MMAR'98 25-29 August 1998, Międzyzdroje, Poland
Fifth International Symposium on Methods and Models in Automation and Robotics

NON-LINEAR SYSTEM IDENTIFICATION UNDER SMALL *A PRIORI* KNOWLEDGE

WŁODZIMIERZ GREBLICKI*, ZYGMUNT HASIEWICZ*

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

Abstract. The paper deals with recovering non-linearities in the single element static systems and dynamical complexes composed of a static non-linear element and a linear dynamic part connected in a cascade. The systems are driven by random signals and are disturbed by additive, white or coloured, random noise. The *a priori* information about the systems is small. To recover non-linear characteristics of examined systems, the non-parametric identification algorithms are proposed and analysed.

Key Words. System identification, non-parametric approach, kernel estimate, orthogonal series estimate, Hammerstein system.

1. INTRODUCTION

In this paper, we present the identification algorithms which may be applied when our *a priori* information about the system to be identified is small - much smaller than in parametric inference. It makes the problems closer to those encountered in applications. We propose two classes of non-parametric algorithms to estimate non-linearities in memoryless systems and cascade connections of a static non-linearity and a linear dynamic element (Hammerstein systems), driven by random signals and disturbed by white or correlated random noise. These are kernel and orthogonal series algoritms. We show that our algorithms successfully recover non-linearities for white and coloured noise as well, and that they easily cope with the possible delay in the dynamical part of the Hammerstein system. The pointwise convergence of the algorithms to the unknown characteristics is established. It is shown that neither the form nor the convergence conditions of the proposed algorithms need any modification for the noise being not white but correlated. The particular versions of the general algoritms are discussed. The paper is an extension and generalization of [6].

The Hammerstein systems, located in the class of block-oriented non-linear dynamical complexes (see, e.g., [1]-[3] or [8] for this and other block-oriented structures), are met for a long time in industrial (particularly, chemical) engineering [4] as well as signal processing, image analysis and biocybernetics. Our attention is focused on estimating non-linear part of such systems for two reasons: first, specific non-linearity is just the discriminating attribute of the concrete Hammerstein system at hand thus playing a truly prominent role in characterizing the system and, second, ambiguity of characteristics under small *a priori* knowledge - when in particular no finite parametrization of the possible non-linearity can be reasonably proposed - makes this problem much more pivotal than identification of a linear dynamic part - with an obvious form of parametrization.

2. BACKGROUND

We shall introduce the basic ideas of this paper on the simplest example of memoryless (static) system. The system, shown in Fig. 1, is non-linear and has a characteristic m. Its input is a stationary white random process $\{U_n; n = ..., -1, 0, 1, 2, ...\}$ with finite variance. The probability density of U_n exists and is denoted by f. The system is disturbed by stationary random noise $\{Z_n; n = ..., -1, 0, 1, 2, ...\}$. The noise is by assumption:

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

(*b*) *coloured* - obtained as an output of a discrete-time time-invariant and asymptotically stable linear filter, with the impulse response $\{\omega_p; p = 0, 1, \dots\}$, operating in steady state and driven by a zero mean stationary white random noise $\{\xi_n; n = \dots, -1, 0, 1, 2, \dots\}$ with finite variance, i.e.

$$Z_n = \sum_{p=0}^{\infty} \omega_p \xi_{n-p}$$

where $E\xi_n = 0$, $\operatorname{var} \xi_n < \infty$.

Processes $\{U_n\}$ and $\{Z_n\}$ ($\{U_n\}$ and $\{\xi_n\}$) are mutually independent. The non-linear characteristic m of the system is completely unknown. We only assume that

$$|m(u)| \le a_1 |u| + a_2$$

some a_1 and a_2. Then $E\{m^2(U_0)\} < \infty$.

Our problem is to estimate the true characteristic from input-output observations $(U_1, Y_1), \dots, (U_n, Y_n)$.

Observe that *a priori* information concerning m is extremely poor, and in result the class of admissible characteristics is so wide that no finite-dimensional parametric representation of the class can be admitted. Thus our identification problem is, in fact, non-parametric and needs an adequate non-parametric approach.

