On the Weak Convergence of the Orthogonal Series-Type Kernel Regresion Neural Networks in a Non-stationary Environment

Meng Joo Er[1] and Piotr Duda[2]

[1] Nanyang Technological University,
School of Electrical and Electronic Engineering, Singapore
[2] Department of Computer Engineering,
Czestochowa University of Technology, Czestochowa, Poland
emjer@ntu.edu.sg, pduda@kik.pcz.pl

Abstract. In the paper general regression neural networks, based on the orthogonal series-type kernel, is studied. Convergence in probability is proved assuming non-stationary noise. The performance is investigated using syntetic data.

1 Introduction

Let $X_1, \ldots, X_n$ be a sequence of independent random variables with a common desity function f. Consider the following model

$$Y_i = \phi(X_i) + Z_i, \qquad i = 1, \ldots, n, \tag{1}$$

where Z_i are random variables such that

$$E(Z_i) = 0, \qquad EZ_i^2 = d_i, \qquad i = 1, \ldots, n, \tag{2}$$

and $\phi(\cdot)$ is an unknown function.

$$f(x) \sim \sum_{j=0}^{\infty} a_j g_j(x), \tag{3}$$

where

$$a_j = \int_A f(x) g_j(x) dx = E g_j(X_j). \tag{4}$$

and $\{g_j(\cdot)\}$, $j = 0, 1, 2, \ldots$ is a complete orthonormal set defined on $A \subset R^p$. Then the estimator of density $f(\cdot)$ takes the form

$$\hat{f}_n(x) = \sum_{j=0}^{N(n)} \hat{a}_j g_j(x), \tag{5}$$

R. Wyrzykowski et al. (Eds.): PPAM 2011, Part I, LNCS 7203, pp. 443–450, 2012.

where $N(n) \xrightarrow{n} \infty$ and

$$\hat{a}_j = \frac{1}{n} \sum_{k=0}^{n} g_j(X_k) \tag{6}$$

Let us define

$$R(x) = f(x)\phi(x). \tag{7}$$

We assume that function $R(\cdot)$ has the representation

$$R(x) \sim \sum_{j=0}^{\infty} b_j g_j(x), \tag{8}$$

where

$$b_j = \int_A \phi(x) f(x) g_j(x) dx = E(Y_k g_j(X_k)) \tag{9}$$

We estimate function $R(\cdot)$ using

$$\hat{R}_n(x) = \sum_{j=0}^{M(n)} \hat{b}_j g_j(x), \tag{10}$$

where $M(n) \xrightarrow{n} \infty$ and

$$\hat{b}_j = \frac{1}{n} \sum_{k=0}^{n} Y_k g_j(X_k). \tag{11}$$

Then the estimator of the regression function is of the following form

$$\hat{\phi}_n(x) = \frac{\hat{R}_n(x)}{\hat{f}_n(x)} = \frac{\sum_{i=1}^{n} \sum_{j=0}^{M(n)} Y_i g_j(X_i) g_j(x)}{\sum_{i=1}^{n} \sum_{j=0}^{N(n)} g_j(X_i) g_j(x)} \tag{12}$$

It should be noted estimate (12) can be presented in the form of general regression neural networks [35] with appropriately selected kernels $K_n^1(x,u) = \sum_{j=0}^{N(n)} g_j(x) g_j(u)$ and $K_n^2(x,u) = \sum_{j=0}^{M(n)} g_j(x) g_j(u)$. In literature nonparametric estiames have been widely studied both in stationary (see e.g. [3], [5], [6],[8], [10], [12] - [15], [21] - [23], [26] - [29]) and time-varying case (see [7], [16] -[20], [24], [25]). In this paper it will be shown that procedure (12) is applicable even if variance of noise diverges to infinity. Block diagram of general regression neural network is shown in Fig. 1.

