Parametrized error function based Banach space valued multivariate multi layer neural network approximations

George A. Anastassiou
Department of Mathematical Sciences
University of Memphis
Memphis, TN 38152, U.S.A.
ganastss@memphis.edu

Abstract

Here we describe multivariate quantitative approximations of Banach space valued continuous multivariate functions on a box or $\mathbb{R}^N$, $N \in \mathbb{N}$, by the multivariate normalized, quasi-interpolation, Kantorovich type and quadrature type neural network operators. We treat also the case of approximation by iterated operators of the last four types, these correspond to hidden multi-layer neural networks. The approximations are derived by establishing multidimensional Jackson type inequalities involving the multivariate modulus of continuity of the engaged function or its high order partial derivatives. Our multivariate operators are defined by using a multidimensional density function induced by a parametrized error function. The approximations are pointwise and uniform. The related feed-forward neural network starts with one hidden layer.

2020 AMS Mathematics Subject Classification: 41A17, 41A25, 41A30, 41A36.

Keywords and Phrases: Parametrized error function, multivariate neural network approximation, quasi-interpolation operator, Kantorovich type operator, quadrature type operator, multivariate modulus of continuity, abstract approximation, iterated multi layer neural network approximation.

1 Introduction

The author in [2] and [3], see chapters 2-5, was the first to establish neural network approximations to continuous functions with rates by very specifically defined neural network operators of Cardaliaguet-Euvrard and "Squashing" types,

RGMIA Res. Rep. Coll. **26** (2023), Art. 11, 20 pp. Received 01/02/23

by employing the modulus of continuity of the engaged function or its high order derivative, and producing very tight Jackson type inequalities. He treats there both the univariate and multivariate cases. The defining these operators "bell-shaped" and "squashing" functions are assumed to be of compact support. Also in [3] he gives the Nth order asymptotic expansion for the error of weak approximation of these two operators to a special natural class of smooth functions, see chapters 4-5 there.

For this article the author is motivated by the article [14] of Z. Chen and F. Cao, also by [4] - [12], [15], [16].

Let h be a general sigmoid function with $h(0)=0$, and $y=\pm 1$ the horizontal asymptotes. Of course h is strictly increasing over $\mathbb{R}$. Let the parameter $0<r<1$ and $x>0$. Then clearly $-x<x$ and $-x<-rx<rx<x$, furthermore it holds $h(-x)<h(-rx)<h(rx)<h(x)$. Consequently the sigmoid $y=h(rx)$ has a graph inside the graph of $y=h(x)$, of course with the same asymptotes $y=\pm 1$. Therefore $h(rx)$ has derivatives (gradients) at more points x than $h(x)$ has different than zero or not as close to zero, thus killing less number of neurons! And of course $h(rx)$ is more distant from $y=\pm 1$, than $h(x)$ it is. A highly desired fact in Neural Networks theory.

Different activation functions allow for different non-linearities which might work better for solving a specific function. So the need to use neural networks with various activation functions is vivid. Thus, performing neural network approximations using different activation functions is not only necessary but fully justified.

The author here performs multivariate parametrized error sigmoid function based neural network approximations to continuous functions over boxes or over the whole $\mathbb{R}^N$, $N\in\mathbb{N}$. Also he does iterated hidden multi layer neural network approximation. All convergences here are with rates expressed via the multivariate modulus of continuity of the involved function or its high order partial derivative and given by very tight multidimensional Jackson type inequalities.

The author here comes up with the "right" precisely defined multivariate normalized, quasi-interpolation neural network operators related to boxes or $\mathbb{R}^N$, as well as Kantorovich type and quadrature type related operators on $\mathbb{R}^N$. Our boxes are not necessarily symmetric to the origin. In preparation to prove our results we establish important properties of the basic multivariate density function induced by a parametrized error sigmoid function and defining our operators.

Feed-forward neural networks (FNNs) with one hidden layer, the starting type of networks we deal with in this article, are mathematically expressed as

$$N_n(x)=\sum_{j=0}^{n} c_j\sigma\left(\langle a_j\cdot x\rangle+b_j\right),\quad x\in\mathbb{R}^s,\quad s\in\mathbb{N},$$

where for $0\leq j\leq n$, $b_j\in\mathbb{R}$ are the thresholds, $a_j\in\mathbb{R}^s$ are the connection weights, $c_j \in \mathbb{R}$ are the coefficients, $\langle a_j \cdot x \rangle$ is the inner product of a_j and x, and σ is the activation function of the network. In many fundamental network models, the activation function is an error sigmoid function. About approximation theory see [1], and about neural networks read [17], [18], [19].

2 Basics

We consider here the parametrized (Gauss) error special function

$$erf \lambda z = \frac{2}{\sqrt{\pi}} \int_0^{\lambda z} e^{-t^2} dt, \quad \lambda > 0,\, z \in \mathbb{R}, \tag{1}$$

which is a sigmoidal type function and a strictly increasing function. It is acting here as an activation function.

Of special interest in neural network theory is when $0 < \lambda < 1$, see 1. Introduction.

It has the basic properties

$$\begin{gathered} erf\left(\lambda 0\right) = 0, \quad erf\left(-\lambda x\right) = -erf\left(\lambda x\right), \\ erf\left(\lambda\left(+\infty\right)\right) = 1, \;\; erf\left(\lambda\left(-\infty\right)\right) = -1, \end{gathered} \tag{2}$$

and

$$\left(erf\left(\lambda x\right)\right)' = \frac{2\lambda}{\sqrt{\pi}} e^{-(\lambda x)^2}, \;\; x \in \mathbb{R}. \tag{3}$$

We consider the function

$$\chi\left(x\right) = \frac{1}{4}\left(erf\left(\lambda\left(x+1\right)\right) - erf\left(\lambda\left(x-1\right)\right)\right), \;\; x \in \mathbb{R}, \tag{4}$$

and we notice that

$$\begin{gathered} \chi\left(-x\right) = \frac{1}{4}\left(erf\left(\lambda\left(-x+1\right)\right) - erf\left(\lambda\left(-x-1\right)\right)\right) = \\ \frac{1}{4}\left(erf\left(-\lambda\left(x-1\right)\right) - erf\left(-\lambda\left(x+1\right)\right)\right) = \\ \frac{1}{4}\left(-erf\left(\lambda\left(x-1\right)\right) + erf\left(\lambda\left(x+1\right)\right)\right) = \\ \frac{1}{4}\left(erf\left(\lambda\left(x+1\right)\right) - erf\left(\lambda\left(x-1\right)\right)\right) = \chi\left(x\right). \end{gathered} \tag{5}$$

Thus χ is an even function.

Since $x+1 > x-1$, then $erf\left(\lambda\left(x+1\right)\right) > erf\left(\lambda\left(x-1\right)\right)$, and $\chi\left(x\right) > 0$, all $x \in \mathbb{R}$.