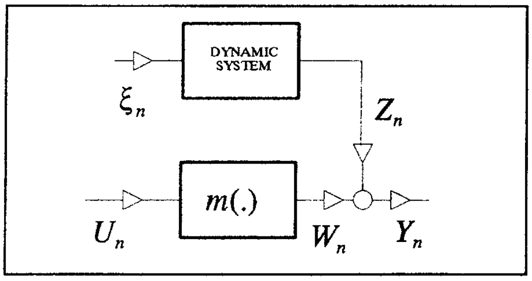

Fig. 1. The memoryless system disturbed by noise Z_n

The key in our derivation of the identification algorithms is the observation that for white as well as coloured noise

$$E\{Y_n | U_n = u\} = m(u), \tag{2.1}$$

i.e., that the unknown non-linearity m is a regression function. Thus estimating non-linear characteristic of the system is factually equivalent to estimating a conditional mean in (2.1).

2.1. Kernel Estimate

To estimate the unknown characteristic m, we apply the following regression estimate (a weighted mean of output measurements):

$$\hat{m}(u) = \sum_{i=1}^{n} Y_i K\left(\frac{u - U_i}{h(n)}\right) \Big/ \sum_{i=1}^{n} K\left(\frac{u - U_i}{h(n)}\right), \tag{2.2}$$

where K is a non-negative kernel function, and $\{h(n)\}$ a positive number sequence. The kernel satisfies the following restrictions:

$$\sup_{u \in (-\infty, \infty)} K(u) < \infty, \quad \int_{-\infty}^{\infty} K(u) du < \infty, \tag{2.3a}$$

and

$$|u| K(u) \to 0 \text{ as } |u| \to \infty. \tag{2.3b}$$

In turn, the positive number sequence is such that

$$h(n) \to 0 \text{ as } n \to \infty, \tag{2.4}$$

$$n h(n) \to \infty \text{ as } n \to \infty. \tag{2.5}$$

The following theorem can be proved for the estimate (2.2).

Theorem 1

Let the non-negative Borel kernel satisfy (2.3) and let the positive number sequence $\{h(n)\}$ fulfil (2.4)-(2.5). Then for white as well as coloured noise

$$\hat{m}(u) \to m(u) \quad \text{as} \quad n \to \infty \quad \text{in probability}$$

at every point u at which both f and m are continuous, and $f(u) > 0$.

2.2. Orthogonal Series Estimate

Observe (see (2.1)) that

$$m(u) = g(u)/f(u), \tag{2.6}$$

where $g(u) = E\{Y_n | U_n = u\} f(u)$, and where f is the probability density of U_n. Let us moreover assume that $\int_D f^2(u) du < \infty$. Then, for the system under consideration, $\int_D g^2(u) du < \infty$, too. Therefore, the nominator g and denominator f in (2.6) may be expanded in a series $\{\varphi_k; k = 0, 1, 2, \dots\}$ of functions composing a complete orthonormal family in the set D (being specified later), as follows

$$g(u) \sim \sum_{k=0}^{\infty} a_k \varphi_k(u) \quad \text{and} \quad f(u) \sim \sum_{k=0}^{\infty} b_k \varphi_k(u),$$

where

$$a_k = E\{Y_0 \varphi_k(U_0)\} \quad \text{and} \quad b_k = E\varphi_k(U_0).$$

This leads to the following natural estimate $\hat{m}(u)$ of $m(u)$:

$$\hat{m}(u) = \sum_{k=0}^{N(n)} \hat{a}_k \varphi_k(u) \Big/ \sum_{k=0}^{N(n)} \hat{b}_k \varphi_k(u), \tag{2.7a}$$

where $\hat{a}_k$ and $\hat{b}_k$ (estimates of a_k's and b_k's) are computed from the (random) observations $\{(U_i,Y_i); i = 1,2,...,n\}$ of the system input and output as the sample means:

$$\hat{a}_k = \frac{1}{n}\sum_{i=1}^{n} Y_i \varphi_k(U_i) \text{ and } \hat{b}_k = \frac{1}{n}\sum_{i=1}^{n} \varphi_k(U_i) \tag{2.7b}$$

and $N(n)$ is a sequence of integers depending on the number of data n.

Assume that

$$|\varphi_k(u)| \le c(u)(k+1)^{\alpha}, \tag{2.8a}$$

some $c(u)$ independent of k, and

$$\sup_{u \in D} |\varphi_k(u)| \le d_1(k+1)^{\beta}, \tag{2.8b}$$

some d_1 independent of k. About the number sequence $N(n)$ we assume that it satisfies the following conditions:

$$N(n) \to \infty \text{ as } n \to \infty, \tag{2.9}$$

and

$$n^{-1}\sum_{k=0}^{N(n)} k^{2\alpha}\sum_{k=0}^{N(n)} k^{2\beta} \to 0 \quad \text{as} \quad n \to \infty \tag{2.10}$$

For the estimate (2.7a)-(2.7b), one can prove the following.

