holds if the scale m in the estimate is fitted to the number of data and appropriately grows with growing N.

Crucial for the convergence is that the numerator $g_N(x;m)$ and the denominator $f_N(x;m)$ of the estimate (4.3)-(4.4), computed from random data $\{(x_k,y_{k+d})\}_{k=1}^N$ and playing the role of estimates of $g(x)$ and $f(x)$ in (4.2), possess the property that

$$\text{bias}\{\, g_N(x;m)\,\} = E\, g_N(x;m) - g(x) \to 0 \quad \text{as} \quad m \to \infty \tag{5.1}$$

$$\text{bias}\{\, f_N(x;m)\,\} = E\, f_N(x;m) - f(x) \to 0 \quad \text{as} \quad m \to \infty \tag{5.2}$$

for each N and each point $x \in (-\infty,\infty)$ at which $g(x)$, respectively $f(x)$, is continuous, and that

$$\text{var}\{\, g_N(x;m)\,\} = O\left(\frac{2^m}{N}\right) \to 0 \quad \text{as} \quad N \to \infty \tag{5.3}$$

$$\text{var}\{\, f_N(x;m)\,\} = O\left(\frac{2^m}{N}\right) \to 0 \quad \text{as} \quad N \to \infty \tag{5.4}$$

for each point x and fixed scale m. Proof of the convergence (5.1) and (5.2) and derivation of the variance bounds in (5.3)-(5.4), valid for large values of m and similar for white and coloured noise, is given in Appendix A. Setting together (5.1), (5.3) and (5.2), (5.4), we conlude that if the scale factor m in the estimate (4.3)-(4.4) is matched to the number N of data in the sample $\{(x_k,y_{k+d})\}_{k=1}^N$, i.e. $m = m(N)$, and increases with N in such a way that

$$m(N) \overset{N}{\to} \infty, \qquad \frac{2^{m(N)}}{N} \overset{N}{\to} 0 \tag{5.5}$$

as $N \to \infty$ then the mean square estimation errors

$$E\left(g_N(x;m) - g(x)\right)^2 = \text{bias}^2\{g_N(x;m)\} + \text{var}\{g_N(x;m)\} \tag{5.6}$$

$$E\left(f_N(x;m) - f(x)\right)^2 = \text{bias}^2\{f_N(x;m)\} + \text{var}\{f_N(x;m)\} \tag{5.7}$$

asymptotically vanish, i.e. we have the convergence

$$E\left(g_N(x;m) - g(x)\right)^2 \to 0 \quad \text{as} \quad N \to \infty \tag{5.8}$$

and

$$E\left(f_N(x;m) - f(x)\right)^2 \to 0 \quad \text{as} \quad N \to \infty \tag{5.9}$$

These limit properties hold at all points x around which $g(x)$ and $f(x)$ are continuous functions, equally for white and coloured noise. The above facts lead to the following theorem.

Theorem 1

Assume that the conditions of Section 2 are satisfied. Let $ER(x) = 0$, and let the scale parameter $m = m(N)$ fulfil the conditions (5.5). Then for white as well as coloured noise the estimate (4.3)-(4.4) converges and we have

$$R_N(x;m) \overset{N}{\to} c\,R(x) \qquad \textit{in probability}$$

as $N \to \infty$ at every point $x \in (-\infty,\infty)$ at which $f(x) > 0$ and both functions $R(x)$ and $f(x)$ are continuous.

Proof: Convergence in probability of the ratio estimate $R_N(x;m) = g_N(x;m)/f_N(x;m)$ follows as a straightforward conclusion from the convergence (5.8) and (5.9) and the relation (4.2). □

The theorem establishes pointwise convergence of the identification algorithm under very mild conditions on the input density $f(x)$ and the unknown non-linearity $R(x)$, quoted in Section 2. Such requirements are met by a broad class of input densities f and non-linearities R, as it was commented earlier. Additional restriction that $ER(x) = 0$ is not essential and is taken in the paper only in order to simplify the derivations in Appendix A. Therefore, from the practical viewpoint, the only important demand is fulfilment of the condition (5.5) imposed on the scale m in the estimate $R_N(x;m)$ (which controls sensitivity of the estimate to the details in the graph of a non-linear characteristic). We notice that according to (5.5) the scale factor $m = m(N)$ should grow with N slower than $\lceil \log_2 N \rceil$. Then $R_N(x;m)$ converges (in probability) in a pointwise manner to the scaled non-linearity $cR(x)$ of the

Hammerstein system generating the data set, as $N \to \infty$. The convergence takes place at the points where the input density does not vanish and the characteristic R and probability density f of the exciting input signal are simultaneously continuous functions. Such domain of convergence coincides with the almost everywhere convergence of classical trigonometric or Hermite orthogonal series regression estimates[15], however our algorithm is much simpler. We emphasize the property that convergence conditions do not need any modification for correlated noise.

6. RATE OF CONVERGENCE

Efficiency of the algorithm depends on the convergence rate. Thus, an important question is how fast does $R_N(x;m)$ tend to $cR(x)$ for each given point x as the number of data grows large, and how fast must the scale $m = m(N)$ increase (cf. condition (5.5)) in order to ensure the best rate of convergence. Obviously, it is essential for practice to answer these questions for a possibly wide class of non-linear characteristics R and input densities f. In what follows we shall examine asymptotic rate of convergence of the estimate $R_N(x;m)$ assuming only (beside the conditions of Theorem 1) that non-linearity $R(x)$ and input density $f(x)$ of the Hammerstein system are locally, i.e. around point x, Lipschitz functions (which will be denoted by $R, f \in Lip(x)$), i.e.

$$|R(x) - R(v)| \le L_R |x - v| \tag{6.1}$$

and

$$|f(x) - f(v)| \le L_f |x - v| \tag{6.2}$$

for $|x - v| < h$, where L_R, L_f are some positive constants and h determines a small neighbourhood around x, and that in addition to $sup_x f(x) = M_f < \infty$ (see Section 2) also $sup_x |R(x)| = M_R < \infty$. Then, as it can easily be verified, the function $g(x) = cR(x)f(x)$ is also locally Lipschitz, $g \in Lip(x)$, i.e. in a small neighbourhood of x we have

$$|g(x) - g(v)| \le L_g |x - v| \tag{6.3}$$

where the Lipschitz constant is $L_g = |c|(M_f L_R + M_R L_f)$. We turn attention to the fact that in our considerations $R(x)$ and $f(x)$ do not need to be continuous functions on the entire real line.