We see

$$\chi\left(0\right) = \frac{erf \lambda}{2}. \tag{6}$$

Let $x > 0$, we have that

$$\chi'(x) = \frac{1}{4}\left((erf(\lambda(x+1)))' - (erf(\lambda(x-1)))'\right) =$$

$$\frac{1}{4}\frac{2\lambda}{\sqrt{\pi}}\left(e^{-\lambda^2(x+1)^2} - e^{-\lambda^2(x-1)^2}\right) = \frac{\lambda}{2\sqrt{\pi}}\left(\frac{1}{e^{\lambda^2(x+1)^2}} - \frac{1}{e^{\lambda^2(x-1)^2}}\right) \tag{7}$$

$$\frac{\lambda}{2\sqrt{\pi}}\left(\frac{e^{\lambda^2(x-1)^2} - e^{\lambda^2(x+1)^2}}{e^{\lambda^2(x+1)^2}e^{\lambda^2(x-1)^2}}\right) < 0,$$

proving $\chi'(x) < 0$, for $x > 0$. That is χ is strictly decreasing on $[0, \infty)$ and it is strictly increasing on $(-\infty, 0]$, and $\chi'(0) = 0$.

Clearly, the x-axis is the horizontal asymptote of χ.

Conclusion, χ is a bell symmetric function with maximum

$$\chi(0) = \frac{erf\lambda}{2}.$$

Let $h : \mathbb{R} \to [-1, 1]$ be a general sigmoid function, such that it is strictly increasing, $h(0) = 0$, $h(-x) = -h(x)$, $h(+\infty) = 1$, $h(-\infty) = -1$. Also h is strictly convex over $(-\infty, 0]$ and strictly concave over $[0, +\infty)$, with $h^{(2)} \in C(\mathbb{R})$, see [13].

So $erf\lambda x$ is a special case of h. Furthermore $\chi(x)$ is a special case of the following general function

$$\psi(x) := \frac{1}{4}(h(x+1) - h(x-1)), \quad x \in \mathbb{R}, \tag{8}$$

see [13].

We have

Theorem 1 *([13]) It holds*

$$\sum_{i=-\infty}^{\infty} \psi(x-i) = 1, \quad \forall\, x \in \mathbb{R}. \tag{9}$$

Thus

Corollary 2 *It holds*

$$\sum_{i=-\infty}^{\infty} \chi(x-i) = 1, \quad \forall\, x \in \mathbb{R}. \tag{10}$$

We have

Theorem 3 *([13]) It holds*

$$\int_{-\infty}^{\infty} \psi(x)\,dx = 1. \tag{11}$$

Thus

Corollary 4 *We have that*

$$\int_{-\infty}^{\infty} \chi(x)\,dx = 1. \tag{12}$$

Hence $\chi(x)$ is a density function on $\mathbb{R}$.
We need

Theorem 5 *([13]) Let $0 < \alpha < 1$, and $n \in \mathbb{N}$ with $n^{1-\alpha} > 2$. It holds*

$$\sum_{\begin{cases} k=-\infty \\ : |nx-k| \geq n^{1-\alpha} \end{cases}}^{\infty} \psi(nx-k) < \frac{\left(1 - h\left(n^{1-\alpha}-2\right)\right)}{2}, \tag{13}$$

with

$$\lim_{n\to+\infty} \frac{\left(1 - h\left(n^{1-\alpha}-2\right)\right)}{2} = 0.$$

Thus we obtain

Corollary 6 *Let $0 < \alpha < 1$, and $n \in \mathbb{N}$ with $n^{1-\alpha} > 2$, $\lambda > 0$. It holds*

$$\sum_{\begin{cases} k=-\infty \\ : |nx-k| \geq n^{1-\alpha} \end{cases}}^{\infty} \chi(nx-k) < \frac{\left(1 - erf\left(\lambda\left(n^{1-\alpha}-2\right)\right)\right)}{2}, \tag{14}$$

with

$$\lim_{n\to+\infty} \frac{\left(1 - erf\left(\lambda\left(n^{1-\alpha}-2\right)\right)\right)}{2} = 0.$$

Denote by $\lfloor\cdot\rfloor$ the integral part and by $\lceil\cdot\rceil$ the ceiling of a number.
Furthermore we need

Theorem 7 *([13]) Let $x \in [a,b] \subset \mathbb{R}$ and $n \in \mathbb{N}$ so that $\lceil na\rceil \leq \lfloor nb\rfloor$. It holds*

$$\frac{1}{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} \psi(nx-k)} < \frac{1}{\psi(1)}, \quad \forall\, x \in [a,b]. \tag{15}$$

Therefore we derive

Corollary 8 *Let $x \in [a,b] \subset \mathbb{R}$, $\lambda > 0$ and $n \in \mathbb{N}$ so that $\lceil na\rceil \leq \lfloor nb\rfloor$. Then*

$$\frac{1}{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} \chi(nx-k)} < \frac{1}{\chi(1)} = \frac{4}{erf(2\lambda)}. \tag{16}$$

Remark 9 *As in [13], we have that*

$$\lim_{n\to\infty}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\chi\left(nx-k\right)\neq 1. \tag{17}$$

Note 10 *For large enough n we always obtain $\lceil na\rceil\leq\lfloor nb\rfloor$. Also $a\leq\frac{k}{n}\leq b$, iff $\lceil na\rceil\leq k\leq\lfloor nb\rfloor$. As in [13], we obtain that*

$$\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\chi\left(nx-k\right)\leq 1. \tag{18}$$

We introduce

$$Z\left(x_1,...,x_N\right):=Z\left(x\right):=\prod_{i=1}^{N}\chi\left(x_i\right),\quad x=\left(x_1,...,x_N\right)\in\mathbb{R}^N,\ N\in\mathbb{N}. \tag{19}$$

It has the properties:

(i) $Z\left(x\right)>0,\ \forall\ x\in\mathbb{R}^N$,

(ii)

$$\sum_{k=-\infty}^{\infty}Z\left(x-k\right):=\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}...\sum_{k_N=-\infty}^{\infty}Z\left(x_1-k_1,...,x_N-k_N\right)=1, \tag{20}$$

where $k:=\left(k_1,...,k_n\right)\in\mathbb{Z}^N$, $\forall\ x\in\mathbb{R}^N$,

hence

(iii)

$$\sum_{k=-\infty}^{\infty}Z\left(nx-k\right)=1, \tag{21}$$

$\forall\ x\in\mathbb{R}^N$; $n\in\mathbb{N}$,

and

(iv)

$$\int_{\mathbb{R}^N}Z\left(x\right)dx=1, \tag{22}$$

that is Z is a multivariate density function.

