Proceedings of the XXIst National Conference on
CIRCUIT THEORY and ELECTRONIC NETWORKS
Poznań-Kiekrz, October 22-24, 1998

Włodzimierz Greblicki
Zygmunt Hasiewicz
Wrocław University of Technology
Institute of Engineering Cybernetics
Wybrzeże S. Wyspiańskiego 27, 50-370 Wrocław

Wavelet Approach to Non-linear System Identification

Keywords: system identification, wavelet approach, multiscale approximation

ABSTRACT

The paper deals with recovering non-linearities in the complex time-series systems composed of a static Non-linear element and a Linear dynamic part connected in a cascade (shortly: **NL** systems). The systems are driven by random signals and are disturbed by additive random noise. The *a priori* information about the systems is non-parametric. To recover non-linear characteristics, a class of non-parametric identification algorithms is proposed and analysed. This class is based on multiscale approximations - a basic concept of wavelet theory.

1. INTRODUCTION

In this paper, we examine wavelet approach to identification of non-linear systems, suited to the case when our *a priori* information about the system to be identified is confined to only some qualitative features of system characteristics. We propose a class of algorithms to estimate non-linearities in the dynamical **NL** (Hammerstein) systems (cascade connections of a static non-linearity and a linear dynamic element), driven by random signals and disturbed by white or correlated random noise. The identification algorithms are based on the idea of multiscale approximations, a leading concept of wavelet theory ([1], [9], or recently [7], [8]). The paper is an extension of [3] and [4].

2. MULTISCALE APPROXIMATION

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

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

and that

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

Such $\phi(x)$ is called a "scaling function" and the sequence of subspaces $\{ V_m \}$ constitutes a multiscale (multiresolution) approximation of $L^2(\mathbb{R})$ (see, e.g., the aforementioned monographs). Observe that the basis functions $\phi_{mn}(x)$ of the resolution spaces V_m are generated in an easy and automatic way from a single initial function $\phi(x)$ by only scaling (the scale factor m) and shifting (the translation factor n). Any function $F(x) \in L^2(\mathbb{R})$ can be approximated in V_m as follows

$$F(x;m) = \sum_{n=-\infty}^{\infty} \alpha_{mn} \phi_{mn}(x) \tag{2.1}$$

where the coefficients α_{mn} are given by

$$\alpha_{mn} = \int_{-\infty}^{\infty} F(x) \phi_{mn}(x) dx. \tag{2.2}$$

The approximation $F(x;m)$ in (2.1) is thus the orthogonal projection of $F(x)$ on V_m and the sequence $\{F_m(x) = F(x;m)\}_{m \in Z}$ is the multiscale approximation of $F(x)$.

In the following, we shall assume that the basis functions $\phi_{mn}(x)$ of the resolution space V_m are such that

$$|\phi_{mn}(x)| \le d(x) 2^{\alpha m}, \qquad \sup_{x \in R} |\phi_{mn}(x)| \le d_1 2^{\beta m}, \qquad \text{all } n \tag{2.3}$$

some $d(x)$ and d_1 independent of m, and that the scaling function $\phi(x)$ is supported in a compact set, $[s_1, s_2]$ say (equals zero outside $[s_1, s_2]$). Then $\phi_{mn}(x)$ are supported in $[(s_1+n)/2^m, (s_2+n)/2^m]$, and consequently the sum in (2.1), for each scale m and each point x, can be truncated to finite set of n, $n_{\min}(x;m) \le n \le n_{\max}(x;m)$, yielding

$$F(x;m) = \sum_{n=n_{\min}(x;m)}^{n_{\max}(x;m)} \alpha_{mn} \phi_{mn}(x) \tag{2.4}$$

with

$$n_{\min}(x;m) = [2^m x - s_2] + 1 \quad \text{and} \quad n_{\max}(x;m) = [2^m x - s_1] \tag{2.5}$$

where $[v]$ stands for the integer part of v. In (2.4), for every m and every x, the number of summands does not exceed $S = [s_2 - s_1] + 1$.

