Trigonometric Neural Networks $L_p, p<1$ Approximation

Eman Samir Bhaya, and Zaineb Hussain Abd Al-sadaa
University of Babylon, College of education for pure sciences, Mathematics Department
e-mail: emanbhaya@itnet.uobabylon.edu.iq

Abstract. Many researchers studied the approximation by neural networks approximation. However only using first or second modulus, that is with low speed approaching zero. Here we define a neural network. Then we use it to approximate functions from $L_p$-quasi normed spaces we prove upper and lower bounds trigonometric neural networks estimations using the modulus of smoothness of order $k$.

1. Introduction

In the recent years, the approximation using neural networks have many good applications. Many results on the density of the FNNs on the space of continuous functions or on the space of integrable functions are introduced see for example [7], [11], [10], [18], [6] and [4]. In these references we can read the result that for any continuous function $f$ of multivariable defined on a compact subset of $\mathbb{R}^n$ we can find a FNN of one hidden layer as best approximation for $f$ of the form

$$N(x) = \sum_{i=1}^{d} c_i \sigma_i(\sum_{j=1}^{d} w_{ij}x_j + \theta_i), x \in \mathbb{R}^d, d \geq 1,$$

where $i = 1, 2, ..., d$, $\theta_i$ is a real threshold, $w_i = (w_{i1}, w_{i2}, ..., w_{is})^T \in \mathbb{R}^d$ is the weight that connect the neuron of index $i$ of the hidden layer and the neuron that we input it., $c_i$ is a real constant that connect the weight and the neuron that it output. and $\sigma_i(.)$ is the activation function of the neural network. In the above formula the $d$ is very important: it draw the topology of the hidden layer of the neural network. In many of the approximation studied of the neural network is very difficult to specify the number $d$, and it is sufficient to say it is existing and large. [8]

We can see many kinds of forward neural networks; all these kinds are different. They are same by: Its input nodes and the links connecting them. We input these nodes then we make processing on them to get the outputs.

The approximation by neural network attracted attentions, especially in the recent years. See for example [1], [13], [16], [14], [5], [3] and [17]. In all studies above the authors study the degree of best neural approximation using modules of smoothness of order 1. The estimation in the above references cannot characterize the ability of the neural network in general. So in this section we will study the order of essential approximation on a special class of neural using trigonometric hidden layer in terms of the $k^{th}$ order modulus of smoothness. We shall use upper and lower bound estimation of neural approximation. After upper and lower
bounded, estimation we can write the order of essential neural approximation. We want to mention that we will use the multivariate function for approximation, and using $k^{th}$ order modulus of smoothness for measuring the approximation order. We clear that there is a relationship between the speed of approximation and the number of hidden units.

2. Some Definitions and Notation

If $N$ is the naturals, and $R$ is the reals. Let $N_0$ be the naturals with the zero number and 0 is the zero vector, $1_t = (0, 0, \ldots, 1^{th}, 0, \ldots, 0) \in N_0^d$. Let $|r| = \sum_{i=1}^d |r_i|$ for $r = (r_1, r_2, \ldots, r_d) \in N_0^d$, $\|t\| = (\sum_{i=1}^d t_i^2)^{\frac{1}{2}}$ for $t = (t_1, t_2, \ldots, t_d) \in R^d$, and $rt = \sum_{i=1}^d r_i t_i$.

Let $f \in L_2^{\pi}$, $0 < p < 1$. Write $C_2\pi$ the space of the continuous functions with $2\pi$ periodic with respect to the variable in $R^d$. If $f \in L_2^{\pi}$, $0 < p < 1$, its quasi norm.

Define the symmetric difference of degree $r$ for the function $f$ as

$$\Delta_h^{(r)} f(x) = \sum_{i=0}^{(r)} (-1)^i f(x + \frac{r}{2} - i)h).$$

Using $\Delta_h f(\cdot)$, we define the modulus of smoothness of order $r$ as:

$$\omega_r(f, t)_p = \sup_{h \subset \mathbb{R}} \|h\|_t \|\Delta_h^{(r)} f(\cdot)\|_p. \quad (2.1)$$

where

$$\|f\|_p = \left(\int_{-\pi}^{\pi} |f(x)|^p \right)^{1/p}.$$

We say that the function $f$ belongs to Lipschitz space of order greater than $r, r \in N_+$. Write $f \in Lip(\alpha)_r$, if $\omega_r(f, t)_p = O(t^\alpha)$, with an $\alpha \in (0, r]$.

