Nonparametric identification of nonlinearities in block-oriented systems by orthogonal wavelets with compact support

Zygmunt Hasiewicz, Mirosław Pawlak and Przemysław Śliwiński

Abstract

The paper addresses the problem of identification of nonlinear characteristics in a certain class of discrete-time block-oriented systems. The systems are driven by random stationary white processes (i.i.d. input sequences) and disturbed by stationary, white or coloured, random noise. The prior information about nonlinear characteristics is nonparametric. In order to construct identification algorithms the orthogonal wavelets of compact support are applied, and a class of wavelet-based models is introduced and examined. It is shown that under moderate assumptions the proposed models converge almost everywhere (in probability) to the identified nonlinear characteristics, irrespective of the noise model. The rule for optimum model size selection is given and the asymptotic rate of convergence of the model error is established. It is demonstrated that in some circumstances the wavelet models are, in particular, superior to classical trigonometric and Hermite orthogonal series models worked out earlier.

Index Terms

block-oriented systems, nonlinearity recovering, nonparametric approach, wavelet-based models.

AFFILIATIONS OF AUTHORS

Zygmunt Hasiewicz and **Przemysław Śliwiński**

Institute of Engineering Cybernetics

Wrocław University of Technology

Janiszewskiego 11/17

50-372 Wrocław, Poland

phone: +48 71 3203277

fax: +48 71 3212677

email: zhas@ict.pwr.wroc.pl and slk@ict.pwr.wroc.pl

Mirosław Pawlak

Department of Electrical and Computer Engineering

University of Manitoba

Winnipeg, MB, Canada R3T 2N2

phone: +1 204 4748881

fax: +1 204 2614639

email: pawlak@ee.umanitoba.ca

I. INTRODUCTION

We consider the problem of recovering nonlinear characteristics of a class of discrete-time block-oriented dynamical systems, i.e. structured objects where nonlinear static elements are separated from the rest of the system and embedded in a composite structure containing discrete-time linear dynamic blocks and other 'nuisance' nonlinearities ([1]). It is assumed that prior information about subsystems is small, and in particular that the nonlinearity of interest is merely a bounded function in the identification region. Since the nonlinear characteristics to be identified are not given in a parametric form, and no finite-dimensional parametric representation of a possible characteristic can be reasonably motivated, our problem is nonparametric and standard parametric identification methods are not applicable.

The problem of recovering static characteristics in interconnected composite systems has been extensively studied in the literature. For steady state systems it has been investigated in a number of papers using parametric and nonparametric approach (e.g. [2], [3]). For block-oriented dynamical systems, the problem has been initially examined under the assumption that the nonlinear characteristics are known up to the parameters and that are typically polynomials of known degree. Various parametric approaches have been developed, with particular attention applied to cascade and parallel systems. A thorough overview of the derived parametric identification methods can be found in [4].

A less demanding nonparametric approach, discarding the restrictive assumption of parametric prior knowledge of the characteristics, has been initiated in [5] for Hammerstein systems. This approach, which originated from nonparametric estimation of a regression function, was next developed bringing about a collection of papers where two types of nonparametric identification algorithms were studied: kernel algorithms (e.g. [5], [6]) and orthogonal series algorithms applying conventional orthogonal series expansions of characteristics (e.g. [7], [8]). The algorithms were elaborated for Hammerstein and Wiener systems, i.e. cascade connections of static nonlinearities and linear dynamic blocks in a suitable order.

In this paper, we propose and examine a class of nonparametric identification algorithms based on wavelet approximations of functions. The algorithms exploit only input-output measurement data collected in the experiment and apply to built of the wavelet models the orthogonal wavelets with compact supports. There are four main reasons for using orthogonal compactly supported wavelets for the construction of the identification algorithms:

1. ease of obtaining the orthogonal wavelet bases by only scaling and translating of a father and mother wavelet,

2. extraordinary approximation capability and adaptivity associated with the ability of shrinking of the

wavelets domain to any small interval,

3. ease of obtaining models of increasing precision and sensitivity to the characteristic details, and

4. existence of fast algorithms for wavelet computations and ready-for-use program packages.

Orthogonality of wavelets enables simple and convenient improvement of the wavelet models by adding standard building blocks and compactness of the supports results in parsimonious representations of nonlinear characteristics, with good localization properties. For reconstruction of nonlinearities the models use only the most informative 'local' data lying in a close neighbourhood of a point at which the estimation is carried out, and hence can be useful in exploring local details in the nonlinear characteristics, and for practical implementation need only a number of coefficients, being merely a small fraction of the whole number of measurements employed for the identification. In this sense, they combine the advantages of kernel and orthogonal series algorithms worked out earlier (see the papers cited above). The present paper refers to the nonparametric identification algorithms based on orthogonal series expansions of functions, and is a generalization of [9] and [10] where only the father wavelets (scaling functions) of compact support have been applied in the identification algorithms, towards application of multiscale approximations which employ both father and mother wavelets and hence fully realize the multiresolution idea.

As in the latter papers, we assume that the identified systems operate in stochastic conditions. By assumption, the systems under consideration are excited by stationary white random processes (i.i.d. random input sequences) for which there exist probability density functions (unknown in our approach). The external stationary random noises acting on the systems can be white or coloured, with arbitrary correlation structure. Owing to the zooming capability of wavelets, we focus on the analysis of local**,** pointwise properties of the wavelet models. We show that under weak assumptions concerning the underlying nonlinearities and input probability density functions, the wavelet-based models provide consistent estimates of nonlinear characteristics which rapidly converge to the true nonlinear functions, regardless of whether the disturbing noise is white or coloured. We demonstrate in particular that the rates of convergence can approach the best possible convergence rates for nonparametric inference, established in [11], and that the wavelet models may offer better rates than, for example, classical trigonometric and Hermite orthogonal series models elaborated earlier. To the best of our knowledge, such an analysis in the system identification context has not yet been presented in the literature. Despite numerous successful applications of wavelets in many fields of engineering, for instance in signal processing, image analysis, communication systems, to quote but a few (see e.g. [12], [13] and the references cited therein), there have not been many papers to date concerning wavelets employment in system identification. Actually,

the collection of contributions in this area does not significantly exceed [14], [15], [16], [17], [18] and [9], [10].

The paper is organized as follows. Section II presents the class of systems under consideration and states the system identification task. Section III gives examples of popular block-oriented systems belonging to the studied class. The general wavelet-based identification algorithm of nonlinear characteristics is constructed in Section IV. In Section V, we consider convergence and rate of convergence of the wavelet models. The conditions ensuring weak pointwise consistency are provided and the guidelines for optimum model size selection (optimizing the convergence speed) are given. Then the rate of convergence of optimum models is studied. We also propose a simplified sub-optimal rule for selection of the size of the wavelet model, recommended for the use when prior knowledge about identified nonlinearity and input density is too small to apply the optimum size selection rule. The results of computer simulations are reported in Section VI, and conclusions in Section VII complete the paper. For the sake of convenience, pertinent facts concerning approximation of functions by using orthogonal wavelets with compact support are collected in Appendix I. Appendices II-V contain technical derivations and lemmas.

II. Problem formulation

We consider a class of discrete-time block-oriented systems with input-output behavior that can be expressed by the equation

$$y_k = R(x_k) + \xi(x_{k-1}, x_{k-2}, \ldots) + z_k \tag{1}$$

where $R(x)$ is a static nonlinearity to be identified, $\{(x_k, y_k)\}$ is the (scalar input, scalar output) sequence, $\{\xi_k = \xi(x_{k-1}, x_{k-2}, \ldots)\}$ is the 'system noise' induced by the system dynamics, and $\{z_k\}$ is the external disturbance (Fig. 1). We assume that the system in (1) is asymptotically stable and operates in steady state. Moreover, the following conditions are imposed.

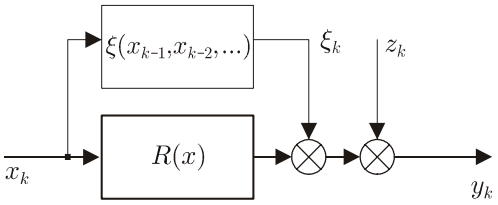

Fig. 1. General system under consideration.

A1. The nonlinearity $R(x)$ is such that $|R(x)| \leq C_{1R}|x| + C_{2R}$, some C_{1R}, $C_{2R} > 0$.

A2. The input process $\{x_k\}_{k\in\mathsf{Z}}$ (Z - the set of integers) is a sequence of independent and identically distributed (i.i.d.) random variables with finite variance σ_x^2, and there exists a probability density function $f(x)$ such that $f(x) \leq M_f < \infty$, some $M_f > 0$. The density $f(x)$ and the variance σ_x^2 are unknown.

A3. The system noise $\{\xi_k\}_{k\in\mathsf{Z}}$ is of the form

$$\xi_k = \xi_{1k} + \xi_{2k} + ... + \xi_{Jk} \tag{2}$$

where

$$\xi_{jk} = \sum_{i=1}^{\infty} \lambda_{ji}\, \gamma_j(x_{k-i}) \tag{3}$$

and where for $j = 1, 2, ..., J$ it holds that $\sum_{i=1}^{\infty} |\lambda_{ji}| < \infty$ and $\gamma_j(x)$ is a nonlinear function such that $\mathsf{E}\,\gamma_j(x_1) = 0$ and $\mathsf{E}\,\gamma_j^2(x_1) = \sigma_j^2 < \infty$. The impulse responses $\{\lambda_{ji}\}$ and functions $\gamma_j(x)$ are not known.

A4. The external noise $\{z_k\}_{k\in\mathsf{Z}}$ is, in general, a correlated process generated from stationary white noise processes $\{\varepsilon_{lk}\}_{k\in\mathsf{Z}}$, $l = 1, 2, ..., L$, with $\mathsf{E}\,\varepsilon_{l1} = 0$ and $\mathsf{E}\,\varepsilon_{l1}^2 = \sigma_{l\varepsilon}^2 < \infty$, such that

$$z_k = z_{1k} + z_{2k} + ... + z_{Lk} \tag{4}$$

where

$$z_{lk} = \sum_{i=0}^{\infty} \omega_{li}\varepsilon_{l,k-i} \tag{5}$$

and $\sum_{i=0}^{\infty} |\omega_{li}| < \infty$ for $l = 1, 2, ..., L$. The impulse responses $\{\omega_{li}\}$ and the noise variances $\sigma_{l\varepsilon}^2$ are unknown.

A5. Processes $\{x_k\}$ and $\{\varepsilon_{lk}\}$, $l = 1, 2, ..., L$, are mutually statistically independent.

A6. Only the external (input, output) measurement data $\{(x_k, y_k)\}$ are available and $R(x)$ should be identified in the bounded region $[a, b]$.

A7. For $x \in [a, b]$, it holds that

$$0 < \delta \leq f(x) \tag{6}$$

some $\delta > 0$.

The aim is to recover $R(x)$ in the region $[a, b]$ using solely the input-output data $\{(x_k, y_k)\}_{k=1}^{N}$.

We note that Assumptions A1-A7 are rather weak and represent realistic system identification conditions. First, prior knowledge about $R(x)$ is poor. We merely presume that $|R(x)|$ does not grow with $|x|$ faster than linearly (Assumption A1), which yields a broad class of admissible nonlinearities. As such nonlinearities cannot be parameterized, our identification problem is nonparametric. Second, the system

input can be any i.i.d. random process possessing a bounded and nonvanishing probability density function in the identification region, with finite variance (Assumptions A2 and A7). Third, the external noise can be correlated and can be any type of coloured noise (Assumption A4). Moreover, the disturbances have not to be bounded. Assumption A6 reflects standard practical situations where (i) the inner signals in the block-oriented systems are not accessible for measurements (see, e.g., [1], [4]) and (ii) the nonlinearity $R(x)$ is required to be known only in some bounded region. We observe for further use that under Assumptions A1-A4 the system output $\{y_k\}$ is a correlated second-order stationary random process of finite variance.

III. EXAMPLES OF SYSTEMS

Although the class of systems in (1)-(5) looks rather specific, there are quite a lot of exemplary block-oriented structures which conform to this description and are often considered in the literature. In the following examples (x), $\eta(x)$ and (x) are static nonlinearities, $\{\lambda_i\}$, $\{\gamma_i\}$ and $\{\ _i\}$ denote impulse responses of linear dynamical elements such that $\sum_{i=0}^{\infty}|\lambda_i|<\infty$, $\sum_{i=0}^{\infty}|\gamma_i|<\infty$, $\sum_{i=0}^{\infty}|\ _i|<\infty$, and z_k is coloured noise produced by linear noise filter $\{\omega_i\}$ from a stationary white noise process $\{\varepsilon_k\}$ according to the equation $z_k=\sum_{i=0}^{\infty}\omega_i\varepsilon_{k-i}$ where $\sum_{i=0}^{\infty}|\omega_i|<\infty$, $\mathsf{E}\,\varepsilon_1=0$, $\mathsf{E}\,\varepsilon_1^2=\ _{\varepsilon}^2<\infty$ and $\{\varepsilon_k\}$ is independent of the input $\{x_k\}$. Moreover, we denote $d\ _{\lambda}=\mathsf{E}\ \ (x_1)\sum_{i=1}^{\infty}\lambda_i$, $d_{\eta\lambda}=\mathsf{E}\,\eta(x_1)\sum_{i=1}^{\infty}\lambda_i$, $d_{\eta\gamma}=\mathsf{E}\,\eta(x_1)\sum_{i=1}^{\infty}\gamma_i$ and $_0(x)=\ \ (x)-\mathsf{E}\ \ (x_1)$, $\eta_0(x)=\eta(x)-\mathsf{E}\,\eta(x_1)$.

