Copyright © IFAC System Identification
Santa Barbara, California, USA, 2000

NONLINEAR SYSTEM IDENTIFICATION VIA WAVELET EXPANSIONS

Z. Hasiewicz* M. Pawlak**

* *Institute of Engineering Cybernetics, Wroclaw University of Technology, Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland, e-mail: zhas@ict.pwr.wroc.pl*
** *Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Manitoba, R3T 5V6, Canada, e-mail: pawlak@ee.umanitoba.ca*

Abstract: The paper deals with the problem of recovering a nonlinearity in a class of nonlinear dynamical systems of block-oriented structure.The class includes a large number of previously examined block-oriented models.The sought nonlinearity is allowed to have singular points like discontinuities and points of non-differentiability. In order to cope with such general nonlinearities the theory of wavelet expansions is applied. A major advantage of these expansions is adaptation to erratic behavior of the nonlinearity and local adaptation to the degree of smoothness of an unknown characteristic. Hence a wavelet-based identification algorithm of the nonlinearity is proposed and conditions for the convergence of the algorithm are given. For nonlinearities satisfying some smoothing conditions the rate of convergence is also evaluated.

Keywords: nonlinear system identification, block-oriented systems, wavelets expansions, convergence analysis.

1. INTRODUCTION

A large class of physical systems is nonlinear or reveal nonlinear behavior if they are considered over a broad operating range. Hence the commonly used linearity assumption can be regarded only as a first-order approximation to the observed process. System identification is the fundamental problem of complete determination of a system description from an analysis of its input and output data. A large class of techniques exist for identification of linear models. Much less attention has been paid to nonlinear system identification, mostly because their analysis is generally harder and because the range of nonlinear model structures and behaviors is much broader than the range of linear model structures and behaviors. There is no universal approach to identification of nonlinear systems, and existing solutions depend strongly on a prior knowledge of the system structure, see (Bendat, 1990), (Billings, 1980), (Chen, 1995), (Hunter and Korenberg, 1986) for some classical techniques for nonlinear system identification. A promising strategy for nonlinear system identification is based on the assumption that the system structure is known. This yields the concept of block-oriented models, i.e., models consisting of linear dynamic subsystems and static nonlinear elements connected together in a certain composite structure. Signals interconnecting the subsystems are not accessible for measurements making the identification problem not reducible to the standard situations, i.e., identification of linear dynamic systems and recovering memoryless nonlinearities separately. Such composite models have found numerous applications in such diverse areas as biology, communication systems, chemical engineering, psychology and sociology

(Bendat, 1990), (Billings, 1980), (Chen, 1995), (Hunter and Korenberg, 1986). A class of cascade/parallel models, i.e., when linear dynamic subsystems are in a tandem/parallel connection with a static element is a popular type of block-oriented structures. Examples of such models include cascade Hammerstein, Wiener and sandwich structures and their parallel counterparts. Traditionally it has been assumed that the nonlinearity in such models can be parameterize, e.g., it can be a polynomial of a finite and known order. The parametric restriction is often too rigid yielding incorrect conclusions about the system structure. A non-parametric identification algorithm for reconstruction of nonlinearities in a cascade block-oriented model has been first proposed in (Greblicki and Pawlak, 1986). We also refer to (Greblicki, 1994), (Greblicki and Pawlak, 1994), (Pawlak and Hasiewicz, 1998) for detailed discussion of nonparametric approaches to identification of the cascade/parallel block-oriented models. The aim of nonparametric methods is to relax assumptions on the form of an underlying nonlinear characteristic, and to let the training data decide which characteristic fits them best. These approaches are powerful in exploring fine details in the nonlinear characteristics (Härdle, 1990), (Juditsky *et al.*, 1995), (Sjoberg *et al.*, 1995).