Theorem 2
Let the orthogonal system satisfy (2.8a)-(2.8b) and let the number sequence $N(n)$ fulfil (2.9)-(2.10). Then for both white and coloured noise

$$\hat{m}(u) \to m(u) \quad \text{as} \quad n \to \infty \quad \text{in probability}$$

at every point $u \in D$ at which $f(u) > 0$, and

$$\sum_{k=0}^{n} a_k \varphi_k(u) \to m(u)f(u) \quad \text{as} \quad n \to \infty \tag{2.11a}$$

and

$$\sum_{k=0}^{n} b_k \varphi_k(u) \to f(u) \quad \text{as} \quad n \to \infty. \tag{2.11b}$$

Applying different kernel functions and various orthogonal series, we can obtain various particular versions of the algorithms (2.2) and (2.7), and the above theorems. Examples are given in Section 5.

Observe that correlation of the system noise does not cause any numerical complications in our algorithms, which preserve the standard form - the same as for the white noise case.

3. THE SYSTEM

The system under consideration is shown in Fig. 2. This is a block-oriented tandem connection of a non-linear memoryless element followed by a linear output dynamics. Such systems are called the Hammerstein systems. The non-linear memoryless component has a characteristic m (i.e. $W_n = m(U_n)$) and the linear dynamic part is by assumption a discrete-time time-invariant and asymptotically stable element operating in steady state, with the impulse response $\{\lambda_p; p = 0, 1, ...\}$. The system is driven by a stationary white random process $\{U_n; n = ...-1, 0, 1, 2, ...\}$ and disturbed by zero mean stationary random noise $\{Z_n; n = ...-1, 0, 1, 2, ...\}$, not necessarily white. Properties of the exciting signal $\{U_n\}$ and the noise $\{Z_n\}$ are the same as in Section 2. So is the *a priori* knowledge about the non-linear characteristic m. The goal is to identify the system non-linearity m from input-output observations $\{(U_i, Y_i)\}$ of the whole system.

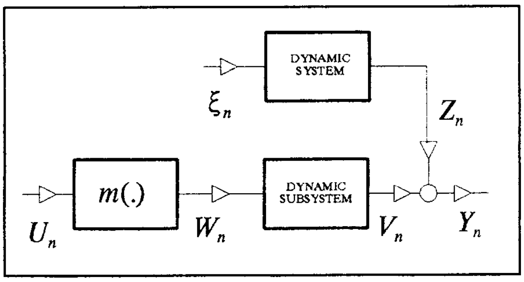

Fig. 2. The Hammerstein system with noise Z_n

The internal signal W_n, interconnecting both parts of the system, is not accessible for measurements.

Observe that inaccessibility of the internal signal W_n (generally assumed in the block-oriented modeling of dynamical systems, see [2], [3], [8]) makes the identification problem implicit (in the sense that not both the input and output signal of the element to be identified can be directly measured) and irreducible to the standard situation, where memoryless non-linearity and linear dynamics may be identified separately from their own input-output data sets.

4. IDENTIFICATION ALGORITHMS

We begin our considerations with the observation that for whichever white or coloured noise it holds

$$E\{Y_n | U_{n-d} = u\} = cm(u), \tag{4.1}$$

where c ($= \lambda_d$) is a constant (we have assumed for the ease of exposition that $EW_n=0$; this is the case for e.g., m odd and f even). Thus recovering the unknown non-linearity of the Hammerstein system (up to a scale factor c) is factually equivalent to estimating the regression in (4.1). Such a fact was originally observed in [7] (see also [5]). At present, however, 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 $\lambda_d \neq 0$, some d), and this is in contrast to the previous papers.

4.1. Kernel Estimate

To recover cm, we use the measurements $\{(U_i, Y_{i+d}); i = 1, 2, ..., n\}$ and propose the following estimate:

$$\bar{m}(u) = \sum_{i=1}^{n} Y_{i+d} K(\frac{u - U_i}{h(n)}) \Big/ \sum_{i=1}^{n} K(\frac{u - U_i}{h(n)}), \tag{4.2}$$

where K and $h(n)$ are the same as in (2.2). Observe that the estimate is an appropriate modification of (2.2) (compare also the basic relations (2.1) and (4.1)). The following counterpart of Theorem 1 holds.