In the above case, the bias expressions in (5.1)-(5.2) can be specified and, as is shown in Appendix B, asymptotically (i.e. for the scale m growing large; cf. condition (5.5)) we obtain

$$|\,\mathrm{bias}\{g_N(x;m)\}\,| \;=\; |\,E\,g_N(x;m) - g(x)\,| \;=\; O\Big(\frac{1}{2^m}\Big) \tag{6.4}$$

and

$$|\,\mathrm{bias}\{f_N(x;m)\}\,| \;=\; |\,E\,f_N(x;m) - f(x)\,| \;=\; O\Big(\frac{1}{2^m}\Big) \tag{6.5}$$

The class of functions satisfying the conditions (6.1) and (6.2) covers, among others, locally constant non-linearities and input densities and smooth (locally) differentiable functions $R(x)$ and $f(x)$. It is worth of noticing that for smoother, p times differentiable functions $R(x)$ and $f(x)$ around the point x ($p \geq 1$), the asymptotic bias error is not improved and for each p the error is of the same order as in (6.4)-(6.5) (see Appendix B). This insensitivity of the bias errors to the smoothness of $R(x)$ and $f(x)$ is an obvious consequence of the fact that $g_N(x;m)$ and $f_N(x;m)$ in the estimate (4.3)-(4.4) are rough piecewise constant approximations of $g(x)$ and $f(x)$.

For the above-mentioned class of non-linear characteristics $R(x)$ and input probability densities $f(x)$, owing to (6.4)-(6.5) and (5.3)-(5.4) in Section 5, we get the following asymptotic expressions for the pointwise mean square errors (5.6)-(5.7) of the factors $g_N(x;m)$ and $f_N(x;m)$:

$$E\big(g_N(x;m) - g(x)\big)^2 \;=\; O\Big(\frac{1}{2^{2m}}\Big) \;+\; O\Big(\frac{2^m}{N}\Big) \tag{6.6}$$

and

$$E\big(f_N(x;m) - f(x)\big)^2 \;=\; O\Big(\frac{1}{2^{2m}}\Big) \;+\; O\Big(\frac{2^m}{N}\Big) \tag{6.7}$$

The error bounds in these expressions (identical for white and coloured noise) are minimized, for large values of N, if the scale factor $m = m(N)$ is selected according to the rule

$$m_{opt}(N) \;=\; \Big[\frac{1}{3}\log_2 N\Big] \tag{6.8}$$

where [ ] denotes the integer part of a number. This follows immediately from direct minimization of the right hand side of (6.6)-(6.7). The rule (6.8) obviously satisfies the condition (5.5) in Section 5 (see also the associated remarks in the end of Section 5). For large number N of data and the scale selected as in (6.8) the (optimum) estimation errors (6.6)-(6.7) are of order

$$E\big(g_N(x;m) - g(x)\big)^2 \;=\; O\big(N^{-2/3}\big) \tag{6.9}$$

$$E\big(f_N(x;m) - f(x)\big)^2 \;=\; O\big(N^{-2/3}\big) \tag{6.10}$$

These ascertainments yield the following theorem establishing local asymptotic rate of convergence of

our ratio estimate $R_N(x;m) = g_N(x;m)/f_N(x;m)$.

Theorem 2

Let the conditions of Theorem 1 be satisfied. Let $R, f \in Lip(x)$ and let the scale parameter $m = m(N)$ be optimally selected according to the rule (6.8). Then asymptotically, for the relative estimation error at the point x, $|R_N(x;m) - cR(x)| / |cR(x)|$, it holds

$$P\{|R_N(x;m) - cR(x)| > \varepsilon |cR(x)|\} = O(N^{-2/3}) \tag{6.11}$$

for each $\epsilon > 0$, and the estimate $R_N(x;m)$ tends to $cR(x)$ as fast as

$$|R_N(x;m) - cR(x)| = O(N^{-1/3}) \quad \textit{in probability} \tag{6.12}$$

without any distinction for white and coloured noise.

Proof: See Appendix B.

In the above theorem, for a sequence of random variables $\{\zeta_N\}$ by $\zeta_N = O(a_N)$ *in probability* we mean that $d_N\zeta_N / a_N \to 0$ *in probability* as $N \to \infty$ for any number sequence $\{d_N\}$ convergent to zero (see, e.g., Reference 39, p. 140).

We point out the fact that the rate of convergence does not depend on correlation of the noise affecting the system. As it follows from inspection of Appendix A and the proof in Appendix B, the asymptotic speed of convergence is not worsen by the fact that the noise is not white and is insensitive to any particular correlation structure of the noise.

The guaranteed speed $O(N^{-1/3})$ at which the estimate $R_N(x;m)$ converges is worse than the typical convergence rate $O(N^{-1/2})$ in probability for parametric methods (cf., e.g., Reference 40; chapter 4.4). In our approach however uncertainty in the system characteristic is incomparably greater than in the standard parametric estimation. Our non-parametric estimate is prepared for implementation when the *a priori* knowledge of the system to be identified is indeed very small. For the non-parametric curve fitting methods the rate of convergence is generally of slower order and for the convergence in probability it holds $O(N^{-r})$, $0 < r < 1/2$ (cf. Reference 38). Our $O(N^{-1/3})$ rate agrees with the best possible non-parametric rate of convergence (in probability) for the Lipschitz functions (Reference 41). The convergence rate remains however the same also for smoother differentiable functions (see the remarks concerning the bias errors (6.4) and (6.5) and suitable derivations in Appendix B). This is a weaker point of the algorithm, following from the fact that the estimate $R_N(x;m)$ is only a piecewise constant

approximation of $cR(x)$ (see Section 4). It can be seen as a price paid for its computational simplicity. To get the picture, the attainable rate of convergence in probability of a non-parametric estimate for p times differentiable non-linear characteristic and independent input-output data, i.e. for a static system, is (as established in Reference 41) of order $O(N^{-p/(2p+1)})$. For $p > 1$ this yields better than $O(N^{-1/3})$ rate of convergence. The pointwise rate of convergence of other non-parametric estimates applied for the Hammerstein system in the case comparable with ours is however not faster. For example, employing the classical Hermite or trigonometric orthogonal series expansions to the numerator and denominator of (4.2) yields the estimates which for the class of non-linear characteristics and input densities with one derivative, and for the white noise, converge to a non-linear characteristic at a rate $O(N^{-1/4})$ in probability, as it was shown in Reference 15. The estimation algorithm is then much more computationally demanding than our. The normalized (relative) estimation error in (6.11) tends to zero in probability as fast as $O(N^{-2/3})$.

