NONLINEAR SYSTEM IDENTIFICATION BY WAVELET MULTIRESOLUTION ANALYSIS

M. Pawlak*
Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba, Canada

Z. Hasiewicz
Institute of Engineering Cybernetics
Technical University of Wrocław
Wrocław, Poland

Abstract

This paper deals with the problem of reconstruction of nonlinearities in a certain class of nonlinear systems of composite structure from their input-output observations when a prior information about the system is poor, thus excluding the standard parametric approach to the problem. The multiresolution idea, being the fundamental concept of modern wavelet theory, is adopted and is applied to construct nonparametric identification techniques of nonlinear characteristics. The pointwise convergence properties of the proposed identification algorithms are established.

1 INTRODUCTION

A large class of physical systems in practice 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 (mathematical model) from an analysis of its input and output data. A large class of techniques exist for identification of linear models. Much less attention, however, has been paid to nonlinear system identification, mostly because their analysis is generally harder and because the range of nonlinear model structures and behaviors is much broader than the range of linear model structures and behaviors. There is no universal approach to identification of nonlinear systems, and existing solutions depend strongly on a prior knowledge of the system structure, see [1], [2], [3], [8] for some techniques for nonlinear system identification. In general, the causal nonlinear (discrete time) system transforms the input data $\{X_t, t \leq n\}$ into the output signal Y_n at the time n. This transformation can be approximated in various ways and an early approach relies on Volterra and Wiener expansions. These representations lead, however, to very complicated identification algorithms since multidimensional Volterra/Wiener kernels must be evaluated, requiring often an extremely large input-output data set. An alternative strategy is based on the assumption that the system structure is to some extent known. This yields the concept of block-oriented models, i.e., models consisting of linear dynamic subsystems and static nonlinear elements connected together in a certain composite structure. Signals interconnecting the subsystems are not accessible for measurements making the identification problem not reducible to standard situations, i.e., identification of linear dynamic systems and recovering memoryless nonlinearities. A class of cascade/parallel models is a popular type of block-oriented structures, i.e., when linear dynamic subsystems are in a tandem/parallel connection with a static element. Examples of such models include cascade Hammerstein, Wiener and sandwich structures and their parallel counterparts, [1], [2], [3], [6], [7], [10]. The popularity of these connections stem not only from their relative simplicity (allowing us to design a constructive identification algorithms) but surprisingly from their ability to approximate closely systems which are not necessarily of this form. This is particularly the case if one allows in the cascade/parallel models 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 [1], [3] for parametric identification techniques of the cascade/parallel block-oriented models with polynomial nonlinearities. The parametric restriction is often too rigid, i.e., if one chooses a parametric family that is not appropriate form then there is a danger of reaching incorrect conclusions in the system identification. In [6], [7], [10] 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 non-

*This work was supported by NSERC

linear characteristic, and to let the training data decide which characteristic fits them best. These approaches are powerful in exploring fine details in the nonlinear characteristics. In this paper we consider the nonparametric approach to the identification of a broad class of nonlinear composite models which includes most previously defined connections. We are mostly interested in recovering a nonlinearity which is embedded in a block oriented structure containing dynamic linear subsystems and other "nuisance" nonlinearities. Our identification approach combines the concept of regression analysis and the theory of orthogonal bases originating from multiresolution and wavelet approximations of square integrable functions. This theory provides elegant techniques for representing the levels of details of the approximated function and consequently gives better results than other approximation methods, see [4], [5], [9], [11], [12] for a full account of the theory and applications of this subject.