3. THE NL SYSTEM

The **NL** system is a tandem connection of a non-linear memoryless element, with a characteristic R, followed by a linear output dynamics. The linear dynamic part is by assumption a discrete-time time-invariant and asymptotically stable element operating in steady state, with the impulse response $\{ \lambda_p ; p = 0, 1, \ldots \}$. We assume that the system input $\{x_k ; k = \ldots, -1, 0, 1, 2, \ldots\}$ is a stationary white random process with finite variance. The probability density of x_k exists and is denoted by f. The internal signal $w_k = R(x_k)$ (output of a static non-linearity), interconnecting both parts of the system, is *not accessible* for measurements. The overall system output y_k is disturbed by additive stationary random noise $\{z_k; k = ..., -1, 0, 1, 2, ...\}$, i.e. the following equation holds

$$y_k = \sum_{p=0}^{\infty} \lambda_p R(x_{k-p}) + z_k \tag{3.1}$$

The noise is by assumption

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

(*b*) *coloured* - obtained as an output of a discrete-time time-invariant and asymptotically stable linear filter operating in steady state and driven by a zero mean stationary white noise $\{\varepsilon_k; k = ..., -1, 0, 1, 2, ...\}$ with finite variance, i.e. $E\varepsilon_k = 0$, $\operatorname{var}\varepsilon_k < \infty$.

Processes $\{x_k\}$ and $\{z_k\}$ ($\{x_k\}$ and $\{\varepsilon_k\}$) are mutually independent. The form of system non-linearity R is completely unknown and we only assume that

$$|R(x)| \le a_1 |x| + a_2$$

some a_1 and a_2, and moreover that $\int_R f^2(x)dx < \infty$.

Our aim is to recover the non-linearity R from input-output observations $\{(x_k, y_k)\}$ of the *whole* system (implicit and non-parametric identification task).

4. IDENTIFICATION ALGORITHM

The key in derivation of our identification algorithm is the recognition that for whichever white or coloured noise it holds (cf. (3.1) and [5])

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

where c ($= \lambda_d$) is a constant (for clarity of exposition, we have here assumed that Ew_k=0, i.e., that for instance the non-linearity R is an odd and the input probability density f is a symmetric (even) function). It is permitted in (4.1) that $\lambda_0 = \lambda_1 = ... = \lambda_{d-1} = 0$, i.e. some delay is allowed in the linear subsystem; we only assume that $\lambda_d \ne 0$. In turn, we can write

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

where $g(x) = E\{y_k | x_{k-d} = x\} f(x)$. Since $\int_R f^2(x)dx < \infty$ and $\int_R g^2(x)dx < \infty$, the numerator g and denominator f in the decomposition (4.2) may be approximated in the resolution space V_m by the series (see (2.1)-(2.2) in section 2)

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

where

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

For the scaling function with compact support this leads to the following natural m-scale estimate $R_N(x;m)$ of $cR(x)$:

$$R_N(x;m) = \sum_{n=n_{\min}(x;m)}^{n_{\max}(x;m)} a_{mn,N}\,\phi_{mn}(x) \Big/ \sum_{n=n_{\min}(x;m)}^{n_{\max}(x;m)} b_{mn,N}\,\phi_{mn}(x) \tag{4.3}$$

where $n_{\min}(x;m)$ and $n_{\max}(x;m)$ are as in (2.5) and where $a_{mn,N}$ and $b_{mn,N}$ (estimates of a_{mn}'s and b_{mn}'s) are computed from N (random) observations $\{(x_k, y_{k+d}); k = 1, 2, ..., N\}$

of the whole system input and output as follows:

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

Observe that the numerator and denominator of (4.3) contains (at each resolution $1/2^m$) the finite number of at most S components (cf. section 2).

5. CONVERGENCE

Assuming that the scale factor m depends on the number N of data in the sample $\{(x_k, y_{k+d})\}_{k=1}^{N}$, i.e. $m = m(N)$, and grows with growing N in such a way that

$$m(N) \rightarrow \infty, \qquad 2^{2(\alpha+\beta)m(N)} / N \rightarrow 0 \tag{5.1}$$

as $N \rightarrow \infty$, one can prove the following theorem.