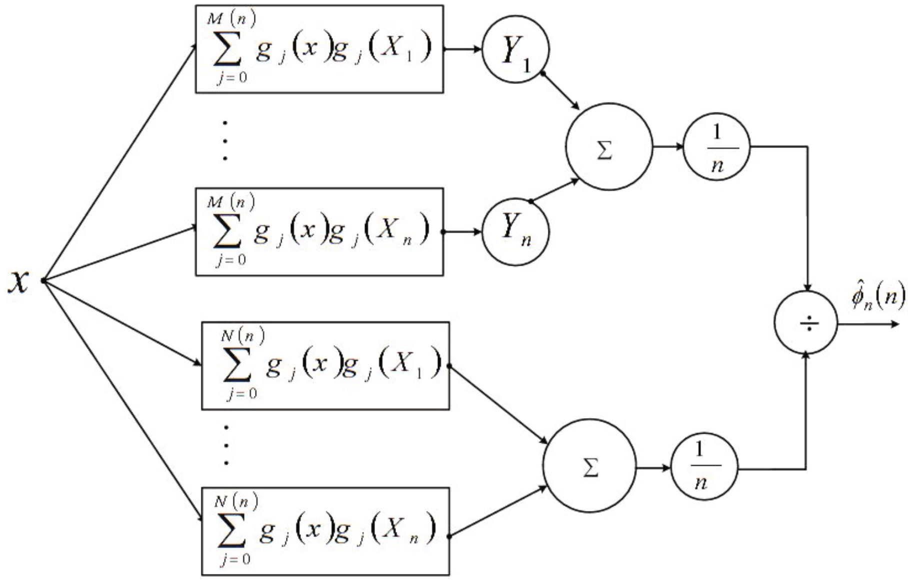

Fig. 1. General regression neural network

2 Main Result

Let us assume that

$$\max_x |g_j| < G_j. \tag{13}$$

Let us denote

$$s_i = d_i + \int_A \phi^2(u) f(u) du \tag{14}$$

Theorem 1. *If $s_i < \infty$, for all $i \geq 1$, and the following conditions hold*

$$\frac{1}{n^2} \left(\sum_{j=0}^{M(n)} G_j^2 \right)^2 \sum_{i=1}^{n} s_i \xrightarrow{n} 0, \quad M(n) \xrightarrow{n} \infty \tag{15}$$

$$\frac{1}{n} \left(\sum_{j=0}^{N(n)} G_j^2 \right)^2 \xrightarrow{n} 0, \quad N(n) \xrightarrow{n} \infty \tag{16}$$

then

$$\hat{\Phi}_n(x) \xrightarrow{n} \Phi(x) \qquad \text{in probability,} \tag{17}$$

at every point $x \in A$ at which series (3) and (8) converge to $f(x)$ and $R(x)$, respectively.

Proof. It is sufficient to show that:

$$E[\hat{R}_n(x) - R(x)]^2 \xrightarrow{n} 0 \tag{18}$$

$$E[\hat{f}_n(x) - f(x)]^2 \xrightarrow{n} 0, \tag{19}$$

in probability, at every point $x \in A$, at which series (3) and (8) are convergent. Observe that

$$E(\hat{R}_n(x) - R(x))^2 \leq \sum_{j=0}^{M(n)} G_j^2 \sum_{j=0}^{M(n)} E(\hat{b}_j - b_j)^2 + (\sum_{j=0}^{M(n)} b_j g_j(x) - R(x))^2. \tag{20}$$

One can see that the mean square error $E(\hat{b_j} - b_j)^2$ is bounded by

$$E(\hat{b}_j - b_j)^2 \leq \frac{G_j^2}{n^2} \sum_{i=1}^{n} (\int_A \phi^2(u) f(u) du + d_i). \tag{21}$$

Then

$$E[\hat{R}_n(x) - R(x)]^2 \leq \frac{1}{n^2} (\sum_{j=0}^{M(n)} G_j^2)^2 \sum_{i=1}^{n} (\int_A \phi^2(u) f(u) du + d_i) + (\sum_{j=0}^{M(n)} b_j g_j(x) - R(x))^2. \tag{22}$$

In view of assumption (15), convergence (18) is established. Convergence (19) can be proved in a similar way. This concludes the proof.

Remark 1. Conditions for convergence of series (3) and (8) can be found in [1],[33],[38].