Modulus of smoothness is a measurement of smoothness. The modulus of smoothness of many variables is an improvement of the modulus of smoothness of one variable. Let us list some properties of the modulus of smoothness. For $f \in L_2^{\pi}$, $0 < p < 1$, we have

(1) $\lim_{\delta \to 0} \omega_r(f, \delta, f)_p = 0$.

(2) $\omega_r(f, \delta, f)_p$ is nondecreasing function of $\delta$.

(3) $\omega_r(f, \delta, f)_p \leq c \lambda_r \omega_r(f, \delta, f)_p$, for $\lambda \geq 1$.

We denote by $f \ast g(x) = \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} f(t) g(x-t) dt$ the convolution of $f$ and $g$, and by $f^r(r) = \langle f, e^{-it} \rangle$ the Fourier transformation of function $f$, where $\langle f, g \rangle = \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} f(t) g(t) dt$ is inner product of $f$ and $g$. The definition of the $r$-th K-functional of $f \in L_2^{\pi}$, $0 < p < 1$, and $\delta > 0, r \in N$, it mean
\[ K_r(f, \delta^r)_p = \inf_{D^\beta g \in \ell_p} \left\{ \| f - g \|_p + \delta^r \sup_{|\beta| = r} \| D^\beta g \|_p \right\}, \]  

(2.2)

where \(|\beta| = \beta_1 + \beta_2 + \cdots + \beta_d, \beta = (\beta_1, \beta_2, \ldots, \beta_d) \in N_0^d\), and \(D^\beta = \frac{\partial^{|\beta|}}{\partial x_1^{\beta_1} \cdots \partial x_d^{\beta_d}}\)

is the operator of derivative. The K-functional operator was defined by K-Peetre in [15]. Then it developed by Johmen and Scherer in [12] and in [9] by Ditzian and Totik. The K-functional operator used to measure the distance between the neural liner space and the approximations space. One of the famous results for the K-functionless is its equivalence with the modulus of smoothness define in (2.2), it mean there are constants \(C_1\) and \(C_2\), satisfy

\[ C_1 \omega_r(f, \delta)_p \leq K_r(f, \delta^r)_p \leq C_2 \omega_r(f, \delta)_p. \]  

(2.3)

Now let us introduce some notations from [16]. We have \(\lambda \in \mathbb{N}\) and \(f_i \in \ell_p^{2\pi}, 0 < p < 1\).

\[ p = (p_1, p_2, \ldots, p_d) \in N_0^d, q = (q_1, q_2, \ldots, q_d) \in N_0^d \]

\[ B_{\lambda} = \left( \frac{2}{\lambda + 2} \right)^d, b_{\lambda, r} = \prod_{i=1}^{d} \sin \frac{\eta_i + 1}{\lambda + 2} \pi. \]

In our article we will use the notation \(c(v_1, v_2)\) to denote such absolute crostatas which are may differ on different occurrences even in the same line, and depending on \(v_1\) and \(v_2\).

3. The Main Results

This section consists of the main results of this article.

Theorem 3.1. For \(f_i \in \ell_p^{2\pi}, 0 \leq p \leq 1\), we have

\[ \left\| E\mathcal{N}(f_i) - f_i \right\|_p \leq c(p, k)W_k(f_i, \frac{1}{\lambda + 2})p \]

Proof.

Suppose \(r = (r_1, r_2, \ldots, r_d) \in N_0^d, \lambda\) is a natural number.