Theorem 1: Let all the assumptions of section 3 be in force. Let R be odd and f even, and let the multiresolution basis functions $\{\phi_{mn}(x)\}$ satisfy the conditions (2.3) in section 2. Let the scale parameter $m = m(N)$ fulfil the conditions (5.1). Then for white as well as coloured noise, for the estimate (4.3)-(4.4) we have the convergence

$$R_N(x;m) \rightarrow cR(x) \qquad \textit{in probability}$$

as $N \rightarrow \infty$, at every point $x \in (-\infty,\infty)$ at which $f(x) > 0$, and both

$$\sum_{n=n_{\min}(x;m)}^{n_{\max}(x;m)} a_{mn}\phi_{mn}(x) \rightarrow cR(x)f(x) \quad \text{and} \quad \sum_{n=n_{\min}(x;m)}^{n_{\max}(x;m)} b_{mn}\phi_{mn}(x) \rightarrow f(x) \quad \text{as} \quad m \rightarrow \infty$$

It should be remarked that by applying the algorithm (4.3)-(4.4) we can only estimate the system non-linearity with accuracy to the scaling constant c. This is however an unavoidable consequence of the cascade complex structure of the system and assembling character of the data $\{(x_k, y_{k+d})\}$ used for identification.

Imposing some additional 'regularity' conditions on R and f, we can establish the rate of convergence of $R_N(x\,;m)$ to $cR(x)$. Assume to this end that $R(x)$ and $f(x)$ are bounded and locally Lipschitz functions, satisfying the conditions

$$\sup_x |R(x)| < \infty, \ \sup_x f(x) < \infty \tag{5.2}$$

and

$$|R(x+h) - R(x)| \le L_R |h|^p, \quad |f(x+h) - f(x)| \le L_f |h|^r \tag{5.3}$$

with exponents $0 < p,r \le 1$ and $|h| < 1$. Denote $\delta = \min(p,r)$. One can prove the theorem.

Theorem 2: Let the conditions of Theorem 1 hold. Let moreover $R(x)$ and $f(x)$ satisfy the assumptions (5.2)-(5.3) and let the scale parameter $m = m(N)$ be optimally selected as

$$m_{opt}(N) = \left[\frac{1}{2(\delta+1)}\log_2 N\right] \tag{5.4}$$

Then asymptotically, without any distinction for white and coloured noise, it holds

$$|R_N(x;m) - cR(x)| = O\left(N^{-(\delta+(1-(\alpha+\beta)))/2(\delta+1)}\right) \quad \textit{in probability} \tag{5.5}$$

where α and β are as in (2.3) of section 2.

The rate of convergence in (5.5) is determined, through the index δ, by the smoothness of more rough function from among $R(x)$ and $f(x)$ (with smaller Lipschitz exponent in the proper Lipschitz condition). This rate also depends on the exponents α and β in the bounds (2.3) characterizing multiscale basis functions $\{\phi_{mn}(x)\}$.

We emphasize the important fact that our algorithm, convergence conditions and rate of convergence remain the same for white and coloured noise.

6. EXAMPLE

For the Haar multiscale approximation, the scaling function has the form

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

(is supported in [0,1]) and the resolution spaces V_m are spanned, for each scale $m \in Z$, by

$$\phi_{H,mn}(x) = 2^{m/2}\, \phi_H(2^m(x - \frac{n}{2^m})) = 2^{m/2}\, I_{[\frac{n}{2^m}, \frac{n+1}{2^m})}(x), \qquad n \in Z \tag{6.1}$$

