Generalized Regression Neural Networks in Time-Varying Environment

Leszek Rutkowski, *Senior Member, IEEE*

***Abstract*—The current state of knowledge regarding nonstationary processes is significantly poorer then in the case of stationary signals. In many applications, signals are treated as stationary only because in this way it is easier to analyze them; in fact, they are nonstationary. Nonstationary processes are undoubtedly more difficult to analyze and their diversity makes application of universal tools impossible. In this paper we propose a new class of generalized regression neural networks working in nonstationary environment. The generalized regession neural networks (GRNN) studied in this paper are able to follow changes of the best model, i.e., time-varying regression functions. The novelty is summarized as follows: 1) We present adaptive GRNN tracking time-varying regression functions. 2) We prove convergence of the GRNN based on general learning theorems presented in Section IV. 3) We design in detail special GRNN based on the Parzen and orthogonal series kernels. In each case we precise conditions ensuring convergence of the GRNN to the best models described by regression function. 4) We investigate speed of convergence of the GRNN and compare performance of specific structures based on the Parzen kernel and orthogonal series kernel. 5) We study various nonstationarities (multiplicative, additive, "scale change," "movable argument") and design in each case the GRNN based on the Parzen kernel and orthogonal series kernel.**

***Index Terms*—Convergence properties, generalized regression neural networks (GRNN), orthogonal series kernel, Parzen kernel, time-varying environment.**

I. Introduction

The generalized regression neural network (GRNN) was introduced by Nadaraya [19] and Watson [41] and rediscovered by Specht [35] to perform general (linear or nonlinear) regressions. The GRNN was applied to solve a variety of problems [22] like prediction, control, plant process modeling or general mapping problems. Other stochastically based neural networks, the so-called probabilistic neural networks, are used for classification [4], [18], [22], [27], [29], [33], [34], [36]. The concept of the GRNN is based on nonparametric estimation commonly used in statistics [10], [12], [16], [21], [23]–[26], [38]. An interesting study presenting a bridge between nonparametric estimation and artificial neural networks is given in [43]. The essence of nonparametric estimation is nonlimiting to an assumed—usually in an arbitrary way—parametric class of models. Such approach was applied by several authors (see, e.g., [16], [23]–[26]) who created nonparametric algorithms based on the Parzen method and orthogonal series. More precisely, in stationary regression analysis we consider a random vector (X, Y), where X is R^p—valued and Y is R—valued. The problem is to find a (measurable) function ϕ: $R^p \to R$ such that the L_2 risk

$$E[\phi(X) - Y]^2 \tag{1}$$

attains minimum. The solution is the regression function

$$\phi^*(x) = E[Y \mid X = x]. \tag{2}$$

Nonparametric procedures and GRNN approach the best solution (2) as the sample size grows large.

The nonparametric methods discussed above could be applied only in stationary situations—where probability distributions do not change with time. However, in many cases the assumption concerning stationarity may be false, because usually properties of various processes depend on time. It is possible to enumerate the following examples: i) the production process in an oil refinery, where nonstationarity is a result of a change of catalyst properties; ii) the process of carbon dioxide conversion, where nonstationarity is also a result of catalyst aging; iii) the vibrations of the atmosphere around a starting space rocket are a nonstationary process, because the force that stimulates the rocket to start is a function of parameters that change quickly, such as the speed of the rocket and the distance from Earth's surface; iv) the converter-oxygen process of steelmaking, when thermal conditions in the converter may change between melts. In literature, there are three best-known parametric methods for modeling nonstationary systems (see, e.g., [17], [32]): a) Movable models method; for modeling of nonstationary systems, the classic method of minimum squares is used and the data set is constantly updated through the elimination of the oldest data and simultaneous feeding of the newest data. The period of time during which the data set is collected is called the observation horizon. b) Method based on the criterion of the minimum weighted sum squares; the minimum squares method is also used, but the elimination of the oldest data is carried out through assigning decreasing weights in the criterion of the minimum weighed sum squares. c) Method of dynamic stochastic approximation; characteristics of the nonstationary plant are approximated by a linear model having time-varying coefficients which are estimated by means of the dynamic stochastic approximation method [11].

An important problem in method a) is the optimization of the observation horizon and in method b), the selection of weight coefficients. Unfortunately, the solution of such problems depends on the possession of a relative number of *a priori* information, such as, e.g., the character of nonstationarity, the variance of disturbances and the form of the input signal. Similarly,

Manuscript received April 17, 2003; revised December 16, 2003.

The author is with the Department of Computer Engineering, Technical University of Czestochowa, Czestochowa. and the Department of Artificial Intelligence, WSHE University, Lodz, Poland (e-mail: lrutko@kik.pcz.czest.pl).

Digital Object Identifier 10.1109/TNN.2004.826127

a disadvantage of method c) is a necessity to know the way in which the linear model coefficients change.

Methods a), b), and c) that were previously discussed do not allow to track the changing characteristics of the best models described by time-varying regression functions. Such a property is possessed by the GRNN constructed in this paper. In the nonstationary regression, we consider a sequence of random variables $\{X_n, Y_n\}, n = 1, 2, \ldots$, having time-varying cumulative probability density functions $f_n(x, y)$. The problem is to find a measurable function ϕ_n: $R^p \to R$ such that the L_2 risk

$$E[\phi_n(X) - Y]^2 \tag{3}$$

attains minimum. The solution is the regression function

$$\phi_n^*(x) = E[Y_n \mid X_n = x], \quad n = 1, 2, \ldots \tag{4}$$

changing with time.

In this paper, we propose a new class of generalized regression neural networks working in a nonstationary environment. The general regression neural networks studied in this paper are able to follow changes of the best model i.e., time-varying regression functions given by (4). The novelty is summarized as follows:

1) We present the adaptive GRNN tracking time-varying regression functions.
2) We prove convergence of the GRNN based on general learning theorems presented in Section IV.
3) We design in detail special GRNN based on the Parzen and the orthogonal series kernels. In each case, we precise conditions ensuring convergence of the GRNN to the best models given by (4).
4) We investigate speed of the convergence of the GRNN and compare performance of specific structures based on the Parzen kernel and the orthogonal series kernel.
5) We study various nonstationarities (multiplicative, additive, "scale change," "movable argument") and design in each case the GRNN based on the Parzen kernel and orthogonal series kernel.

As aforementioned, the current state of knowledge regarding nonstationary processes is significantly poorer than in the case of stationary signals. In many applications signals are treated as stationary because only in this way it is easier to analyze them; in fact, they are nonstationary. Nonstationary processes are undoubtedly more difficult to analyze and their diversity makes application of universal tools impossible. In this context our paper seems to be a significant contribution to the development of new techniques in the area of nonstationary signals. More specifically, the paper advances the current state of knowledge in the following fields: a) stochastic—based neural networks; b) nonparametric regression estimation; c) modeling of time-varying plants. It should be emphasized that the methodology proposed in this paper allows to solve problems that earlier could have been treated as " impossible to solve." For illustration of the capability of our GRNN we may consider an application to modeling of nonstationary plants described by

$$Y_n = \phi_n^*(X_n) + Z_n$$

where ϕ_n^* is given by (4) and noise Z_n is a sequence of i.i.d. random variables. Suppose that

i)
$$\phi_n^*(x) = \alpha_n \phi(x)$$

ii)
$$\phi_n^*(x) = \phi(x) + \beta_n$$

iii)
$$\phi_n^*(x) = \phi(x\omega_n)$$

iv)
$$\phi_n^*(x) = \phi(x - \lambda_n)$$

where $\alpha_n, \beta_n, \omega_n$, and λ_n are sequences of real numbers. In the paper, based on the learning sequence $(X_1, Y_1), (X_2, Y_2), \ldots$, we design the GRNN that allow to track $\phi_n^*(x)$ in cases i)–iv) despite the fact that we do not known the function ϕ and sequences $\alpha_n, \beta_n, \omega_n$ or λ_n. It will be shown that it is possible to design the GRNN tracking, e.g., the following nonstationarites in the above models

$$\alpha_n = \beta_n = \omega_n = \lambda_n = \begin{cases} c_1 n^t \\ c_2 \cos n^t, \\ c_3 n^{t_1} \cos n^{t_2}. \end{cases}$$

This paper is organized into fourteen sections. In Section II we introduce kernels functions on which the construction of GRNN will be based. In Section III we review GRNN working in stationary environment. In the same section we extend the formula of the classical GRNN suggested by Specht [35] to the recursive GRNN with a gain $(1/n)$. In the next sections we will replace the sequence $(1/n)$ in the recursive GRNN by a more general sequence a_n used in stochastic approximation methods [1]. Due to such replacement, the recursive GRNN will be able to follow changes of time-varying regression functions (4). Since the existing theories do not allow to study recursive GRNN in time-varying environment, in Section IV we give the appropriate theorems which are very useful in the next sections. In Section V we introduce the GRNN studied in this paper and describe its relation with previous results concerning stochastic—based neural networks in stationary case. In Section VI we formulate a theorem for convergence of the GRNN in probability and with probability one to regressions (4). The GRNN based on the Parzen kernel and orthogonal series kernel are studied in Sections VII and VIII, respectively. The speed of convergence is investigated in Section IX. In Sections X–XII we design the GRNN tracking various nonstationarities. Section XIII presents simulation results. The proofs of all the theorems are given in the Appendix.

II. Kernel Functions for the GRNN Construction

All probabilistic neural networks studied in this paper are based on a sequence $\{K_n\}, n = 1, 2, \ldots$, of bivariate

Borel—measurable functions (so-called general kernel functions) defined on $A \times A, A \subset R^p, p \geq 1$. The concept of general kernel functions stems from the theory of nonparametric density estimation. We will use ideas of the two methods: Parzen's approach and orthogonal series.

A. Application of the Parzen Kernel

Sequence K_n based on the Parzen kernel in the multidimensional version takes the following form

$$K_n(x,u) = h_n^{-p} K\left(\frac{x-u}{h_n}\right) \tag{5}$$

where $h_n > 0$ is a certain sequence of numbers and K is an appropriately selected function. Precise assumptions concerning the sequence h_n and function K that ensure convergence will be given in the next sections. It is convenient to assume that function K can be presented in the form

$$K(x) = \prod_{i=1}^{p} H\left(x^{(i)}\right).$$

Then, sequence K_n is expressed by means of

$$K_n(x,u) = h_n^{-p} \prod_{i=1}^{p} H\left(\frac{x^{(i)} - u^{(i)}}{h_n}\right). \tag{6}$$

The most popular is Gaussian kernel given by

$$H(v) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}v^2} \tag{7}$$

and

$$K_n(x,u) = h_n^{-p} (2\pi)^{-\frac{1}{2}} \prod_{i=1}^{p} \exp\left(\frac{x^{(i)} - u^{(i)}}{h_n}\right)^2. \tag{8}$$

B. Application of Orthogonal Series

Let $g_j(\cdot), j = 0, 1, 2, \ldots$, be a complete orthonormal system in $L_2(\Delta), \Delta \in R$, such that

$$\max_x |g_j(x)| \leq G_j. \tag{9}$$

It is well known that, the system composed of all possible products

$$\left\{\Psi_{j_1,\ldots,j_p}\left(x^{(1)},\ldots,x^{(p)}\right) = g_{j_1}\left(x^{(1)}\right)\cdots g_{j_p}\left(x^{(p)}\right) \quad j_k = 0,1,2,\ldots, k = 1,\ldots,p\right\} \tag{10}$$

is a complete orthonormal system in $L_2(A)$, where

$$A = \underbrace{\Delta \times \cdots \times \Delta}_{p-\text{times}}.$$

It constitutes the basis for construction of the following sequence K_n

$$K_n(x,u) = \sum_{j_1=0}^{q} \cdots \sum_{j_p=0}^{q} g_{j_1}\left(x^{(1)}\right) g_{j_1}\left(u^{(1)}\right) \cdots g_{j_p}\left(x^{(p)}\right) g_{j_p}\left(u^{(p)}\right) \tag{11}$$

where q depends on the length of the learning sequence, i.e., $q = q(n)$. It can be given in a shortened form as

$$K_n(x,u) = \sum_{|\underline{j}| \leq q} \Psi_{\underline{j}}(x) \Psi_{\underline{j}}(u) \tag{12}$$

where $\underline{j} = (j_1, \ldots, j_p)$ and $|\underline{j}| = \max_{1 \leq k \leq p}(j_k)$.

If $\Delta = (-\infty, \infty)$, then we design the PNN based on the Hermite series given by

$$g_j(x) = \left(2^j j! \pi^{\frac{1}{2}}\right)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} H_j(x)$$

where

$$H_{j+1}(x) = 2xH_j(x) - 2jH_{j-1}(x)$$

and $H_0(x) = 1, H_1(x) = 2x$. It is easly seen that the orthonormal functions of the Hermite series can be recursively generated by

$$\left.\begin{array}{c} g_0(x) = \pi^{-\frac{1}{4}} e^{-\frac{x}{2}} \\ g_1(x) = 2^{\frac{1}{2}} \pi^{-\frac{1}{4}} x e^{-\frac{x}{2}} = 2^{\frac{1}{2}} x g_0(x) \\ g_{j+1}(x) = (2/(j+1))^{\frac{1}{2}} x g_j(x) \\ -(j/(j+1))^{\frac{1}{2}} g_{j-1}(x) \\ \text{for } j = 1, 2, \ldots, \end{array}\right\}.$$

It is known [39] that for the Hermite series $G_j = \text{const} \cdot j^{-12}$.

If $\Delta = [0, \infty)$, then we design the PNN based on the Laguerre series given by

$$g_j(x) = e^{-\frac{x}{2}} L_j(x)$$

where

$$(j+1)L_{j+1}(x) = (2j+1-x)L_j(x) - jL_{j-1}(x)$$

and $L_0(x) = 1, L_1(x) = 1 - x$. It is easly seen that the orthonormal functions of the Laguerre series can be recursively generated by

$$\left.\begin{array}{c} g_0(x) = e^{-\frac{x}{2}} \\ g_1(x) = e^{-\frac{x}{2}}(1-x) = g_0(x)(1-x) \\ (j+1)g_{j+1}(x) = (2j+1-x)g_j(x) \\ -jg_{j-1}(x) \\ \text{for } j = 1, 2, \ldots, \end{array}\right\}.$$

It is known [39] that for the Laguerre series $G_j = \text{const} \cdot j^{-1/4}$.