The Fejer - korovkin kernel \(k_{0, \lambda}\) of dimension \(d\) is defined by

\[ K_{\lambda}(t) = B_{\lambda} \left| \sum_{0 \leq r \leq \lambda} b_{\lambda, r} \right|^2, \]

where

\[ b_{\lambda, r} = \prod_{i=1}^{d} \sin \frac{\eta_i + 1}{\lambda + 2} \pi, B_{\lambda} = (\sum_{0 \leq r \leq \lambda} (b_{\lambda, r})^2)^{-1}. \]

Then

\[ B_{\lambda} = (\sum_{0 \leq r \leq \lambda} (\prod_{i=1}^{d} \sin \frac{\eta_i + 1}{\lambda + 2} \pi)^2)^{-1} = \left( \frac{2}{\lambda + 2} \right)^d, \]

\(2\)}
and

\[ K_j(t) = B_\lambda \sum_{0 \leq \ell \leq \lambda} b_{\lambda, \ell} e^{iq\ell t} |^2 = 1 + 2B_\lambda \sum_{p \neq q \in \mathbb{N}_0^d} b_{\lambda, p} b_{\lambda, q} \cos(p - q)t. \]

We define the operator

\[ EN_j(f_i) = \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} \left( \sum_{j=0}^{r} f_i(x + \frac{r}{2}) \right) K_j(t) dt \]

\[ + \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} \left( \sum_{j=0}^{r} f_i(x + \frac{r}{2}) \right) K_j(t) dt + \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} C_r(-1)^\frac{r}{2} f_i(x) K_j(t) dt, \]

where \( c_r = \frac{r!((\frac{r}{2})!)^2}{(\frac{r}{2})^2} \).

\[ K_j(0) = 1, K_j(r) = B_\lambda \sum_{\delta p=q \in \mathbb{N}_0^d} b_{\lambda, p} b_{\lambda, q} \]

and \( K_j(1) = \cos \frac{\pi}{\lambda + 2} \) [16].

Using (1), (2), and (3) to get

\[ \| EN_j(f_i) - f_i \| \]

\[ = \| \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} \left( \sum_{j=0}^{r} f_i(x + \frac{r}{2}) \right) K_j(t) dt \]

\[ + \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} \left( \sum_{j=0}^{r} f_i(x + \frac{r}{2}) \right) K_j(t) dt + \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} C_r(-1)^\frac{r}{2} f_i(x) K_j(t) dt - f_i \|_p \]

\[ \leq \frac{c(p)}{(2\pi)^d} \int_{-\pi}^{\pi} K_j(t) \omega_k(f_i, \|t\|)_p \]

\[ \leq \frac{c(p)}{(2\pi)^d} \int_{-\pi}^{\pi} K_j(t) \omega_k(f_i, \|t\|)_p \]

\[ \leq \frac{c(p)}{(2\pi)^d} \omega_k(f_i, \|t\|)_p \int_{-\pi}^{\pi} K_j(t) (\delta^{-1} \|t\|)^k dt \]

\[ \leq c(p) \omega_k(f_i, \|t\|)_p \left( \frac{1}{(2\pi)^d} \sum_{j=1}^{\pi} \int_{-\pi}^{\pi} t_j^{2k} K_j(t) dt \right)^\frac{1}{2}. \]

Since

\[ t \leq \pi \sin \frac{t}{2}, \ 0 \leq t \leq \pi, \ \pi \sin \frac{t}{2} \leq t, -\pi \leq t \leq 0, \]

so
\[ t^{2k} \leq \pi^{2k} \left( \sin \frac{t}{2} \right)^{2k}, \quad -\pi \leq t \leq \pi. \]

Therefore, \[ t^{2k} \leq \pi^{2k} \left( \sin \frac{\pi}{2} \right)^{2k}. \]

Consequently,

\[
\frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} t_j^{2k} K_\lambda(t) \, dt \leq \frac{1}{(2\pi)^d} \pi^{2k} \int_{-\pi}^{\pi} (\sin t_j)^{2k} K_\lambda(t) \, dt = \pi^k (1 - \cos \frac{\pi}{\lambda+2})^k.
\]