In this paper we consider the nonparametric approach to the identification of a broad class of nonlinear composite models which includes most previously defined connections. We are mostly interested in recovering the system nonlinearity which is embedded in a block oriented structure containing dynamic linear subsystems and other "nuisance" nonlinearities. Our approach is based on regression analysis and we propose the identification algorithms originating from the area of nonparametric regression techniques (Härdle, 1990). The identification algorithm is convergent for a large class of nonlinear characteristics and under very mild conditions on the model dynamics. The proposed estimates are based on the theory of orthogonal bases originating from wavelet approximations of square integrable functions. This theory provides elegant techniques for representing the levels of details of the approximated function (Daubechies, 1992), (Mallat, 1998), (Vetterli and Kovacevic, 1995), (Walter, 1994). The wavelet theory has found recently applications in a remarkable diversity of disciplines. A little attention, however, has been paid to the application of the wavelet methodology to control theory and system identification in particular(Juditsky *et al.*, 1995), (Pawlak and Hasiewicz, 1998), (Sjoberg *et al.*, 1995), (Sureshbabu and Farrell, 1999).

In this paper we apply the wavelet analysis to the identification of the proposed nonlinear composite systems. In particular a class of wavelet orthogonal expansions with scaling and wavelet functions of compact support is taken into considerations. We give conditions for the identification algorithms to be pointwise convergent and find its optimal rate of convergence. As a result of these studies the optimal local choice of the resolution level is calculated. This optimal value depends on some unknown characteristics of the system and therefore the problem of estimating the resolution level from data is also addressed. It is worth mentioning that we are dealing with the dependent and non-Gaussian observations as the data are generated from non-linear and dynamical systems. For the sake of further developments let us give a brief overview of multiresolution and wavelet decompositions, see (Daubechies, 1992), (Mallat, 1998), (Vetterli and Kovacevic, 1995), (Walter, 1994) for the detailed treatment of this subject. Let Z denote the set of all integers in R. The multiresolution representation of a function f from $L_2(R)$ at the resolution m is given by

$$f_m(x) = \sum_{k\in Z} \alpha_{mk}\varphi_{mk}(x), \qquad (1.1)$$

where $\alpha_{mk} = \int_{-\infty}^{\infty} f(x)\varphi_{mk}(x)dx$ is the kth Fourier coefficient, $\varphi_{mk}(x) = 2^{m/2}\varphi(2^m x - k)$ and $\varphi(x)$ is the so-called father wavelet function. This function is such that $\{\varphi(x-k), k \in Z\}$ is orthonormal system in $L_2(R)$ and $\{\varphi_{mk}(x), k \in Z\}$, $m = 0, \pm 1, \ldots$ generates a sequence of nested spaces $\{V_m, m \in Z\}$ in $L_2(R)$ which union is dense in $L_2(R)$. A number of practical father wavelet functions with various properties have been proposed in the literature, culminating in the seminal work of Daubechies (Daubechies, 1992) on compactly supported father wavelet functions. It should be noted that for such functions the infinite sum in (1.1) contains only a finite number of terms. Hence for compactly supported scaling functions the multiscale basis $\{\varphi_{mk}(x)\}$ consists of functions which are non-zero in a finite interval and as m increases the support of $\varphi_{mk}(x)$ shrinks, i.e., $\varphi_{mk}(x)$ becomes taller and thinner. The wavelet analysis characterizes the detail information hidden between two consecutive resolution levels. This is quantitatively described by the fact that there exists a mother function $\psi(x)$ such that $\{2^{m/2}\psi(2^m x - k), k \in Z\}$ forms an orthonormal basis in the orthogonal complement of V_{m+1} in V_m. The mother wavelet function $\psi(x)$ can be derived from a given scaling function $\varphi(x)$. As a result any $f \in L_2(R)$ can be approximated at the resolution level $m_0 + r + 1$ as follows

$$f_{m_0+r+1}(x) = \sum_{k\in Z} a_{m_0k}\varphi_{m_0k}(x) + \sum_{s=m_0}^{m_0+r}\sum_{k\in Z} b_{sk}\psi_{sk}(x). \qquad (1.2)$$

The first term in (1.2) represents our initial guess about $f(x)$, whereas the second one adds further layers of information about $f(x)$. It is common to call the $\{b_{sk}\}$ as the detail coefficients. On the other hand they are Fourier coefficients corresponding to $\psi_{sk}(x)$.

