Parametrized Gudermannian function induced Banach space valued ordinary and fractional neural networks approximations

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

Abstract

Here we examine the univariate quantitative approximation, ordinary and fractional, of Banach space valued continuous functions on a compact interval or all the real line by quasi-interpolation Banach space valued neural network operators. These approximations are derived by establishing Jackson type inequalities involving the modulus of continuity of the engaged function or its Banach space valued high order derivative or fractional derivatives. Our operators are defined by using a density function generated by a parametrized Gudermannian sigmoid function. The approximations are pointwise and of the uniform norm. The related Banach space valued feed-forward neural networks are with one hidden layer.

2020 AMS Mathematics Subject Classification : 26A33, 41A17, 41A25, 41A30, 46B25.
Keywords and Phrases: Parametrized Gudermannian sigmoid function, Banach space valued neural network approximation, Banach space valued quasi-interpolation operator, modulus of continuity, Banach space valued Caputo fractional derivative, Banach space valued fractional approximation.

1 Introduction

The author in [1] and [2], see Chapters 2-5, was the first to establish neural network approximation to continuous functions with rates by very specifically defined neural network operators of Cardaliagnet-Euvrard and "Squashing" types, 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 suport. Also in [2] 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.

The author inspired by [15], continued his studies on neural networks approximation by introducing and using the proper quasi-interpolation operators of sigmoidal and hyperbolic tangent type which resulted into [3] - [7], by treating both the univariate and multivariate cases. He did also the corresponding fractional case [8].

In this article we are greatly inspired by the related works [16], [17].

The author here performs parametrized Gudermannian function based neural network approximations to continuous functions over compact intervals of the real line or over the whole $\mathbb{R}$ with values to an arbitrary Banach space $(X, \|\cdot\|)$. Finally he treats completely the related X-valued fractional approximation. All convergences here are with rates expressed via the modulus of continuity of the involved function or its X-valued high order derivative, or X-valued fractional derivatives and given by very tight Jackson type inequalities.

Our compact intervals are not necessarily symmetric to the origin. Some of our upper bounds to error quantity are very flexible and general. In preparation to prove our results we establish important properties of the basic density function defining our operators which is induced by a parametrized Gudermannian sigmoid function.

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

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

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, $c_j \in X$ 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 neural network models, the activation function is derived by the Gudermannian sigmoid functions. About neural networks in general read [18], [19], [21]. See also [9] for a complete study of real valued approximation by neural network operators.

2 Background

Here we consider the Gudermannian function ([23]) $gd(x)$ which is defined as follows

$$gd(x) := \int_0^x \frac{dt}{\cosh t} = 2\arctan\left(\tanh\left(\frac{x}{2}\right)\right), \forall\, x \in \mathbb{R}. \tag{1}$$

Let $\lambda > 0$, then

$$gd(\lambda x) = \int_0^{\lambda x} \frac{dt}{\cosh t} = 2\arctan\left(\tanh\left(\frac{\lambda x}{2}\right)\right). \tag{2}$$

We will use the following normalized and parametrized function

$$f_\lambda(x) := \frac{2}{\pi} gd(\lambda x) = \frac{4}{\pi}\arctan\left(\tanh\left(\frac{\lambda x}{2}\right)\right) = \tag{3}$$

$$\frac{2}{\pi}\int_0^{\lambda x} \frac{dt}{\cosh t} = \frac{4}{\pi}\int_0^{\lambda x} \frac{dt}{e^t + e^{-t}}, \ x \in \mathbb{R}$$

We will prove that f_λ is a generator sigmoid function with the general properties as in [14]. When $0 < \lambda < 1$, f_λ is expected to outperform ReLu and Leaky ReLu activation functions.

We notice that

$$\left(\frac{2}{\pi} gd(x)\right)' = \frac{2}{\pi \cosh x} > 0,$$

and

$$f_\lambda'(x) = \left(\frac{2}{\pi} gd(\lambda x)\right)' = \frac{2\lambda}{\pi \cosh \lambda x} > 0, \quad \forall\, x \in \mathbb{R}. \tag{4}$$

Hence f_λ is strictly increasing on $\mathbb{R}$.