7. NUMERICAL EXAMPLE

Numerical experiment has been performed in order to illustrate the behaviour of the estimate $R_N(x;m)$ for finite number of measurements in the sample $\{(x_k, y_{k+d})\}_{k=1}^{N}$, $N < \infty$. In the experiment, the exciting input signal $\{x_k; k = ..., -1, 0, 1, 2, ...\}$ and the white noise process $\{\varepsilon_k; k = ..., -1, 0, 1, 2, ...\}$ (see Fig. 1) were distributed uniformly, over the intervals [-0.5,0.5] and [-l,l] respectively, yielding $E\varepsilon_k = 0$, see Section 2, and the noise dispersion σ_ϵ ($= l/\sqrt{3}$) depending on l in the latter case. The example non-linear characteristics $R(x)$ have been selected as:

Function 1: arctan non-linearity

$$R(x) = \arctan(6x), \qquad x \in [-0.5, 0.5]$$

Function 2: quantizer non-linearity

$$R(x) = 0.2\left[10(x+0.05)\right], \qquad x \in [-0.5, 0.5]$$

Function 1 represents smooth differentiable characteristics. Function 2, piecewise constant with jumps at the points $x_j^0 = \pm 0.05j$, $j = 1,3,5,7,9$, is an example of non-smooth discontinuous non-linearities which however fit well (locally and beyond jump points) the Haar approximation (in our case for the scale $m \geq 5$). Since the non-linear characteristics are odd and the probability density function of the input signal

is symmetric, we have $e = 0$ in the example (cf. Section 4). Functions 1 and 2 and the input density $f(x)$ satisfy the conditions of Section 2 (in particular (2.3)) and the conditions (6.1) and (6.2) of Section 6, excluding the break-points of Function 2. The linear output dynamics in the Hammerstein system has been selected as follows (see Fig. 1 and equation (2.2))

$$v_k = 0.5\,w_{k-1} + 0.25\,w_{k-2} + 0.125\,w_{k-3} + 0.0625\,w_{k-4}$$

i.e. *FIR*(4) subsystem with one-step delay was chosen ($\lambda_0 = 0$, $d = 1$). The subsystem is obviously asymptotically stable (see assumptions in Section 2). The correlated output noise $\{z_k;\, k = ..., -1, 0, 1, 2, ...\}$ was generated from the white noise $\{\varepsilon_k\}$ as *MA*(3) process, using the filter (see Fig. 1*b* and equation (2.1))

$$z_k = \varepsilon_k + 0.6\varepsilon_{k-1} + 0.4\varepsilon_{k-2} + 0.2\varepsilon_{k-3}$$

The level of the noise (white or correlated) was varied, by changing l, so as to give the noise dispersion value (σ_ϵ or σ_z) of 1%, 5%, and 10% of the maximum noiseless output signal value, $\max|v_k|$. In the estimate $R_N(x;m)$ we assumed $d = 1$ (see (4.3)-(4.4)) which resulted in the non-linearity scaling constant $c = \lambda_1 = 0.5$ (see Section 4). The length of data sequence was chosen as $N = 50, 100, 150, ..., 500$ and the corresponding data $\{(x_k,y_{k+1})\}_{k=1}^{N}$ were obtained by computer simulation. For each number N of data, the appropriate estimate $R_N(x;m)$ has been computed (according to (4.3a)-(4.4a)) and evaluated. As the estimate accuracy index, an empirical counterpart

$$MISE_{emp}(m;N) = \frac{1}{P}\sum_{r=1}^{P}\left\{\frac{1}{1000}\sum_{i=-500}^{500}\left[cR(x_i) - R_N^r(x_i;m)\right]^2\right\} \qquad (7.1)$$

of the *MISE* error

$$MISE(m;N) = E\int_{-0.5}^{0.5}\left[cR(x) - R_N(x;m)\right]^2 dx$$

over [-0.5,0.5] has been used. The latter for the interval [-0.5,0.5] (of the unit length) is also an average pointwise estimation error. The error (7.1) has been calculated at the points $x_i = i/1000$, $i = -500,...,500$, from the equispaced x-grid on [-0.5,0.5] and for $P = 100$ independent trials. To compute $MISE_{emp}(m;N)$, the data generation process was repeated $P = 100$ times for each data length N, yielding the estimate $R_N^r(x;m)$ in the r-th repetition. For each number N of data the scale factor m in the estimate has been selected according to the law $m(N) = \left[C\, m_{opt}(N)\right]$, where $m_{opt}(N)$ is the asymptotically optimal rule

(6.8) and where we have assumed $C = 1.5$. The constant $C = 1.5$, established experimentally, has minimized the error (7.1) in our finite-length experiment. The denotation $MISE_{emp}(N)$ in Figs. 2 and 3 is abbreviation of the $MISE_{emp}(m;N)$ error for this choice of m.

The run of the $MISE_{emp}(N)$ error for growing number N of data for Function 1 is shown in Fig. 2, and for Function 2 - in Fig. 3. The plots illustrate accuracy of the estimate in dependence on the

< *Fig*. 2 >

< *Fig*. 3 >

length of the set of measurements for various noise levels and for white (Figs. 2, 3*a*) and correlated (Figs. 2, 3*b*) noise. From the plots we see that the estimation error quickly decreases with growing N, and that - in each case - for $N > 250$ the error does not factually depend on the noise level, σ_ϵ or σ_z, in the output data. This is an obvious effect of smoothing inherent in our non-parametric estimate (cf. Section 4). The error value does not also visibly depend on the correlation of noise. For the white noise (Figs. 2, 3*a*) and *MA*(3) noise (Figs. 2, 3*b*) the corresponding plots coincide very closely, so that no essential difference can be observed for $N > 250$. These empirical observations confirm the theory developed in Sections 5 and 6.

It is also apparent from the plots in Figs. 2 and 3 that the error values for piecewise constant quantizer non-linearity (Function 2) are, in the same conditions, smaller than for smooth arctan non-linearity (Function 1). For example, for $N > 150$ we have $MISE_{emp}(N) < 0.005$ in the case of Function 2, for both white and coloured noise, whereas only $MISE_{emp}(N) < 0.01$ is quaranteed for Function 1. This better estimation accuracy is a clear effect of the fact that the Haar basis functions (cf. (3.6) in Section 3), and consequently the piecewise constant estimate $R_N(x;m)$ (see (4.3a)-(4.4a) in Section 4), are better matched with quantizer characteristic which results in a smaller bias (approximation) error.

8. CONCLUDING REMARKS

The identification algorithm presented in the paper has been based on the Haar multiresolution approximation. The algorithm is non-parametric, i.e., may be applied to estimate a non-linear characteristic of the Hammerstein system when the *a priori* knowledge of the system to be identified is very small, and in particular no parametric form of the characteristic is given. The only essential

restriction imposed on the non-linearity is that $|R(x)| \leq a_1|x| + a_2$ for some a_1 and a_2. As regards the probability density function of the random input signal we only require that it is bounded and that $\int_{-\infty}^{+\infty} f^2(x)\,dx < \infty$. From the engineering viewpoint these restrictions are of no importance and without exaggeration we can say that no *a priori* information is practically needed in order to apply the algorithm and successfully recover the unknown characteristic. Additional assumption that $ER(x) = 0$ (which is met if, e.g., R is an odd and f is an even function; Section 4) has been taken only for simplification of the proofs. Factually, one can apply the approach to a general case (i.e. to any density and any non-linearity). The main advantages of the algorithm can be summarized as follows.

1. The algorithm - consisting in essence in estimating regression - is very simple and requires only elementary computations; for the use of the algorithm only a set of coefficients $a_{mn,N}$ and $b_{mn,N}$ must be calculated from experimental data and these computations can be done in a short time using standard computing procedures (cf. (4.4a) in Section 4).