Here denote $\left\|x\right\|_\infty:=\max\left\{\left|x_1\right|,...,\left|x_N\right|\right\}$, $x\in\mathbb{R}^N$, also set $\infty:=\left(\infty,...,\infty\right)$, $-\infty:=\left(-\infty,...,-\infty\right)$ upon the multivariate context, and

$$\begin{aligned}\lceil na\rceil &:= \left(\lceil na_1\rceil,...,\lceil na_N\rceil\right),\\ \lfloor nb\rfloor &:= \left(\lfloor nb_1\rfloor,...,\lfloor nb_N\rfloor\right),\end{aligned} \tag{23}$$

where $a:=\left(a_1,...,a_N\right)$, $b:=\left(b_1,...,b_N\right)$.

We obviously see that

$$\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} Z(nx-k) = \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\left(\prod_{i=1}^{N}\chi(nx_i-k_i)\right) =$$

$$\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor}\cdots\sum_{k_N=\lceil na_N\rceil}^{\lfloor nb_N\rfloor}\left(\prod_{i=1}^{N}\chi(nx_i-k_i)\right) = \prod_{i=1}^{N}\left(\sum_{k_i=\lceil na_i\rceil}^{\lfloor nb_i\rfloor}\chi(nx_i-k_i)\right). \quad (24)$$

For $0<\beta<1$ and $n\in\mathbb{N}$, a fixed $x\in\mathbb{R}^N$, we have that

$$\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} Z(nx-k) =$$

$$\sum_{\begin{cases} k=\lceil na\rceil \\ \left\|\frac{k}{n}-x\right\|_\infty \le \frac{1}{n^\beta}\end{cases}}^{\lfloor nb\rfloor} Z(nx-k) + \sum_{\begin{cases} k=\lceil na\rceil \\ \left\|\frac{k}{n}-x\right\|_\infty > \frac{1}{n^\beta}\end{cases}}^{\lfloor nb\rfloor} Z(nx-k). \quad (25)$$

In the last two sums the counting is over disjoint vector sets of k's, because the condition $\left\|\frac{k}{n}-x\right\|_\infty > \frac{1}{n^\beta}$ implies that there exists at least one $\left|\frac{k_r}{n}-x_r\right| > \frac{1}{n^\beta}$, where $r\in\{1,\dots,N\}$.

(v) As in [10], pp. 379-380, we derive that

$$\sum_{\begin{cases} k=\lceil na\rceil \\ \left\|\frac{k}{n}-x\right\|_\infty > \frac{1}{n^\beta}\end{cases}}^{\lfloor nb\rfloor} Z(nx-k) \overset{(7)}{<} \frac{1-\mathrm{erf}\lambda\left(n^{1-\beta}-2\right)}{2}, \quad 0<\beta<1, \quad (26)$$

with $n\in\mathbb{N}: n^{1-\beta}>2$, $x\in\prod_{i=1}^{N}[a_i,b_i]$.

(vi) By Corollary 8 we get that

$$0 < \frac{1}{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} Z(nx-k)} < \left(\frac{4}{\mathrm{erf}(2\lambda)}\right)^N, \quad (27)$$

$\forall\, x\in\left(\prod_{i=1}^{N}[a_i,b_i]\right)$, $n\in\mathbb{N}$.

It is also clear that

(vii)

$$\sum_{\begin{cases} k=-\infty \\ \left\|\frac{k}{n}-x\right\|_\infty > \frac{1}{n^\beta}\end{cases}}^{\infty} Z(nx-k) < \frac{1-\mathrm{erf}\lambda\left(n^{1-\beta}-2\right)}{2}, \quad (28)$$

$0<\beta<1$, $n\in\mathbb{N}: n^{1-\beta}>2$, $x\in\mathbb{R}^N$.

Furthermore it holds

$$\lim_{n\to\infty}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} Z(nx-k)\neq 1, \tag{29}$$

for at least some $x\in\left(\prod_{i=1}^{N}[a_i,b_i]\right)$.

Here $\left(X,\|\cdot\|_{\gamma}\right)$ is a Banach space.

Let $f\in C\left(\prod_{i=1}^{N}[a_i,b_i],X\right)$, $x=(x_1,...,x_N)\in\prod_{i=1}^{N}[a_i,b_i]$, $n\in\mathbb{N}$ such that $\lceil na_i\rceil\leq\lfloor nb_i\rfloor$, $i=1,...,N$.

We introduce and define the following multivariate linear normalized neural network operator ($x:=(x_1,...,x_N)\in\left(\prod_{i=1}^{N}[a_i,b_i]\right)$):

$$A_n(f,x_1,...,x_N):=A_n(f,x):=\frac{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} f\left(\frac{k}{n}\right)Z(nx-k)}{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} Z(nx-k)}=$$

$$\frac{\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor}\sum_{k_2=\lceil na_2\rceil}^{\lfloor nb_2\rfloor}...\sum_{k_N=\lceil na_N\rceil}^{\lfloor nb_N\rfloor} f\left(\frac{k_1}{n},...,\frac{k_N}{n}\right)\left(\prod_{i=1}^{N}\chi(nx_i-k_i)\right)}{\prod_{i=1}^{N}\left(\sum_{k_i=\lceil na_i\rceil}^{\lfloor nb_i\rfloor}\chi(nx_i-k_i)\right)}. \tag{30}$$

For large enough $n\in\mathbb{N}$ we always obtain $\lceil na_i\rceil\leq\lfloor nb_i\rfloor$, $i=1,...,N$. Also $a_i\leq\frac{k_i}{n}\leq b_i$, iff $\lceil na_i\rceil\leq k_i\leq\lfloor nb_i\rfloor$, $i=1,...,N$.

When $g\in C\left(\prod_{i=1}^{N}[a_i,b_i]\right)$ we define the companion operator

$$\widetilde{A}_n(g,x):=\frac{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} g\left(\frac{k}{n}\right)Z(nx-k)}{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} Z(nx-k)}. \tag{31}$$

Clearly $\widetilde{A}_n$ is a positive linear operator. We have that

$$\widetilde{A}_n(1,x)=1,\ \ \forall\ x\in\left(\prod_{i=1}^{N}[a_i,b_i]\right).$$

Notice that $A_n(f)\in C\left(\prod_{i=1}^{N}[a_i,b_i],X\right)$ and $\widetilde{A}_n(g)\in C\left(\prod_{i=1}^{N}[a_i,b_i]\right)$.

Furthermore it holds

$$\|A_n(f,x)\|_{\gamma}\leq\frac{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\left\|f\left(\frac{k}{n}\right)\right\|_{\gamma}Z(nx-k)}{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} Z(nx-k)}=\widetilde{A}_n\left(\|f\|_{\gamma},x\right), \tag{32}$$

$\forall\ x\in\prod_{i=1}^{N}[a_i,b_i]$.

Clearly $\|f\|_{\gamma}\in C\left(\prod_{i=1}^{N}[a_i,b_i]\right)$.

So, we have that

$$\left\| A_n(f,x) \right\|_\gamma \leq \widetilde{A}_n\left(\|f\|_\gamma, x \right), \tag{33}$$

$\forall\ x \in \prod_{i=1}^N [a_i, b_i]$, $\forall\ n \in \mathbb{N}$, $\forall\ f \in C\left(\prod_{i=1}^N [a_i, b_i], X \right)$.