Furthermore we have

$$f_\lambda''(x) = -\frac{2\lambda^2}{\pi}\frac{\sinh \lambda x}{(\cosh \lambda x)^2}, \quad \forall\, x \in \mathbb{R}. \tag{5}$$

Notice that

$$f_\lambda''(x) > 0 \ \text{ for } x < 0, \text{ and}$$

$$f_\lambda''(x) < 0 \ \text{ for } x > 0, \text{ and}$$

$$f_\lambda''(0) = 0.$$

Therefore f_λ is stritly concave up for $x < 0$, and f_λ is striclty concave down for $x > 0$, and $f_\lambda(0) = 0$, with $(0, 0)$ the inflection point.

Let $x \to +\infty$, then $\tanh\left(\frac{\lambda x}{2}\right) \to 1$ and $\arctan\left(\tanh\left(\frac{\lambda x}{2}\right)\right) \to \frac{\pi}{4}$. Let $x \to -\infty$, then $\tanh\left(\frac{\lambda x}{2}\right) \to -1$ and $\arctan\left(\tanh\left(\frac{\lambda x}{2}\right)\right) \to -\frac{\pi}{4}$.

Clearly, then $f_\lambda(+\infty) = 1$ and $f_\lambda(-\infty) = -1$, so that $y = \pm 1$ are horizontal asymptotes for f_λ.

Also it is $f_\lambda(x) \geq 0$ for $x \geq 0$, and $f_\lambda(x) < 0$ for $x < 0$. Obviously then $f_\lambda : \mathbb{R} \to [-1,1]$, with $f_\lambda'' \in C(\mathbb{R})$.

Notice that $\tanh(-x) = -\tanh x$ and $\arctan(-x) = -\arctan x$, $x \in \mathbb{R}$.

We have that

$$f_\lambda(-x) = \frac{4}{\pi}\arctan\left(\tanh\left(-\frac{\lambda x}{2}\right)\right) = \frac{4}{\pi}\arctan\left(-\tanh\left(\frac{\lambda x}{2}\right)\right) =$$

$$-\frac{4}{\pi}\arctan\left(\tanh\left(\frac{\lambda x}{2}\right)\right) = -f_\lambda(x),$$

proving

$$f_\lambda(-x) = -f_\lambda(x), \quad \forall\, x \in \mathbb{R}. \tag{6}$$

So, indeed, f_λ is a sigmoid function as in [14].

So, all the theory of [14] applies here for f_λ, etc.

We consider the activation function

$$\psi(x) := \frac{1}{4}\left(f_\lambda(x+1) - f_\lambda(x-1)\right), \quad x \in \mathbb{R}, \tag{7}$$

As in [13], p. 285, and [14], we get that $\psi(-x) = \psi(x)$, thus ψ is an even function. Since $x+1 > x-1$, then $f_\lambda(x+1) > f_\lambda(x-1)$, and $\psi(x) > 0$, all $x \in \mathbb{R}$.

We see that

$$\psi(0) = \frac{f_\lambda(1)}{2} = \frac{gd(\lambda)}{\pi}. \tag{8}$$

Let $x > 1$, we have that

$$\psi'(x) = \frac{1}{4}\left(f_\lambda'(x+1) - f_\lambda'(x-1)\right) < 0,$$

by f_λ' being strictly decreasing over $[0,+\infty)$.

Let now $0 < x < 1$, then $1-x > 0$ and $0 < 1-x < 1+x$. It holds $f_\lambda'(x-1) = f_\lambda'(1-x) > f_\lambda'(x+1)$, so that again $\psi'(x) < 0$. Consequently ψ is stritly decreasing on $(0,+\infty)$.

Clearly, ψ is strictly increasing on $(-\infty,0)$, and $\psi'(0) = 0$.

See that

$$\lim_{x\to+\infty}\psi(x) = \frac{1}{4}\left(f_\lambda(+\infty) - f_\lambda(+\infty)\right) = 0, \tag{9}$$

and

$$\lim_{x\to-\infty}\psi(x) = \frac{1}{4}\left(f_\lambda(-\infty) - f_\lambda(-\infty)\right) = 0. \tag{10}$$

That is the x-axis is the horizontal asymptote on ψ.

Conclusion, ψ is a bell symmetric function with maximum

$$\psi(0) = \frac{gd(\lambda)}{\pi}.$$

We need

Theorem 1 *(by [14]) We have that*

$$\sum_{i=-\infty}^{\infty} \psi(x-i) = 1, \quad \forall\, x \in \mathbb{R}. \tag{11}$$

Theorem 2 *(by [14]) It holds*

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

Thus $\psi(x)$ is a density function on $\mathbb{R}$.