Theorem 3

Let m be odd and f even. Let the non-negative Borel kernel satisfy (2.3) and the positive number sequence $\{h(n)\}$ fulfil (2.4)-(2.5). Then for white/coloured noise

$$\bar{m}(u) \rightarrow cm(u) \quad \text{as} \quad n \rightarrow \infty \quad \text{in probability}$$

at every point u at which both f and m are continuous, and $f(u) > 0$.

4.2. Orthogonal Series Estimate

Assume that $\int_D f^2(u)du < \infty$. To identify cm, we again use the measurements $\{(U_i, Y_{i+d}); i = 1, 2, ..., n\}$ and apply the same algorithm as in (2.7a):

$$\bar{m}(u) = \sum_{k=0}^{N(n)} \bar{a}_k \varphi_k(u) \Big/ \sum_{k=0}^{N(n)} \bar{b}_k \varphi_k(u), \tag{4.3a}$$

with the obvious modification that the weighting coefficients $\bar{a}_k$ and $\bar{b}_k$ are now computed as

$$\bar{a}_k = \frac{1}{n}\sum_{i=1}^{n} Y_{i+d}\varphi_k(U_i) \quad \text{and} \quad \bar{b}_k = \frac{1}{n}\sum_{i=1}^{n} \varphi_k(U_i) \tag{4.3b}$$

The counterpart of Theorem 2 is now as follows.

Theorem 4

Let m be odd and f even. Let all the assumptions of Theorem 2 hold. Then for white as well as coloured noise we have the convergence

$$\bar{m}(u) \rightarrow cm(u) \quad \text{as} \quad n \rightarrow \infty \quad \text{in probability}$$

at every point $u \in D$ at which $f(u) > 0$, and (2.11) holds - with $cm(u)$ instead of $m(u)$.

Observe that identification algorithms are now not more complicated than for memoryless system and that they are only slight modifications of the latter. Also the convergence conditions are the same as in Section 2. This is despite the fact that at present, contrary to the case of a static system, the output observations $\{Y_{i+d}\}$ are dependent (correlated) random variables, as the outputs of a linear dynamical element. It is the case for white and coloured noise alike. We also emphasize that the algorithms (4.2)-(4.3) are identical for white and correlated noise, and that the form of the algorithms does not depend on any specific correlation structure of the disturbances corrupting the system, in particular. We want to stress strongly this remarkable property of our non-parametric algorithms because as long as parametric methods are considered, correlation of the noise requires generally a substantial revision of estimation algorithms to assure convergence and results in more demanding computer procedures than for the white noise case (see e.g. [9], [11]).

It should be emphasized that by applying the algorithms (4.2)-(4.3) we can estimate the unknown characteristic only up to some multiplicative (scaling) constant c, i.e., we may discover only the 'sketchy' shape of non-linearity. This is however not a drawback of the algorithms but an unavoidable consequence of the cascade structure of the system and assembling character of the data $\{(U_i, Y_{i+d})\}$ used for identification.

5. EXAMPLES

There exist kernel functions satisfying (2.3). One can choose, e.g., a rectangular kernel equal 1 or 0 according to $|u| \leq 1$ or $|u| > 1$, respectively. Other kernels are, e.g., $\exp(-|u|)$, $1/(1+u^2)$. As far as the number sequence is concerned, we can apply $h(n) = cn^{-\alpha}$, some positive α. Restrictions (2.4)-(2.5) are met for $0 < \alpha < 1$.

Examples of classical orthogonal systems satisfying (2.8a)-(2.8b) are: *1.* the trigonometric series $\phi_k(u) = \exp(jku)$ - orthonormal in the interval $D = [-\pi,\pi]$, for which these conditions hold for $\alpha = \beta = 0$, *2.* the Legendre series $p_k(u) = \sqrt{(2k+1)/2}\,P_k(u)$, where the P_k's are Legendre polynomials ($P_0(u) = 1$, $P_1(u) = u$, $P_2(u) = (3/2)u^2 - 1/2$, $P_3(u) = (5/2)u^3 - (3/2)u$ and so on) - orthonormal in the interval $D=[-1,1]$ for which we have $\alpha = \beta = 1/2$, and *3.* the Hermite series $h_k(u) = 1/\sqrt{2^k k!\sqrt{\pi}}\exp(-u^2/2)H_k(u)$, where H_k's are the Hermite polynomials ($H_0(u) = 1$, $H_1(u) = -2u$, $H_2(u) = 4u^2 - 2$, $H_3(u) = -8u^3 + 12u$, and so on) - orthonormal in the whole real line, $D = R$, for which $\alpha = -1/4$, $\beta = -1/12$ - see [10] and [12] for the details.