If $\Delta = [-1, 1]$, then we design the PNN based on the Legendre series given by

$$g_j(x) = \sqrt{\frac{2j+1}{2}} P_j(x)$$

where

$$(j+1)P_{j+1}(x) = (2j+1)xP_j(x) - jP_{j-1}(x)$$

and $P_0(x) = 1, P_1(x) = x$. It is easly seen that the orthonormal functions of the Legendre series can be recursively generated by

$$\left.\begin{array}{c} g_0(x) = \sqrt{\frac{1}{2}} \\ g_1(x) = \sqrt{\frac{3}{2}} x \\ (j+1)g_{j+1}(x) = \sqrt{(2j+1)(2j+3)} x g_j(x) \\ -\sqrt{\frac{(2j+3)}{(2j-1)}} j g_{j-1}(x) \\ \text{for } j = 1, 2, \ldots, \end{array}\right\}.$$

It is known [39] that for the Legendre series $G_j = \text{const} \cdot j^{1/2}$.

In some applications it is convenient to use multiple Fourier series. We present two different multiple Fourier series.

1) Expansions Based on Dirichlet's Kernel: It is well known [20] that the functions

$$\left\{ (2\pi)^{-p/2} e^{ikx}, k = (k_1, \ldots, k_p), kx = \sum_{j=1}^{p} k_j x^{(j)} \right. \\ \left. (k_j = 0, \pm 1, \pm 2, \ldots, j = 1, \ldots, p) \right\}$$

are orthonormal and complete over the p-dimensional cube

$$Q = \{-\pi \le x^{(j)} \le \pi, j = 1, \ldots, p\}.$$

The Dirichlet's kernel of order q is given by

$$D_q(x) = \prod_{j=1}^{p} D_q\left(x^{(j)}\right),$$
$$D_q(u) = \frac{1}{2} + \sum_{n=1}^{q} \cos \nu = \frac{\sin\left(q + \frac{1}{2}\right) u}{2 \sin \frac{u}{2}}.$$

2) Expansion Based on Fejer's Kernel: The multidimensional Fejer's kernel of order q is given by

$$F_q(x) = \prod_{j=1}^{p} \phi_q\left(x^{(j)}\right)$$

where

$$\phi_q(u) = \frac{1}{2(q+1)} \left(\frac{\sin \frac{1}{2}(q+1)u}{\sin \frac{1}{2} u} \right)^2.$$

III. Generalized Regression Neural Networks in Stationary Environment

Let (X, Y) be a pair of random variables. X takes values in a Borel set $A, A \subset R^p$, whereas Y takes values in R. Let f be the marginal Lebesgue density of X. Based on a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ of i.i.d. observations of (X, Y) we wish to estimate the regression ϕ of Y on X given by (2).

To estimate function (2) we propose the following

$$\tilde{\phi}_n(x) = \frac{\tilde{R}_n(x)}{\tilde{f}_n(x)} \tag{13}$$

where

$$\tilde{R}_n(x) = \frac{1}{n} \sum_{i=1}^{n} Y_i K_n(x, X_i) \tag{14}$$

and estimator $\tilde{f}_n$ is given by

$$\tilde{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_n(x, X_i). \tag{15}$$

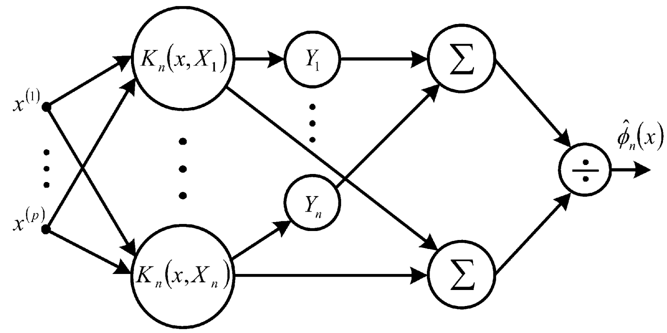

Fig. 1. Scheme of GRNN.

In Fig. 1 we show the neural-network implementation of estimator (13).

Example 1: (Nadaraya [19] and Watson [41]): Applying the Parzen kernel to estimator (13) one gets

$$\tilde{\phi}_n(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)}. \tag{16}$$

Several results concerning convergence of estimator (16) can be found in [9], [13], and [14].

The recursive version of procedure (13) is given as follows:

$$\hat{\phi}_n(x) = \frac{\hat{R}_n(x)}{\hat{f}_n(x)} \tag{17}$$

where

$$\hat{R}_n(x) = \frac{1}{n} \sum_{i=1}^{n} Y_i K_i(x, X_i) \tag{18}$$

and

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_i(x, X_i). \tag{19}$$

Observe that procedures (18) and (19) differ from (14) and (15). Formulas (18) and (19) can be expressed in the recursive form

$$\hat{R}_{n+1}(x) = \hat{R}_n(x) + \frac{1}{n+1} \cdot [Y_{n+1} K_{n+1}(x, X_{n+1}) - \hat{R}_n(x)] \tag{20}$$

and

$$\hat{f}_{n+1}(x) = \hat{f}_n(x) + \frac{1}{n+1} \cdot (K_{n+1}(x, X_{n+1}) - \hat{f}_n(x)) \tag{21}$$

where $\hat{R}_0(x) = 0$ and $\hat{f}_0(x) = 0$. The block diagram of the recursive GRNN corresponding to (17), (20), and (21) is depicted in Fig. 2.

Example 2: (Rutkowski [23], [24]): Since the orthogonal series kernel is less popular than the Parzen kernel we will explain how the orthogonal series method leads to density and regression function estimators. Let us define

$$R(x) = \phi(x) f(x).$$

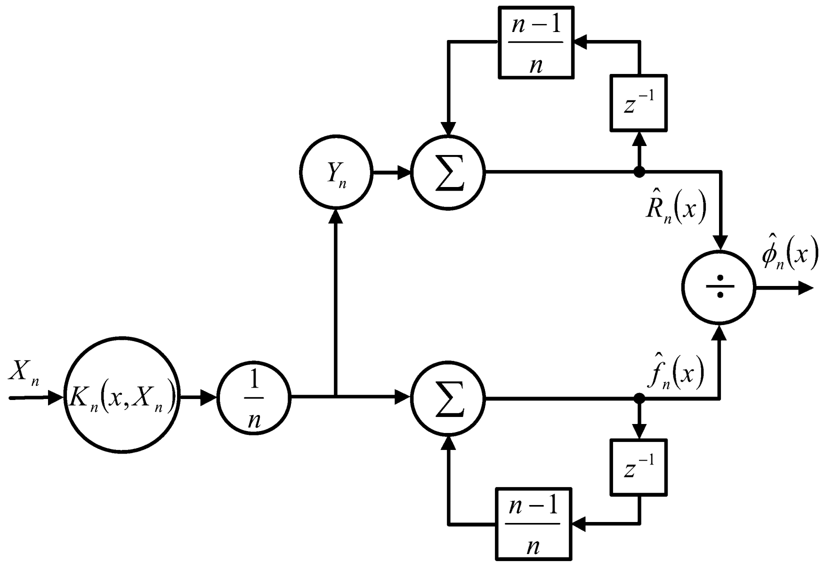

Fig. 2. Recursive GRNN.

We assume that functions R and f have the representations

$$R(x) \sim \sum_{k=0}^{\infty} A_k g_k(x)$$
$$f(x) \sim \sum_{k=0}^{\infty} B_k g_k(x)$$

where

$$A_k = \int_A R(x) g_k(x)\, dx = E[Y g_k(X)]$$
$$B_k = \int_A f(x) g_k(x) dx = E[g_k(X)].$$

We truncate the infinite orthogonal expansions as follows:

$$R_q(x) = \sum_{k=0}^{q} A_k g_k(x)$$
$$f_q(x) = \sum_{k=0}^{q} B_k g_k(x).$$

The coefficients A_k and $B_k$can be estimated by unbiased estimators

$$A_{kn} = \frac{1}{n}\sum_{i=1}^{n} Y_i g_k(X_i)$$
$$B_{kn} = \frac{1}{n}\sum_{i=1}^{n} g_k(X_i).$$

Replacing in $R_q(x)$ and $f_q(x)$ coefficients A_k and B_k by their estimates A_{kn} and B_{kn} we get

$$\tilde{R}_n(x) = \sum_{k=0}^{q(n)} A_{kn} g_k(x)$$
$$\tilde{f}_n(x) = \sum_{k=0}^{q(n)} B_{kn} g_k(x)$$

where q depends on the length of the learning sequence, i.e., $q = q(n)$. Finally we get a nonrecursive estimate of regression $\phi(x)$ given by (13). We will now derive the recursive orthogonal series regression estimate. Note that $\tilde{R}_n(x)$ and $\tilde{f}_n(x)$ can be expressed in the form

$$\tilde{R}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{q(n)} Y_i g_k(X_i) g_k(x)$$
$$\tilde{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{q(n)} g_j(X_i) g_j(x)$$

Replacing $q(n)$ by $q(i)$ in the last two expressions we get

$$\hat{R}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{q(i)} Y_i g_k(X_i) g_k(x)$$
$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{q(i)} g_j(X_i) g_j(x).$$

The above formulas can be presented in the recursive forms

$$\hat{R}_{n+1}(x) = \hat{R}_n(x) + \frac{1}{n+1} \times \left[\sum_{k=0}^{q(n+1)} Y_{n+1} g_k(X_{n+1}) g_k(x) - \hat{R}_n(x)\right]$$
$$\hat{f}_{n+1}(x) = \hat{f}_n(x) + \frac{1}{n+1} \times \left[\sum_{j=0}^{q(n+1)} g_j(X_{n+1}) g_j(x) - \hat{f}_n(x)\right]$$

and where $\hat{R}_0(x) = 0$ and $\hat{f}_0(x) = 0$. Thus, the unknown regression function $\phi(x)$ is estimated recursively by (17) based on the orthogonal series method.

It should be noted that the orthogonal series method is also applicable when the inputs $X_1, \ldots, X_n$ in model (4) are not random and are contained in the interval $A = [0,1]$ (see Rutkowski [25], [26]). Let us partition interval $A = [0,1]$ into n regions $A_1, \ldots, A_n$, where $A_i = [d_{i-1}, d_i], d_0 = 0, d_n = 1$, and $\cup A_i = A$. Assume that the input signals X_i are selected so that $X_i \in A_i$. We expand the regression function $\phi(x)$ in the orthogonal series

$$\phi(x) \sim \sum_{k=0}^{\infty} C_k g_k(x)$$

where

$$C_k = \int_A g_k(x)\phi(x)\,dx = \sum_{i=1}^{n} \int_{A_i} g_k(x)\phi(x)\,dx.$$

As the estimator of C_k we take

$$C_{kn} = \sum_{i=1}^{n} Y_i \int_{A_i} g_k(x)\,dx.$$

We truncate the infinite orthogonal expansion as follows:

$$\phi_q(x) = \sum_{k=0}^{q} C_k g_k(x)$$

and estimate $\phi(x)$ by

$$\hat{\phi}_n(x) = \sum_{k=0}^{q(n)} C_{kn} g_k(x)$$

where $q(n)$ is a sequence of integers. The convergence properties of $\hat{\phi}_n(x)$ are investigated in [25] and [26].

IV. Preliminaries to GRNN in Time-Varying Environment

Here we study a general problem of learning in the nonstationary environment. The results and theorems will be a starting point for construction of the GRNN in the next sections. Let us consider a sequence $\{(X_n, Y_n)\}, n = 1, 2, \ldots$, of independent pairs of random variables, where X_n represents random variables having the probability density f_n taking values in the set $A \subset R^p$ and Y_n represents random variables taking values in the set $B \subset R$.

We assume that time-varying probability distributions of the above random variables are completely unknown.

Let us define the following function

$$R_n(x) \stackrel{df}{=} f_n(x) E[Y_n \,|\, X_n = x] \quad n = 1, 2, \ldots \tag{22}$$

From the assumption that the probability distributions are completely unknown, it follows that the sequence of functions (22) is also unknown. In this paper, the goal of learning will be tracking the changing function $R_n, n = 1, 2, \ldots$.

Let $\{a_n\}$ be a sequence of numbers satisfying the following conditions:

$$a_n > 0, \quad a_n \xrightarrow{n} 0, \quad \sum_{n=1}^{\infty} a_n = \infty. \tag{23}$$

We will consider a nonparametric learning procedure of the following type:

$$\hat{R}_{n+1}(x) = \hat{R}_n(x) + a_{n+1}[Y_{n+1}K_{n+1}(x, X_{n+1}) - \hat{R}_n(x)]$$
$$n = 0, 1, 2, \ldots \hat{R}_0(x) = 0. \tag{24}$$

Comparing (24) to (21) we realize that algorithm (21) is a special case of the general procedure (24); if $a_n = n^{-1}$ and $Y_n = 1$. Similarly, procedure (24) reduces to recursion (20) if $a_n = n^{-1}$.

