Parametrized error function based Banach space valued univariate neural network approximation

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

Abstract

Here we research the univariate quantitative approximation of Banach space valued continuous functions on a compact interval or all the real line by quasi-interpolation Banach space valued neural network operators. We perform also the related Banach space valued fractional approximation. 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 derivaties. Our operators are defined by using a density function induced by a parametrized error function. The approximations are pointwise and with respect to the uniform norm. The related Banach space valued feed-forward neural networks are with one hidden layer. We finish with a convergence analysis.

2020 AMS Mathematics Subject Classification: 26A33, 41A17, 41A25, 41A30, 46B25.

Keywords and Phrases: parametrized error 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 Cardaliaguet-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

RGMIA Res. Rep. Coll. **26** (2023), Art. 10, 16 pp. Received 30/01/23

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].

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 parametrized error 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 error 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. Here we work for a parametrized error function. 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 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

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

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 [14].

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 [14].

We have

Theorem 1 *([14]) 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 *([14]) 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 *([14]) 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 *([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*

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

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 [14], we have that*

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

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 [14], we obtain that*

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

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

Definition 11 *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*

$$A_n(f,x) := \frac{\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} f\left(\frac{k}{n}\right) \chi(nx-k)}{\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \chi(nx-k)}, \quad x \in [a,b]. \tag{19}$$

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

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

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

$$A_n(f,x) = \frac{A_n^*(f,x)}{\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \chi(nx-k)}. \tag{21}$$

So that

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

$$= \frac{A_n^*(f,x) - f(x)\left(\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \chi(nx-k)\right)}{\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \chi(nx-k)}. \tag{22}$$

Consequently we derive

$$\left\|A_n(f,x)-f(x)\right\| \leq \frac{4}{erf(2\lambda)}\left\|A_n^*(f,x)-f(x)\left(\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\chi(nx-k)\right)\right\|. \quad (23)$$

That is

$$\left\|A_n(f,x)-f(x)\right\| \leq \frac{4}{erf(2\lambda)}\left\|\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\left(f\left(\frac{k}{n}\right)-f(x)\right)\chi(nx-k)\right\|. \quad (24)$$

We will estimate the right hand side of (24).

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

$$\omega_1(f,\delta)_{[a,b]} := \omega_1(f,\delta) := \sup_{\substack{x,y\in[a,b] \\ |x-y|\leq\delta}} \left\|f(x)-f(y)\right\|, \quad \delta>0. \quad (25)$$

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

$$\overline{A}_n(f,x) := \sum_{k=-\infty}^{\infty} f\left(\frac{k}{n}\right)\chi(nx-k), \quad n\in\mathbb{N},\ x\in\mathbb{R}, \quad (26)$$

the X-valued quasi-interpolation neural network operator.

Remark 13 *We have that*

$$\left\|f\left(\frac{k}{n}\right)\right\| \leq \|f\|_{\infty,\mathbb{R}} < +\infty,$$

and

$$\left\|f\left(\frac{k}{n}\right)\right\|\chi(nx-k) \leq \|f\|_{\infty,\mathbb{R}}\,\chi(nx-k), \quad (27)$$

and

$$\sum_{k=-\lambda}^{\lambda}\left\|f\left(\frac{k}{n}\right)\right\|\chi(nx-k) \leq \|f\|_{\infty,\mathbb{R}}\left(\sum_{k=-\lambda}^{\lambda}\chi(nx-k)\right),$$

and finally

$$\sum_{k=-\infty}^{\infty}\left\|f\left(\frac{k}{n}\right)\right\|\chi(nx-k) \leq \|f\|_{\infty,\mathbb{R}}, \quad (28)$$

a convergent in $\mathbb{R}$ series.

So the series $\sum_{k=-\infty}^{\infty} f\left(\frac{k}{n}\right)\chi(nx-k)$ 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 14 *Let $f \in C([a,b], X)$, $0 < \alpha < 1$, $n \in \mathbb{N} : n^{1-\alpha} > 2$, $\lambda > 0$, $x \in [a,b]$. Then*

i)

$$\|A_n(f,x) - f(x)\| \leq \frac{4}{erf(2\lambda)} \left[\omega_1\left(f, \frac{1}{n^\alpha}\right) + \left(1 - erf\lambda\left(n^{1-\alpha} - 2\right)\right) \|f\|_\infty \right] =: \rho, \tag{29}$$

and

ii)

$$\|A_n(f) - f\|_\infty \leq \rho. \tag{30}$$

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

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

Proof. As similar to [13], p. 293 is omitted. ■

Next we give

Theorem 15 *Let $f \in C_B(\mathbb{R}, X)$, $0 < \alpha < 1$, $n \in \mathbb{N} : n^{1-\alpha} > 2$, $\lambda > 0$, $x \in \mathbb{R}$. Then*

i)

$$\left\|\overline{A}_n(f,x) - f(x)\right\| \leq \omega_1\left(f, \frac{1}{n^\alpha}\right) + \left(1 - erf\lambda\left(n^{1-\alpha} - 2\right)\right) \|f\|_\infty =: \mu, \tag{31}$$

and

ii)

$$\left\|\overline{A}_n(f) - f\right\|_\infty \leq \mu. \tag{32}$$

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 $\max\left(\frac{1}{n^\alpha}, \left(1 - erf\lambda\left(n^{1-\alpha} - 2\right)\right)\right)$.

