NON-PARAMETRIC IDENTIFICATION OF MULTI-CHANNEL SYSTEMS BY MULTISCALE EXPANSIONS

M. Pawlak

Electrical and Computer Engineering,
University of Manitoba,
Winnipeg, Manitoba,
R3T5V6 Canada
pawlak@ee.umanitoba.ca

Z. Hasiewicz

Institute of Engineering Cybernetic,
Wroclaw University of Technology,
Janiszewskiego 11/17,
50-372 Wroclaw, Poland,
zhas@ict.pwr.wroc.pl

ABSTRACT

The paper deals with the problem of reconstruction of nonlinearities in a certain class of nonlinear dynamical systems of the multichannel form. A prior information about the system is very limited excluding the standard parametric approach to the problem. The system structure consists of nonlinearities being embedded in a block oriented structure containing dynamic linear subsystems and other "nuisance" nonlinearities. The multiresolution idea, being the fundamental concept of the modern wavelet theory, is adopted and multiscale expansions associated with a large class of scaling functions are applied to construct nonparametric identification techniques of the nonlinearities. The pointwise convergence properties of the proposed identification algorithms are established. Conditions for the convergence are given; and for nonlinearities satisfying a local Lipschitz condition, the rate of convergence is evaluated.

1. INTRODUCTION

A large class of physical systems are 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 problem of complete determination of a system description from 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 [1, 2]. A promising approach is based on 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. Such composite models have found numerous applications in such distant areas as biology, communication systems, chemical engineering, psychology and sociology [1, 3, 2]. A class of cascade/parallel models is a particularly popular type of block-oriented structures. Examples of such models include cascade Hammerstein, Wiener and sandwich structures and their parallel counterparts, [4, 1, 5, 6, 7]. These models become especially attractive if one allows a general class of nonlinear characteristics not being able to be parametrized and smooth, e.g., not being just a polynomial of a finite order. We refer to [4, 1, 2, 7] for parametric identification techniques of the cascade/parallel block-oriented models with polynomial nonlinearities. The parametric restriction is, however, often too rigid. In [5, 6, 8, 9, 10, 11] the nonparametric approach to identification of the cascade/parallel block-oriented models has been proposed. 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.

In this paper we extend the nonparametric identification approach to the case of multi-channel nonlinear, block-oriented models. This kind of models appear in areas as diverse as multisensor systems, data fusion, multiuser detection, and biological systems with multiple excitations [3, 12, 13]. Surprisingly there has been a little effort made in identifying multi-channel nonlinear systems. In fact most of the theoretical developments of nonlinear identification to date have dealt with single-channel systems. In this paper we are interested in recovering the system nonlinearities in each channel which are embedded in a block oriented structure containing dynamic linear subsystems and other "nuisance" nonlinearities. Our approach to function recovering is based on regression analysis and we propose the identification algorithms originating from the area of nonparametric regression [14]. The proposed estimate is based on the theory of orthogonal bases originating from multiscale and wavelet approximations. This theory provides elegant techniques for representing the levels of details of the approximated function. Multiresolution and wavelet theory has found recently applications in a remarkable diversity of disciplines [15]. A little attention, however, has been paid to the application of the multiresolution and wavelet methodology to system theory and to nonlinear system identification in particular, see [16, 10, 11, 17] for some preliminary studies into this direction. For the proposed identification algorithms we show that they converge for a large class of nonlinear characteristics and under very mild conditions on the model dynamics. The rigorous convergence properties are established and the best possible pointwise rate of convergence is found.

2. MULTI-CHANNEL NONLINEAR SYSTEMS

Let us now introduce a class of nonlinear multi-channel dynamical systems examined in this paper. A system of this class is characterized by the general property that memoryless nonlinear characteristics are separated from the dynamical part the system. Although