If we take \( \delta = \frac{1}{\lambda+2} \)

and since \( 1 - \cos \frac{\pi}{\lambda+2} \) is bounded set so

\[
\sum_{i=1}^{d} (1 - \cos \frac{\pi}{\lambda+2})^k = c \left( \sum_{i=1}^{d} (1 - \cos \frac{\pi}{\lambda+2})^k \right) \leq c \left( \frac{\pi}{\lambda+2} \right)^{2k}, \quad c \text{ a positive constant.}
\]

Thus take \( \delta = \frac{1}{\lambda+2} \)

\[
\| EN_\lambda(f_i) - f_i \|_p \leq c(p) \left( \frac{1}{\lambda+2} \right)^k \left( c(p) \frac{1}{\lambda+2} \right)^{1/2} \omega_k(f_i, \frac{1}{\lambda+2})_p \\
\leq c(p) \left( (\lambda + 2) \right)^k \left( \frac{\pi^k}{(\lambda+2)^k} \right) \omega_k(f_i, \frac{1}{\lambda+2})_p \leq \omega_k(f_i, \frac{1}{\lambda+2})_p \leq c(p) \omega_k(f_i, \frac{1}{\lambda+2})_p.
\]

**Lemma 3.2.** [8]. For \( f_i \in L_{2\pi}^p, 0 < p < 1 \), we have

\[
\lim_{\lambda \to \infty} \| EN_\lambda[f_i] - f_i \|_p = \lim_{\lambda \to \infty} \frac{1}{\lambda^k} \sum_{k=1}^{\lambda} K \| EN_k[f_i] - [f_i] \|_p.
\]

**Theorem 3.3.** [2]. If \( f_i \in L_{2\pi}^p, 0 < p < 1, n \in \mathbb{N}, \lambda \in \mathbb{N} \), then

\[
\omega_k \left( f_i, \frac{1}{\lambda+2} \right) \leq c(p) \sum_{\lambda=1}^{n} \| EN_\lambda(f_i) - (f_i) \|_p.
\]

**Corollary 3.4.** for \( f_i \in L_{2\pi}^p, 0 < p < 1 \), we get
\[ c(p)\omega_k(f_i, \frac{1}{\lambda+2}) \leq EN_k[f_i] - [f_i] \leq c(p, k)\omega_k(f_i, \frac{1}{\lambda+2}) \]

**Proof.**

By lemma (3.2)

\[ \lim_{\lambda \to \infty} \| EN_k(f_i) - f_i \|_p \leq \frac{1}{\lambda^2} \sum_{k=1}^{\lambda} k \| EN_k(f_i) - f_i \|_{p'} \]

and by using theorem (3.1)

\[ \| EN_k(f_i) - f_i \|_p \leq c(p, k)\omega_k(f_i, \frac{1}{\lambda+2}) \]

Therefore, we get

\[ c(p)\omega_k(f_i, \frac{1}{\lambda+2}) \leq \| EN_k[f_i] - [f_i] \|_p \leq c(p, k)\omega_k(f_i, \frac{1}{\lambda+2}) \]

\[ \Box \]

**Theorem 3.5.** If \( f_i \in L^p_{2\pi}, 0 < p < 1 \), then \( \| EN_k[f_i] - f_i \|_p = O(\lambda^{-a}), 0 < a < k - 1 \), if and only if \( f_i \in Lip(\alpha)_k \)

**Proof.**

Let \( f_i \in Lip(\alpha)_k \) where \( 0 < \alpha < k - 1 \) we must prove that

\[ \| EN_k[f_i] - f_i \|_p = O(\lambda^{-a}). \]

Since \( f_i \in Lip(\alpha)_k \), then \( \omega_k(f, \lambda)_p = O(\lambda^{-a}). \) (3.5.1)

Using theorem 3.1 and (3.5.1) we get

\[ \| EN_k[f_i] - f_i \|_p \leq c(p)O(\lambda^{-a}). \]

Then

\[ \| EN_k[f_i] \|_p = O(\lambda^{-a}). \]

Let \( f_i \in L^p_{2\pi}, 0 < p < 1 \), then \( \| EN_k[f_i] - f_i \|_p = O(\lambda^{-a}). \)

We must prove that \( f_i \in Lip(\alpha)_k \).