The measure of quality of the learning process in a given point $x \in A$ can be

$$I_n(x) = |\hat{R}_n(x) - R_n(x)|. \tag{25}$$

Of course, sequence $I_n(x)$ in a given point $x \in A$ is a sequence of random variables. We will show that

$$EI_n^2(x) \xrightarrow{n} 0 \quad \text{and} \quad I_n(x) \xrightarrow{n} 0 \text{ with pr. } 1$$

Define

$$r_n(x) = E[Y_n K_n(x, X_n)]. \tag{26}$$

Theorem 1: If in a certain point x, the following conditions are satisfied:

$$a_n \operatorname{var}[Y_n K_n(x, X_n)] \xrightarrow{n} 0 \tag{27}$$
$$a_n^{-1}|r_n(x) - R_n(x)| \xrightarrow{n} 0 \tag{28}$$
$$a_n^{-1}|R_{n+1}(x) - R_n(x)| \xrightarrow{n} 0 \tag{29}$$

then

$$EI_n^2(x) \xrightarrow{n} 0. \tag{30}$$

Theorem 2: If in a certain point x, the following conditions are satisfied:

$$\sum_{n=1}^{\infty} a_n^2 \operatorname{var}[Y_n K_n(x, X_n)] < \infty \tag{31}$$
$$\sum_{n=1}^{\infty} a_n^{-1}(r_n(x) - R_n(x))^2 < \infty \tag{32}$$
$$\sum_{n=1}^{\infty} a_n^{-1}(R_{n+1}(x) - R_n(x))^2 < \infty \tag{33}$$

then

$$I_n(x) \xrightarrow{n} 0 \text{ with pr. } 1 \tag{34}$$

Theorem 3: If the following conditions are satisfied:

$$\operatorname{var}[Y_n K_n(x, X_n)] = 0(n^A), \quad A > 0 \tag{35}$$
$$|R_{n+1}(x) - R_n(x)| = 0(n^{-B}), \quad B > 0 \tag{36}$$
$$|r_n(x) - R_n(x)| = 0(n^{-C}), \quad C > 0 \tag{37}$$
$$a_n = \frac{k}{n^a}, \quad k > 0, \quad 0 < a \le 1 \tag{38}$$

then

$$EI_n^2(x) \le l_1 n^{-2C} + l_2 n^{-r} \tag{39}$$

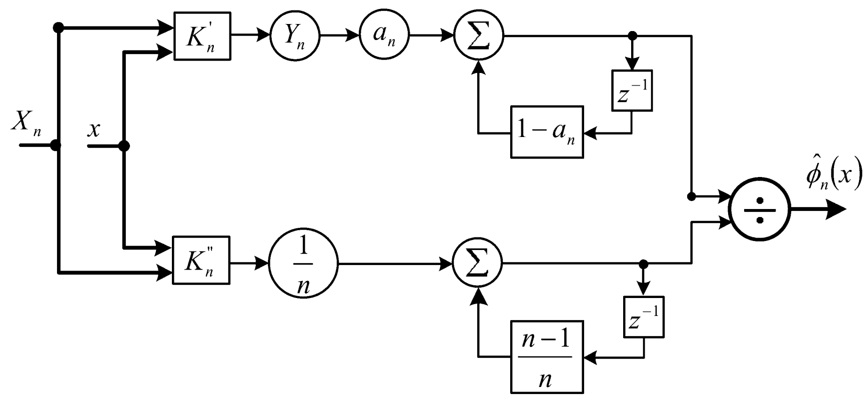

Fig. 3. Structual scheme of GRNN in time-varying environment.

where l_1, l_2—positive constants and

$$r = \min\{a - A, (B - a), 2(C - a)\} \tag{40}$$

with $r > 0$ for $0 < a < 1$ and $0 < r < 1$ for $a = 1$.

V. Problem Description and Presentation of the GRNN in Time-Varying Environment

The problem of nonparametric regression boils down to finding an adaptive algorithm that could follow the changes of optimal characteristics expressed by (4). This algorithm should be constructed on the basis of a learning sequence, i.e., observations of random variables

$$(X_1, Y_1), \quad (X_2, Y_2), \ldots$$

We assume that pairs of the above random variables are independent. In points x, where $f(x) \neq 0$, the characteristics of the best model (4) can be expressed as

$$\phi_n^*(x) = \frac{R_n(x)}{f(x)}, \quad n = 1, 2, \ldots \tag{41}$$

where $R_n(x) = \phi_n^*(x) f(x)$ corresponds to (22) if $f_n = f$. Because of this, the adaptive algorithm that is able to follow changes of unknown characteristics of the best model ϕ_n^*, will be constructed on the basis of a general procedure (24). The algorithm has the form

$$\hat{\phi}_n(x) = \frac{\hat{R}_n(x)}{\hat{f}_n(x)} \tag{42}$$

where $\hat{R}_n$ is expressed by means of (24) and $\hat{f}_n$ is a recurrent estimator of the density f given by (21). Observe that procedures (42) and (17) are equivalent when $a_n = n^{-1}$, i.e., in the stationary case. It is understandable that sequences K_n that are in the numerator and denominator of (42) can be of a different type. If sequences K_n are of the same type (e.g., based on the Parzen kernel), they generally should meet different conditions. In the block diagram (Fig. 3) of the GRNN that realizes algorithm (42) sequence K'_n present in the numerator of (41) was differentiated from sequence K''_n present in the denominator of that expression. In situation where there is no doubt, corresponding indexes at sequences h'_n and h''_n as well as $q'(n)$ and $q''(n)$ will be omitted.

VI. Convergence of the GRNN in Time-Varying Environment

The theorem presented below describes general conditions ensuring convergence of algorithm (42).

Theorem 4: (Pointwise Convergence of Algorithm (42) in Probability and With pr. 1): Let us assume that the following conditions are satisfied:

i) Condition A:

$$|\hat{R}_n(x) - R_n(x)| \xrightarrow{n} 0 \quad \text{in prob. (with prob. 1)} \tag{43}$$

ii) Condition B:

$$\hat{f}_n(x) \xrightarrow{n} f(x) > 0 \quad \text{in prob. (with prob. 1)} \tag{44}$$

iii) Condition C:

$$|\phi_n^*(x)||\hat{f}_n(x) - f(x)| \xrightarrow{n} 0 \quad \text{in prob. (with prob. 1)} \tag{45}$$

Then, for algorithm (42) we have

$$|\hat{\phi}_n(x) - \phi_n^*(x)| \xrightarrow{n} 0 \quad \text{in prob. (with prob. 1)} \tag{46}$$

Let us point out that condition A is satisfied when conclusions of Theorems 1 and 2 are true. Condition B reflects the requirement of the convergence of the estimator of the density function [expressed by (21)] and condition C imposes certain assumptions on the speed of this convergence. Of course, when ϕ_n^* is a bounded sequence, condition C boils down to condition B.

Now we will consider two methods of construction of algorithm (42). We will present procedures based on the Parzen kernel and on the orthogonal series method. In both cases we will present assumptions that guarantee satisfaction of conditions A, B, and C and, as a result, convergence (46). In this paper, we use the following symbols:

$$m'_n = \sup_x \left\{ \left(\text{var}\,[Y_n \mid X_n = x] + \phi_n^{*2}(x) \right) f(x) \right\} \tag{47}$$

and

$$m''_n = E[Y_n - \phi_n^*(X_n)]^2 + \int \phi_n^{*2}(x) f(x)\, \mathrm{d}x \tag{48}$$

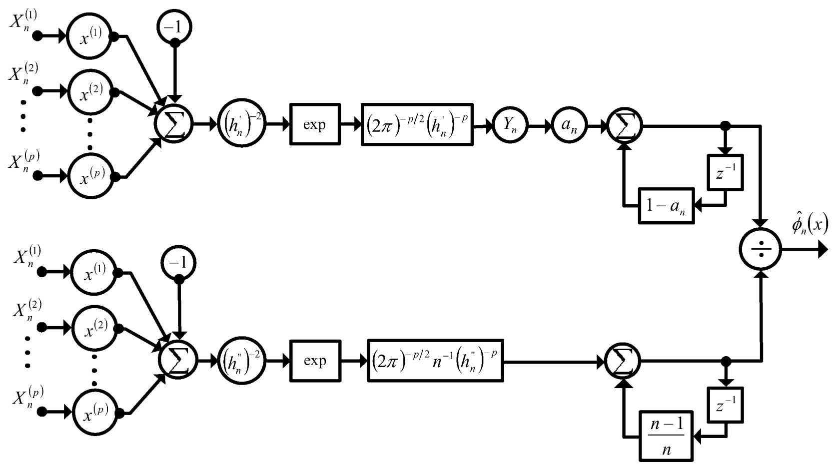

Fig. 4. Recursive generalized regression neural network based on the Parzen kernel in time-varying environment.

In Sections X–XII, we will discuss in detail regressions described by the equation

$$Y_n = \phi_n^*(X_n) + Z_n \tag{49}$$

where

$$EZ_n = 0, \quad EZ_n^2 = \sigma_z^2. \tag{50}$$

In such a situation, (47) and (48) take the form

$$m_n' = \sup_x \left\{ \left(\sigma_z^2 + \phi_n^{*2}(x)\right) f(x) \right\} \tag{51}$$

and

$$m_n'' = \sigma_z^2 + \int \phi_n^{*2}(x) f(x)\,\mathrm{dx}. \tag{52}$$

VII. The GRNN Based on the Parzen Kernel

The structural scheme of the system that realizes algorithm (42) on the basis of the Parzen kernel is depicted in Fig. 4. assuming use of kernel (6) and the normalization of vectors x and X_i. In order to differentiate sequences h_n and functions K present in the numerator and denominator of (42), symbols h_h' and h_n'' as well as K' and K'' are used. Condition A will be connected with the selection of sequence h_n' and conditions B and C with the selection of sequence h_n''. Now, we will present assumptions that guarantee satisfaction of conditions A, B, and C of Theorem 4.

a) Condition A As we remember (Section II), kernel K can be expressed in the following way:

$$K(x) = \prod_{i=1}^{p} H\left(x^{(i)}\right). \tag{53}$$

Let us assume that

$$\sup_{v \in R} |H(v)| < \infty \tag{54}$$

$$\int_R H(v)\,\mathrm{dv} = 1 \tag{55}$$

$$\int_R H(v) v^j\,\mathrm{dv} = 0, \quad j = 1, \ldots, r-1 \tag{56}$$

$$\int_R |H(v) v^k|\,\mathrm{dv} < \infty, \quad k = 1, \ldots, r. \tag{57}$$

For $r = 2$, the above conditions are satisfied by kernel (8). For $r = 4$, the conditions are met by the function

$$H(v) = \frac{3}{2\sqrt{2\pi}} \left(1 - \frac{v^2}{3}\right) e^{-\frac{1}{2}v^2}.$$

Let us introduce the following symbol:

$$D_n^{\underline{i}} = \sup_x \left| \frac{\delta^r}{\delta x^{(i_1)} \ldots \delta x^{(i_r)}} R_n(x) \right| \tag{58}$$

where $\underline{i} = (i_1, \ldots, i_r), i_k = 1, \ldots, p, k = 1, \ldots, r$.

We will associate parameter r with smooth properties of function R_n $(n = 1, 2, \ldots)$. The following theorems guarantee satisfaction of condition A.

Theorem 5: Let us assume that function K satisfies conditions (53)–(57), $h_n \xrightarrow{n} 0$ and one of the following assumptions holds:

$$a_n h_n^{-p} m_n' \xrightarrow{n} 0 \tag{59}$$

or

$$a_n h_n^{-2p} m_n'' \xrightarrow{n} 0. \tag{60}$$

If function ϕ_n^* changes with time in such a way that

$$a_n^{-1} |\phi_{n+1}^*(x) - \phi_n^*(x)| \xrightarrow{n} 0 \tag{61}$$

and

$$a_n^{-1} h_n^r D_n^{\underline{i}} \xrightarrow{n} 0 \tag{62}$$

then

$$E[\hat{R}_n(x) - R_n(x)]^2 \xrightarrow{n} 0.$$

Theorem 6: Let us assume that function K satisfies conditions (53)–(57), $h_n \xrightarrow{n} 0$ and one of the following assumptions holds:

$$\sum_{n=1}^{\infty} a_n^2 h_n^{-p} m_n' < \infty \tag{63}$$

or

$$\sum_{n=1}^{\infty} a_n^2 h_n^{-2p} m_n'' < \infty. \tag{64}$$

If function ϕ_n^* changes with time in such a way that

$$\sum_{n=1}^{\infty} a_n^{-1} (\phi_{n+1}^*(x) - \phi_n^*(x))^2 < \infty \tag{65}$$

and

$$\sum_{n=1}^{\infty} a_n^{-1} h_n^r \left(D_n^i\right)^2 < \infty \tag{66}$$

then

$$|\hat{R}_n(x) - R_n(x)| \xrightarrow{n} 0 \quad \text{with pr. 1.}$$

It is worth mentioning that (59) and (63) are weaker than the alternative (60) and (64) as far as the selection of sequence h_n is concerned. However, while designing the system that realizes algorithm (42) one should also take into account information that may be possessed about functions ϕ_n^* and f, because the mentioned assumptions depend also on m_n' and m_n'' [(51) and (52)]. Assumptions (62) and (66) concern certain conditions of smooth properties of function R_n. As we will see later (Sections X–XII), for regressions with various types of nonstationarity, they take a more concrete form.

b) Condition B A recurrent estimator (21) of the density function f with the Parzen kernel was studied in [7], [8], [13], [42]. The convergence of this estimator could be obtained by the use of general Theorems 1 and 2. However, one should remember that these theorems concern a nonstationary situation, so the conditions obtained would be quite strong. Therefore, results concerning the convergence of estimator (21) will be taken from [8] and [13] that were mentioned above. Let us assume that $h_n \xrightarrow{n} 0$ and kernel K satisfies conditions

$$K(x) \geq 0, \quad \int K(x)\,\mathrm{dx} = 1, \quad \sup K(x) < \infty \tag{67}$$

Devroye [8] showed that

$$n^{-2} \sum_{i=1}^{n} h_i^{-p} \xrightarrow{n} 0 \tag{68}$$

implies

$$\hat{f}_n(x) \xrightarrow{n} f(x) \tag{69}$$

in probability, and

$$\sum_{n=1}^{\infty} n^{-2} h_n^{-p} < \infty \tag{70}$$

implies convergence (69) with pr. 1, whereas both these convergences occur in the following points x:

i) in every point of continuity of f if

$$\|x\|^p K(x) \xrightarrow{n} 0 \quad \text{when} \quad \|x\| \xrightarrow{n} 0 \tag{71}$$

ii) in Lebesque points of function f if f is bounded,

iii) in Lebesque points of function f if $Kx(x)$ satisfies the condition

$$\begin{aligned} K(x) &\neq 0 \quad \text{when} \quad x \in B \subset R^p, \quad \mu(B) < \infty \\ K(x) &= 0 \quad \text{when} \quad x \in R^p - B. \end{aligned} \tag{72}$$

It is worth reminding that Lebesque points are points of continuity of the function and almost all points x. The speed of the convergence of procedure (21) can be evaluated by means of expression (see [13])

$$E[\hat{f}_n(x) - f(x)]^2 \leq c_1 n^{-2} \sum_{i=1}^{n} h_i^{-p} + c_2 n^{-2} \left(\sum_{i=1}^{n} h_i^2\right)^2 \tag{73}$$

if density function f has continuous partial derivatives up to the third order.