0-7803-7402-9/02/$17.00 ©2002 IEEE

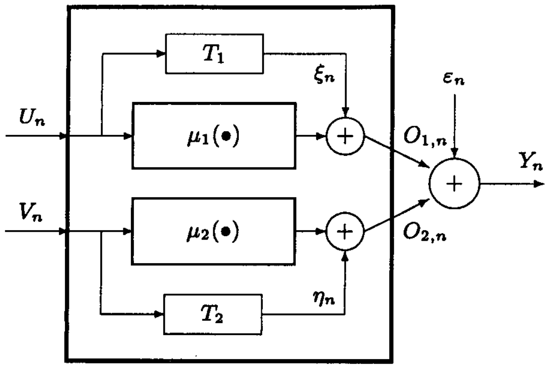

Fig. 1. Two-channel nonlinear dynamic system

the system structure is known we assume no prior model of the system characteristics, i.e., we are interested in a nonparametric approach to system identification. Throughout the paper we examine, without much loss of generality, the two-channel nonlinear model (depicted in Fig. 1) of the following form

$$\begin{cases} O_{1,n} = \mu_1(U_n) + \xi_n \\ O_{2,n} = \mu_2(V_n) + \eta_n \\ Y_n = O_{1,n} + O_{2,n} + \varepsilon_n \end{cases}, \qquad (2.1)$$

where $(U_n, V_n; Y_n)$ are (input; output) signals, $\mu_1(u)$, $\mu_2(v)$ represent the unknown system nonlinearities in the each channel, $\{\xi_n\}$, $\{\eta_n\}$ are the "system noise" processes characterizing the history of the system, and $\{\varepsilon_n\}$ is the measurement noise. It is assumed that the blocks T_1, T_2 in Fig. 1, representing the system noise processes ξ_n, η_n, are measurable transformations of past inputs and are given by

$$\xi_n = \sum_{j=1}^{p} s_{1,j}\lambda_{1,j}(U_{n-j}), \eta_n = \sum_{j=1}^{q} s_{2,j}\lambda_{2,j}(V_{n-j}) \qquad (2.2)$$

Here $\{\lambda_{1,j}(u)\}$, $\{\lambda_{2,j}(v)\}$ are some nonlinearities and $\{s_{1,j}\}$, $\{s_{2,j}\}$ are some weight factors. All these quantities are assumed to be unknown. It should be noted that the memory length for ξ_n is p whereas for η_n is q. The memory lengths p, q are unknown and take values between zero and infinity. A large number of different nonlinear systems can be obtained by choosing various combinations of the aforementioned parameters. For instance $p = q = 0$ yields $Y_n = \mu_1(U_n) + \mu_2(V_n) + \varepsilon_n$. This is an additive model, a well studied structure in the area of non-parametric multiple regression modeling and estimation [18, 19, 20].

Our principal goal in this paper is to recover the non-linear characteristics $\mu_1(u)$, $\mu_2(v)$ in (2.1) from the input-output training data $\{(U_1, V_1; Y_1), \ldots, (U_n, V_n; Y_n)\}$.

The following assumptions concerning the model in (2.1) are required in the paper.

Assumption 1 *The inputs $\{(U_1, V_1), (U_2, V_2), \ldots\}$ form a sequence of independent and identically distributed random variables which are independent of $\{\varepsilon_n\}$. The joint probability density f_{UV} of (U, V) exists and moreover $f_{UV} \in L_2(R^2)$. We also assume that f_{UV} is strictly bounded away from zero and infinity.*

Assumption 2 *For the system noise processes $\{\xi_n\}$, $\{\eta_n\}$ we have $E\lambda_{1,j}(U) = 0$, $E\lambda_{2,i}(V) = 0$ with $E\lambda_{1,j}^2(U) < \infty$, $E\lambda_{2,i}^2(V) < \infty$, for $j = 1, 2, \ldots, p$ and $i = 1, 2, \ldots, q$.*

Assumption 3 *The nonlinear characteristics μ_1, μ_2 satisfy $E\mu_1^2(U) < \infty$ and $E\mu_2^2(V) < \infty$.*

Assumption 4 *The measurement noise $\{\varepsilon_n\}$ is uncorrelated and such that $E\varepsilon_n = 0$, var $\varepsilon_n < \infty$.*

Let us elaborate on the role of the above conditions. Assumption 1 is required since we use the $L_2(R^2)$ multiscale decomposition of f_{UV}. The boundness of f_{UV} away from zero is necessary since we can only consider the estimation problem in such points where the input density is sufficiently high. Assumption 2 is necessary for $\{\xi_n\}$ to be the second order covariance stationary stochastic process with $E\xi_n = 0$, var $\xi_n < \infty$. The condition on $\{\eta_n\}$ can be interpreted analogously. This along with Assumptions 3, 4 make the output process $\{Y_n\}$ well defined, i.e., it is a second order covariance stationary stochastic process.