It is easy to show that both $f_m(x)$ in (1.1) and $f_{m_0+r+1}(x)$ in (1.2) tend to $f(x)$ in the $L_2(R)$ norm as $m \to \infty$ and $r \to \infty$, respectively. The pointwise convergence of $f_m(x)$ and $f_{m_0+r+1}(x)$ is less trivial (Daubechies, 1992), (Mallat, 1998).

A popular choice of scaling and wavelet functions is the one corresponding to the Haar system. Here $\varphi(x) = \mathbf{1}_{[0,1)}(x)$ and $\psi(x) = \mathbf{1}_{[0,1/2)}(x) - \mathbf{1}_{[1/2,1)}(x)$, where $\mathbf{1}_A(x)$ denotes the indicator function of A. Hence the resolution level m contains all functions being constant on all intervals $\{[k2^{-m},(k+1)2^{-m}), k \in Z\}$. Smooth scaling and wavelet functions with compact support have been proposed by Daubechies (Daubechies, 1992).A popular class of non-compact supported wavelets is the Meyer-type wavelets. This type of wavelets belongs to the class of band-limited wavelets, i.e., the Fourier transform of $\varphi(x), \psi(x)$ is compactly supported.

2. THE NONLINEAR BLOCK-ORIENTED MODEL

Let us now introduce a class of nonlinear block-oriented models examined in this paper. The class is characterized by the general property that the nonlinear characteristic of our interest can be isolated from the rest of the system. Moreover the model under study is discrete time and time-invariant.

Hence our general nonlinear model is of the following input-output form

$$\begin{cases} O_n = \mu(X_n) + \xi_n \\ \xi_n = \sum_{j=1}^{p} s_j \lambda_j(X_{n-j}) \ , \\ Y_n = O_n + \varepsilon_n \end{cases} \tag{2.1}$$

where (X_n, Y_n) is the (input, output) pair, $\mu(x)$ represents the unknown system nonlinearity to be recovered, $\{\xi_n\}$ is the "system noise" process characterizing the history of the system and $\{\varepsilon_n\}$ is the measurement noise. The system noise process ξ_n is a measurable transformation of $\{X_{n-1}, X_{n-2}, \dots, X_{n-p}\}$, where $p, 1 \le p < \infty$ is the memory length. It is worth noting that p need not be known. Hence ξ_n has the following nonlinear moving average representation

$$\xi_n = \sum_{j=1}^{p} s_j \lambda_j(X_{n-j}). \tag{2.2}$$

The following assumptions concerning the model in (2.1) are used in the paper:

Assumption 1 : The inputs $\{X_1, X_2, \dots\}$ form a sequence of independent and identically distributed random variables which are independent of $\{\varepsilon_n\}$. The probability density f of $\{X_1, X_2, \dots\}$ exists but is unknown and satisfies the following restrictions:

$$\int_{-\infty}^{\infty} f^2(x)dx < \infty, \tag{A1.1}$$

and

$$0 < \eta \le f(x) \tag{A1.2}$$

for all $x \in R$ and some unknown η.

Assumption 2 : For the system noise process $\{\xi_n\}$ let $\{\lambda_j(x)\}$ be a sequence of measurable functions such that

$$E\lambda_j(X) = 0, \ j = 1, 2, \dots \tag{A2.1}$$

Assumption 3 : The nonlinear characteristic $\mu(x)$ is a measurable function satisfying the following conditions:

$$E\mu^2(X) < \infty, \tag{A3.1}$$

$$\int_{-\infty}^{\infty} (\mu(x)f(x))^2 dx < \infty, \tag{A3.2}$$

Assumption 4 : The measurement noise $\{\varepsilon_n\}$ is uncorrelated and such that:

$$E\,\varepsilon_n = 0, \qquad \operatorname{var} \varepsilon_n = \sigma^2 < \infty.$$

