MMAR'96 10-13 September 1996, Międzyzdroje, Poland
Third International Symposium on Methods and Models in Automation and Robotics

SYSTEM IDENTIFICATION IN THE PRESENCE OF CORRELATED NOISE; NON-PARAMETRIC APPROACH

WŁODZIMIERZ GREBLICKI*, ZYGMUNT HASIEWICZ*

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

Abstract. The paper deals with the non-parametric identification of dynamical systems based on the idea of orthogonal expansions. A general orthogonal series algorithm for the identification of Hammerstein systems is proposed. Conditions for its in-probability convergence to the true non-linear characteristic are given. The additive noise is correlated. Then various versions of the algorithm, corresponding to classical orthogonal series (trigonometric, Legendre and Hermite polynomials) as well as to the compactly supported orthogonal wavelets, are presented and discussed.

Key Words. Hammerstein system, correlated noise, system identification, non-parametric estimation.

1. INTRODUCTION

It is well-known that the identification of systems in the presence of correlated noise is much more difficult than identification in the white noise environment, both in theoretical and practical aspects, and generally requires special steps to obtain efficient procedures. For instance, in the case of linear dynamical systems the popular least squares method fails for correlated noise and, therefore, essential modifications are necessary. They depend on the particular correlation structure of the noise and lead to variety of methods of local applicability. The problem is by far more complex for non-linear systems.

In the paper, we propose the class of algorithms for non-linear system identification, which
(*i*) easily cope with correlation of the noise,
(*ii*) are effective for a wide class of noise models,
(*iii*) can be implemented under poor a priori knowledge of the system,
(*iv*) are computationally simple.
The algorithms are proposed to identify Hammerstein systems. These systems are non-linear dynamical complexes, which consist of a static non-linearity followed by a linear dynamics. Such tandem connections are often met in control theory as well as in communication theory, image processing, or biocybernetics, e.g. [5]. It is assumed that the internal signal, interconnecting two parts of the system is not accessible for measurements and only input-output signals of the overall system are measured. It is also assumed that disturbances affecting the system are produced by a linear dynamics from white noise, i.e., in general, are non white. The attention is focused on estimating non-linear part of the system under poor *a priori* knowledge (which is rather standard in real applications). Consequently, the problem is non-parametric, i.e., no finite parametrization is pre-selected for the static characteristc of the unknown non-linearity. The identification of the linear dynamic part of the system is simpler is omitted here.

2. THE SYSTEM

The Hammerstein system is presented in Fig. 1. We shall assume that the system is driven by a sequence $\{U_n; n = \ldots,-1,0,1,\ldots\}$ of independent identically distributed (i.i.d.) random variables with zero mean and finite variance. The nonlinear static characteristic m is by assumption a Borel measurable function. Consequently, the interconnecting internal signal (not accessible for measurements)

$W_n = m(U_n)$

is a random variable and $\{W_n ; n = ..., -1, 0, 1, ...\}$ is a stationary random noise. We make the following

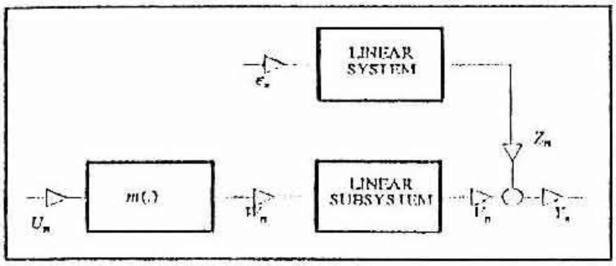

Fig. 1. The Hammerstein system

Assumption 1.
In the system,

$|m(u)| \le c_1 |u| + c_2$

some c_1 and c_2.