There is a large class of nonlinear multi-channel models which fall in the description given by (2.1). This includes multi-channel Hammerstein models consisting of nonlinear static elements followed by linear dynamic systems, a multi-channel parallel system where each channel is of the form of the parallel connection of a memoryless nonlinear element and a linear dynamic system, and multi-channel systems being combinations of the aforementioned connections.

3. IDENTIFICATION ALGORITHMS AND THEIR ACCURACY

Our reconstruction methods relay on the theory of the nonparametric regression aiming to find a relationship between the input and output variables of the system [14]. The regression function of Y_n on (U_n, V_n), being the orthogonal projection of Y_n on the subspace spanned by (U_n, V_n), is given by $m(u,v) = E\{Y_n|U_n = u, V_n = v\}$. By Assumptions 1, 2, and 4 we obtain that $m(u,v) = \mu_1(u)+\mu_2(v)$. Hence one cannot recover the nonlinearities $\mu_1(u)$, $\mu_2(v)$ from the regression function $m(u,v)$. Furthermore, using one dimensional regression functions $m_1(u) = E\{Y_n|U_n = u\}$, $m_2(v) = E\{Y_n|V_n = v\}$ we get $m_1(u) = \mu_1(u) + E\{\mu_2(V_n)|U_n = u\}$ and $m_2(v) = \mu_2(v) + E\{\mu_1(U_n)|V_n = v\}$. Once again the recovery is impossible unless the inputs U_n and V_n are statistically independent, not very realistic assumption in practice. To overcome these difficulties one can use the method of marginal integration introduced originally in [19] for the case of memoryless models. The underlying idea of the integration method is to estimate $\mu_1(u)$, $\mu_2(v)$ by integrating a suitable pilot estimate of $m(u,v)$ with respect to a specific density function. In particular by using the marginal densities f_U, f_V of f_{UV} in the integration process we obtain the following fundamental identities

$$M_1(u) = \int_{-\infty}^{\infty} m(u,v) f_V(v) dv = \mu_1(u) + c_2 \qquad (3.1)$$

and

$$M_2(v) = \int_{-\infty}^{\infty} m(u,v) f_U(u) du = \mu_2(u) + c_1, \qquad (3.2)$$

where $c_1 = E\mu_1(U)$ and $c_2 = E\mu_2(V)$. Therefore the nonlinearities $\mu_1(u)$, $\mu_2(v)$ are identifiable from $M_1(u)$ and $M_2(v)$ up to the additive constants c_1, c_2. It is worth mentioning that under some mild conditions we have $c_1 = c_2 = 0$. This can happen if, e.g., $f_U(u)$ and $f_V(v)$ are even whereas $\mu_1(u)$ and $\mu_2(v)$ are odd. Formulas (3.1), (3.2) suggest a simple estimation strategy, i.e., first estimate $m(u, v)$ by $\hat{m}(u, v)$ and then identify $\mu_1(u)$, $\mu_2(v)$ from

$$\begin{aligned} \hat{\mu_1}(u) &= n^{-1}\sum_{s=1}^{n}\hat{m}(u, V_s) \\ \hat{\mu_2}(v) &= n^{-1}\sum_{s=1}^{n}\hat{m}(U_s, v) \end{aligned}, \tag{3.3}$$

Here and through the paper we assume that $c_1 = c_2 = 0$. Note that we replaced the integration by an empirical average. Thus (3.3) provides a generic estimation scheme for estimating $\mu_1(u)$, $\mu_2(v)$. In this paper we utilize the multiresolution representations to form a pilot estimate $\hat{m}(u, v)$ of $m(u, v)$. Hence let $\Phi(u, v)$ be a scaling function in $L_2(R^2)$ with the associated wavelet functions $\{\Psi^\lambda(u, v), \lambda = 1, 2, 3\}$ such that the so-called multiresolution decomposition of $L_2(R^2)$ takes place [15]. It can be shown that in this case $\{\Phi_{J,\mathbf{k}}(u, v), J \in Z, \mathbf{k} \in Z^2\}$ as well as $\{\Psi_{J,\mathbf{k}}(u, v), J \in Z, \mathbf{k} \in Z^2, \lambda = 1, 2, 3\}$ form orthonormal bases for $L_2(R^2)$, where

$$\begin{aligned} \Phi_{J,\mathbf{k}}(u, v) &= 2^J\Phi(2^J u - k_1, 2^J v - k_2), \\ \Psi^\lambda_{J,\mathbf{k}}(u, v) &= 2^J\Psi^\lambda(2^J u - k_1, 2^J v - k_2). \end{aligned}$$