2 NONLINEAR COMPOSITE SYSTEMS

A class of nonlinear composite systems examined in this paper is described by the following equation:

$$\begin{cases} O_n = \mu(X_n) + \xi_n \\ \xi_n = \sum_{j=-\infty}^{n-1} s_{n-j}\lambda_{n-j}(X_j)\,, \\ Y_n = O_n + \varepsilon_n \end{cases} \tag{1}$$

where (X_n, Y_n) is the input- output pair, $\mu(\cdot)$ represents the unknown system nonlinearity, $\{\xi_n\}$ is the system "noise" process characterizing the system history and $\{\varepsilon_n\}$ is the measurement noise. The system noise process $\{\xi_n\}$ has an infinite convolution representation with the weight sequence $\{s_j\}$ and the transformed input sequence $\{\lambda_j(X_{n-j})\}$. It is important to note that the nonlinear functions $\{\lambda_j(x)\}$ need not be known. The following assumptions concerning the model in (2.1) are used in the paper:

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

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

$$0 \quad < \quad \eta \;\leq\; f(x) \tag{3}$$

for all $x \in R$ and some unknown η.

Assumption 2: The parameters characterizing the system noise process $\{\varepsilon_n\}$ satisfy the following conditions:

$$E\lambda_j(X) = 0\,,\; j = 1, 2, \ldots \tag{4}$$

$$\sum_{j=1}^{\infty} s_j^2 E\lambda_j^2(X) < \infty \tag{5}$$

$$\sum_{j=1}^{\infty} |s_j||\lambda_j(x)| < \infty\,,\text{ for almost } x \in R \tag{6}$$

$$\sum_{t=1}^{\infty}\sum_{j=1}^{\infty} |s_j s_{t+j}|E\{|\lambda_j(X)\lambda_{t+j}(X)|\} < \infty \tag{7}$$

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

$$E\mu^2(X) < \infty \tag{8}$$

$$\int_{-\infty}^{\infty} (\mu(x)f(x))^2\; dx < \infty \tag{9}$$

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

$$E\varepsilon_n \quad = \quad 0\,,\; \operatorname{var}\varepsilon_n < \infty \tag{10}$$

The restriction (A1.1) is required since we use the $L_2(R)$ multiresolution decomposition of $f(x)$. The condition (A1.2) says that we consider the estimation problem in such points on R where the input density is high, i.e., where $f(x)$ is strictly bounded away from zero. The assumptions (A2.1) and (A2.2) are necessary for $\{\xi_n\}$ to be the second order covariance stationary stochastic process. This along with Assumption (A3.1) and Assumption 4 makes the output process $\{Y_n\}$ well defined, i.e., it is also a second order stochastic process. It is worth noting that $\{Y_n\}$ is not strictly stationary process. The conditions (A2.3), (A2.4), (A3.2) put some restrictions on the system dynamics and they are required for the convergence property of our identification procedure for recovering $\mu(\cdot)$. Let us note that (A2.3) is meant in the Lebesque measure sense, i.e., it holds at all points $x \in R$, except sets with zero Lebesque measure. In particular (A2.3) is true at all points where $\{\lambda_j(x), j = 1, 2, \ldots\}$ are continuous functions. Surprisingly there is a large class of block-oriented nonlinear models which fall into the description given in (2.1). This includes, e.g., the following popular connections: memoryless system, cascade Hammerstein system, parallel system, parallel-series structures, cascade Wiener system.

3 IDENTIFICATION ALGORITHMS

It is a fundamental fact for our paper to observe that

$$E\{Y_n \mid X_n = x\} = \mu(x),$$

i.e., the system nonlinearity is equal to the standard regression function. Thus by estimating the regression we can recover the non-linearity $\mu(x)$. Due to this fundamental property we can treat $\mu(x)$ as a standard regression function of Y_n on $X_n = x$. In order to construct an estimate of the regression function let us first observe that $\mu(x) = g(x)/f(x)$, where $g(x) = \mu(x)\, f(x)$ for every x where the assumption (A1.2) holds. Owing to the assumptions in (A1.1), (A3.2) we can approximate $g(x)$ and $f(x)$ by their projections on the mth multiresolution subspace of $L_2(R)$ as follows:

$$\begin{aligned} g_m(x) &= \sum_{k\in Z} a_{mk}\phi_{mk}(x) \\ f_m(x) &= \sum_{k\in Z} b_{mk}\phi_{mk}(x) \end{aligned}$$

