The Essential Order of $L_p, p < 1$ Approximation Using Regular Neural Networks

Eman S. Bhaya
Omar A. Al-sammak
University of Babylon, College of Education for Pure Sciences, Department of Mathematics
emanbhaya@itnet.uobabylon.edu.iq  omar.alhelly@itnet.uobabylon.edu.iq

Abstract
This paper is concerned with essential degree of approximation using regular neural networks and how a multivariate function in $L_p(K)$ spaces for $p < 1$ can be approximated using a forward regular neural network. So, we can have the essential approximation ability of a multivariate function in $L_p(K)$ spaces for $p < 1$ using regular FFN.

Keywords. Neural network approximation, Modulus of smoothness, $L_p(K)$ Spaces, best approximation.

1. Introduction
Various papers on feasibility of approximation by forward neural networks have been made in past years (see [Cardaliaguet & Euvrard, 1992; Chen, 1995; Chen, 1994; Chui, 1992; Cybenko, 1989; Gallant, 1992; Hornik, 1989; Hornik & Stinchcombe, 1990; Leshno et al., 1993; Mhaskr & Michelli, 1992].

The most important result among these papers is that:

If we have a continuous function with multivariable and compact domain subset of $\mathbb{R}^d$ there exist a feed forward neural networks (FNNs) as an approximation for it. By a sigmoidal we can be approximated arbitrarily well. A three-layer of the FNNs with $d$ input units and one hidden and one output units can be mathematically expressed as

$$N_n(x) = \sum_{i=1}^{m} c_i \sigma (\omega_i x_i + \theta_i), \quad x \in \mathbb{R}^d, \quad d \geq 1,$$

where $1 \leq i \leq m$, $\theta_i \in \mathbb{R}$ is the threshold, $\omega_i = (\omega_{i1}, \omega_{i2}, \ldots, \omega_{id})^T \in \mathbb{R}^d$ are connection weights of neuron $i$ in the hidden layer with input neurons, $c_i \in \mathbb{R}$ are the connection strength of neuron $i$ with the output neuron, and $\sigma$ is the sigmoidal activation function used in the network.

In this paper we prove direct and inverse estimation and saturation problem for the approximation of multivariate function in $L_p(K)$ spaces for $p < 1$ using a forward regular neural network.
2. Notations and Definitions

Let \( \mathbb{R} \) be the set of reals, \( \mathbb{R}^d \) be the d-dimensional Euclidean space \((d \geq 1)\), and let \( K \) be any subset of \( \mathbb{R}^d \).

**Definition 2.1**

Let \( K \) be a multiple cell in \( d \)-dimension Euclidean space \( \mathbb{R}^d \) \((d \geq 1)\), the \( L_p \) space for \( p < 1 \) defined by:

\[
L_p(K) = \{ f: K \to \mathbb{R} \mid \|f\|_p = \left( \int_K |f|^p \right)^{\frac{1}{p}} < \infty \}.
\]

For any \( x = (x_1, x_2, \ldots, x_d) \) and \( y = (y_1, y_2, \ldots, y_d) \), let \( d(x, y) \) be the Euclidean distance of \( x \) and \( y \), that is,

\[
d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_d - y_d)^2}.
\]

**Definition 2.2**

The \( r \)th order difference of function \( f \in L_p(K) \) is\( \Delta^r f(x) = (1 - T^h)^r f(x) \), \( x = x_1, x_2, \ldots, x_d \in K \), \( r \in \mathbb{N} \) and \( h > 0 \).

We use \( I \) for the unit operator.

For any positive integer \( r \) we define the generalized module of smoothness of the \( r \)th order by the formula

\[
\omega_r(f, \delta)_p = \sup_{0 < h < \delta} \|\Delta^r f\|_p, \quad \delta > 0, f \in L_p(K).
\]

**Definition 2.3**

we denoted by the Lipschitzian class \( \text{Lip}(\alpha)_r \) defined by the space of all functions \( f \) in \( L_p(K) \) spaces satisfies \( \omega_r(f, \delta)_p = O(t^\alpha), \) where \( 0 < \alpha \leq r \).