Let us consider memoryless system, and apply orthogonal series estimate (2.7a)-(2.7b) with the trigonometric series. From Theorem 2 and standard results concerning convergence of trigonometric expansions [10], we get

Corollary
If (2.9) holds and $N^2(n)/n \to 0$ as $n \to \infty$, then

$$\hat{m}(u) \to m(u) \quad \text{as } n \to \infty \quad \text{in probability}$$

at every point $u \in [-\pi,\pi]$ at which $f(u) > 0$, and both m and f are differentiable.

We can similarly proceed for Hammerstein systems (using Theorem 4) and other orthogonal series.

6. FINAL REMARKS

Algorithms presented here are non-parametric, which means that they may be implemented when the *a priori* knowledge of the identified system is very small, and in particular no parametric hypothesis concerning unknown system characteristic is preassumed. The nice feature of our approach, and advantage over parametric methods, is that our algorithms need no modification when the noise becomes correlated. Moreover, our algorithms - consisting generally in estimating regression - need only elementary computations for white or coloured noise.

Two alternative classes of identification algorithms are presented: kernel and orthogonal series algorithms. Kernel algorithms are more intuitive, need however storing the whole number n of measurement data in a computer memory. The orthogonal series algorithms are in turn more formal, but have some computational advantages since they need a much smaller amount of computer memory (then, a finite number ($\sim N(n)$) of coefficients in (2.7) and (4.3) is sufficient to be memorized instead of the whole set of n measurements - with $N(n) \ll n$; see Corollary in Section 5). Notice that the algorithms converge at different sets of points of the system characteristic. In the case of kernel estimates these are the points of continuity of the characteristic. In the case of orthogonal series estimate, domain of convergence depends on the particular orthogonal system being implemented in the algorithm - see examples in Section 5 and [6].

The presented algorithms may be useful for solving system identification tasks when standard parametric methods fail, e.g., because of lack of appropriate preliminary information about the system characteristic or numerical complications introduced by parametric methods - especially in the case of correlated noise.

7. APPENDIX

7.1. Proof of Theorem 1

Let

$$\hat{g}(u) = (1/nh(n))\textstyle\sum_{i=1}^{n} Y_i K[(u-U_i)/h(n)]$$

and

$$\hat{f}(u) = (1/nh(n))\textstyle\sum_{i=1}^{n} K[(u-U_i)/h(n)].$$

Observe that $\hat{m}(u) = \hat{g}(u)/\hat{f}(u)$. We have

$$E\hat{g}(u) = (1/h(n))E\{m(U_0)K[(u-U_0)/h(n)]\}.$$

Since, for $E\{m(U_0)\} < \infty$ and a Borel measurable kernel K satisfying (2.3), it holds

$$\frac{1}{h}E\{m(U_0)K(\frac{u-U_0}{h})\} \to m(u)f(u)\int_{-\infty}^{\infty} K(v)dv \quad \text{(A.1)}$$

as $h \to 0$, at every point u at which both f and m are continuous (cf. Theorem 9.9 in [13, p.150]) thus under (2.4), we obtain

$$E\hat{g}(u) \to f(u)m(u)\int_{-\infty}^{\infty} K(v)dv \text{ as } n \to \infty. \quad \text{(A.2)}$$

In turn, $\mathrm{var}[\hat{g}(u)] = V_1(u) + V_2(u)$, where

$$V_1(u) = \frac{1}{n^2h^2(n)}\mathrm{var}[\sum_{i=1}^{n} m(U_i)K(\frac{u-U_i}{h(n)})],$$

$$V_2(u) = \frac{1}{n^2h^2(n)}\mathrm{var}[\sum_{i=1}^{n} Z_i K(\frac{u-U_i}{h(n)})],$$

respectively. Including that $\{U_n\}$ is a white noise process, $E\{m^2(U_0)\} < \infty$, and applying (A.1), we find $V_1(u) = O(1/nh(n))$ as $n \to \infty$, at every point at which both f and m are continuous. Examining $V_2(u)$ is tougher for $\{Z_n\}$ being not white (which is a general case in our consideration). We then have

$$V_2(u) = \frac{1}{n^2 h^2(n)} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{cov}[Z_i K(\frac{u-U_i}{h(n)}), Z_j K(\frac{u-U_j}{h(n)})$$