We give

Theorem 3 *(by [14]) Let $0 < \alpha < 1$, and $n \in \mathbb{N}$ with $n^{1-\alpha} > 2$. It holds*

$$\sum_{\substack{k=-\infty \\ : |nx-k| \geq n^{1-\alpha}}}^{\infty} \psi(nx-k) < \frac{\left(1 - f_\lambda\left(n^{1-\alpha}-2\right)\right)}{2} = \frac{\left(\pi - 2gd\left(\lambda\left(n^{1-\alpha}-2\right)\right)\right)}{2\pi}. \tag{13}$$

Notice that

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

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

We further give

Theorem 4 *(by [14]) Let $x \in [a,b] \subset \mathbb{R}$ and $n \in \mathbb{N}$ so that $\lceil na \rceil \leq \lfloor nb \rfloor$. It holds*

$$\frac{1}{\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \psi(nx-k)} < \frac{1}{\psi(1)} = \frac{4}{f_\lambda(2)} = \frac{2\pi}{gd(2\lambda)}, \quad \forall\, x \in [a,b]. \tag{14}$$

Remark 5 *(by [14]) We have that*

$$\lim_{n\to\infty} \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \psi(nx-k) \neq 1, \tag{15}$$

for at least some $x \in [a,b]$.

See also [13], p. 290, same reasoning.

Note 6 *For large enough n we always obtain $\lceil na \rceil \leq \lfloor nb \rfloor$. Also $a \leq \frac{k}{n} \leq b$, iff $\lceil na \rceil \leq k \leq \lfloor nb \rfloor$. In general it holds (by (11))*

$$\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \psi(nx-k) \leq 1. \tag{16}$$

Let $(X, \|\cdot\|)$ be a Banach space.

Definition 7 *Let $f \in C([a,b], X)$ and $n \in \mathbb{N} : \lceil na \rceil \leq \lfloor nb \rfloor$. We introduce and define the X-valued linear neural network operators*

$$A_n(f,x) := \frac{\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} f\left(\frac{k}{n}\right) \psi(nx-k)}{\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \psi(nx-k)}, \quad x \in [a,b]. \tag{17}$$

Clearly here $A_n(f,x) \in C([a,b], X)$. For convenience we use the same A_n for real valued function when needed. We study here the pointwise and uniform convergence of $A_n(f,x)$ to $f(x)$ with rates.

For convenience also we call

$$A_n^*(f,x) := \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} f\left(\frac{k}{n}\right) \psi(nx-k), \tag{18}$$

(similarly A_n^* can be defined for real valued function) that is

$$A_n(f,x) = \frac{A_n^*(f,x)}{\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \psi(nx-k)}. \tag{19}$$

So that

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

$$= \frac{A_n^*(f,x) - f(x)\left(\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \psi(nx-k)\right)}{\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \psi(nx-k)}. \tag{20}$$

Consequently we derive

$$\|A_n(f,x) - f(x)\| \leq \frac{2\pi}{gd(2\lambda)} \left\| A_n^*(f,x) - f(x)\left(\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \psi(nx-k)\right) \right\|. \tag{21}$$

That is

$$\|A_n(f,x) - f(x)\| \leq \frac{2\pi}{gd(2\lambda)} \left\| \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \left(f\left(\frac{k}{n}\right) - f(x)\right) \psi(nx-k) \right\|. \tag{22}$$

We will estimate the right hand side of (22).

For that we need, for $f \in C([a,b], X)$ the first modulus of continuity

$$\omega_1(f,\delta)_{[a,b]} := \omega_1(f,\delta) := \sup_{\substack{x,y \in [a,b] \\ |x-y| \leq \delta}} \|f(x) - f(y)\|, \quad \delta > 0. \tag{23}$$

Similarly, it is defined ω_1 for $f \in C_{uB}(\mathbb{R}, X)$ (uniformly continuous and bounded functions from $\mathbb{R}$ into X), for $f \in C_B(\mathbb{R}, X)$ (continuous and bounded X-valued) and for $f \in C_u(\mathbb{R}, X)$ (uniformly continuous).