2. There exist simple recursive routines for updating $a_{mn,N}$ and $b_{mn,N}$ for each scale m (Section 4).

3. The algorithm requires moderate amount of computer memory; practically, to memorize the estimate only a number $O(2^m)$ of pairs $(a_{mn,N}, b_{mn,N})$ is required to be stored for the scale m (see Section 4) instead of the whole set of N input-output measurement data pairs $\{(x_k, y_{k+d}); k = 1, 2, \ldots, N\}$ (which is, e.g., the case for non-parametric kernel estimates[9-14]). This yields savings which rapidly grow with growing N as we have $2^m/N \to 0$ as $N \to \infty$ (see condition (5.5) in Section 5).

4. The algorithm is very flexible and easily copes with a delay in the system; to recover the unknown characteristic in the presence of a (known) delay it suffices to use in the general routine only the appropriately shifted output measurements (cf. equation (4.4a)).

5. The algorithm deals successfully with correlation of the noise; the form of the non-linearity estimate, convergence conditions and asymptotic rate of convergence are the same for white and coloured noise (of arbitrary correlation structure; Sections 4-6). This is in contrast to parametric methods.

6. The presented algorithm has satisfactory domain of convergence; it tends to the unknown non-linearity (up to a scale c) at the set of continuity points of R and f (Theorem 1).

7. For (locally) Lipschitz non-linearities and input densities R and f the algorithm attains the appropriate best possible non-parametric rate of convergence $O(N^{-1/3})$ in probability (Theorem 2).

Some weaker points of the algorithm are that:

1. It yields the estimate of only the scaled $cR(x)$ (or scaled and translated $cR(x) + e$) non-linearity $R(x)$; this is a consequence of limited measurement possibilities in the system (Sections 2 and 4).

2. There is no a closed-form expression for the estimate $R_N(x;m)$. To compute $R_N(x;m)$ for a set of x (e.g. for x in an interval $[x_{\min},x_{\max}]$) the calculations as in (4.3a)-(4.4a) have to be performed for each separate sub-interval $[n/2^m , (n+1)/2^m)$ of the resolution grid covered by $[x_{\min},x_{\max}]$; after changing (growing) scale m the calculations must be repeated. This is a result of small prior knowledge of the system.

3. We can reach in practice, i.e. for finite number N of data and finite scale m, only a crude step-wise approximation of non-linearity; this is a price paid for computational simplicity of the algorithm.

4. Convergence rate is not improved for smoother (differentiable) functions R and f; this is an effect of a rough nature of the Haar basis functions and insensitivity to the smoothness of the bias (approximation) error (see Section 6 and Appendix B).

Taking account of the above list of advantages and disadvantages, the proposed non-parametric identification routine may be recognized as a convenient and useful technique for recovering non-linearities in the Hammerstein systems. The algorithm can be particularly recommended when standard parametric methods fail, e.g., because of lack of appropriate preliminary information about the system or numerical complications introduced by parametric methods - especially in the case of correlated noise. Efficiency of the algorithm was confirmed by computer simulations.

In the paper we have focused on the Hammerstein systems and the Haar multiresolution approximation. Future research can be associated with generalization of the obtained results to more complex system structures and more general multiresolution (or wavelet) bases. Even a sketchy review of the theory presented in the paper leads us to a conviction that the approach can be also used for obtaining non-linear characteristics of more complex block-oriented systems, where however the non-linearity to be recovered (or its version) can be expressed as a regression function, similarly as was the case in (4.1). Such systems, composed of non-linear static and linear dynamic elements and described by the general equation

$$y_k = [\alpha R(x_{k-d}) + \beta] + \sum_{\substack{p=0 \\ p \neq d}}^{\infty} \eta_p \, \xi(x_{k-p}) + z_k$$

are characterized by the property that the non-linearity of interest (or its version $\alpha R(x) + \beta$) can be

isolated from the rest of the system as a separate static component and the remaining part of the system acts, in a sense, as a source of a zero-mean correlated nuisance noise $\zeta_k = \sum_{p \neq d} \eta_p \xi(x_{k-p})$ where $\sum_{p \neq d} |\eta_p| < \infty$ and $E\xi(x_k) = 0$. Examples are parallel, series-parallel and other systems without output feedback. Non-parametric identification by the use of the Haar multiresolution approximation of such more general complexes in the case of white noise and without delay in the system has been recently discussed in Reference 45. On the other hand, a refinement of the guaranteed rate (6.12) may be expected for the Hammerstein system if more smooth than Haar multiresolution basis functions $\{\phi_{mn}(x)\}$ will be implemented in the case of smooth functions R and f (e.g. Daubechies bases of higher numbers[29,31]). Selection of the basis $\{\phi_{mn}(x)\}$ adequately matching smoothness of R and f requires however in practice much more precise prior information about R and f than that assumed in the present paper. Moreover, for smoother functions $\phi_{mn}(x)$ the identification algorithm will be more complicated as, in contrast to the Haar basis functions, such functions are not given in an explicit form but need recursive calculation for each point x (see Reference 29). Nevertheless, the relationships between smoothness of R, f and basis functions $\phi_{mn}(x)$, and their impact on estimation accuracy (in particular bias error) and the rate of convergence of the generalized version of the estimate (4.3)-(4.4) are, at least for standard classes of non-linearities $R(x)$ and input densities $f(x)$, of an obvious interest and should be a subject of forthcoming analysis. Such studies may be then extended to more general classes of systems, indicated above.

ACKNOWLEDGEMENTS

The author would like to thank the reviewers for their helpful comments and M.Sc. P. Śliwiński for his assistance in preparing the numerical example.

APPENDIX A

It is fundamental for the subsequent analysis (particularly in the part concerning variance errors) to recall that in the sample means (4.4) the measurements x_k are independent (and identically distributed) random variables (see assumptions in Section 2) whereas the output observations y_{k+d}, for white as well as coloured noise, are stationary but dependent (correlated) quantities, as the outputs of an asymptotically stable linear output dynamics. Moreover, according to the assumption in Section 4, $ER(x_k) = 0$ whence

$e = 0$.

The estimate (4.3)-(4.4) can be rewritten in the following equivalent form:

$$R_N(x;m) \;=\; \frac{g_N(x;m)}{f_N(x;m)} \;=\; \frac{\dfrac{1}{N}\displaystyle\sum_{k=1}^{N} q_m(x_k,x)\,y_{k+d}}{\dfrac{1}{N}\displaystyle\sum_{k=1}^{N} q_m(x_k,x)} \tag{A.1}$$

where $q_m(v,x)$ is the kernel function of the Haar system $\{\phi_{H,mn}(x),\ n \in \mathrm{Z}\}$ (cf. (3.6) in Section 3)

$$q_m(v,x) \;=\; \sum_{n=-\infty}^{\infty} \varphi_{H,mn}(v)\,\varphi_{H,mn}(x) \;=\; 2^m\, I_{\left[\frac{[2^m x]}{2^m},\ \frac{[2^m x]+1}{2^m}\right)}(v) \tag{A.2}$$

In the above expression $[x]$ denotes the integer part of x, and $I_{[a,b)}(v)$ and m are the indicator function of $[a,b)$ and the scale factor as before.