c) Condition C) As already mentioned, this condition imposes certain assumptions on the speed of convergence of estimator (21). Let us assume that density function f has continuous partial derivatives up to the third order. With the use of reasoning similar to that in [8] and using results of [13] it is possible to show that convergence (45) in version "in probability" is implied by

$$|\phi_n^*(x)| n^{-1} \sum_{i=1}^{n} h_i^2 \xrightarrow{n} 0 \tag{74}$$

and

$$\phi_n^{*2}(x) n^{-2} \sum_{i=1}^{n} h_i^{-p} \xrightarrow{n} 0 \tag{75}$$

whereas convergence (45) with pr. 1 is implied by (74) and

$$\sum_{n=1}^{\infty} \phi_n^{*2}(x) n^{-2} h_n^{-p} < \infty. \tag{76}$$

Analyzing the above assumptions, we may raise the following problem: how fast can ϕ_n^* grow to infinity (if ϕ_n^* is an unbounded

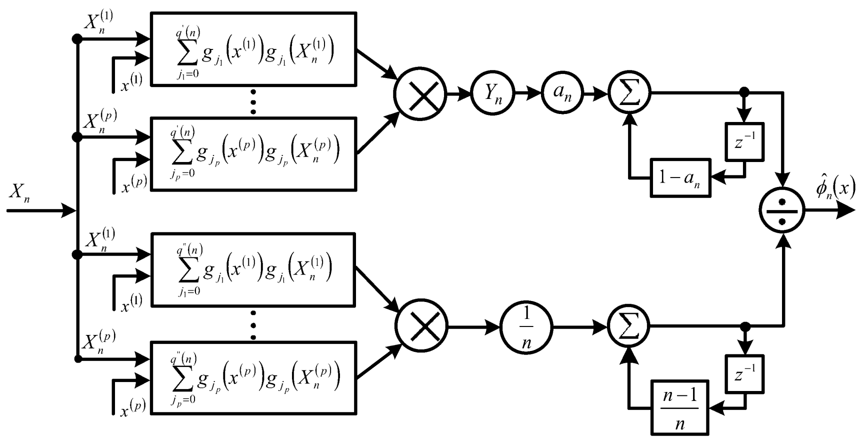

Fig. 5. Recursive GRNN based on the orthogonal series kernel in time-varying environment.

sequence) so that condition C could be satisfied at all? Let, e.g., in a certain point x

$$|\phi_n^*(x)| = 0(n^\alpha) \quad \alpha > 0.$$

Of course, (74) and (75) are now satisfied when

$$0 < \alpha < 1$$

while (74) and (76) are satisfied when

$$0 < \alpha < \frac{1}{2}.$$

In other words, algorithm (42) has tracking properties if function ϕ_n^* does not grow to infinity too fast. Other limitations regarding sequence ϕ_n^* result from assumptions formulated in Therorems 5 and 6. This problem will be discussed in detail in the subsequent sections, considering regressions of various types of nonstationarity.

VIII. The GRNN Based on the Orthogonal Series Kernel

The structural scheme of the GRNN that realizes algorithm (42) on the basis of the orthogonal series kernel is shown in Fig. 5. In order to differentiate between sequences q_n that appear in the numerator and denominator of (42), symbols $q'(n)$ and $q''(n)$ were used. Like in the previous section we will specify assumptions ensuring satisfaction of conditions A, B, and C of Theorem 4. Condition A will be connected with the selection of sequence $q'(n)$ and the conditions B and C will be connected with the selection of sequence $q''(n)$.

a) Condition A Let us denote

$$S_n(x) = \left| \sum_{|\underline{j} \le q|} b_{\underline{j}n} \Psi_{\underline{j}}(x) - R_n(x) \right| \tag{77}$$

where

$$b_{jn} = \int R_n(x) \Psi_{\underline{j}}(x) \, \mathrm{dx}$$

The following two theorems guarantee the meeting of condition A.

Theorem 7: Let us assume that $q(n) \xrightarrow{n} \infty$, (9) and (61) hold and one of the following two conditions is satisfied

$$a_n \left(\sum_{j=0}^{q(n)} G_j^2 \right)^p m_n' \xrightarrow{n} 0 \tag{78}$$

or

$$a_n \left(\sum_{j=0}^{q(n)} G_j^2 \right)^{2p} m_n'' \xrightarrow{n} 0. \tag{79}$$

If

$$a_n^{-1} S_n(x) \xrightarrow{n} 0 \tag{80}$$

then

$$E[\hat{R}_n(x) - R_n(x)]^2 \xrightarrow{n} 0$$

Theorem 8: Let us assume that $q(n) \xrightarrow{n} \infty$, assumptions (9) and (65) hold and one of the following two conditions is satisfied

$$\sum_{n=1}^{\infty} a_n^2 \left(\sum_{j=0}^{q(n)} G_j^2 \right)^p m_n' < \infty \tag{81}$$

or

$$\sum_{n=1}^{\infty} a_n^2 \left(\sum_{j=0}^{q(n)} G_j^2 \right)^{2p} m_n'' < \infty. \tag{82}$$

If

$$\sum_{n=1}^{\infty} a_n^{-1} S_n^2(x) < \infty \tag{83}$$

then

$$|\hat{R}_n(x) - R_n(x)| \xrightarrow{n} 0 \quad \text{with pr. 1}$$

The problem of convergence (80) and (83) is not a standard problem of the orthogonal series theory because the expanded function R_n changes with the increase of n. Even if n is fixed, the problem of convergence of multidimensional series

$$\sum_{\underline{j}} b_{\underline{j}n}\Psi_{\underline{j}}(x) \longrightarrow R_n(x) \tag{84}$$

is not trivial. It was investigated in more detail for the Fourier series (e.g., Sjolin [31]) but it is less known for the multidimensional Hermite series. For these two series, we will specify conditions (80) and (83). It is possible to show (the proof can be found in the Appendix) that these conditions take the following form:

$$a_n^{-1}\left\|t_n^l\right\|_{L_2} q^{-ps}(n) \xrightarrow{n} 0 \tag{85}$$

$$\sum_{n=1}^{\infty} a_n^{-1}\left\|t_n^l\right\|_{L_2}^2 q^{-2ps}(n) < \infty \tag{86}$$

where

$$t_n^l(x,R_n) = \begin{cases} \prod_{k=1}^p \left(x^{(k)} - \frac{\partial}{\partial x^{(k)}}\right)^l R_n\left(x^{(1)},\ldots,x^{(p)}\right) \\ \quad \text{for the Hermite series} \\ \prod_{k=1}^p \frac{\partial^l}{\partial x^{(k)l}} R_n\left(x^{(1)},\ldots,x^{(p)}\right) \\ \quad \text{for the Fourier series} \end{cases}$$

and

$$s = \begin{cases} \frac{l}{2} - \frac{5}{12} & \text{for the Hermite series} \\ l - \frac{1}{2} & \text{for the Fourier series.} \end{cases}$$

Parameter l we will be associated with smooth properties of function $R_n, (n = 1, 2, \ldots)$.

The aforementioned conditions were derived under the following assumptions:

i) Hermite series: $t_n^l \in L_2(R^p), l \geq 1$ and condition (84) holds

ii) Fourier series: $t_n^l \in L_2(A), l \geq 1, A = [-\pi,\pi]^p$, (84) holds and it is assumed that function t_n^l and its partial derivatives up to the order of $p-1$ are equal 0 on the boundary of A.

Conditions (85) and (86) are connected with the assessment of the "tail" of series (84); it follows from (77). Unfortunately, the assessment of the "tail" of an orthogonal series requires making rather complicated assumptions concerning functions expanded into this series. However, assumptions of this type are typical in all works devoted to the orthogonal series theory (see, e.g., monograph [30]).

b) Condition B A recurrent estimator (21) of a density function constructed on the basis of orthogonal series was proposed by Rutkowski in [23] and [24]. Let us assume that $q(n) \xrightarrow{n} \infty$. Now, the condition

$$n^{-2}\sum_{i=1}^{n}\left(\sum_{j=0}^{q(i)} G_j^2\right)^{2p} \xrightarrow{n} 0 \tag{87}$$

implies a weak convergence (44) and

$$\sum_{n=1}^{\infty} n^{-2}\left(\sum_{j=0}^{q(n)} G_j^2\right)^{2p} < \infty \tag{88}$$

implies a strong convergence (44) at every point x where

$$\sum_{|\underline{j}|\leq q} \Psi_{\underline{j}}(x)d_j \to f(x) \tag{89}$$

where $d_j = \int \Psi_{\underline{j}}(x)f(x)\,\mathrm{dx}$. If (89) is true for almost every x, then a weak and a strong convergence of procedure (21) is true also almost everywhere. Convergence (89) depends on the orthonormal series used and on the properties of function f. In the one-dimensional (1-D) case, $p = 1$, the following results are known.

— For the Fourier, Legendre, Laguerre, and Hermite series, various conditions imposed on function f, ensuring point and uniform convergence (89) were given by Sansone [30];

— for the Haar series, (89) is true in almost all points x for any function $f \in L_1$(Alexits [2]);

— for the Fourier series, (89) is true in almost all points x for any function $f \in L_2$ (Carleson [5]); a similar result may be obtained for Legendre, Laguerre, and Hermite series using theorems of equivalent convergence (Szegö [37]);

— for the Fourier series with Fejer's kernel (see Section II), convergence of (89) is true in almost all points x for any function $f \in L_1$ (Sansone [30]); this result may be extended for Laguerre and Hermite series with the help of the above mentioned theorems on equivalent convergence.

Unfortunately, in the multidimensional case, the conditions for the convergence (89) are known only for the Fourier series

— for any function $f \in L_2$ (89) is true in almost all points x (Sjolin [31]).

— for the multidimensional Fourier series with Fejer's kernel, convergence (89) is true uniformly if f is a continuous function (Nikolski [20]).

The speed of convergence of estimate (21) can be evaluated by

$$E[\hat{f}_n(x) - f(x)]^2 \leq c_1 n^{-2}\sum_{i=1}^{n} q^{2pw}(i) + c_2 n^{-2}\left(\sum_{i=1}^{n} q^{-ps}(i)\right)^2 \tag{90}$$

where

$$s = \begin{cases} \frac{l}{2} - \frac{5}{12} & \text{for the Hermite series} \\ l - \frac{1}{2} & \text{for the Fourier series} \end{cases}$$

$$w = \begin{cases} \frac{5}{6} & \text{for the Hermite series} \\ 1 & \text{for the Fourier series} \end{cases}$$

c) Condition C Concrete assumptions imposed on sequence $q(n)$ that appear in the denominator of (42) that guarantee the satisfaction of condition C can be derived with the use of reasoning similar to that in Rutkowski [24]. Now, we obtain a weak convergence (45) if

$$n^{-2}\phi_n^{2*}(x)\sum_{i=1}^{n}\left(\sum_{j=0}^{q(i)}G_j^2\right)^{2p} \xrightarrow{n} 0 \tag{91}$$

and a strong convergence (45) if

$$\sum_{n=1}^{\infty} n^{-2}\phi_n^{2*}(x)\left(\sum_{j=0}^{q(i)}G_j^2\right)^{2p} < \infty. \tag{92}$$

In both cases, we should also assume that

$$|\phi_n^*(x)|n^{-1}\left|\sum_{i=1}^{n}\left(\sum_{|\underline{j}|\le q(i)}\Psi_{\underline{j}}(x)d_{\underline{j}} - f(x)\right)\right| \xrightarrow{n} 0. \tag{93}$$

Condition (93) can take a concrete form depending on the orthogonal series applied and assumptions imposed on function f. Assuming that the orthogonal expansion of function f is convergent in point x (or in almost every point x), i.e., (89) holds we will use the Hermite and Fourier orthogonal systems. Let us define

$$t^l(x;f) = \begin{cases} \prod_{k=1}^{p}\left(x^{(k)} - \frac{\partial}{\partial x^{(k)}}\right)^l f(x) \\ \qquad \text{for the Hermite series}, l \ge 1 \\ \prod_{k=1}^{p}\frac{\partial^l}{\partial x^{(k)l}} f(x) \\ \qquad \text{for the Fourier series.} \end{cases}$$

Let us assume that $t^l \in L_2$. Now, (93) takes the form

$$|\phi_n^*(x)|n^{-1}\sum_{i=1}^{n}[q(i)]^{-ps} \xrightarrow{n} 0 \tag{94}$$

where

$$s = \begin{cases} \frac{l}{2} - \frac{5}{12} & \text{for the Hermite series} \\ l - \frac{1}{2} & \text{for the Fourier series.} \end{cases}$$

In other words, the satisfaction of condition C depends on the smooth properties of an unknown function f. From (91) and (92) it follows that $|\phi_n^*|$ cannot grow to infinity too fast. If, e.g., $|\phi_n^*(x)| = 0(n^\alpha), \alpha > 0$, then parameter α should be contained within the same bounds as in the case of the use of the algorithm based on the Parzen kernel.

IX. Speed of Convergence

The problem of investigating the speed of the convergence of procedure (42) which is a quotient of two algorithms is a relatively complex one. The following theorem allows us to assess the speed of the convergence of procedure (42) on the basis of the knowledge of the speed of the convergence of procedures (21) and (24):

Theorem 9: For any $\varepsilon > 0$, the following inequality holds

$$\begin{aligned} P(|\hat{\phi}_n(x) - \phi_n^*(x)| > \varepsilon) \\ \le \left(\frac{\varepsilon+2}{\varepsilon f(x)}\right)^2 \Big[E[\hat{R}_n(x) - R_n(x)]^2 \\ + \left(\phi_n^{*^2}(x)+1\right)E[\hat{f}_n(x) - f(x)]^2\Big]. \end{aligned} \tag{95}$$

Using the above inequality, we will later assess the speed of convergence of procedure (42) used for regressions with particular types of nonstationarity.

In the context of stationary problems, other authors also considered nonparametric procedures (based on the Parzen kernel) that are a quotient of two algorithms. Moreover, they carried out an optimization of the speed of convergence. However, they assumed that $h_n' = h_n''$ which is justified in a stationary case. In a nonstationary case, sequences h_n' and h_n'' as well as $q'(n)$ and $q''(n)$ which are present in the numerator and denominator of (42) should usually satisfy various conditions, which makes their optimal selection difficult [e.g., in the sense of minimizing the right-hand side (RHS) of (95)]. The matter is further complicated by the necessity of selection of sequence a_n (in the stationary case $a_n = n^{-1}$) and by the influence of nonstationarity in an expression that estimates the speed of convergence of the numerator of procedure (42) and in the RHS of (95). That is why sequences $a_n, q(n)$, and $h(n)$ will not be the subject of optimization. We will be satisfied with the fact that the conditions given in this paper allow us to design a system that realizes algorithm (42), i.e., to select sequences a_n, h_n', and h_n'' (or $a_n, q'(n), q''(n)$) implying possession of tracking properties by this algorithm, which is by no means an easy task in a nonstationary situation. Expression (95), as aforementioned, will be used several times in the next sections for analysis of the influence of various factors on the speed of convergence of procedure (42).