The fact $f \in C([a,b], X)$ or $f \in C_u(\mathbb{R}, X)$, is equivalent to $\lim_{\delta \to 0} \omega_1(f, \delta) = 0$, see [11].

Definition 8 *When $f \in C_{uB}(\mathbb{R}, X)$, or $f \in C_B(\mathbb{R}, X)$, we define*

$$\overline{A}_n(f, x) := \sum_{k=-\infty}^{\infty} f\left(\frac{k}{n}\right) \psi(nx - k), \quad n \in \mathbb{N}, \ x \in \mathbb{R}, \tag{24}$$

the X-valued quasi-interpolation neural network operator.

Remark 9 *(by [14]) We have that the series $\sum_{k=-\infty}^{\infty} f\left(\frac{k}{n}\right) \psi(nx-k)$ is absolutely convergent in X, hence it is convergent in X and $\overline{A}_n(f,x) \in X$.*

We denote by $\|f\|_\infty := \sup_{x \in [a,b]} \|f(x)\|$, for $f \in C([a,b], X)$, similarly is defined for $f \in C_B(\mathbb{R}, X)$.

3 Main Results

We present a series of X-valued neural network approximations to a function given with rates.

We first give

Theorem 10 *Let $f \in C([a,b], X)$, $0 < \alpha < 1$, $n \in \mathbb{N} : n^{1-\alpha} > 2$, $x \in [a,b]$. Then*

i)

$$\|A_n(f,x) - f(x)\| \leq \frac{2\pi}{gd(2\lambda)} \left[\omega_1\left(f, \frac{1}{n^\alpha}\right) + \left(1 - f_\lambda\left(n^{1-\alpha} - 2\right)\right) \|f\|_\infty\right] =: \rho, \tag{25}$$

and

ii)

$$\|A_n(f) - f\|_\infty \leq \rho. \tag{26}$$

We notice $\lim_{n \to \infty} A_n(f) = f$, pointwise and uniformly.

The speed of convergence is $\max\left(\frac{1}{n^\alpha}, \left(1 - f_\lambda\left(n^{1-\alpha} - 2\right)\right)\right)$.

Proof. As similar to [13], p. 293 is omitted, see also [14]. ∎

Next we give

Theorem 11 *Let $f \in C_B(\mathbb{R}, X)$, $0 < \alpha < 1$, $n \in \mathbb{N} : n^{1-\alpha} > 2$, $x \in \mathbb{R}$. Then*

i)

$$\left\| \overline{A}_n(f, x) - f(x) \right\| \leq \omega_1\left(f, \frac{1}{n^\alpha}\right) + \left(1 - f_\lambda\left(n^{1-\alpha} - 2\right)\right) \|f\|_\infty =: \mu, \tag{27}$$

and

ii)

$$\left\| \overline{A}_n(f) - f \right\|_\infty \leq \mu. \tag{28}$$

For $f \in C_{uB}(\mathbb{R}, X)$ we get $\lim_{n\to\infty} \overline{A}_n(f) = f$, pointwise and uniformly.

The speed of convergence is $\max\left(\frac{1}{n^\alpha}, \left(1 - f_\lambda\left(n^{1-\alpha} - 2\right)\right)\right)$.

Proof. As similar to [13], p. 294 is omitted, see also [14]. ∎

In the next we discuss high order neural network X-valued approximation by using the smoothness of f.

Theorem 12 *Let $f \in C^N([a, b], X)$, $n, N \in \mathbb{N}$, $0 < \alpha < 1$, $x \in [a, b]$ and $n^{1-\alpha} > 2$. Then*

i)

$$\|A_n(f, x) - f(x)\| \leq \frac{2\pi}{gd(2\lambda)} \left\{ \sum_{j=1}^{N} \frac{\left\|f^{(j)}(x)\right\|}{j!} \left[\frac{1}{n^{\alpha j}} + \frac{\left(1 - f_\lambda\left(n^{1-\alpha} - 2\right)\right)}{2} (b-a)^j \right] + \right. \tag{29}$$

$$\left. \left[\omega_1\left(f^{(N)}, \frac{1}{n^\alpha}\right) \frac{1}{n^{\alpha N} N!} + \frac{\left(1 - f_\lambda\left(n^{1-\alpha} - 2\right)\right) \left\|f^{(N)}\right\|_\infty (b-a)^N}{N!} \right] \right\},$$