Example. Let assume that

$$M(n) = [c_1 n^{q_M}] \quad N(n) = [c_2 n^{q_N}] \quad d_n = c_3 n^{\alpha} \quad G_j = c_4 j^d, \tag{23}$$

where q_M, q_N and α are positive numbers. It is easily seen that if

$$4dq_M + 2q_M + \alpha < 1, \qquad 4dq_N + 2q_N < 1, \tag{24}$$

then Theorem 1 holds. It should be noted that $d = -\frac{1}{12}$ for the Hermite sytem, $d = -\frac{1}{4}$ for the Laguerre system, $d = 0$ for the Fourier system, $d = \frac{1}{2}$ for the Legendre and Haar systems (see [33]).

3 Experimental Results

For computer simulations we use synthetic data. Distribution of random variables X_i is uniform on the interval $[0; \pi]$, for $i = 1, \ldots, n$. Consider the following model

$$\phi(x) = \exp(\sin(2x)), \tag{25}$$

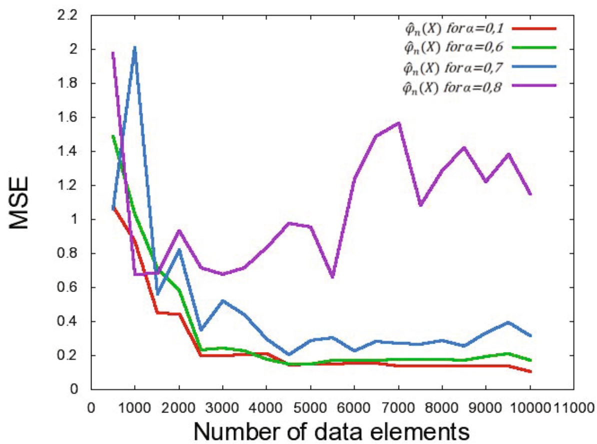

Fig. 2. The MSE as a function of n

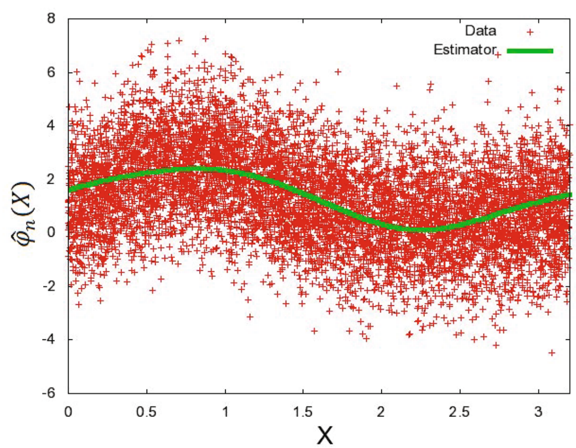

Fig. 3. Training set and obtained estimator

with Z_i which are realizations of random variables $N(0, d_i)$, $d_i = i^\alpha$, $\alpha > 0$. All constants (c_1, c_2, c_3) in (23) are equal to 1. Parameters q_M and q_N are both equal to $0, 25$. The Hermite orthonormal system is chosen to perform calculations. Number of data set is taken from the interval $[500; 10000]$ and parameter α is tested in the interval $[\frac{1}{10}, \frac{12}{10}]$.

Figure 2 shows how the MSE (Mean Square Error) changes with the number of data elements n for different values of parameter α. For parameter $\alpha \in [0, 1; 0, 7]$ we can see that, when n goes to infinity, the MSE goes to 0. For $\alpha = 0, 8$ this trend is not maintained. Moreover, for $\alpha = 0, 8$, value of the MSE is much bigger than for lower values of parameter α. Experimental results show that for higher values of α the MSE is growing. For $\alpha = 1, 2$ and $n = 10^4$, the MSE is equal to 7,37.