Let us elaborate on the role of the above conditions. Restriction (A1.1) is required since we use the $L_2(R)$ multiscale decomposition of $f(x)$. Condition (A1.2) says that we consider the estimation problem in such points on R where the input density is high, i.e., where $f(x)$ is strictly bounded away from zero. Assumption (A2.1) implies that $\{\xi_n\}$ is the second order covariance stationary stochastic process with $E\xi_n = 0$, $\operatorname{var}\xi_n < \infty$ and $\operatorname{cov}(\xi_n, \xi_{n+r}) = \sum_{j=1}^{p} s_j s_{j+r} E\{\lambda_j(X)\lambda_{j+r}(X)\}$, $|r| > 1$. This along with Assumption (A3.1) and Assumption 4 makes the output process $\{Y_n\}$ well defined, i.e., it is also a second order covariance stationary stochastic process. Note that if (A2.1) is not met then an additional bias term in estimating $\mu(x)$ is present. The only condition concerning $\mu(x)$ required in this paper is (A3.2). This condition is related to our identification procedure for recovering $\mu(x)$ using the wavelets decomposition in $L_2(R)$.

There is a large class of block-oriented nonlinear systems which fall into the description given in (2.1) and (2.2). For instance the following system with two nonlinearities and one dynamical subsystem meets our requirements.

$$\begin{cases} O_n &= \sum_{j=0}^{p} h_j(\theta(X_{n-j}) + \theta_0(X_{n-j-1})) \\ Y_n &= O_n + \varepsilon_n \end{cases}.$$

Here $\theta(x)$ is the nonlinearity to be estimated and $\theta_0(x)$ is a "nuisance" nonlinearity (known or not). For $\theta_0(x) \equiv 0$ we recover the so called Hammerstein model being extensively studied in the system identification literature (Bendat, 1990), (Billings, 1980), (Chen, 1995), (Greblicki and Pawlak, 1986), (Greblicki and Pawlak, 1994), (Hunter and Korenberg, 1986).

It can be easily verified that the above system is of the form (2.1), (2.2) with $\mu(x) = \theta(x) + E\theta_0(X) + E(\theta(X) + \theta_0(X))\sum_{j=1}^{p} h_j$ and $\xi_n = (\theta_0(X_{n-1}) - E\theta_0(X_{n-1})) + \sum_{j=1}^{p} h_j\gamma_{n-j}$, where $\gamma_j = \theta(X_{j-1})+\theta_0(X_{j-1})-E(\theta(X_{j-1})+\theta_0(X_{j-1}))$. Assumption 3 holds if $E\theta^2(X) < \infty$, $E\theta_0^2(X) < \infty$. Note also that if $\theta(0) = 0$ then $\theta(x) = \mu(x) - \mu(0)$ for all x. Hence the estimation of $\mu(x)$ is equivalent to estimation of $\theta(x)$.

The key observation for designing our identification algorithms is that

$$E\{Y_n|X_n = x\} = \mu(x), \tag{2.3}$$

i.e., that the system nonlinearity to be identified is equal to the standard regression function of the system output Y_n on the system input X_n. Thus by estimating the regression in (2.3) we recover the non-linearity $\mu(x)$.

The problem of estimation of the regression function from the input-output training data $\{(X_t, Y_t)\}$ when $\{(X_t, Y_t)\}$ is a sequence of independent and identically distributed (*iid*) random variables has been extensively studied in the statistical literature (Antoniadis and Oppenheim, 1995), (Härdle, 1990). In this paper it is assumed that the system is excited by the *iid* signal $\{X_t\}$ (Assumption 2), whereas $\{Y_t\}$ being an output of a nonlinear time-invariant dynamic system is a dependent stationary stochastic process which is in contrast to the papers cited above. Furthermore let us note that $\{Y_n\}$ is neither strictly stationary process nor mixing. The latter condition has been commonly used in publications on nonparametric estimation from dependent data (Härdle, 1990).