ii) assume further $f^{(j)}(x_0) = 0$, $j = 1, ..., N$, for some $x_0 \in [a, b]$, it holds

$$\|A_n(f, x_0) - f(x_0)\| \leq \frac{2\pi}{gd(2\lambda)}$$

$$\left\{ \omega_1\left(f^{(N)}, \frac{1}{n^\alpha}\right) \frac{1}{n^{\alpha N} N!} + \frac{\left(1 - f_\lambda\left(n^{1-\alpha} - 2\right)\right) \left\|f^{(N)}\right\|_\infty (b-a)^N}{N!} \right\}, \tag{30}$$

and

iii)

$$\|A_n(f) - f\|_\infty \leq \frac{2\pi}{gd(2\lambda)} \left\{ \sum_{j=1}^{N} \frac{\left\|f^{(j)}\right\|_\infty}{j!} \left[\frac{1}{n^{\alpha j}} + \frac{\left(1 - f_\lambda\left(n^{1-\alpha} - 2\right)\right)}{2} (b-a)^j \right] + \right.$$

$$\left. \left[\omega_1\left(f^{(N)}, \frac{1}{n^\alpha}\right) \frac{1}{n^{\alpha N} N!} + \frac{\left(1 - f_\lambda\left(n^{1-\alpha} - 2\right)\right) \left\|f^{(N)}\right\|_\infty (b-a)^N}{N!} \right] \right\}. \tag{31}$$

Again we obtain $\lim_{n\to\infty} A_n(f) = f$, pointwise and uniformly.

8

Proof. As similar to [13], pp. 296-301 is omitted, see also [14]. ■

All integrals from now on are of Bochner type [20].

We need

Definition 13 *([12]) Let $[a,b] \subset \mathbb{R}$, X be a Banach space, $\alpha > 0$; $m = \lceil \alpha \rceil \in \mathbb{N}$, ($\lceil \cdot \rceil$ is the ceiling of the number), $f : [a,b] \to X$. We assume that $f^{(m)} \in L_1([a,b], X)$. We call the Caputo-Bochner left fractional derivative of order α:*

$$\left(D_{*a}^{\alpha} f\right)(x) := \frac{1}{\Gamma(m-\alpha)} \int_a^x (x-t)^{m-\alpha-1} f^{(m)}(t)\, dt, \quad \forall\, x \in [a,b]. \tag{32}$$

If $\alpha \in \mathbb{N}$, we set $D_{*a}^{\alpha} f := f^{(m)}$ the ordinary X-valued derivative (defined similar to numerical one, see [22], p. 83), and also set $D_{*a}^{0} f := f$.

By [12], $\left(D_{*a}^{\alpha} f\right)(x)$ exists almost everywhere in $x \in [a,b]$ and $D_{*a}^{\alpha} f \in L_1([a,b], X)$.

If $\left\| f^{(m)} \right\|_{L_\infty([a,b],X)} < \infty$, then by [12], $D_{*a}^{\alpha} f \in C([a,b], X)$, hence $\left\| D_{*a}^{\alpha} f \right\| \in C([a,b])$.

Definition 14 *([10]) Let $[a,b] \subset \mathbb{R}$, X be a Banach space, $\alpha > 0$, $m := \lceil \alpha \rceil$. We assume that $f^{(m)} \in L_1([a,b], X)$, where $f : [a,b] \to X$. We call the Caputo-Bochner right fractional derivative of order α:*

$$\left(D_{b-}^{\alpha} f\right)(x) := \frac{(-1)^m}{\Gamma(m-\alpha)} \int_x^b (z-x)^{m-\alpha-1} f^{(m)}(z)\, dz, \quad \forall\, x \in [a,b]. \tag{33}$$

We observe that $\left(D_{b-}^{m} f\right)(x) = (-1)^m f^{(m)}(x)$, for $m \in \mathbb{N}$, and $\left(D_{b-}^{0} f\right)(x) = f(x)$.

By [10], $\left(D_{b-}^{\alpha} f\right)(x)$ exists almost everywhere on $[a,b]$ and $\left(D_{b-}^{\alpha} f\right) \in L_1([a,b], X)$.