**Remarks 2.4**

1) In this paper we deal with the approximation by neural network with special type of neural activation functions \( J_k \) of which each function \( \sigma: \mathbb{R} \to [0,1] \) has up to \( K + 1 \) order continuous derivatives \( \sigma^k, k = 1, 2, \ldots, K + 1 \).
2) The regular neural activation functions are the normal sigmoidal activation functions \( \sigma(x) = \frac{1}{1 + e^{-ax}} \) for positive .
3) Any neural network whose neural activation functions are regular (satisfies the conditions of part 1 of this remark) will be called a regular neural network.

**Definition 2.5**

Let \( r \) be an integer number and

\[
P_r(x) = a_r x^r, x \in [a, b] \subset (-\infty, \infty)
\]

be a homogeneous univariate polynomial of degree \( r \).

XU Zongben & CAO Feilong (2004) prove the lemma "Let \([a, b] \) be a compact interval, \( \sigma \in J_r \) be a regular neural activation function and \( P_r(x) \) a homogeneous univariate polynomial of the form \((2.6)\). Then for any \( \epsilon > 0 \), there is a neural network of the form \((1.1)\) the number of whose hidden units is not less than \( (r + 1) \) such that \( |N_n(x) - P_r(x)| < \epsilon \)"

As a direct consequence of above lemma we introduce the following theorem.

**Theorem 2.7**

For any regular neural activation function \( \sigma \in J_r \) and a homogeneous univariate polynomial \( P_r(x) \) and a given \( \epsilon > 0 \) there is a neural network of the form \((1.1)\) with not less than \( r + 1 \) hidden layers such that \( \|N_n(x) - P_r(x)\|_p < \epsilon \)

Proof:
Since $\sigma \in J_\ell$ be a regular neural activation function and $P_\ell(x)$ a homogeneous univariate polynomial then by above lemma we have for any $\epsilon > 0$ there is a neural network of the form (1.1) the number of whose hidden units is not less than $(r + 1)$ such that:

$$|N_n(x) - P_\ell(x)| < \frac{\epsilon}{(\mu(k))^{1/p}}$$

$$\Rightarrow |N_n(x) - P_\ell(x)|^p < \frac{\epsilon}{(\mu(k))^{1/p}}^p$$

$$\Rightarrow \int_K |N_n(x) - P_\ell(x)|^p < \int_K \left(\frac{\epsilon}{(\mu(k))^{1/p}}\right)^p$$

$$\Rightarrow \left(\int_K |N_n(x) - P_\ell(x)|^p\right)^{1/p} < \left(\int_K \left(\frac{\epsilon}{(\mu(k))^{1/p}}\right)^p\right)^{1/p}$$

$$\Rightarrow \|N_n(x) - P_\ell(x)\|_p < \frac{\epsilon}{(\mu(k))^{1/p}}^p$$

$$\Rightarrow \|N_n(x) - P_\ell(x)\|_p < \epsilon$$

In this section we construct an FNN to realize universal approximation to any integral multivariate functions in $L_p(K), P < 1$ we will use the Bernstein-Durrmeger operation as base tools.

**Definition 2.8** [X. Zongben and C. Feilong, (2004)]

Let $K$ be any subset of $R^d$ the Bernstein-Durremger operator $B_n$ in $L_1(T)$ defined by:

$$(B_n f)(\alpha) = \sum_{|k| \leq n} P_{n,k}(x) \phi_{n,k}(f) \quad (2.9)$$

Where $x \in K, f \in L_1(K)$ and 

$$\phi_{n,k}(f) = (n + d)!/n! \int K P_{n,k}(u) f(u)du$$

**Lemma 2.10**

If $f \in L_p(K)$ for $p < 1$ then

$$\|B_n f - f\|_p \leq c(p) \omega_r(f, 1/n)^p$$

**Proof:**

$$\|B_n f - f\|_p \leq c(p) K_1(f, (1/n)^r)^p$$

$$\leq c(p) \omega_r(f, \frac{1}{n})^p \quad [E.S. Belkina and S.S. Platoov (2008)]$$

**Lemma 2.11**

If $f \in L_p(K)$ for $p < 1$ then

$$\omega_r(f, 1/n) \leq c(p) \sum_{i=1}^n \|B_i f - f\|_p$$

**Proof://**

$$\omega_r(f, \delta)^p = \omega_r(f - B_n f + B_n f, \delta)^p \leq c(p) \omega_r(f - B_n f, \delta) + \omega_r(B_n f, \delta) \leq c(p) \|f - B_n f\|_p + \omega_r(f, \delta)$$

$$B_n(f) - B_0(f) = B_n(f) - B_2^1(f) + (B_2^1(f) - B_2^1(f)) + \cdots + (B_4(f) - B_0(f))$$