Consider (A.1) for a random sample $\{(X_k, Y_{k+d})\}_{k=1}^{N}$ (further on random variables will be denoted by capitals for clarity). We have

$$\begin{aligned} g_N(x;m) \;&=\; \frac{1}{N}\sum_{k=1}^{N} q_m(X_k,x)\,Y_{k+d} \\ f_N(x;m) \;&=\; \frac{1}{N}\sum_{k=1}^{N} q_m(X_k,x) \end{aligned} \tag{A.3}$$

A. Bias Error - limit properties for $m \to \infty$ (see (5.1)-(5.2) in Section 5)

Including that $E\{Y_k \,|\, X_{k-d} = v\} = cR(v)$ (see (4.1)) and that $g(v) = cR(v)f(v)$ (see (4.2)), we obtain

$$E\,g_N(x;m) \;=\; E\{q_m(X_{k-d},x)\,Y_k\} \;=\; \int_{-\infty}^{\infty} q_m(v,x)\,g(v)\,dv$$

i.e. the bias error of $g_N(x;m)$ playing the role of the estimate of $g(x)$ is

$$\mathrm{bias}\{g_N(x;m)\} \;=\; E\,g_N(x;m) - g(x) \;=\; \int_{-\infty}^{\infty} q_m(v,x)\,g(v)\,dv - g(x) \tag{A.4}$$

Since $g(v) \in L^2(\mathbb{R})$ (see Section 4) and the kernel sequence $\{q_m\}$ with q_m given by (A.2) is, for growing m, a positive-type delta sequence on $\mathbb{R}$ (see: Reference 42; chapter 2, section 9.1, p. 319), we have the convergence

$$\int_{-\infty}^{\infty} q_m(v,x)\,g(v)\,dv \;\to\; g(x)$$

as $m \to \infty$ for each point $x \in (-\infty,\infty)$ at which g is continuous (see Reference 42; chapter 2, section 9.3, pp. 325-327). Hence

$$\text{bias}\{g_N(x;m)\} \to 0 \quad \text{as} \quad m \to \infty$$

for each N and at every point $x \in \text{Cont}(g)$ - the set of continuity points of g, which concludes (5.1) in Section 5. This convergence holds equally for white and coloured noise.

By similar arguments, noticing that

$$E f_N(x;m) = E\{q_m(X_k, x)\} = \int_{-\infty}^{\infty} q_m(v,x) f(v)\, dv$$

and that

$$\text{bias}\{f_N(x;m)\} = E f_N(x;m) - f(x) = \int_{-\infty}^{\infty} q_m(v,x) f(v)\, dv - f(x) \tag{A.5}$$

we infer that for each N it holds

$$\text{bias}\{f_N(x;m)\} \to 0 \quad \text{as} \quad m \to \infty$$

at any point $x \in \text{Cont}(f)$, which proves (5.2).

B. Variance Error

Observe that

$$\text{var}\{g_N(x;m)\} = \frac{1}{N^2} \text{var}\Big[\sum_{k=1}^{N} Y_{k+d}\, q_m(X_k,x)\Big]$$

$$= \frac{1}{N^2}\sum_{k=1}^{N} \text{var}\Big[Y_{1+d}\, q_m(X_1,x)\Big] + \frac{2}{N^2}\sum_{k=1}^{N-1}(N-k)\,\text{cov}\Big[Y_{k+1+d}\, q_m(X_{k+1},x), Y_{1+d}\, q_m(X_1,x)\Big]$$

$$= Q_1 + Q_2 \tag{A.6}$$

say, where we have used stationarity of the processes $\{X_k\}$ and $\{Y_k\}$. Since $Y_{1+d} = V_{1+d} + Z_{1+d}$ and V_{1+d}, Z_{1+d} are independent random variables such that $\text{E}\, V_{1+d}^2 < \infty$, $\text{E}\, Z_{1+d}^2 < \infty$ for both white and coloured noise (see (2.1), (2.2) and assumptions in Section 2), we obtain as the first step

$$Q_1 = \frac{1}{N}\text{var}\Big[Y_{1+d}\, q_m(X_1,x)\Big] = \frac{1}{N}\text{var}\Big[V_{1+d}\, q_m(X_1,x)\Big] + \frac{1}{N}\text{var}\Big[Z_{1+d}\, q_m(X_1,x)\Big]$$

Clearly,

$$\text{var}\Big[V_{1+d}\, q_m(X_1,x)\Big] \le E\Big[V_{1+d}^2\, q_m^2(X_1,x)\Big] \tag{A.7}$$

$$\text{var}\left[Z_{1+d}\, q_m(X_1,x) \right] \;=\; E\, Z_{1+d}^2\, E\, q_m^2(X_1,x) \tag{A.8}$$

the latter by independence of $\{X_k\}$ and $\{Z_k\}$ and the fact that $\mathrm{E}\, Z_{1+d} = 0$. Noticing that (see (2.2))

$$V_{1+d} \;=\; \lambda_d R(X_1) \;+\; \sum_{\substack{p=0 \\ p \neq d}}^{\infty} \lambda_p R(X_{1+d-p})$$

and including the fact that $\{W_k = R(X_k)\}$, the input signal of the dynamical subsystem with the impulse response $\{\lambda_p; p = 0, 1, \ldots\}$, is a stationary white process with $EW_k = 0$ and finite variance (see assumptions in Section 2), we find after some algebra that

$$E\left[V_{1+d}^2\, q_m^2(X_1,x) \right] \;=\; \lambda_d^2\, E\left[R^2(X_1)\, q_m^2(X_1,x) \right] \;+\; E\left[q_m^2(X_1,x) \right] E\, W_1^2 \sum_{\substack{p=0 \\ p \neq d}}^{\infty} \lambda_p^2 \tag{A.9}$$

In turn, one can ascertain that for each x, a fixed point, and large scale m it holds (see the definition (A.2), (2.3) and the assumptions in Section 2 concerning the density $f(x)$)

$$E\, q_m^2(X_1, x) \;=\; \int_{-\infty}^{\infty} q_m^2(v,x)\, f(v)\, dv \;\leq\; M_f\, 2^m \tag{A.10}$$

$$E\left[R^2(X_1)\, q_m^2(X_1, x) \right] \;=\; \int_{-\infty}^{\infty} q_m^2(v,x)\, R^2(v)\, f(v)\, dv \;\leq\; d^2(x)\, M_f\, 2^m \tag{A.11}$$

where $d(x) = a_1 |x| + a_2$. Hence, including (A.7)-(A.8) and the expression (A.9), we realize that for each point x, we have

$$\text{var}\left[V_{1+d}\, q_m(X_1,x) \right] \leq\; 2^m M_f \left[\lambda_d^2 d^2(x) \;+\; E\, W_1^2 \sum_{\substack{p=0 \\ p \neq d}}^{\infty} \lambda_p^2 \right]$$

and

$$\text{var}\left[Z_{1+d}\, q_m(X_1,x) \right] \leq \begin{cases} 2^m M_f E \varepsilon_1^2 & , \; \textit{for white noise} \\ 2^m M_f E \varepsilon_1^2 \sum_{p=0}^{\infty} \omega_p^2 & , \; \textit{for coloured noise} \end{cases}$$

For asymptotically stable system and noise filter dynamics this yields as a straightforward conclusion that

$$Q_1 \;=\; O\left(2^m / N\right) \tag{A.12}$$

for white and coloured noise.