3. IDENTIFICATION ALGORITHMS

Due to the fundamental property established in (2.3) we can treat $\mu(x)$ as a standard regression function of Y_n on $X_n = x$. In order to construct an estimate of the regression function let us first observe that

$$\mu(x) = \frac{g(x)}{f(x)},$$

where $g(x) = \mu(x)f(x)$ for every x where the assumption (A1.2) holds. Owing to the assumptions (A1.1), (A3.2) and using the results of Section 1 we can approximate $g(x)$ and $f(x)$ at the resolution m as follows:

$$\begin{aligned} g_m(x) &= \sum_{k\in Z} a_{mk}\varphi_{mk}(x), \\ f_m(x) &= \sum_{k\in Z} c_{mk}\varphi_{mk}(x), \end{aligned} \tag{3.1}$$

where one can easily observe that

$$\begin{aligned} a_{mk} &= \int_{-\infty}^{\infty} \mu(x)\varphi_{mk}(x)f(x)dx \\ &= E\{Y_n\varphi_{mk}(X_n)\} \end{aligned}$$

and

$$c_{mk} = \int_{-\infty}^{\infty} \varphi_{mk}(x)f(x)dx = E\{\varphi_{mk}(X_n)\}.$$

Empirical counterparts of $g_m(x)$ and $f_m(x)$ in (3.1) can be easily constructed by replacing the expected values in the formulas for a_{mk} and c_{mk} by their natural estimates

$$\begin{aligned} \hat{a}_{mk} &= n^{-1}\sum_{i=1}^{n} Y_i\varphi_{mk}(X_i), \\ \hat{c}_{mk} &= n^{-1}\sum_{i=1}^{n} \varphi_{mk}(X_i). \end{aligned} \tag{3.2}$$

All these considerations yield the following initial estimate of $\mu(x)$ at the resolution m

$$\hat{\mu}_m(x) = \frac{\sum_{k\in Z} \hat{a}_{mk}\varphi_{mk}(x)}{\sum_{k\in Z} \hat{c}_{mk}\varphi_{mk}(x)} \tag{3.3}$$

It is worth noting that $\hat{a}_{mk}$, $\hat{c}_{mk}$ are unbiased estimators of a_{mk}, c_{mk}, i.e., $E\hat{a}_{mk} = a_{mk}$, $E\hat{c}_{mk} = c_{mk}$. Let us note that for compact supported father wavelet functions $\varphi(x)$ there is a finite number of terms in the sums in (3.3). If $\varphi(x)$ is not compact supported one has to truncate the sums in (3.3) to some finite limits.

A wavelet-based estimate of $\mu(x)$ can be proposed based on the representation in (1.2). Proceeding as in (3.1) and (3.2) we can define the following estimate

$$\tilde{\mu}(x) = \left(\sum_{k \in Z} \hat{a}_{m_0 k} \varphi_{m_0 k}(x) + \sum_{s=m_0}^{m_0+r} \sum_{k \in Z} \tilde{b}_{sk} \psi_{sk}(x) \right) \div \left(\sum_{k \in Z} \hat{c}_{m_0 k} \varphi_{m_0 k}(x) + \sum_{s=m_0}^{m_0+r} \sum_{k \in Z} \tilde{d}_{sk} \psi_{sk}(x) \right), \quad (3.4)$$

where $\hat{a}_{m_0 k}$, $\hat{c}_{m_0 k}$ are defined as in (3.2) and

$$\begin{aligned} \tilde{b}_{sk} &= n^{-1} \sum_{i=1}^{n} Y_i \psi_{sk}(X_i), \\ \tilde{d}_{sk} &= n^{-1} \sum_{i=1}^{n} \psi_{sk}(X_i) \end{aligned} \quad (3.5)$$

are estimates of the Fourier coefficients corresponding to the wavelet function.

The estimate in (3.4) has an advantage of being able to incorporate the *a prior* knowledge about $\mu(x)$.

4. CONVERGENCE ANALYSIS

The parameters m in (3.3) and r in (3.4) play important role in both asymptotic and finite sample size performance of the estimators. For the convergence property, i.e., that $\tilde{\mu}(x) \to \mu(x)$ as $n \to \infty$ in probability for almost all $x \in R$ it can be shown that the resolution level r must be chosen as a function of the sample size n, i.e., $r = r(n)$ in such a way that $r(n) \to \infty$ and $2^{r(n)}/n \to 0$ as $n \to \infty$. In order to establish the rate of convergence we need some smooth conditions on $\mu(x)$and $f(x)$. Hence let

$$\mu(x) \text{ and } f(x) \text{ possess } s \text{ derivatives} \quad (4.1)$$

We also need some conditions on the father wavelet function $\varphi(x)$.Let $K(x,y) = \sum_j \varphi(x-j)\varphi(y-j)$. The following conditions on $K(x,y)$ are required.