Now, \( \| EN_k[f_i] - f_i \|_p = O(\lambda^{-a}) \)

\[ \leq c(\lambda^{-a}), \]

and since \( \| EN_k[f_i] - f_i \|_p \leq c(p, k)\omega_k(f_i, \lambda) \). Then

\[ c(p)\omega_k(f_i, \lambda) = c(\lambda^{-a}) \]
\[ \omega_k(f_i, \lambda) = O(\lambda^{-\alpha}). \]

Therefore, using definition of Lipschitian function we get \( f_i \in Lip(\alpha)_K \). 

References

1. Bhaya, E.S., and Walla. H. "Lp, p<1 Approximation Using Radial Basis Functions Neural Networks on Ordered Space", Journal of Engineering and Applied Sciences, vol.13, pp.4771-4773, 2018.

2. Bhaya, E.S., Abd Al-sadaa, Z.H., "Stechkin-Marchaud Inequality in Terms of Neural Networks Approximation in \( L_p \) Spaces for \( 0 < p < 1 \)." (to appear).

3. Bhaya, E.S, and Hawraa. A. A.," Neural Network Trigonometric Approximation", Journal of University of Babylon/ Pure and Aoolied Sciences, vol24, no.9, pp. 2395-2399, 2016.

4. Bolcke. H., Grohs. P., Kutyniok. G., and Petersen. P., " Optimal Approximation with Sparsely Connected Deep Neural Networks", ARXIV, vol. 4, pp.1705-01714, 2017.

5. Chen, X.H., White, H., "Improved Rates and Asymptotic Normality for Nonparametric Neural Network Estimators", IEEE Trans, Inform Theory, vol.45, vol. 682-691, 1999.

6. Chui, C.K., Li, X., "Approximation by Ridge Functions and Neural Networks with One Hidden Layer", Approx. Theory, vol.70, pp.131-341, 1992.

7. Cybenko, G., " Approximation by Superpositions of Sigmoidal Function", math Of control signals and system, vol.2, pp.303-314,1989.

8. Ding, C., Cao, F., Xu, Z., "The Essential Approximation Order for Neural Networks with Trigonometric Hidden Layer Units", Springer-Verlag Berlin Heidelberg, pp.72-79, 2006.

9. Ditzian, Z., Totik, V., Moduli of Smoothness, New York, Springer-Verlag, Berlin Heidelberg, 1987.

10. Dmitry. Y, Optimal Approximation of Continuous Function by Very Deep Relu Networks, Arxiv:1802.03620, vol.2, 6 Jun 2018.

11. Dmitry. Y, "Error Bounds for Approximation with Deep Relu Networks ", Neural Networks, vol.94, pp.103-114, 2017.

12. Johnen, H., Scherer, K., "On the Equivalence of the K-Functional and the Moduli of Continuity and Some Applications", Springer-Verlag, Berlin Heidelberg New York, vol. 571, pp.119-140, 1977.
13. Monica. B., "On the Complexity of Neural Network Classifiers", IEEE Explore Digital Linbrary, vol.25, no.8, pp.1553-1565, 2014.

14. Nagler. J, Cerejeiras. P. and Forster. B., " Lower Bounds for the Approximation with Variation - Diminishing Splines", Journal of Complexity, vol.32, no.1, pp. 81-91, 2016.

15. Peetre, J., "On the Connection between the Theory of Interpolation Spaces and Approximation Theory". In: Alexits, G., Stechkin, S.B.(eds): Proc. Conf. Construction of Function. Budapest, pp.351-363, 1969.

16. Philipp. P. and Felix. V. Optimal Approximation of Piecewise Smooth Function Using Deep Relu Neural Networks, Arxiv: 1709.0528, vol. 4 last Revised 22 May 2018., pp.1049-1058,1998.

17.

18. Walla H., "Constrained Approximation on Ordered Spaces," M.Sc. Thesis, University of Babylon, Babylon, Iraq, 2017.

18. Xu, Z.B., Cao, F.L., "Simultaneous $L^p$-Approximation Order for Neural Networks", Neural Networks, vol.18, pp.914-923, 2005.