The covariance terms and the summand Q_2 in (A.6) will be considered separately for white and coloured noise. Generally, by independence of $\{V_k\}$ and $\{Z_k\}$, we have

$$\operatorname{cov}\left[\, Y_{k+1+d}\, q_m(X_{k+1},x), Y_{1+d}\, q_m(X_1,x)\,\right]$$

$$= \operatorname{cov}\left[\, V_{k+1+d}\, q_m(X_{k+1},x), V_{1+d}\, q_m(X_1,x)\,\right] + \operatorname{cov}\left[\, Z_{k+1+d}\, q_m(X_{k+1},x), Z_{1+d}\, q_m(X_1,x)\,\right] \quad \text{(A.13)}$$

(*a*) *White Noise Case*

For the white noise $\{Z_k\}$ and $k \geq 1$, we have

$$\operatorname{cov}\left[\, Y_{k+1+d}\, q_m(X_{k+1},x), Y_{1+d}\, q_m(X_1,x)\,\right] = \operatorname{cov}\left[\, V_{k+1+d}\, q_m(X_{k+1},x), V_{1+d}\, q_m(X_1,x)\,\right]$$

Using the same arguments as for (A.9), we get after a routine but rather vast calculus that for $k \geq 1$ and large values of m the following bound is valid

$$\left|\operatorname{cov}\left[\, V_{k+1+d}\, q_m(X_{k+1},x), V_{1+d}\, q_m(X_1,x)\,\right]\right| \leq \left(\left|\lambda_d \lambda_{d+k}\right| + \left|\lambda_d \lambda_{d-k}\right|\right) E^2\left[R(X_1)\, q_m(X_1,x)\,\right]$$

$$+ E^2\left[q_m(X_1,x)\right] E W_1^2 \sum_{\substack{p \neq d \\ p \neq d-k}}^{\infty} \left|\lambda_p \lambda_{p+k}\right|$$

Hence, including that

$$E^2\left[q_m(X_1,x)\right] \leq E\, q_m^2(X_1,x) \quad \text{(A.14)}$$

$$E^2\left[R(X_1)\, q_m(X_1,x)\,\right] \leq E\left[R^2(X_1)\, q_m^2(X_1,x)\,\right]$$

and taking account of (A.10)-(A.11), we obtain

$$\left|\operatorname{cov}\left[\, V_{k+1+d}\, q_m(X_{k+1},x), V_{1+d}\, q_m(X_1,x)\,\right]\right| \leq 2^m M_f \left(\left(\left|\lambda_d \lambda_{d+k}\right| + \left|\lambda_d \lambda_{d-k}\right|\right) d^2(x)\right.$$

$$\left. + E W_1^2 \sum_{\substack{p \neq d \\ p \neq d-k}}^{\infty} \left|\lambda_p \lambda_{p+k}\right|\right) \quad \text{(A.15a)}$$

This bound and the bound

$$\left|Q_2^{\,white\ noise}\right| \leq \frac{2}{N^2} \sum_{k=1}^{N-1} (N-k) \left|\operatorname{cov}\left[\, V_{k+1+d}\, q_m(X_{k+1},x), V_{1+d}\, q_m(X_1,x)\,\right]\right|$$

yield together

$$\left|Q_2^{\,white\ noise}\right| \leq 2^m M_f \frac{2}{N^2} \sum_{k=1}^{N-1} (N-k) \left(\left(\left|\lambda_d \lambda_{d+k}\right| + \left|\lambda_d \lambda_{d-k}\right|\right) d^2(x)\right.$$

$$\left. + E W_1^2 \sum_{\substack{p \neq d \\ p \neq d-k}}^{\infty} \left|\lambda_p \lambda_{p+k}\right|\right) \triangleq B Q_2^{\,white\ noise}(m,N)$$

which for asymptotically stable dynamics in the Hammerstein system gives eventually

$$\left|Q_2^{\,white\ noise}\right| = O\left(2^m / N\right) \quad \text{(A.16a)}$$

(*b*) *Coloured Noise Case*

In the case of coloured noise both components in (A.13) are present. For the first component, the bound (A.15a) holds - identically as for the white noise case.

For the second component, considering that the input noise $\{\varepsilon_k\}$ of the noise filter (with the impulse response $\{\omega_p; p = 0, 1, \ldots\}$; cf. (2.1)) is a stationary white process, with $E\varepsilon_k = 0$ and finite variance, independent of $\{X_k\}$ (see assumptions in Section 2), we obtain for $k \geq 1$ that

$$|\operatorname{cov}\left[Z_{k+1+d}\, q_m(X_{k+1}, x), Z_{1+d}\, q_m(X_1, x)\right]| \quad = \quad E^2\left[q_m(X_1, x)\right] E\varepsilon_1^2 \sum_{p=0}^{\infty} |\omega_p \omega_{p+k}|$$

Hence, taking account of (A.14) and (A.10), we get

$$|\operatorname{cov}\left[Z_{k+1+d}\, q_m(X_{k+1}, x), Z_{1+d}\, q_m(X_1, x)\right]| \quad \leq \quad 2^m M_f E\varepsilon_1^2 \sum_{p=0}^{\infty} |\omega_p \omega_{p+k}| \qquad \text{(A.15b)}$$

Since (cf. (A.6) and (A.13))

$$|Q_2^{\text{coloured noise}}| \quad \leq \quad \frac{2}{N^2} \sum_{k=1}^{N-1} (N-k)\, |\operatorname{cov}\left[V_{k+1+d}\, q_m(X_{k+1}, x), V_{1+d}\, q_m(X_1, x)\right]|$$

$$+ \quad \frac{2}{N^2} \sum_{k=1}^{N-1} (N-k)\, |\operatorname{cov}\left[Z_{k+1+d}\, q_m(X_{k+1}, x), Z_{1+d}\, q_m(X_1, x)\right]|$$

thus, owing to (A.15a) and (A.15b), we obtain

$$|Q_2^{\text{coloured noise}}| \quad \leq \quad BQ_2^{\text{white noise}}(m, N) \; + \; 2^m M_f E\varepsilon_1^2 \frac{2}{N^2} \sum_{k=1}^{N-1} (N-k) \sum_{p=0}^{\infty} |\omega_p \omega_{p+k}|$$

which for asymptotically stable dynamics in the Hammerstein system and asymptotically stable noise filter yields

$$|Q_2^{\text{coloured noise}}| \quad = \quad O\left(2^m / N\right) \qquad \text{(A.16b)}$$

just as for the white noise case.