For an integer $S > 0$ we have

$$| K(x,y) | \leq F(x-y) \quad (4.2)$$

with $\int | x |^{S+1} F(x) \, dx < \infty$ and

$$\int (y-x)^l K(x,y) \, dy = 0 \quad (4.3)$$

for $l = 1, 2, \dots, S$.

Theorem 1. Let Assumptions 1–4 be satisfied. Let $\mu(x)$ and $f(x)$ satisfy (4.1). Let the father wavelet $\varphi(x)$ be compact supported and satisfy (4.2) and (4.3) with $F \in L_2(R)$. Let S be such that $s \leq S+1$.

If the resolution level $r(n)$ is selected as $r(n) \approx log_2(n)/(2s+1)$ then

$$E(\tilde{\mu}(x) - \mu(x))^2 = O\left(n^{-2s/(2s+1)}\right).$$

Theorem 1 gives the conditions for the mean squared error convergence of the identification algorithm. It is of considerable interest to establish conditions for the strong convergence, i.e., the convergence for which $\tilde{\mu}(x)$ converges to $\mu(x)$ with probability one.

It can be shown that if $r(n) \approx log_2^2(n)/(2s+1)$ then

$$\tilde{\mu}(x) = \mu(x) + O\left(n^{-s/(2s+1)} log(n)\right)$$

with probability one.

Theorem 1 holds for a class of compactly supported wavelet systems satisfying conditions (4.2), (4.3). As an example of non compact supported wavelets one can use a Coifman wavelet system (Mallat, 1998) of degree S. Hence if the following moment conditions are met

$$\int x^l \varphi(x) \, dx = 0 \text{ for l=1,\dots,S}$$

and

$$\int x^l \psi(x) \, dx = 0 \text{ for l=0,\dots,S.}$$

Under these conditions the rate of convergence as in Theorem 1 can be obtained provided that (4.1) holds with $s = S$. Let us finally note that the popular Haar wavelet system satisfies (4.2), (4.3) with $S = 1$. Hence in this case the rate obtained in Theorem 1 is of order $O(n^{-2/3})$ with $r(n)$ selected as $log_2(n)/3$.

The identification algorithm in (3.3) and (3.4) is in the form of the ratio of orthogonal series expansions. In (Greblicki and Pawlak, 1994) estimates based on classical orthogonal polynomials have been studied for a particular class of block-oriented models. In particular, it has been proved that the convergence holds if $\mu(x)$ is differentiable at x which is consistent with the well known fact that there are examples of continuous functions whose orthogonal series diverge. On the contrary the wavelet expansions converge for continuous functions and consequently they can be applied to a broader class of nonlinear characteristics. Nevertheless, the estimates (3.3) and (3.4) are not adaptive in the sense that they cannot reach an optimal rate of convergence for a large class of nonlinear characteristics (Antoniadis and Oppenheim, 1995). From practical point of view one would like to exclude those terms in the series in (3.4) which cannot be accurately estimated due to the existing distortions, i.e., the system noise ξ_n, the measurement noise ε_n and the sparsity of the input signal $\{X_n\}$. Considerations of this nature suggest the following modification of $\tilde{\mu}(x)$ using the concept of thresholding (Antoniadis and Oppenheim, 1995), (Mallat, 1998)

$$\check{\mu}(x) = \left(\sum_{k\in Z} \hat{a}_{m_0k}\varphi_{m_0k}(x) + \sum_{s=m_0}^{m_0+r} \sum_{k\in Z} t_{sk}(\tilde{b}_{sk})\psi_{sk}(x) \right) \div \left(\sum_{k\in Z} \hat{c}_{m_0k}\varphi_{m_0k}(x) + \sum_{s=m_0}^{m_0+r} \sum_{k\in Z} t_{sk}(\tilde{d}_{sk})\psi_{sk}(x) \right),$$