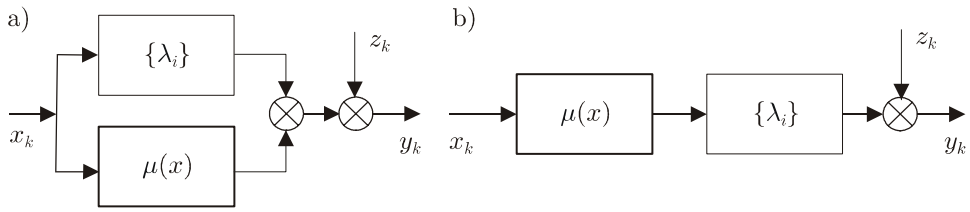

Fig. 2. **a)** Parallel system, **b)** Hammerstein system.

Parallel system [19]: For a parallel system as in Fig. 2a and $\lambda_0=0$ (one step delay), we get

$$y_k=\ \ (x_k)+d+\sum_{i=1}^{\infty}\lambda_i\,(x_{k-i}-\mathsf{E}\,x_1)+z_k$$

where $d=\mathsf{E}\,x_1\sum_{i=1}^{\infty}\lambda_i$ which maps (1)-(5) for $R(x)=\ \ (x)+d$ and $J=L=1$ with $\lambda_{1i}=\lambda_i$, $\omega_{1i}=\omega_i$, $\varepsilon_{1k}=\varepsilon_k$ and $_1(x)=x-\mathsf{E}\,x_1$, provided that $\mathsf{E}\,x_1^2<\infty$ (Assumption A2).

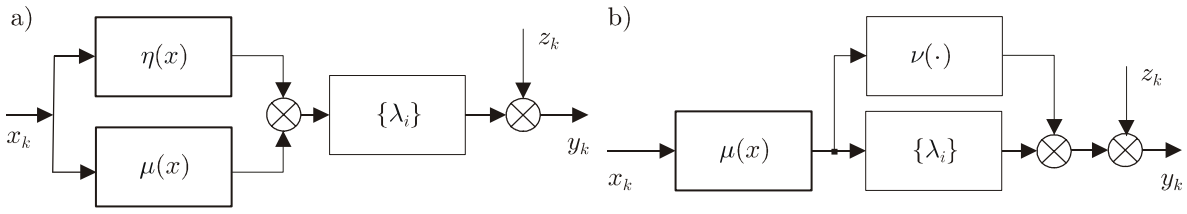

Fig. 3. **a)** Hammerstein system with two-segment nonlinearity, **b)** Hammerstein system with nuisance nonlinearity.

Hammerstein system [7]: For the Hammerstein system (Fig. 2b) we have

$$y_k = \lambda_0 \ (x_k) + d \ _\lambda + \sum_{i=1}^{\infty} \lambda_i \ _0(x_{k-i}) + z_k$$

This gives (1)-(5) for $R(x) = \lambda_0 \ (x) + d \ _\lambda$ and $J = L = 1$ with the same specifications as above and $_1(x) = \ _0(x)$, provided that $\mathsf{E} \ ^2(x_1) < \infty$ (Assumptions A1 and A2).

Hammerstein system with two-segment nonlinearity [20]: For a Hammerstein system with two-segment nonlinearity (Fig. 3a) we obtain

$$y_k = \lambda_0[\ (x_k) + \eta(x_k)] + (d \ _\lambda + d_{\eta\lambda}) + \sum_{i=1}^{\infty} \lambda_i \ _0(x_{k-i}) + \sum_{i=1}^{\infty} \lambda_i \eta_0(x_{k-i}) + z_k$$

This falls into the description (1)-(5) for $R(x) = \lambda_0[\ (x) + \eta(x)] + d$ where $d = d \ _\lambda + d_{\eta\lambda}$, $J = 2$ with $\lambda_{1i} = \lambda_{2i} = \lambda_i$, $_1(x) = \ _0(x)$, $_2(x) = \eta_0(x)$ and $L = 1$ with the noise filter defined above, provided that $\mathsf{E} \ ^2(x_1) < \infty$, $\mathsf{E}\eta^2(x_1) < \infty$ (Assumptions A1 and A2).

Hammerstein system with nuisance nonlinearity: For a Hammerstein system with nuisance nonlinearity $(.)$ (Fig. 3b) it holds that

$$y_k = \lambda_0 \ (x_k) + \eta(x_k) + d \ _\lambda + \sum_{i=1}^{\infty} \lambda_i \ _0(x_{k-i}) + z_k$$

where $\eta(x) = \ (\ (x))$. This conforms with the description (1)-(5) for $R(x) = \lambda_0 \ (x) + \eta(x) + d \ _\lambda$ and further specifications as for the Hammerstein system.

Hammerstein system with nuisance dynamics: For a Hammerstein system with nuisance dynamics $\{ \ _i\}$ (Fig. 4a) and $_0 = 0$ we have

$$y_k = \lambda_0 \ (x_k) + d + \sum_{i=1}^{\infty} \lambda_i \ _0(x_{k-i}) + \sum_{i=1}^{\infty} \gamma_i (x_{k-i} - \mathsf{E}\, x_1) + z_k$$

where $d = d \ _\lambda + \mathsf{E}\, x_1 \sum_{i=1}^{\infty} \gamma_i$ and $\gamma_i = \sum_{p=0}^{\infty} \lambda_p \ _{i-p}$. This fits (1)-(5) for $R(x) = \lambda_0 \ (x) + d$, $J = 2$ with $\lambda_{1i} = \lambda_i$, $_1(x) = \ _0(x)$, $\lambda_{2i} = \gamma_i$, $_2(x) = x - \mathsf{E}\, x_1$ (provided that $\mathsf{E} \ ^2(x_1) < \infty$ and $\mathsf{E}\, x_1^2 < \infty$), and $L = 1$ with the previous noise model.

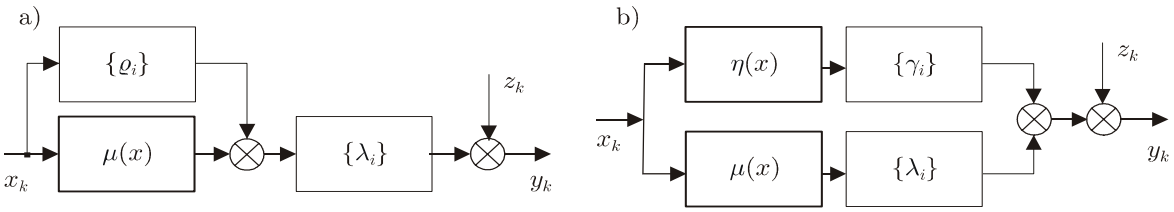

Fig. 4. **a)** Hammerstein system with nuisance dynamics, **b)** Uryson system.

Uryson system [4]: For the Uryson system in Fig. 4b we obtain

$$y_k = \lambda_0 \quad (x_k) + \gamma_0 \eta(x_k) + d + \sum_{i=1}^{\infty} \lambda_i \quad _0(x_{k-i}) + \sum_{i=1}^{\infty} \gamma_i \eta_0(x_{k-i}) + z_k$$

where $d = d \quad _\lambda + d_{\eta\gamma}$. This fulfils (1)-(5) for $R(x) = \lambda_0 \quad (x) + \gamma_0 \eta(x) + d$, $J = 2$, $L = 1$ and obvious further specifications.

Two-channel system [21]: For a two-channel system (Fig. 5a) with stationary white inputs $\{x_k\}$ and $\{u_k\}$ (mutually independent and independent of the noise $\{\varepsilon_k\}$) we get

$$y_k = \lambda_0 \quad (x_k) + d + \sum_{i=1}^{\infty} \lambda_i \quad _0(x_{k-i}) + \sum_{i=0}^{\infty} \gamma_i \left(\eta(u_{k-i}) - \mathsf{E}\,\eta(u_1)\right) + z_k$$

where $d = d \quad _\lambda + \mathsf{E}\,\eta(u_1) \sum_{i=0}^{\infty} \gamma_i$. This can be put in the form of (1)-(5) for $R(x) = \lambda_0 \quad (x) + d$, $J = 1$ with $\lambda_{1i} = \lambda_i$, $\quad _1(x) = \quad _0(x)$ and $L = 2$ with $\omega_{1i} = \gamma_i$, $\varepsilon_{1k} = \eta(u_k) - \mathsf{E}\,\eta(u_1)$, $\omega_{2i} = \omega_i$, $\varepsilon_{2k} = \varepsilon_k$, provided that $\mathsf{E} \quad ^2(x_1) < \infty$ and $\mathsf{E}\,\eta^2(u_1) < \infty$. The last two examples can be easily generalized.

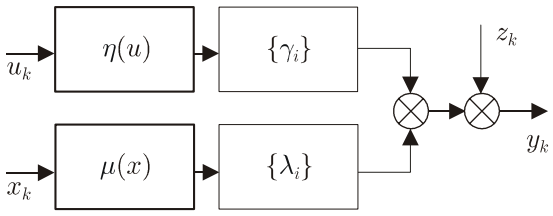

Fig. 5. Two-channel system.

The above systems can be found in diverse areas**,** as for example nonlinear control, telecommunication, acoustics, signal processing, chemical engineering, or biomedical engineering. A comprehensive bibliography concerning the practical applications of such and other structures can be found in [22]. We also refer to [4, Chapter 7] and the references cited therein for some specific case studies.

The presented examples reveal that the nonlinearity $R(x)$ in equation (1) can generally differ from any true nonlinear characteristic existing in a system. However, as it follows from (1), only $R(x)$ of such a type as in the examples can be attempted to be identified from just external input-output observations $\{(x_k, y_k)\}$, unless some additional structural conditions are fulfilled. For instance, for the Hammerstein system we gain $R(x) = \quad (x)$ if $\lambda_0 = 1$ and $\mathsf{E} \quad (x_1) = 0$ where the latter is fulfilled if, e.g., the system nonlinearity $\quad (x)$ is an odd and the input density $f(x)$ is an even function. This limitation, following from the composite structure of block-oriented systems, is well realized in the system identification literature, cf. [1], [7], [4].

IV. Identification algorithm

We shall construct a class of algorithms for the identification of the nonlinearity $R(x)$ in equation (1), based on orthogonal wavelets with compact support. The motivation for the use of compactly supported orthogonal wavelets for the construction of the identification algorithms are their beneficial properties quoted in Section I and in more detail presented in Appendix I. These properties announce the flexibility, precision, and good local behaviour of the wavelet models, along with their parsimony and ease of calculation.

Denote $g(x) = R(x) f(x)$ where $R(x)$ is the nonlinear characteristic to be identified and $f(x)$ is the probability density function of the input signal. Under Assumptions A1 and A2, the functions $g(x)$ and $f(x)$ are square integrable, $g(x), f(x) \in L^2(\mathsf{R})$, as

$$\int_{-\infty}^{+\infty} g^2(x)\,dx \quad \leq \quad M_f \int_{-\infty}^{+\infty} R^2(x)\, f(x) dx \leq 2M_f C_R^2 (\ {}^2_x + m_x^2 + 1) < \infty$$

$$\int_{-\infty}^{+\infty} f^2(x)\,dx \quad \leq \quad M_f \int_{-\infty}^{+\infty} f(x) dx = M_f < \infty$$

where $C_R = \max\{C_{1R}, C_{2R}\}$, $m_x = E\, x_1$, ${}^2_x = var x_1$. Hence we ascertain the following.

1. Both $g(x)$ and $f(x)$ can be approximated by the wavelet-based models as in (39) in Appendix I:

$$g(x; K) = \sum_{n=n_{\min}(\varphi,x,M)}^{n_{\max}(\varphi,x,M)} \alpha^g_{Mn} \varphi_{Mn}(x) + \sum_{m=M}^{K-1} \sum_{n=n_{\min}(\ ,x,m)}^{n_{\max}(\ ,x,m)} {}^g_{mn} \quad _{mn}(x) \tag{7}$$

$$f(x; K) = \sum_{n=n_{\min}(\varphi,x,M)}^{n_{\max}(\varphi,x,M)} \alpha^f_{Mn} \varphi_{Mn}(x) + \sum_{m=M}^{K-1} \sum_{n=n_{\min}(\ ,x,m)}^{n_{\max}(\ ,x,m)} {}^f_{mn} \quad _{mn}(x) \tag{8}$$

with the coefficients α^g_{Mn}, ${}^g_{mn}$, α^f_{Mn}, ${}^f_{mn}$ as in (40) and the summation limits $n_{\min}(\varphi, x, M)$, $n_{\max}(\varphi, x, M)$, $n_{\min}(\ , x, m)$, $n_{\max}(\ , x, m)$ as in (41).

2. Since $f(x)$ is a probability density function, the α_{Mn}^{g}'s, β_{mn}^{g}'s, α_{Mn}^{f}'s and β_{mn}^{f}'s in (7) and (8) are the following expectations (see (35) and (40) in Appendix I):

$$\alpha_{Mn}^{g} = E[R(x_1)\varphi_{Mn}(x_1)] \; ; \qquad \beta_{mn}^{g} = E[R(x_1)\psi_{mn}(x_1)] \tag{9}$$

$$\alpha_{Mn}^{f} = E[\varphi_{Mn}(x_1)] \; ; \qquad \beta_{mn}^{f} = E[\psi_{mn}(x_1)] \tag{10}$$

By virtue of (1) and Assumptions A2-A5 it further holds that

$$E[R(x_1)\varphi_{Mn}(x_1)] = E[y_1\varphi_{Mn}(x_1)] \tag{11}$$

$$E[R(x_1)\psi_{mn}(x_1)] = E[y_1\psi_{mn}(x_1)] \tag{12}$$

3. The approximators $g(x;K)$ and $f(x;K)$, when substituted in the relationship $R(x) = g(x)/f(x)$ for the exact $g(x)$ and $f(x)$ for x where $f(x) > 0$, yield a wavelet-based model of $R(x)$:

$$R(x;K) = g(x;K)/f(x;K) \tag{13}$$

for each point x where $f(x;K) \neq 0$.