Proof. As similar to [13], p. 294 is omitted. ■

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

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

i)

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

$$\frac{4}{erf(2\lambda)} \left\{ \sum_{j=1}^{N} \frac{\left\|f^{(j)}(x)\right\|}{j!} \left[\frac{1}{n^{\alpha j}} + \frac{\left(1 - erf\lambda\left(n^{1-\alpha} - 2\right)\right)}{2} (b-a)^j \right] + \right. \tag{33}$$

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

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

$$\|A_n(f,x_0) - f(x_0)\| \leq \frac{4}{erf(2\lambda)}$$

$$\left\{ \omega_1 \left(f^{(N)}, \frac{1}{n^\alpha} \right) \frac{1}{n^{\alpha N} N!} + \frac{\left(1 - erf\lambda\left(n^{1-\alpha} - 2\right)\right) \left\|f^{(N)}\right\|_\infty (b-a)^N}{N!} \right\}, \tag{34}$$

and

iii)

$$\|A_n(f) - f\|_\infty \leq \frac{4}{erf(2\lambda)} \left\{ \sum_{j=1}^{N} \frac{\left\|f^{(j)}\right\|_\infty}{j!} \left[\frac{1}{n^{\alpha j}} + \frac{\left(1 - erf\lambda\left(n^{1-\alpha} - 2\right)\right)}{2} (b-a)^j \right] + \right.$$

$$\left. \left[\omega_1 \left(f^{(N)}, \frac{1}{n^\alpha} \right) \frac{1}{n^{\alpha N} N!} + \frac{\left(1 - erf\lambda\left(n^{1-\alpha} - 2\right)\right) \left\|f^{(N)}\right\|_\infty (b-a)^N}{N!} \right] \right\}. \tag{35}$$

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

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

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

We need

Definition 17 *([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:*

$$\left(D_{*a}^\alpha f\right)(x) := \frac{1}{\Gamma(m-\alpha)} \int_a^x (x-t)^{m-\alpha-1} f^{(m)}(t)\, dt, \quad \forall\, x \in [a,b]. \tag{36}$$

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], $(D_{*a}^{\alpha}f)(x)$ 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 $\|D_{*a}^{\alpha}f\| \in C([a,b])$.

We mention

Lemma 18 *([11]) Let $\alpha > 0$, $\alpha \notin \mathbb{N}$, $m = \lceil \alpha \rceil$, $f \in C^{m-1}([a,b],X)$ and $f^{(m)} \in L_\infty([a,b],X)$. Then $D_{*a}^{\alpha}f(a) = 0$.*

We mention

Definition 19 *([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:*

$$\left(D_{b-}^{\alpha}f\right)(x) := \frac{(-1)^m}{\Gamma(m-\alpha)} \int_x^b (z-x)^{m-\alpha-1} f^{(m)}(z)\,dz, \quad \forall\, x \in [a,b]. \tag{37}$$

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 need

Lemma 20 *([11]) Let $f \in C^{m-1}([a,b],X)$, $f^{(m)} \in L_\infty([a,b],X)$, $m = \lceil \alpha \rceil$, $\alpha > 0$, $\alpha \notin \mathbb{N}$. Then $D_{b-}^{\alpha}f(b) = 0$.*

Convention 21 *We assume that*

$$D_{*x_0}^{\alpha}f(x) = 0, \text{ for } x < x_0, \tag{38}$$

and

$$D_{x_0-}^{\alpha}f(x) = 0, \text{ for } x > x_0, \tag{39}$$

for all $x, x_0 \in [a,b]$.

We mention

Proposition 22 *([11]) Let $f \in C^n([a,b],X)$, $n = \lceil \nu \rceil$, $\nu > 0$. Then $D_{*a}^{\nu}f(x)$ is continuous in $x \in [a,b]$.*

Proposition 23 *([11]) Let $f \in C^m([a,b],X)$, $m = \lceil \alpha \rceil$, $\alpha > 0$. Then $D_{b-}^{\nu}f(x)$ is continuous in $x \in [a,b]$.*

We also mention

Proposition 24 *([11]) Let $f \in C^{m-1}([a,b],X)$, $f^{(m)} \in L_\infty([a,b],X)$, $m = \lceil \alpha \rceil$, $\alpha > 0$ and*

$$D_{*x_0}^{\alpha} f(x) = \frac{1}{\Gamma(m-\alpha)} \int_{x_0}^{x} (x-t)^{m-\alpha-1} f^{(m)}(t)\, dt, \tag{40}$$

for all $x, x_0 \in [a,b] : x \geq x_0$.