Let $f_J(u, v)$ be the orthogonal projection of $f \in L_2(R^2)$ onto the J-th resolution subspace, i.e.,

$$f_J(u, v) = \sum_{\mathbf{k}\in Z^2} a_{J,\mathbf{k}}\Phi_{J,\mathbf{k}}(u, v),$$

where $a_{J,\mathbf{k}} = \langle f, \Phi_{J,\mathbf{k}}\rangle$ is the Fourier coefficient. Returning to our estimation problem let us first project the functions $g(u, v) = m(u, v)f_{UV}(u, v)$ and $f_{UV}(u, v)$ onto the J-th resolution subspace. Then expressing the Fourier coefficients of $g(u, v)$, $f_{UV}(u, v)$ in terms of expected values [11] we can easily derive the following estimate of $m(u, v)$

$$\hat{m}_J(u, v) = \frac{\hat{g}_J(u, v)}{\hat{f}_{UV,J}(u, v)}, \tag{3.4}$$

where

$$\begin{aligned} \hat{g}_J(u, v) &= \sum_{\mathbf{k}\in Z^2} \hat{a}_{J,\mathbf{k}}\Phi_{J,\mathbf{k}}(u, v), \\ \hat{f}_{UV,J}(u, v) &= \sum_{\mathbf{k}\in Z^2} \hat{\alpha}_{J,\mathbf{k}}\Phi_{J,\mathbf{k}}(u, v) \end{aligned}$$

are the estimates of $g(u, v)$, $f_{UV}(u, v)$, respectively. Here $\hat{a}_{J,\mathbf{k}} = n^{-1}\sum_{i=1}^{n} Y_i\Phi_{J,\mathbf{k}}(U_i, V_i)$, $\hat{\alpha}_{J,\mathbf{k}} = n^{-1}\sum_{i=1}^{n}\Phi_{J,\mathbf{k}}(U_i, V_i)$ are the empirical Fourier coefficients. Combining (3.3) and (3.4) we can define the following estimates of the nonlinear characteristics of the two-channel system

$$\begin{aligned} \hat{\mu}_{1J}(u) &= n^{-1}\sum_{s=1}^{n}\hat{m}_J(u, V_s) \\ \hat{\mu}_{2J}(v) &= n^{-1}\sum_{s=1}^{n}\hat{m}_J(U_s, v) \end{aligned}. \tag{3.5}$$

It is common to determine $\Phi(u, v)$ and $\{\Psi^\lambda(u, v), \lambda = 1, 2, 3\}$ from the univariate multiresolution analysis. Thus let ϕ and ψ be one-dimensional scaling and wavelet functions yielding the muliresolution decomposition of $L_2(R)$. Then one can define $\Phi(u, v) = \phi(u)\phi(v)$ and $\Psi^1(u, v) = \psi(u)\phi(v)$, $\Psi^2(u, v) = \phi(u)\psi(v)$, $\Psi^3(u, v) = \psi(u)\psi(v)$. Such a class of scaling and wavelet functions is assumed in the paper. Moreover we assume [21] the following.

Assumption 5 *The scaling and wavelet functions $\phi(u)$, $\psi(u)$ are radially bounded by decreasing functions decaying like $O((1 + |u|)^{-D})$, $D > 1$.*

A number of practical scaling and wavelet functions satisfy Assumption 5. For instance the Lemarie and Meyer scaling [15] function satisfies Assumption 5 with $D \geq 2$. The spline wavelet analysis corresponds to scaling functions with the exponential radial bounds. Assumption 5 is trivially satisfied by the Daubechie class of compactly supported scaling/wavelet functions.