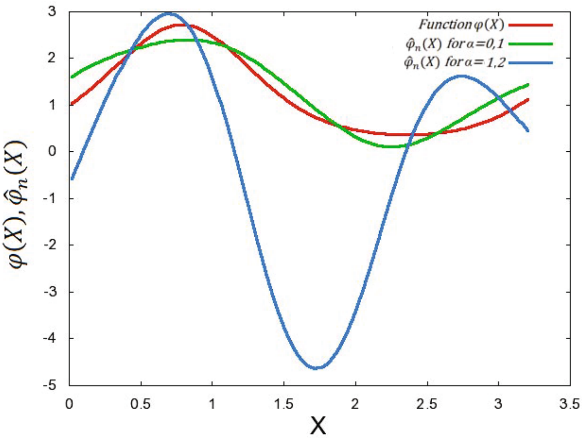

Fig. 4. Function $\phi(\cdot)$ and its estimators for different values of parameter α

In Figure 3 input data and the result of estimation for $n = 10^4$ and $\alpha = 0,1$ is indicated. As we can see the estimator found in the appropriate manner center of data and maintained its trend.

Figure 3 shows the course of the function given by (25) and estimators obtained for $n = 10^4$, with parameters α equal to $0,1$ and $1,2$.

4 Conclusions

In this paper we studied general regression neural networks, based on the orthogonal series-type kernel. We proved convergence in probability assuming nonstationary noise. In future works alternative methods, based on neural networks [2], [4], [11] and neuro-fuzzy structures [9], [30] - [32], [34], [36], [37], will be adopted to handle nonstationary noise.

Acknowledgments. The paper was prepared under project operated within the Foundation for Polish Science Team Programme co-financed by the EU European Regional Development Fund, Operational Program Innovative Economy 2007-2013, and also supported by the National Science Center NCN.

References

1. Alexits, G.: Convergence Problems of Orthogonal Series, Budapest, Academia and Kiado, pp. 261–264 (1961)
2. Bilski, J., Rutkowski, L.: A fast training algorithm for neural networks. IEEE Transactions on Circuits and Systems II 45, 749–753 (1998)

3. Cacoullos, P.: Estimation of a multivariate density. Annals of the Institute of Statistical Mathematics 18, 179–190 (1965)
4. Cierniak, R., Rutkowski, L.: On image compression by competitive neural networks and optimal linear predictors. Signal Processing: Image Communication - a Eurasip Journal 15(6), 559–565 (2000)
5. Gałkowski, T., Rutkowski, L.: Nonparametric recovery of multivariate functions with applications to system identification. Proceedings of the IEEE 73, 942–943 (1985)
6. Gałkowski, T., Rutkowski, L.: Nonparametric fitting of multivariable functions. IEEE Transactions on Automatic Control AC-31, 785–787 (1986)
7. Greblicki, W., Rutkowska, D., Rutkowski, L.: An orthogonal series estimate of time-varying regression. Annals of the Institute of Statistical Mathematics 35, Part A, 147–160 (1983)
8. Greblicki, W., Rutkowski, L.: Density-free Bayes risk consistency of nonparametric pattern recognition procedures. Proceedings of the IEEE 69(4), 482–483 (1981)
9. Nowicki, R.: Rough Sets in the Neuro-Fuzzy Architectures Based on Monotonic Fuzzy Implications. In: Rutkowski, L., Siekmann, J.H., Tadeusiewicz, R., Zadeh, L.A. (eds.) ICAISC 2004. LNCS (LNAI), vol. 3070, pp. 510–517. Springer, Heidelberg (2004)
10. Parzen, E.: On estimation of a probability density function and mode. Analysis of Mathematical Statistics 33(3), 1065–1076 (1962)
11. Patan, K., Patan, M.: Optimal training strategies for locally recurrent neural networks. Journal of Artificial Intelligence and Soft Computing Research 1(2), 103–114 (2011)
12. Rutkowski, L.: Sequential estimates of probability densities by orthogonal series and their application in pattern classification. IEEE Transactions on Systems, Man, and Cybernetics SMC-10(12), 918–920 (1980)
13. Rutkowski, L.: Sequential estimates of a regression function by orthogonal series with applications in discrimination, New York, Heidelberg, Berlin. Lectures Notes in Statistics, vol. 8, pp. 236–244 (1981)
14. Rutkowski, L.: On system identification by nonparametric function fitting. IEEE Transactions on Automatic Control AC-27, 225–227 (1982)
15. Rutkowski, L.: Orthogonal series estimates of a regression function with applications in system identification. In: Probability and Statistical Inference, pp. 343–347. D. Reidel Publishing Company, Dordrecht (1982)
16. Rutkowski, L.: On Bayes risk consistent pattern recognition procedures in a quasi-stationary environment. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-4(1), 84–87 (1982)
17. Rutkowski, L.: On-line identification of time-varying systems by nonparametric techniques. IEEE Transactions on Automatic Control AC-27, 228–230 (1982)
18. Rutkowski, L.: On nonparametric identification with prediction of time-varying systems. IEEE Transactions on Automatic Control AC-29, 58–60 (1984)
19. Rutkowski, L.: Nonparametric identification of quasi-stationary systems. Systems and Control Letters 6, 33–35 (1985)
20. Rutkowski, L.: The real-time identification of time-varying systems by nonparametric algorithms based on the Parzen kernels. International Journal of Systems Science 16, 1123–1130 (1985)
21. Rutkowski, L.: A general approach for nonparametric fitting of functions and their derivatives with applications to linear circuits identification. IEEE Transactions Circuits Systems CAS-33, 812–818 (1986)