Owing to that, $EW_n^2 < \infty$. The linear subsystem, in turn, is described by the following stata equation

$$\left.\begin{aligned} X_{n+1} &= AX_n + bW_n \\ V_n &= c^T X_n, \end{aligned}\right\} \qquad (1)$$

where X_n is the state vector at time n. Assuming that the linear subsystem is asymptotically stable and recalling $EW_n^2 < \infty$, we find V_n a random variable. The output of the linear subsystem is disturbed by an additive stationary random noise Z_n, i.e. we observe

$Y_n = V_n + Z_n$

however the noise Z_n is not necessarily white. Contrary, we assume that it is itself the output of an asymptotically stable linear system (filter) excited by a stationary white noise, i.e., that it is formed as follows

$$\left.\begin{aligned} \xi_{n+1} &= P\xi_n + q\epsilon_n \\ Z_n &= s^T \xi_n, \end{aligned}\right\} \qquad (2)$$

where ξ_n is a filter state vector, and $\{\epsilon_n ; n = ..., -1, 0, 1, ...\}$ a sequence of i.i.d. random variables with zero mean and finite variance. Moreover, by assumption, random processes $\{U_n ; n = ..., -1, 0, 1, ...\}$ and $\{\epsilon_n ; n = ..., -1, 0, 1, ...\}$ are mutually independent.

The problem is to recover the characteristic m of a static non-linearity from input-output observations $(U_0, Y_0), (U_1, Y_1), ..., (U_n, Y_n)$ of the whole system.

3. IDENTIFICATION ALGORITHM

Observe that

$E\{Y_{n+1} | U_n = u\} = a_1 + am(u),$

where $a = c^T b$ and $a_1 = c^T AEX_n$. For the sake of simplicity, we assume that

the distribution of U_n is symmetric,
m is an odd function.

Owing to that, $a_1 = 0$ and

$$E\{Y_{n+1} | U_n = u\} = am(u). \qquad (3)$$

Thus, recovering the nonlinear characteristic m is actually equivalent to regression estimation. Let us now observe that

$$am(u) = \frac{g(u)}{f(u)},$$

where

$$g(u) = E\{Y_{n+1} | U_n = u\} f(u),$$

and where f is the probability density of U_n (assumed to exist). The nominator g and denominator f of the above expression will be expanded in a series $\{\varphi_k ; k = 0, 1, 2, ...\}$ of functions orthonormal in a set D (being specified later). We assume that all φ_k's vanish outside D. The orthogonality means that

$$\int_D \varphi_k(u)\varphi_m(u)du = \begin{cases} 1 & \text{for } k = m \\ 0 & \text{otherwise} \end{cases}$$

On the the density f, we impose the following

Assumption 2.
In the system,

$\int_D f^2(u)du < \infty.$

Observing that, moreover, $\int_D g^2(u)du < \infty$, we can write

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

where

$$a_k = E\{Y_1\,\varphi_k(U_0)\}, \quad \text{and} \quad b_k = E\,\varphi_k(U_0),$$

respectively. This leads to the following natural estimate $\hat{\mu}(u)$ of $am(u)$:

$$\hat{\mu}(u) = \frac{\sum_{k=0}^{N(n)} \hat{a}_k\,\varphi_k(u)}{\sum_{k=0}^{N(n)} \hat{b}_k\,\varphi_k(u)}, \tag{4}$$

where $\hat{a}_k$ and $\hat{b}_k$ (estimates of a_k and b_k) are computed from the random observations $\{(U_i, Y_i);\ i = 0,1,\ldots,n\}$ of the system input and output as follows:

$$\hat{a}_k = \frac{1}{n}\sum_{i=0}^{n-1} Y_{i+1}\,\varphi_k(U_i) \quad \text{and} \quad \hat{b}_k = \frac{1}{n}\sum_{i=0}^{n-1} \varphi_k(U_i)$$

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

Respectively, for compactly supported orthogonal wavelets $\{\psi_{kl};\ k,l = \ldots,-1,0,1,\ldots\}$ we can write

$$g(u) \sim \sum_{|k|=0}^{\infty}\sum_{|l|=0}^{\infty} a_{kl}\,\psi_{kl}(u) \quad \text{and}$$

$$f(u) \sim \sum_{|k|=0}^{\infty}\sum_{|l|=0}^{\infty} b_{kl}\,\psi_{kl}(u),$$

where

$$a_{kl} = E\,\{Y_1\,\psi_{kl}(U_0)\} \quad \text{and} \quad b_{kl} = E\,\psi_{kl}(U_0)$$