X. Multiplicative Nonstationarity

Let us consider regressions described by (49), where

$$\phi_n^*(x) = \alpha_n\phi(x) \tag{96}$$

where α_n—unknown sequence of numbers, ϕ—unknown function. In Fig. 6, we illustrate an application of the GRNN for modeling a plant described by (49) with nonstationarity (96).

In Tables I–IV, based on the results of Sections VII and VIII, we present the conditions implying convergence of algorithm (42) used for tracking regressions with multiplicative nonstationarity. Tables I and II give proper conditions for the algorithm based on the Parzen kernel, whereas Tables III and IV give similar conditions for the algorithm based on the orthogonal series kernel. In order to specify these conditions more precisely, in Tables III and IV two specific multidimensional orthogonal series were considered: the Fourier series and the Hermite series. We should notice that now

$$R_n(x) = \alpha_n f(x)\phi(x). \tag{97}$$

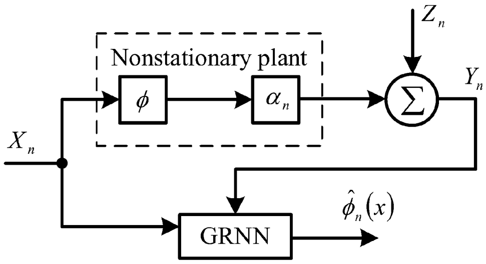

Fig. 6. GRNN for modeling a plant with multiplicative nonstationarity.

TABLE I
CONDITIONS FOR WEAK CONVERGENCE OF GRNN BASED ON THE PARZEN KERNEL—MULTIPLICATIVE NONSTATIONARITY

Condition	$\left\|\widehat{\phi}_n(x) - \varphi_n^*(x)\right\| \xrightarrow{n} 0$ wg pr.
(59)	$a_n h_n^{-p}\left(\alpha_n^2+1\right) \xrightarrow{n} 0$
(60)	$a_n h_n^{-2p}\left(\alpha_n^2+1\right) \xrightarrow{n} 0$
(61)	$a_n^{-1}\left\|\alpha_{n+1}-\alpha_n\right\| \xrightarrow{n} 0$
(62)	$a_n^{-1} h_n^r \left\|\alpha_n\right\| \xrightarrow{n} 0$
(74)	$\left\|\alpha_n\right\| n^{-1}\sum_{i=1}^n h_i^2 \xrightarrow{n} 0$
(68, 75)	$\left(\alpha_n^2+1\right) n^{-2}\sum_{i=1}^n h_i^{-p} \xrightarrow{n} 0$

Because of the "separability" of the nonstationary factor α_n, assumptions (62) and (66) [as well as (80) and (83)], connected with smooth properties of function R_n $(n=1,2,\ldots)$ reduce to assumptions concerning smooth properties of function $f\phi$, which simplifies significantly the convergence conditions described in Sections VIII and VIII.

Remark 1: Conditions (59)–(62) and (63)–(66) presented in Tables I and II concern the selection of sequence h_n'. Conditions (74), (68), (75), and (70), (76) concern the selection of sequence h_n'' (according to symbols in Fig. 7). In a similar way, (61), (78), (79), (80), and (65), (81), (82), (83) presented in Tables III and IV concern the selection of sequence $q'(n)$. Conditions (87), (92), (91), and (88), (93) concern the selection of sequence $q''(n)$ (according to symbols in Fig. 8).

It is obvious that sequence α_n cannot change in a completely arbitrary way so that algorithm (42) could possess the tracking property. Nevertheless, we will show that the class of considered sequences α_n is quite wide. From the point of view of maintaining tracking properties by algorithm (42), situations when characteristic (96) is for a certain x divergent to infinity or does not have a finite limit, seem to be particularly difficult. Such cases are illustrated by the following examples of sequences α_n:

a)

$$\alpha_n = c_0 + c_1 n^{t_1} + c_2 n^{t_2} + \cdots + c_k n^{t_k}$$

where $c_0, c_1, \ldots, c_k$—real numbers and $t_j > 0, j = 1,\ldots,k$.

TABLE II
CONDITIONS FOR STRONG CONVERGENCE OF GRNN BASED ON THE PARZEN KERNEL—MULTIPLICATIVE NONSTATIONARITY

Condition	$\left\|\widehat{\phi}_n(x) - \varphi_n^*(x)\right\| \xrightarrow{n} 0$ with pr. 1
(63)	$\sum_{n=1}^\infty a_n^2 h_n^{-p}\left(\alpha_n^2+1\right) < \infty$
(64)	$\sum_{n=1}^\infty a_n^2 h_n^{-2p}\left(\alpha_n^2+1\right) < \infty$
(65)	$\sum_{n=1}^\infty a_n^{-1}\left(\alpha_{n+1}-\alpha_n\right)^2 < \infty$
(66)	$\sum_{n=1}^\infty a_n^{-1} h_n^{2r}\alpha_n^2 < \infty$
(74)	$\left\|\alpha_n\right\| n^{-1}\sum_{i=1}^n h_i^2 \xrightarrow{n} 0$
(70, 76)	$\sum_{n=1}^\infty \left(\alpha_n^2+1\right) n^{-2} h_n^{-p} < \infty$

TABLE III
CONDITIONS FOR WEAK CONVERGENCE OF GRNN BASED ON THE ORTHOGONAL SERIES METHOD—MULTIPLICATIVE NONSTATIONARITY

Condition	$\left\|\widehat{\phi}_n(x) - \varphi_n^*(x)\right\| \xrightarrow{n} 0$ in probability
(61)	$a_n^{-1}\left\|\alpha_{n+1}-\alpha_n\right\| \xrightarrow{n} 0$
(78)	$a_n q^{(2d+1)p}(n)\left(\alpha_n^2+1\right) \xrightarrow{n} 0$
(79)	$a_n q^{(2d+1)2p}(n)\left(\alpha_n^2+1\right) \xrightarrow{n} 0$
(80)	$a_n^{-1}\left\|\alpha_n\right\| q^{-ps}(n) \xrightarrow{n} 0$
(87, 92)	$\left(\alpha_n^2+1\right) n^{-2}\sum_{i=1}^n q^{(2p+1)2p}(n) \xrightarrow{n} 0$
(91)	$\left\|\alpha_n\right\| n^{-1}\sum_{i=1}^n q^{-ps}(i) \xrightarrow{n} 0$

TABLE IV
CONDITIONS FOR STRONG CONVERGENCE OF GRNN BASED ON THE ORTHOGONAL SERIES METHOD—MULTIPLICATIVE NONSTATIONARITY

Condition	$\left\|\widehat{\phi}_n(x) - \varphi_n^*(x)\right\| \xrightarrow{n} 0$ with pr. 1
(65)	$\sum_{n=1}^\infty a_n^{-1}\left(\alpha_{n+1}-\alpha_n\right)^2 < \infty$
(81)	$\sum_{n=1}^\infty a_n^2 q^{(2d+1)p}(n)\left(\alpha_n^2+1\right) < \infty$
(82)	$\sum_{n=1}^\infty a_n^2 q^{(2d+1)2p}(n)\left(\alpha_n^2+1\right) < \infty$
(83)	$\sum_{n=1}^\infty a_n^{-1}\alpha_n^2 q^{-2ps}(n) < \infty$
(88, 92)	$\sum_{n=1}^\infty \left(\alpha_n^2+1\right) n^{-2} q^{(2d+1)2p}(n) < \infty$
(91)	$\left\|\alpha_n\right\| n^{-1}\sum_{i=1}^n q^{-ps}(i) \xrightarrow{n} 0$

b)

$$\alpha_n = c_1 n^t + c_2 \log n + c_3$$

where c_1, c_2, c_3, are real numbers, $t > 0$

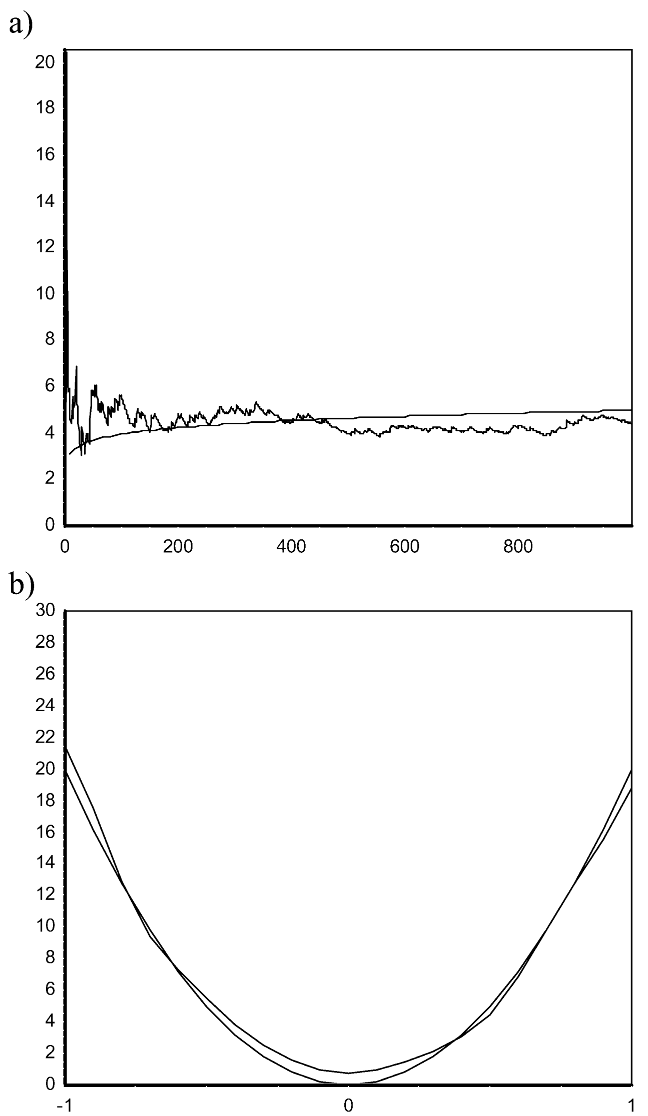

Fig. 7. Illustration of simulation 1 ($a = 0.8, H' = 0.3, H'' = 0.4, k_1' = k_2'' = 1.5$).

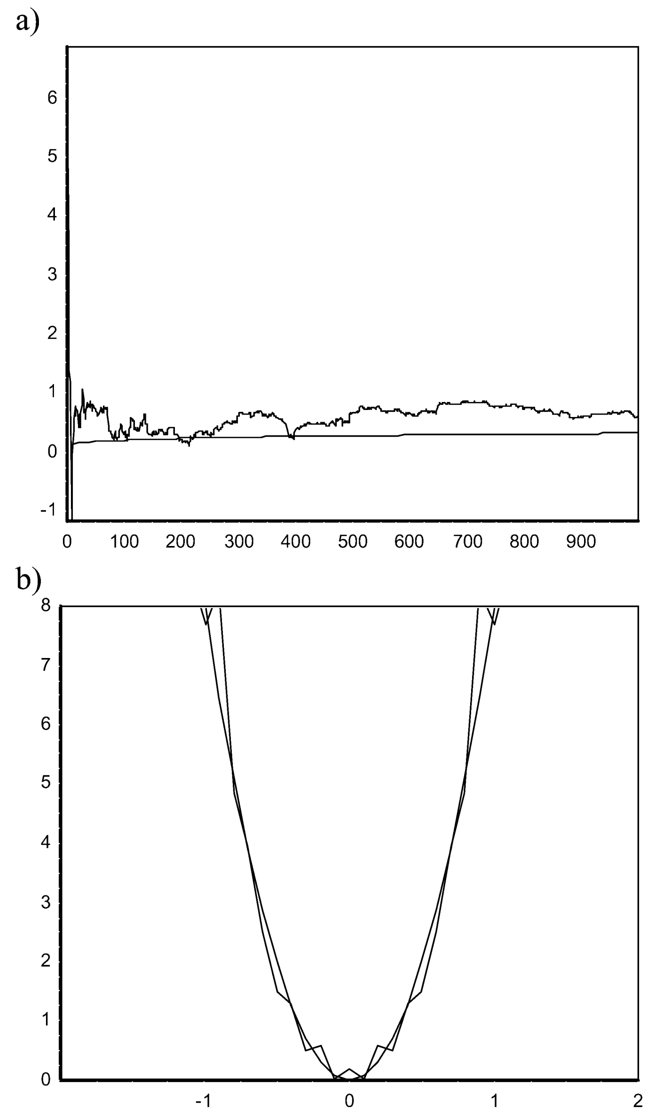

Fig. 8. Illustration of simulation 2 ($a = 0.9, Q' = 0.2, Q'' = 0.2, k_1' = k_2'' = 1$).

c)

$$\alpha_n = c_1 \sin A_n + c_2 \cos B_n + c_3$$

where $A_n = k_1 n^{t_1}, B_n = k_2 n^{t_2}, c_1, c_2, c_3, k_1, k_2$ are real numbers, $t_1, t_2 > 0$.

d)

$$\alpha_n = c_1 n^{t_1} \sin A_n + c_2$$

where $A_n = k_n{}^{-t_2}, c_1, c_2, k$ are real numbers, $t_1, t_2 > 0$, and $t_1 > t_2$.

e)

$$\alpha_n = c_1 n^{t_1} \sin A_n + c_2 n^{t_2} \cos B_n + c_3$$

where $A_n = k_1 n^{\tau_1}, B_n = k_2 n^{\tau_2}, c_1, c_2, c_3, k_1, k_2$ are real numbers, $t_1, t_2, \tau_1, \tau_2 > 0$.

f)

$$\alpha_n = c_1 n^{t_1} + c_2 \sin A_n + c_3 \cos B_n + c_4$$

where $A_n = k_1 n^{\tau_1}, B_n = k_2 n^{\tau_2}, c_1, c_2, c_3, c_4, k_1, k_2$ are any real numbers, $t_1, t_2, \tau_1, \tau_2 > 0$.