Based on these observations and applying the plug-in routine, we get the following three-step algorithm to identify the nonlinearity $R(x)$ from the empirical measurement data $\{(x_k, y_k)\}_{k=1}^{N}$.

Step 1: Compute the sample means (see (9)-(12))

$$\hat{\alpha}_{Mn}^{g} = N^{-1}\sum_{k=1}^{N} y_k\varphi_{Mn}(x_k) \; ; \qquad \hat{\beta}_{mn}^{g} = N^{-1}\sum_{k=1}^{N} y_k\psi_{mn}(x_k) \tag{14}$$

and

$$\hat{\alpha}_{Mn}^{f} = N^{-1}\sum_{k=1}^{N} \varphi_{Mn}(x_k) \; ; \qquad \hat{\beta}_{mn}^{f} = N^{-1}\sum_{k=1}^{N} \psi_{mn}(x_k) \tag{15}$$

Step 2: Plug in $\hat{\alpha}_{Mn}^{g}$, $\hat{\beta}_{mn}^{g}$, $\hat{\alpha}_{Mn}^{f}$, $\hat{\beta}_{mn}^{f}$ into the approximators in (7) and (8) obtaining the empirical wavelet-based models of $g(x)$ and $f(x)$:

$$\hat{g}(x;K) = \sum_{n=n_{\min}(\varphi,x,M)}^{n_{\max}(\varphi,x,M)} \hat{\alpha}_{Mn}^{g}\varphi_{Mn}(x) + \sum_{m=M}^{K-1}\sum_{n=n_{\min}(\psi,x,m)}^{n_{\max}(\psi,x,m)} \hat{\beta}_{mn}^{g}\psi_{mn}(x) \tag{16}$$

and

$$\hat{f}(x;K) = \sum_{n=n_{\min}(\varphi,x,M)}^{n_{\max}(\varphi,x,M)} \hat{\alpha}_{Mn}^{f}\varphi_{Mn}(x) + \sum_{m=M}^{K-1}\sum_{n=n_{\min}(\psi,x,m)}^{n_{\max}(\psi,x,m)} \hat{\beta}_{mn}^{f}\psi_{mn}(x) \tag{17}$$

Step 3: Put the models $\hat{g}(x;K)$ and $\hat{f}(x;K)$ in (13), resulting in the empirical wavelet-based model of $R(x)$:

$$\hat{R}(x;K) = \hat{g}(x;K)/\hat{f}(x;K) \tag{18}$$

for each point x where $\hat{f}(x;K) \neq 0$.

Remark 1: Due to the fractional form of the model $\hat{R}(x;K)$ and compactness of the supports of the applied wavelet functions, the coefficients $\hat{\alpha}^g_{Mn}$, $\hat{}^g_{mn}$, $\hat{\alpha}^f_{Mn}$, $\hat{}^f_{mn}$ can be calculated using simplified rules

$$\hat{\alpha}^g_{Mn} = 2^{M/2} \sum_{\{k:u_{Mn,k}\in[s_1,s_2]\}} y_k \varphi(u_{Mn,k}) \quad \text{and} \quad \hat{}^g_{mn} = 2^{m/2} \sum_{\{k:u_{mn,k}\in[t_1,t_2]\}} y_k \quad (u_{mn,k}) \tag{19}$$

and

$$\hat{\alpha}^f_{Mn} = 2^{M/2} \sum_{\{k:u_{Mn,k}\in[s_1,s_2]\}} \varphi(u_{Mn,k}) \quad \text{and} \quad \hat{}^f_{mn} = 2^{m/2} \sum_{\{k:u_{mn,k}\in[t_1,t_2]\}} \quad (u_{mn,k}) \tag{20}$$

where $u_{mn,k} = 2^m x_k - n$ and $[s_1, s_2]$ and $[t_1, t_2]$ is, respectively, the support of the father and mother wavelet $\varphi(x)$ and (x) (cf. (36) and (37) in Appendix I).

Remark 2: Due to compactness of the support of $\varphi(x)$ and (x) there is a finite number of terms in the sums in (16) and (17), and for $R(x)$ to be estimated over the interval $[a, b]$ the whole number of the required wavelet coefficients $\hat{\alpha}^g_{Mn}$, $\hat{}^g_{mn}$, $\hat{\alpha}^f_{Mn}$, $\hat{}^f_{mn}$ does not exceed

$$\overline{L}(b-a, M, K) = 2[S(K - M + 1) + 2^K \lceil b - a \rceil] \tag{21}$$

where $S = \lfloor s \rfloor + 1$ and s is the support size of $\varphi(x)$ and (x) (see (41) and (42) in Appendix I).

Example 1: Consider the version of the wavelet identification algorithm (14)-(18), obtained for the Haar wavelets. In such a case $\varphi(x) = I_{[0,1]}(x)$ and $(x) = I_{[0,1]}(x) sgn(1/2 - \rfloor x \lfloor)$ and hence

$$\varphi_{Mn}(x) = 2^{M/2} I_{[n/2^M,(n+1)/2^M]}(x)$$

$$_{mn}(x) = 2^{m/2} I_{[n/2^m,(n+1)/2^m]}(x) sgn(1/2 - \rfloor 2^m x - n \lfloor)$$

where $\rfloor v \lfloor = v - \lfloor v \rfloor$ stands for the fractional part of v. By including that $\rfloor v - \lfloor v \rfloor \lfloor = \rfloor v \lfloor$ and that

$$n_{\min}(\varphi, x, M) = n_{\max}(\varphi, x, M) = \lfloor 2^M x \rfloor \ , \ n(x, M)$$

$$n_{\min}(\ , x, m) = n_{\max}(\ , x, m) = \lfloor 2^m x \rfloor \ , \ n(x, m)$$

we get the models (see (16) and (17))

$$\hat{g}(x;K) = \hat{\alpha}^g_{Mn(x,M)} + \sum_{m=M}^{K-1} \hat{}^g_{mn(x,m)} sgn(1/2 - \rfloor 2^m x \lfloor)$$

$$\hat{f}(x;K) = \hat{\alpha}^f_{Mn(x,M)} + \sum_{m=M}^{K-1} \hat{}^f_{mn(x,m)} sgn(1/2 - \rfloor 2^m x \lfloor)$$

where the wavelet coefficients are computed from the measurement data $\{(x_k, y_k)\}$ as follows (cf. (19), (20))

$$\hat{\alpha}^g_{Mn(x,M)} = 2^M \sum_{\{k:u_{Mn(x,M),k}\in[0,1]\}} y_k \quad \text{and} \quad \hat{}^{\,g}_{mn(x,m)} = 2^m \left[\sum_{\{k:u_{mn(x,m),k}\in[0,1/2)\}} y_k - \sum_{\{k:u_{mn(x,m),k}\in[1/2,1]\}} y_k \right]$$

and

$$\begin{aligned} \hat{\alpha}^f_{Mn(x,M)} &= 2^M \,\#\left\{k : u_{Mn(x,M),k} \in [0,1]\right\} \\ \hat{}^{\,f}_{mn(x,m)} &= 2^m \big[\#\left\{k : u_{mn(x,m),k} \in [0,1/2)\right\} - \#\left\{k : u_{mn(x,m),k} \in [1/2,1]\right\}\big] \end{aligned}$$

where $u_{Mn(x,M),k} = 2^M x_k - n(x,M)$, $u_{mn(x,m),k} = 2^m x_k - n(x,m)$, i.e. only elementary calculations are needed. Finally, it remains to compute the ratio as in (18).

We note that computations required by the algorithm are the same for white and coloured noise $\{z_k\}$ (see Assumption A4 in Section II). This stands in contrast to parametric methods where a correlation of the disturbances acting on the system usually requires a substantial modification of identification routines and leads to far more demanding computation procedures than for the white noise case; see, e.g., [23], [4] for appropriate examples concerning parametric identification techniques.

V. Convergence analysis

We shall prove that the empirical wavelet models $\hat{g}(x;K)$, $\hat{f}(x;K)$ and $\hat{R}(x;K)$ converge, in a probabilistic sense, to the true functions $g(x)$, $f(x)$ and $R(x)$ if the number N of data points (x_k, y_k) grows large and the scale factor K is appropriately fitted to the number of data, and establish asymptotic rate of convergence. To this end we evaluate the variance error of the empirical wavelet coefficients $\hat{\alpha}^g_{Mn}$, $\hat{}^{\,g}_{mn}$, $\hat{\alpha}^f_{Mn}$, $\hat{}^{\,f}_{mn}$ and then of the wavelet models $\hat{g}(x;K)$, $\hat{f}(x;K)$, and calculate the approximation error of the approximators $g(x;K)$ and $f(x;K)$. All technical derivations (rather involved) are presented in Appendices II, III and IV.

A. Convergence of the empirical wavelet models

For the sake of conciseness, let $(\hat{F}(x;K), F(x;K), F(x))$ denote a generic element of the set $\{(\hat{g}(x;K), g(x;K), g(x)), (\hat{f}(x;K), f(x;K), f(x))\}$. According to (55), (60) and (61) in Appendix III, for the model $\hat{F}(x;K)$ it holds that $E\hat{F}(x;K) = F(x;K)$ and for each $x \in [a,b]$ and each fixed scale factor K the variance error is bounded by

$$var\left\{\hat{F}(x;K)\right\} \le (C_F 2^K) N^{-1} \tag{22}$$

some $C_F > 0$ independent of K and N. Consequently, the following bound of the mean square error of $\hat{F}(x;K) \in \{\hat{g}(x;K), \hat{f}(x;K)\}$ is valid

$$E\left[\hat{F}(x;K) - F(x)\right]^2 \leq AE^2(F;x;K) + (C_F 2^K)N^{-1} \tag{23}$$

where $AE(F;x;K) = F(x) - F(x;K)$ is the approximation error. This bound and (45) in Appendix I show that certainly the choice of the parameter K plays a crucial role for the asymptotic behaviour of the model. Values of K which are too small may result in large approximation error (cf. (44) in Appendix I) whereas K values which are too large cause an increase in variance. One can easily ascertain that if the scale factor (resolution level) K is taken as a function of the sample size $K = K(N)$ and properly adjusted to the number N of measurement data $\{(x_k, y_k)\}_{k=1}^N$, then the models $\hat{g}(x;K)$, $\hat{f}(x;K)$ and $\hat{R}(x;K)$ converge to $g(x)$, $f(x)$ and $R(x)$ as $N \to \infty$. The following theorem holds.

Theorem 1: Let the assumptions of Section II be in force. If $K = K(N)$ and

$$K(N) \to \infty, \ 2^{K(N)}/N \to 0 \quad \text{as} \quad N \to \infty \tag{24}$$

then $\hat{g}(x;K(N)) \to g(x)$ and $\hat{f}(x;K(N)) \to f(x)$ in mean square, and $\hat{R}(x;K(N)) \to R(x)$ in probability as $N \to \infty$ for almost all $x \in [a,b]$.

Proof: For $\hat{g}(x;K(N))$ and $\hat{f}(x;K(N))$ the conclusion is obvious, including (23) and (45) in Appendix I. For $\hat{R}(x;K(N))$ the convergence follows from the convergence of $\hat{g}(x;K(N))$ and $\hat{f}(x;K(N))$, fractional form of the model $\hat{R}(x;K(N))$ (cf. (18)), and the fact that $g(x) = R(x)f(x)$ and $f(x) > 0$ on $[a,b]$ (Assumption A7). ■

Remark 3: Under weak assumptions as in Section II, from the mean square consistency of the models $\hat{g}(x;K(N))$ and $\hat{f}(x;K(N))$ we can only infer weak consistency of the ratio model $\hat{R}(x;K(N)) = \hat{g}(x;K(N))/\hat{f}(x;K(N))$, i.e. that $P\left\{\left|\hat{R}(x;K(N)) - R(x)\right| > \varepsilon\right\} \to 0$ as $N \to \infty$ for each $\varepsilon > 0$. However, as it follows from the Markov inequality and the Lebesgue dominated convergence theorem, convergence in probability is equivalent to more intuitive convergence of $\hat{R}(x;K(N))$ in mean, i.e. $E\left|\hat{R}(x;K(N)) - R(x)\right| \to 0$ as $N \to \infty$, provided that for the model $\hat{R}(x;K(N))$ it holds that $E\left\{\sup_N \left|\hat{R}(x;K(N))\right|\right\} \leq C_R < \infty$ almost everywhere on $[a,b]$, some $C_R > 0$. This is achieved if $E\{\sup_N |\hat{g}(x;K(N))|\} \leq C_g < \infty$ and $\left|\hat{f}(x;K(N))\right| \geq \delta_f$ (compare (6) in Assumption A7) almost everywhere over $[a,b]$, some $C_g, \delta_f > 0$ each N. The former is obtained under more restrictive demand that the overall noise corrupting the system (1) is bounded (see (1), Assumption A1, (14), (16) and (38), (42) in Appendix I), and the latter is ensured by a slight modification of the density model $\hat{f}(x;K(N))$ to the form $\hat{f}_{\text{mod}}(x;K(N)) = \hat{f}(x;K(N))I_{\{x:\left|\hat{f}(x;K(N))\right| \geq \delta_f\}}(x) + \delta_f \, I_{\{x:\left|\hat{f}(x;K(N))\right| < \delta_f\}}(x)$

where $0 < \delta_f < \delta$ (see (6)). The examination of the appropriate modified model $\hat{R}_{\mathsf{mod}}(x; K(N)) = \hat{g}(x; K(N))/\hat{f}_{\mathsf{mod}}(x; K(N))$ is beyond the scope of this paper. See [24] for some hints concerning analysis of the modified density model $\hat{f}_{\mathsf{mod}}(x; K(N))$.