Then $D_{*x_0}^{\alpha} f(x)$ is continuous in x_0.

Proposition 25 *([11]) Let $f \in C^{m-1}([a,b],X)$, $f^{(m)} \in L_\infty([a,b],X)$, $m = \lceil \alpha \rceil$, $\alpha > 0$ and*

$$D_{x_0-}^{\alpha} f(x) = \frac{(-1)^m}{\Gamma(m-\alpha)} \int_{x}^{x_0} (\zeta - x)^{m-\alpha-1} f^{(m)}(\zeta)\, d\zeta, \tag{41}$$

for all $x, x_0 \in [a,b] : x_0 \geq x$.

Then $D_{x_0-}^{\alpha} f(x)$ is continuous in x_0.

Corollary 26 *([11]) Let $f \in C^m([a,b],X)$, $m = \lceil \alpha \rceil$, $\alpha > 0$, $x, x_0 \in [a,b]$. Then $D_{*x_0}^{a} f(x)$, $D_{x_0-}^{a} f(x)$ are jointly continuous functions in (x, x_0) from $[a,b]^2$ into X, X is a Banach space.*

We need

Theorem 27 *([11]) Let $f : [a,b]^2 \to X$ be jointly continuous, X is a Banach space. Consider*

$$G(x) = \omega_1(f(\cdot, x), \delta, [x,b]), \tag{42}$$

$\delta > 0$, $x \in [a,b]$.

Then G is continuous on $[a,b]$.

Theorem 28 *([11]) Let $f : [a,b]^2 \to X$ be jointly continuous, X is a Banach space. Then*

$$H(x) = \omega_1(f(\cdot, x), \delta, [a,x]), \tag{43}$$

$x \in [a,b]$, is continuous in $x \in [a,b]$, $\delta > 0$.

We present the following X-valued fractional approximation result by $erf\lambda$ based neural networks.

Theorem 29 *Let $\alpha > 0$, $N = \lceil \alpha \rceil$, $\alpha \notin \mathbb{N}$, $f \in C^N([a,b],X)$, $0 < \beta < 1$, $\lambda > 0$, $x \in [a,b]$, $n \in \mathbb{N} : n^{1-\beta} > 2$. Then*

i)

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

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

$$\left.\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)\left(\left\|D_{x-}^{\alpha}f\right\|_{\infty,[a,x]}(x-a)^{\alpha}+\left\|D_{*x}^{\alpha}f\right\|_{\infty,[x,b]}(b-x)^{\alpha}\right)\right\},\tag{44}$$

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

$$\left\|A_n(f,x)-f(x)\right\|\leq\frac{4}{\Gamma(\alpha+1)\,erf(2\lambda)}$$

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

$$\left.\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)\left(\left\|D_{x-}^{\alpha}f\right\|_{\infty,[a,x]}(x-a)^{\alpha}+\left\|D_{*x}^{\alpha}f\right\|_{\infty,[x,b]}(b-x)^{\alpha}\right)\right\},\tag{45}$$

iii)

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

$$\frac{4}{erf(2\lambda)}\left\{\sum_{j=1}^{N-1}\frac{\left\|f^{(j)}(x)\right\|}{j!}\left\{\frac{1}{n^{\beta j}}+(b-a)^{j}\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)\right\}+\right.$$

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

$$\left.\left.\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)\left(\left\|D_{x-}^{\alpha}f\right\|_{\infty,[a,x]}(x-a)^{\alpha}+\left\|D_{*x}^{\alpha}f\right\|_{\infty,[x,b]}(b-x)^{\alpha}\right)\right\}\right\},\tag{46}$$

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

and

iv)

$$\left\|A_nf-f\right\|_{\infty}\leq$$

$$\frac{4}{erf(2\lambda)}\left\{\sum_{j=1}^{N-1}\frac{\left\|f^{(j)}\right\|_{\infty}}{j!}\left\{\frac{1}{n^{\beta j}}+(b-a)^{j}\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)\right\}+\right.$$

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

$$\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)(b-a)^{\alpha}\left(\sup_{x\in[a,b]}\left\|D_{x-}^{\alpha}f\right\|_{\infty,[a,x]}+\sup_{x\in[a,b]}\left\|D_{*x}^{\alpha}f\right\|_{\infty,[x,b]}\right)\Bigg\}\Bigg\}. \tag{47}$$

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. ■

Next we apply Theorem 29 for $N=1$.