Let $c \in X$ and $g \in C\left(\prod_{i=1}^N [a_i, b_i] \right)$, then $cg \in C\left(\prod_{i=1}^N [a_i, b_i], X \right)$.

Furthermore it holds

$$A_n(cg, x) = c\widetilde{A}_n(g, x), \ \ \forall\ x \in \prod_{i=1}^N [a_i, b_i]. \tag{34}$$

Since $\widetilde{A}_n(1) = 1$, we get that

$$A_n(c) = c, \ \ \forall\ c \in X. \tag{35}$$

We call $\widetilde{A}_n$ the companion operator of A_n.

For convinience we call

$$A_n^*(f,x) := \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} f\left(\frac{k}{n} \right) Z(nx - k) =$$

$$\sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor} \sum_{k_2=\lceil na_2 \rceil}^{\lfloor nb_2 \rfloor} \dots \sum_{k_N=\lceil na_N \rceil}^{\lfloor nb_N \rfloor} f\left(\frac{k_1}{n}, \dots, \frac{k_N}{n} \right) \left(\prod_{i=1}^N \chi(nx_i - k_i) \right), \tag{36}$$

$\forall\ x \in \left(\prod_{i=1}^N [a_i, b_i] \right)$.

That is

$$A_n(f,x) := \frac{A_n^*(f,x)}{\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} Z(nx - k)}, \tag{37}$$

$\forall\ x \in \left(\prod_{i=1}^N [a_i, b_i] \right)$, $n \in \mathbb{N}$.

Hence

$$A_n(f,x) - f(x) = \frac{A_n^*(f,x) - f(x) \left(\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} Z(nx - k) \right)}{\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} Z(nx - k)}. \tag{38}$$

Consequently we derive

$$\left\| A_n(f,x) - f(x) \right\|_\gamma \overset{(27)}{\leq} \left(\frac{4}{\operatorname{erf}(2\lambda)} \right)^N \left\| A_n^*(f,x) - f(x) \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} Z(nx - k) \right\|_\gamma, \tag{39}$$

$\forall\ x \in \left(\prod_{i=1}^N [a_i, b_i] \right)$.

We will estimate the right hand side of (39).

For the last and others we need

Definition 11 *([11], p. 274) Let M be a convex and compact subset of $\left(\mathbb{R}^N, \|\cdot\|_p\right)$, $p \in [1, \infty]$, and $\left(X, \|\cdot\|_\gamma\right)$ be a Banach space. Let $f \in C(M, X)$. We define the first modulus of continuity of f as*

$$\omega_1(f, \delta) := \sup_{\substack{x, y \in M: \\ \|x-y\|_p \leq \delta}} \|f(x) - f(y)\|_\gamma, \quad 0 < \delta \leq diam(M). \tag{40}$$

If $\delta > diam(M)$, then

$$\omega_1(f, \delta) = \omega_1(f, diam(M)). \tag{41}$$

Notice $\omega_1(f, \delta)$ is increasing in $\delta > 0$. For $f \in C_B(M, X)$ (continuous and bounded functions) $\omega_1(f, \delta)$ is defined similarly.

Lemma 12 *([11], p. 274) We have $\omega_1(f, \delta) \to 0$ as $\delta \downarrow 0$, iff $f \in C(M, X)$, where M is a convex compact subset of $\left(\mathbb{R}^N, \|\cdot\|_p\right)$, $p \in [1, \infty]$.*

Clearly we have also: $f \in C_U\left(\mathbb{R}^N, X\right)$ (uniformly continuous functions), iff $\omega_1(f, \delta) \to 0$ as $\delta \downarrow 0$, where ω_1 is defined similarly to (40). The space $C_B\left(\mathbb{R}^N, X\right)$ denotes the continuous and bounded functions on $\mathbb{R}^N$.

When $f \in C_B\left(\mathbb{R}^N, X\right)$ we define,

$$B_n(f, x) := B_n(f, x_1, ..., x_N) := \sum_{k=-\infty}^{\infty} f\left(\frac{k}{n}\right) Z(nx - k) :=$$

$$\sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} ... \sum_{k_N=-\infty}^{\infty} f\left(\frac{k_1}{n}, \frac{k_2}{n}, ..., \frac{k_N}{n}\right) \left(\prod_{i=1}^{N} \chi(nx_i - k_i)\right), \tag{42}$$

$n \in \mathbb{N}$, $\forall\, x \in \mathbb{R}^N$, $N \in \mathbb{N}$, the multivariate quasi-interpolation neural network operator.

Also for $f \in C_B\left(\mathbb{R}^N, X\right)$ we define the multivariate Kantorovich type neural network operator

$$C_n(f, x) := C_n(f, x_1, ..., x_N) := \sum_{k=-\infty}^{\infty} \left(n^N \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(t)\, dt\right) Z(nx - k) =$$

$$\sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} ... \sum_{k_N=-\infty}^{\infty} \left(n^N \int_{\frac{k_1}{n}}^{\frac{k_1+1}{n}} \int_{\frac{k_2}{n}}^{\frac{k_2+1}{n}} ... \int_{\frac{k_N}{n}}^{\frac{k_N+1}{n}} f(t_1, ..., t_N)\, dt_1 ... dt_N\right)$$

$$\cdot \left(\prod_{i=1}^{N} \chi(nx_i - k_i)\right), \tag{43}$$

$n \in \mathbb{N}$, $\forall \ x \in \mathbb{R}^N$.

Again for $f \in C_B\left(\mathbb{R}^N, X\right)$, $N \in \mathbb{N}$, we define the multivariate neural network operator of quadrature type $D_n(f,x)$, $n \in \mathbb{N}$, as follows.

Let $\theta = (\theta_1, ..., \theta_N) \in \mathbb{N}^N$, $r = (r_1, ..., r_N) \in \mathbb{Z}_+^N$, $w_r = w_{r_1,r_2,...r_N} \geq 0$, such that $\sum\limits_{r=0}^{\theta} w_r = \sum\limits_{r_1=0}^{\theta_1} \sum\limits_{r_2=0}^{\theta_2} ... \sum\limits_{r_N=0}^{\theta_N} w_{r_1,r_2,...r_N} = 1$; $k \in \mathbb{Z}^N$ and

$$\delta_{nk}(f) := \delta_{n,k_1,k_2,...,k_N}(f) := \sum_{r=0}^{\theta} w_r f\left(\frac{k}{n} + \frac{r}{n\theta}\right) =$$

$$\sum_{r_1=0}^{\theta_1} \sum_{r_2=0}^{\theta_2} ... \sum_{r_N=0}^{\theta_N} w_{r_1,r_2,...r_N} f\left(\frac{k_1}{n} + \frac{r_1}{n\theta_1}, \frac{k_2}{n} + \frac{r_2}{n\theta_2}, ..., \frac{k_N}{n} + \frac{r_N}{n\theta_N}\right), \tag{44}$$

where $\frac{r}{\theta} := \left(\frac{r_1}{\theta_1}, \frac{r_2}{\theta_2}, ..., \frac{r_N}{\theta_N}\right)$.