and as a wavelet estimate $\hat{\mu}(u)$ of $am(u) = g(u)/f(u)$, by analogy to (4), we take

$$\hat{\mu}(u) = \frac{\sum_{|k|=0}^{N(n)}\sum_{|l|=0}^{\infty} \hat{a}_{kl}\,\psi_{kl}(u)}{\sum_{|k|=0}^{N(n)}\sum_{|l|=0}^{\infty} \hat{b}_{kl}\,\psi_{kl}(u)}, \tag{5}$$

where

$$\hat{a}_{kl} = \frac{1}{n}\sum_{i=0}^{n-1} Y_{i+1}\,\psi_{kl}(U_i) \text{ and } \hat{b}_{kl} = \frac{1}{n}\sum_{i=0}^{n-1} \psi_{kl}(U_i).$$

Actually, the basis functions are of the form

$$\psi_{kl}(u) = 2^{k/2}\,\psi(2^k u - l)$$

i.e. are derived from a single initial function $\psi(u)$ ("mother" wavelet) and indexed by two labels (k - scale, l - translation), where $\psi(u)$ has compact support (i.e. equals zero outside some compact set, $[s_1, s_2]$ say). They constitute a series of functions orthonormal in the set $D = R$ (an orthonormal basis for $L^2(R)$). The orthogonality property means now that

$$\int_R \psi_{kl}(u)\,\psi_{mn}(u)\,du = \begin{cases} 1 & \text{for } k = m,\ l = n \\ 0 & \text{otherwise.} \end{cases}$$

Notice that instead of (4) we have now the double series in the nominator and denominator of (5), with infinite inner sum in general. However, for compactly supported wavelets (in $[s_1, s_2]$) this inner sum, for each k and each u, is truncated to give

$$\sum_{|l|=0}^{\infty} = \sum_{l=L_{\min}}^{L_{\max}}$$

where

$$L_{\min} = [2^k u - s_2] + 1 \quad \text{and} \quad L_{\max} = [2^k u - s_1]$$

with $[v]$ the integer part of v, and the number of contributing terms is

$$\sum_{l=L_{\min}}^{L_{\max}} 1 \le [s_2 - s_1] + 1$$

for every k and u (the 'zoom in' property of wavelets). This simplifies (5) to the form

$$\hat{\mu}(u) = \frac{\sum_{|k|=0}^{N(n)}\sum_{l=L_{\min}}^{L_{\max}} \hat{a}_{kl}\,\psi_{kl}(u)}{\sum_{|k|=0}^{N(n)}\sum_{l=L_{\min}}^{L_{\max}} \hat{b}_{kl}\,\psi_{kl}(u)}. \tag{5a}$$

4. CONVERGENCE CONDITIONS

Throughout the paper, in particular, in all next Theorems and Corollaries, Assumptions 1-2 hold. For the sake of shortness, this fact will not be repeated each time.

For orthogonal series , we assume that

$$|\varphi_k(u)| \le c(k+1)^{\alpha}, \tag{6}$$

some c independent of k,

$$\sup_{u \in D} |\varphi_k(u)| \le d_1 (k+1)^{\beta}, \tag{7}$$

some d_1 independent of k. As far as wavelets are concerned,

$$|\psi_{kl}(u)| \le c\,2^{\alpha k}, \quad \text{all } l \tag{8}$$

some c independent of k,

$$\sup_{u \in R} |\psi_{kl}(u)| \le d_1 2^{\beta k}, \qquad \text{all } l \tag{9}$$

some d_1 independent of k (observe that (8)-(9) correspond to the conditions (6)-(7) with 2^k in the role of $k+1$ and ψ_{kl} in place of φ_k). Then, see [4],

$$E\hat{a}_k = a_k \quad \text{and} \quad \mathrm{var}\,\hat{a}_k = O((k+1)^{2\beta}/n) \tag{10}$$

$k=0,1,2,\ldots$ for standard orthogonal systems, and

$$E\hat{a}_{kl} = a_{kl} \quad \text{and} \quad \mathrm{var}\,\hat{a}_{kl} = O(2^{2\beta k}/n)$$

for wavelets. Notice that the latter is the same as (10), with $k+1$ replaced by 2^k. Using these facts, the following general convergence theorem can be proved for the estimates (4) and (5)-(5a) ([4]).