If $\left\| f^{(m)} \right\|_{L_\infty([a,b],X)} < \infty$, and $\alpha \notin \mathbb{N}$, by [10], $D_{b-}^{\alpha} f \in C([a,b], X)$, hence $\left\| D_{b-}^{\alpha} f \right\| \in C([a,b])$.

We present the following X-valued fractional approximation result by neural networks.

Theorem 15 *Let $\alpha > 0$, $N = \lceil \alpha \rceil$, $\alpha \notin \mathbb{N}$, $f \in C^N([a,b], X)$, $0 < \beta < 1$, $x \in [a,b]$, $n \in \mathbb{N} : n^{1-\beta} > 2$. Then*

i)

$$\left\| A_n(f,x) - \sum_{j=1}^{N-1} \frac{f^{(j)}(x)}{j!} A_n\left((\cdot - x)^j\right)(x) - f(x) \right\| \leq$$

$$\frac{2\pi}{gd(2\lambda)\,\Gamma(\alpha+1)} \left\{ \frac{\left(\omega_1\left(D_{x-}^{\alpha} f, \frac{1}{n^\beta}\right)_{[a,x]} + \omega_1\left(D_{*x}^{\alpha} f, \frac{1}{n^\beta}\right)_{[x,b]}\right)}{n^{\alpha\beta}} + \right.$$

9

$$\left.\left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)\left(\left\|D_{x-}^{\alpha}f\right\|_{\infty,[a,x]}(x-a)^{\alpha}+\left\|D_{*x}^{\alpha}f\right\|_{\infty,[x,b]}(b-x)^{\alpha}\right)\right\}, \tag{34}$$

ii) if $f^{(j)}(x)=0$, *for* $j=1,\dots,N-1$, *we have*

$$\left\|A_{n}(f,x)-f(x)\right\| \leq \frac{2\pi}{gd(2\lambda)\,\Gamma(\alpha+1)}$$

$$\left\{\frac{\left(\omega_{1}\left(D_{x-}^{\alpha}f,\frac{1}{n^{\beta}}\right)_{[a,x]}+\omega_{1}\left(D_{*x}^{\alpha}f,\frac{1}{n^{\beta}}\right)_{[x,b]}\right)}{n^{\alpha\beta}}+\right.$$

$$\left.\left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)\left(\left\|D_{x-}^{\alpha}f\right\|_{\infty,[a,x]}(x-a)^{\alpha}+\left\|D_{*x}^{\alpha}f\right\|_{\infty,[x,b]}(b-x)^{\alpha}\right)\right\}, \tag{35}$$

iii)

$$\left\|A_{n}(f,x)-f(x)\right\| \leq \frac{2\pi}{gd(2\lambda)}$$

$$\left\{\sum_{j=1}^{N-1}\frac{\left\|f^{(j)}(x)\right\|}{j!}\left\{\frac{1}{n^{\beta j}}+(b-a)^{j}\left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)\right\}+\right.$$

$$\frac{1}{\Gamma(\alpha+1)}\left\{\frac{\left(\omega_{1}\left(D_{x-}^{\alpha}f,\frac{1}{n^{\beta}}\right)_{[a,x]}+\omega_{1}\left(D_{*x}^{\alpha}f,\frac{1}{n^{\beta}}\right)_{[x,b]}\right)}{n^{\alpha\beta}}+\right.$$

$$\left.\left.\left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)\left(\left\|D_{x-}^{\alpha}f\right\|_{\infty,[a,x]}(x-a)^{\alpha}+\left\|D_{*x}^{\alpha}f\right\|_{\infty,[x,b]}(b-x)^{\alpha}\right)\right\}\right\}, \tag{36}$$

$\forall\ x\in[a,b]$,

and

iv)

$$\left\|A_{n}f-f\right\|_{\infty} \leq \frac{2\pi}{gd(2\lambda)}$$

$$\left\{\sum_{j=1}^{N-1}\frac{\left\|f^{(j)}\right\|_{\infty}}{j!}\left\{\frac{1}{n^{\beta j}}+(b-a)^{j}\left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)\right\}+\right.$$

$$\frac{1}{\Gamma(\alpha+1)}\left\{\frac{\left(\sup_{x\in[a,b]}\omega_{1}\left(D_{x-}^{\alpha}f,\frac{1}{n^{\beta}}\right)_{[a,x]}+\sup_{x\in[a,b]}\omega_{1}\left(D_{*x}^{\alpha}f,\frac{1}{n^{\beta}}\right)_{[x,b]}\right)}{n^{\alpha\beta}}+\right.$$

10

$$\left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)(b-a)^{\alpha}\left(\sup_{x\in[a,b]}\left\|D_{x-}^{\alpha}f\right\|_{\infty,[a,x]}+\sup_{x\in[a,b]}\left\|D_{*x}^{\alpha}f\right\|_{\infty,[x,b]}\right)\Bigg\}\Bigg\}. \tag{37}$$

Above, when $N=1$ the sum $\sum_{j=1}^{N-1}\cdot=0$.