We set

$$D_n(f,x) := D_n(f, x_1, ..., x_N) := \sum_{k=-\infty}^{\infty} \delta_{nk}(f) Z(nx-k) = \tag{45}$$

$$\sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} ... \sum_{k_N=-\infty}^{\infty} \delta_{n,k_1,k_2,...,k_N}(f) \left(\prod_{i=1}^{N} \chi(nx_i - k_i)\right),$$

$\forall \ x \in \mathbb{R}^N$.

In this article we study the approximation properties of A_n, B_n, C_n, D_n neural network operators and as well of their iterates. That is, the quantitative pointwise and uniform convergence of these operators to the unit operator I.

3 Multivariate general sigmoid Neural Network Approximations

Here we present several vectorial neural network approximations to Banach space valued functions given with rates.

We give

Theorem 13 *Let $f \in C\left(\prod_{i=1}^N [a_i, b_i], X\right)$, $0 < \beta < 1$, $\lambda > 0$, $x \in \left(\prod_{i=1}^N [a_i, b_i]\right)$, $N, n \in \mathbb{N}$ with $n^{1-\beta} > 2$. Then*

1)

$$\|A_n(f,x) - f(x)\|_\gamma \leq$$

$$\left(\frac{4}{erf(2\lambda)}\right)^N \left[\omega_1\left(f, \frac{1}{n^\beta}\right) + \left(1 - erf\lambda\left(n^{1-\beta} - 2\right)\right) \left\|\|f\|_\gamma\right\|_\infty\right] =: \lambda_1(n), \tag{46}$$

and

2)

$$\left\| \left\| A_n(f) - f \right\|_\gamma \right\|_\infty \leq \lambda_1(n). \tag{47}$$

We notice that $\lim_{n\to\infty} A_n(f) \overset{\|\cdot\|_\gamma}{=} f$*, pointwise and uniformly.*

Above ω_1 *is with respect to* $p = \infty$ *and the speed of convergnece is* $\max\left(\frac{1}{n^\beta}, \left(1 - erf\lambda\left(n^{1-\beta} - 2\right)\right)\right)$.

Proof. As similar to [12] is omitted. ■

We continue with

Theorem 14 *Let* $f \in C_B\left(\mathbb{R}^N, X\right)$, $0 < \beta < 1$, $\lambda > 0$, $x \in \mathbb{R}^N$, $N, n \in \mathbb{N}$ *with* $n^{1-\beta} > 2$, ω_1 *is for* $p = \infty$. *Then*

1)

$$\left\| B_n(f, x) - f(x) \right\|_\gamma \leq \omega_1\left(f, \frac{1}{n^\beta}\right) + \left(1 - erf\lambda\left(n^{1-\beta} - 2\right)\right) \left\| \left\| f \right\|_\gamma \right\|_\infty =: \lambda_2(n), \tag{48}$$

2)

$$\left\| \left\| B_n(f) - f \right\|_\gamma \right\|_\infty \leq \lambda_2(n). \tag{49}$$

Given that $f \in \left(C_U\left(\mathbb{R}^N, X\right) \cap C_B\left(\mathbb{R}^N, X\right)\right)$*, we obtain* $\lim_{n\to\infty} B_n(f) = f$*, uniformly. The speed of convergence above is* $\max\left(\frac{1}{n^\beta}, \left(1 - erf\lambda\left(n^{1-\beta} - 2\right)\right)\right)$.

Proof. As similar to [12] is omitted. ■

We give

Theorem 15 *Let* $f \in C_B\left(\mathbb{R}^N, X\right)$, $0 < \beta < 1$, $\lambda > 0$, $x \in \mathbb{R}^N$, $N, n \in \mathbb{N}$ *with* $n^{1-\beta} > 2$, ω_1 *is for* $p = \infty$. *Then*

1)

$$\left\| C_n(f, x) - f(x) \right\|_\gamma \leq \omega_1\left(f, \frac{1}{n} + \frac{1}{n^\beta}\right) + \left(1 - erf\lambda\left(n^{1-\beta} - 2\right)\right) \left\| \left\| f \right\|_\gamma \right\|_\infty =: \lambda_3(n), \tag{50}$$

2)

$$\left\| \left\| C_n(f) - f \right\|_\gamma \right\|_\infty \leq \lambda_3(n). \tag{51}$$

Given that $f \in \left(C_U\left(\mathbb{R}^N, X\right) \cap C_B\left(\mathbb{R}^N, X\right)\right)$, *we obtain* $\lim_{n\to\infty} C_n(f) = f$*, uniformly.*

Proof. As similar to [12] is omitted. ■

We also present

Theorem 16 *Let $f \in C_B\left(\mathbb{R}^N, X\right)$, $0<\beta<1$, $\lambda>0$, $x \in \mathbb{R}^N$, $N, n \in \mathbb{N}$ with $n^{1-\beta}>2$, ω_1 is for $p=\infty$. Then*

1)

$$\left\|D_n(f, x)-f(x)\right\|_\gamma \leq \omega_1\left(f, \frac{1}{n}+\frac{1}{n^\beta}\right)+\left(1-erf\lambda\left(n^{1-\beta}-2\right)\right)\left\|\|f\|_\gamma\right\|_{\infty}=\lambda_4(n), \tag{52}$$

2)

$$\left\|\left\|D_n(f)-f\right\|_\gamma\right\|_{\infty} \leq \lambda_4(n). \tag{53}$$

Given that $f \in\left(C_U\left(\mathbb{R}^N, X\right) \cap C_B\left(\mathbb{R}^N, X\right)\right)$, we obtain $\lim_{n \rightarrow \infty} D_n(f)=f$, uniformly.

Proof. As similar to [12] is omitted. ■

Let $f \in C^m\left(\prod_{i=1}^N\left[a_i, b_i\right]\right)$, $m, N \in \mathbb{N}$. Here f_α denotes a partial derivative of f, $\alpha:=\left(\alpha_1, \ldots, \alpha_N\right)$, $\alpha_i \in \mathbb{Z}_{+}$, $i=1, \ldots, N$, and $|\alpha|:=\sum_{i=1}^N \alpha_i=l$, where $l=0,1, \ldots, m$. We write also $f_\alpha:=\frac{\partial^\alpha f}{\partial x^\alpha}$ and we say it is of order l.

We denote

$$\omega_{1, m}^{\max}\left(f_\alpha, h\right):=\max_{\alpha:|\alpha|=m} \omega_1\left(f_\alpha, h\right). \tag{54}$$

Call also

$$\left\|f_\alpha\right\|_{\infty, m}^{\max}:=\max_{|\alpha|=m}\left\{\left\|f_\alpha\right\|_{\infty}\right\}, \tag{55}$$

$\|\cdot\|_{\infty}$ is the supremum norm.