where $t_{sk}(x)$ is a threshold function such that $t_{sk}(x) = 0$ for $|x| \le \tau_{sk}$ and τ_{sk} is a threshold constant. For instance $t_{sk}(x) = x\mathbf{1}_{\{|x|\ge\tau_{sk}\}}(x)$ and $t_{sk}(x) = \max(x - \tau_{sk}, 0)\text{sign}(x)$ are two popular choices often referred to as soft and hard thresholding, respectively. The choice of thresholds τ_{sk} has been suggested as $c\sqrt{\log n/n}$, $c\sqrt{(s-m_0)/n}$ or $c\sqrt{s/n}$ for all (s,k), where $c > 0$ is suitable chosen constant. It is worth noting that these propositions have been derived for the case of independent data and they may be far to be optimal in our case. Yet another alternative is to keep (or delete) a whole resolution level of detail coefficients in (3.4). Hence $\sum_{s=m_0}^{m_0+r}\sum_{k\in Z}\tilde{b}_{sk}\psi_{sk}(x)$ in (3.4) is replaced by $\sum_{s=m_0}^{m_0+r} t_s(\sum_{k\in Z}\tilde{b}_{sk}\psi_{sk}(x))$, where $t_s(x)$ is some threshold type function.

Finally let us suggest a further extension of our identification algorithms based on the concept of multiple wavelets (Geronimo *et al.*, 1994) by introducing several scaling and wavelet functions. This approach offers further improvements in the accuracy of wavelet based estimation algorithms, i.e., one can estimate nonlinear characteristics with highly variable smoothness, e.g., with multiple discontinuities and other singular points.

5. REFERENCES

Antoniadis, A. and G. Oppenheim (1995). Wavelets and statistics. In: *Lecture Notes in Statistics*. Vol. 103. Springer-Verlag. New York.

Bendat, J. S. (1990). *Nonlinear System Analysis and Identification*. Wiley. New York.

Billings, S. A. (1980). Identification of nonlinear systems — a survey. *Proc. IEE* **127**, 272–285.

Chen, H. W. (1995). Modeling and identification of parallel nonlinear systems: structural classification and parameter estimation methods. *Proc. IEEE* **83**, 39–66.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM. Philadelphia.

Geronimo, J. S., D. P. Hardin and P. R. Massopust (1994). Fractal functions and wavelet expansions based on several scaling functions. *J. of Approximation Theory* **78**, 373–401.

Greblicki, W. (1994). Nonparametric identification of wiener systems by orthogonal series. *IEEE Transactions on Automatic Control* **10**, 2077–2086.

Greblicki, W. and M. Pawlak (1986). Identification of discrete hammerstein systems using kernel regression estimates. *IEEE Trans. Automatic Control* **31**, 74–77.

Greblicki, W. and M. Pawlak (1994). Dynamic system identification with order statistics. *IEEE Trans. Inform. Theory* **40**, 1474–1489.

Härdle, W. (1990). *Applied Nonparametric Regressionn*. Cambridge University Press. Cambridge, U.K.

Hunter, I. W. and M. J. Korenberg (1986). The identification of nonlinear biological systems: Wiener and hammerstein cascade models. *Biol. Cybern.* **55**, 135–144.

Juditsky, A., H. Hjalmaarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjoberg and Q. Zhang (1995). Nonlinear black-box models in system identification: mathematical foundations. *Automatica* **31**, 1725–1750.

Mallat, S. (1998). *A Wavelet Tour of Signal Processing*. Academic Press. San Diego.

Pawlak, M. and Z. Hasiewicz (1998). Nonlinear system identification by the haar muliresolution analysis. *IEEE Trans. Circuits and Systems* **45**, 945–961.

Sjoberg, J., Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Y. Glorennec, H. Hjalmaarsson and A. Juditsky (1995). Nonlinear black-box modeling in system identification: a unified overview. *Automatica* **31**, 1691–1724.

Sureshbabu, N. and J. A. Farrell (1999). Wavelet-based system identification for nonlinear control. *IEEE Trans. on Automatic Control* **44**, 412–417.

Vetterli, M. and J. Kovacevic (1995). *Wavelets and Subband Coding*. Prentice Hall.

Walter, G. (1994). *Wavelets and Other Orthogonal Systems with Applications*. CRC Press. Bocan Raton.