where one can easily observe that

$$a_{mk} = \int_{-\infty}^{\infty} \mu(x)\phi_{mk}(x)f(x)\,dx = E\{Y_n\phi_{mk}(X_n)\}$$

and

$$b_{mk} = \int_{-\infty}^{\infty} \phi_{mk}(x)f(x)\,dx = E\{\phi_{mk}(X_n)\}$$

Here $\{\phi_{mk}(x), k \in Z\}$ is the orthonormal basis for the mth resolution subspace. Empirical counterparts for $g_m(x)$ and $f_m(x)$ given above can be easily constructed first by replacing the expected values in the formulas for a_{mk} and b_{mk} by their natural estimates

$$\begin{aligned} \hat{a}_{mk} &= n^{-1}\sum_{i=1}^{n} Y_i\phi_{mk}(X_i) \\ \hat{b}_{mk} &= n^{-1}\sum_{i=1}^{n} \phi_{mk}(X_i) \end{aligned}$$

and next by cutting off the number of terms to some finite value referred to in this paper as a truncation value q. All these things yield the following estimator of $\mu(x)$ utilizing $2q+1$ terms at the resolution level m:

$$\hat{\mu}_m(x) = \frac{\sum_{|k|\le q} \hat{a}_{mk}\phi_{mk}(x)}{\sum_{|k|\le q} \hat{b}_{mk}\phi_{mk}(x)}$$

It is worth noting that the truncation value should be sufficiently large to have $\hat{\mu}_m(x)$ well defined. The resolution level m plays the most important role in both asymptotic and finite sample size performance of the estimators. In fact it is required thatthe resolution level m must be chosen as a function of the sample size n, i.e., $m = m(n)$ in such a way that

$$m(n) \rightarrow \infty$$

and

$$\frac{2^{m(n)}}{n} \rightarrow \infty$$

as $n \to \infty$. Then under Assumptions 1-4 the following convergence property can be established:

$$\hat{\mu}_{m(n)}(x) \rightarrow \mu(x)$$
$$\text{as } n \rightarrow \infty \text{ in probability for almost all } x \in R.$$

This property holds for all input densities $f(\cdot)$ and all measurable nonlinearities $\mu(\cdot)$ which satisfy Assumption 1 and Assumption (A3.2). No continuity conditions for the characteristic $\mu(\cdot)$ are required. Under further smoothness conditions on $\mu(\cdot)$ and $f(\cdot)$ we demonstrate that $m(n)$ can be specified as $m(n) = \frac{1}{3}\log_2(n)$ yielding the rate $O\left(n^{-1/3}\right)$, in probability. The latter result holds for the first order multiresolution basis as, e.g. the Haar system. A faster rate of convergence can be obtained with the higher order multiresolution orthonormal systems.

References

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

[2] D. R. Brillinger, The identification of a particular nonlinear time series system, Biometrika, vol.64, pp.509-515, 1977.

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

[4] I. Daubechies, Ten Lectures on Wavelets. Philadelphia: SIAM, 1992.

[5] D. L. Donoho and I. M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika, vol.81, pp.425-455, 1994.

[6] W. Greblicki, Nonparametric identification of Wiener systems by orthogonal series, IEEE Transactions on Automatic Control, vol.10, pp.2077-2086, 1994.

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

[8] A. Juditsky, H. Hjalmaarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjoberg and Q. Zhang, Nonlinear black-box models in system identification:mathematical foundations, Automatica, vol.31, pp.1725-1750, 1995.

[9] S. Mallat, Multiresolution approximation and wavelets, Transactions of the American Mathematical Society, vol.315, pp.69-88, 1989.

[10] M. Pawlak, On the series expansion approach to the identification of Hammerstein systems, IEEE Trans. Autom. Control, vol.36, pp.763-767, 1991.

[11] M. Vetterli and J. Kovacevic, Wavelets and Subband Coding. Prentice Hall, 1995.

[12] G. Walter, Wavelets and Other Orthogonal Systems with Applications. Bocan Raton: CRC Press, 1994.