In the next we discuss high order of approximation by using the smoothness of f.

We give

Theorem 17 *Let $f \in C^m\left(\prod_{i=1}^N\left[a_i, b_i\right]\right)$, $0<\beta<1$, $n, m, N \in \mathbb{N}$, $n^{1-\beta} \geq 2$, $\lambda>0$, $x \in\left(\prod_{i=1}^N\left[a_i, b_i\right]\right)$. Then*

i)

$$\left|A_n(f, x)-f(x)-\sum_{j=1}^m\left(\sum_{|\alpha|=j}\left(\frac{f_\alpha(x)}{\prod_{i=1}^N \alpha_i!}\right) A_n\left(\prod_{i=1}^N\left(\cdot-x_i\right)^{\alpha_i}, x\right)\right)\right| \leq \tag{56}$$

$$\left(\frac{4}{erf(2\lambda)}\right)^N\left\{\frac{N^m}{m!n^{m\beta}} \omega_{1, m}^{\max}\left(f_\alpha, \frac{1}{n^\beta}\right)+\right.$$

$$\left.\left(\frac{\|b-a\|_{\infty}^m\left\|f_\alpha\right\|_{\infty, m}^{\max} N^m}{m!}\right)\left(1-erf\lambda\left(n^{1-\beta}-2\right)\right)\right\},$$

ii)

$$\left|A_n(f, x)-f(x)\right| \leq\left(\frac{4}{erf(2\lambda)}\right)^N \cdot \tag{57}$$

$$\left\{\sum_{j=1}^{m}\left(\sum_{|\alpha|=j}\left(\frac{|f_{\alpha}(x)|}{\prod_{i=1}^{N}\alpha_i!}\right)\left[\frac{1}{n^{\beta j}}+\left(\prod_{i=1}^{N}(b_i-a_i)^{\alpha_i}\right)\cdot\right.\right.\right.$$

$$\left.\left.\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)\right]\right)+\frac{N^m}{m!n^{m\beta}}\omega_{1,m}^{\max}\left(f_{\alpha},\frac{1}{n^{\beta}}\right)$$

$$\left.+\left(\frac{\|b-a\|_{\infty}^{m}\|f_{\alpha}\|_{\infty,m}^{\max}N^m}{m!}\right)\left(1-erf\lambda\left(n^{1-\beta}-2\right)\right)\right\},$$

iii)

$$\|A_n(f)-f\|_{\infty}\leq\left(\frac{4}{erf(2\lambda)}\right)^{N}\cdot \tag{58}$$

$$\left\{\sum_{j=1}^{m}\left(\sum_{|\alpha|=j}\left(\frac{\|f_{\alpha}\|_{\infty}}{\prod_{i=1}^{N}\alpha_i!}\right)\left[\frac{1}{n^{\beta j}}+\left(\prod_{i=1}^{N}(b_i-a_i)^{\alpha_i}\right)\cdot\right.\right.\right.$$

$$\left.\left.\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)\right]\right)+\frac{N^m}{m!n^{m\beta}}\omega_{1,m}^{\max}\left(f_{\alpha},\frac{1}{n^{\beta}}\right)$$

$$\left.+\left(\frac{\|b-a\|_{\infty}^{m}\|f_{\alpha}\|_{\infty,m}^{\max}N^m}{m!}\right)\left(1-erf\lambda\left(n^{1-\beta}-2\right)\right)\right\},$$

iv) assume $f_{\alpha}(x_0)=0$, for all $\alpha:|\alpha|=1,...,m$; $x_0\in\left(\prod_{i=1}^{N}[a_i,b_i]\right)$, then

$$|A_n(f,x_0)-f(x_0)|\leq\left(\frac{4}{erf(2\lambda)}\right)^{N}\left\{\frac{N^m}{m!n^{m\beta}}\omega_{1}^{\max}\left(f_{\alpha},\frac{1}{n^{\beta}}\right)+\right. \tag{59}$$

$$\left.\left(\frac{\|b-a\|_{\infty}^{m}\|f_{\alpha}\|_{\infty,m}^{\max}N^m}{m!}\right)\left(1-erf\lambda\left(n^{1-\beta}-2\right)\right)\right\},$$

notice in the last the extremely high rate of convergence at $n^{-\beta(m+1)}$.

Proof. As similar to [10], pp. 389-396, it is omitted. ■

We make

Definition 18 *Let $f\in C_B\left(\mathbb{R}^N,X\right)$, $N\in\mathbb{N}$, where $\left(X,\|\cdot\|_{\gamma}\right)$ is a Banach space. We define the general neural network operator*

$$F_n(f,x):=\sum_{k=-\infty}^{\infty}l_{nk}(f)\,Z(nx-k)=$$

$$\begin{cases}B_n(f,x), & \text{if } l_{nk}(f)=f\left(\frac{k}{n}\right),\\ C_n(f,x), & \text{if } l_{nk}(f)=n^N\int_{\frac{k}{n}}^{\frac{k+1}{n}}f(t)\,dt,\\ D_n(f,x), & \text{if } l_{nk}(f)=\delta_{nk}(f).\end{cases} \tag{60}$$

Clearly $l_{nk}(f)$ is an X-valued bounded linear functional such that $\|l_{nk}(f)\|_\gamma \leq \left\| \|f\|_\gamma \right\|_\infty$.

Hence $F_n(f)$ is a bounded linear operator with $\left\| \|F_n(f)\|_\gamma \right\|_\infty \leq \left\| \|f\|_\gamma \right\|_\infty$.

We need

Theorem 19 *Let $f \in C_B\left(\mathbb{R}^N, X\right)$, $N \geq 1$. Then $F_n(f) \in C_B\left(\mathbb{R}^N, X\right)$.*

Proof. Very lengthy and as similar to [10], pp. 396-400, it is omitted. ■

Remark 20 *By (30) it is obvious that $\left\| \|A_n(f)\|_\gamma \right\|_\infty \leq \left\| \|f\|_\gamma \right\|_\infty < \infty$, and $A_n(f) \in C\left(\prod_{i=1}^N [a_i, b_i], X\right)$, given that $f \in C\left(\prod_{i=1}^N [a_i, b_i], X\right)$.*

Call L_n any of the operators A_n, B_n, C_n, D_n.

Clearly then

$$\left\| \left\|L_n^2(f)\right\|_\gamma \right\|_\infty = \left\| \|L_n(L_n(f))\|_\gamma \right\|_\infty \leq \left\| \|L_n(f)\|_\gamma \right\|_\infty \leq \left\| \|f\|_\gamma \right\|_\infty, \tag{61}$$

etc.

Therefore we get

$$\left\| \left\|L_n^k(f)\right\|_\gamma \right\|_\infty \leq \left\| \|f\|_\gamma \right\|_\infty, \ \ \forall\ k \in \mathbb{N}, \tag{62}$$

the contraction property.