The above result states that under mild conditions on the system dynamics and system noise, specified in Section II, the wavelet model $\hat{R}(x; K(N))$ converges, in a pointwise manner, to $R(x)$ if only weak boundedness requirements in Assumptions A1, A2, and A7 are satisfied, i.e. the convergence is guaranteed for a rather large class of nonlinear characteristics and input densities. In order to ensure convergence of $\hat{R}(x; K(N))$ to $R(x)$ as $N \to \infty$ the models $\hat{g}(x; K(N))$ and $\hat{f}(x; K(N))$ in the numerator and denominator should merely duly expand with the number of data. The convergence condition (24) imposed on the scale factor $K(N)$ is not restrictive and is fulfilled for $K(N) = \lfloor c \log_2 N \rfloor$ with arbitrary $0 < c < 1$. However the rate of convergence of $\hat{R}(x; K(N))$ strongly depends on the specific choice of a constant c and upon the behaviour of the approximation errors $AE(g; x; K)$ and $AE(f; x; K)$ as will be seen in the following. Since convergence of the wavelet models $\hat{g}(x; K(N))$, $\hat{f}(x; K(N))$ and $\hat{R}(x; K(N))$ takes place for almost every x, and in particular at all continuity points of $R(x)$ and $f(x)$ in the region $[a, b]$, hence the range of convergence is the same as for the conventional trigonometric and Hermite orthogonal series models (see [25], [7]). Nevertheless, wavelet models can offer faster rate of convergence, as will be demonstrated below.

B. Rate of convergence

Assuming some local smoothness of $R(x)$ and $f(x)$ we can establish the asymptotic rate of convergence of the wavelet models and yield some recommendations for the choice of the scale factor $K(N)$. For comparison purposes with conventional orthogonal series models, we shall discuss convergence rate of the model $\hat{R}(x; K(N))$ for local smoothness classes of $R(x)$ and $f(x)$ typically considered in the literature. Namely, we shall assume that $R(x) \in C^{r_R}(x_0)$ and $f(x) \in C^{r_f}(x_0)$ where $r_R, r_f \in (0, 1]$ or $r_R, r_f \in \mathsf{N}$ (the set of natural numbers), i.e. that $R(x)$ and $f(x)$ are locally (around a point x_0) Lipschitz continuous (with the exponents r_R and r_f) or r_R, r_f times differentiable functions. The standard Lipschitz functions $Lip(x_0, 1)$, with $r_R, r_f = 1$, and once continuously differentiable functions from $C^1(x_0)$ are not distinguished in the denotation because of the lack of the difference between these two smoothness classes from the viewpoint of the approximation error bound given in Appendix IV. To avoid ambiguity, the particular smoothness of $R(x), f(x) \in C^1(x_0)$ will be individually specified, if necessary. In the following, for a sequence of random variables $\{\vartheta_N\}$ and a positive number sequence $\{b_N\}$ convergent to zero, by $\vartheta_N = O(b_N)$ in probability we mean that $r_N(\vartheta_N/b_N)$ tends to zero in probability as $N \to \infty$,

i.e.

$$P\{|r_N|\,(|\vartheta_N|/b_N) > \varepsilon\} \to 0 \quad \text{as} \quad N \to \infty \tag{25}$$

each $\varepsilon > 0$, where $\{r_N\}$ is a number sequence arbitrarily slowly tending to zero.

Let $R(x) \in C^{R}(x_0)$ and $f(x) \in C^{f}(x_0)$. Then the product function $g(x) = R(x)f(x) \in C^{g}(x_0)$ where $g = \min\{R, f\}$, which is straightforward to see. Let further $(\hat{F}(x; K(N)), F(x; K(N)), F(x))$ be a generic triplet as in Section V-A and let $F(x) \in C^{F}(x_0)$ with appropriate smoothness index $F \in (0, 1]$ or $F \in \mathbb{N}$. Taking into account (23), (63) in Appendix IV and Lemma 3 in Appendix V, we ascertain that asymptotically (for large values of N and hence large $K(N)$; cf. (24)) it holds that

$$\hat{F}(x_0; K(N)) = F(x_0) + O\left(\left(2^{-2\lambda_F K(N)} + 2^{K(N)}/N\right)^{1/2}\right) \tag{26}$$

in probability, where $\lambda_F = \min\{F, r + 1\}$ and r is a number of vanishing moments of the wavelet function (x) applied in the wavelet model (cf. (46) in Appendix I). Since in particular $\lambda_g = \min\{g, r + 1\} = \min\{R, f, r + 1\}$ and $\lambda_f = \min\{f, r + 1\}$, and hence $\lambda_g \leq \lambda_f$, by Lemma 4 in Appendix V we conclude for the ratio model $\hat{R}(x_0; K(N))$ that its asymptotic behaviour is governed by the rule

$$\hat{R}(x_0; K(N)) = R(x_0) + O\left(\left(2^{-2\lambda_g K(N)} + 2^{K(N)}/N\right)^{1/2}\right) \tag{27}$$

in probability. Clearly, the performance of the model depends on the particular choice of c in the scale factor $K(N) = \lfloor c \log_2 N \rfloor$ (see Section V-A), and the best asymptotic rate of convergence in (27) is achieved when the two (antagonistic) components of the model error are in balance. This leads to the following.

Theorem 2: Suppose that $R(x) \in C^{R}(x_0)$, $f(x) \in C^{f}(x_0)$ and let the mother wavelet (x) of the wavelet family applied in the model $\hat{R}(x_0; K(N))$ possess r vanishing moments. If the scale factor $K(N)$ is selected according to the rule

$$K_{opt}(N) = \lfloor (1/(2\lambda_g + 1)) \log_2 N \rfloor \tag{28}$$

where $\lambda_g = \min\{R, f, r + 1\}$ then

$$\hat{R}(x_0; K_{opt}(N)) = R(x_0) + O\left(N^{-\lambda_g/(2\lambda_g + 1)}\right) \tag{29}$$

in probability, and this is the best guaranteed asymptotic rate of convergence of the wavelet model $\hat{R}(x; K(N))$ for the triple $(R(x), f(x), (x))$.

Proof: The conclusion follows immediately from (27). ■

The rate in (29) depends, through the index λ_g , on local smoothness of more crude function from among $R(x)$ and $f(x)$ (i.e. a function with a smaller Lipschitz exponent or a smaller number of bounded derivatives) and regularity of the wavelet function (x). Increase of smoothness of $R(x)$ and $f(x)$ and adequate choice of (x), matching the smoothness of the functions $R(x)$ and $f(x)$, result in improvement of the convergence speed. The optimal matching of the wavelets in the model $\hat{R}(x; K_{opt}(N))$ to the smoothness of $R(x)$ and $f(x)$ is achieved when $r + 1 = \lceil \min\{ _R, _f\} \rceil$. Then we obtain the fastest rate of order $O\left(N^{-\min\{ _R, _f\}/(2\min\{ _R, _f\}+1)}\right)$ for a given $_R$ and $_f$. Higher regularity of wavelets (larger number of vanishing moments r) does not improve convergence and lower regularity decreases the rate in (29). Since, however, $\lambda_g / (2\lambda_g + 1) < 1/2$ for each index λ_g , the rate $O\left(N^{-\lambda_g / (2\lambda_g +1)}\right)$ guaranteed in (29) is always smaller than $O(N^{-1/2})$, i.e. the best-possible parametric rate of convergence in probability. This is an effect of the approximation errors $AE(g; x; K)$ and $AE(f; x; K)$ present in (23), (26) and (27) (compare (63) in Appendix IV). The parametric rate $O(N^{-1/2})$ can be achieved only in a specific case when accurate representation of the functions $g(x)$ and $f(x)$ is achievable in some model space V_K (cf. Appendix I), and consequently just a finite number of unknown parameters have to be estimated in the wavelet models of the fixed scale K. Then there is no approximation error and the only variance error of the models is present in (23). The deterioration of convergence speed in nonparametric inference, where permitted uncertainty in the identified characteristics is incomparably greater than in the standard parametric estimation, is a typical occurence and achievable rate of convergence in probability is generally of smaller order $O(N^{-r})$, $0 < r < 1/2$ ([11]). Nevertheless, the following advantageous properties of the optimum wavelet model $\hat{R}(x; K_{opt}(N))$ can be concluded from (29).

1. If $R(x)$ and $f(x)$ are Lipschitz functions on $[a, b]$ ($_R = _f = 1$) then the rate in (29) is $O(N^{-1/3})$ in probability for each wavelet family. As established in [11], this is the best possible nonparametric rate of convergence for Lipschitz nonlinearities.
2. If $R(x) \in C^{ R}(x_0)$, $f(x) \in C^{ f}(x_0)$, some $_R$, $_f \in \mathbb{N}$, and moreover $_f \geq _R$ and $r + 1 \geq _R$ then the rate in (29) is $O(N^{- R/(2 R+1)})$ in probability. This is the best possible nonparametric rate of convergence for differentiable nonlinearities $R(x)$ [11].

As a result of these optimality properties and freedom in the choice of wavelet functions, the wavelet-based model $\hat{R}(x; K_{opt}(N))$ can outperform conventional orthogonal series models elaborated earlier in the literature. For example, for Lipschitz $R(x)$ and $f(x)$ around x_0 the attainable rate $O(N^{-1/3})$ in probability is better than $O(N^{-1/4})$ in probability guaranteed for more smooth differentiable $R(x), f(x) \in C^1(x_0)$ by the trigonometric and Hermite orthogonal series models (see [25], [7]). The latter rate

$O(N^{-1/4})$ can, in turn, be achieved by the wavelet model $\hat{R}(x; K_{opt}(N))$ for considerably less smooth $R(x), f(x) \in C^{1/2}(x_0)$. For differentiable $R(x), f(x) \in C^d(x_0)$, $d \in \mathbb{N}$, and wavelet-based models with $r + 1 \geq d$ we can reach the convergence rate $O(N^{-d/(2d+1)})$ which is again faster than $O(N^{-(2d-1)/4d})$ in probability attained by the models using trigonometric or Hermite series approximations (see the references above). Specifically, for locally constant functions $R(x)$ and $f(x)$ around x_0, i.e. $R(x), f(x) \in C^{\infty}(x_0)$ in our denotation, the model $\hat{R}(x; K_{opt}(N))$ can potentially achieve the parametric rate of convergence $O(N^{-1/2})$ in probability.

It is noteworthy that the particular system and noise dynamics as well as level (variance) of the noise do not influence the order of the asymptotic rate of convergence and all the above rates remain the same for white and correlated noise of a rather general correlation structure (Assumption A4 in Section II).

C. Practical scale selection strategy

The scale selection rule in (28) needs for the employment the values of local smoothness indices $_R$ and $_f$ and hence (i) requires advanced prior knowledge of the regularity of the nonlinearity $R(x)$ and the input density $f(x)$ and (ii) can vary from point to point, along with the local smoothness of $R(x)$ and $f(x)$, yielding unstable values of $K_{opt}(N)$. Because of this disadvantage and since the indices $_R$ and $_f$ are usually unknown the rule in (28) is rather of only theoretical meaning, and hence – based on the analysis in Section V-B – we propose the following 'rule of thumb' for selecting the scale in the wavelet model, independent of local regularity of $R(x)$ and $f(x)$:

$$\bar{K}(N) = \lfloor (1/3) \log_2 N \rfloor \tag{30}$$

Owing to (27) and the obvious inequality $\min\{_R, _f, r + 1\} \geq \min\{_R, _f, 1\}$, for this rule we obtain the asymptotic guaranteed rate

$$\hat{R}(x_0; \bar{K}(N)) = R(x_0) + O\left(N^{-\min\{_R, _f, 1\}/3}\right) \quad \text{in probability} \tag{31}$$

irrespective of the applied wavelet family, which ensures convergence $O(N^{-1/3})$ in probability provided that $_R, _f \geq 1$, i.e. $R(x)$ and $f(x)$ are at least Lipschitz functions around x_0. For $\min\{_R, _f\} \geq 1$ the rule assures balance of both error components in (27), and no other rule in the class $K(N) = \lfloor (1/n) \log_2 N \rfloor$, integer $n > 1$, yields better warranted convergence of the error for each wavelet model (each $r \geq 0$). Since Lipschitz regularity of nonlinearity $R(x)$ and input density $f(x)$ can be reliably expected in most real-life situations and the rate $O(N^{-1/3})$ seems quite satisfactory as compared with the theoretically best possible (parametric) rate of convergence $O(N^{-1/2})$ in probability (see Section V-B),

we recommend the rule (30) for use in the case of lack of prior knowledge of $_R$ and $_f$. When considering rougher functions $R(x)$ or $f(x)$ around x_0, with $\min\{ _R, _f\} < 1$, the asymptotic guaranteed rate in (31) is deteriorated to the order $O\left(N^{-\min\{ _R, _f\}/3}\right)$ in probability. For fractional $\min\{ _R, _f\} = k/l$ with $k \ll l$ we get asymptotically very slow rate $O(N^{-k/3l})$. Such a roughness of $R(x)$ and $f(x)$ seems to be however a rare occurrence in practice. Practical utility of the sub-optimal rule (30) for small and moderate number of data is verified empirically in Section VI. The results presented there confirm the fair efficiency of the strategy.