As we see here we obtain X-valued fractionally type pointwise and uniform convergence with rates of $A_n \to I$ the unit operator, as $n\to\infty$.

Proof. It is very lengthy, as similar to [13], pp. 305-316, is omitted, see also [14]. ■

Next we apply Theorem 15 for $N=1$.

Theorem 16 *Let $0<\alpha,\beta<1$, $f\in C^{1}([a,b],X)$, $x\in[a,b]$, $n\in\mathbb{N}: n^{1-\beta}>2$. Then*

i)

$$\left\|A_n(f,x)-f(x)\right\|\leq$$

$$\frac{2\pi}{gd(2\lambda)\Gamma(\alpha+1)}\left\{\frac{\left(\omega_1\left(D_{x-}^{\alpha}f,\frac{1}{n^{\beta}}\right)_{[a,x]}+\omega_1\left(D_{*x}^{\alpha}f,\frac{1}{n^{\beta}}\right)_{[x,b]}\right)}{n^{\alpha\beta}}+\right.$$

$$\left.\left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)\left(\left\|D_{x-}^{\alpha}f\right\|_{\infty,[a,x]}(x-a)^{\alpha}+\left\|D_{*x}^{\alpha}f\right\|_{\infty,[x,b]}(b-x)^{\alpha}\right)\right\}, \tag{38}$$

and

ii)

$$\left\|A_n f-f\right\|_{\infty}\leq\frac{2\pi}{gd(2\lambda)\Gamma(\alpha+1)}$$

$$\left\{\frac{\left(\sup_{x\in[a,b]}\omega_1\left(D_{x-}^{\alpha}f,\frac{1}{n^{\beta}}\right)_{[a,x]}+\sup_{x\in[a,b]}\omega_1\left(D_{*x}^{\alpha}f,\frac{1}{n^{\beta}}\right)_{[x,b]}\right)}{n^{\alpha\beta}}+\right.$$

$$\left.\left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)(b-a)^{\alpha}\left(\sup_{x\in[a,b]}\left\|D_{x-}^{\alpha}f\right\|_{\infty,[a,x]}+\sup_{x\in[a,b]}\left\|D_{*x}^{\alpha}f\right\|_{\infty,[x,b]}\right)\right\}. \tag{39}$$

When $\alpha=\frac{1}{2}$ we derive

Corollary 17 *Let $0<\beta<1$, $f\in C^{1}([a,b],X)$, $x\in[a,b]$, $n\in\mathbb{N}: n^{1-\beta}>2$. Then*

i)

$$\left\|A_n(f,x)-f(x)\right\|\leq$$

$$\frac{4\sqrt{\pi}}{gd\left(2\lambda\right)}\left\{\frac{\left(\omega_1\left(D_{x-}^{\frac{1}{2}}f,\frac{1}{n^{\beta}}\right)_{[a,x]}+\omega_1\left(D_{*x}^{\frac{1}{2}}f,\frac{1}{n^{\beta}}\right)_{[x,b]}\right)}{n^{\frac{\beta}{2}}}+\right.$$

$$\left.\left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)\left(\left\|D_{x-}^{\frac{1}{2}}f\right\|_{\infty,[a,x]}\sqrt{(x-a)}+\left\|D_{*x}^{\frac{1}{2}}f\right\|_{\infty,[x,b]}\sqrt{(b-x)}\right)\right\}, \tag{40}$$

and

ii)

$$\left\|A_n f-f\right\|_{\infty}\leq\frac{4\sqrt{\pi}}{gd\left(2\lambda\right)}$$