Putting together (A.6), (A.12) and (A.16a)-(A.16b), we conclude that for each point x and both white and coloured noise

$$\text{var}\,\{\, g_N(x; m)\,\} \quad = \quad O\left(2^m / N\right)$$

which is the variance bound (5.3) in Section 5.

In turn, including that (cf. (A.3))

$$\text{var}\{ f_N(x;m)\} \quad = \quad \frac{1}{N}\ \text{var}\{ q_m(X_1,x)\} \tag{A.17}$$

and that

$$\text{var}\left[q_m(X_1,x)\right] \quad \le \quad E\, q_m^2(X_1,x)$$

and next using (A.10), we obtain

$$\text{var}\left[q_m(X_1,x)\right] \quad \le \quad 2^m M_f$$

which together with (A.17) yields

$$\text{var}\{ f_N(x;m)\} \quad = \quad O\left(2^m/N\right)$$

i.e. the variance bound (5.4).

APPENDIX B

A. Asymptotic Bias Error for Lipschitz and Differentiable Functions

We have (cf. (A.4))

$$|\ \text{bias}\{ g_N(x;m)\}\ | \quad = \quad |\ E\, g_N(x;m) \ - \ g(x)\ | \quad = \quad \left|\ \int_{-\infty}^{\infty} q_m(v,x)\, g(v)\, dv \ - \ g(x)\right|$$

Taking account of the definition (A.2) in Appendix A, we ascertain that

$$\int_{-\infty}^{\infty} q_m(v,x)\, g(v)\, dv \quad = \quad 2^m \int_{\frac{[2^m x]}{2^m}}^{\frac{[2^m x]}{2^m}+\frac{1}{2^m}} g(v)\, dv$$

and hence

$$|\ \text{bias}\{ g_N(x;m)\}\ | \quad = \quad |\ 2^m \int_{\frac{[2^m x]}{2^m}}^{\frac{[2^m x]}{2^m}+\frac{1}{2^m}} \big(g(v)-g(x)\big)dv\ | \quad \le \quad 2^m \int_{\frac{[2^m x]}{2^m}}^{\frac{[2^m x]}{2^m}+\frac{1}{2^m}} |\ g(v)-g(x)\,|\, dv \tag{B.1}$$

For large values of *m*, local properties of *g* around *x* refer to each point *v* in the interval $\left[[2^m x]/2^m\,,\,[2^m x]/2^m + 1/2^m\right]$.

1. Lipschitz Functions

For locally Lipschitz function *g*, $g \in Lip(x)$, and large scale *m* we have (cf. (6.3) in Section 6)

$$|\,\text{bias}\{g_N(x;m)\}\,| \;\le\; L_g\, 2^m \int_{\frac{[2^m x]}{2^m}}^{\frac{[2^m x]}{2^m}+\frac{1}{2^m}} |v-x|\,dv \;\le\; \frac{L_g}{2^m}$$

where we have used that $|v-x| \le 1/2^m$. This is the asymptotic bound (6.4).

2. Differentiable Functions

For locally p times differentiable functions R and f around the point x ($p \ge 1$), the function $g = cRf$ is also p times differentiable around x (by the Leibnitz formula). Applying the Taylor series expansion technique with respect to $g(v)$ at x, a fixed point, we obtain

$$|\,g(v) - g(x)\,| \;\le\; C_1\,|v-x| + C_2\,|v-x|^2 + \cdots + C_p\,|v-x|^p$$

where C_i's are positive constants. Hence, for large values of m (cf. (B.1))

$$|\,\text{bias}\{g_N(x;m)\}\,| \;\le\; C_1\Big(\frac{1}{2^m}\Big) + C_2\Big(\frac{1}{2^m}\Big)^2 + \cdots + C_p\Big(\frac{1}{2^m}\Big)^p$$

where we have exploited that $|v-x| \le 1/2^m$. Since however for large m higher-order terms in the above expression can be omitted, we obtain eventually

$$|\,\text{bias}\{g_N(x;m)\}\,| \;\le\; \frac{C_1}{2^m}$$

i.e. the asymptotic bias error (6.4), the same for each p.

Taking account of (A.5), similar asymptotic bias errors can be established for the estimate $f_N(x;m)$, yielding for locally Lipschitz/differentiable density function f the bound (6.5).

B. Proof of Theorem 2

The proof follows from (6.9) and (6.10) in Section 6 and the following lemma for the ratio estimates.

Lemma (cf. Reference 43; lemma 2 or Reference 44; lemma B1)

If $R_N(x;m) = g_N(x;m)/f_N(x;m)$ is an estimate of $cR(x) = g(x)/f(x)$ and at the point x it holds

$$E\big(g_N(x;m) - g(x)\big)^2 \;=\; O(\alpha_N)$$

$$E\big(f_N(x;m) - f(x)\big)^2 \;=\; O(\beta_N)$$

then for arbitrary $\epsilon > 0$

$$\mathrm{P}\big\{\,|R_N(x;m) - cR(x)| > \varepsilon\,|cR(x)|\,\big\} \;=\; O(\gamma_N)$$

where $\gamma_N = \max(\alpha_N, \beta_N)$.

The first part, (6.11), of the theorem is an immediate consequence of the above lemma and the rates (6.9)-(6.10). The second part, (6.12), follows easily from the relation (see the proof of lemma 2 in Reference 43)

$$P\{|R_N(x;m) - cR(x)| > \varepsilon |cR(x)|\} \le P\{|g_N(x;m) - g(x)| > |g(x)|\varepsilon/(2+\varepsilon)\} + P\{|f_N(x;m) - f(x)| > |f(x)|\varepsilon/(2+\varepsilon)\}$$

when including the rates (6.9) and (6.10) and Chebyshev's inequality. We recall that in the theorem, for a sequence of random variables $\{\zeta_N\}$ by $\zeta_N = O(a_N)$ *in probability* we mean that $d_N\zeta_N/a_N \to 0$ *in probability* as $N \to \infty$ for any number sequence $\{d_N\}$ convergent to zero (see Reference 39, p. 140). □

REFERENCES

1. Billings, S. A., 'Identification of non-linear systems: a survey', *Proc. IEE*, **127**, 272-285 (1980).
2. Billings, S. A. and S. Y. Fakhouri, 'Identification of systems containing linear dynamic and static non-linear elements', *Automatica*, **18**, 15-26 (1982).
3. Haber, R. and H. Unbehauen, 'Structure identification of non-linear dynamic systems: a survey on input/output approaches', *Automatica*, **26**, 651-677 (1990).
4. Eskinat, E., S. H. Johnson and W. L. Luyben, 'Use Hammerstein models in identification of non-linear systems', *Amer. Inst. Chemical Engineering*, **37**, 255-268 (1991).
5. De Boer, E., 'Cross-correlation function of a bandpass non-linear network', *Proc. IEEE*, **64**, 1443-1446 (1976).
6. Sawchuk, A. A. and T. C. Strand, 'Fourier optics in non-linear image processing', in: *Applications of Optical Fourier Transforms* (edited by H. Stark), Academic Press, New York, 1982, pp. 371-429.
7. Hall, C. F. and E. L. Hall, 'A non-linear model for the spatial characteristics of the human visual system', *IEEE Trans. System Man and Cybernetics*, **7**, 161-170 (1977).
8. Hunter, I. W. and M. J. Korenberg, 'The identification of non-linear biological systems: Wiener and Hammerstein cascade models', *Biolog. Cybern.*, **55**, 135-144 (1986).