Remark 4: Convergence properties of particular wavelet models can be easily concluded from (27), Theorem 2 and discussion concerning the asymptotic rate in (31). Specifically, if for the construction of the wavelet models we applied wavelets collected in Table II in Appendix I, they follow from the possessed number of vanishing moments r , given in Table II. In particular, a general conclusion concerning whole class of Daubechies wavelet-based models is that for $R(x)$ and $f(x)$ with $\min\{ _R, _f\} < 1$ employing Daubechies wavelets with the wavelet numbers $p > 1$ is pointless from the rate of convergence viewpoint as the rate in (29) cannot be then faster than for the Haar wavelets, with $p = 1$, i.e. the wavelet model of Example 1.

VI. Numerical experiment

In the experiment, we examine as an example the performance of the above-mentioned class of the Daubechies wavelet-based models, denoted further by $\hat{R}_D^p(x; K(N))$, obtained for the Daubechies wavelets $(\varphi_D^p(x), \ _D^p(x))$ of the numbers $p = 1, 3, 5$, with, accordingly, $r \ (= p-1) = 0, 2, 4$ vanishing moments (see Table II in Appendix I). Since the Daubechies wavelets have the shortest supports among all orthogonal compactly supported wavelet functions with a given number of vanishing moments, and hence require the smallest number of components in the generic wavelet models (16) and (17), the Daubechies wavelet models $\hat{R}_D^p(x; K(N))$ are most parsimonious within the class considered in the paper (see [26], [12], Table II and (42) in Appendix I). In our simulations we assume $[a, b] = [0, 1]$ (identification region) and use Hammerstein systems which represent well the systems reported in Section III, with three nonlinear characteristics (x):

$$\begin{aligned} \text{polynomial} &: \quad _1(x) = 10\left(2x^3 - 3x^2 + x\right) \\ \text{cube root} &: \quad _2(x) = \sqrt[3]{x - 1/2} \\ \text{quantizer} &: \quad _3(x) = 1/8 + \lfloor 8x - 4 \rfloor /4 \end{aligned} \tag{32}$$

The polynomial nonlinearity illustrates smooth infinitely times continuously differentiable functions with finite power series representation, for which the approximation error is particularly sensitive to the number of vanishing moments of the applied wavelet functions (see (40), (44) and (46) in Appendix I). The cube root nonlinearity is a smooth, almost everywhere differentiable function with infinite power series representation and hence with less sensitive approximation error to the number of vanishing moments of applied wavelets. In turn, the quantizer nonlinearity (piecewise-constant) is an example of discontinuous functions with jumps which match well the dyadic grid points and thus are well-fitted to the implementation of the Haar wavelet models ($p = 1$). The linear output dynamics was selected as (i) FIR element with the impulse response $\{\lambda_i = 1 - i/4\}_{i=0}^{5}$ and (ii) IIR subsystem with the description $v_k - 0.75v_{k-1} + 0.5v_{k-2} = w_k + 0.75w_{k-1} + 0.5w_{k-2}$ (where $w_k = \quad (x_k)$ is the input and v_k is the output of the dynamics). The systems were driven by white, stationary random sequence $\{x_k\}$ of the uniform distribution $x_k \sim U[0, 1]$ (Assumption A2 in Section II), and thus the input probability density function $f(x)$ was infinitely times continuously differentiable in each internal point of the identification region. Since in our tests the nonlinearities $\quad (x)$ are anti-symmetric and the input density is symmetric with respect to the center $x = 0.5$ of the region $[0, 1]$, we get in each case $E \quad (x_1) = 0$. Since moreover $\lambda_0 = 1$, we have $R(x) = \quad (x)$ (see Section III) i.e. in our experiment we identify true nonlinear characteristics of the Hammerstein systems. Finally, the external correlated zero mean output noise $\{z_k\}$ was generated as an output of MA(l) filter $z_k = \sum_{i=0}^{l} \omega_i \varepsilon_{k-i}$, driven by the uniformly distributed white noise process $\varepsilon_k \sim U[-\delta, \delta]$, independent of $\{x_k\}$ (see Assumption A4 for $L = 1$ and Assumption A5). In simulations we assumed $\{\omega_i = 1 - i/(l + 1)\}_{i=0}^{l}$ and $l = 0, 2, 4$ obtaining $\{z_k\}$ white for $l = 0$ and correlated for $l = 2$ and 4, with increasing correlation for growing l. For each identified nonlinearity $R(x) = \quad (x)$ and each value of l, the white noise parameter δ was selected as to give constant the noise-to-signal ratio:

$$NSR \stackrel{\text{def}}{=} \frac{\max |z_k|}{\max_{x \in [0,1]} |R(x)|}$$

In all tests we set $NSR = 10\%$. In fact, according to (1), the noise blurring the nonlinearity $R(x)$ was enlarged by the influence of the system dynamics (system noise), and factually it was $y_k = R(x_k) + \xi_k + z_k$ with $\xi_k = \sum_{i=1}^{\infty} \lambda_i \quad (x_{k-i})$ (see Section III) yielding a substantial additional disturbance for large values of $\sum_{i=1}^{\infty} |\lambda_i|$. Such influence was discussed in more detail in [18].

In the models $\hat{R}_D^p(x; K(N))$ we assumed the initial scale $M = 1$ and used the sub-optimal scale selection rule $\bar{K}(N)$ in (30) to verify its practical applicability. Thereby, we did not exploit in the simulations any prior knowledge about the smoothness of $R(x)$ and $f(x)$, and hence likened the

experimental conditions to typical practical circumstances. In what follows, we denote $\hat{R}_D^p(x;\bar{K}(N)) = R_N(x;p)$ for shortness. The accuracy of the models was evaluated using the empirical average pointwise identification error of the form

$$error\,(N) = \frac{1}{T}\sum_{r=1}^{T}\left\{\frac{1}{1000}\sum_{q=1}^{1000}\left[R\,(x_q) - R_N^r\,(x_q;p)\right]^2\right\} \tag{33}$$

where T is the number of independent trials for each sample size N, $R_N^r\,(x;p)$ is the empirical model computed in the r-th run for N data points, and $x_q = q/1000$, $q = 1, 2, \ldots, 1000$, are the equidistant estimation points from the interval $[0, 1]$. In our experiment we assumed $T = 100$ and $N = 2^5, 2^6, \ldots, 2^{13} = 8192$.

The performance of the models is illustrated in Figs. 6 and 7 where the $error(N)$ is shown for FIR and IIR dynamics, $p = 1, 3, 5$ and growing number N of data for each test nonlinearity. For comparison purposes, we present also the estimation error obtained for the trigonometric series model, elaborated in [7] and invoked in Section V-B. We see that even for the sub-optimal scale selection rule (30) the error visibly decreases with growing N, and the accuracy better than 0.02, is achieved for each wavelet model and each sample nonlinear characteristic for $N = 2^9 = 512$ data points. As it was noticed by one of the reviewers, this however does not actually guarantee that a very good estimate for the nonlinearity is obtained when we use merely $N = 512$ observation as the relative average inaccuracy $(\sqrt{error\,(N)}/R_{\max}) \times 100\%$, where $R_{\max} = \max_{x\in[0,1]}|R\,(x)|$ is the magnitude of the characteristic in the identification region, is still of about 14%. Thus, far larger number of data is in fact required in

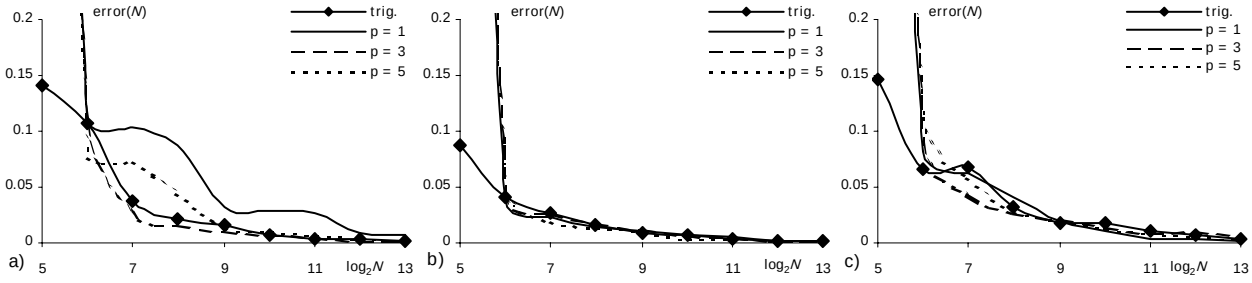

Fig. 6. Accuracy of the models $R_N(x;p)$ and the trigonometric series model against the sample size N (in log-scale) and the wavelet number p for **a)** polynomial, **b)** cube root, **c)** quantizer non-linearity; FIR dynamics and MA(2) output noise.

our method. It should be recalled at this point that asymptotic rate of convergence, in our experiment of order $O\left(N^{-1/3}\right)$ in probability for each p (see (31) in Section V-C and the discussion concerning

smoothness of the input density and test nonlinearities) is smaller than typical parametric rate, $O\left(N^{-1/2}\right)$ in probability, which is a general weakness of nonparametric inference and has been already explained in Section V-B. In effect, for our models the same accuracy as for parametric approaches can be obtained solely for much larger number of data. This can be considered as the disadvantage of the method.

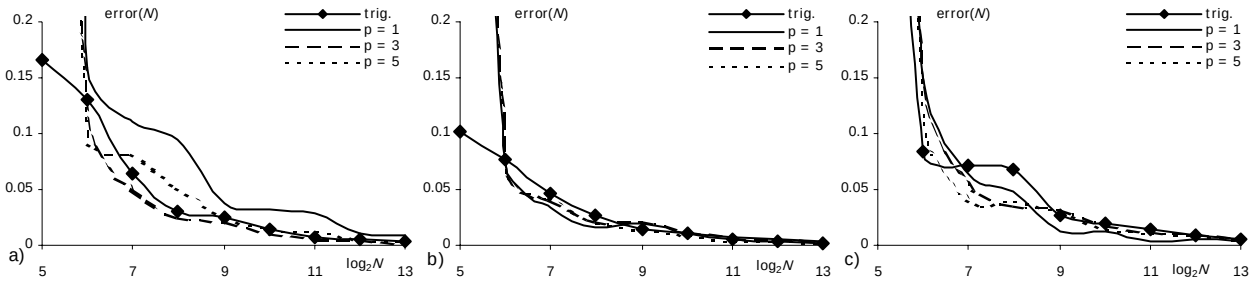

Fig. 7. Accuracy of the models $R_N(x;p)$ and the trigonometric series model against the sample size N (in log-scale) and the wavelet number p for **a)** polynomial, **b)** cube root, **c)** quantizer non-linearity; IIR dynamics and MA(2) output noise.

Wavy shape of the plots for the wavelet models is an cumulative effect caused by stepwise (abrupt) decrease of the approximation error and parallel increase of the variance error of the models for $\log_2 N = 6, 9$ and 12, connected with the stepwise growth of the scale factor from $\bar{K}(N)$ to $\bar{K}(N)+1$ (cf. (30) and (60), (61) and (63) in Appendix III and IV), see [18] for the details. Despite the same order $O\left(N^{-1/3}\right)$ of asymptotic convergence rate of the error, for each nonlinear characteristic one can point out $p = p^*$ minimizing the empirical $error(N)$ for smaller number of data. In our experiment, $p^* = 3$ for polynomial nonlinearity, $p^* = 5$ for cube root nonlinearity, and $p^* = 1$ for quantizer nonlinearity. This is an effect of various adaptivity of the wavelet models $R_N(x;p)$ of various p to individual nonlinear characteristics, and in particular of various diminution rate of the approximation error of the models (particularly important for small and medium scales $\bar{K}(N)$), and is in good agreement with our expectations (cf. (63) in Appendix IV and the discussion regarding approximation error of test nonlinearities).