Since $\{U_n\}$ and $\{Z_n\}$ are mutually independent and EZ_n=0, applying (A.1) once again, we obtain this quantity of order $O(1/nh(n))$ at every point at which f is continuous. Thus, $\mathrm{var}[\hat{g}(u)] = O(1/nh(n))$ at every continuity point of f. In this way, we have shown that

$$\hat{g}(u) \to f(u)m(u)\int_{-\infty}^{\infty} K(v)dv \quad \text{as} \quad n \to \infty$$

in probability, at every point u at which both f and m are continuous. Since, using similar arguments, we can show that $\hat{f}(u)$ converges to $f(u)\int_{-\infty}^{\infty} K(v)dv$ (in probability) as n tends to infinity, the proof is completed. □

7.2. Proof of Theorem 2

It is obvous that $E\hat{a}_k = a_k$. Using (2.8b), one can verify that $\mathrm{var}\,\hat{a}_k = O((k+1)^{2\beta}/n), k$=0,1,2,... We then easily ascertain that

$$E\hat{g}(u) = \sum_{k=0}^{N(n)} a_k \varphi_k(u),$$

where $\hat{g}(u)$ is the nominator in (2.7a): $\hat{g}(u) = \sum_{k=0}^{N(n)} \hat{a}_k \varphi_k(u)$ .Thus, under (2.9), we have

$$\lim_{n\to\infty} E\hat{g}(u) = g(u) \; (= m(u)f(u))$$

at every point $u \in D$ at which (2.11a) holds. In turn,

$$\mathrm{var}\,\hat{g}(u) = \mathrm{var}[\sum_{k=0}^{N(n)} \hat{a}_k \varphi_k(u)] \le \sum_{k=0}^{N(n)} \mathrm{var}\,\hat{a}_k \sum_{k=0}^{N(n)} \varphi_k^2(u).$$

From (2.8a) and the variance $\mathrm{var}\,\hat{a}_k$ bound, we get immediately

$$\mathrm{var}\,\hat{g}(u) = O\left(n^{-1} \sum_{k=0}^{N(n)} k^{2\alpha} \sum_{k=0}^{N(n)} k^{2\beta}\right).$$

Hence, including (2.10), $\hat{g}(u) \to g(u)$ as $n \to \infty$ in probability, at every point $u \in D$ at which (2.11a) is in force. The same can be shown for the denominator $\hat{f}(u)$ of the estimate (2.7a). □

7.3. Proof of Theorem 3

The theorem may be proved using similar arguments as in the proof of Theorem 1, however the fact that output measurements are then dependent (correlated) random variables must be included in a reckoning.

7.4. Proof of Theorem 4

The theorem can be proved following the lines of the proof of Theorem 2, with the modification as in the proof of Theorem 3.

8. REFERENCES

1. Bendat J. S.: Nonlinear system analysis and identification, Wiley, New York, 1990
2. Billings S. A.: Identification of non-linear systems: a survey, Proc. IEE, vol. 127, 1980, pp. 272-285
3. Billings S. A., Fakhouri S. Y.: Identification of systems containing linear dynamic and static non-linear elements, Automatica, vol. 18, 1982, pp. 15-26
4. Eskinat E., Johnson S. H., Luyben W. L.: Use Hammerstein models in identification of non-linear systems, Amer. Instit. Chem. Engng, vol. 37, 1991, pp. 255-268
5. Greblicki W.: Nonparametric orthogonal series identification of Hammerstein systems, Int. J. Systems Sci., vol. 20, 1989, pp. 2355-2367
6. Greblicki W., Hasiewicz Z.: System identification in the presence of correlated noise; non-parametric approach, Proc. 3rd Int. Symp. on Methods and Models in Automation and Robotics - MMAR'96, vol. 2, Międzyzdroje 1996, pp. 619-624
7. Greblicki W., Pawlak M.: Identification of discrete Hammerstein systems using kernel regression estimates, IEEE Trans. Autom. Control, vol. 31, 1986, pp. 74-77
8. Haber R., Unbehauen H.: Structure identification of non-linear dynamic systems: a survey on input/output approaches, Automatica, vol. 26, 1990, pp. 651-677
9. Norton J. P.: An introduction to identification, Academic Press, London, 1986
10. Sansone G.: Orthogonal functions, Interscience Publishers Inc., 1959
11. Söderström T., Stoica P.: System identification, Prentice-Hall, New York, 1989
12. Szegö G.: Orthogonal polynomials, Amer. Math. Soc. Coll. Publ., 1959
13. Wheeden R. L., Zygmund A.: Measure and integral, Dekker, New York, 1977