$$\left\{\frac{\left(\sup_{x\in[a,b]}\omega_1\left(D_{x-}^{\frac{1}{2}}f,\frac{1}{n^{\beta}}\right)_{[a,x]}+\sup_{x\in[a,b]}\omega_1\left(D_{*x}^{\frac{1}{2}}f,\frac{1}{n^{\beta}}\right)_{[x,b]}\right)}{n^{\frac{\beta}{2}}}+\right.$$

$$\left.\left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)\sqrt{(b-a)}\left(\sup_{x\in[a,b]}\left\|D_{x-}^{\frac{1}{2}}f\right\|_{\infty,[a,x]}+\sup_{x\in[a,b]}\left\|D_{*x}^{\frac{1}{2}}f\right\|_{\infty,[x,b]}\right)\right\}<\infty. \tag{41}$$

We finish with

Remark 18 *Some convergence analysis follows:*

Let $0<\beta<1$, $f\in C^1\left([a,b],X\right)$, $x\in[a,b]$, $n\in\mathbb{N}:n^{1-\beta}>2$. *We elaborate on (41). Assume that*

$$\omega_1\left(D_{x-}^{\frac{1}{2}}f,\frac{1}{n^{\beta}}\right)_{[a,x]}\leq\frac{K_1}{n^{\beta}}, \tag{42}$$

and

$$\omega_1\left(D_{*x}^{\frac{1}{2}}f,\frac{1}{n^{\beta}}\right)_{[x,b]}\leq\frac{K_2}{n^{\beta}}, \tag{43}$$

$\forall\ x\in[a,b]$, $\forall\ n\in\mathbb{N}$, *where* $K_1,K_2>0$.

Then it holds

$$\frac{\left[\sup_{x\in[a,b]}\omega_1\left(D_{x-}^{\frac{1}{2}}f,\frac{1}{n^{\beta}}\right)_{[a,x]}+\sup_{x\in[a,b]}\omega_1\left(D_{*x}^{\frac{1}{2}}f,\frac{1}{n^{\beta}}\right)_{[x,b]}\right]}{n^{\frac{\beta}{2}}}\leq$$

$$\frac{\frac{(K_1+K_2)}{n^{\beta}}}{n^{\frac{\beta}{2}}}=\frac{(K_1+K_2)}{n^{\frac{3\beta}{2}}}=\frac{K}{n^{\frac{3\beta}{2}}}, \tag{44}$$

where $K:=K_1+K_2>0$.

12

The other summand of the right hand side of (41), for large enough n, converges to zero at the speed $\left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)$.

Then, for large enough $n \in \mathbb{N}$, by (41) and (44) and the last comment, we obtain that

$$\|A_n f - f\|_{\infty} \leq M \max\left(\frac{1}{n^{\frac{3\beta}{2}}}, \left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)\right), \tag{45}$$

where $M > 0$.

If $\frac{1}{n^{\frac{3\beta}{2}}} \geq \left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)$, *then* $\frac{1}{n^{\beta}} \geq \left(\frac{1-f_{\lambda}\left(n^{1-\beta}-2\right)}{2}\right)$, *and consequently* $\|A_n f - f\|_{\infty}$ *in (45) converges to zero faster than in Theorem 10. This because the differentiability of* f.

References

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

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

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

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

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

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

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

[8] G.A. Anastassiou, *Fractional neural network approximation*, Computers and Mathematics with Applications, 64 (2012), 1655-1676.

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

[10] G.A. Anastassiou, *Strong Right Fractional Calculus for Banach space valued functions*, 'Revista Proyecciones, Vol. 36, No. 1 (2017), 149-186.

[11] G.A. Anastassiou, *Vector fractional Korovkin type Approximations*, Dynamic Systems and Applications, 26 (2017), 81-104.

[12] G.A. Anastassiou, *A strong Fractional Calculus Theory for Banach space valued functions*, Nonlinear Functional Analysis and Applications (Korea), 22(3)(2017), 495-524.

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

[14] G.A. Anastassiou, *General sigmoid based Banach space valued neural network approximation*, accepted, J.of Computational Analysis and Applications, 2022.

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

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

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

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

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

[20] J. Mikusinski, *The Bochner integral*, Academic Press, New York, 1978.

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

[22] G.E. Shilov, *Elementary Functional Analysis*, Dover Publications, Inc., New York, 1996.

[23] E.W. Weisstein, *Gudermannian*, MathWorld.