22. Rutkowski, L.: Sequential pattern recognition procedures derived from multiple Fourier series. Pattern Recognition Letters 8, 213–216 (1988)
23. Rutkowski, L.: Nonparametric procedures for identification and control of linear dynamic systems. In: Proceedings of 1988 American Control Conference, June 15-17, pp. 1325–1326 (1988)
24. Rutkowski, L.: An application of multiple Fourier series to identification of multivariable nonstationary systems. International Journal of Systems Science 20(10), 1993–2002 (1989)
25. Rutkowski, L.: Nonparametric learning algorithms in the time-varying environments. Signal Processing 18, 129–137 (1989)
26. Rutkowski, L., Rafajłowicz, E.: On global rate of convergence of some nonparametric identification procedures. IEEE Transaction on Automatic Control AC-34(10), 1089–1091 (1989)
27. Rutkowski, L.: Identification of MISO nonlinear regressions in the presence of a wide class of disturbances. IEEE Transactions on Information Theory IT-37, 214–216 (1991)
28. Rutkowski, L.: Multiple Fourier series procedures for extraction of nonlinear regressions from noisy data. IEEE Transactions on Signal Processing 41(10), 3062–3065 (1993)
29. Rutkowski, L., Gałkowski, T.: On pattern classification and system identification by probabilistic neural networks. Applied Mathematics and Computer Science 4(3), 413–422 (1994)
30. Rutkowski, L.: A New Method for System Modelling and Pattern Classification. Bulletin of the Polish Academy of Sciences 52(1), 11–24 (2004)
31. Rutkowski, L., Cpałka, K.: A general approach to neuro - fuzzy systems. In: Proceedings of the 10th IEEE International Conference on Fuzzy Systems, Melbourne, December 2-5, vol. 3, pp. 1428–1431 (2001)
32. Rutkowski, L., Cpałka, K.: A neuro-fuzzy controller with a compromise fuzzy reasoning. Control and Cybernetics 31(2), 297–308 (2002)
33. Sansone, G.: Orthogonal functions, vol. 9. Interscience Publishers In., New York (1959)
34. Scherer, R.: Boosting Ensemble of Relational Neuro-fuzzy Systems. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Żurada, J.M. (eds.) ICAISC 2006. LNCS (LNAI), vol. 4029, pp. 306–313. Springer, Heidelberg (2006)
35. Specht, D.F.: Probabilistic neural networks. Neural Networks 3, 109–118 (1990)
36. Starczewski, J., Rutkowski, L.: Interval type 2 neuro-fuzzy systems based on interval consequents. In: Rutkowski, L., Kacprzyk, J. (eds.) Neural Networks and Soft Computing, pp. 570–577. Physica-Verlag, Springer-Verlag Company, Heidelberg, New York (2003)
37. Starczewski, J., Rutkowski, L.: Connectionist Structures of Type 2 Fuzzy Inference Systems. In: Wyrzykowski, R., Dongarra, J., Paprzycki, M., Waśniewski, J. (eds.) PPAM 2001. LNCS, vol. 2328, pp. 634–642. Springer, Heidelberg (2002)
38. Zygmund, A.: Trigonometric Series. Cambridge University Press, Cambridge (1959)