Theorem 30 *Let* $0<\alpha,\beta<1$, $f\in C^{1}([a,b],X)$, $x\in[a,b]$, $n\in\mathbb{N}:n^{1-\beta}>2$, $\lambda>0$. *Then*

i)

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

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

$$\left.\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)\left(\left\|D_{x-}^{\alpha}f\right\|_{\infty,[a,x]}(x-a)^{\alpha}+\left\|D_{*x}^{\alpha}f\right\|_{\infty,[x,b]}(b-x)^{\alpha}\right)\right\}, \tag{48}$$

and

ii)

$$\|A_nf-f\|_{\infty}\leq\frac{4}{\Gamma(\alpha+1)\,erf(2\lambda)}$$

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

$$\left.\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)(b-a)^{\alpha}\left(\sup_{x\in[a,b]}\left\|D_{x-}^{\alpha}f\right\|_{\infty,[a,x]}+\sup_{x\in[a,b]}\left\|D_{*x}^{\alpha}f\right\|_{\infty,[x,b]}\right)\right\}. \tag{49}$$

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

Corollary 31 *Let* $0<\beta<1$, $f\in C^{1}([a,b],X)$, $x\in[a,b]$, $n\in\mathbb{N}:n^{1-\beta}>2$, $\lambda>0$. *Then*

i)

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

$$\frac{8}{erf\left(2\lambda\right)\sqrt{\pi}}\left\{\frac{\left(\omega_{1}\left(D_{x-}^{\frac{1}{2}}f,\frac{1}{n^{\beta}}\right)_{[a,x]}+\omega_{1}\left(D_{*x}^{\frac{1}{2}}f,\frac{1}{n^{\beta}}\right)_{[x,b]}\right)}{n^{\frac{\beta}{2}}}+\right.$$

$$\left.\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)\left(\left\|D_{x-}^{\frac{1}{2}}f\right\|_{\infty,[a,x]}\sqrt{(x-a)}+\left\|D_{*x}^{\frac{1}{2}}f\right\|_{\infty,[x,b]}\sqrt{(b-x)}\right)\right\}, \tag{50}$$

and

ii)

$$\left\|A_{n}f-f\right\|_{\infty}\leq\frac{8}{erf\left(2\lambda\right)\sqrt{\pi}}$$

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

$$\left.\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)\sqrt{(b-a)}\left(\sup_{x\in[a,b]}\left\|D_{x-}^{\frac{1}{2}}f\right\|_{\infty,[a,x]}+\sup_{x\in[a,b]}\left\|D_{*x}^{\frac{1}{2}}f\right\|_{\infty,[x,b]}\right)\right\}<\infty. \tag{51}$$

We finish with

Remark 32 *Some convergence analysis follows:*

Let $0<\beta<1$, $f\in C^{1}\left([a,b],X\right)$, $x\in[a,b]$, $n\in\mathbb{N}:n^{1-\beta}>2$, $\lambda>0$. *We elaborate on (51). Assume that*

$$\omega_{1}\left(D_{x-}^{\frac{1}{2}}f,\frac{1}{n^{\beta}}\right)_{[a,x]}\leq\frac{K_{1}}{n^{\beta}}, \tag{52}$$

and

$$\omega_{1}\left(D_{*x}^{\frac{1}{2}}f,\frac{1}{n^{\beta}}\right)_{[x,b]}\leq\frac{K_{2}}{n^{\beta}}, \tag{53}$$

$\forall\ x\in[a,b]$, $\forall\ n\in\mathbb{N}$, *where* $K_{1},K_{2}>0$.

Then it holds

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

$$\frac{\frac{(K_{1}+K_{2})}{n^{\beta}}}{n^{\frac{\beta}{2}}}=\frac{(K_{1}+K_{2})}{n^{\frac{3\beta}{2}}}=\frac{K}{n^{\frac{3\beta}{2}}}, \tag{54}$$

where $K:=K_{1}+K_{2}>0$.

The other summand of the right hand side of (51), for large enough n, converges to zero at the speed $\left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)$.

Then, for large enough $n \in \mathbb{N}$, by (51) and (54) and the last comment, we obtain that

$$\|A_n f - f\|_\infty \leq M \max\left(\frac{1}{n^{\frac{3\beta}{2}}}, \left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)\right), \tag{55}$$

where $M > 0$.

If $\frac{1}{n^{\frac{3\beta}{2}}} \geq \left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)$, *then* $\frac{1}{n^{\beta}} \geq \left(\frac{1-erf\lambda\left(n^{1-\beta}-2\right)}{2}\right)$, *and consequently* $\|A_n f - f\|_\infty$ *in (55) converges to zero faster than in Theorem 14. 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*, J. Computational Analysis and Applications, 31 (4) (2023), 520-534.

[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.