To get deeper insight into the accuracy of the wavelet models of various p, exemplary numerical values of the $error(N)$ for larger number of data are presented in Table I for two contrasting nonlinearities, smooth polynomial $_1(x)$ and nonsmooth quantizer $_3(x)$, for FIR dynamics and MA(2) output noise together with the error of conventional trigonometric series model. As we see, in accordance with the theory, the smallest error is obtained, respectively, for $p = 3$ and $p = 1$, and even for the sub-optimal scale selection rule $\bar{K}(N)$ applied in the experiment accuracy of the adequate wavelet models is better

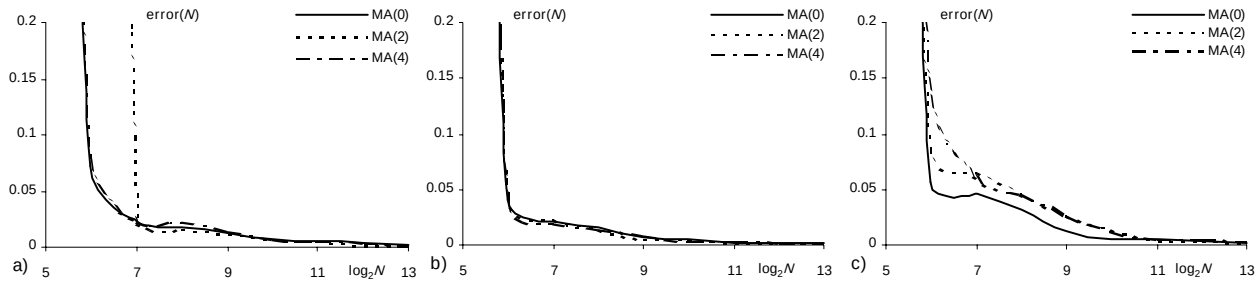

Fig. 8. Effect of the correlation of the output noise on the accuracy of the models $R_N(x;p^*)$ for **a)** polynomial, **b)** cube root, **c)** quantizer non-linearity; MA(l) output noise model, FIR dynamics.

than of the trigonometric series model. For instance, employing $N = 4096$ data points for recovery of polynomial nonlinearity $_1(x)$ and using the wavelet model with $p = 3$ we obtain the error 0.0022 against 0.0031 for the trigonometric series model. It yields approximately 4.7% and 5.6% relative average inaccuracy $(\sqrt{error(N)}/R_{\max}) \times 100\%$ (since for the nonlinearity $R(x) = _1(x)$ the magnitude in the identification region is $R_{\max} = 1$). This, on the one hand, confirms practical utility of the sub-optimal rule $\bar{K}(N)$ for selecting the scale of the wavelet models and, on the other, testifies superior accuracy of optimum wavelet models $\hat{R}(x;K_{opt}(N))$, established in Section V-B. Although the error for IIR dynamics is about twice as large because of the larger system noise, the similarity of the shape of plots in Figs. 6 and 7 also confirms the robustness of the order of the identification error against system dynamics, noticed in Section V-B.

To illustrate the performance of the models for various correlation patterns of the noise, Fig. 8 presents the plots of the $error(N)$ for the wavelet models $R_N(x;p^*)$ against the order l of the external noise MA(l) for FIR dynamics and each test nonlinearity. It is seen that for $N \geq 2^7 = 128$ and each $l = 0, 2, 4$ the error is nearly the same, i.e. the computed models are indeed insensitive to correlation of the output noise $\{z_k\}$ yielding, in particular, similar identification error for white noise ($l = 0$) and the noise MA(4).

VII. Conclusions

We have studied a class of nonparametric wavelet-based identification algorithms for reconstruction of nonlinear characteristics of a class of discrete-time block-oriented dynamical systems operating in a stochastic environment. The algorithms apply orthogonal wavelets with compact support. They can be used when a priori knowledge about the system and noise model is very small, and in particular when no

		polynomial				quantizer			
N		$p=1$	$p=3$	$p=5$	trig.	$p=1$	$p=3$	$p=5$	trig.
2048	$=2^{11}$	0.0269	**0.0045**	0.0077	0.0043	**0.0034**	0.0084	0.0095	0.0115
4096	$=2^{12}$	0.0083	**0.0022**	0.0028	0.0031	**0.0032**	0.0099	0.0061	0.0077
8192	$=2^{13}$	0.0075	**0.0016**	0.0018	0.0016	**0.0019**	0.0063	0.0036	0.0044

TABLE I

THE ESTIMATION ERROR $error(N)$ FOR POLYNOMIAL AND QUANTIZER NONLINEARITY (FIR DYNAMICS AND MA(2) OUTPUT NOISE)

parametric representation of unknown nonlinear characteristic is given. We have focused on the analysis of local, pointwise properties of the wavelet models. It was demonstrated that for an appropriate choice of the scale factor (resolution level) K in relation to the number of measurement data N, the models converge almost everywhere in probability for a wide class of nonlinear characteristics and the convergence holds irrespective of the system and noise dynamics. Assuming some smoothness of the nonlinear characteristic and the input probability density function we have determined the rule for optimum scale selection of the wavelet model, minimizing the asymptotic identification error, and have established an asymptotic rate of convergence of the optimum model. It was shown that the wavelet models can attain the best possible nonparametric rate of convergence, and that may perform better than trigonometric and Hermite orthogonal series models worked out earlier in the literature (Section V-B). For the use of the models, only a number of coefficients $\{\hat{\alpha}^g_{Mn}, \hat{}^g_{mn}\}$ and $\{\hat{\alpha}^f_{Mn}, \hat{}^f_{mn}\}$ must be calculated from experimental data, of order $O\left(2^K\right)$ (equation (21)), much smaller than the number N of measured raw data, and the 'data compression' grows with growing N as $2^K/N \to 0$ for $N \to \infty$ (Theorem 1).

The presented algorithms can be truly recommended for the identification purposes when precise pointwise properties of nonlinear characteristics are of particular interest and traditional parametric methods fail, i.e. when a priori information about nonlinearity is poor and the noise corrupting the system cannot be modeled exactly.

APPENDIX I
BASIC PROPERTIES OF COMPACTLY SUPPORTED WAVELETS

Here, we have collected basic facts concerning wavelet approximation of functions, used in the paper.

1. Any function $F(x) \in L^2(\mathsf{R})$ can be approximated with the help of orthogonal wavelets $\{\varphi_{Mn}(x)\}_{n\in\mathsf{Z}} \cup \{\ _{mn}(x)\}_{M\le m\le K-1, n\in\mathsf{Z}}$, $\mathsf{Z}-$the set of integers, and the wavelet approximator (wavelet model) in the

adequate approximation (model) space $V_K = V_M \oplus W_M \oplus W_{M+1} \oplus \ldots \oplus W_{K-1} \subset L^2(\mathsf{R})$ has the form

$$F(x;K) = \sum_{n=-\infty}^{\infty} \alpha_{Mn}\varphi_{Mn}(x) + \sum_{m=M}^{K-1}\sum_{n=-\infty}^{\infty} \quad _{mn} \quad _{mn}(x) \tag{34}$$

where

$$\alpha_{Mn} = \int_{-\infty}^{+\infty} F(x)\,\varphi_{Mn}(x)\,dx\,; \qquad _{mn} = \int_{-\infty}^{+\infty} F(x) \quad _{mn}(x)\,dx \tag{35}$$

In the wavelet model K is called the scale, or resolution level, and

$$\varphi_{Mn}(x) = 2^{M/2}\varphi\left(2^M x - n\right)\,; \qquad _{mn}(x) = 2^{m/2} \quad (2^m x - n) \tag{36}$$

are scaled (factor m) and translated (factor n) versions of a father and mother wavelet, $\varphi(x)$ and (x), spanning the initial approximation space V_M and the orthogonal detail (wavelet) spaces W_m:

$$V_M = \mathsf{span}\{\varphi_{Mn}(x)\,, n \in \mathsf{Z}\}\,; \quad W_m = \mathsf{span}\{\ _{mn}(x)\,, n \in \mathsf{Z}\}\,.$$

2. For bounded wavelets with compact support

$$|\varphi(x)| \le M_\varphi\, I_{[s_1,s_2]}(x)\,; \quad |\ (x)| \le M\ I_{[t_1,t_2]}(x) \tag{37}$$

$$|\varphi_{Mn}(x)| \le 2^{M/2} M_\varphi\ I_{[\frac{s_1+n}{2^M},\frac{s_2+n}{2^M}]}(x)\,; \quad |\ _{mn}(x)| \le 2^{m/2} M \quad I_{[\frac{t_1+n}{2^m},\frac{t_2+n}{2^m}]}(x) \tag{38}$$

some $M_\varphi, M > 0$, the approximator (34) takes the form

$$F(x;K) = \sum_{n=n_{\min}(\varphi,x,M)}^{n_{\max}(\varphi,x,M)} \alpha_{Mn}\varphi_{Mn}(x) + \sum_{m=M}^{K-1}\sum_{n=n_{\min}(\ ,x,m)}^{n_{\max}(\ ,x,m)} \quad _{mn} \quad _{mn}(x) \tag{39}$$

where

$$\alpha_{Mn} = \int_{(s_1+n)/2^M}^{(s_2+n)/2^M} F(x)\,\varphi_{Mn}(x)\,dx\,; \qquad _{mn} = \int_{(t_1+n)/2^m}^{(t_2+n)/2^m} F(x) \quad _{mn}(x)\,dx \tag{40}$$

and

$$\begin{aligned} n_{\min}(\varphi,x,M) &= \lceil 2^M x - s_2 \rceil\,; \quad n_{\max}(\varphi,x,M) = \lfloor 2^M x - s_1 \rfloor \\ n_{\min}(\ ,x,m) &= \lceil 2^m x - t_2 \rceil\,; \quad n_{\max}(\ ,x,m) = \lfloor 2^m x - t_1 \rfloor \end{aligned} \tag{41}$$

indicate active (nonvanishing) wavelets at a point x ($\lfloor . \rfloor$ and $\lceil . \rceil$ are respectively the 'floor' and 'ceiling' function).

3. The number of components (nonzero wavelet coefficients) in (39) depends on the support size $s = s_2 - s_1 = t_2 - t_1$ of $\varphi(x)$ and (x) (which is the same for both wavelet functions, and moreover $t_1 = -(s_2 - s_1 - 1)/2$ and $t_2 = (s_2 - s_1 + 1)/2$, cf. [12]) and for each M, m and x it holds that

$$\begin{aligned} n_{\max}(\varphi,x,M) - n_{\min}(\varphi,x,M) + 1 &\le S \\ n_{\max}(\ ,x,m) - n_{\min}(\ ,x,m) + 1 &\le S \end{aligned} \tag{42}$$

where $S = \lfloor s \rfloor + 1$.

4. The supports of wavelets present in (39) (active at a point x) are included in the interval

$$[x_{\min}(x, M), x_{\max}(x, M)] = \left[x - s/2^M, x + s/2^M\right] \tag{43}$$

5. The pointwise approximation error of the approximator (39) is (cf. (38) and (42))

$$|AE(F; x; K)| = |F(x) - F(x; K)| = \left| \sum_{m=K}^{\infty} \sum_{n=n_{\min}(\,,x,m)}^{n_{\max}(\,,x,m)} \quad mn \quad mn(x) \right| \leq$$

$$\leq S M \sum_{m=K}^{\infty} 2^{m/2} \max\{ | \quad mn| : n_{\min}(\,, x, m) \leq n \leq n_{\max}(\,, x, m) \} \tag{44}$$

6. For the wavelet functions as in (37) it holds that

$$|AE(F; x; K)| \to 0 \quad \text{as} \quad K \to \infty \tag{45}$$

for almost all x (in the sense of Lebesgue measure), and in particular the convergence holds at the continuity points of $F(x)$ [27, Theorem 2.1(ii)].

7. By definition, the mother wavelet (x) has r vanishing moments if

$$\int_{t_1}^{t_2} x^k \quad (x)\, dx = 0 \quad \text{for} \quad k = 0, 1, \ldots, r \quad . \tag{46}$$

Then the wavelet support size is $s \geq 2r + 1$ [12, Proposition 7.4] and for each $m, n \in \mathbb{Z}$ also [12]

$$\int_{(t_1+n)/2^m}^{(t_2+n)/2^m} x^k \quad mn(x)\, dx = 0 \quad \text{for} \quad k = 0, 1, \ldots, r$$

Other details can be found in the rich wavelet literature (e.g. [26], [12]). Exemplary orthogonal wavelets with compact support are characterized in Table II.

	Daubechies/symmlet	coiflet
Support of $\varphi(x)$	$[0, 2p-1]$	$[-2p, 4p-1]$
Support of (x)	$[1-p, p]$	$[1-3p, 3p]$
$n_{\min}(\varphi, x, m)$	$\lfloor 2^m x \rfloor - 2p + 2$	$\lfloor 2^m x \rfloor - 4p + 2$
$n_{\max}(\varphi, x, m)$	$\lceil 2^m x \rceil - 1$	$\lceil 2^m x \rceil + 2p - 1$
$n_{\min}(\,, x, m)$	$\lfloor 2^m x \rfloor - p + 1$	$\lfloor 2^m x \rfloor - 3p + 1$
$n_{\max}(\,, x, m)$	$\lceil 2^m x \rceil + p - 2$	$\lceil 2^m x \rceil + 3p - 2$
r	$p - 1$	$2p - 1$

TABLE II

BASIC PROPERTIES OF TYPICAL ORTHOGONAL WAVELET FUNCTIONS WITH COMPACT SUPPORT (p – WAVELET NUMBER)

APPENDIX II
VARIANCE ERROR OF THE EMPIRICAL WAVELET COEFFICIENTS

For shortness, the derivations are given for $J = L = 1$ and consequently we omit in the denotations the superfluous subscript '1'. Extension to $J, L > 1$ is straightforward. Due to Assumption A6, x varies over the interval $[a, b]$.