The accuracy of our estimates depends on the smoothness of $\mu_1(u)$, $\mu_2(v)$ as well as $f_U(u)$, $f_V(v)$. In this paper we use the local Lipschitz condition. Hence let $Lip(\alpha; x_0)$ denote the collection of functions $f \in L_2(R)$ such that $|f(x_0 + \delta) - f(x_0)| \leq L_f|\delta|^\alpha$, where $0 < \alpha \leq 1$, L_f is some positive constant and δ, $|\delta| < 1$, determines the size of a small neighborhood around x_0. If $f \in Lip(\alpha; x_0)$ one can say that f has a fractional derivative of order α at x_0. Note if $f \in Lip(\alpha; x_0)$ then f need not be continuous on R. The following theorem gives the rate of convergence for $\hat{\mu}_{1J}(u)$, $\hat{\mu}_{2J}(v)$ in the mean squared error sense.

Theorem 1 *Let Assumptions 1–4 be satisfied. Let Assumption 5 hold with $D > 1$. Assume that $\mu_1 \in Lip(\alpha_1, u_0)$, $f_U \in Lip(\beta_1, u_0)$ and $\mu_2 \in Lip(\alpha_2, v_0)$, $f_V \in Lip(\beta_2, v_0)$. Then selecting*

$$J_1 = \frac{\delta_1 + D - 1}{\delta_1(2D - 1) + D - 1}\log_2(n)$$

we have

$$E(\hat{\mu}_{1J_1}(u_0) - \mu_1(u_0))^2 = O\left(n^{\frac{-2\delta_1(D-1)}{\delta_1(2D-1)+D-1}}\right),$$

and selecting

$$J_2 = \frac{\delta_2 + D - 1}{\delta_2(2D - 1) + D - 1}\log_2(n)$$

we have

$$E(\hat{\mu}_{2J_2}(v_0) - \mu_2(v_0))^2 = O\left(n^{\frac{-2\delta_2(D-1)}{\delta_2(2D-1)+D-1}}\right),$$

where $\delta_1 = \min(\alpha_1, \beta_1)$, $\delta_2 = \min(\alpha_2, \beta_2)$.

Remark 1 *Let us observe that for compactly supported scaling functions the above rates become $O(n^{\frac{-2\delta_1}{2\delta_1+1}})$, $O(n^{\frac{-2\delta_2}{2\delta_2+1}})$. This is the optimal rate of convergence since it agrees with the best possible rate attained by any linear nonparametric regression estimation [20]. In particular, if $\alpha_i = \beta_i = 1$, $i = 1, 2$ then the rate is of order $O(n^{-2/3})$, where the resolution level J must be selected as $J = 3^{-1}\log_2(n)$.*

Remark 2 *We have applied the same scaling/wavelet functions to form the pilot estimate $\hat{m}_J$ used for defining $\hat{\mu}_{1J}$, $\hat{\mu}_{2J}$. Generally the nonlinear characteristics μ_1, μ_2 can have much different smoothness and a different pilot estimate is needed for each estimated nonlinearity. This can be achieved by selecting scaling/wavelet functions of different forms and orders (quantified by a number of vanishing moments).*

Remark 3 *The detail analysis reveals that the estimates $\hat{\mu}_{1J}$, $\hat{\mu}_{2J}$ are negatively biased. This is due to the fact that in (3.5) one uses the same data for learning and empirical averaging. The reduced bias estimate can be obtained if we apply the leave-one-out strategy. Thus the estimate $\hat{\mu}_{1J}$ takes the following form $\tilde{\mu}_{1J}(u) = n^{-1}\sum_{i=1}^{n}\hat{m}_{J,-i}(u, V_i)$, where $\hat{m}_{J,-i}(u,v)$ is the estimate $\hat{m}_J(u,v)$ with the observation $(U_i, V_i; Y_i)$ deleted.*

Thus far we have examined the two-channel nonlinear system. The extension to M-channel case is straightforward and the result of Theorem 1 still holds. Nevertheless, there is a change in the asymptotic constants which may depend on M. This seems yields some degradation in the performance of the integrated estimates. Therefore for very large M one would like to eliminate some "weak" channels. This can be done by forming some measure of strength for μ_s, $s = 1, \ldots, M$ using the estimates like in (3.3). This interesting problem is beyond the scope of this paper. Further improvement in the accuracy of our estimates might be expected if nonlinear multiscale approximations (employing some kind of wavelet thresholding rules) are applied [15, 22]. This has been extremely successful strategy for signal and image denoising. Nevertheless in this paper we consider dynamical systems, i.e., we deal with dependent data, and it is not clear whether wavelet thresholding may play any significant role.