Obviously, for the basis functions $\phi_{H,mn}(x)$ as in (6.1) the conditions (2.3) of section 2 are satisfied for $\alpha = \beta = 1/2$. Application of the Haar basis $\{\phi_{H,mn}(x)\}$ in the algorithm (4.3)-(4.4) yields for each $x \in [n/2^m, (n+1)/2^m)$ the estimate

$$R_{H,N}(x;m) = a_{H,mn,N} \,/\, b_{H,mn,N} \tag{6.2}$$

where

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

and where # denotes the cardinality of a collection (in the above quotient, the common factor $2^m/N$ was omitted). The denominator of the estimate (6.2) counts the number of measurements x_k in the interval $[n/2^m, (n+1)/2^m)$, containing the particular reference point x at which the estimation is carried out, and the numerator selects and sums up the corresponding output measurements y_{k+d}, including the possible d-step delay in the system. This gives the sample mean of the output y in the respective interval of the length $1/2^m$ as the estimate of $cR(x)$ at x. The scale factor m (precisely: the resolution level $1/2^m$) controls the size of the neighbourhood around the point x in which averaging is made and determines thereby sensitivity of the estimate to the details in the run of $cR(x)$.

Using Theorem 1 and the well known facts concerning convergence of the Haar multiscale approximations [6, Theorem 2.1], we can conclude the following:

Corollary 1: Let the assumptions of Theorem 1 hold. If the scale factor m increases with N in such a way that $m(N) \to \infty$, $2^{2m(N)}/N \to 0$ as $N \to \infty$, then for both white and coloured noise

$$R_{H,N}(x;m) \overset{N}{\to} cR(x) \qquad \textit{in probability}$$

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

Employing in turn Theorem 2 and assuming that in particular the Lipschitz exponents in (5.3) are $p = r = 1$ (i.e. $\delta = 1$), we infer that in such a case the (optimum) in-probability rate of convergence of $R_{H,N}(x;m)$ to $cR(x)$ is of order $O(N^{-1/4})$, where the scale factor m

is chosen as (cf. (5.4))

$$m_{opt}(N) = \left[\frac{1}{4}\log_2 N\right]$$

This rate is comparable with that obtained by employing classical Hermite or trigonometric orthogonal series expansions in the case of differentiable R and f, however the identification algorithm is then much more complicated than (6.2) (see [2] for comparison).

7. CONCLUSIONS

The identification algorithm presented in the paper is based on multiscale approximations associated with scaling functions of compact support. The algorithm is non-parametric, i.e., may be applied to estimate **NL** system non-linearity when the *a priori* knowledge about the system is very small, and in particular no parametric representation of the unknown characteristic of non-linear static block is known. The algorithm is very simple and requires only elementary computations. For the use of the algorithm, only a set of weighting coefficients $a_{mn,N}$ and $b_{mn,N}$ must be calculated from experimental data and these computations can be done automatically and in short time - due to the specific 'reproducible' form of the multiscale basis functions, all of which are generated from a single initial scaling function by merely dilation and translation. Another advantage of the algorithm is that it easily copes with correlation of the noise. Both the form of the algorithm and the convergence conditions and properties (domain and rate of convergence) are the same for white and coloured noise (of arbitrary correlation structure). This is in sharp contrast to parametric methods and significantly broadens applicability of the algorithm.

REFERENCES

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

[2] W. Greblicki, Nonparametric orthogonal series identification of Hammerstein systems, *Int. J. Systems Sci.*, vol. 20, 1989, pp. 2355-2367.

[3] W. Greblicki and Z. Hasiewicz, "Identification of non-linear systems corrupted by colored noise", Proc. XIXth National Conf. Circuit Theory and Electronics Circuits, Kraków-Krynica, October 1996, vol. 1, pp. II/33-II/38.

[4] W. Greblicki and Z. Hasiewicz, "Identification of non-linear systems by orthogonal series", Proc. XXth National Conf. Circuit Theory and Electronics Circuits, Kołobrzeg, October 1997, vol. 1, pp. 85-90.

[5] W. Greblicki and M. Pawlak, Identification of discrete Hammerstein systems using kernel regression estimates, *IEEE Trans. Autom. Contr.*, vol. 31, 1986, pp. 74-77.

[6] S. Kelly, M. Kon and L. Raphael, Pointwise convergence of wavelet expansions, *Bull. Amer. Math. Soc.*, vol. 30, 1994, pp. 87-94.

[7] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, New York, 1998.

[8] R. T. Ogden, *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhäuser, Boston, 1997.

[9] G. Walter, *Wavelets and Other Orthogonal Systems with Applications*, CRC Press Inc., Boca Raton, 1994.