Theorem
(*A*) *Standard orthogonal systems*:
If

$$N(n) \to \infty \text{ as } n \to \infty, \tag{11}$$

and

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

then

$$\hat{\mu}(u) \to am(u) \text{ as } n \to \infty \text{ in probability}$$

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

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

and

$$\sum_{k=0}^{n} a_k \varphi_k(u) \to am(u)f(u) \quad \text{as} \quad n \to \infty.$$

(*B*) *Wavelets*:
If (11) holds and

$$n^{-1} 2^{2(\alpha+\beta)N(n)} \to 0 \quad \text{as} \quad n \to \infty$$

then

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

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

$$\sum_{|k|=0}^{n} \sum_{l=L_{min}}^{L_{max}} b_{kl} \psi_{kl}(u) \to f(u) \quad \text{as} \quad n \to \infty$$

and

$$\sum_{|k|=0}^{n} \sum_{l=L_{min}}^{L_{max}} a_{kl} \psi_{kl}(u) \to am(u)f(u) \quad \text{as} \quad n \to \infty$$

(here the inner series is reduced to a finite set of l as in (5a); $L_{min} \le l \le L_{max}$).

5. EXAMPLES

Now we shall present particular versions of the algorithm (4)-(5)-(5a) and the above theorem, obtained by applying the trigonometric, Legendre, and Hermite series and, respectively, the Daubechies families of wavelets. These orthogonal systems satisfy the conditions (6) - (7) and (8) - (9) with various α and β.

5.1. The Trigonometric Series Algorithm

Using the trigonometric complex series

$$\phi_k(u) = e^{jku},$$

$k = \ldots,-1,0,1,\ldots$, orthogonal in the interval $D = [-\pi,\pi]$, we obtain the algorithm:

$$\hat{\mu}(u) = \frac{\sum_{|k|=0}^{N(n)} \tilde{a}_k e^{jku}}{\sum_{|k|=0}^{N(n)} \tilde{b}_k e^{jku}},$$

where

$$\tilde{a}_k = \frac{1}{2\pi n} \sum_{i=0}^{n-1} Y_{i+1} e^{-jkU_i}, \quad \text{and}$$

$$\tilde{b}_k = \frac{1}{2\pi n} \sum_{i=0}^{n-1} e^{-jkU_i}.$$

From the general theorem (Part *A*), the fact that (6) - (7) actually hold for $\alpha = \beta = 0$ and standard results concerning convergence of trigonometric expansions [7], we get

Corollary 1. If (11) holds and

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

then

$$\tilde{\mu}(u) \to am(u) \quad \text{as} \quad n \to \infty \quad \text{in probability}$$

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

5.2. The Legendre Series Algorithm

Denote

$$P_k(u) = \frac{1}{2^k k!} \frac{d^k}{du^k}(u^2 - 1)^k,$$

$k = 0,1,2,\ldots$. The P_k's are Legendre polynomials and we have $P_0(u) = 1$, $P_1(u) = u$, $P_2(u) = (3/2)u^2 - 1/2$,

$P_3(u) = (5/2)u^3 - (3/2)u$ and so on. It is well known that $\{p_k; k = 0,1,2,...\}$, where

$$p_k(u) = \sqrt{\frac{2k+1}{2}} P_k(u),$$

$k = 0,1,2,...$, is a system orthonormal in the interval $D = [-1,1]$, ([7, Ch. 4]). The respective Legendre series algorithm has the following form:

$$\bar{\mu}(u) = \frac{\sum_{k=0}^{N(n)} \bar{a}_k p_k(u)}{\sum_{k=0}^{N(n)} \bar{b}_k p_k(u)},$$

where

$$\bar{a}_k = \frac{1}{n} \sum_{i=0}^{n-1} Y_{i+1} p_k(U_i), \text{ and } \bar{b}_k = \frac{1}{n} \sum_{i=0}^{n-1} p_k(U_i).$$

Using the fact that (6)-(7) hold for $\alpha = \beta = 1/2$, [7], applying results on pointwise convergence of Legendre expansions given in [7], and taking account of the Theorem (Part A), we obtain

Corollary 2. Let f' and m' be of bounded variation. If (11) holds and

$$N^4(n)/n \to 0 \quad \text{as} \quad n \to \infty,$$

then

$$\bar{\mu}(u) \to am(u) \quad \text{as} \quad n \to \infty \quad \text{in probability}$$

at every point $u \in [-1,1]$ at which $f(u) > 0$.

5.3. The Hermite Series Algorithm

Hermite polynomials are defined in the following way

$$H_k(u) = e^{u^2} \frac{d^k}{du^k} e^{-u^2},$$

$k = 0,1,2,...$ [7, Ch. 4]. For example, $H_0(u) = 1$, $H_1(u) = -2u$, $H_2(u) = 4u^2 - 2$, $H_3(u) = -8u^3 + 12u$, and so on. It is known that the series

$$h_k(u) = \frac{1}{\sqrt{2^k k! \sqrt{\pi}}} e^{-u^2} H_k(u),$$

$k = 0,1,2,...$ is orthonormal in the whole real line, $D = R$. Applying the Hermite series, we get the following algorithm:

$$\check{\mu}(u) = \frac{\sum_{k=0}^{N(n)} \check{a}_k h_k(u)}{\sum_{k=0}^{N(n)} \check{b}_k h_k(u)},$$

where

$$\check{a}_k = \frac{1}{n} \sum_{i=0}^{n-1} Y_{i+1} h_k(U_i), \text{ and } \check{b}_k = \frac{1}{n} \sum_{i=0}^{n-1} h_k(U_i).$$

For the series, $\alpha = -1/4$ and $\beta = -1/12$, see [8]. Invoking pointwise equconvergence result in [8] and applying the Theorem (Part A), we find

Corollary 3. If (11) holds and

$$N^{\frac{4}{3}}(n)/n \to 0 \quad \text{as} \quad n \to \infty,$$

then

$$\check{\mu}(u) \to am(u) \quad \text{as} \quad n \to \infty \quad \text{in probability}$$

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

5.4. The Daubechies Wavelet Algorithm

The Daubechies wavelets are families of functions $\mathcal{D}^s = \{D^s_{kl}; \ |k|, |l| = 0,1,2,...\}$, $s = 1,2,3,...$, with the basic (of mother wavelet) supports in $[1-s, s]$ and the support widths $2s-1$ growing with the smoothness of members (with the index s). There, as usual

$$D^s_{kl}(u) = 2^{k/2} D^s(2^k u - l)$$

however for $s > 1$ the mother wavelet $D^s(u)$ is not given by an explicit formula but is computed from an iterative procedure ([1], [9]). For $s = 1$ this gives the Haar system $\{H_{kl}; \ |k|, |l| = 0,1,2,...\}$, where the mother wavelet is supported in $[0,1]$ and given by

$$H(u) = \begin{cases} 1 & \text{for } u \in [0, 1/2) \\ -1 & \text{for } u \in [1/2, 1) \\ 0 & \text{otherwise}. \end{cases}$$

It is well known that $\{D^s_{kl}\}$ are orthonormal in the whole real line (form complete orthonormal bases of $L^2(R)$) i.e. $D = R$ ([1], [9]). Using the Daubechies families of wavelets $\mathcal{D}^s$ (for arbitrary s), we obtain

$$\hat{\mu}^s_D(u) = \frac{\sum_{|k|=0}^{N(n)} \sum_{l=L_{min}}^{L_{max}} \hat{a}^s_{kl,D} D^s_{kl}(u)}{\sum_{|k|=0}^{N(n)} \sum_{l=L_{min}}^{L_{max}} \hat{b}^s_{kl,D} D^s_{kl}(u)}$$

where for given u and given s,

$$L_{min} = [2^k u - s] + 1, \qquad L_{max} = [2^k u + s - 1]$$

and

$$\hat{a}^s_{kl,D} = \frac{1}{n}\sum_{i=0}^{n-1} Y_{i+1} D^s_{kl}(U_i)$$

$$= \frac{2^{k/2}}{n} \sum_{\{i:\Delta_i \in [1-s,s)\}} Y_{i+1} D^s(\Delta_i)$$

and

$$\hat{b}^s_{kl,D} = \frac{1}{n}\sum_{i=0}^{n-1} D^s_{kl}(U_i)$$

$$= \frac{2^{k/2}}{n} \sum_{\{i:\Delta_i \in [1-s,s)\}} D^s(\Delta_i)$$

with $\Delta_i = 2^k U_i - l$ and $D^s(\Delta_i)$ being calculated numerically (for $s > 1$). For Haar wavelets ($s = 1$) we have respectively

$$L_{min} = L_{max} = [2^k u]$$

(since two Haar wavelets of the same scale k do not overlap) and

$$\hat{a}_{kl,H} = \frac{2^{k/2}}{n}\Big[\sum_{\{i:\Delta_i \in [0,1/2)\}} Y_{i+1} - \sum_{\{i:\Delta_i \in [1/2,1)\}} Y_{i+1} \Big]$$

and

$$\hat{b}_{kl,H} = \frac{2^{k/2}}{n}[\#\{\Delta_i \in [0,1/2)\} - \#\{\Delta_i \in [1/2,1)\}]$$

where $\#$ denotes the cardinality of a collection. Taking account of the general theorem (Part *B*), including the fact that the conditions (8) and (9) are now satisfied for $\alpha = \beta = 1/2$ [9], and exploiting the known results concerning convergence of the Daubechies expansions [6], we obtain that

Corollary 4. If (11) holds and

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

then

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

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

6. CONCLUSIONS

1. Concerning restrictions imposed on the nonlinearity, we have only assumed that Assumption 1 holds. The class of all possible characteristics is very wide and includes, e.g., discontinuous functions. Algorithms derived from various series converge, however, at various sets of points.
2. Convergence rate has not been studied here. It can be, however, shown that the asymptotic rate is not worsen by the fact that the noise is not white but correlated.
3. Our estimates need only elementary calculations based on explicitly given formulas of the orthogonal basis functions, except for the Daubechies wavelets for $s > 1$ where the basis functions (mother wavelets) must be computed numerically from an iterative procedure.
4. The approach successfully copes with correlation of the noise. The form of the algorithms given in the paper is insensitive to the correlation structure of the noise. This property becomes evident when invoking the results given previously in [2] and [3], where only white noise was admitted.
5. The algorithms can be used for solving system identification tasks under poor *a priori* knowledge of the system, when no parametrization of the characteristic is known.

7. REFERENCES

1. Daubechies I.: Ten Lectures on Wavelets, SIAM Edition, Philadelphia, 1992
2. Greblicki W.: Non-parametric orthogonal series identification of Hammerstein systems, International Journal of Systems Science, vol. 20, 1989, pp. 2355-2367
3. Greblicki W., Pawlak M.: Non-parametric identification of Hammerstein systems, IEEE Trans. Inform. Theory, vol. 35, 1989, pp. 409-419
4. Greblicki W., Hasiewicz Z.: Non-linear system identification in the presence of correlated noise, Reports of the Inst. Engng. Cybern., Tech. Univ. of Wroclaw, No. 37/95, 1995
5. Hall C. F., Hall E. L.: A nonlinear model for the spatial characteristics of the human visual system, IEEE Trans. System Man Cybern., vol. 7, 1977, pp. 161-170
6. Kelly S., Kon M., Raphael L.: Pointwise convergence of wavelet expansions, Bulletin of the American Mathematical Society, vol. 30, 1994, pp. 87-94
7. Sansone G.: Orthogonal Functions, Interscience Publishers Inc., 1959
8. Szegö G.: Orthogonal Polynomials, Amer. Math. Soc. Coll. Publ., 1959
9. Walter G.: Wavelets and Other Orthogonal Systems with Applications, CRC Press Inc., Boca Raton, 1994