Let us now choose in algorithm (42) the following parameters:

$$h_n' = k_1' n^{-H'}, \quad h_n'' = k_1'' n^{-H''}, \quad H', H'' > 0, \quad k_1', k_1'' > 0$$

for the algorithm based on the Parzen kernel and

$$q'(n) = [k_2' n^{Q'}],$$
$$q''(n) = [k_2'' n^{Q''}], \quad Q', Q'' > 0, \quad k_2', k_2'' > 0$$

for the algorithm based on the orthogonal series method, ($[a]$ stands for the integer part of a). In both cases we take

$$a_n = k/n^a, \quad 0 < a \leq 1, \quad k > 0$$

Analyzing all the conditions given in Tables I–IV it is possible to specify precisely within what limits the constants t, t_j, τ_j present in examples (a)–(f) should be contained so that algorithm (42) could possess tracking properties. The results are shown in Tables V and VI.

TABLE V
CONDITIONS IMPOSED ON CONSTANTS t, t_j, τ_j FROM EXAMPLES (a)–(f); WEAK CONVERGENCE

Example	$\left\vert \widehat{\phi}_n(x) - \phi_n^*(x) \right\vert \xrightarrow{n} 0$ in probability
a	$0 < t_j < \frac{1}{3}, \; j = 1, ..., k$
b	$0 < t < \frac{1}{3}$
c	$0 < t_j < 1, \quad j = 1, 2,$
d	$0 < t_1 - t_2 < \frac{1}{3}$
e	$0 < t_j < \frac{1}{3}, \quad j = 1, 2,$ $0 < t_j + \tau_j < 1, \quad j = 1, 2,$
f	$0 < t_1 < \frac{1}{3}$ $0 < \tau_j < 1, \quad j = 1, 2,$

TABLE VI
CONDITIONS IMPOSED ON CONSTANTS t, t_j, τ_j FROM EXAMPLES (a)–(f); STRONG CONVERGENCE

Example	$\left\vert \widehat{\phi}_n(x) - \phi_n^*(x) \right\vert \xrightarrow{n} 0$ with pr. 1
a	$0 < t_j < \frac{1}{6}, \; j = 1, ..., k$
b	$0 < t < \frac{1}{6}$
c	$0 < t_j < \frac{1}{2}, \quad j = 1, 2,$
d	$0 < t_1 - t_2 < \frac{1}{6}$
e	$0 < t_j < \frac{1}{6}, \quad j = 1, 2,$ $0 < t_j + \tau_j < \frac{1}{2}, \quad j = 1, 2,$
f	$0 < t_1 < \frac{1}{6}$ $0 < \tau_j < \frac{1}{2}, \quad j = 1, 2,$

It is worth emphasizing that for the designing of a system that would realize algorithm (42), i.e., for a proper selection of sequences $h_n, q(n)$, and a_n, it is not necessary to precisely know sequences $\alpha_n$that were specified in examples (a)–(e) but only to know the information contained in Tables V and VI. For example, in order to track changes in the model

$$Y_n = (c_1 n^t + c_2 \log n + c_3)\phi(X_n) + Z_n$$

where t—unknown parameter, ϕ—unknown function, it is possible to use algorithm (42) if $0 < t < (1/3)$ for week convergence and $0 < t < (1/6)$ for strong.

We will now investigate the speed of convergence of algorithm (42). For this purpose one should:

1) use dependence (95),
2) determine constants A, B, and C that are present in assumptions of Theorem 3 and then, with use of this theorem, evaluate the speed of convergence of
$$E[\hat{R}_n(x) - R_n(x)]^2 \xrightarrow{n} 0$$
3) on the basis of inequalities (73) or (90), evaluate the speed of convergence of
$$E[\hat{f}_n(x) - f(x)]^2 \xrightarrow{n} 0$$

Example 3: Assuming that

$$\alpha_n = \text{const.}\, n^t, \quad t > 0$$

we will evaluate the speed of convergence of algorithm (42) based on the Parzen kernel and the orthogonal series kernel. We will assume that sequences $h_n, q(n)$, and a_n are of a power type.

A. Speed of Convergence of Algorithm (42) Based on the Parzen Kernel

In this case we have

$$A = 2t + H'p$$
$$B = 1 - t$$
$$C = rH' - t,$$

where parameter r is connected with smooth properties of function f. Omitting some simple calculations, we obtain

$$\begin{aligned} &P(|\hat{\phi}_n(x) - \phi_n^*(x)| > \varepsilon) \\ &\quad \le \left(\frac{\varepsilon + 2}{\varepsilon f(x)}\right)^2 n^{2t} \Big(c_1 n^{-2rH'} + c_2 n^{-r_1} \\ &\qquad + c_3 n^{-4H''} + c_4 n^{-(1-H''p)}\Big) \end{aligned} \tag{98}$$

where

$$r_1 = \min[a - H'p, 2(1-a), 2(rH' - a)]$$

An optimal selection of parameters H', H'', and a, minimizing the RHS of (98) seems to be a complicated problem. However, this expression may be used for designing a system that realizes algorithm (42) in such a way that it could possess tracking properties. Analyzing (98) we realize that parameters H' and H'' should satisfy

$$\frac{t + a}{r} < H' < \frac{a - 2t}{p}, \quad a < 1 - t$$
$$\frac{t}{2} < H'' < \frac{1 - 2t}{p}.$$

We should point out that the maximum value of t, with which the algorithm has tracking properties, depends on parameter r

specifying smooth properties of function ϕf. From the above inequalities it follows that if

$$t \in \left(0, \frac{1}{3} - \frac{p}{3r}\right), \quad r > p$$

then algorithm (42) is convergent. Let us notice that an increase of dimension p results in a decrease of the above range and an increase of smooth properties results in an increase of this range.

B. The Speed of Convergence of Algorithm (42) Based on the Orthogonal Series Kernel

Referring to symbols from Theorem 3, we obtain

$$\begin{aligned} A &= 2t + Q'(2d+1)p \\ B &= 1 - t \\ C &= pQ's - t \end{aligned}$$

where

$d = -(1/12), s = (l/2) - (5/12)$ for the Hermite series
$d = 0, s = l - (1/2)$ for the Fourier series,

and parameter l is connected with smooth properties of function ϕf. The speed of convergence of procedure (42) can be now expressed in the following way

$$\begin{aligned} &P(|\hat{\phi}_n(x) - \phi_n^*(x)| > \varepsilon) \\ &\quad \le \left(\frac{\varepsilon + 2}{\varepsilon f(x)}\right)^2 n^{2t} \left(c_1 n^{-2pQ's} + c_2 n^{-r_1}\right. \\ &\qquad \left. + c_3 n^{2pQ''(2d+1)-1} + c_4 n^{-2Q''ps}\right) \end{aligned} \tag{99}$$

where

$$r = \min[a - Q'(2d+1)p, 2(1-a), 2(pQ's - a)].$$

Analyzing the above inequality it is possible to say that algorithm (42) has tracking properties if

$$t \in \left(0, \frac{1}{3} - \frac{1}{3s}\right)$$

with use of the Fourier system and

$$t \in \left(0, \frac{1}{3} - \frac{5}{18}\frac{1}{s}\right)$$

with use of the Hermite system.

Of special interest is the fact that the maximum value of t with which the algorithm still has tracking properties does not depend on the dimension p.

XI. Additive Nonstationarity

Let us consider plants described by (49), where

$$\phi_n^*(x) = \phi(x) + \beta_n \tag{100}$$

β_n is an unknown sequence of numbers, ϕ is an unknown function. Observe that in this case

$$R_n(x) = f(x)\phi(x) + f(x)\beta_n. \tag{101}$$

Presently, assumptions (62) and (66) as well as (80) and (83) connected with smooth properties of function R_n $(n = 1, 2, \ldots)$ are replaced by assumptions concerning smooth properties of functions $f\phi$ and f. This fact significantly simplifies the convergence conditions described in Sections VII and VIII and facilitates the designing of a system that realizes algorithm (42).

The conditions for convergence of procedure (42) are very similar to those that are given in Tables I–IV concerning multiplicative nonstationarity. So, as examples of sequences β_n that satisfy convergence conditions, we may mention sequences specified in examples (a)–(e) of the previous section.

XII. Nonstationarity of the Type "Scale Change" and "Movable Argument"

Let us consider regressions described by (49), where

$$\phi_n^*(x) = \phi(\omega_n x) \tag{102}$$

or

$$\phi_n^*(x) = \phi(x - \lambda_n) \tag{103}$$

where

ω_n— unknown sequence of numbers;
λ_n— unknown sequence of vectors $\lambda_n = [\lambda_n^{(1)}, \ldots, \lambda_n^{(p)}]^T$,
ϕ— unknown function.

With reference to the nonstationarity (102) we use the expression "scale change," whereas the nonstationarity (103) is referred to as "movable argument." Of course

$$R_n(x) = f(x)\phi(\omega_n x) \tag{104}$$

for model (102) and

$$R_n(x) = f(x)\phi(x - \lambda_n) \tag{105}$$

for model (103). In Section X nonstationary factor α_n and in Section XI nonstationary component β_n were "separable" from function ϕ, significantly simplifying convergence conditions. The present situation is more complicated. Particularly, (80) and (83), taking form (85) and (86) for the Fourier and Hermite series, now are more complicated. That is why with reference to regressions (102) and (103) we will use the GRNN based only on the Parzen kernel. With the help of results of Section VII, in Tables VII and VIII are shown conditions ensuring the convergence of algorithm (42) tracking changing characteristics (102) and (103). Conditions (59), (60), (63), and (64) concern the selection of sequence h_n' and (62), (67), (66), and (70) concern the selection of sequence h_n'' (according to symbols in Fig. 7). Let us now assume that sequences ω_n and λ_n are of the following type.

i) $\omega_n = k_1 n^t, t > 0$
ii) $\lambda_n = k_2 n^t, t > 0$.

Employing Theorem 9 and arguments similar to those in Section X, we obtain the following expressions defining the speed of convergence of algorithm (42).

i) If $\omega_n = k_1 n^t, t > 0$ then

$$\begin{aligned} &P(|\hat{\phi}_n(x) - \phi_n^*(x)| > \varepsilon) \\ &\quad \le \left(\frac{\varepsilon + 2}{\varepsilon f(x)}\right)^2 \left(c_1 n^{2r(t - H't)}\right. \\ &\qquad \left. + c_2 n^{-r_1} + c_3 n^{-4H''} + c_4 n^{-(1 - H''p)}\right) \end{aligned} \tag{106}$$

TABLE VII
CONDITIONS FOR WEAK CONVERGENCE OF GRPNN BASED ON THE PARZEN KERNEL—NONSTATIONARITY OF THE TYPE "SCALE CHANGE" AND "MOVABLE ARGUMENT"

Condition	$\|\phi(x) - \varphi_n^*(x)\| \xrightarrow{n} 0$ in probability
(59)	$a_n h_n^{-p} \xrightarrow{n} 0$
(60)	$a_n h_n^{-2p} \longrightarrow 0$
(61)	$\begin{cases} a_n^{-1}\|\omega_{n+1} - \omega_n\| \xrightarrow{n} 0 \\ a_n^{-1}\|\lambda_{n+1} - \lambda_n\| \xrightarrow{n} 0 \end{cases}$
(62)	$\begin{cases} a_n^{-1} h_n^r (\omega_n^2 + 1) \xrightarrow{n} 0 \\ a_n^{-1} h_n^r \xrightarrow{n} 0 \end{cases}$
(67)	$n^{-2}\sum_{i=1}^{n} h_i^{-p} \xrightarrow{n} 0$

TABLE VIII
CONDITIONS FOR STRONG CONVERGENCE OF GRPNN BASED ON THE PARZEN KERNEL—NONSTATIONARITY OF THE TYPE "SCALE CHANGE" AND "MOVABLE ARGUMENT"

Condition	$\|\phi_n(x) - \varphi_n^*(x)\| \xrightarrow{n} 0$ with pr. 1
(63)	$\sum_{n=1}^{\infty} a_n^2 h_n^{-p} < \infty$
(64)	$\sum_{n=1}^{\infty} a_n^2 h_n^{-p} < \infty$
(65)	$\begin{cases} \sum_{n=1}^{\infty} a_n^{-1}(\omega_{n+1} - \omega_n)^2 < \infty \\ \sum_{n=1}^{\infty} a_n^{-1}(\lambda_{n+1} - \lambda_n)^2 < \infty \end{cases}$
(66)	$\begin{cases} \sum_{n=1}^{\infty} a_n^{-1} h_n^{2r}(\omega_n^4 + 1) < \infty \\ \sum_{n=1}^{\infty} a_n^{-1} h_n^{2r} < \infty \end{cases}$
(70)	$\sum_{n=1}^{\infty} n^{-2} h_n^{-p} < \infty$

where

$$r_1 = \min\begin{bmatrix} a - H'p, \\ 2(1-a-t), \quad 2(r - H' - rt - a) \end{bmatrix}.$$

Presently, algorithm (42) has tracking properties if

$$t + \frac{a}{r} < H' < \frac{a}{p},$$
$$a < 1 - t, \quad 0 < H'' < \frac{1}{p}$$

It means that parameter t should be contained within the range

$$t \in \left(0, 1 - \frac{1}{1 + \frac{1}{p} - \frac{1}{r}}\right)$$

where $r > p$. In a 1-Dl case

$$t \in \left(0, 1 - \frac{r}{2r-1}\right).$$

Along with an increase of parameter r specifying smooth properties of functions f and ϕ, the range in which parameter t is contained widens, not exceeding the interval $(0, 1/2)$. The increase of dimension p results in the decrease of the above mentioned range.

ii) If $\lambda_n = k_2 n^t, t > 0$ then

$$\begin{aligned} &P(|\hat{\phi}_n(x) - \phi_n^*(x)| > \varepsilon) \\ &\quad \le \left(\frac{\varepsilon + 2}{\varepsilon f(x)}\right)^2 \Big(c_1 n^{-2rH'} + c_2 n^{-r_1} \\ &\qquad + c_3 n^{-4H''} + c_4 n^{-(1-H''p)}\Big) \end{aligned} \tag{107}$$

where

$$r_1 = \min\begin{bmatrix} a - H'p, \\ 2(1-a-t), \quad 2(rH' - a) \end{bmatrix}.$$

In other words, algorithm (42) has tracking properties if for $r > p$

$$\frac{a}{r} < H' < \frac{a}{p},$$
$$a < 1 - t, \quad 0 < H'' < \frac{1}{p}$$

Assuming that $t \in (0, 1)$ it is possible to select such parameters H' and H'' in algorithm (42) that would satisfy the above inequalities.