9. Greblicki, W. and M. Pawlak, 'Identification of discrete Hammerstein systems using kernel regression estimates', *IEEE Trans. Automatic Control*, **31**, 74-77 (1986).
10. Greblicki, W. and M. Pawlak, 'Hammerstein system identification by non-parametric regression estimation', *Int. J. Control*, **45**, 343-354 (1987).
11. Greblicki, W. and M. Pawlak, 'Non-parametric identification of Hammerstein systems', *IEEE Trans. Information Theory*, **35**, 409-418 (1989).
12. Greblicki, W. and M. Pawlak, 'Recursive non-parametric identification of Hammerstein systems', *J. Franklin Inst.*, **326**, 461-481 (1989).
13. Greblicki, W. and M. Pawlak, 'Non-parametric identification of a cascade non-linear time series system', *Signal Processing*, **22**, 61-75 (1991).
14. Krzyzak, A., 'On estimation of a class of non-linear systems by the kernel regression estimate', *IEEE Trans.Information Theory*, **36**, 141-152 (1990).
15. Greblicki, W., 'Nonparametric orthogonal series identification of Hammerstein systems', *Int. J. Systems Sci.*, **20**, 2355-2367 (1989).
16. Krzyzak, A., 'Identification of discrete Hammerstein systems by the Fourier series regression estimate', *Int. J. Systems Sci.*, **20**, 1729-1744 (1989).
17. Pawlak, M., 'On the series expansion approach to the identification of Hammerstein systems', *IEEE Trans. Automatic Control*, **36**, 763-767 (1991).
18. Norton, J. P., *An Introduction to Identification*, Academic Press, London, 1986.
19. Ljung, L., *System Identification - Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
20. Söderström, T. and P. Stoica, *System Identification*, Prentice-Hall, New York, 1989.
21. Antoniadis, A., G. Gregoire and I. W. McKeague, 'Wavelet methods for curve estimation', *J. Amer. Statist. Assoc.*, **89**, 1340-1353 (1994).
22. Cambanis, S. and E. Masry, 'Wavelet approximation of deterministic and random signals: convergence properties and rates', *IEEE Trans. Information Theory*, **40**, 1013-1029 (1994).
23. Kelly, S., M. Kon and L. Raphael, 'Pointwise convergence of wavelet expansions', *Bull. Amer. Math. Soc.*, **30**, 87-94 (1994).
24. Zhang, J., G. G. Walter, Y. Miao and W. N. W. Lee, 'Wavelet neural networks for function learning', *IEEE Trans. Signal Processing*, **43**, 1485-1497 (1995).

25. Sjöberg, J., Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P-Y. Glorennec, H. Hjalmarsson and A. Juditsky, 'Nonlinear black-box modeling in system identification: a unified overview', *Automatica*, **31**, 1691-1724 (1995).
26. Juditsky, A., H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjöberg and Q. Zhang, 'Nonlinear black-box models in system identification: mathematical foundations', *Automatica*, **31**, 1725-1750 (1995).
27. Chui, C. K., *An Introduction to Wavelets*, Academic Press, New York, 1992.
28. Chui, C. K., *Wavelets: A Tutorial in Theory and Applications*, Academic Press, New York, 1992.
29. Daubechies, I., *Ten Lectures on Wavelets*, SIAM Edition, Philadelphia, 1992.
30. Meyer, Y., *Wavelets - Algorithms and Applications*, SIAM Edition, Philadelphia, 1993.
31. Walter, G. G., *Wavelets and Other Orthogonal Systems with Applications*, CRC Press, Boca Raton, 1994.
32. Mallat, S., *Wavelet Tour of Signal Processing*, Academic Press, San Diego, 1998.
33. Mallat, S., 'Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$', *Trans. Amer. Math. Soc.*, **315**, 69-87 (1989).
34. Mallat, S., 'A theory for multiresolution signal decomposition: the wavelet representation', *IEEE Trans. Pattern Analysis and Machine Intelligence*, **11**, 674-693 (1989).
35. Bendat, J. S., *Nonlinear System Analysis and Identification*, Wiley-Interscience, New York, 1990.
36. Prakasa Rao, B. L. S., *Nonparametric Functional Estimation*, Academic Press, Orlando, 1983.
37. Eubank, R. L., *Spline Smoothing and Nonparametric Regression*, Marcel-Dekker, New York, 1988.
38. Härdle, W., *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, UK, 1990.
39. Greblicki, W. and M. Pawlak, 'Cascade non-linear system identification by a non-parametric method', *Int. J. Systems Sci.*, **25**, 129-153 (1994).
40. Bickel, P. J. and K. A. Doksum, *Mathematical Statistics. Basic Ideas and Selected Topics*, Hoden-Day, San Francisco, 1977.
41. Stone, C. J., 'Optimal rates of convergence for nonparametric estimators', *Ann. Statist.*, **8**, 1348-1360 (1980).
42. Korevaar, J., *Mathematical Methods. Vol. 1*, Academic Press, New York, 1968.
43. Greblicki, W. and M. Pawlak, 'Fourier and Hermite series estimates of regression functions', *Ann.*

Inst. Statist. Math., **37A**, 443-454 (1985).

44. Greblicki, W., 'Nonparametric identification of Wiener systems by orthogonal series', *IEEE Trans. Automatic Control*, **39**, 2077-2086 (1994).
45. Pawlak, M. and Z. Hasiewicz, 'Non-linear system identification by the Haar multiresolution analysis', *IEEE Trans. Circuits and Systems - I: Fundamental Theory and Appl.*, **45**, 945-961 (1998).

CAPTIONS FOR FIGURES

Figure 1. The Hammerstein system disturbed by the noise z_k: (*a*) - white noise case, (*b*) - coloured noise case.

Figure 2. Estimation accuracy versus number of measurement data for different noise levels; Function 1: (*a*) - white noise, (*b*) - correlated noise.

Figure 3. Estimation accuracy versus number of measurement data for different noise levels; Function 2: (*a*) - white noise, (*b*) - correlated noise.