$2^l = n, l = \max i, 2^l < n$

$J = \omega_r(B_n f, \delta) = \omega_r(\sum_{i=1}^l B_2^i f - B_2^i f, \delta)$

$$\leq c(p) \sum_{i=1}^l \|f - B_2^i f\|_p$$

$$\leq c(p) \sum_{i=1}^n \|f - B_i f\|_p$$

$$\omega_r(f, \delta) \leq c(p) \sum_{i=1}^n \|B_i f - f\|_p \quad \blacksquare$$

To explain Lemma 2.13 we need the flowing notations: [X. Zongben and C. Feilong, (2004)]

1. Let $Z^d_\ell$ be the set of all non-negative multi-integers in $R^d$
2. For any \( x = (x_1, x_2, \ldots, x_d) \in R^d \) and \( k = (k_1, k_2, \ldots, k_d) \in Z^d \), Let \( |x| = \sum_{i=1}^{d} x_i \), \(|k| = \sum_{i=1}^{d} k_i\), and \( x^k = x_1^{k_1} \cdots x_d^{k_d} \). If \( j_1! j_2! \cdots j_d! = k_1! k_2! \cdots k_d! \) then each \( i_l^{(p)} = p \) is multi-integer in \( Z^d \) that satisfies \( |i_l^{(p)}| = p \).

3. We say that \( x \leq y \) for any \( y \in R^s \), iff \( x_i \leq y_i \) for any \( 1 \leq i \leq s \).

4. Let \( N_p = \binom{p + d - 1}{d - 1} \) be the number of multi-integers \( i = (i_1, i_2, \ldots, i_d) \) in \( Z^d \) that satisfy \( i_1 + i_2 + \cdots + i_d = p \).

5. Let \( L_p = \binom{p + d - 2}{d - 2} \) be the number of multi-integers \( j = (j_1, j_2, \ldots, j_d) \) in \( Z^{d-1} \) that satisfy \( j_1 + j_2 + \cdots + j_{d-1} = p \).

6. Denote by \( J_{N_p-1} + l, 1 \leq l \leq L_p \) a generic multi-integers \( j \) in \( Z^{d-1} \) satisfying \( j_1 + j_2 + \cdots + j_{d-1} = p \).

7. \( j_1, 1 \leq l \leq N_p \) a generic multi-integers \( j \) in \( Z^{d-1} \) satisfying \( j_1 + j_2 + \cdots + j_{d-1} = p \).

8. Any \( 1 \leq l \leq N_p \) Let \( i_l^{(p)} \) then \( i_l^{(p)} = p \) is multi-integer in \( Z^d \) that satisfies \( |i_l^{(p)}| = p \).

9. Define \( p_i = (j_1, 1 \leq l \leq p_i) \) and \( p_i^{(p)} = \frac{1}{2^{(1+p)}} p_1, 1 \leq l \leq N_p \).

The following lemma provides an equivalent expression of Bernstein-durrmeyer operator \( B_n \).

**Lemma 2.12** \([X. Zongben and C. Feilong, (2004)](https://example.com)\)

For any \( \in L_p(\mathbb{T}) \), the Berustin–Durrmeyer operator \( B_n f \) in (2.9) can be expressed as:

\[
B_n f(x) = \sum_{p=0}^{n} \sum_{l=1}^{N_p} d_l^{(p)} (x, p_1^{(n)})^p \tag{2.13}
\]

where \( (x, p_1^{(n)}) \) is inner product \( x \) and \( p_1^{(n)}, q_1^{(p)} \) are uniquely determined by:

\[
\binom{\left(\begin{array}{c}
(1)\left(\begin{array}{c}
(j_1)_{i_1} \\
(j_2)_{i_2} \\
\vdots \\
(j_{N_p})_{i_{N_p}}
\end{array}\right)
\end{array}\right)}{\left(\begin{array}{c}
(d_1)^{(p)} \\
(d_2)^{(p)} \\
\vdots \\
(d_{N_p})^{(p)}
\end{array}\right)} = \binom{\left(\begin{array}{c}
(i_1)^{(p)}! \\
(i_2)^{(p)}! \\
\vdots \\
(i_{N_p})^{(p)}!
\end{array}\right)}{\left(\begin{array}{c}
c_1^{(p)}(f) \\
c_2^{(p)}(f) \\
\vdots \\
c_{N_p}^{(p)}(f)
\end{array}\right)}
\]

With \( c_1^{(p)}(f) = \frac{n!}{(n-p)!} \sum_{q \leq i_1^{(p)}} (f) \frac{1}{q! (i_1^{(p)} - q)!} (-1)^{i_1^{(p)} - q} \).

1 \leq l \leq N_p.

**Remark 2.14**

We observe that in expression (2.14) each term \( (x, p_1^{(n)}) \) can be viewed as homogeneous univariate polynomial of \( (x, p_1^{(n)}) \) with order \( p \), and so by Theorem 2.8, it can be approximated arbitrarily well by a network of the form:

\[
N_{p+1}(x) = \sum_{i=1}^{N_p} c_{i,p} \sigma(\omega_{i,p}(x, p_1^{(n)}) + \theta), K_p \geq p + 1
\]

**Remark 2.15**

1) As \( B_n f(x) \) can be approximate \( f \), the following neural networks:

\[
N_n(x) = \sum_{p=0}^{n} \sum_{l=1}^{N_p} c_{i,p} \sigma(\omega_{i,p}(x, p_1^{(n)}) + \theta) \tag{2.16}
\]
Then can approximate \( f \) to any accuracy.

2) the network (2.17) will be the FNN models we propose use in this paper.

3) the network (2.17) are clearly of from (1.1) and contain hidden units where

\[
m = N_0 + k_1N_1 + \cdots + k_nN_n \\
\geq N_0 + 2N_1 + \cdots + (n + 1)N_n \\
= \sum_{k=0}^{n}(k + 1) \left( \frac{k + d - 1}{d - 1} \right) = m_0(n) .
\]

**Theorem 2.17**

For any \( f \in L_p(K) \), \( p < 1 \) there is a regular, one hidden layer FNN, \( N_n(x) \), of the form (1.1) with \( \sigma \in J_n \) and the hidden unit number \( m \geq \sum_{k=0}^{n}(k + 1) \left( \frac{k + d - 1}{d - 1} \right) = m_0(n) \) such that

\[
\| N_n - f \|_p \leq c(p) \omega_r(f, \frac{1}{n})_p 
\]

Proof: we assume \( f \in L_p(K) \)

Then, by Lemma 2.12, the Bernstein – durrmeyer operator \( B_n f \) can be defined and expressed as

\[
B_n f(x) = \sum_{p=0}^{n} \sum_{l=1}^{N_p} d_l^{(p)}(x, p_1^{(n)})^p 
\]

and, furthermore, it approximates \( f \) in the following sense

\[
\| B_n f - f \|_p \leq c(p) \omega_r(f, \frac{1}{n})_p \quad [\text{lemma 2.10}]
\]

And by remark (2.15) part 1 , we have

\[
\| N_n - f \|_p \leq c(p) \| B_n f - f \|_p \leq c(p) \omega_r(f, \frac{1}{n})_p . \quad \square
\]

**Theorem 2.18**

For any \( f \in L_p(K) \), \( p < 1 \) there is a regular, one hidden layer FNN, \( N_n(x) \), of the form (1.1) with \( \sigma \in J_n \) and the hidden unit number \( m \geq \sum_{k=0}^{n}(k + 1) \left( \frac{k + d - 1}{d - 1} \right) = m_0(n) \) such that

\[
c(p)\omega_r(f, \frac{1}{n})_p \leq \sum_{l=1}^{N_1} \| N_l - f \|_p 
\]

Proof: we assume \( f \in L_p(K) \)

Then, by Lemma 2.12, the Bernstein – durrmeyer operator \( B_n f \) can be defined and expressed as

\[
B_n f(x) = \sum_{p=0}^{n} \sum_{l=1}^{N_p} d_l^{(p)}(x, p_1^{(n)})^p 
\]

and since

\[
\left| p_1^{(n)} \right| \leq \frac{1}{2} , \quad \text{we have} \quad \langle x, p_1^{(n)} \rangle \leq 1 
\]

i.e \((-1 \leq \langle x, p_1^{(n)} \rangle \leq 1)\)