4. CONCLUDING REMARKS

In this paper the problem of identifying nonlinearities in a broad class of multi-channel nonlinear models has been addressed. No a prior information about the nonlinear characteristics and input signal probability density function is required making the identification problem of a nonparametric type. Using the concept of the multivariate regression function estimation and the marginal integration idea, the nonparametric identification algorithms using multiscale expansions are formed. Their rigorous convergence properties are established and, in particular, the best possible local rate of convergence is obtained. The convergence results hold under very mild restrictions on the nonlinear characteristic and the input density function as well as on the system dynamics.

5. REFERENCES

[1] H. W. Chen, "Modeling and identification of parallel nonlinear systems: structural classification and parameter estimation methods," *Proc. IEEE*, vol. 83, pp. 39–66, 1995.

[2] O. Nelles, *Nonlinear System Identification*, Springer Verlag, 2001.

[3] P. Z. Marmarelis and V. Z. Marmarelis, *Analysis of Physiological Systems. The White Noise Approach*, Plenum Press, 1978.

[4] N. J. Bershad, P. Celka, and S. McLaughlin, "Analysis of stochastic gradient identification of Wiener-Hammerstein systems for nonlinearities with Hermite polynomial expansions," *IEEE Trans. on Signal Processing*, vol. 49, pp. 1060–1072, 2001.

[5] W. Greblicki, "Nonparametric identification of Wiener systems," *IEEE Trans. Infor. Theory*, vol. 38, pp. 1487–1493, 1992.

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

[7] J. C. Ralston, A. M. Zoubir, and B. Boashash, "Identification of class of nonlinear systems under stationary non-Gaussian excitation," *IEEE Trans. on Signal Processing*, vol. 45, pp. 719–735, 1997.

[8] W. Greblicki and M. Pawlak, "Nonparametric identification of a particular nonlinear time series system," *IEEE Trans. Signal Processing*, vol. 40, pp. 985–989, 1992.

[9] W. Greblicki and M. Pawlak, "Dynamic system identification with order statistics," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1474–1489, 1994.

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

[11] M. Pawlak and Z. Hasiewicz, "Nonlinear system identification by the Haar multiresolution analysis," *IEEE Trans. Circuits and Systems*, vol. 45, pp. 945–961, 1998.

[12] S. Verdu, *Multiuser Detection*, Cambridge University Press, 1998.

[13] R. Viswanathan and P. S. Varshney, "Distributed detection with multiple sensors - Part I : Fundamentals," in *Proc. IEEE*, 1997, vol. 85, pp. 54–63.

[14] S. Efromovich, *Nonparametric Curve Estimation*, Springer-Verlag, 1999.

[15] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2nd edition, 1999.

[16] A. Benveniste, A. A. Juditsky, Q. Zhang B. Delyon, and P. Y. Glorennec, "Wavelets in identification," in *Identification, Adaptation, Learning: The Science of Learning Models from Data*, S. Bittanti and G. Picci, Eds. Springer-Verlag, 1996.

[17] N. Sureshbabu and J. A. Farrel, "Wavelet-based system identification for non-linear control," *IEEE Trans. on Automatic Control*, vol. 44, no. 412–417, 1999.

[18] T. Hastie and R. Tibshirani, *Generalized Additive Models*, Chapman and Hall, 1990.

[19] O. Linton and J. P. Nielsen, "A kernel method of estimating structured nonparametric regression based on marginal integration," *Biometrika*, vol. 82, pp. 93–100, 1995.

[20] C. J. Stone, "Additive regression and other nonparametric models," *Annals of Statistics*, vol. 13, pp. 685–705, 1985.

[21] M. A. Kon and L. A. Raphael, "Convergence rates of multiscale and wavelet expansions," in *Wavelet Transforms and Time-Frequency Signal Analysis*, L. Debnath, Ed., pp. 37–65. Birkhauser, 2001.

[22] R. DeVore, "Nonlinear approximation," *Acta Numerica*, vol. 2, pp. 51–150, 1998.