Also we see that

$$\left\| \left\|L_n^k(f)\right\|_\gamma \right\|_\infty \leq \left\| \left\|L_n^{k-1}(f)\right\|_\gamma \right\|_\infty \leq \ldots \leq \left\| \|L_n(f)\|_\gamma \right\|_\infty \leq \left\| \|f\|_\gamma \right\|_\infty. \tag{63}$$

Here L_n^k are bounded linear operators.

Notation 21 *Here $N \in \mathbb{N}$, $0 < \beta < 1$. Denote by*

$$c_N := \begin{cases} \left(\frac{4}{erf(2\lambda)}\right)^N, & \text{if } L_n = A_n, \\ 1, & \text{if } L_n = B_n, C_n, D_n, \end{cases} \tag{64}$$

$$\varphi(n) := \begin{cases} \frac{1}{n^\beta}, & \text{if } L_n = A_n,\ B_n, \\ \frac{1}{n} + \frac{1}{n^\beta}, & \text{if } L_n = C_n, D_n, \end{cases} \tag{65}$$

$$\Omega := \begin{cases} C\left(\prod_{i=1}^N [a_i, b_i], X\right), & \text{if } L_n = A_n, \\ C_B\left(\mathbb{R}^N, X\right), & \text{if } L_n = B_n, C_n, D_n, \end{cases} \tag{66}$$

and

$$Y := \begin{cases} \prod_{i=1}^N [a_i, b_i], & \text{if } L_n = A_n, \\ \mathbb{R}^N, & \text{if } L_n = B_n, C_n, D_n. \end{cases} \tag{67}$$

We give the condensed

Theorem 22 *Let $f \in \Omega$, $0 < \beta < 1$, $x \in Y$; n, $N \in \mathbb{N}$ with $n^{1-\beta} > 2$, $\lambda > 0$. Then*

(i)

$$\left\| L_n(f, x) - f(x) \right\|_\gamma \leq c_N \left[\omega_1(f, \varphi(n)) + \left(1 - erf \lambda \left(n^{1-\beta} - 2\right)\right) \left\| \|f\|_\gamma \right\|_\infty \right] =: \tau(n), \tag{68}$$

where ω_1 is for $p = \infty$,

and

(ii)

$$\left\| \left\| L_n(f) - f \right\|_\gamma \right\|_\infty \leq \tau(n) \to 0, \text{ as } n \to \infty. \tag{69}$$

For f uniformly continuous and in Ω we obtain

$$\lim_{n \to \infty} L_n(f) = f,$$

pointwise and uniformly.

Proof. By Theorems 13, 14, 15, 16. ■

Next we do iterated neural network approximation (see also [9]).

We make

Remark 23 *Let $r \in \mathbb{N}$ and L_n as above. We observe that*

$$L_n^r f - f = \left(L_n^r f - L_n^{r-1} f\right) + \left(L_n^{r-1} f - L_n^{r-2} f\right) +$$

$$\left(L_n^{r-2} f - L_n^{r-3} f\right) + \ldots + \left(L_n^2 f - L_n f\right) + \left(L_n f - f\right).$$

Then

$$\left\| \left\| L_n^r f - f \right\|_\gamma \right\|_\infty \leq \left\| \left\| L_n^r f - L_n^{r-1} f \right\|_\gamma \right\|_\infty + \left\| \left\| L_n^{r-1} f - L_n^{r-2} f \right\|_\gamma \right\|_\infty +$$

$$\left\| \left\| L_n^{r-2} f - L_n^{r-3} f \right\|_\gamma \right\|_\infty + \ldots + \left\| \left\| L_n^2 f - L_n f \right\|_\gamma \right\|_\infty + \left\| \left\| L_n f - f \right\|_\gamma \right\|_\infty =$$

$$\left\| \left\| L_n^{r-1} (L_n f - f) \right\|_\gamma \right\|_\infty + \left\| \left\| L_n^{r-2} (L_n f - f) \right\|_\gamma \right\|_\infty + \left\| \left\| L_n^{r-3} (L_n f - f) \right\|_\gamma \right\|_\infty$$

$$+ \ldots + \left\| \left\| L_n (L_n f - f) \right\|_\gamma \right\|_\infty + \left\| \left\| L_n f - f \right\|_\gamma \right\|_\infty \leq r \left\| \left\| L_n f - f \right\|_\gamma \right\|_\infty. \tag{70}$$

That is

$$\left\| \left\| L_n^r f - f \right\|_\gamma \right\|_\infty \leq r \left\| \left\| L_n f - f \right\|_\gamma \right\|_\infty. \tag{71}$$

We give

Theorem 24 *All here as in Theorem 22 and* $r \in \mathbb{N}$, $\tau(n)$ *as in (68). Then*

$$\left\| \|L_n^r f - f\|_\gamma \right\|_\infty \leq r\tau(n). \tag{72}$$

So that the speed of convergence to the unit operator of L_n^r *is not worse than of* L_n.

Proof. By (71) and (68). ■

We make

Remark 25 *Let* $m_1, ..., m_r \in \mathbb{N} : m_1 \leq m_2 \leq ... \leq m_r$, $0 < \beta < 1$, $f \in \Omega$. *Then* $\varphi(m_1) \geq \varphi(m_2) \geq ... \geq \varphi(m_r)$, φ *as in (65).*

Therefore

$$\omega_1(f, \varphi(m_1)) \geq \omega_1(f, \varphi(m_2)) \geq ... \geq \omega_1(f, \varphi(m_r)). \tag{73}$$

Assume further that $m_i^{1-\beta} > 2$, $i = 1, ..., r$. *Then*

$$\frac{1 - erf\lambda\left(m_1^{1-\beta} - 2\right)}{2} \geq \frac{1 - erf\lambda\left(m_2^{1-\beta} - 2\right)}{2} \geq ... \geq \frac{1 - erf\lambda\left(m_r^{1-\beta} - 2\right)}{2}. \tag{74}$$

Let L_{m_i} *as above,* $i = 1, ..., r$, *all of the same kind.*

We write

$$\begin{gathered} L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_2}\left(L_{m_1}f\right)\right)\right) - f = \\ L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_2}\left(L_{m_1}f\right)\right)\right) - L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_2}f\right)\right) + \\ L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_2}f\right)\right) - L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_3}f\right)\right) + \\ L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_3}f\right)\right) - L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_4}f\right)\right) + ... + \\ L_{m_r}\left(L_{m_{r-1}}f\right) - L_{m_r}f + L_{m_r}f - f = \\ L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_2}\right)\right)\left(L_{m_1}f - f\right) + L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_3}\right)\right)\left(L_{m_2}f - f\right) + \\ L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_4}\right)\right)\left(L_{m_3}f - f\right) + ... + L_{m_r}\left(L_{m_{r-1}}f - f\right) + L_{m_r}f - f. \end{gathered} \tag{75}$$