From stationarity of the processes $\{x_k\}$ and $\{y_k\}$ (see equation (1) and Assumptions A2-A4 in Section II) it follows that (cf. (14))

$$\begin{aligned} var\left(\hat{\alpha}^g_{Mn}\right) &= var\left[N^{-1}\sum_{k=1}^{N} y_k\varphi_{Mn}(x_k)\right] = N^{-1}var\left[y_1\varphi_{Mn}(x_1)\right] \\ &\quad + 2N^{-1}\sum_{k=1}^{N-1}(1-k/N)\,cov\left[y_{k+1}\varphi_{Mn}(x_{k+1}), y_1\varphi_{Mn}(x_1)\right] \\ &= N^{-1}(A+B), \quad \text{say} \end{aligned} \tag{47}$$

Since $var\left[y_1\varphi_{Mn}(x_1)\right] \le E\left[y_1^2\varphi^2_{Mn}(x_1)\right]$ and $y_1^2 \le 3\left(R^2(x_1) + \xi_1^2 + z_1^2\right)$ (by (1) and Cauchy inequality), we get for the first component

$$A \le 3\left[E\left[R^2(x_1)\varphi^2_{Mn}(x_1)\right] + \left(E\xi_1^2 + Ez_1^2\right)E\left[\varphi^2_{Mn}(x_1)\right]\right]$$

owing to independence of x_1 and ξ_1, z_1 (Assumptions A2, A3 and A5). Further, because of the boundedness of the density function $f(x)$ (Assumption A2), orthonormality of wavelets $\{\varphi_{Mn}(x)\}$ and the fact that for x ranging over the region $[a, b]$ the supports of all active wavelets $\varphi_{Mn}(x)$ in the model (16) are contained in the interval $[x_{\min}(a, M), x_{\max}(b, M)] = \left[a - s/2^M, b + s/2^M\right]$ (see (43) in Appendix I), we recognize that for each $n_{\min}(\varphi, x, M) \le n \le n_{\max}(\varphi, x, M)$ where $x \in [a, b]$ (see (16)) it holds that

$$E\left[\varphi^2_{Mn}(x_1)\right] = \int_{x_{\min}(a,M)}^{x_{\max}(b,M)} \varphi^2_{Mn}(x)\,f(x)dx \le M_f \tag{48}$$

$$E\left[R^2(x_1)\varphi^2_{Mn}(x_1)\right] = \int_{x_{\min}(a,M)}^{x_{\max}(b,M)} \varphi^2_{Mn}(x)\,R^2(x)\,f(x)dx \le M_R^2 M_f \tag{49}$$

In the above expressions we have exploited that for active wavelets $\int_{x_{\min}(a,M)}^{x_{\max}(b,M)} \varphi^2_{Mn}(x)\,dx = 1$ and that for $x \in [x_{\min}(a, M), x_{\max}(b, M)]$ we have $|R(x)| \le M_R$, where

$$M_R = C_{1R}\max\left\{|x_{\min}(a, M)|, |x_{\max}(b, M)|\right\} + C_{2R}$$

(see Assumption A1 in Section II). Using (48) and (49) we obtain eventually

$$A = var\left[y_1\varphi_{Mn}(x_1)\right] \le 3M_f\left[M_R^2 + \sigma_\xi^2 + \sigma_z^2\right] < \infty \tag{50}$$

where $\overset{2}{\xi} = E\xi_1^2 = \ ^2\sum_{i=1}^{\infty}\lambda_i^2 < \infty$ and $\overset{2}{z} = Ez_1^2 = \ \overset{2}{\varepsilon}\sum_{i=0}^{\infty}\omega_i^2 < \infty$ (see Assumptions A2-A4).

Concerning the second term in (47), denote $r_{Mn,\varphi}^g(k) = cov\left[y_{k+1}\varphi_{Mn}(x_{k+1}), y_1\varphi_{Mn}(x_1)\right]$. Using equation (1) and Assumptions A2-A5 (in particular, stationarity and independence of appropriate quantities along with the fact that $E\xi_1 = Ez_1 = 0$), we ascertain after standard calculation that

$$\begin{aligned} r_{Mn,\varphi}^g(k) \ &= \ \lambda_k E\left[\ (x_1)\, R(x_1)\, \varphi_{Mn}(x_1)\right] E\left[\varphi_{Mn}(x_1)\right] \\ &\qquad + \left(E\left[\xi_{k+1}\xi_1\right] + E\left[z_{k+1}z_1\right]\right) E^2\left[\varphi_{Mn}(x_1)\right] \end{aligned}$$

Since (see the definitions of $\{\xi_k\}$ and $\{z_k\}$ in Assumptions A3 and A4 for $J = L = 1$) $\ _{\xi}(k) = E\left[\xi_{k+1}\xi_1\right] = \ ^2\sum_{i=1}^{\infty}\lambda_i\lambda_{i+k}$, $\ _z(k) = E\left[z_{k+1}z_1\right] = \ \overset{2}{\varepsilon}\sum_{i=0}^{\infty}\omega_i\omega_{i+k}$ and (by Schwarz inequality)

$$\begin{aligned} \left|E\left[\ (x_1)\, R(x_1)\, \varphi_{Mn}(x_1)\right]\right| \ &\leq \ \left(E\left[R^2(x_1)\, \varphi_{Mn}^2(x_1)\right]\right)^{1/2} \\ \left|E\left[\varphi_{Mn}(x_1)\right]\right| \ &\leq \ \left(E\left[\varphi_{Mn}^2(x_1)\right]\right)^{1/2} \end{aligned}$$

thus, owing to (48) and (49), we get therefrom that

$$\begin{aligned} \left|r_{Mn,\varphi}^g(k)\right| \ &\leq \ M_f\left(M_R \quad |\lambda_k| + \left|\ _{\xi}(k)\right| + \left|\ _z(k)\right|\right) \\ &\leq \ M_f\left(M_R \quad |\lambda_k| + \ ^2\sum_{i=1}^{\infty}|\lambda_i\lambda_{i+k}| + \ \overset{2}{\varepsilon}\sum_{i=0}^{\infty}|\omega_i\omega_{i+k}|\right) \end{aligned} \tag{51}$$

Hence (cf. (47))

$$\begin{aligned} |B| \ &\leq \ 2\sum_{k=1}^{N-1}(1-k/N)\left|r_{Mn,\varphi}^g(k)\right| \leq 2\sum_{k=1}^{\infty}\left|r_{Mn,\varphi}^g(k)\right| \\ &\leq 2M_f\left(c_1 M_R \quad + c_2\ ^2 + c_3\ \overset{2}{\varepsilon}\right) < \infty \end{aligned} \tag{52}$$

as under Assumptions A3 and A4 it holds that $\sum_{k=1}^{\infty}|\lambda_k| = c_1$, $\sum_{k=1}^{\infty}\sum_{i=1}^{\infty}|\lambda_i\lambda_{i+k}| = c_2$ and $\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}|\omega_i\omega_{i+k}| = c_3$ some $0 < c_1, c_2, c_3 < \infty$. Putting together (47), (50) and (52) yields

$$var\left(\hat{\alpha}_{Mn}^g\right) \leq C_{\alpha}^g\, N^{-1} \tag{53}$$

where $C_{\alpha}^g = M_f\left(\bar{c}_1 M_R \max\{M_R, \quad\} + \bar{c}_2\ ^2 + \bar{c}_3\ \overset{2}{\varepsilon}\right)$ and $\bar{c}_1 = 2c_1 + 3$, $\bar{c}_2 = 2c_2 + 3\sum_{i=1}^{\infty}\lambda_i^2$, $\bar{c}_3 = 2c_3 + 3\sum_{i=0}^{\infty}\omega_i^2$. After similar steps, we get (cf. (14))

$$var\left(\hat{\ }_{mn}^{g}\right) \leq C^g\, N^{-1} \tag{54}$$

where $C^g = C_{\alpha}^g$. As regards the coefficients $\hat{\alpha}_{Mn}^f$ and $\hat{\ }_{mn}^{f}$ (cf. (15)), from Assumption A2 we obtain immediately

$$var\left(\hat{\alpha}_{Mn}^f\right) = var\left[N^{-1}\sum_{k=1}^{N}\varphi_{Mn}(x_k)\right] = N^{-1}var\left[\varphi_{Mn}(x_1)\right]$$

and then using $var\left[\varphi_{Mn}\left(x_1\right)\right] \leq E\left[\varphi_{Mn}^2\left(x_1\right)\right]$ and the bound in (48), we conclude that $var\left(\hat{\alpha}_{Mn}^f\right) \leq M_f\, N^{-1}$. Analogously, $var\left(\hat{\ }_{mn}^{f}\right) \leq M_f\, N^{-1}$.

APPENDIX III

VARIANCE ERROR OF THE EMPIRICAL WAVELET MODELS

Let $(\hat{c}, c) \in \{(\hat{\alpha}_{Mn}^g, \alpha_{Mn}^g), (\hat{\ }_{mn}^{g}, \ {}_{mn}^{g}), (\hat{\alpha}_{Mn}^f, \alpha_{Mn}^f), (\hat{\ }_{mn}^{f}, \ {}_{mn}^{f})\}$. Owing to (9)-(12) and (14), (15) along with stationarity of the processes $\{x_k\}$ and $\{y_k\}$, we get $E\hat{c} = c$. Hence, for $(\hat{F}(x;K), F(x;K)) \in \{(\hat{g}(x;K), g(x;K)), (\hat{f}(x;K), f(x;K))\}$ we conclude that (cf. (7), (8) and (16), (17))

$$E\,\hat{F}(x;K) = F(x;K) \tag{55}$$

Consider the variance of $\hat{g}(x;K)$ for $x \in [a,b]$. By virtue of (55), we have

$$var\ \{\hat{g}(x;K)\} = E\,[\hat{g}(x;K) - g(x;K)]^2 \tag{56}$$

where (cf. (7) and (16))

$$\hat{g}(x;K) - g(x;K) = \sum_{n=n_{\min}(\varphi,x,M)}^{n_{\max}(\varphi,x,M)} \left(\hat{\alpha}_{Mn}^g - \alpha_{Mn}^g\right) \varphi_{Mn}\left(x\right) + \sum_{m=M}^{K-1} \sum_{n=n_{\min}(\ ,x,m)}^{n_{\max}(\ ,x,m)} \left(\hat{\ }_{mn}^{g} - \ {}_{mn}^{g}\right) \ {}_{mn}\left(x\right)$$

Owing to (38) in Appendix I, we obtain immediately

$$|\hat{g}(x;K) - g(x;K)| \leq M_\varphi 2^{M/2} \sum_{n=n_{\min}(\varphi,x,M)}^{n_{\max}(\varphi,x,M)} \left|\hat{\alpha}_{Mn}^g - \alpha_{Mn}^g\right| + M \sum_{m=M}^{K-1} 2^{m/2} \sum_{n=n_{\min}(\ ,x,m)}^{n_{\max}(\ ,x,m)} \left|\hat{\ }_{mn}^{g} - \ {}_{mn}^{g}\right|$$

and then (by Cauchy inequality)

$$\begin{aligned}
[\hat{g}(x;K) - g(x;K)]^2 \ &\leq\ 2M_\varphi^2 2^M \left(\sum_{n=n_{\min}(\varphi,x,M)}^{n_{\max}(\varphi,x,M)} \left|\hat{\alpha}_{Mn}^g - \alpha_{Mn}^g\right|\right)^2 \\
&\qquad + 2M^2 \left(\sum_{m=M}^{K-1} 2^{m/2} \sum_{n=n_{\min}(\ ,x,m)}^{n_{\max}(\ ,x,m)} \left|\hat{\ }_{mn}^{g} - \ {}_{mn}^{g}\right|\right)^2 \\
&=\ 2\left(e_1^2 + e_2^2\right), \quad \text{say}
\end{aligned} \tag{57}$$

For the first component, using again Cauchy inequality and including (42) in Appendix I, we get

$$e_1^2 \leq M_\varphi^2\, S\, 2^M \sum_{n=n_{\min}(\varphi,x,M)}^{n_{\max}(\varphi,x,M)} \left(\hat{\alpha}_{Mn}^g - \alpha_{Mn}^g\right)^2 \tag{58}$$

For the second component, after appropriate rearrangements, we have

$$e_2^2 = M^2 \sum_{m=M}^{K-1} 2^{m/2} \sum_{m^\theta=M}^{K-1} 2^{m^\theta/2} \sum_{n=n_{\min}(\ ,x,m)}^{n_{\max}(\ ,x,m)} \sum_{n^\theta=n_{\min}(\ ,x,m^\theta)}^{n_{\max}(\ ,x,m^\theta)} \left|\hat{\ }_{mn}^{g} - \ {}_{mn}^{g}\right| \left|\hat{\ }_{m^\theta n^\theta}^{g} - \ {}_{m^\theta n^\theta}^{g}\right| \tag{59}$$

Now, taking into account (56)-(59), the fact that

$$E\left(\hat{\alpha}^g_{Mn} - \alpha^g_{Mn}\right)^2 \;=\; var\left(\hat{\alpha}^g_{Mn}\right)$$

$$E\left|\hat{}^{g}_{mn} - {}^{g}_{mn}\right|\left|\hat{}^{g}_{m^0n^0} - {}^{g}_{m^0n^0}\right| \;\leq\; \max\left\{var\left(\hat{}^{g}_{mn}\right), var\left(\hat{}^{g}_{m^0n^0}\right)\right\}$$

and (42) in Appendix I along with the bounds (53) and (54) in Appendix II, we see that for $x \in [a,b]$ it holds that

$$var\ \{\hat{g}(x;K)\} \leq 2\,S^2 \max\left\{M^2_\varphi, M^2\right\} C^g_\alpha \left[2^M + \left(\sum_{m=M}^{K-1} 2^{m/2}\right)^2\right] N^{-1}$$

Since

$$2^M + \left(\sum_{m=M}^{K-1} 2^{m/2}\right)^2 = 2^K \left[2^{-(K-M)} + \left(1/\left(\sqrt{2}-1\right)\right)^2 \left(1 - 2^{-(K-M)/2}\right)^2\right]$$

and in the wavelet models $K > M$, we get eventually

$$var\ \{\hat{g}(x;K)\} \leq \left(C_g\, 2^K\right) N^{-1} \tag{60}$$

where $C_g = 2(1 + 1/(\sqrt{2}-1)^2) S^2 \max\{M^2_\varphi, M^2\} C^g_\alpha$, for each point $x \in [a,b]$ and each scale K.