XIII. SIMULATION EXAMPLES

In this section we present four simulations in order to test the GRNN studied in the paper.

1) Simulation 1 (Parzen Kernel): We consider the following nonstationary regression

$$\phi_n^*(x) = 10x^2 n^{0.1}$$

The GRNN based on the Parzen kernel has been applied with the following parameters

$$a = 0.8, H' = 0.3, \quad H'' = 0.4, \quad k_1' = k_2'' = 1.5.$$

The results are depicted in Fig. 7(a) and (b). Fig. 7(a) shows tracking the nonstationary regression with changing n in the point $x = 0.5$, whereas Fig. 7(b) displays comparison of a true regression and estimated by the GRNN for $n = 1000$.

2) Simulation 2 (Orthogonal Series Kernel): We consider the following nonstationary regression

$$\phi_n^*(x) = 2x^2 n^{0,2}$$

The GRNN based on the Fourier orthogonal series kernel has been applied with the following parameters:

$$a = 0.9, \quad Q' = 0.2, \quad Q'' = 0.2, \quad k_2' = k_2'' = 1.$$

The results are depicted in Fig. 8(a) and (b). Fig. 8(a) shows tracking the nonstationary regression with changing n in the point $x = 0.2$, whereas Fig. 8(b) displays comparison of a true regression and estimated by the GRNN for $n = 1000$.

3) Simulation 3 (Orthogonal Series Kernel): We consider the following nonstationary regression

$$\phi_n^*(x) = 2\cos(xn^{0.2})$$

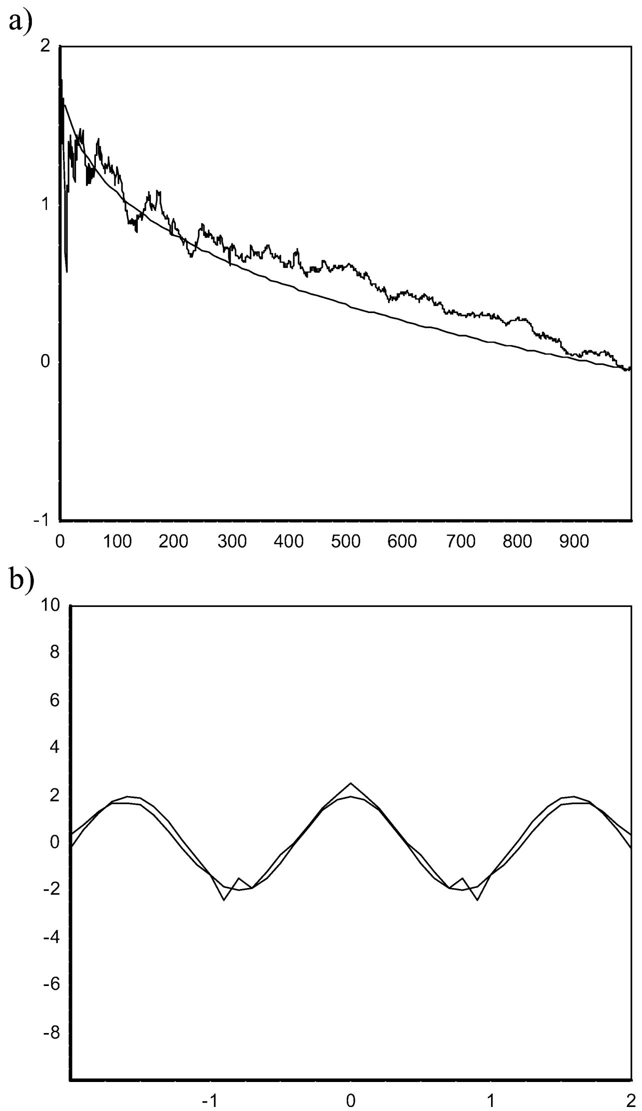

Fig. 9. Illustration of simulation 3 ($a = 0.8, Q' = 0.1, Q'' = 0.1, k_1' = k_2'' = 1$).

The GRNN based on the Fourier orthogonal series kernel has been applied with the following parameters:

$$a = 0.8, \quad Q' = 0.1, \quad Q'' = 0.1, \quad k_2' = k_2'' = 1.$$

The results are depicted in Fig. 9(a) and (b). Fig. 9(a) shows tracking the nonstationary regression with changing n in the point $x = 0.2$, whereas Fig. 9(b) displays comparison of a true regression and estimated by the GRNN for $n = 1000$.

4) Simulation 4 (Parzen Kernel): We consider the following nonstationary regression

$$\phi_n^*(x) = 10\cos(x - n^{0.1})$$

The GRNN based on the Parzen kernel has been applied with the following parameters

$$a = 0.8, \quad H' = 0.5, \quad H'' = 0.5, \quad k_1' = k_2'' = 5.$$

The results are depicted in Fig. 10(a) and (b). Fig. 10(a) shows tracking the nonstationary regression with changing n in the point $x = 0.2$, whereas Fig. 10(b) displays comparison of a true regression and estimated by the GRNN for $n = 1000$.

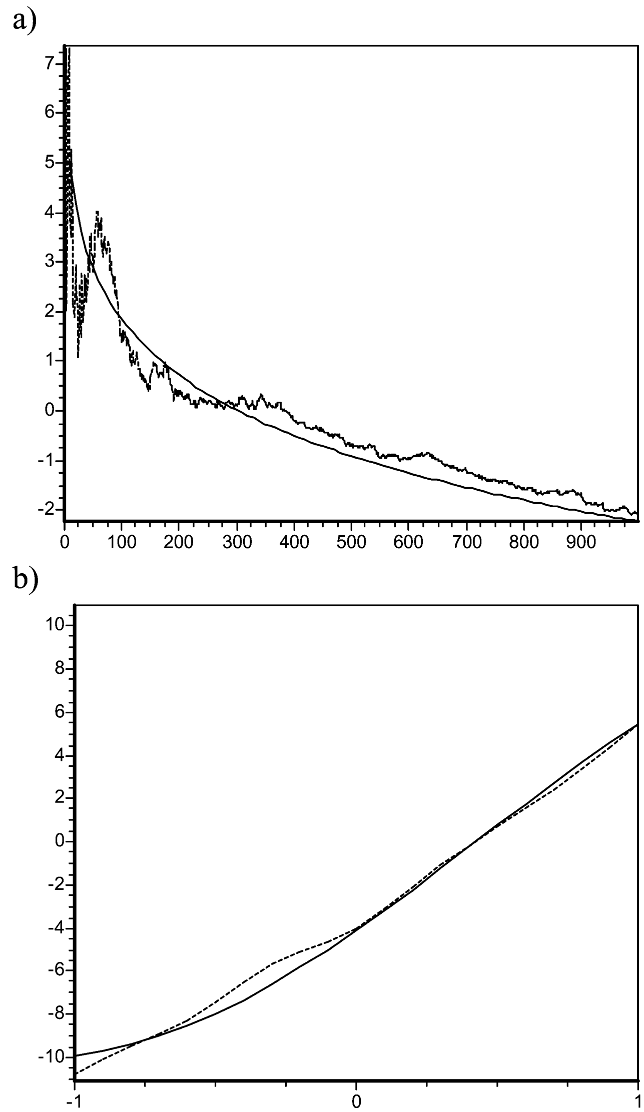

Fig. 10. Illustration of simulation 4 ($a = 0.8, H' = 0.5, H'' = 0.5, k_1' = k_2'' = 5$).

XIV. Summary and Discussions

In this paper, two types of the GRNN were studied: a) the GRNN based on the Parzen kernel, b) the GRNN based on the orthogonal series method.

Generally, we should say that the convergence of the GRNN b) is connected with both the convergence of the orthogonal series and with the speed of this convergence (assessment of the "tail" of the series). The problem of convergence of the orthogonal series is less complicated in the scalar case (e.g., Sansone [30]), but more complicated in the multidimensional case, that was investigated in more detail only for the Fourier series (e.g., Sjolin [31]). An additional problem is the examination of the orthogonal series speed of convergence because even in the 1-D case, appropriate results can be obtained with quite complicated assumptions regarding expanded functions which should be "smooth" enough. The selection of a particular series depends on both boundlessness or unboundedness of the input signal and on the dimensionality of the problem; if $A = R^p, p \geq 1$ then the proper series is the Hermite series. If $A \subset R^p, p \geq 1, \mu(A) < \infty$, then it is reasonable to employ the Fourier series because its properties are well known. In the scalar case, when the input signal is bounded, it is possible to use, among others, the Fourier, the Haar and the Legendre series. The later series is particularly interesting because the assumptions concerning smooth properties connected with the application of this series are much weaker than in the case of application of the Fourier series (Sansone [30]). The above considerations suggest that we should rather use algorithm a) that is based on the Parzen kernel. However, the orthogonal series method has a very desirable advantage: if, e.g., sequence α_n (or β_n) is of type n^t, then the maximum value of t at which the algorithm still has tracking properties does not depend on dimension p (contrary to the algorithm based on the Parzen kernel). Moreover, simulations that were conducted do not discredit the orthogonal series method, especially when it is used for the identification of plants with multiplicative and additive nonstationarity. A certain problem in the case of applying the algorithm based on the Parzen kernel may be the selection of function H meeting conditions (54)–(57). This problem will arise for some types of nonstationarity when a high degree of smooth properties of functions R_n $(n = 1, 2, \ldots)$ will be required. In this paper examples of function H were given for which parameter r connected with smooth properties of R_n $(n = 1, 2, \ldots)$ takes on the value of 2 and 4.

The examples of particular types of nonstationarity given in Sections X–XII do not exhaust all the possibilities of application of the GRNN. In particular, the results can be used for modeling of plants with nonstationarity that is a combination of cases discussed in these sections, i.e.,

$$\phi_n^*(x) = \alpha_n \phi(\omega_n x - \lambda_n) + \beta_n$$

In the future research we plan to describe nonstationary changes linguistically and modify flexible neurofuzzy systems [28] for their modeling.

APPENDIX

Proof of Theorem 1 and 2: Observe that

$$(\hat{R}_n(x) - R_n(x))^2 \leq 2(\hat{R}_n(x) - r_n(x))^2 + 2(r_n(x) - R_n(x))^2.$$

By making use of (24) and (26) we get

$$\begin{aligned} E[(\hat{R}_{n+1}(x) - r_{n+1}(x))^2 \,|\, X_1, Y_1, \ldots, X_n, Y_n] &= (1 - a_{n+1})^2 (\hat{R}_n(x) - r_n(x))^2 \\ &\quad + a_{n+1}^2 E[Y_{n+1} K_{n+1}(x, X_{n+1}) - r_{n+1}(x)]^2 \\ &\quad + (1 - a_{n+1})^2 (r_{n+1}(x) - r_n(x))^2 \\ &\quad + 2(1 - a_{n+1})^2 (r_n(x) - r_{n+1}(x))(\hat{R}_n(x) - r_n(x)). \end{aligned}$$

Of course,

$$E[Y_{n+1} K_{n+1}(x, X_{n+1}) - r_{n+1}(x)]^2 = \operatorname{var}[Y_{n+1} K_{n+1}(x, X_{n+1})]$$

Using $2ab \leq a^2 k + b^2 k^{-1}$, true for any $k > 0$ and setting $k = (a_{n+1} c_1)^{-1}, 0 < c_1 < 1$, we obtain

$$\begin{aligned} 2(r_{n+1}(x) - r_n(x))(\hat{R}_n(x) - r_n(x)) &\leq c_1 a_{n+1} (\hat{R}_n(x) - r_n(x))^2 \\ &\quad + c_1^{-1} a_{n+1}^{-1} (r_{n+1}(x) - r_n(x))^2. \end{aligned}$$

The following inequality is true:

$$(r_{n+1}(x) - r_n(x))^2 \leq 3(r_{n+1}(x) - R_{n+1}(x))^2 + 3(R_{n+1}(x) - R_n(x))^2 + 3(R_n(x) - r_n(x))^2.$$

Consequently

$$\begin{aligned} E[(\hat{R}_{n+1}(x) - r_{n+1}(x))]^2 [X_1, Y_1, \ldots, X_n, Y_n] &\leq (1 - a_{n+1}(1 - c_1))(\hat{R}_n(x) - r_n(x))^2 \\ &\quad + a_{n+1}^2 \operatorname{var}[Y_{n+1} K_{n+1}(x, X_{n+1})] + c_2 a_{n+1}^{-1} \\ &\quad \cdot (r_{n+1}(x) - R_{n+1}(x))^2 + c_3 a_{n+1}^{-1} (R_{n+1}(x) \\ &\quad - R_n(x))^2 + c_4 a_{n+1}^{-1} (R_n(x) - r_n(x))^2. \end{aligned} \tag{108}$$

We will now use the following lemma [3].