Hence by the triangle inequality property of $\left\| \|\cdot\|_\gamma \right\|_\infty$ *we get*

$$\begin{gathered} \left\| \left\| L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_2}\left(L_{m_1}f\right)\right)\right) - f \right\|_\gamma \right\|_\infty \leq \\ \left\| \left\| L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_2}\right)\right)\left(L_{m_1}f - f\right) \right\|_\gamma \right\|_\infty + \\ \left\| \left\| L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_3}\right)\right)\left(L_{m_2}f - f\right) \right\|_\gamma \right\|_\infty + \\ \left\| \left\| L_{m_r}\left(L_{m_{r-1}}\left(...L_{m_4}\right)\right)\left(L_{m_3}f - f\right) \right\|_\gamma \right\|_\infty + ... + \end{gathered}$$

$$\left\| \left\| L_{m_r} \left(L_{m_{r-1}} f - f \right) \right\|_\gamma \right\|_\infty + \left\| \left\| L_{m_r} f - f \right\|_\gamma \right\|_\infty$$

(repeatedly applying (61))

$$\leq \left\| \left\| L_{m_1} f - f \right\|_\gamma \right\|_\infty + \left\| \left\| L_{m_2} f - f \right\|_\gamma \right\|_\infty + \left\| \left\| L_{m_3} f - f \right\|_\gamma \right\|_\infty + ...+$$

$$\left\| \left\| L_{m_{r-1}} f - f \right\|_\gamma \right\|_\infty + \left\| \left\| L_{m_r} f - f \right\|_\gamma \right\|_\infty = \sum_{i=1}^{r} \left\| \left\| L_{m_i} f - f \right\|_\gamma \right\|_\infty . \tag{76}$$

That is, we proved

$$\left\| \left\| L_{m_r} \left(L_{m_{r-1}} \left(...L_{m_2} \left(L_{m_1} f \right) \right) \right) - f \right\|_\gamma \right\|_\infty \leq \sum_{i=1}^{r} \left\| \left\| L_{m_i} f - f \right\|_\gamma \right\|_\infty . \tag{77}$$

We give

Theorem 26 *Let $f \in \Omega$; $N, m_1, m_2, ..., m_r \in \mathbb{N} : m_1 \leq m_2 \leq ... \leq m_r$, $0 < \beta < 1$; $m_i^{1-\beta} > 2$, $i = 1, ..., r$, $x \in Y$, and let $(L_{m_1}, ..., L_{m_r})$ as $(A_{m_1}, ..., A_{m_r})$ or $(B_{m_1}, ..., B_{m_r})$ or $(C_{m_1}, ..., C_{m_r})$ or $(D_{m_1}, ..., D_{m_r})$, $p = \infty$. Then*

$$\left\| L_{m_r} \left(L_{m_{r-1}} \left(...L_{m_2} \left(L_{m_1} f \right) \right) \right) (x) - f (x) \right\|_\gamma \leq$$

$$\left\| \left\| L_{m_r} \left(L_{m_{r-1}} \left(...L_{m_2} \left(L_{m_1} f \right) \right) \right) - f \right\|_\gamma \right\|_\infty \leq$$

$$\sum_{i=1}^{r} \left\| \left\| L_{m_i} f - f \right\|_\gamma \right\|_\infty \leq$$

$$c_N \sum_{i=1}^{r} \left[\omega_1 \left(f, \varphi \left(m_i \right) \right) + \left(1 - erf\lambda \left(m_i^{1-\beta} - 2 \right) \right) \left\| \left\| f \right\|_\gamma \right\|_\infty \right] \leq$$

$$r c_N \left[\omega_1 \left(f, \varphi \left(m_1 \right) \right) + \left(1 - erf\lambda \left(m_1^{1-\beta} - 2 \right) \right) \left\| \left\| f \right\|_\gamma \right\|_\infty \right] . \tag{78}$$

Clearly, we notice that the speed of convergence to the unit operator of the multiply iterated operator is not worse than the speed of L_{m_1}.

Proof. Using (77), (73), (74) and (68), (69). ■

References

[1] G.A. Anastassiou, *Moments in Probability and Approximation Theory*, Pitman Research Notes in Math., Vol. 287, Longman Sci. & Tech., Harlow, U.K., 1993.

[2] G.A. Anastassiou, *Rate of convergence of some neural network operators to the unit-univariate case*, J. Math. Anal. Appli. 212 (1997), 237-262.

[3] G.A. Anastassiou, *Quantitative Approximations*, Chapman&Hall/CRC, Boca Raton, New York, 2001.

[4] G.A. Anastassiou, *Inteligent Systems: Approximation by Artificial Neural Networks*, Intelligent Systems Reference Library, Vol. 19, Springer, Heidelberg, 2011.

[5] G.A. Anastassiou, *Univariate hyperbolic tangent neural network approximation*, Mathematics and Computer Modelling, 53(2011), 1111-1132.

[6] G.A. Anastassiou, *Multivariate hyperbolic tangent neural network approximation*, Computers and Mathematics 61(2011), 809-821.

[7] G.A. Anastassiou, *Multivariate sigmoidal neural network approximation*, Neural Networks 24(2011), 378-386.

[8] G.A. Anastassiou, *Univariate sigmoidal neural network approximation*, J. of Computational Analysis and Applications, Vol. 14, No. 4, 2012, 659-690.

[9] G.A. Anastassiou, *Approximation by neural networks iterates,* Advances in Applied Mathematics and Approximation Theory, pp. 1-20, Springer Proceedings in Math. & Stat., Springer, New York, 2013, Eds. G. Anastassiou, O. Duman.

[10] G.A. Anastassiou, *Intelligent Systems II: Complete Approximation by Neural Network Operators*, Springer, Heidelberg, New York, 2016.

[11] G.A. Anastassiou, *Intelligent Computations: Abstract Fractional Calculus, Inequalities, Approximations*, Springer, Heidelberg, New York, 2018.

[12] G.A. Anastassiou, General Multivariate arctangent function activated neural network approximations, J. Numer. Anal. Approx Theory, 51(1) (2022), 37-66.

[13] G.A. Anastassiou, *General sigmoid based Banach space valued neural network approximation*, J. Computational Analysis and Applications, 31 (4) (2023), 520-534.

[14] Z. Chen and F. Cao, *The approximation operators with sigmoidal functions*, Computers and Mathematics with Applications, 58 (2009), 758-765.

[15] D. Costarelli, R. Spigler, *Approximation results for neural network operators activated by sigmoidal functions*, Neural Networks 44 (2013), 101-106.

[16] D. Costarelli, R. Spigler, *Multivariate neural network operators with sigmoidal activation functions*, Neural Networks 48 (2013), 72-77.

[17] S. Haykin, *Neural Networks: A Comprehensive Foundation* (2 ed.), Prentice Hall, New York, 1998.

[18] W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bulletin of Mathematical Biophysics, 7 (1943), 115-133.

[19] T.M. Mitchell, *Machine Learning*, WCB-McGraw-Hill, New York, 1997.