Due to the similarity of the models (8), (17) to those in (7), (16), and similarity of the variance bounds of $\hat{\alpha}^f_{Mn}$ and $\hat{}^{f}_{mn}$ to those in (53), (54), we obtain after analogous steps that

$$var\left\{\hat{f}(x;K)\right\} \leq \left(C_f\, 2^K\right) N^{-1} \tag{61}$$

where $C_f = 2(1 + 1/(\sqrt{2}-1)^2) S^2 \max\{M^2_\varphi, M^2\} M_f$, for any $x \in [a,b]$ and any K.

APPENDIX IV

APPROXIMATION ERROR OF THE WAVELET MODELS

Let $F(x) \in C^{\ F}(x_0)$ where $_F \in (0,1]$ or $_F \in \mathbb{N}$ (the set of natural numbers). For the wavelet functions (x) with r vanishing moments (cf. (46) in Appendix I) and large values of m, for the wavelet coefficients $_{mn}$ ((40) in Appendix I) it holds that (cf. e.g. [12])

$$|\ _{mn}| \leq C^F\ 2^{-(\lambda_F\ + \frac{1}{2})m} \qquad \text{all } n \tag{62}$$

some $C^F > 0$ independent of m where $\lambda_F = \min\{\ _F, r\ + 1\}$. Hence the approximation error $AE(F;x;K)$ as in (44) in Appendix I at the point x_0 is bounded for large K as follows (see (62) and (44) in Appendix I)

$$|AE(F;x_0;K)| \leq S\,M\ \ C^F \sum_{m=K}^{\infty} 2^{m/2} 2^{-(\lambda_F\ + \frac{1}{2})m}$$

yielding

$$|AE(F;x_0;K)| \leq C_{AE,F}\ \ 2^{-\lambda_F\ K} \tag{63}$$

where $C_{AE,F} = S\,M\ \ C^F/(1 - 2^{-\lambda_F}\ )$ is a constant.

APPENDIX V
TECHNICAL LEMMAS

Let $\{\vartheta_N\}$ and $\{\eta_N\}$ be sequences of random variables and $\{a_N\}, \{b_N\}$ be sequences of positive numbers such that $a_N, b_N \to 0$ as $N \to \infty$.

Lemma 3: If $E\vartheta_N^2 \leq Ca_N$ some $C > 0$ independent of N then $\vartheta_N = O(a_N^{1/2})$ in probability.

Proof: From $E\,\vartheta_N^2 \leq Ca_N$, by Chebychev's inequality, $P\{|r_N|\,(|\vartheta_N|\,/b_N) > \varepsilon\} \leq \left(C/\varepsilon^2\right)\, r_N^2\, \left(a_N/b_N^2\right)$ for each $\varepsilon > 0$, i.e. (25) holds for $b_N = a_N^{1/2}$. ■

Lemma 4: If $\vartheta_N = a + O(a_N)$ in probability and $\eta_N = b + O(b_N)$ in probability where $b \neq 0$, then $\vartheta_N/\eta_N = a/b + O\left(\max\{a_N, b_N\}\right)$ in probability.

Proof: It holds that (see proof of Lemma 2 in [25])

$$\left|\frac{\vartheta_N}{\eta_N} - \frac{a}{b}\right| \leq \left|\frac{\vartheta_N}{\eta_N}\right| \frac{|\eta_N - b|}{|b|} + \frac{|\vartheta_N - a|}{|b|}$$

Assume $|\vartheta_N - a| \leq |a|\,(\varepsilon/\,(2+\varepsilon))\,(a_N/\,|r_N|)$ and $|\eta_N - b| \leq |b|\,(\varepsilon/\,(2+\varepsilon))\,(b_N/\,|r_N|)$. Making use of the above inequality, we get that

$$\left|\frac{\vartheta_N}{\eta_N} - \frac{a}{b}\right| \left(1 - \left(\frac{\varepsilon}{2+\varepsilon}\right)\frac{b_N}{|r_N|}\right) \leq 2\left|\frac{a}{b}\right| \left(\frac{\varepsilon}{2+\varepsilon}\right) \frac{\max\{a_N, b_N\}}{|r_N|}$$

and for r_N arbitrarily slowly tending to zero this yields for large N that

$$\left|\frac{\vartheta_N}{\eta_N} - \frac{a}{b}\right| \leq \left|\frac{a}{b}\right| \varepsilon\, \frac{\max\{a_N, b_N\}}{|r_N|}$$

Therefore

$$\begin{aligned} P\left\{|r_N|\left|\frac{\vartheta_N}{\eta_N} - \frac{a}{b}\right| / \max\{a_N, b_N\} > \left|\frac{a}{b}\right|\varepsilon\right\} \;\leq\; & P\left\{|r_N|\,\frac{|\vartheta_N - a|}{a_N} > |a|\left(\frac{\varepsilon}{2+\varepsilon}\right)\right\} \\ & + P\left\{|r_N|\,\frac{|\eta_N - b|}{b_N} > |b|\left(\frac{\varepsilon}{2+\varepsilon}\right)\right\} \end{aligned}$$

each $\varepsilon > 0$, which - including (25) in Section V - ends the proof. ■

Acknowledgments

The authors wish to thank the Reviewers for their helpful comments and Professor Alexander A. Georgiev from the Louisiana State University for his friendly support.

REFERENCES

[1] J. S. Bendat, *Nonlinear System Analysis and Identification*. New York: Wiley, 1990.

[2] Z. Hasiewicz, "Applicability of least-squares to the parameter estimation of large-scale no-memory linear composite systems," *International Journal of Systems Science*, vol. 20, pp. 2427–2449, 1989.

[3] W. Greblicki and A. Krzyżak, "Nonparametric identification of a memoryless system with a cascade structure," *International Journal of Systems Science*, vol. 10, pp. 1301–1310, 1979.

[4] R. Haber and L. Keviczky, *Nonlinear System Identification: Input-Output Modeling Approach*. Dordrecht-Boston-London: Kluwer Academic Publishers, 1999.

[5] W. Greblicki and M. Pawlak, "Identification of discrete Hammerstein system using kernel regression estimates," *IEEE Transactions on Automatic Control*, vol. 31, pp. 74–77, 1986.

[6] W. Greblicki and M. Pawlak, "Nonparametric identification of Hammerstein systems," *IEEE Transactions on Information Theory*, vol. 35, pp. 409–418, 1989.

[7] W. Greblicki, "Nonparametric orthogonal series identification of Hammerstein systems," *International Journal of Systems Science*, vol. 20, pp. 2355–2367, 1989.

[8] M. Pawlak, "On the series expansion approach to the identification of Hammerstein systems," *IEEE Transactions on Automatic Control*, vol. 36, pp. 763–767, 1991.

[9] M. Pawlak and Z. Hasiewicz, "Nonlinear system identification by the Haar multiresolution analysis," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 45, pp. 945–961, 1998.

[10] Z. Hasiewicz, "Non-parametric estimation of non-linearity in a cascade time series system by multiscale approximation," *Signal Processing*, vol. 81, pp. 791–807, 2001.

[11] C. J. Stone, "Optimal rates of convergence for nonparametric regression," *Annals of Statistics*, vol. 8, pp. 1348–1360, 1980.

[12] S. G. Mallat, *A Wavelet Tour of Signal Processing*. San Diego: Academic Press, 1998.

[13] A. Chambolle, R. A. Devore, N. Y. Lee, and B. J. Lucier, "Nonlinear wavelet image-processing - variational-problems, compression, and noise removal through wavelet shrinkage," *IEEE Transactions on Image Processing*, vol. 7, pp. 319–335, 1998.

[14] Y. Kitada, "Identification of nonlinear structural dynamic systems using wavelets," *Journal of Engineering Mechanics-ASCE*, vol. 124, pp. 1059–1066, 1998.

[15] N. Sureshbabu and J. A. Farrell, "Wavelet based system identification for non-linear control," *IEEE Transactions on Automatic Control*, vol. 44, pp. 412–417, 1999.

[16] D. Coca and S. A. Billings, "Non-linear system identification using wavelet multiresolution models," *International Journal of Control*, vol. 74, pp. 1718–1736, 2001.

[17] R. Ghanem and F. Romeo, "A wavelet-based approach for model and parameter identification of non-linear systems," *International Journal of Nonlinear Mechanics*, vol. 36, pp. 835–859, 2001.

[18] Z. Hasiewicz and P. Śliwiński, "Identification of non-linear characteristics of a class of block-oriented non-linear systems via Daubechies wavelet-based models," *International Journal of Systems Science*, vol. 14, pp. 1121–1144, 2002.

[19] H. W. Chen, "Modelling and identification of parallel non-linear systems: Structural classification and parameter estimation methods," *Proceedings of IEEE*, vol. 83, pp. 39–66, 1995.

[20] J. Vörös, "Iterative algorithm for parameter identification of Hammerstein systems with two-segment nonlinearities," *IEEE Transactions on Automatic Control*, vol. 44, pp. 2145–2149, 1999.

[21] S. Verdu, *Multiuser Detection*. London: Cambridge University Press, 1998.

[22] G. B. Giannakis and E. Serpedin, "A bibliography on nonlinear system identification," *Signal Processing*, vol. 81, pp. 533–580, 2001.

[23] L. Ljung, *System Identification - Theory For the User*. Englewood Cliffs: Prentice-Hall, 1987.

[24] Y. Yang, "Mixing strategies for density estimation," *Annals of Statistics*, vol. 28, pp. 75–87, 2000.

[25] W. Greblicki and M. Pawlak, "Fourier and Hermite series estimates of regression function," *Annals of The Institute of Statistical Mathematics*, vol. 37, pp. 443–455, 1985.

[26] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: SIAM Edition, 1992.

[27] S. Kelly, M. Kon, and M. Raphael, "Pointwise convergence of wavelet expansions," *Bulletin of The American Mathematical Society*, vol. 30, pp. 87–94, 1994.

CAPTIONS OF FIGURES

Fig. 1. General system under consideration.

Fig. 2.

(a) Parallel system,

(b) Hammerstein system.

Fig. 3.

(a) Hammerstein system with two-segment nonlinearity,

(b) Hammerstein system with nuisance nonlinearity.

Fig. 4.

(a) Hammerstein system with nuisance dynamics,

(b) Uryson system.

Fig. 5. Two-channel system.

Fig. 6. Accuracy of the models $R_N(x;p)$ and the trigonometric series model against the sample size N (in log-scale) and the wavelet number p for:

(a) polynomial,

(b) cube root,

(c) quantizer nonlinearity□

FIR dynamics and MA(2) output noise.

Fig. 7. Accuracy of the models $R_N(x;p)$ and the trigonometric series model against the sample size N (in log-scale) and the wavelet number p for:

(a) polynomial,

(b) cube root,

(c) quantizer nonlinearity□

IIR dynamics and MA(2) output noise,

Fig. 8. Effect of the correlation of the output noise on the accuracy of the models $R_N(x;p^*)$ for:

(a) polynomial,

(b) cube root,

(c) quantizer nonlinearity□

MA(l) output noise model, FIR dynamics.

CAPTIONS OF TABLES

Table 1. THE ESTIMATION ERROR $error(N)$ FOR POLYNOMIAL AND QUANTIZER NONLINEARITY (FIR DYNAMICS AND MA(2) OUTPUT NOISE)

Table 2. BASIC PROPERTIES OF TYPICAL ORTHOGONAL WAVELET FUNCTIONS WITH COMPACT SUPPORT (p – WAVELET NUMBER)

BRIEF TECHNICAL BIOGRAPHIES

Zygmunt Hasiewicz was born in 1948 in Poland. He received the M.Sc. and Ph.D. degrees in control engineering from the Wroclaw University of Technology, Poland, in 1971 and 1974, respectively, and the D.Sc. degree from the Warsaw University of Technology, Poland, in 1993. In 1971 he joined the Institute of Engineering Cybernetics, Wroclaw University of Technology, where he is currently an Associate Professor. In 1976 he held visiting appointments at Lille University, France, and in 1995 and 2001 he was a visiting professor at the University of Manitoba, Winnipeg, Canada. His research interests include nonlinear system modeling and statistical methods in composite system identification. He is a reviewer for *Zentralblatt für Mathematik*.

Miroslaw Pawlak (M'85) received the M.Sc. and Ph.D. degrees in computer engineering from the Wroclaw University of Technology, Poland, in 1978 and 1982, respectively. In 1982 he joined the Institute of Engineering Cybernetics, Wroclaw University of Technology as an Assistant Professor. From 1984 to 1985, he was a visiting Research Associate at the Department of Computer Science, Concordia University, Montreal. Since 1985 he has been with the University of Manitoba, Winnipeg, Canada, where he is currently a Professor at the Department of Electrical and Computer Engineering. He is also holding a scientific position at the Telecommunication Research Laboratories (*TRLabs*). In 1992 he was at the University of Ulm, Germany as an *Alexander von Humboldt Foundation* fellow. He has also held a number of visiting positions in universities in Europe, USA and Australia. He has been an associate editor of *Pattern Recognition* and *Sampling Theory in Signal and Image Processing*. His research interests are in statistical signal processing, pattern analysis, communication theory and nonparametric estimation. He published over 150 papers on pattern recognition, signal processing, system identification, and nonparametric estimation.

Przemysław Śliwiński was born in 1971 in Poland. He received the M.Sc. and Ph.D. degrees in computer engineering from the Wroclaw University of Technology, Poland, in 1996 and 2000, respectively. In 2000 he joined the Institute of Engineering Cybernetics, Wroclaw University of Technology, where he is currently an Assistant Professor. His research interests are in nonparametric system identification, numerical analysis and artificial neural networks. He is a reviewer for *IEEE Transactions on Signal Processing*.