Lemma: Let W_n be a certain sequence of random variables. Let us introduce a sequence of functions $U_n = U_n(W_1, \ldots, W_n)$. Let a_n, s_n, and t_n be sequences of numbers. Let us assume that

i)
$$U_n \geq 0, \quad n = 1, 2, \ldots, \text{with pr. } 1$$

ii)
$$EU_1 < \infty$$

iii)
$$a_n \geq 0, \quad a_n \xrightarrow{n} 0, \quad \sum_{n=1}^{\infty} a_n = \infty$$

a) If
$$E[U_{n+1} \,|\, W_1, \ldots, W_n] \leq (1 - a_n) U_n + a_n s_n$$
where
$$s_n \xrightarrow{n} 0$$
then
$$EU_n \xrightarrow{n} 0.$$

b) If
$$E[U_{n+1} \,|\, W_1, \ldots, W_n] \leq (1 - a_n) U_n + t_n$$
where
$$\sum_{n=1}^{\infty} t_n < \infty$$
then
$$U_n \xrightarrow{n} 0 \text{ with pr. } 1.$$

Applying the above lemma to inequality (108), we obtain the conclusion of Theorems 1 and 2. ■

Proof of Theorem 3: We will apply the following lemma [6]:

Lemma: Let $p_1, p_2, \ldots$ be real numbers such that for $n \le n_0$

$$p_{n+1} \le (1 - c/n^{\omega})p_n + c'/n^t$$

where $0 < \omega < 1, c > 0, c' > 0, t$ real. Then

$$\lim_{n\to\infty} \sup n^{t-\omega} p_n \le c'/c.$$

The theorem is a consequence of direct application of that lemma to (108). Alternatively one can use Watanabe's [40] result (for $0 < \omega \le 1$). ■

Proof of Theorem 4: Conclusion of this theorem results immediately from the following inequality:

$$|\hat{\phi}_n(x) - \phi_n^*(x)| \le \frac{1}{|\hat{f}_n(x)|}(|\hat{R}_n(x) - R_n(x)| + |\phi_n^*(x)(\hat{f}_n(x) - f(x))|). \quad (109)$$

■

Proof of Theorem 5 and 6: Observe that

$$\begin{aligned} &\mathrm{var}\,[Y_n K_n(x, X_n)] \\ &\le EY_n^2 K_n^2(x, X_n) \\ &= h_n^{-2p} \int E\left[Y_n^2 \mid X_n = u\right] \\ &\qquad \times f_n(u) K^2\left(\frac{x-u}{h_n}\right) \mathrm{du} \\ &\le 2 \sup_u |K(u)| h_n^{-p} m_n'. \end{aligned} \quad (110)$$

Assessment (110) can be alternatively carried out in the following manner:

$$\begin{aligned} \mathrm{var}\,[Y_n K_n(x, X_n)] &\le EY_n^2 K_n^2(x, X_n) \\ &\le 2(\sup K(x))^2 h_n^{-2p} m_n''. \end{aligned}$$

Of course,

$$r_n(x) = h_n^{-p} EY_n K\left(\frac{x - X_n}{h_n}\right).$$

Denote

$$P_n(x) = |r_n(x) - R_n(x)|. \quad (111)$$

Expression (111) can be written as

$$P_n(x) = \left| \int K(u)[R_n(x - h_n u) - R_n(x)] \,\mathrm{du} \right|.$$

Expanding functions $R_n(x - h_n u)$ in the multidimensional Taylor's series, we obtain

$$\begin{aligned} P_n(x) = h_n^r \int \cdots \int \prod_{i=1}^{p} H\left(u^{(i)}\right) \\ \times \left[u^{(1)} \frac{\partial}{\partial x^{(1)}} + \cdots + u^{(p)} \frac{\partial}{\partial x^{(p)}}^T\right] \\ \times R_n(x - h_n u \theta_n) \,\mathrm{du}^{(1)} \ldots \mathrm{du}^{(p)} \end{aligned}$$

where $0 < \theta_n < 1$. As a result

$$|P_n(x)| \le \text{const.}\, h_n^r \sum_{i_1=1}^{p} \cdots \sum_{i_r=1}^{p} \sup_x \left| \frac{\partial^r}{\partial x^{(i_1)} \ldots \partial x^{(i_r)}} R_n(x) \right|$$

assuming that functions R_n have continuous partial derivatives up to the rth order. Now, Theorem 5 and 6 are consequences of Theorems 1 and 2. ■

Proof of Theorems 7 and 8: Observe that

$$\begin{aligned} &\mathrm{var}\,[Y_n K_n(x, X_n)] \le EY_n^2 K_n^2(x, X_n) \\ &\quad = \int E\left[Y_n^2 \mid X_n = u\right] f(u) \\ &\qquad \times \left(\sum_{|\underline{j}| \le q} \Psi \underline{j}(x) \Psi_{\underline{j}}(u)\right)^2 \mathrm{du} \le 2 \left(\sum_{j=0}^{q(n)} G_j^2\right)^p m_n'. \end{aligned} \quad (112)$$

Assessment (112) can be alternatively expressed as

$$\begin{aligned} \mathrm{var}\,[Y_n K_n(x, X_n)] &\le EY_n^2 K_n^2(x, X_n) \\ &\le \left(\sum_{j=0}^{q(n)} G_j^2\right)^{2p} 2m_n'' \end{aligned}$$

Observe that

$$r_n(x) = \sum_{|\underline{j}| \le q} b_{\underline{j}n} \Psi \underline{j}(x)$$

where

$$b_{\underline{j}n} = EY_n \Psi_{\underline{j}}(X_n).$$

Now both Theorems follow directly from Theorems 1 and 2.■

Proof of conditions (85) and (86): Let B_{jn} be the coefficient of expansion of function $t_n^l, l \ge 1, n = 1, 2, \ldots$, into multidimensional Hermite series. If $t_n^l \in L_2(R^p)$, then

$$b_{j_p,\ldots,j_p,n} \le \frac{|B_{j_1+1,n}, \ldots, B_{j_p+1,n}|}{j_1^{1/2} \ldots j_p^{1/2}}. \quad (113)$$

For $p = 1$ the above inequality was presented in [39] and its generalization for multidimensional case is straightforward. Assuming that (84) is true, under (113) and (9) with $G_j = \text{const.}\, j^{-(1/12)}$(see [37]), we obtain

$$\begin{aligned} S_n(x) &\le \left| \sum_{|\underline{j}| \le q} b_{jn} \Psi_{\underline{j}}(x) \right| \\ &\le \left\|t_n^l\right\|_{L_2} \left(\sum_{i=q+1}^{\infty} i^{-l+1/6}\right)^{\frac{p}{2}} \\ &\le \left\|t_n^l\right\|_{L_2} [q(n)]^{\frac{-p(l-5/6)}{2}}. \end{aligned}$$

Carrying out similar considerations for the multidimensional Fourier series (see [15]), we obtain (85) and (86). ■

Proof of Theorem 9: We will use ideas presented in work [13]. Let us consider the following occurrences:

$$A_1 : |\hat{f}_n(x) - f(x)| < f(x)\frac{\varepsilon}{\varepsilon+2}$$
$$A_2 : |\hat{R}_n(x) - R_n(x)| < f(x)\frac{\varepsilon}{\varepsilon+2}$$
$$A_3 : |\phi_n^*(x)(\hat{f}_n(x) - f(x))| < f(x)\frac{\varepsilon}{\varepsilon+2}.$$

Under (109), occurrences A_1, A_2, and A_3 imply the occurrence

$$B : |\hat{\phi}_n(x) - \phi_n^*(x)| < \varepsilon.$$

As a result

$$P(\bar{B}) \le P(\overline{A_1 \cap A_2 \cap A_3}) \le P(\bar{A}_1) + P(\bar{A}_2) + P(\bar{A}_3)$$

which concludes the proof of Theorem 9. ∎

ACKNOWLEDGMENT

The author would like to thank the Associate Editor and the reviewers for their helpful comments.

REFERENCES

[1] A. E. Albert and L. A. Gardner, *Stochastic Approximation and Nonlinear Regression*. Cambridge, MA: The MIT Press, 1967.

[2] G. Alexits, *Convergence Problems of Ortogonal Series*, Akademiai Kiado, Hungary: Budapest, 1961.

[3] E. M. Braverman and L. I. Rozonoer, "Convergence of random processes in machine learning theory," *Automat. Remote Contr.*, vol. 30, pp. 44–64, 1969.

[4] P. Burrascano, "Learning vector quantization for the probabilistic neural network," *IEEE Trans. Neural Networks*, vol. 2, pp. 458–461, 1991.

[5] L. Carleson, "On convergence and growth of partial sums of Fourier series," *Acta Math.*, vol. 116, pp. 135–137, 1966.

[6] K. L. Chung, "On a stochastic approximation method," *Ann. Math. Statist.*, vol. 25, pp. 463–483, 1954.

[7] H. I. Davies, "Strong consistency of a sequential estimator of a probability density function," *Bull. Math. Statist.*, vol. 15, pp. 49–53, 1973.

[8] L. P. Devroye, "On the pointwise and the integral convergence of recursive kernel estimates of probability densities," *Utilitias Math.*, vol. 15, pp. 113–128, 1979.

[9] L. P. Devroye, "On the almost everywhere convergence of nonparametric regression function estimates," *Ann. Stat.*, vol. 9, pp. 1310–1309, 1981.

[10] L. P. Devroye and L. Györfi, *Nonparametric Density Estimation: The L_1 View*, U.K.: Wiley, 1983.

[11] S. Dupač, "A dynamic stochastic approximation method," *Ann. Math. Stat.*, vol. 36, pp. 1695–1702, 1965.

[12] R. L. Eubank, *Spline Smoothing and Nonparametric Regression*. New York and Basel: Marcel Dekker, 1988.

[13] W. Greblicki and A. Krzyżak, "Asymptotic properties of kernel estimates of a regression function," *J. Statist. Planning Inference*, vol. 4, pp. 81–90, 1980.

[14] W. Greblicki, A. Krzyżak, and M. Pawlak, "Distribution-free pointwise consistency of kernel regression estimate," *Ann. Stat.*, vol. 12, pp. 1570–1575, 1984.

[15] W. Greblicki and M. Pawlak, "Classification using the Fourier series estimate of multivariate density function," *IEEE Trans. Syst., Man., Cybern.*, vol. 11, pp. 726–730, 1981.

[16] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*. New York, NY: Springer Verlag, 2002.

[17] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[18] K. Z. Mao, K.-C. Tan, and W. Ser, "Probabilistic neural-network structure determination for pattern classification," *IEEE Trans. Neural Networks*, vol. 11, pp. 1009–1016, 2000.

[19] E. A. Nadaraya, "On estimating regression," *Theory of Probab. Applicat.*, vol. 9, pp. 141–142, 1964.

[20] S. M. Nikolsky, *A Course of Mathematical Analysis*, Moscow: Mir, 1977.

[21] A. Pagan and A. Ullah, *Nonparametric Econometrics*. London: Cambridge Univ. Press., 1999.

[22] D. W. Patterson, *Artificial Neural Networks, Theory, and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[23] L. Rutkowski, "Sequential estimates of probability densities by orthogonal series and their application in pattern classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-10, no. 12, pp. 918–920, 1980.

[24] ——, "Sequential estimates of a regression function by orthogonal series with applications in discrimination," in *Lectures Notes in Statistics*, vol. 8. New York, 1981, pp. 236–244.

[25] ——, "Identification of MISO nonlinear regressions in the presence of a wide class of disturbances," *IEEE Trans. Inform. Theory*, vol. 37, pp. 214–216, 1991.

[26] ——, "Multiple Fourier series procedures for extraction of nonlinear regressions from noisy data," *IEEE Trans. Signal Processing*, vol. 41, pp. 3062–3065, 1993.

[27] ——, *New Soft Computing Techniques for System Modeling, Pattern Classification, and Image Processing*. New York: Springer Verlag, 2004.

[28] L. Rutkowski and K. Cpałka, "Flexible neuro-fuzzy systems," *IEEE Trans. Neural Networks*, vol. 14, pp. 554–574, 2003.

[29] L. Rutkowski, "Adaptive probabilistic neural networks for pattern classification in time-varying environment," *IEEE Trans. Neural Networks*, vol. 15, 2004.

[30] G. Sansone, *Orthogonal Functions*. New York: Interscience, 1959.

[31] P. Sjölin, "Convergence almost everywhere of certain singular integrals and multiple fourier series," *Ark. Math.*, vol. 9, pp. 65–90, 1971.

[32] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[33] D. F. Specht, "Probabilistic neural networks," *Neural Netw.*, vol. 3, pp. 109–118, 1990.

[34] ——, "Probabilistic neural networks and the polynomial adaline as complementary techniques for classification," *IEEE Trans. Neural Networks*, vol. 1, pp. 111–121, 1990.

[35] ——, "A general regression neural network," *IEEE Trans. Neural Networks*, vol. 2, pp. 568–576, 1991.

[36] ——, "Enhancements to the probabilistic neural networks," in *Proc. IEEE Int. Joint Conf., Neural Networks*, Baltimore, MD, 1992, pp. 761–768.

[37] G. Szegö, "Orthogonal polynomials," *Amer. Math. Soc. Coll. Publ.*, vol. 23, 1959.

[38] J. R. Thompson and R. A. Tapia, "Nonparametric function estimation, modeling and simulation," in *SIAM*, Philadelphia, PA, 1990.

[39] G. G. Walter, "Properties of Hermite series estimation of probability density," *Ann. Stat.*, vol. 5, pp. 1258–1264, 1977.

[40] W. Watanabe, "On convergence of asymptotically optimal discriminant functions for pattern classification problems," *Bull. Math. Statist.*, vol. 16, pp. 23–34, 1974.

[41] G. S. Watson, "Smooth regression analysis," *Sankhya Series A*, vol. 26, pp. 359–372, 1964.

[42] H. Yamato, "Sequential estimation of a continuous probability density function and the mode," *Bull. Math. Statist.*, vol. 14, pp. 1–12, 1971.

[43] P. V. Yee and S. Haykin, *Regularized Radial Basis Function Networks, Theory, and Applications*. New York: Wiley, 2001.

Leszek Rutkowski (M'92–SM'94) was born in Wrocław, Poland, in 1952. He received the M.Sc., Ph.D., and D.Sc. degrees in 1977, 1980, and 1986, respectively, all from the Technical University of Wrocław.

Since 1980, he has been with the Technical University of Częstochowa where he is currently a Professor and Chairman of the Computer Engineering Department. From 1987 to 1990, he held a visiting position with the School of Electrical and Computer Engineering, Oklahoma State University. His research interests include neural networks, fuzzy systems, computational intelligence, pattern recognition, and systems identification. He published more than 100 technical papers including 16 in various series IEEE TRANSACTIONS. He is the author of the books *New Soft Computing Techniques For System Modeling, Pattern Classification, and Image Processing* (New York: Springer), *Flexible Neuro-Fuzzy Systems* (Norwell, MA: Kluwer Academic), and *Adaptive Filters and Adaptive Signal Processing* (in Polish), and is the coauthor of two others (in Polish) *Neural Networks, Genetic Algorithms and Fuzzy Systems* and *Neural Networks for Image Compression*.

Dr. Rutkowski is also President and Founder of the Polish Neural Networks Society. He organized and served as General Chair of the Polish Neural Networks Conferences held in 1996, 1997, 1999, 2000, and 2002. He is an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS.