Republic of Iraq
Ministry of Higher Education
and Scientific Research
University of Babylon
College of Education for Pure
Science
Department of Mathematics

New Neural Networks for New Weighted Approximation

A Thesis Submitted to the Council of College of Education for Pure Sciences, University Babylon in Partial Fulfillment of the Requirements for the Degree of Master in Education /Mathematics.

By

Ahmad AbdulKarim Hassan Mardan

Supervised by

Asst.Prof.Dr. Hawraa Abbas Almurieb

2022 A.D 1444 A.H

بِسْمِ اللَّــهِ الرَّحْمَـٰنِ الرَّحِيمِ

يَا بُنَيَّ إِنَّهَا إِن تَكُ مِثْقَالَ حَبَّةٍ مِّنْ خَرْدَلٍ فَتَكُن
فِي صَخْرَةٍ أَوْ فِي السَّمَاوَاتِ أَوْ فِي الْأَرْضِ يَأْتِ
بِهَا اللَّهُ إِنَّ اللَّهَ لَطِيفٌ خَبِيرٌ

صدق الله العليّ العظيم

(من سورة لقمان , الآية 16)

Supervisor Certification

I certify that this thesis entitled **" New Neural Networks for New Weighted Approximation "**, was prepared by the student**" Ahmad AbdulKarim Hassan Mardan"** under my supervision at the Department of mathematics , College of Education for Pure Sciences, University of Babylon as a partial fulfillment of the requirement the Degree for Master in Education /Mathematics.

Signature:

Name: **Asst.Prof.Dr. Hawraa Abbas Almurieb**

Date :   /   / 2022

According to the available recommendation, I forward this thesis for debate by the examining committee.

Signature

Name: **Dr. Azal Jaafar Musa**

Scientific grade: Assistant Prof.

Date :   /   /2022

Head of Department of Mathematics

Scientific Supervisor's Certification

This is to certify that I have read this thesis entitled **" New Neural Networks for New Weighted Approximation "** and I found it is qualified for debate.

Signature:

Name:

Date: / /2022

Linguistic Supervisor's Certification

This is to certify that I have read this thesis entitled **" New Neural Networks for New Weighted Approximation "** and I found it is qualified for debate.

Signature:

Name:

Date: / /2022

Examining Committee Certification

We certify that we have read the thesis entitled" **New Neural Networks for New Weighted Approximation** ", and as committee, examined the student" **Ahmad AbdulKarim Hassan Mardan** " in its contents and that in our opinion it meets the standard of a dissertation for the Degree of Master in Education/ Mathematics.

Signature:
Name: Dr. Eman Samir Bhaya
Title: Professor
Date: /10/2022
Chairman

Signature:
Name: Dr. Nadia Mohammed Jwad Ibrahem
Title: Assistant Professor
Date: /10/2022
Member

Signature:
Name: Dr. Zainab Abdulmunim Sharba
Title: Assistant Professor
Date: /10/2022
Member

Signature:
Name: Dr. Hawraa Abbas Almurieb
Title: Assistant Professor
Date: /10/2022
Member / Supervisor

Signature:
Name: Dr. Bahaa Hussien Salih Rabee
Scientific Grade: Professor
Address: Dean of the College of Education for Pure Sciences
Date: / /2022

Dedication

To the great Messenger and to the seal of prophets "Mohammed" May Allah bless him and his family and grant him peace.

To the prince of believers Ali.

To Madam Fatima Al- Zahra.

To Aba Al-Hessen Al-Mahdi

To my supervisor" **Dr. Hawraa Abbas Almurieb**".

To my family.

With love and Appreciation

I dedicate this work

Ahmad

Acknowledgements

Thanks for God, the most mercy, to provide me the opportunity and to desire to achieve this work.

I would like to present my thanks and appreciation to my supervisor **"Dr. Hawraa Abbas Almurieb"** who supported and provided me with suggestions throughout the work period. More thanks are to the Department of Mathematics professors.

Thanks are to my family for their support and save the right at me for my success in this work

Finally, thanks are for my dear colleagues.

Table of Symbols

Symbols	Definition
$P_n(\cdot)$	**Polynomial of degree at most n**
e	**Backpropagation error**
$\omega_{\alpha,\beta}(\cdot)$	**The classical Jacobian weights**
$\omega_{k,r}^{\varphi}(\cdot)_{\alpha,\beta,p}$	**Weighted DT modulus of smoothness**
$\varphi(\cdot)$	$\varphi(x) = \sqrt{1-x^2}\ , x \in [-1,1]$
L_p	**Lebesgue space**
$\Delta_h^k(\cdot)$	**The kth symmetric difference**
$\mathbb{N}$	**The space of natural numbers**
$\mathbb{N}_0$	$\mathbb{N} \cup \{0\}$
$\mathcal{W}_\delta^J(\cdot)$	**Generalized Jacobian Weight**
$\omega_{k,r}^{\varphi}(\cdot,\cdot)_p$	**General Jacobian Weighted Modulus of Smoothness**
$\omega_{k,r}^{*\varphi}(\cdot,\cdot)_p$	**Averaged moduli of smoothness**
$K_{k,r}^{\varphi}(\cdot,\cdot)_{\alpha,\beta,p}$	**K-functional**
$\mathbb{R}^d$	**The space of real numbers of dimension d**
$\mathbb{R}$	**The space of real numbers**
$\mathrm{R}(\mathrm{x})$	**Activation Function**
$N_{d,r}(\cdot,\cdot)$	**Neural Network**
Ω	**The family of all neural networks for functions from $\mathbb{B}_p^r$**

$E(\cdot)_p$	The degree of best approximation of $\mathbb{B}_p^r$ from Ω
δ_{im}	$\delta_{im} = \begin{cases} 1 & , \text{ if both } i \text{ and } m \text{ are even or both are odd} \\ 0 & , \quad o.w. \end{cases}$
$x_j(\cdot)$	The jth component of vector input
$t(\cdot)$	Actual results vector
$o(\cdot)$	The vector output
ζ	Local gradients
η	The learning-rate parameter
α	The momentum constant

Abbreviations

Symbols	Definition	Page
UAT	Universal Approximation Theorem	2
NNs	Neural Networks	2
MLP	Multilayer Perceptron	2
GJWMS	General Jacobian Weighted Modulus of Smoothness	9
ANN	Artificial Neural Networks	25
SNNs	Simulated Neural Networks	25
BP	Backpropagation	39

Abstract

The topic of modulus of smoothness still gets the interest of many researchers, for its applicable usage in different fields, especially for function approximation. Few researchers used General Jacobian form as a weight in the modulus of smoothness. Therefore, in thesis work a new General Jacobian Weight was defined, through which we define the modulus of smoothness of weighted type, named as *General Jacobian Weighted Modulus of Smoothness*. Properties of our modulus are studied here to be easily used with functions from spaces $L^r_{p,\varphi}$, especially when $0 < p < 1$. The properties of our weight were also studied, in addition to the relationship with classical Jacobian weights .

One of the most important results, whose usage is wide in the theory of approximation of functions, is the equivalence between our General Jacobian Weighted Modulus of Smoothness and the K-functional One. The role of that equivalence is to prove the major approximation theories later .

For the need of approximation functions with neural networks, the expertise of approximation is used to define a new activation function has been defined in addition to a new neural network. Many properties of our activation function are studied especially its differentiability. Also, functions from spaces $L^r_{p,\varphi}$ are approximated by the neural networks with the above activation function. Moreover, the degree of approximation is estimated using a previously defined modulus of smoothness in order that we relate concepts to each other eventually.

Lastly, a realistic evidence on the importance of approximation in learning through a widely used backpropagation algorithm is provided, and some examples show the approximation abilities through MATLAB.

Contents

List of Figures

Publications

1) A. A. Al-Dhahabi and H. A. Almurieb, “Generalized Jacobian Weighted Modulus of Smoothness”, Accepted in the Iraqi Journal of Science, 2022, to be published in January 2023.

2) A. A. Al-Dhahabi and H. A. Almurieb, “Neural Network Essential Approximation in Besov Spaces $\mathbb{B}_p^r$”, Journal of Optoelectronics Laser, vol. 41, no. 4, pp. 33–38, 2022.

Chapter One

Brief Introduction to Approximation and Neural Networks

1.1 Brief History of Approximation Functions

Approximation functions are important in many fields, including applied mathematics and computer science. The necessity of approximating functions arises in different situations, for example when there is a need to replace a certain function with a smoother or simpler one. Also, it is useful when there is a requirement to establish a functional dependence on the basis of empirical data.

Chebyshev is considered to be the first to lay the foundations for the approximation of functions, through his book [1] in 1853 which he dealt with laying the mathematical foundations for the theory of machines.

In 1885 , Weierstrass [2] proved his famous theorem on the approximation of continuous functions by algebraic polynomials which states that :

"Any continuous function on a bounded interval can be uniformly approximated by polynomial functions."[2]

In 1912 , Sergei Bernstein [3] introduced his famous polynomials to prove the Weierstrass Approximation theorem:

" If $F(x)$ is any continuous function in the interval $[0,1]$, it is always possible, regardless how small ε, to determine a polynomial $P_n(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n$ of degree at most n high enough such that we have

$$|F(x) - P_n(x)| < \varepsilon,$$

for every point in the interval under consideration."[3]

In 1946 , Salem [4] approximated continuous functions by Partial Sums of Fourier Series.

In 1968 , De Boor [5] approximated continuous functions by polynomial spline functions on $[a, b]$ of degree $k - 1$ on π .

In 1974 , Motornyi [6] approximated periodic function by trigonometric polynomials.

Since 1989, researchers began using neural networks to approximate functions which is preferred to be used for approximation because it is a universal approximator to learn any continuous function to reach to the approximation target.

Where in 1989, Cybenko [7] was the first who used neural networks in approximation theory, where he proved that any continuous function is approximated by a single hidden-layer feedforward neural network. In other words, UAT concludes that the principle of NNs is function approximation. Many other researchers have used neural networks to approximate functions with different activation functions such as **Sigmoid** [8], **Logsig** [9], **ReLU** [10], **Tanh** [11].

1.2 The History of Neural Networks

The history of neural networks is much older than most people think, the idea of a "thinking machine" can be traced back to ancient Greece as well as the ancient Egyptians, we'll focus on the key events that led to the development of thinking about neural networks. In the following section, we introduce important stations for development of neural networks :

In 1943, McCulloch and Pitts[12] ,were the first who proposed the MLP model of neurons. The model is based on known principles of the biological processes of neurons and it is the first mathematical model of neurons. Also MLP model is the first attempt to describe the working principle of the brain in human history.

In 1958, Frank Rosenblatt [13] was credited with developing the Perceptron model. He took McCulloch and Pitts ' work one step further by introducing weights into the equation. Using the IBM 704, Rosenblatt made the computer able to learn how to characterize cards that were placed on the left versus the cards that are placed on the right.

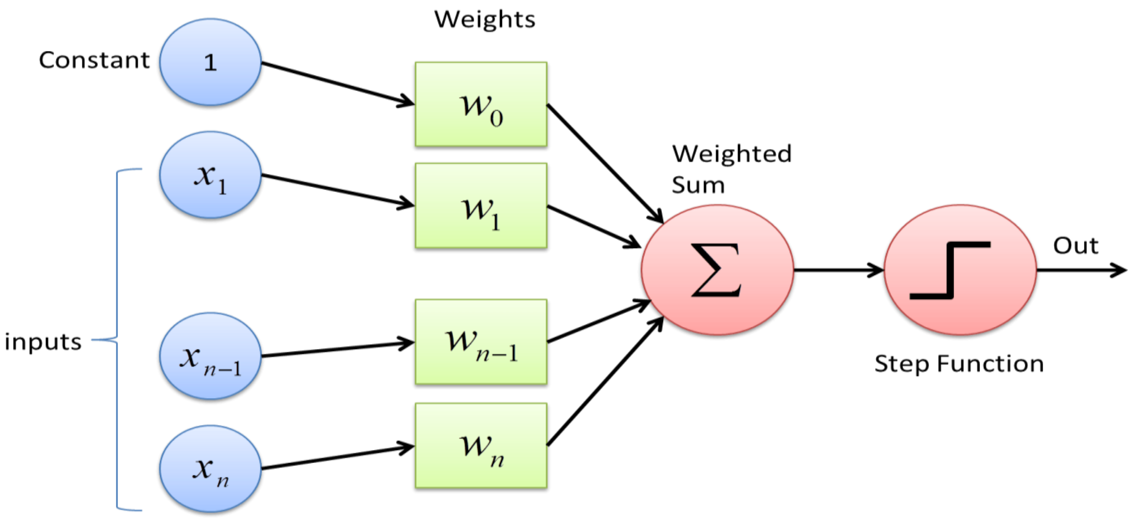


Figure 1.1 The Perceptron Model

In 1969, Marvin Minsky [14] ,in his book Perceptron, discussed the impossibility of creating a multi-layer neural network, disrupts the networks, deep learning, and artificial intelligence in general.

The failure of solving some problems by the perceptron networks had led to find other networks capable of solving such problems .Where as in 1974, Paul Werbos [15] discovered the backpropagation algorithm for such networks, it was not widely published, until 1986 when Rumelhart, Geoffrey and Ronald [16] have modeled a new formula for the algorithm known as backpropagation with error :

$$e = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{j,i} - d_{j,i})^2,$$

where i is an index over cases (input - output pairs) , j is an index over output units , y is the actual state of an output unit and d is the desired state

Until 2006, researchers didn't know how to train neural networks to bypass traditional methods, except for some specialized problems. What changed in 2006 was the discovery of learning techniques in so-called deep neural networks. These technologies are now known as deep learning .

In 2006, Hinton, Osindero and Teh [17] demonstrated the initial training technique for neural networks. He pointed out that it is possible to train a multi-layer neural network initially and effectively, if each layer is trained individually using a finite Boltzmann machine, and then he continued the training through the method of backpropagation of errors.

Today, deep neural networks and deep learning achieve outstanding performance on many important problems in computer vision, speech recognition, and natural language processing. It is widely published by companies such as Google, Microsoft, and Facebook.

1.3 Modului of Smoothness

Moduli of smoothness vary among researches in their structures. The idea began with defining the weighted norm by Hunt [18] in 1973. It is not clear when the first weighted modulus of smoothness was introduced, but it is probably defined first by Ditizian and Totik in their book [19] in 1987, they were working on linking the weighted moduli of smoothness to the weighted approximation [20] in weighted spaces such as in [21].

In 1999, Ditzian Define a modulus of smoothness on the unit sphere [22], also relate this modulus to best approximation by spherical harmonics of smaller order .

In 2002, Draganov [23] defined a new modulus of smoothness for trigonometric polynomial approximation .

In 2007, Jianjun defined weighted modulus in terms of the classical Jacobi weights [24]

$$\omega_{\alpha,\beta}(x) = (1-x)^{\alpha}(1+x)^{\beta} \,, \tag{1.3.1}$$

where $x \in [-1,1]$, α and $\beta \in J_\delta$ which is given by

$$J_\delta = \begin{cases} \left(-\frac{1}{p}, \infty\right), & \text{If } 0 < p < \infty \\ [0, \infty)\ , & \text{If } p = \infty \end{cases}$$

Since 2014, Koputon , Leviatan and Shevchuk have made many generalizations to the modulus defined with weights

$$\mathcal{W}_\delta^{\frac{r}{2}+\alpha,\frac{r}{2}+\beta}(x) = (1 - x - \delta\varphi(x)/2)^{\frac{r}{2}+\alpha}(1 + x - \delta\varphi(x)/2)^{\frac{r}{2}+\beta},$$

in their papers [25] , [26] and [27], as follow

$$\omega_{k,r}^{\varphi}\left(f^{(r)}, t\right)_{\alpha,\beta,p} = \sup_{0 \le h \le t} \left\| \mathcal{W}_{kh}^{\frac{r}{2}+\alpha,\frac{r}{2}+\beta}(\cdot)\Delta_{h\varphi(\cdot)}^{k}(f^{(r)},\cdot) \right\|_p \qquad ,$$

where $f \in L_{p,\varphi}^r$, $\varphi(x) = \sqrt{1 - x^2}$ and $\alpha, \beta \in J_p, k \in N$.

1.4 Results Summary

In this work, in addition to the study of the approximation of functions, the modulus of smoothness has been studied. The question here is: what is the relationship between the modulus of smoothness and the approximation of functions? To answer this question, when using the approximation process, it is important to have a calculated measure of its error. For the next chapter, we define a new modulus of smoothness through a weight which is called the generalized Jacobian weight is define as follow:

$$\mathcal{W}_\delta^J(x) = \prod_{j=1}^{M} \left| x - z_j - \frac{\delta\varphi(x)}{M} \right|^{\lambda j} \qquad ,$$

where $-1 = z_1 \le \cdots \le z_M = 1, \lambda j \in J_\delta$. So we define the modulus

$$\omega_{k,r}^{\varphi}\left(f^{(r)}, \delta\right)_p = \sup_{0 \le h \le \delta} \left\| \mathcal{W}_{kh}^J(\cdot)\Delta_{h\varphi(x)}^{k}(f^{(r)},\cdot) \right\|_p \qquad ,$$

we can use the General Jacobian Weighted Modulus of Smoothness (GJWMS) which is used as an accurate measure of the structural properties of the function, for getting the best approximation and the fast approach to zero .

It includes some important properties that show the relationship among our weight , the classical Jacobian weight and the generalized Jacobian weight. Moreover, we discuss the relationship between our generalized jacobian weighted modulus of smoothness and the K-functional, in the following theorem, we get the equivalence between them **as follow:**

Theorem I

let $0 < p < 1$ and if $k \in \mathbb{N}, r \in \mathbb{N}_0, f \in L^r_{p,\varphi}$ then for $0 < \delta \le 2/k$:

$$c(k,r,p)K^{\varphi}_{k,r}\left(f^{(r)},\delta^k\right)_p \le \omega^{*\varphi}_{k,r}\left(f^{(r)},\delta\right)_p \le \omega^{\varphi}_{k,r}\left(f^{(r)},\delta\right)_p$$
$$\le c(k,r,p)K^{\varphi}_{k,r}\left(f^{(r)},\delta^k\right)_p .$$

In chapter three , we introduce a new activation function through which the neural network was defined . Some properties of the activation function was included to get theories related to the approximation of functions using neural networks, through the equivalence between our General Jacobian Weighted Modulus of Smoothness and the K-functional .

Definition II (Activation Function)

For any $x \in [-1,1]$, define

$$R(x) = \begin{cases} 0 & if\ x \le 0 \\ 1 - \varphi(x) & if\ x > 0 \end{cases} ,$$

where $\varphi(x) = \sqrt{1 - x^2}$ **.**

Lemma III (Activation Function Properties)

1. The domain and rang of R are both $[-1,1]$

2. $|\, R(x)| \leq 1$, $\quad \forall\, x \in [-1,1]$

3. $\|\, R(x)\|_p \leq (2)^{\frac{1}{p}}$, $\quad \forall\, x \in [-1,1]$

4. R is differentiable and the first derivative of R is:

$$R'(x) = \frac{x}{\varphi(x)}$$

where $\varphi(x) = \sqrt{1-x^2}$ and $x \in (-1,1)$.

5. The general form of derivatives of R is:

$$R^{(m)}(x) = \sum_{i=0}^{m} \frac{cx^i}{\left(\varphi(x)\right)^{m+i-1}}\, \delta_{im} \quad ,$$

where

$$\delta_{im} = \begin{cases} 1 & , \quad \text{if both } i \text{ and } m \text{ are even or both are odd} \\ 0 & , \quad o.w. \end{cases}$$

Definition IV (Neural Network)

Let $x \in [-1,1]$ and $f \in L^r_{p,\varphi}$

$$N_{d,r}(f^{(r)}, \boldsymbol{x}) = \sum_{j=1}^{d} \sum_{i=1}^{k} \binom{k}{i} (-1)^{k-i} f^{(r)}(h\varphi(a)) R(\boldsymbol{x}) \; ,$$

where $a \in [-1,1]$ and $\boldsymbol{x} = \omega_j x_j + b_j$.

In the following theorem, functions were approximated from the space $L^r_{p,\varphi}$ by neural network with degree of approximation lower than the modulus of smoothness.

Theorem V (Direct Theorem)

Let $k \in N, r \in \mathbb{N}_0, 0 < p < 1$ then for any $f \in L^r_{p,\varphi}$, there exist $N_{d,r} \in \Omega$ in **(Definition IV)**, that satisfies

$$\left\|f^{(r)} - N_{d,r}(f^{(r)})\right\|_p \leq c(k,p)\omega^{\varphi}_{k,r}\left(f^{(r)}, \delta\right)_p$$

for $\delta > 0$.

Theorem III (Inverse Theorem)

Let $k \in N, r \in \mathbb{N}_0, 0 < p < 1$ then for any $f \in L^r_{p,\varphi}$, there exist $N_{d,r}$ of form in **Definition IV**, satisfies

$$\omega^{\varphi}_{k,r}\left(f^{(r)}, \delta\right)_p \leq c(k,p)\left\|f^{(r)}(x) - N_{d,r}\left(f^{(r)}, \boldsymbol{x}\right)\right\|_p$$

If the function is differentiable with degree $r+1$, it can be approximated using the derivative of the neural network with degree of approximation lower than the modulus of smoothness, The following theorem shows this :

Theorem IV

Let $r \in \mathbb{N}_0, 0 < p < 1$ and $f \in L^{r+1}_{p,\varphi}$

$$\left\|f^{(r+1)}(x) - \acute{N}(f^{(r+1)}, \boldsymbol{x})\right\|^p_p \leq c\omega^{\varphi}_{k,r}\left(f^{(r+1)}, \frac{1}{n}\right)$$

Theorem V

Let $r \in \mathbb{N}_0, 0 < p < 1$ and $f \in L^{r+1}_{p,\varphi}$

$$\omega^{\varphi}_{k,r}\left(f^{(r+1)}, \frac{1}{n}\right)_p \leq c\left\|f^{(r+1)}(x) - \acute{N}\left(f^{(r+1)}, x\right)\right\|_p.$$

In chapter four , we deal with an introduction to learning of the neural network, in addition to how to learn the neural network, and then we approximated some functions using the backpropagation algorithm applied in MATLAB.

Chapter Two

Generalized Jacobian Weighted Modulus of Smoothness

2.1 Introduction and preliminaries

Weighted modulus of smoothness has several uses in function approximation, especially for estimating the degree of approximation, so many authors were investigating other weights for different spaces and moduli. The weighted modulus of smoothness is much better than others for its ability to approach to zero as fast as the good weight is chosen. The standards of choosing the weights is studied in details in [19]. Jacobian weights [24] are of the most important defined weights for their usage in different purposes in approximation. The classical Jacobin weights are given by (1.3.1). They were used widely to define weighted moduli of smoothness. Some generalizations were made but the last-defined weighted DT modulus of smoothness is given by [27]

$$\omega_{k,r}^{\varphi}\left(f^{(r)},t\right)_{\alpha,\beta,p} = \sup_{0\leq h\leq t}\left\|\mathcal{W}_{kh}^{\frac{r}{2}+\alpha,\frac{r}{2}+\beta}(\cdot)\Delta_{h\varphi(\cdot)}^{k}(f^{(r)},\cdot)\right\|_{p} \quad , \qquad (2.1.1)$$

where $f \in L_{p,\varphi}^{r}, \varphi(x) = \sqrt{1-x^2}$ and $\alpha,\beta \in J_\delta, k \in N$.

On the other hand, generalized Jacobian weight is a good choice for usage in our work. Studying approximation on finer partition of the interval $[-1,1]$ may led to more accurate degree of approximation. For this purpose , we study generalized Jacobian weight with $M \geq 2$, given by [28] :

$$\mathcal{W}(x) = \prod_{j=1}^{M}\left|x-z_j\right|^{\lambda j} \quad , \qquad (2.1.2)$$

where $-1 = z_1 \leq \cdots \leq z_M = 1, \lambda j \in J_\delta$.

Since we need to use approximation in L_p space, L_p space is defined in the following definition :

Definition 2.1.1 :

For a measurable function $f: [-1,1] \to \mathbb{R}$ and an interval $I \subseteq [-1,1]$, the $L_{\mathcal{W},p}$ space for $0 < p < \infty$ and a weight function $\mathcal{W}$ is defined as follow:

$$L_{\mathcal{W},p}(I) = \left\{ f \mid \ \|\mathcal{W}f\|_{L_p(I)} < \infty \right\},$$

where

$$\|f\|_{L_p(I)} = \left(\int_I |f|^p \, dx \right)^{1/p} . [27]$$

Definition 2.1.2 :

Let $0 < p \le \infty$, $\varphi(x) = \sqrt{1 - x^2}$ and for $r \in \mathbb{N}_0$,we have

$$L^r_{p,\varphi}(\mathcal{W}I) = \{ f \mid \ f^{(r)}\varphi^r \in L_{\mathcal{W},p}(I) \}, \qquad r \ge 1$$

where $L^0_{p,\varphi}(\mathcal{W}) = L_{\mathcal{W},p}$.

Definition 2.1.3 : [27]

For $k \in \ \mathbb{N}_0$, $h \ge 0$, $x \in [-1,1]$, and $f : [-1,1] \ \mapsto \mathbb{R}$, let

$$\Delta^k_h(f, x, [-1,1]) = \begin{cases} \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} f\left(x - \frac{kh}{2} + ih\right) & if \ \left[x - \frac{kh}{2}, x + \frac{kh}{2}\right] \subseteq [-1,1] \\ 0 & otherwise \, , \end{cases}$$

be the kth symmetric difference, and let $\Delta^k_h(f, x) = \Delta^k_h(f, x, [-1,1])$.

By (2.1.2) , we define our new weights as follow :

Definition 2.1.4:

For $j \in \mathbb{N}, \delta > 0, M \ge 2, \varphi(x) = \sqrt{1 - x^2}, \lambda j \in J_\delta, x \in (-1,1)$, we have :

$$\mathcal{W}_{\delta}^{J}(x) = \prod_{j=1}^{M} \left| x - z_j - \frac{\delta\varphi(x)}{M} \right|^{\lambda j} \quad . \qquad (2.1.3)$$

So we define our new weighted modulus of smoothness is given as follow :

Definition 2.1.5 :

For $k \in \mathbb{N}, h > 0, f \in L_{p,\varphi}^{r}$ and $x \in \mathfrak{D}_{\delta}$, we define the generalized Jacobian weighted modulus of smoothness as follow

$$\omega_{k,r}^{\varphi}\left(f^{(r)}, \delta\right)_{p} = \sup_{0 \le h \le \delta} \left\| \mathcal{W}_{kh}^{J}(x) \Delta_{h\varphi(x)}^{k}\left(f^{(r)}, x\right) \right\|_{p} \quad , \quad (2.1.4)$$

where

$$\mathfrak{D}_{\delta} = \left\{ x \mid 1 - \frac{\delta\varphi(x)}{M} \ge |x| \right\} \setminus \{\pm 1\},$$

and

$$\Delta_{h\varphi(x)}^{k}(f, x, [-1,1]) = \begin{cases} \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} f\left(x - \frac{kh\varphi}{M} + ih\varphi\right) & if \left[x - \frac{kh}{M}, x + \frac{kh}{M}\right] \subseteq [-1,1] \\ 0 & otherwise . \end{cases}$$

Remarks 2.1.6

- Note that, when $\delta = 0$, our improved generalized Jacobian weight (2.1.3) returns to the generalized Jacobi weight (2.1.2).
- Moreover, when $\delta = 0, M = 2$, (2.1.3) returns to (1.3.1) (the classical Jacobian weight).
- If $J = \{\alpha, \beta\}$, with $M = 2$, (2.1.4) returns to (2.1.1), so we conclude a general case of (2.1.4).

For the proofs of equivalence to the K-functional, we can be helped by the following averaged moduli of smoothness.

Definition 2.1.7

Let $k \in \mathbb{N}, r \in \mathbb{N}_0, 0 < p < 1$ and $f \in L^r_{p,\varphi}$,

$$\omega^{*\varphi}_{k,r}\left(f^{(r)},\delta\right)_p = \left(\frac{1}{t}\int_0^t \int_{\mathfrak{D}_{kh}} \left|\mathcal{W}^J_{kh}(x)\Delta^k_{\tau\varphi(x)}(f^{(r)},x)\right|^p dx d\tau\right)^{1/p},$$

Definition 2.1.8 [26]

Let $k \in \mathbb{N}, r \in \mathbb{N}_0, 1 \le p < \infty$ and $f \in L^r_{p,\varphi}$,

$$\omega^{*k}_{\varphi}\left(f^{(r)},\delta\right)_{w,p} = \left(\frac{1}{t}\int_0^t \int_{-1+t^*}^{1-t^*} \left|w(x)\Delta^k_{\tau\varphi(x)}\left(f^{(r)},x\right)\right|^p dx d\tau\right)^{\frac{1}{p}}$$

$$+\left(\frac{1}{t^*}\int_0^{t^*} \int_{-1}^{-1+At^*} \left|w(x)\vec{\Delta}^k_{u\varphi(x)}\left(f^{(r)},x\right)\right|^p dx du\right)^{\frac{1}{p}}$$

$$+\left(\frac{1}{t^*}\int_0^{t^*} \int_{1-At^*}^{1} \left|w(x)\overleftarrow{\Delta}^k_{u\varphi(x)}(f^{(r)},x)\right|^p dx du\right)^{1/p},$$

where $t^* = 2k^2t^2$, A is a constant and $w = \omega_{\alpha,\beta}\varphi^r$.

The role of K-functional comes from the relationship among moduli, properties of GJWMS and for function approximation too..

Definition 2.1.9 : [27]

For $k \in \mathbb{N}, r \in \mathbb{N}_0$ and $f \in L^r_{p,\varphi}(\omega_{\alpha,\beta})$, $0 < p \le \infty$, define

$$K_{k,r}^{\varphi}\left(f^{(r)},\delta^k\right)_{\alpha,\beta,p} = \inf_{g\in L_{p,\varphi}^{k+r}(\omega_{\alpha,\beta})}\left\{\left\|\omega_{\alpha,\beta}\varphi^r\left(f^{(r)}-g^{(r)}\right)\right\|_p + \delta^k\left\|\omega_{\alpha,\beta}\varphi^{k+r}g^{(k+r)}\right\|_p\right\}.$$

Kopoton in his paper [27] studied all cases of p, he confined g to be a polynomial for the case of $0<p<1$ since K-functional may become identically zero.

To have a closer look to the modulus of smoothness, we give the following example for the modulus of smoothness with General Jacobian Weight of a function from $L_{p,\varphi}^{r}[-1,1]$.

Example 2.1.10

Let $f(x)=x^2+1, M=3$ the general Jacobian weight $\mathcal{W}_\delta^J(x)$ can be estimated with $J=\{1,1,1\}$, and taking the maximum value of h as δ so that the modulus of smoothness of order $k=2$ of f is as follow

$$\omega_2^{\varphi}(f,\delta)_p^p = \sup_{0\le h\le\delta}\left\|\mathcal{W}_\delta^J(x)\Delta_{h\varphi(x)}^2(f,x)\right\|_p^p$$

$$= \sup_{0\le h\le\delta}\int_{-1}^{1}\left(\left(x^3-\delta\sqrt{(1-x^2)}x^2+\frac{\delta\sqrt{1-x^2}}{3}+\frac{\delta^2x(1-x^2)}{3}-x-\frac{\delta^3\left(\sqrt{1-x^2}\right)^3}{27}\right)\left(2h^2(1-x^2)\right)\right)^p dx$$

$$\le \int_{-1}^{1}\left(x^3-\delta\sqrt{(1-x^2)}x^2+\frac{\delta\sqrt{1-x^2}}{3}+\frac{\delta^2x(1-x^2)}{3}-x-\frac{\delta^3\left(\sqrt{1-x^2}\right)^3}{27}\right)\left(2\delta^2(1-x^2)\right)dx$$

$$= \int_{-1}^{1} \left(2x^3\delta^2(1-x^2)\right) - 2\delta^3 x^2 (1-x^2)^{\frac{3}{2}}$$

$$+ \frac{2\delta^3\sqrt{(1-x^2)}}{3} - \frac{2\delta^3 x^2\sqrt{(1-x^2)}}{3} + \frac{2\delta^4 x(1-x^2)^2}{3}$$

$$- 2x\delta^2 + 2x^3\delta^2 - \frac{1}{27}\left(2\delta^5(1-x^2)^2\sqrt{(1-x^2)}\right) dx$$

When Integral the previous amount, we will get inverse trigonometric functions, through which we get the following result:

$$= 0 - \frac{\pi}{8}\delta^3 + \frac{\pi}{3}\delta^3 - \frac{\pi}{12}\delta^3 + 0 - 0 + 0 - \frac{5\pi}{216}\delta^5$$

$$= \frac{-3\pi + 8\pi - 2\pi}{24}\delta^3 - \frac{5\pi}{216}\delta^5$$

$$= \frac{3\pi}{24}\delta^3 - \frac{5\pi}{216}\delta^5 \ ,$$

for $\delta < 1$, for example $\delta = 0.1$ we get :

$\omega_2^{\varphi}(f,\delta)_p^p \leq \ 0.00039$.

2.2 Properties of General Jacobian Weight

Here are some general properties of Jacobi weight (1.3.1) from [26], they are useful to our work.

Lemma 2.2.1:

Kopotun [26] studied the properties of φ in his set $\left\{x \mid 1 - \frac{\delta\varphi(x)}{2} \geq |x|\right\} \setminus \{\pm 1\}$, by using some similar steps, for $\mathfrak{D}_\delta = \left\{x \mid \ 1 - \frac{\delta\varphi(x)}{M} \geq |x|\right\} \setminus \{\pm 1\}$ $M \geq 2$, we get :

i. $(1-x) \leq 2(1-u)$ and

$$(1+x) \leq 2(1+u) \qquad , \ if \ \ u \ \in \ [min\{0,x\}, max\{0,x\}]$$

ii. $\varphi(x) \le \varphi(u) \quad , if\ |u| \le |x| \le 1 \quad , u \in [min\{0,x\}, max\{0,x\}]$ and $x \in \mathfrak{D}_\delta$

iii. $\delta|\acute{\varphi}(x)| \le 1$ for $x \in \mathfrak{D}_\delta$

iv. $If\ y(x) = x + \frac{\delta_1 \varphi(x)}{M}$ and $|\delta_1| \le \delta$ then $y'(x) \le \frac{1}{2}$ for all $x \in \mathfrak{D}_\delta$

v. $If\ \delta_1 > \delta_2$ then $\mathfrak{D}_{\delta_1} \subset \mathfrak{D}_{\delta_2}$

The following lemma relates our improved generalized Jacobi weight $\mathcal{W}$ given in (2.1.3) to the classical Jacobian weight (1.3.1), we find that $\mathcal{W}$ is equivalent to φ .

Lemma 2.2.2:

i. $\mathcal{W}_\delta^J(x) \le \varphi(u)$ for $x \in \mathfrak{D}_\delta$ and $u \in \left[-|x| - \frac{\delta\varphi(x)}{M}, |x| + \frac{\delta\varphi(x)}{M}\right]$

ii. $\mathcal{W}_\delta^J(x) \le \varphi(x)$ for $x \in \mathfrak{D}_\delta$

iii. $\varphi^M(x) \le M^M \mathcal{W}_\delta^J(x) \quad for\ x \in \mathfrak{D}_\delta$

Proof :

i) By **lemma 2.2.1 (ii)**

$$\varphi^2(u) - \mathcal{W}_\delta^2(x) \ge \varphi^2\left(|x| + \frac{\delta\varphi(x)}{M}\right) - \prod_{j=1}^{M} \left|x - z_j - \frac{\delta\varphi(x)}{M}\right|^{2\lambda j}$$

$$= 1 - \left(|x| + \frac{\delta\varphi(x)}{M}\right)^2 - \prod_{j=1}^{M} \left|x - z_j - \frac{\delta\varphi(x)}{M}\right|^{2\lambda j}$$

$$\ge 0 \qquad , \forall\, x \in \mathfrak{D}_\delta .$$

Proof of **(ii)** is a special case of (i) when $u = x$, that is if $x \in \mathfrak{D}_\delta$,then :

$$\mathcal{W}_\delta^J(x) \le \varphi(x) .$$

iii) Let $z_j \in I_j = [z_{j-1}, z_j]$, then

Since $1 + |x| \ge |x - z_j|$,

but for $M \ge 2$ we get :

$$1 + |x| \leq M|x - z_j| \ ,$$

and

$$1 - |x| \leq M|x - z_j|$$

So

$$\varphi^2(x) \leq M^2 \left| x - z_j - \frac{\delta\varphi(x)}{M} \right|^2$$

$$\prod_{j=1}^{M} \varphi^2(x) \leq \prod_{j=1}^{M} M^2 \left| x - z_j - \frac{\delta\varphi(x)}{M} \right|^{2\lambda_j}$$

$$\varphi^{2M}(x) \leq \left(M^M \prod_{j=1}^{M} |x - z_j - \frac{\delta\varphi(x)}{M}|^{\lambda_j} \right)^2$$

$$\varphi^M(x) \leq M^M \mathcal{W}_\delta^J(x) \ . \quad \blacksquare$$

More properties are proved in the next lemma , we get an equivalence between $\mathcal{W}_\delta^J$ and $\omega_{\alpha,\beta}$.

Lemma 2.2.3 : For $x \in \mathfrak{D}_\delta$, $\alpha, \beta \in J_\delta$ and $M \geq 2$:

i. $\mathcal{W}_\delta^J(x) \leq M^{|\alpha|+|\beta|}\omega_{\alpha,\beta}(x)$

ii. $\omega_{\alpha,\beta}(x) \leq M^{|\alpha|+|\beta|}\mathcal{W}_\delta^J(x)$

Proof :

i) Let $\lambda j = \{\alpha, \beta, 0,0,0,0,0, \ldots\}$, then

$$\mathcal{W}_\delta^J(x) = \prod_{j=1}^{M} \left| x - z_j - \frac{\delta\varphi(x)}{M} \right|^{\lambda j}$$

$$= \left| x - z_j - \frac{\delta\varphi(x)}{M} \right|^{\alpha} \left| x - z_j - \frac{\delta\varphi(x)}{M} \right|^{\beta}$$

$$\leq \left|x-1-\frac{\delta\varphi(x)}{M}\right|^{\alpha}\left|x+1-\frac{\delta\varphi(x)}{M}\right|^{\beta}$$

$$=\left|1-\left(x-\frac{\delta\varphi(x)}{M}\right)\right|^{\alpha}\left|1+\left(x-\frac{\delta\varphi(x)}{M}\right)\right|^{\beta}$$

$$=\ \omega_{\alpha,0}\left(x-\frac{\delta\varphi(x)}{M}\right)\omega_{0,\beta}\left(x-\frac{\delta\varphi(x)}{M}\right)$$

$$\leq\ M^{|\alpha|}\omega_{\alpha,0}(x)M^{|\beta|}\omega_{0,\beta}(x)$$

$$=M^{|\alpha|+|\beta|}\omega_{\alpha,\beta}(x)\,.$$

ii) $\omega_{\alpha,\beta}(x)\ =\omega_{\alpha,0}(x)\omega_{0,\beta}(x)$

$$\leq\ M^{|\alpha|}\omega_{\alpha,0}\left(x-\frac{\delta\varphi(x)}{M}\right)M^{|\beta|}\omega_{0,\beta}\left(x-\frac{\delta\varphi(x)}{M}\right)$$

$$=M^{|\alpha|+|\beta|}\mathcal{W}_{\delta}^{J}(x)\ .$$

2.3 Properties of Generalized Weighted Modulus of Smoothness

Now , we begin studying properties of our improved generalized weighted modulus of smoothness (2.1.4) in the following lemmas:

Lemma 2.3.1:

Let $k\in\mathbb{N}, r\in\mathbb{N}_0$, $0<p<1, f,g\in L_{p,\varphi}^{r}$ then for all $\delta>0$:

i. $\omega_{k,r}^{\varphi}\left(f^{(r)}+g^{(r)},\delta\right)_p\leq c(p)\left(\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p+\omega_{k,r}^{\varphi}\left(g^{(r)},\delta\right)_p\right)$

ii. $\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p\leq c(p,k)\left\|f^{(r)}\right\|_p$

iii. $\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)\ \leq c(p,k)\ \left\|\omega_{\alpha,\beta}f^{(r)}\right\|_p$

iv. $\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)\leq\ \omega_{k,r}^{\varphi}\left(f^{(r)},\acute{\delta}\right)$, for $\delta\leq\acute{\delta}$

v. $\omega_{k,r}^{\varphi}\left(f^{(r)},\gamma\delta\right)\leq(1+\gamma)^{k}\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)$, for $0<\gamma<1$

vi. $\omega_{k,r}^{*\varphi}\left(f^{(r)},\delta\right)_{\alpha,\beta,p}\leq c(p,k)\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_{\alpha,\beta,p}$

Where c is a positive constant depends $p\,,k$

Proof :

i) $\omega_{k,r}^{\varphi}\left(f^{(r)}+g^{(r)},\delta\right)_p = \sup_{0\le h\le\delta}\left\|\mathcal{W}_{kh}^{J}(x)\Delta_{h\varphi(x)}^{k}(f^{(r)}+g^{(r)},x)\right\|_p$

$$= \sup_{0\le h\le\delta}\left\|\mathcal{W}_{kh}^{J}(x)\left(\Delta_{h\varphi(x)}^{k}f^{(r)}+\Delta_{h\varphi(x)}^{k}g^{(r)},x\right)\right\|_p$$

By quasi-triangle inequality of $\|\cdot\|_p$, $when\ 0<p<1$ we have :

$$\omega_{k,r}^{\varphi}\left(f^{(r)}+g^{(r)},\delta\right)_p$$

$$\le c\left(\sup_{0\le h\le\delta}\left\|\mathcal{W}_{kh}^{J}(x)\Delta_{h\varphi(x)}^{k}(f^{(r)},x)\right\|_p+\sup_{0\le h\le\delta}\left\|\mathcal{W}_{kh}^{J}(x)\Delta_{h\varphi(x)}^{k}(g^{(r)},x)\right\|_p\right)$$

$$= c(p)\left(\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p+\omega_{k,r}^{\varphi}\left(g^{(r)},\delta\right)_p\right)$$

ii) $\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p = \sup_{0\le h\le\delta}\left\|\mathcal{W}_{kh}^{J}(x)\Delta_{h\varphi(x)}^{k}(f^{(r)},x)\right\|_{L_p(\mathfrak{D}_{kh})}$

$$= \sup_{0\le h\le\delta}\left\{\int_{\mathfrak{D}_{kh}}\left|\mathcal{W}_{kh}^{J}(x)\sum_{i=0}^{k}\binom{k}{i}(-1)^{k-i}f^{(r)}\left(x+\left(i-\frac{k}{M}\right)h\varphi(x)\right)\right|^p dx\right\}^{1/p}$$

$$\le\left(\sum_{i=0}^{k}\left|\binom{k}{i}\right|^p\int_{\mathfrak{D}_{kh}}\sup_{0\le h\le\delta}\left|\mathcal{W}_{kh}^{J}(x)f^{(r)}\left(x+\left(i-\frac{k}{M}\right)h\varphi(x)\right)\right|^p dx\right)^{1/p}$$

$$\le\ c\,(p,k)\left\|f^{(r)}\right\|_{L_p(\mathfrak{D}_{kh})}$$

iii) $\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p = \sup_{0\le h\le\delta}\left\|\mathcal{W}_{kh}^{J}(x)\Delta_{h\varphi(x)}^{k}(f^{(r)},x)\right\|_{L_p(\mathfrak{D}_{kh})}$

from (ii) and Lemma 2.2.3(i) we get :

$$\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)\le c(p,k)\left\|\omega_{\alpha,\beta}f^{(r)}\right\|_p .$$

iv) Since $\mathcal{W}_{\delta}^{J}(x)$ is monotone non-decreasing with respect to δ, then

$$\begin{aligned}\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p &= \sup_{0\le h\le\delta}\left\|\mathcal{W}_{kh}^{J}(x)\Delta_{h\varphi(x)}^{k}(f^{(r)},x)\right\|_p \\ &\le \sup_{0\le h\le\acute{\delta}}\left\|\mathcal{W}_{kh}^{J}(x)\Delta_{h\varphi(x)}^{k}(f^{(r)},x)\right\|_p \\ &= \omega_{k,r}^{\varphi}\left(f^{(r)},\acute{\delta}\right)_p .\end{aligned}$$

v) Noting that $\gamma\delta\le\delta$, v comes immediately from (iv)

$$\omega_{k,r}^{\varphi}\left(f^{(r)},\gamma\delta\right)_p \le \omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p \le (1+\gamma)^k\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p$$

vi) It is clear by **Definition 2.1.5** , **Definition 2.1.7** and **Lemma 2.2.3.(ii)** that

$$\omega_{k,r}^{*\varphi}\left(f^{(r)},\delta\right)_{\alpha,\beta,p} \le c(p,k)\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_{\alpha,\beta,p} \quad . \blacksquare$$

The most important property of any modulus, is its convergence to zero because it is strongest guarantee of approximation, so we prove that in the next lemma, we follow some assumptions from Kopotun [26] , but with essential different techniques due to our modulus of smoothness .

Lemma 2.3.2 :

If $r\in\mathbb{N}_0, 0<p<1$ and $f\in L_{p,\varphi}^{r}$, then

$$\lim_{\delta\to0^+}\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p = 0$$

Proof:

If $\epsilon>0$ then $\exists\ \delta>0$ such that

$$\int_{[-1,1]\backslash\mathfrak{D}_\delta}\left|\varphi^r(x)f^{(r)}(x)\right|^p dx < \left(\frac{\epsilon}{2^{k+2}}\right)^p \text{ [27]}$$

Set

$$g^{(r)}(x)=\begin{cases} f^{(r)}(x) & , \quad if\ x\in\mathfrak{D}_\delta \\ 0 & , \quad otherwise \end{cases}$$

Since $g^{(r)}\in L_P[-1,1\,]$, $\exists\ \delta_0>0$ such that

$$\omega_k^{\varphi}\left(g^{(r)},\delta\right)_p<\frac{\epsilon}{2}\,, \quad 0<\delta\leq\delta_0$$

For each $h>0$ by **Lemma 2.3.1 (i)** ,we have

$$\left\|\mathcal{W}_{kh}^{j}(x)\Delta_{h\varphi(x)}^{k}\left(f^{(r)},x\right)\right\|_p$$

$$\leq c(p)\left(\left\|\mathcal{W}_{kh}^{j}(x)\Delta_{h\varphi(x)}^{k}\left(g^{(r)},x\right)\right\|_p\right.$$
$$\left.+\left\|\mathcal{W}_{kh}^{j}(x)\Delta_{h\varphi(x)}^{k}\left(f^{(r)}-g^{(r)},x\right)\right\|_p\right)$$

$$\leq c(p)\left(\left\|\Delta_{h\varphi(x)}^{k}\left(g^{(r)},x\right)\right\|_p+\left\|\mathcal{W}_{kh}^{j}(x)\Delta_{h\varphi(x)}^{k}\left(f^{(r)}-g^{(r)},x\right)\right\|_p\right)$$

$$\leq c(p)\left(\frac{\epsilon}{2}+\sum_{i=0}^{k}\binom{k}{i}\left(\left(\int_{\mathfrak{D}_{kh}}(\mathcal{W}_{kh}^{j}(x)\left|f^{(r)}(x+\left(i-\frac{k}{M}\right)h\varphi(x)\right)-g^{(r)}(x\right.\right.\right.$$
$$\left.\left.\left.+\left(i-\frac{k}{M}\right)h\varphi(x)\right|^p\,dx\right)^{\frac{1}{p}}\right)$$

By **lemma 2.2.2 (i)** and letting $u(x)=x+\left(i-\frac{k}{M}\right)h\varphi(x)$, we get from $M\geq2$ and monotonicity of φ that:

$$\left\|\mathcal{W}_{kh}^{j}(x)\Delta_{h\varphi(x)}^{k}\left(f^{(r)},x\right)\right\|_p\leq$$

$$c(p)\left(\frac{\epsilon}{2}+\textstyle\sum_{i=0}^{k}\binom{k}{i}\left(\int_{\mathfrak{D}_{kh}}\left(\varphi^r\left(x+\left(i-\frac{k}{M}\right)h\varphi(x)\right)\right)\left|f^{(r)}\left(x+\left(i-\right.\right.\right.\right.\right.$$
$$\left.\left.\left.\left.\frac{k}{M}\right)h\varphi\right)-g^{(r)}\left(x+\left(i-\frac{k}{M}\right)h\varphi\right|^p\,dx\right)^{\frac{1}{p}}\right)$$

$$\leq c(p)\left(\frac{\epsilon}{2}+2\sum_{i=0}^{k}\binom{k}{i}\left(\int_{-1}^{1}\left(\varphi^{r}(u)\left|f^{(r)}(u)-g^{(r)}(u)\right|\right)^{p}du\right)^{\frac{1}{p}}\right)$$

$$\leq c(p)\left(\frac{\epsilon}{2}+2\sum_{i=0}^{k}\binom{k}{i}\left(\int_{[-1,1]\setminus\mathfrak{D}_{\delta}}\left|\varphi^{r}(u)f^{(r)}(u)\right|^{p}\right)^{\frac{1}{p}}\right)\leq\frac{\epsilon}{2}+\frac{\epsilon}{2}\leq\epsilon$$

Then the proof is done .■

Kopotun [26] proved the following property of φ for $1 \leq p \leq \infty$, later Sharba and Bhaya proved the same property for $0 < p < 1$ in [29]. It is important to guarantee φ .

Lemma 2.3.3: [26][29]

Let $0 < p \leq \infty$ and $r \in \mathbb{N}_0$, if $g \in L_{p,\varphi}^{r+1}$ then

$$\left\|\varphi^{\gamma} g^{(r)}\right\|_{p} < \infty$$

for any $\gamma \geq 0$ such that $\gamma > r-1$.

The following property is always true for previous defined moduli of smoothness , it is true here to for clear .

Lemma 2.3.4: [26] [29]

Let $k \in \mathbb{N}$, $r \in \mathbb{N}_0$, $0 < p \leq \infty$,if $f \in \mathbb{B}_p^r$ then for $\delta \geq 2/k$

$$\omega_{k,r}^{\varphi}\left(f^{(r)},2/k\right)_{p} = \omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_{p}$$

Lemma 2.3.5

If $k \in \mathbb{N}$, $r \in \mathbb{N}_0$, $\alpha,\beta \in J_p$, $0 < p < \infty$ and $f \in L_{p,\varphi}^{r}$ then

$$\omega_{\varphi}^{*k}\left(f^{(r)},\delta\right)_{\omega_{\alpha,\beta}\varphi^{r},p} \leq c(k,r,\alpha,\beta)\,\omega_{k,r}^{*\varphi}\left(f^{(r)},c(k)\delta\right)_{\alpha,\beta,p} \quad ,0<\delta<c(k)$$

Proof:

For $0 < p < 1$, by **Definition 2.1.7** and **Definition 2.1.8**

$$\omega_{\varphi}^{*k}\left(f^{(r)},\delta\right)_{\omega_{\alpha,\beta}(x)\varphi^r(x),p} = \left(\frac{1}{t}\int_0^t \int_{-1+t^*}^{1-t^*} \left|\omega_{\alpha,\beta}(x)\varphi^r(x)\Delta_{\tau\varphi(x)}^k\left(f^{(r)},x\right)\right|^p dxd\tau\right)^{\frac{1}{p}}$$

$$+\left(\frac{1}{t^*}\int_0^{t^*} \int_{-1}^{-1+At^*} \left|\omega_{\alpha,\beta}(x)\varphi^r(x)\vec{\Delta}_{u\varphi(x)}^k\left(f^{(r)},x\right)\right|^p dxdu\right)^{\frac{1}{p}}$$

$$+\left(\frac{1}{t^*}\int_0^{t^*} \int_{1-At^*}^{1} \left|\omega_{\alpha,\beta}(x)\varphi^r(x)\overleftarrow{\Delta}_{u\varphi(x)}^k(f^{(r)},x)\right|^p dxdu\right)^{1/p}$$

Now, by **lemma 2.2.3 (ii)** and **lemma 2.2.2(iii)** we get:

$$\omega_{\varphi}^{*k}\left(f^{(r)},\delta\right)_{\omega_{\alpha,\beta}(x)\varphi^r(x),p}$$

$$\leq c(k,r,\alpha,\beta)\left(\frac{1}{t}\int_0^t \int_{-1+t^*}^{1-t^*} \left|\mathcal{W}_{kh}^J(x)\Delta_{\tau\varphi(x)}^k\left(f^{(r)},x\right)\right|^p dxd\tau\right)^{\frac{1}{p}}$$

$$+\left(\frac{1}{t^*}\int_0^{t^*} \int_{-1}^{-1+At^*} \left|\mathcal{W}_{kh}^J(x)\vec{\Delta}_{u\varphi(x)}^k\left(f^{(r)},x\right)\right|^p dxdu\right)^{\frac{1}{p}}$$

$$+\left(\frac{1}{t^*}\int_0^{t^*} \int_{1-At^*}^{1} \left|\mathcal{W}_{kh}^J(x)\overleftarrow{\Delta}_{u\varphi(x)}^k(f^{(r)},x)\right|^p dxdu\right)^{1/p}$$

So, we obtain that:

$$\omega_{\varphi}^{*k}\left(f^{(r)},\delta\right)_{\omega_{\alpha,\beta}\varphi^r,p} \leq c(k,r,\alpha,\beta)\omega_{k,r}^{*\varphi}\left(f^{(r)},c(k)\delta\right)_{\alpha,\beta,p}.$$

we get the last inequality by the same steps from [19]

The following lemma is true from [27] , for $p \geq 1$.

Lemma 2.3.6 [27]

If $k \in \mathbb{N}, r \in \mathbb{N}_0, \alpha, \beta \in J_\delta, 1 \leq p < \infty$ and $f \in L^r_{p,\varphi}$, then

i. $K_{k,\varphi}(f, \delta^k)_{w,p} \leq c\omega^{*k}_{\varphi}(f, \delta)_{w,p} \quad , \quad 0 < \delta < \delta_0$

ii. $K^{\varphi}_{k,r}\left(f^{(r)}, \delta^k\right)_{\alpha,\beta,p} \leq \left\|\omega_{\alpha,\beta}\varphi^r f^{(r)}\right\|_p < \infty$

Using **Lemma 2.2.3** we obtain that **Lemma 2.3.6** is valid for $0 < p < 1$.

2.4 Equivalence between K-functional and GJWMS

In this section, we discuss the relationship between our General Jacobian Weighted Modulus of Smoothness and the K-functional of Koputon [27]. This relationship may open the doors wide to improve the degree of best approximation in terms of our above modulus.

Theorem 2.4.1 :

let $0 < p < 1, k \in \mathbb{N}, r \in \mathbb{N}_0, f \in L^r_{p,\varphi}$ *and for* $0 < \delta \leq 2/k$ we get :

$$c(k,r,p)K^{\varphi}_{k,r}\left(f^{(r)}, \delta^k\right)_p \leq \omega^{*\varphi}_{k,r}\left(f^{(r)}, \delta\right)_p \leq \omega^{\varphi}_{k,r}\left(f^{(r)}, \delta\right)_p$$
$$\leq c(k,r,p)K^{\varphi}_{k,r}\left(f^{(r)}, \delta^k\right)_p .$$

Proof of :

The Upper estimate of Theorem 2.4.1 :

If $k \in \mathbb{N}, r \in \mathbb{N}_0, 0 < p < 1$ and $f \in L^r_{p,\varphi}$, then

$$\omega^{\varphi}_{k,r}\left(f^{(r)}, \delta\right)_p \leq c(k,r,p)K^{\varphi}_{k,r}\left(f^{(r)}, \delta^k\right)_p \qquad , \forall\, \delta > 0$$

By lemma 2.3.5 and monotonicity of K-functional, we suppose that $\delta \geq 2/k$ and take $g \in L^{r+1}_{p,\varphi}$, so we get $g \in L^r_{p,\varphi}$ from Lemma 2.3.4, wherever:

$$\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p \leq \omega_{k,r}^{\varphi}\left(f^{(r)}-g^{(r)},\delta\right)_p + \omega_{k,r}^{\varphi}\left(g^{(r)},\delta\right)_p$$

Let $0 < h \leq \delta$ and $y_i := x + \left(i - \frac{k}{M}\right) h\varphi(x)$ for $0 \leq i \leq k$ and $M \geq 2$ .From **Lemma 2.2.1(ii)** we get $\acute{y_i}(x) \leq 1/2$ for $x \in \mathfrak{D}_{kh}$.

For $0 < p < 1$

$$\omega_{k,r}^{\varphi}\left(f^{(r)}-g^{(r)},\delta\right)_p = \sup_{0\leq h\leq\delta} \left\|\mathcal{W}_{kh}^{J}(y_i)\Delta_{h\varphi(x)}^{k}(f^{(r)}-g^{(r)},x)\right\|_p^p$$

$$= \sup_{0\leq h\leq\delta} \int_{\mathfrak{D}_{kh}} \left|\sum_{i=0}^{k}\binom{k}{i}\mathcal{W}_{kh}^{J}(y_i)\left(f^{(r)}(y_i)-g^{(r)}(y_i)\right)\right|^p dx$$

Since $\mathcal{W}_{kh}^{J}(y) \leq M^{|\alpha|+|\beta|}\omega_{\alpha,\beta}(y)$ for $y \in \left[x - \frac{\delta\varphi(x)}{M}, x + \frac{\delta\varphi(x)}{M}\right]$ and $0 < \delta \leq 2$

We get:

$$\omega_{k,r}^{\varphi}\left(f^{(r)}-g^{(r)},\delta\right)_p \leq$$

$$M^{|\alpha|+|\beta|} \sup_{0\leq h\leq\delta} \left|\sum_{i=0}^{k}\binom{k}{i}\right|^p \int_{\mathfrak{D}_{kh}} \varphi^{rp}(y_i)\varphi^{-rp}(y_i)\omega_{\alpha,\beta}^{p}(y_i)\left|f^{(r)}(y_i) - g^{(r)}(y_i)\right|^p dx$$

$$= M^{|\alpha|+|\beta|} \sup_{0\leq h\leq\delta} \left|\sum_{i=0}^{k}\binom{k}{i}\right|^p \sup_{0\leq i\leq k}|\varphi^{-r}(y_i)|\left\|\varphi^{r}(y_i)\omega_{\alpha,\beta}(y_i)\left|f^{(r)}(y_i) - g^{(r)}(y_i)\right|\right\|_p^p$$

$$= c(p,k)\left\|\varphi^{r}\omega_{\alpha,\beta}\left(f^{(r)}-g^{(r)}\right)\right\|_p^p .$$

To estimate the second term $\omega_{k,r}^{\varphi}\left(g^{(r)},\delta\right)_p$ we can use the identity:

$$\Delta_h^k(f\,,x) = \int_{-\frac{h}{M}}^{\frac{h}{M}} \dots \int_{-\frac{h}{M}}^{\frac{h}{M}} f^{(k)}(x+u_1+\cdots+u_k)\, d_{u_1} \dots d_{u_k}, [26]$$

we have

$$\omega_{k,r}^{\varphi}\left(g^{(r)},\delta\right)_p = \sup_{0\le h\le\delta} \left\| \mathcal{W}_{kh}^{J}(\cdot)\Delta_{h\varphi(x)}^{k}(g^{(r)},\cdot)\right\|_{L_p(\mathfrak{D}_{kh})}$$

$$= \sup_{0\le h\le\delta} \left\| \mathcal{W}_{kh}^{J} \int_{-\frac{h\varphi}{M}}^{\frac{h\varphi}{M}} \dots \int_{-\frac{h\varphi}{M}}^{\frac{h\varphi}{M}} g^{(k+r)}(x+u_1+\cdots+u_k)\, d_{u_1} \dots d_{u_k} \right\|_{L_p(\mathfrak{D}_{kh})}$$

By **Lemma 2.2.3(i)** we get

$\omega_{k,r}^{\varphi}\left(g^{(r)},\delta\right)_p \le$

$$M^{|\alpha|+|\beta|} \sup_{0\le h\le\delta} \left\| \int_{-\frac{h\varphi}{M}}^{\frac{h\varphi}{M}} \dots \int_{-\frac{h\varphi}{M}}^{\frac{h\varphi}{M}} \left(\omega_{\alpha,\beta}\, g^{(k+r)}\right)(x+u_1+\ldots+u_k)\, d_{u_1} \dots d_{u_k} \right\|_{L_p(\mathfrak{D}_{kh})}$$

$$\le\ M^{|\alpha|+|\beta|}\, c(p) \sup_{0\le h\le\delta} \int_{-\frac{h\varphi}{M}}^{\frac{h\varphi}{M}} \dots \int_{-\frac{h\varphi}{M}}^{\frac{h\varphi}{M}} \left\|\left(\omega_{\alpha,\beta}\, g^{(k+r)}\right)(x\right.$$

$$\left. +\,u_1+\ldots+u_k)\right\|_{L_p(\mathcal{A}(x,u))} d_{u_1} \dots d_{u_k}$$

For each u satisfying $-1 < x+u-\frac{h\varphi(x)}{M} < x+u+\frac{h\varphi(x)}{M} <\ 1$, we have

$$\int_{-\frac{h\varphi}{M}}^{\frac{h\varphi}{M}} \left\|\omega_{\alpha,\beta}\, g^{(k+r)}(x+u+u_k)\right\|_p d_{u_k}$$

$$= \int_{x+u-\frac{h\varphi}{M}}^{x+u+\frac{h\varphi}{M}} \left\| \varphi^{-(k+r)}(a)\varphi^{k+r}(a)\omega_{\alpha,\beta}g^{(k+r)}(a) \right\|_{L_p(\mathcal{A}(x,u))} da$$

$$\leq M \sup_{a\in\mathcal{A}(x,u)} \left|\varphi^{-(k+r)}(a)\right| \int_{x+u-\frac{h\varphi}{M}}^{x+u+\frac{h\varphi}{M}} \left\| \varphi^{k+r}(a)\omega_{\alpha,\beta}g^{(k+r)}(a) \right\|_{L_p(\mathcal{A}(x,u))} da,$$

where

$$\mathcal{A}(x,u) := \left[x + u - \frac{h\varphi(x)}{M}, x + u + \frac{h\varphi(x)}{M}\right]$$

To complete the proof, we have

$$M^{|\alpha|+|\beta|} \sup_{0\leq h\leq\delta} \int_{\mathfrak{D}_{kh}} \left(\int_{-\frac{h\varphi}{M}}^{\frac{h\varphi}{M}} \cdots \int_{-\frac{h\varphi}{M}}^{\frac{h\varphi}{M}} \sup_{u_1u_2\ldots u_k} \left|\varphi^{-(k+r)}\right| \left\|\varphi^{k+r}\omega_{\alpha,\beta}g^{(k+r)}\right\|_p d_{u_1} \ldots d_{u_k} \right)^p dx$$

$$\leq c(k,r,p)\left\|\varphi^{k+r}\omega_{\alpha,\beta}g^{(k+r)}\right\|_p^p .$$

So

$$\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p \leq \omega_{k,r}^{\varphi}\left(f^{(r)} - g^{(r)},\delta\right)_p + \omega_{k,r}^{\varphi}\left(g^{(r)},\delta\right)_p$$

$$\leq c(p)\left(\left\|\omega_{\alpha,\beta}\varphi^{r}\left(f^{(r)} - g^{(r)}\right)\right\|_p + \delta^k\left\|\omega_{\alpha,\beta}\varphi^{k+r}g^{(k+r)}\right\|_p\right) . \blacksquare$$

The next prove ,we follow some assumptions from Kopotun [27] , but with essential different techniques due to our modulus of smoothness .

The Lower Estimate of Theorem 2.4.1

let $0 < p < 1$ *and if* $k \in \mathbb{N}, r \in \mathbb{N}_0, f \in L_{p,\varphi}^r$ *then*

$$K_{k,r}^{\varphi}\left(f^{(r)},\delta^k\right)_{\alpha,\beta,p} \leq c(p)\omega_{k,r}^{*\varphi}\left(f^{(r)},\delta\right)_p \leq c(p)\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p \qquad (2.4.1)$$

Proof :

By **lemma 2.3.6 (i)** with weight $w = \omega_{\alpha,\beta}\varphi^{r}$ and **lemma 2.3.5** we obtain , for $0 < p < 1$,

$$K_{k,r}^{\varphi}\left(f^{(r)},\delta^{k}\right)_{\alpha,\beta,p} = K_{k,\varphi}\left(f^{(r)},\delta^{k}\right)_{\omega_{\alpha,\beta}\varphi^{r},p} \leq c\omega_{\varphi}^{*k}\left(f^{(r)},\delta\right)_{\omega_{\alpha,\beta}\varphi^{r},p}$$

$$\leq c\omega_{k,r}^{*\varphi}\left(f^{(r)},c(k)\delta\right)_{p} \qquad , \qquad 0 < \delta < c(k).$$

Hence, we have

$$K_{k,r}^{\varphi}\left(f^{(r)},\delta^{k}\right)_{\alpha,\beta,p} \leq c\omega_{k,r}^{*\varphi}\left(f^{(r)},c(k)\delta\right)_{p} \qquad , \qquad 0 < \delta < c(k) , \tag{2.4.2}$$

Now we assume that $0 < \delta \leq 2/\delta$, and let $M = \max\{1, c_1, 2/kc(k)\}$ then from **(2.4.2)** we obtain

$$K_{k,r}^{\varphi}\left(f^{(r)},\delta^{k}\right)_{\alpha,\beta,p} \leq M^{k}K_{k,r}^{\varphi}\left(f^{(r)},\left(\frac{\delta}{M}\right)^{k}\right)_{\alpha,\beta,p} \leq c\omega_{k,r}^{*\varphi}\left(f^{(r)},\frac{c(k)\delta}{M}\right)_{p}$$

$$\leq c\omega_{k,r}^{*\varphi}\left(f^{(r)},\delta\right)_{p},$$

then by **Lemma 2.3.1 (vi)** we get :

$$c\omega_{k,r}^{*\varphi}\left(f^{(r)},\delta\right)_{p} \leq c\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_{p} . \quad \blacksquare$$

Chapter Three

Neural Network Approximation

3.1 Introduction and Preliminaries

Neural networks, also known as artificial neural networks (ANN) or simulated neural networks (SNNs), are a subsets of machine learning and lie at the heart of deep learning algorithms.

Neural networks perform their tasks like neurons in the human brain using certain algorithms and those algorithms recognize patterns hidden in the raw data, divide and classify them into groups. Over time, these networks learn and gradually improve their performance.

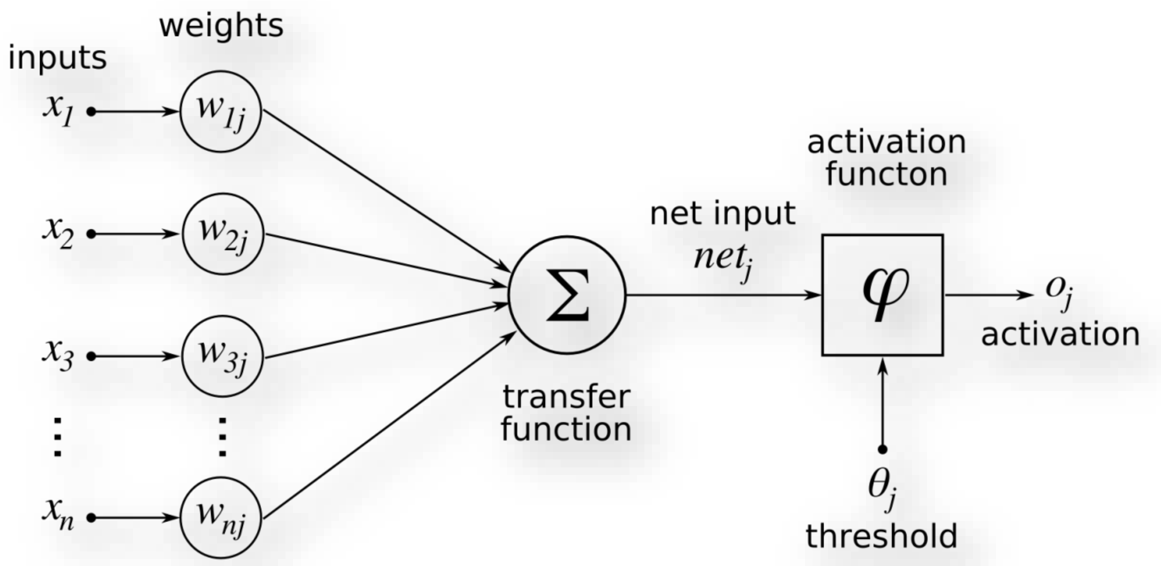


Figure 3. 1 Neural networks

Artificial Neural Networks (ANNs) contains several layers of neurons, an input layer $\boldsymbol{x} = (x_1, \dots \dots, x_d) \in \mathbb{R}^d$, one or more hidden layers with an activation function σ, and an output layer. The general mathematical formula of any neural network is:

$$N_d(\boldsymbol{x}) = \sum_{j=1}^{d} c_j \sigma\left(w_j x_j + b_j\right) \quad ,$$

where

w_j are the weights, $c_j \in \mathbb{R}$ are coffients and $b_j \in \mathbb{R}$ are bases.

The range of functions that are approximated by neural networks is very wide, due to their applications in different fields, they can approximate any function with some conditions, such as continuity, integrability and sufficient training sets.

To understand more about the relationship of neural networks in the approximation of functions, we first need to the activation function and its importance in neural networks. In neural networks we have neurons, each neuron receives inputs and performs weighted summation operations on them, then passes the resulted summation into the activation function. That turns it into an output. The question now is, what if no activation function is used and the neurons are allowed to give the sum of data to the inputs as well as to the outputs?. In this case, the calculation will be very simple because the weighted sum of the inputs has no range. Hence, an important use of activation functions is to keep the output data restricted to a certain range. Its activation functions help neural networks learn complex relationships in data. Another use of the activation function is to add nonlinearity to the data. Nonlinear functions are always chosen as the activation functions.

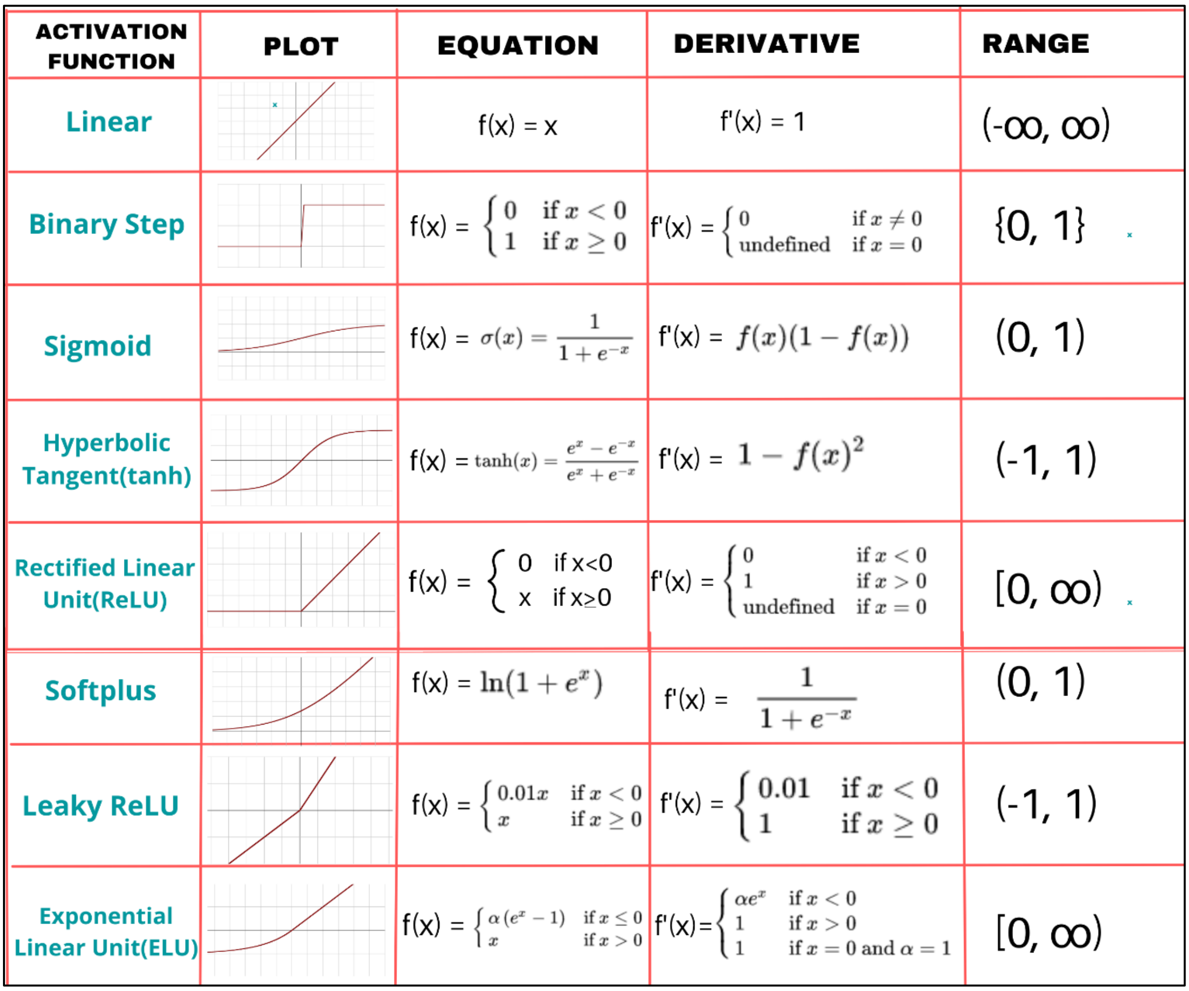
After we have shown the importance of the activation function, the universal approximation theory [30] will tell us about the relationship between the neural network and the approximation of functions, it states as follow:

“The standard multi-layer feed-forward network with one hidden layer, which includes an infinite number of hidden neurons, is a universal approximation between continuous functions in the subgroups that are compact in $\mathbb{R}^n$ ”. [30]

In many papers, the degree of best approximation is given by modulus of continuity such as [31] [32], or modulus of smoothness of order two [33].

A few little papers studied the degree of best neural approximation with kth order modulus of smoothness such as [34].

There are many activation functions(see Figure 3.2) , each type has advantages and disadvantages, as well as the way it works.



ACTIVATION FUNCTION	PLOT	EQUATION	DERIVATIVE	RANGE
Linear		f(x) = x	f'(x) = 1	(-∞, ∞)
Binary Step		f(x) = $\begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$	f'(x) = $\begin{cases} 0 & \text{if } x \neq 0 \\ \text{undefined} & \text{if } x = 0 \end{cases}$	{0, 1}
Sigmoid		f(x) = $\sigma(x) = \frac{1}{1+e^{-x}}$	f'(x) = $f(x)(1 - f(x))$	(0, 1)
Hyperbolic Tangent(tanh)		f(x) = $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	f'(x) = $1 - f(x)^2$	(-1, 1)
Rectified Linear Unit(ReLU)		f(x) = $\begin{cases} 0 & \text{if } x<0 \\ x & \text{if } x \geq 0 \end{cases}$	f'(x) = $\begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \\ \text{undefined} & \text{if } x = 0 \end{cases}$	[0, ∞)
Softplus		f(x) = $\ln(1 + e^x)$	f'(x) = $\frac{1}{1+e^{-x}}$	(0, 1)
Leaky ReLU		f(x) = $\begin{cases} 0.01x & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$	f'(x) = $\begin{cases} 0.01 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$	(-1, 1)
Exponential Linear Unit(ELU)		f(x) = $\begin{cases} \alpha(e^x - 1) & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases}$	f'(x)= $\begin{cases} \alpha e^x & \text{if } x < 0 \\ 1 & \text{if } x > 0 \\ 1 & \text{if } x = 0 \text{ and } \alpha = 1 \end{cases}$	[0, ∞)

Figure 3. 2 Types of Activation Functions

Since to the drawbacks of some activation functions, we define a new formula, as follow :

Definition 3.1.1

For any $x \in [-1,1]$, define

$$R(x) = \begin{cases} 0 & if\ x \leq 0 \\ 1 - \varphi(x) & if\ x > 0 \end{cases} \quad , \qquad (3.1.1)$$

where $\varphi(x) = \sqrt{1 - x^2}$.

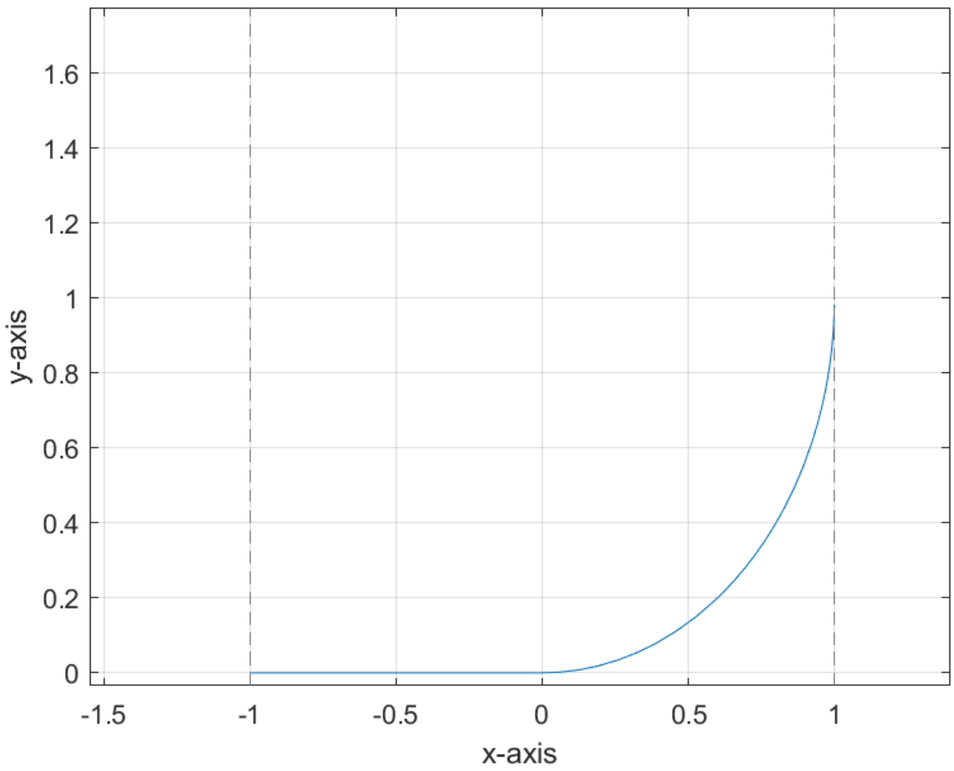


Figure 3. 3 Activation function R(x)

By the above activation function, we define a new neural network :

Definition 3.1.2

Let $x \in [-1,1]$ and $f \in L^r_{p,\varphi}$

$$N_{d,r}(f^{(r)}, \boldsymbol{x}) = \sum_{j=1}^{d} \sum_{i=1}^{k} \binom{k}{i} (-1)^{k-i} f^{(r)}(h\varphi(a)) R(\boldsymbol{x}) \ , \qquad (3.1.2)$$

where $a \in [-1,1]$ and $\boldsymbol{x} = \omega_j x_j + b_j$.

Set $\Omega_{d,r}$ to be the family of all neural networks of degree d of any function from $\mathbb{B}^r_p$ of the form (3.1.2).

The degree of best approximation by neural network in **(3.1.2)** is given by:

Definition 3.1.3:

Let $d \in \mathbb{N}, r \in \mathbb{N}_0, 0 < p < 1$ and $f \in L^r_{p,\varphi}$

$$E_n(f)_p = \inf_{N\in \Omega_{d,r}} \left\| f^r(x) - N_{d,r}(f^{(r)}, \boldsymbol{x}) \right\|_p.$$

3.2 Activation Function Properties

We study the following properties of R that are important in neural networks and approximation

Lemma 3.2.1

1. The domain and rang of R are both $[-1,1]$
2. $|\ R(x)| \leq 1$, $\qquad \forall\ x \in [-1,1]$
3. $\|\ R(x)\|_p \leq (2)^{\frac{1}{p}}$, $\qquad \forall\ x \in [-1,1]$
4. R is differentiable and the first derivative of R is:

$$R'(x) = \frac{x}{\varphi(x)}$$

where $\varphi(x) = \sqrt{1 - x^2}$ and $x \in (-1,1)$.

5. The general form of derivatives of R is:

$$R^{(m)}(x) = \sum_{i=0}^{m} \frac{cx^i}{(\varphi(x))^{m+i-1}}\ \delta_{im} \quad ,$$

where

$$\delta_{im} = \begin{cases} 1 & , \quad \text{if both } i \text{ and } m \text{ are even or both are odd} \\ 0 & , \quad o.w. \end{cases}$$

Lemma 3.2.2

Let $x \in (-1,1)$ and $0 < p < 1$, then it's clear that from (3.1.1)

$$\|\ R'(x)\|_p^p \leq c(p).$$

On the other hand, the norm of the general form of derivatives of R is unbounded.

lemma 3.2.3

Let $x \in (-1,1)$, $0 \ < p < 1$ and $m \geq 2$

$$\left\| R^{(m)}(x) \right\|_p^p \geq c$$

where c is positive constant.

Proof:

$$\left\| \sum_{i=0}^{m} \frac{c x^i}{\left(\phi(x)\right)^{m+i-1}} \ \delta_{im} \right\|_{\boldsymbol{p}}^{\boldsymbol{p}}$$

$$= \int_{-1}^{1} \left| \sum_{i=0}^{m} \frac{c x^i}{\left(\phi(x)\right)^{m+i-1}} \ \delta_{im} \ \right|^p dx$$

$$\leq \sum_{i=0}^{m} c \int_{-1}^{1} \left| \frac{x^i}{\left(\phi(x)\right)^{m+i-1}} \right|^p dx$$

Sine

$$\int_{-1}^{1} \left| \frac{x^i}{\left(\phi(x)\right)^{m+i-1}} \right|^p dx$$

is unbounded, then $\left\| R^{(m)}(x) \right\|_p^p \geq c$.■

3.3 Auxiliary Results

Lemma 3.3.1

Let α , $\beta \ \in J_p$, $r \in \mathbb{N}_0$, $0 < p < 1$ and $f \in L_{p,\varphi}^r$,then there exists $N \in \Omega_{d,r}$, s.t.

$$\left\| \omega_{\alpha,\beta}(x) \varphi^r(x) \left(f^{(r)}(x) - N_{d,r}(f^{(r)}, \boldsymbol{x}) \right) \right\|_p^p \leq c(p) \left\| f^{(r)}(x) - N_{d,r}(f^{(r)}, \boldsymbol{x}) \right\|_p^p$$

Proof:

$$\left\|\omega_{\alpha,\beta}(x)\varphi^{r}(x)\left(f^{(r)}(x)-N_{d,r}(f^{(r)},\boldsymbol{x})\right)\right\|_{p}^{p}$$

$$=\int_{-1}^{1}\left|\omega_{\alpha,\beta}(x)\varphi^{r}(x)\left(f^{(r)}(x)-N_{d,r}(f^{(r)},x)\right)\right|^{p}dx$$

From **lemma 2.2.2 (iii) and lemma 2.2.3 (ii)** we get:

$$\int_{-1}^{1}\left(\omega_{\alpha,\beta}(x)\varphi^{r}(x)\left(f^{(r)}(x)-N_{d,r}(f^{(r)},\boldsymbol{x})\right)\right)^{p}dx$$

$$\leq\int_{-1}^{1}\left(c\left(\mathcal{W}_{\delta}^{J}(x)\right)^{2}\left(f^{(r)}(x)-N_{d,r}(f^{(r)},\boldsymbol{x})\right)\right)^{p}dx$$

then by **lemma 2.2.2 (i)** we get:

$$\leq c(p)\int_{-1}^{1}\left(\left(\varphi(u)\right)^{2}\left(f^{(r)}(x)-N_{d,r}(f^{(r)},\boldsymbol{x})\right)\right)^{p}dx$$

$$\leq c(p)\left\|\left(f^{(r)}(x)-N_{d,r}(f^{(r)},\boldsymbol{x})\right)\right\|_{p}^{p}\ .\ \blacksquare$$

Lemma 3.3.2

Let $\alpha,\beta\in J_{p}$, $r\in\mathbb{N}_{0}$, $0<p<1$, $M\leq n$ and $f\in L_{p,\varphi}^{r}$, then for $N\in\Omega$

$$\left\|\omega_{\alpha,\beta}\varphi^{r}N_{d,r}(f^{(r)})\right\|_{p}^{p}\leq c\left\|f^{(r)}\right\|_{p}^{p}$$

Proof:

$$\left\|\omega_{\alpha,\beta}\varphi^{r}N_{d,r}(f^{(r)})\right\|_{p}^{p}=$$

$$\int_{-1}^{1}\left|\omega_{\alpha,\beta}(x)\varphi^{r}(x)N_{d,r}(f^{(r)},\boldsymbol{x})\right|^{p}dx$$

From **Lemma 2.2.2 (iii) and Lemma 2.2.3 (ii)** we get :

$$\leq \int_{-1}^{1} \left| c \left(\mathcal{W}_{\delta}^{J}(x) \right)^{2} N_{d,r}(f^{(r)}, \boldsymbol{x}) \right|^{p} dx$$

$$= c \int_{-1}^{1} \left| \left(\mathcal{W}_{\delta}^{J}(x) \right)^{2} \sum_{j=1}^{d} \sum_{i=1}^{k} \binom{k}{i} (-1)^{k-i} f^{(r)}(h\varphi(a)) R(\boldsymbol{x}) \right|^{p} dx$$

$$\leq c \int_{-1}^{1} \left| \left(\mathcal{W}_{\delta}^{J}(x) \right)^{2} \sum_{i=1}^{k} \binom{k}{i} (-1)^{k-i} f^{(r)}(h\varphi(a))(1 - \varphi(x)) \right|^{p} dx$$

$$= c \int_{-1}^{1} \left| \left(\mathcal{W}_{\delta}^{J}(x) \right)^{2} \sum_{i=1}^{k} \binom{k}{i} (-1)^{k-i} f^{(r)}(h\varphi(a))(\varphi(x) - 1) \right|^{p} dx$$

In the last inequality, we use the fact that $\varphi(x) - 1 \leq \mathcal{W}_{\delta}^{J}(x)$

$$\leq c \int_{-1}^{1} \left| \left(\mathcal{W}_{\delta}^{J}(x) \right)^{2} \sum_{i=1}^{k} \binom{k}{i} (-1)^{k-i} f^{(r)}(h\varphi(a)) \mathcal{W}_{\delta}^{J}(x) \right|^{p} dx$$

$$= c \int_{-1}^{1} \left| \left(\mathcal{W}_{\delta}^{J}(x) \right)^{3} \sum_{i=1}^{k} \binom{k}{i} (-1)^{k-i} f^{(r)}(h\varphi(a)) \right|^{p} dx$$

And then by **lemma 2.2.2 (i)** we get:

$$\leq c(p) \int_{-1}^{1} \left| (\varphi(u))^3 \sum_{i=1}^{k} \binom{k}{i} (-1)^{k-i} f^{(r)}(h\varphi(a)) \right|^p dx$$

$$= c \left\| f^{(r)} \right\|_p^p . \ \blacksquare$$

Lemma 3.3.3

For any $k \in \mathbb{N}, r \in \mathbb{N}_0, 0 < p < 1, f \in L_{p,\varphi}^r$

$$\left\| N_{d,r}(f^{(r)}, \boldsymbol{x}) \right\|_p^p \leq c\omega_{k,r}^{\varphi}\left(f^{(r)}, \delta\right)_p^p$$

Proof:

Let $x \in [-1,1]$

$$\left\| N_{d,r}(f^{(r)}, \boldsymbol{x}) \right\|_p^p$$

$$= \int_{-1}^{1} \left| \sum_{j=1}^{d} \sum_{i=1}^{k} \binom{k}{i} (-1)^{k-i} f^{(r)}(h\varphi(a)) R(\boldsymbol{x}) \right|^p dx$$

$$\leq \int_{-1}^{1} \left| \sum_{i=1}^{k} \binom{k}{i} (-1)^{k-i} f^{(r)}(h\varphi(a))(\varphi(x) - 1) \right|^p dx$$

In the next inequality, we use again the fact that $\varphi(x) - 1 \leq \mathcal{W}_\delta^J(x)$

$$\leq \int_{-1}^{1} \left| \mathcal{W}_\delta^J(x) \sum_{i=1}^{k} \binom{k}{i} (-1)^{k-i} f^{(r)}(h\varphi(a)) \right|^p dx$$

$= c\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p^p$ ■

We also need the following lemma that is immediately comes from the same techniques of **Lemma 3.3.1** with **Lemma 2.2.3 (i)** $\left(c\mathcal{W}_{\delta}^{J}(x) \le \omega_{\alpha,\beta}(x)\right)$ and **Lemma 2.2.2 (ii)** .

Lemma 3.3.4

Let $\alpha\,,\beta\ \in J_p\ ,r \in \mathbb{N}_0\,,0 < p < 1\,,M \le n$ and $f \in\ L_{p,\varphi}^r$

$$\left\|\left(f^{(r)}(x) - N_{d,r}(f^{(r)},\boldsymbol{x})\right)\right\|_p^p \le \left\|\omega_{\alpha,\beta}(x)\varphi^r(x)\left(f^{(r)}(x) - N_{d,r}\left(f^{(r)},\boldsymbol{x}\right)\right)\right\|_p^p.$$

The following Theorems show the approximation of functions from the spaces $L_{p,\varphi}^r$ to neural networks from (3.1.2).

3.4 $L_{p,\varphi}^r$ Function Approximation by NNs

Theorem 3.4.1

Let $k \in \mathbb{N}, r \in \mathbb{N}_0\,, 0 < p < 1$ then for any $f \in\ L_{p,\varphi}^r$, there exist $N_{d,r} \in \Omega$ in **(3.1.2)**, that satisfies

$$\left\|f^{(r)} - N_{d,r}(f^{(r)})\right\|_p^p \le c(k,p)\omega_{k,r}^{\varphi}\left(f^{(r)}\,,\delta\right)$$

for $\delta > 0$.

Proof:

From **Lemma 3.3.4** we get :

$$\left\|\left(f^{(r)}(x) - N_{d,r}(f^{(r)},\boldsymbol{x})\right)\right\|_p^p \le \left\|\omega_{\alpha,\beta}(x)\varphi^r(x)\left(f^{(r)}(x) - N_{d,r}(f^{(r)},\boldsymbol{x})\right)\right\|_p^p$$

By **Definition 2.1.9** we get :

$$\left\|\omega_{\alpha,\beta}(x)\varphi^{r}(x)\left(f^{(r)}(x)-N_{d,r}(f^{(r)},\boldsymbol{x})\right)\right\|_{p}^{p} \leq K_{k,r}^{\varphi}\left(f^{(r)},\delta^{k}\right)_{\alpha,\beta,p}$$

Then by **Theorem 2.4.1** we obtain:

$$K_{k,r}^{\varphi}\left(f^{(r)},\delta^{k}\right)_{\alpha,\beta,p} \leq c\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right) .$$

So $\left\|\left(f^{(r)}(x)-N_{d,r}(f^{(r)},\boldsymbol{x})\right)\right\|_{p}^{p} \leq c(p,k)\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)$ ■

Theorem 3.4.2

Let $k \in \mathbb{N}, r \in \mathbb{N}_0$, $0 < p < 1$ then for any $f \in L_{p,\varphi}^{r}$, there exist $N_{d,r}$ of form in **definition 3.1.2**, satisfies

$$\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_{p}^{p} \leq c(k,p)\left\|f^{(r)}(x)-N_{d,r}\left(f^{(r)},\boldsymbol{x}\right)\right\|_{p}^{p}$$

Proof:

In this proof, we benefit from the equivalence between modulus (2.1.5) and K- functional to estimate the degree of approximation .

By using **lemmas 3.3.1**, **3.3.2, 3.3.3** we obtain

$$\left\|f^{(r)}(x)\right\|_{p}^{p} \leq \left\|\left(f^{(r)}(x)-N_{d,r}\left(f^{(r)},\boldsymbol{x}\right)\right)\right\|_{p}^{p}+\left\|N_{d,r}(f^{(r)},\boldsymbol{x})\right\|_{p}^{p}$$

$$\leq \left\|\left(f^{(r)}(x)-N_{d,r}\left(f^{(r)},\boldsymbol{x}\right)\right)\right\|_{p}^{p}+c\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_{p}^{p}$$

since

$$K_{k,r}^{\varphi}\left(f^{(r)},\delta^{k}\right)_{\alpha,\beta,p} \leq \left\|f^{(r)}(x)-N_{d,r}\left(f^{(r)},\boldsymbol{x}\right)\right\|_{p}^{p}+\delta^{k}\left\|f^{(r)}(x)-N_{d,r}\left(f^{(r)},\boldsymbol{x}\right)\right\|_{p}^{p}+c(p,k)\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_{p}^{p},$$

then by Theorem 2.4.1 we get :

$$\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_{p}^{p} \leq c(k)\left\|\left(f^{(r)}(x)-N_{d,r}\left(f^{(r)},\boldsymbol{x}\right)\right)\right\|_{p}^{p}+c(p)\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_{p}^{p}$$

$$\omega_{k,r}^{\varphi}\left(f^{(r)},\delta\right)_p^p \le c(k,p)\left\|f^{(r)}(x)-N_{d,r}\left(f^{(r)},\boldsymbol{x}\right)\right\|_p^p \ . \blacksquare$$

In the following theorem, we prove that if a function f has a derivative greater than r, then the function can be approximated by the derivative of the neural network.

Theorem 3.4.3

Let $r \in \mathbb{N}_0$, $0 < p < 1$ and $f \in L_{p,\varphi}^{r+1}$

$$\left\|f^{(r+1)}(x)-\acute{N}(f^{(r+1)},\boldsymbol{x})\right\|_p^p \le c\omega_{k,r}^{\varphi}\left(f^{(r+1)},\frac{1}{n}\right)$$

Proof :

$$\left\|f^{(r+1)}(x)-\acute{N}(f^{(r+1)},\boldsymbol{x})\right\|_p^p$$

$$= \int_{-1}^{1}\left|f^{(r+1)}(x)-\sum_{i=1}^{k}\binom{k}{i}(-1)^{k-i}f^{(r+1)}\left(\frac{h\varphi(a)}{n}\right)\frac{x}{\varphi(x)}\right|^p dx$$

$$\le \sum_{i=1}^{k}\left|\binom{k}{i}(-1)^{k-i}\right|^p \int_{-1}^{1}\left|f^{(r+1)}(x)-f^{(r+1)}\left(\frac{h\varphi(a)}{n}\right)\right|^p \left|\frac{x}{\varphi(x)}\right|^p dx,$$

where we use the fact that $\left|\frac{x}{\varphi(x)}\right|^p \le \ |M\varphi(x)^M|^p$

$$\left\|f^{(r+1)}(x)-\acute{N}(f^{(r+1)},\boldsymbol{x})\right\|_p^p$$

$$\le \sum_{i=1}^{k}\left|\binom{k}{i}(-1)^{k-i}\right|^p \int_{-1}^{1}\left|f^{(r+1)}(x)-f^{(r+1)}\left(\frac{h\varphi(a)}{n}\right)\right|^p |M\varphi(x)^M|^p\, dx$$

By **Lemma 2.2.2 (ii)** we get:

$$\left\|f^{(r+1)}(x)-\acute{N}(f^{(r+1)},\boldsymbol{x})\right\|_p^p$$

$$\leq c \left\| \mathcal{W}_\delta^J \sum_{i=1}^{k} \binom{k}{i} (-1)^{k-i} \left(f^{(r+1)}(x) - f^{(r+1)}\left(\frac{h\varphi(a)}{n}\right)\right) \right\|_p^p$$

$$= c\omega_{k,r}^{\varphi}\left(f^{(r+1)}, x - \frac{h\varphi(a)}{n}\right)$$

By **Lemma 2.3.1 (v)** we get :

$$\leq c\big(1 + nx - h\varphi(a)\big)^k \omega_{k,r}^{\varphi}\left(f^{(r+1)}, \frac{1}{n}\right)$$

$$\leq c\omega_{k,r}^{\varphi}\left(f^{(r+1)}, \frac{1}{n}\right). \blacksquare$$

From **Theorem 3.4.3** we can readily prove the next converse theorem by using the same steps of Theorem **3.4.2** .

Corollary 3.4.4

Let $r \in \mathbb{N}_0$, $0 < p < 1$ and $f \in L_{p,\varphi}^{r+1}$

$$\omega_{k,r}^{\varphi}\left(f^{(r+1)}, \frac{1}{n}\right) \leq c\left\|f^{(r+1)}(x) - \acute{N}\big(f^{(r+1)}, x\big)\right\|_p^p.$$

Chapter four

Function Approximation in Deep Learning

4.1 Introduction of Deep Learning

To solve problems in a more qualified way or use different methods to get the best result, the neural network needs to be trained all the time. So when we introduce new information to the network, it learns how to operate according to the new situation. If the tasks we solve are more difficult, then the learning will be deeper.

In the traditional work of programming, a computer is programmed to do what it does, by breaking down large problems into many small problems that a computer can easily do. On the contrary, in a neural network, we do not program the computer to solve the problem. In lieu, the network learns from the monitoring data and discovers a solution to its problem.

The basic principle of learning neural networks is to reduce the error that results from the difference between the obtained value and the actual value by updating the weights and biases that we have entered at the beginning of the neural network.

As with many problems that NNs solve, function approximation gets benefits from learning to get the best approximation with a better degree of approximation.

According to UAT, learning can be applied in the backward stage so that the results are more accurate from those without learning.

Learning neural networks is being done in three different ways or strategies:

1) **Supervised learning:** Also called the training stage, different patterns of data (visual, audio or text) are recognized, and then the results of processing these data is compared with the desired ideal results through a reverse process in which the data goes backwards from the layers. This process of putting the output as input is called (Backpropagation) and then correct these errors to reach the desired result.

2) **Unsupervised learning:** This learning method is used when processing data for which there is no previous information about its outcome. The network analyzes the data and builds a function to determine the error to reduce it as much as possible to reach the highest possible degree of accuracy.

3) **Reinforcement Learning:** This learning style is based mainly on observation. The information is processed and the results reached, are estimated by the neural network. If it is positive, it is treated each time in the same way, but if it is negative, the network will process it in different ways next time to reach positive results.

4.2 Backpropagation Algorithm

There are many learning algorithms for neural networks, for example : Particle Swarm Optimization algorithm [35] , Genetic Algorithm [36] , GA-PSO algorithm[37] , etc. But the backpropagation algorithm (BP) is the most successful algorithm for training neural networks, in practical tasks .

The BP algorithm is a supervised learning method, which includes two operations: forward propagation and backward propagation of the error. That is, the error output is calculated in the forward propagation process, and then we make an adjustment to the weights and rules in order to reduce this error and this is done in the back propagation process.

BP algorithm consists of three layers, the first layer is a set of data (the input). The second layer is the hidden layer where there is a loop through which the output is calculated. The third layer is a set of data (the output). When the data passes through those layers, it is processed to give what is called the error function, which is the resulting difference between the output we get and the accurate output.

In the following steps, we introduce the main stages of BP algorithm for function approximation.

4.2.1 Backpropagation Algorithm Steps : [38]

i. **Forward Computation** : First of all, a function f is given here with a training example with an input layer denoted by the vector $(x_1, x_2, x_3, \dots \dots, x_m)$ and the required output layer represented by the vector $(t_1, t_2, t_3, \dots \dots, t_m) = \big(f(x_1), f(x_2), \dots \dots \dots, f(x_m)\big)$, the induced local output in layer l for neuron j is:

$$v_j^{(l)}(k) = \sum_{i=1}^{m} \omega_{ji,k}^{(l)}\, y_i^{(l-1)}(k)\,,$$

where $y_i^{(l-1)}(k)$ is the output in the previous layer $l-1$ of neuron i at iteration k.

$\omega_{ji}^{(l)}$ is the weight that is fed from neuron i of neuron j in layer l.

If $l = 1$, the first hidden layer, gives :

$$y_j^0\,(k) = x_j(k),$$

Where $x_j \in \mathbb{R}$ is the jth component of the vector input $(x_1, x_2, x_3, \dots \dots, x_m)$.

By using an activation function R of any type, the output signal of neuron j in layer l is :

$$y_j^{(l)} = R_j\big(v_{j,k}\big).$$

If $l = L$ (where L is the depth of the network), we get :

$$y_j^{(L)}(k) = o_j(k).$$

We compare the obtained results with the actual results. If they do not match, we calculate the value of the difference between them for each neuron of the output layer, which represents the error value:

$$e_j(k) = t_j(k) - o_j(k)\,.$$

If the error value is large, we will reduce the error value by updating the weights value through the next step.

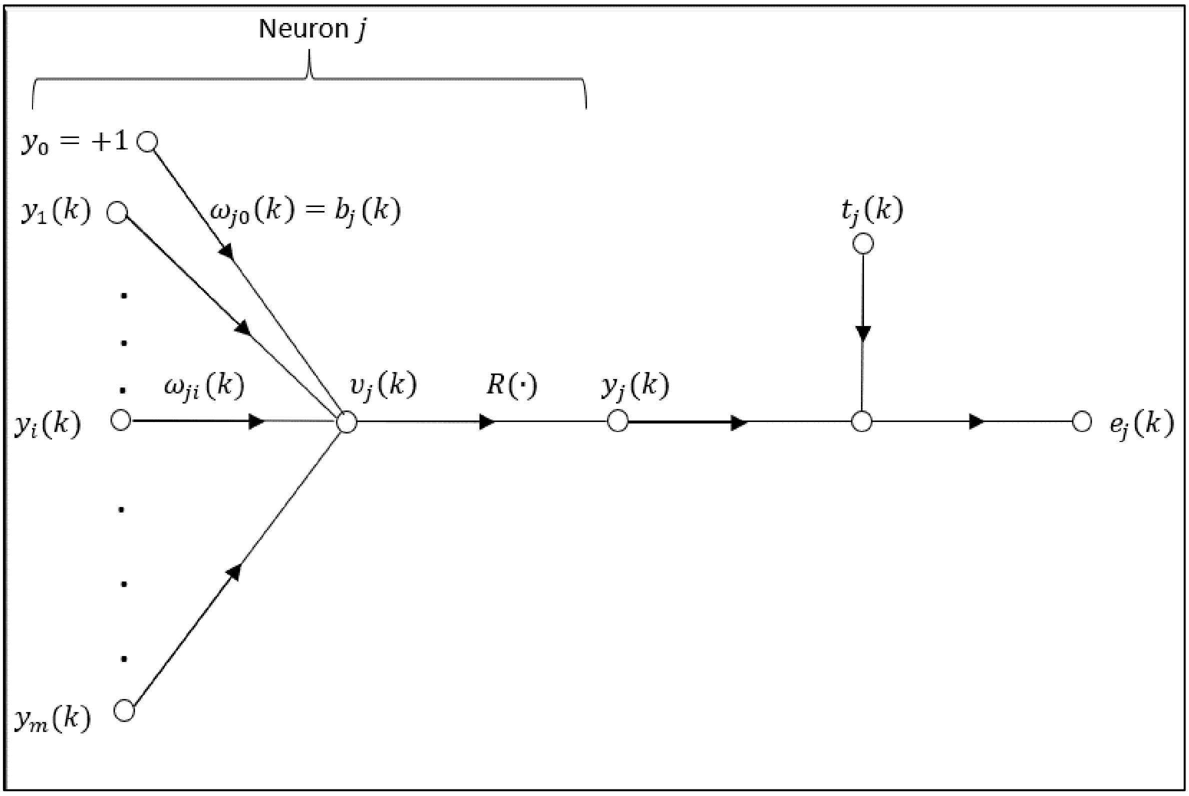


Figure 4. 1 Forward Layer

ii. Backward Computation : This stage begins with calculating $\zeta_j^{(l)}$ (i.e. local gradients) :

$$\zeta_j^{(l)} = \begin{cases} e_j^{(L)}(k)R'\left(v_j^{(L)}(k)\right) , & \text{for neuron } j \text{ in output layer } L \\ R'\left(v_j^{(L)}(k)\right)\sum_{n=1}^{m}\zeta_n^{(l+1)}(k)\omega_{ji}^{(l+1)}(k) , & \text{for neuron } j \text{ in hidden layer } l, \end{cases} \quad \textbf{(4.2.1)}$$

where

R_j' : Derivative of the activation function .

After that, we update the weights by the following rule **:**

$$\omega_{ji}^{(l)}(k+1) = \omega_{ji}^{(l)}(k) + \alpha\,\Delta\omega_{ji}^{(l)}(k+1) + \eta\zeta_j^{(l)}(k)y_i^{(l-1)}(k) ,$$

where

$$\Delta\omega_{ji}^{(l)}(k-1) = \eta\zeta_j^{(l)}(k-1)y_i^{(l-1)}(k-1),$$

η is the learning-rate parameter, and α is constant.

If the value of η is smaller, the change in network weights are smaller from one iteration to the next. Thus, we make η very large in order to speed up the learning rate, as well as for the changes in the network weights to be large . Then we calculate the error and iterations are repeated until we reach the acceptable error value.

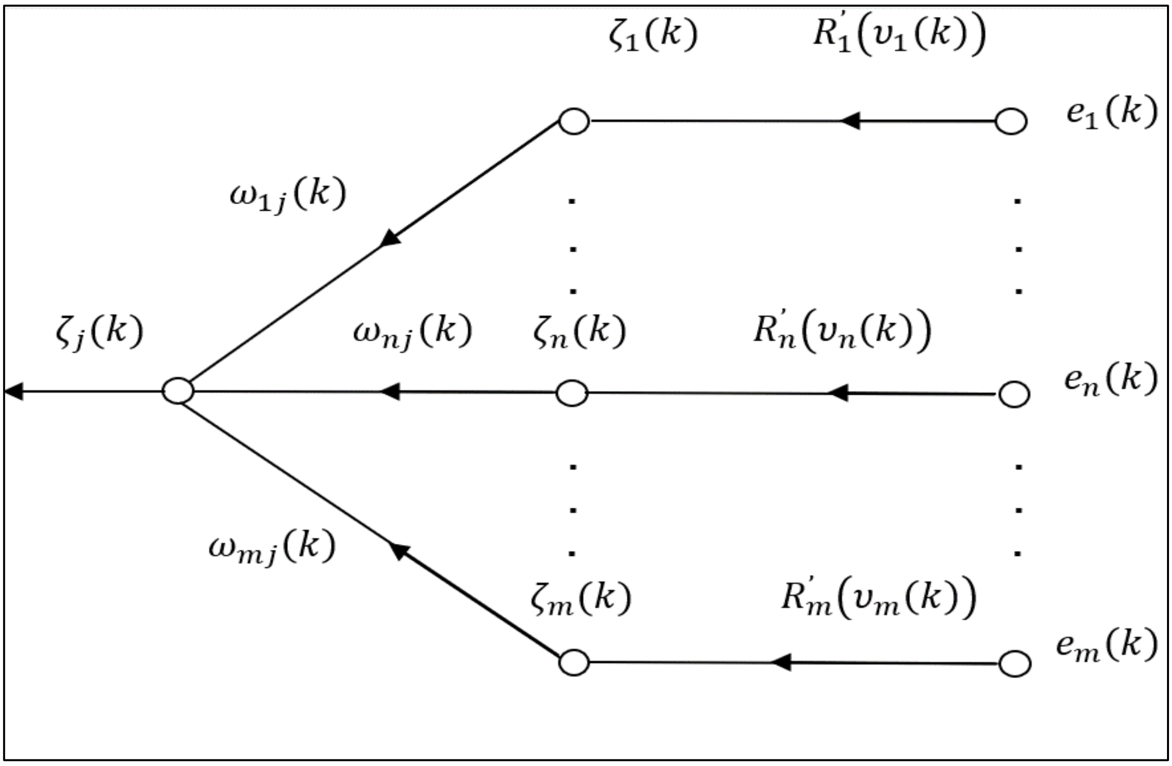


Figure 4. 2 Backward layer

Backpropagation Algorithm

1: Input m , f , x , n

2: for $j = 1 \rightarrow m$ do

3: $t_j = (t_1, t_2, t_3, \ldots \ldots, t_m)$, $t_j = f(x_j)$

4: for $k = 1 \rightarrow n$ do

5: $v_j^{(l)}(k) = \sum_{i=1}^{m} \omega_{ji,k}^{(l)} \, y_i^{(l-1)}(k)$

if $l = 1$ then

$y_j^0(k) = x_j(k)$ input value

for $l = 1 \rightarrow L$ do

if $l = L$ then

$y_j^{(L)}(k) = o_j(k)$ output value

if $e_j(k) = t_j(k) - o_j(k) < \epsilon$ then

go to end

else

$$\zeta_j^{(l)} = \begin{cases} e_j^{(L)}(k) R_j'\left(v_j^{(L)}(k)\right) \\ R_j'\left(v_j^{(L)}(k)\right) \sum_{n=1}^{m} \zeta_n^{(l+1)}(k)\, \omega_{ji}^{(l+1)}(k) \end{cases}$$

$$\omega_{ji}^{(l)}(k+1) = \omega_{ji}^{(l)}(k) + \alpha\, \Delta\omega_{ji}^{(l)}(k+1) + \eta \zeta_j^{(l)}(k)\, y_i^{(l-1)}(k)$$

return to $v_j^{(l)}(k)$

end if

end for

end if

end for

end for

end

4.3 Experimental Results

In this section, we introduce some numerical example for function that are learned to be approximated by NNs with BP algorithm.

4.3.1 Periodic Function

We apply BP algorithm to approximate the function :

$$f(x) = 1 + sin(\frac{\pi}{4}x), \quad x \in [-2,2]$$

as shown in **Figure 4.3**.

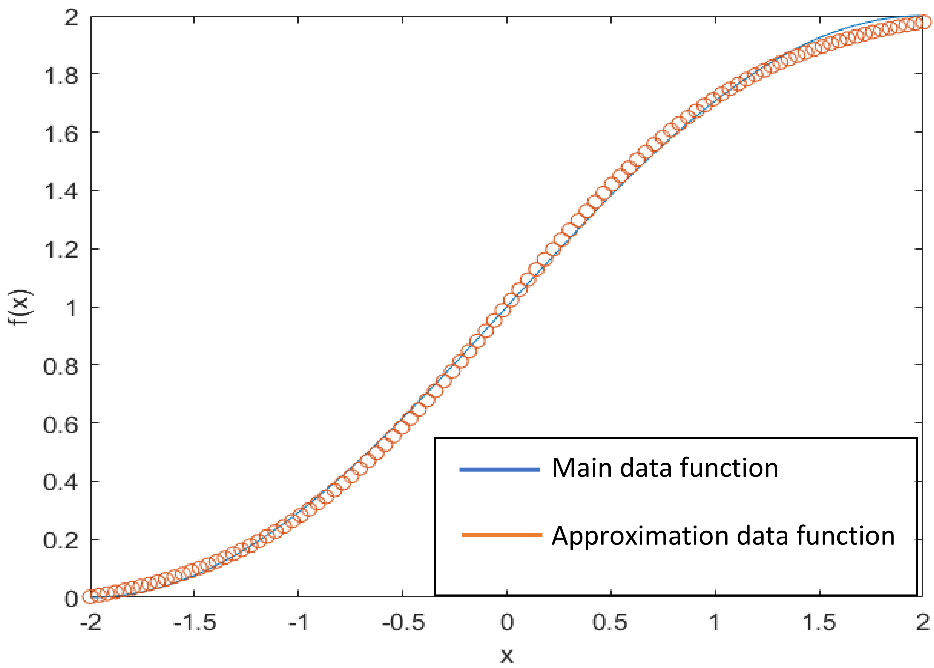


Figure 4. 3 Approximation of the Function in Example 4.3.1

The calculated error are shown in **Figure 4.4**, we notice that it approaches to zero after about 10^3 iteration .

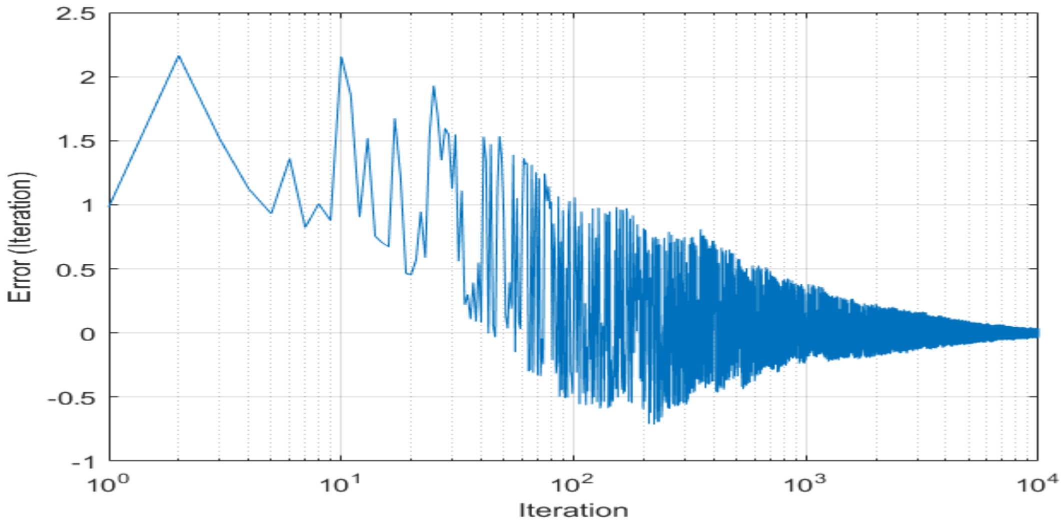


Figure 4. 4 Approximation error value of Example 4.3.1

4.3.2 Quadratic Function

We apply BP algorithm to approximate the function :

$$f(x) = x^2 , \quad x \in [0,1]$$

as shown in **Figure 4.5**.

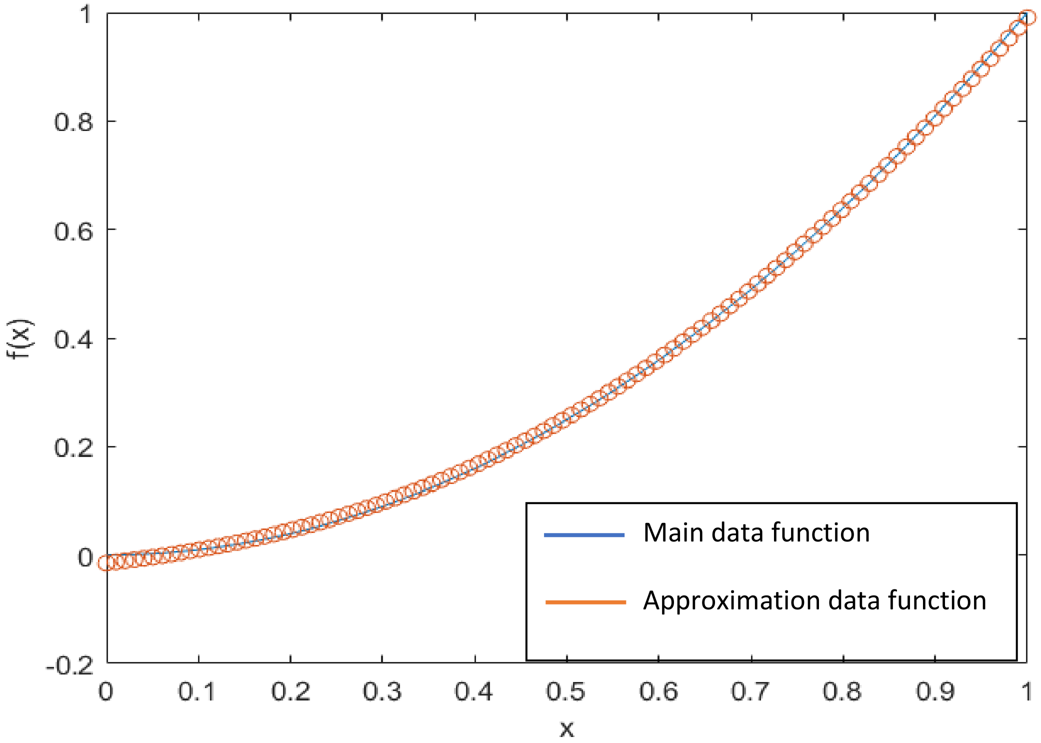


Figure 4. 5 Approximation of the function in Example 4.3.2

The calculated error are shown in **Figure 4.6**, we notice that it approaches to zero after about 10^3 iteration .

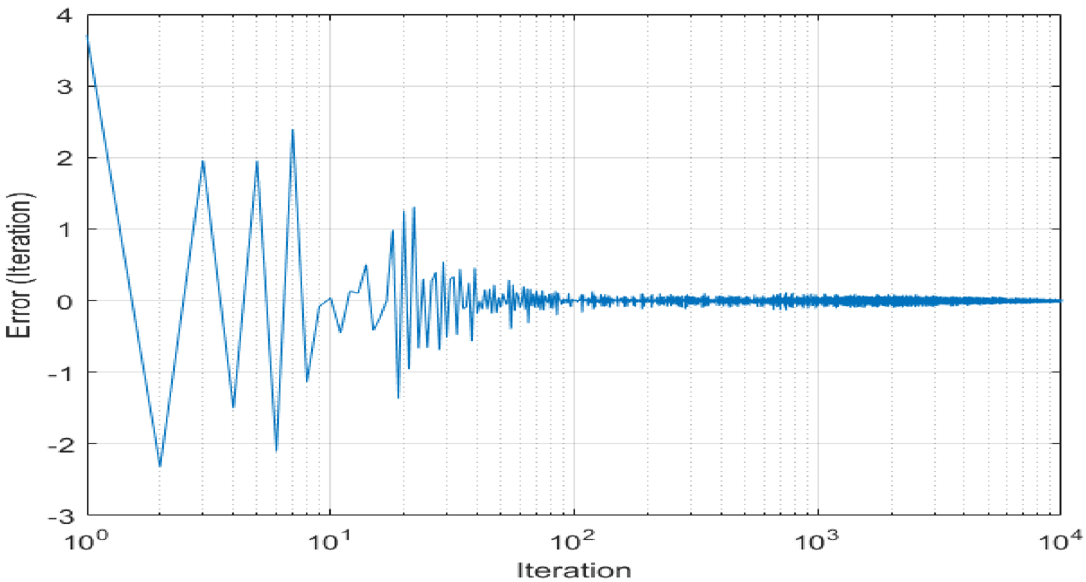


Figure 4. 6 Approximation error value of Example 4.3.2

4.3.3 Exponential Function

We apply BP algorithm to approximate the function :

$$f(x) = e^x, \quad x \in [0,1]$$

as shown in **Figure 4.7**.

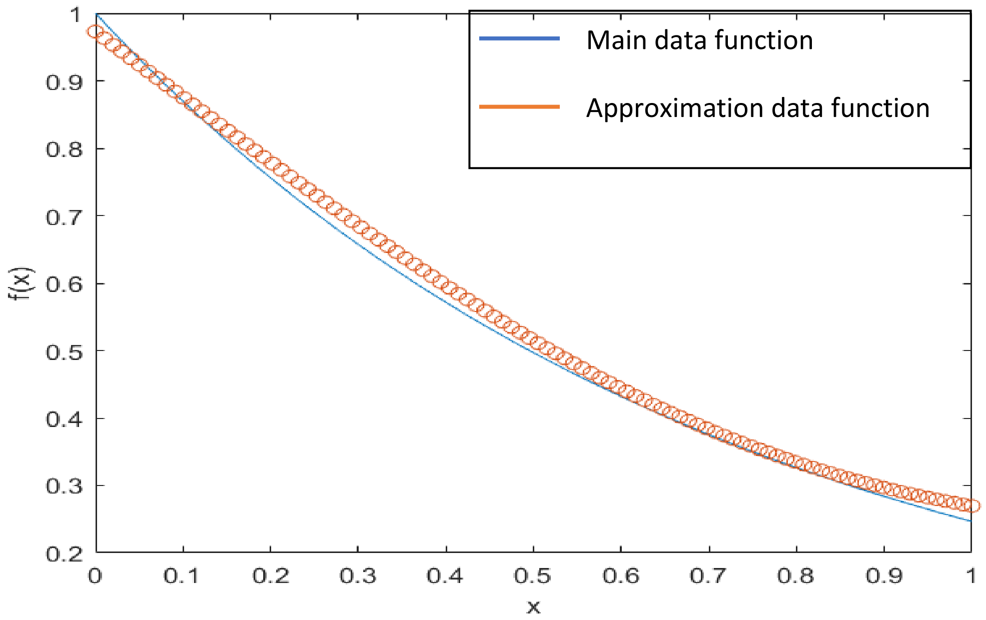


Figure 4. 7 Approximation of the function in Example 4.3.3

The calculated error are shown in **Figure 4.8**, we notice that it approaches to zero after about 10^3 iteration .

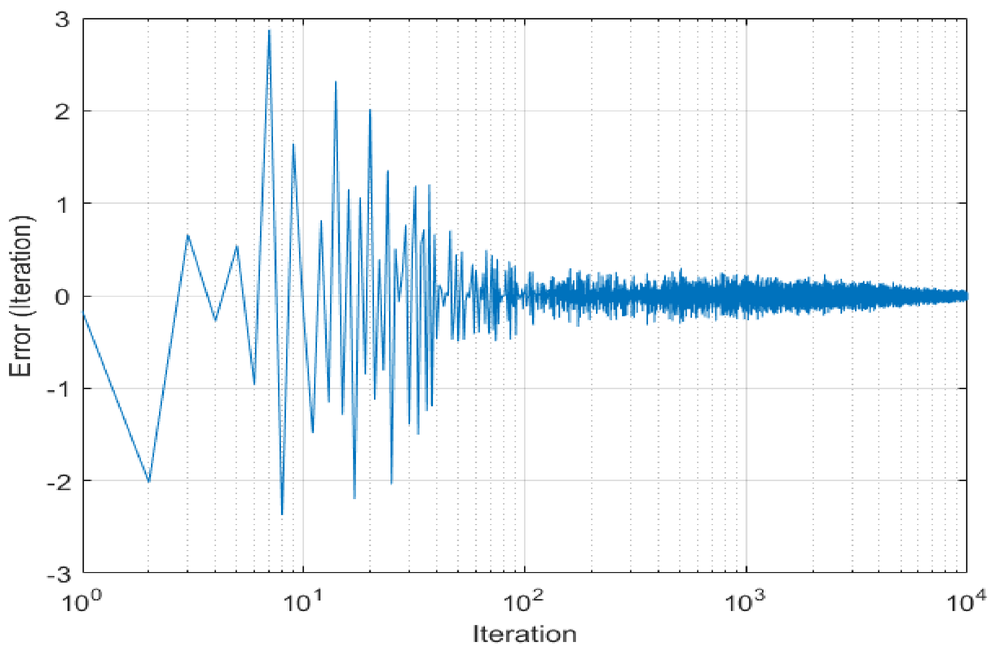


Figure 4. 8 Approximation error value of Example 4.3.3

4.4 Future Work about Learning Approximation

Since in chapter three we defined a neural network with a new activation function and we studied the derivative of this function, we found that it is bounded, so :

1. We can study the degree of best approximation of this activation function of the neural network that we defined in **(3.1.2)** with the Backward layer using the derivative and compare the degree of approximation that results from this layer with the degree of approximation that results from **(4.2.1)** , Where in this process we need to adjust the weights .
2. The intering of the activation function that we defined in **(3.1.1)** in the BP algorithm, as well as the comparison between its work and other types of activation functions.

Conclusions

For the importance of moduli of smoothness improvements were made to the existing moduli, so new generalized models of smoothness were defined related to very important tool, which is Jacobian weight. Furthermore, this modulus was used as a measurement of approximation of functions from the space $L^r_{p,\varphi}$ by using neural networks, which are also defined.

A new definition of activation function is presented because of the drawback of previous activation functions and the continuous need for more effective over. In addition, neural networks are considered an universal approximators, where it can approximation any Lp function.

For their ability to learn, NNs have been taken advantage of to obtain more accurate numerical results using the BP algorithm, some examples of that property have been given using MATLAB.

References

[1] P. Chebyshev, “Théorie des mécanismes connus sous le nom de parallélogrammes,” Imprimerie de l’Académie impériale des sciences, 1853.

[2] K. Weierstrass, “Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen.,” Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin, vol. 2, pp. 633–639, 1885.

[3] S. Bernstein, “Demonstration of a theorem of Weierstrass based on the calculus of probabilities,” Communications of the Kharkov Mathematical Society, vol. 13, no. 1, 1912.

[4] R. Salem and A. Zygmund, “The Approximation by Partial Sums of Fourier Series,” Source: Transactions of the American Mathematical Society, vol. 59, no. 1, pp. 14–22, 1946.

[5] C. de Boor, “On uniform approximation by splines,” J Approx Theory, vol. 1, no. 1, pp. 219–235, 1968.

[6] V. P. Motornyi, “approximation of by trigonometric,” Matematieheskie Zametki, vol. 16, no. 1, pp. 15–26, 1974.

[7] G. Cybenkot, “Approximation by Superpositions of a Sigmoidal Function,” Mathematics of control, signals and systems, vol. 2, pp. 303–314, 1989.

[8] Y. Ito, “Approximation of Functions on a Compact Set by Finite Sums of a Sigmoid Function Without Scaling,” Neural Network, vol. 4, pp. 817–826, 1991.

[9] B. C. Csáji, “Approximation with Artificial Neural Networks Huub ten Eikelder,” Faculty of Sciences, Etvs Lornd University, Hungary, vol. 24, no. 48, 2001.

[10] I. Ohn and Y. Kim, “Smooth function approximation by deep neural networks with general activation functions,” Entropy, vol. 21, no. 7, Jul. 2019, doi: 10.3390/e21070627.

[11] T. de Ryck, S. Lanthaler, and S. Mishra, “On the approximation of functions by tanh neural networks,” Neural Networks, vol. 143, pp. 732–750, Nov. 2021, doi: 10.1016/j.neunet.2021.08.015.

[12] W. S. Mcculloch and W. Pitts, “a logical calculus of the ideas immanent in nervous activity,” Bull Math Biol, vol. 5, 1943.

[13] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain,” Psychol Rev, vol. 65, no. 6, pp. 19–27, 1958.

[14] M. Minsky, “An introduction to computational geometry,” Cambridge tiass, 1969.

[15] P. Werbos, “Beyond regression : new tools for prediction and analysis in the behavioral sciences,” Ph. D. dissertation, Harvard University, 1974.

[16] D. E. Rumelhart, E. H. Geoffrey, and J. W. Ronald, “Learning representations by back-propagating errors,” Nature, vol. 323, no. 6088, pp. 533–536, 1986.

[17] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” Neural Comput, 2006.

[18] R. Hunt, B. Muckenhoupt, and R. Wheeden, “Weighted norm inequalities for the conjugate function and Hilbert transform,” Trans Am Math Soc, vol. 167, pp. 227–251, 1973.

[19] Z. Ditzian and V. Totik, Moduli of Smoothness, vol. 9. New York, NY: Springer New York, 1987. doi: 10.1007/978-1-4612-4778-4.

[20] M. M. Dzbasyan, “On weighted best polynomial approximation on the whole real axis,” Dokl . Akad . Nauk SSSR, vol. 84, 1952.

[21] R. Akgün, R. Akg¨un, and A. Akg¨un, “Realization and characterization of modulus of smoothness in weighted Lebesgue spaces,” CRM Preprint Series, p. 1170, 2013.

[22] Z. Ditzian, “a modulus of smoothness on the unit sphere,” Journal d’Analyse Mathématique, vol. 79, no. 1, 1999.

[23] B. R. Draganov, “A new modulus of smoothness for trigonometric polynomial approximation,” East Journal on Approximations, vol. 8, no. 4, 2002.

[24] J. Wang, Y. Xue, and F. Li, “Stechkin-Marchaud-Type inequalities with Jacobi weights for Bernstein operators,” J. Appl. Math. & Computing, vol. 24, no. 2, pp. 343–355, 2007.

[25] K. A. Kopotun, D. Leviatan, and I. A. Shevchuk, “New moduli of smoothness,” Publications de l’Institut Mathematique, vol. 96, no. 110, pp. 169–180, 2014, doi: 10.2298/PIM1410169K.

[26] K. A. Kopotun, D. Leviatan, and I. A. Shevchuk, “New Moduli of Smoothness: Weighted DT Moduli Revisited and Applied,” Constr Approx, vol. 42, no. 1, pp. 129–159, Aug. 2015, doi: 10.1007/s00365-014-9270-2.

[27] K. A. Kopotun, D. Leviatan, and I. A. Shevchuk, “On the Moduli of Smoothness with Jacobi Weights,” Ukrainian Mathematical Journal, vol. 70, no. 3, pp. 437–466, Aug. 2018, doi: 10.1007/s11253-018-1509-9.

[28] N. Paul G., “Orthogonal polynomials,” American Mathematical Society, vol. 18, no. 0065–9290, pp. 0–184, 1979.

[29] Z. A. Sharba and E. S. Bhaya, “Moduli of Smoothness for Weighted Lp,0 < p < 1 Approximation,” in Journal of Physics: Conference Series, Mar. 2021, vol. 1804, no. 1. doi: 10.1088/1742-6596/1804/1/012081.

[30] B. C. Csáji, “Approximation with Artificial Neural Networks,” Faculty of Sciences, Etvs Lornd University, Hungary, 2001.

[31] C. Ding, F. Cao, and Z. Xu, “LNCS 3971 - The Essential Approximation Order for Neural Networks with Trigonometric Hidden Layer Units,” 2006.

[32] G. Anastassiou, Intelligent Systems: Approximation by Artificial Neural Networks. Springer, 2009.

[33] B. Pirzamanbein, Modulus of continuity and its application in classifying the smoothness of images. 2011.

[34] E. S. Bhaya and N. A. Hadi, “Jacobi Weighted Moduli of Smoothness for Approximation by Neural Networks Application,” Journal of University of Babylon for Pure and Applied Sciences, no. 28, 2020.

[35] M. Meissner, M. Schmuker, and G. Schneider, “Optimized Particle Swarm Optimization (OPSO) and its application to artificial neural

network training,” BMC Bioinformatics, vol. 7, pp. 125–131, Mar. 2006, doi: 10.1186/1471-2105-7-125.

[36] L. Wang, “A hybrid genetic algorithm-neural network strategy for simulation optimization,” Appl Math Comput, vol. 170, no. 2, pp. 1329–1343, Nov. 2005, doi: 10.1016/j.amc.2005.01.024.

[37] S. Du, W. Li, and K. Cao, “A Learning Algorithm of Artificial Neural Network Based on GA-PSO,” 2006 6th World Congress on Intelligent Control and Automation, pp. 3633–3637, 2006.

[38] S. Haykin, “Neural networks and learning machines,” Pearson Education India, p. 906, 2009.

الخلاصة

يحظى موضوع معامل النعومة باهتمام العديد من الباحثين ، لاستخداماته القابلة للتطبيق في مجالات عدة ، ولاسيما فيما يخص تقريب الدوال. استخدم قلة من الباحثين الصيغة العامة للوزن الجاكوبي في معامل النعومة. مما دفعنا ، في هذا العمل ، لتعريف وزن جاكوبي معمم جديد ، والذي عرفنا من خلاله معامل نعومة من النوع الموزون ، المسمى بمقياس النعومة الجاكوبي الموزون المعمم. هذا وقد تمت دراسة خصائص المعامل أعلاه لاستخدامه بسهولة في مجالات مختلفة بدوال من فضاء $L^r_{p,\varphi}$، خاصة عندما $0 < p < 1$. كما درسنا خصائص ذلك الوزن ، بالإضافة إلى العلاقة بينه وبين الوزن الجاكوبي الكلاسيكي.

إنّ واحدة من أهم النتائج التي تم التوصل لها في نظرية تقريب الدوال، هي التكافؤ بين أي معامل نعومة والدالي k، الأمر الذي يدفعنا إلى اثبات ذلك التكافؤ مع معيار نعومة جاكوبي العام أهميته البالغة في نظريات التقريب الرئيسية.

من ناحية أخرى ، ونظرًا لحاجتنا إلى الشبكات العصبية للعمل على تقريب الدوال، فقد عملنا على تعريف دالة تنشيط جديدة بالإضافة إلى شبكة عصبية جديدة، لتتم دراسة خصائص دالة التنشيط بما في ذلك قابليتها للاشتقاق في الفترة $(-1,1)$. الأمر الذي يقودنا لتقريب الدوال من فضاء $L^r_{p,\varphi}$ بواسطة الشبكات العصبية بدالة التنشيط المذكورة أعلاه. علاوة على ذلك ، تم تقدير درجة التقريب باستخدام معامل النعومة المحدد مسبقًا.

في الجزء الأخير من العمل تم تقديم مصداق واقعي على أهمية التعلم في التقريب الدالي، من خلال خوارزمية تستخدم بشكل واسع، تسمى بخوارزمية الانتشار العكسي. تعطي بعض الأمثلة نتائج دقيقة من خلال تطبيق الخوارزمية في برنامج الماتلاب.

جمهورية العراق

وزارة التعليم العالي والبحث العلمي

جامعة بابل

كلية التربية للعلوم الصرفة

قسم الرياضيات

شبكات عصبية جديدة لتقريب موزون جديد

رسالة مقدمة إلى مجلس كلية التربية للعلوم الصرفة / جامعة بابل كجزء من متطلبات الحصول على درجة الماجستير في التربية / الرياضيات

من قبل

احمد عبد الكريم حسن مردان

بإشراف

أ.م.د. حوراء عباس المرعب

1444 هـ

2022 م