Each term \( \langle x, p_1^{(n)} \rangle^p \) in (2.13) is univariate homogeneous polynomial of \( \langle x, p_1^{(n)} \rangle \) with order \( p \) defined on \([-1,1]\)

Now by Theorem 2.7 we have \( \langle x, p_1^{(n)} \rangle^p \) can be approximated by neural network

\[
N_{K_p} = \sum_{i=1}^{N_p} c_{p,i} \sigma \left( \omega_{p,i}(x, p_1^{(n)}) + \theta \right) , c_{p,i}, \omega_{p,i} \in R , K_p \geq p + 1 \quad (2.19)
\]

With accuracy

\[
\left\| N_{K_p} - \langle x, p_1^{(n)} \rangle^p \right\|_p \leq \epsilon \quad (2.20)
\]

Then for constructed FNN

\[
N_n(x) = \sum_{p=0}^{n} \sum_{l=1}^{N_p} d_l^{(p)} \sum_{i=1}^{K_p} c_{p,i} \sigma \left( \omega_{k,p}(x, p_1^{(n)}) + \theta \right) , c_{p,i}, \omega_{k,p} \in R , K_p \geq p + 1 
\]

and we have

\[
\| N_n - f \|_p = \| N_n f - B_n f + B_n f - f \|_p
\]

81
\[ \leq \|N_n f - B_n f\|_p + \|B_n f - f\|_p \]
\[ \leq c(p) \omega_r \left( f, \frac{1}{n} \right)_p + \|N_n f - B_n f\|_p \]  
(22.1)

The term \( \|N_n f - B_n f\|_p \) in above can be arbitrarily small, because (2.19) and (2.20) imply

\[ \|N_n f - B_n f\|_p = \left\| \sum_{p=0}^{n} \sum_{i=1}^{N_p} d_i^{(p)} \left\{ (x, p_i^{(n)})^p - \sum_{i=1}^{kp} c_{i,p} \sigma \left( \omega_{i,p}(x, p_i^{(n)}) + \theta \right) \right\} \right\|_p \]
\[ = \sum_{p=0}^{n} \sum_{i=1}^{N_p} \left| d_i^{(p)} \right| \max \left\{ (x, p_i^{(n)})^p - N_{K_p} \right\} \leq \varepsilon \sum_{p=0}^{n} \sum_{i=1}^{N_p} |d_i^{(p)}| \]

Then inequality (2.21) imply \( \|B_n f - f\|_p \leq c(p) \omega_r \left( f, \frac{1}{n} \right)_p \)
and by (Lemma 2.11) \( c(p) \omega_r \left( f, \frac{1}{n} \right)_p \leq c(p) \sum_{i=1}^{n} \|B_i f - f\|_p \)

Then \( c(p) \omega_r \left( f, \frac{1}{n} \right)_p \leq \sum_{i=1}^{n} \|N_n f - f\|_p \). ■

**Theorem 2.23**

For any \( f \in L_p(K), p < 1 \) there is a regular, one hidden layer FNN, \( N_n(x) \), of the form (1.1) with \( \alpha \in ]1, \infty[ \) and the hidden unit number \( m \geq \sum_{k=0}^{n}(k + 1)(k+1...1)(d-1) \) such that

\[ \|N_n f - B_n f\|_p = O(\sqrt{\alpha}) \]

if and only if \( f \in Lip(\alpha)_r \)

Proof: let \( f \in Lip(\alpha)_r \)

From \( \|N_n f - f\|_p \leq c(p) \omega_r \left( f, \frac{1}{n} \right)_p \Rightarrow \|N_n f - f\|_p = O(\sqrt{\alpha}) \)

And by \( c(p) \omega_r \left( f, \frac{1}{n} \right)_p \leq \sum_{i=1}^{n} \|B_i f - f\|_p \)

We obtain \( \omega_r \left( f, \frac{1}{n} \right)_p = O \left( \frac{1}{n^2} \right) \). ■

**References**


Leshno, M., Lin, V. Y., Pinkas, A., (1993), Multilayer feedforward networks with a non-polynomial activation function can approximate any function, Neural Networks, 6: 861—867.
