Analele Universităţii Oradea
Fasc. Matematica, Tom XXXIII (2026), Issue No. 1, 129–142

MULTIPLE-COMPOSITE MULTIVARIATE QUANTITATIVE APPROXIMATION BY KANTOROVICH-SHILKRET NEURAL NETWORKS

GEORGE A. ANASTASSIOU

Abstract. In this work we research the multivariate quantitative approximation by multi-composite Kantorovich-Shilkret type quasi-interpolation neural network operators with respect to supremum and L_p norms. This is achieved with rates via the first multivariate modulus of continuity. We approximate continuous and bounded non-negative functions on $\mathbb{R}^N$, $N \in \mathbb{N}$. When they are also uniformly continuous we have pointwise, uniform and L_p convergences. Complex approximation is also discussed. Our multi-composite activation functions are formed by general sigmoid functions.

1. Introduction

Overactivation in artificial neural network (ANNs) offers benefits such as improved learning capacity, increased non-linearity, and robustness to noise. In the **Biological brain**, moderate overactivation can be a beneficial compensatory mechanism to maintain performance, while excessive, chronic overactivation is generally detrimental and linked to various disorders.

In Artificial Neural Networks

In machine learning, "overactivation" typically refers to techniques that promote higher activation valued or more complex activation functions, offering potential advantages. These include:

- **Improved Learning Capacity**: Assisting models in learning more complex data patterns.
- **Increased Non-linearity**: Helping networks capture intricate data relationships for better generalization.
- **Robustness to Noise**: Making networks more resilient to noisy input data.
- **Implicit Regularization**: Potentially preventing overfitting by encouraging diverse feature representations.

The next concept relates to the above:

Multi-composition of activation functions involves stacking or combining multiple, often different, non-linear functions (e.g., $\sigma_2(\sigma_1(x))$) to enhance a neural network's capacity to learn complex relationships, improve approximation accuracy,

2020 *Mathematics Subject Classification.* 41A17, 41A25, 41A30, 41A35.

Key words and phrases. multi-composite general sigmoid activation functions, multivariate quasi-interpolation neural network approximation, multi-composite Kantorovich-Shilkret type operators.

and potentially reduce computational effort. This technique allows for more flexible modeling, such as creating compact support functions from wider ones to focus on relevant data.

Key Aspects of Multi-Composition Activation Functions:

- **Enhanced Representation**: Composing functions, such as $\sigma(x) =$ReLU $(\tanh(x))$, allows the network to combine the strengths of different nonlinearities.
- **Reduced Domain/Compact Support**: Compositions can be designed to limit the output to a specific range (e.g., using "cusp" functions), effectively filtering noise and focusing on crucial information.
- **Architectural Flexibility**: Hidden layer functions L_i are often composed as a linear mapping followed by an activation, $L_i = \sigma_i \circ g_i$, allowing for learnable, complex transformations within a single layer.
- **Improved Approximation:** Theoretical work suggests that composing functions can enhance the approximation abilities of neural networks, leading to better mathematical modeling of data.
- "**Faster Training":** Combining functions (fusion) can improve training speed in deep networks.

Common Examples and Usage:

- **Swish-ReLU**: Combining functions like SiLU (Swish) with ReLU can enhance gradient flow.
- **Multilayer Perceptrons (MLP):** Inherently, an MLP with multiple hidden layers is a functional composition ($N_d = \overline{L} \circ L_d \circ ... \circ L_1$) of these activated layers.
- **Hybrid Functions:** Using different activations in different layers, or even within the same layer, helps optimize for non-saturating gradients and efficiency.

The author in [1] and [2], 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, 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 [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 Chaptes 4-5 there.

The author inspired by [8], 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] - [6], by treating both the univariate and multivariate cases. He did also the corresponding fractional case [4, 6].

The author here performs multi-composite and multivariate general sigmoid activation functions based neural network approximation to continuous non-negative functions over the whole $\mathbb{R}^N$, $N \in \mathbb{N}$, then he extends his results to complex valued functions, L_p approximations. All convergences here are with rates expressed via the modulus of continuity of the involved function and given by very tight Jackson type inequalities.

The author comes up with the "right" precisely defined flexible quasi-interpolation, Kantorovich-Shilkret type integral coefficient neural networks operators associated with: multi-composite general sigmoid activation functions. In preparation to prove his results he mentions important properties of the multi-composite general density functions defining the operators.

In recent years non-additive integrals, like the N. Shilkret one [9], have become trendy with great applications to Economics, etc.

Feed-forward neural networks (FNNs) with one hidden layer, the only type of networks he deals with in this work, are mathematically expressed as

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

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. About neural networks in general read [10] - [14]. For recent related works see [15]-[24].

2. Basics

2.1. **Description of Shilkret integral.** Here we follow [9].

Let $\mathcal{F}$ be a σ-field of subsets of an arbitrary set Ω. An extended non-negative real valued function μ on $\mathcal{F}$ is called maxitive if $\mu(\varnothing) = 0$ and

$$\mu\left(\cup_{i\in I} E_i\right) = \sup_{i\in I} \mu\left(E_i\right), \tag{1}$$

where the set I is of cardinality at most countable. We also call μ a maxitive measure. Here f stands for a non-negative measurable function on Ω. In [9], Niel Shilkret developed his non-additive integral defined as follows:

$$(N^*) \int_D f d\mu := \sup_{y\in Y} \left\{y \cdot \mu\left(D \cap \{f \geq y\}\right)\right\}, \tag{2}$$

where $Y = [0, m]$ or $Y = [0, m)$ with $0 < m \leq \infty$, and $D \in \mathcal{F}$. Here we take $Y = [0, \infty)$.

It is easily proved that

$$(N^*) \int_D f d\mu = \sup_{y>0} \left\{y \cdot \mu\left(D \cap \{f > y\}\right)\right\}. \tag{3}$$

The Shilkret integral takes values in $[0, \infty]$.

The Shilkret integral ([9]) has the following properties:

$$(N^*) \int_\Omega \chi_E d\mu = \mu(E), \tag{4}$$

where χ_E is the indicator function on $E \in \mathcal{F}$,

$$(N*) \int_D c f d\mu = c\,(N^*) \int_D f d\mu, \quad c \geq 0, \tag{5}$$

$$(N^*) \int_D \sup_{n \in \mathbb{N}} f_n d\mu = \sup_{n \in \mathbb{N}} (N^*) \int_D f_n d\mu, \tag{6}$$

where f_n, $n \in \mathbb{N}$, is an increasing sequence of elementary (countably valued) functions converging uniformly to f. Furthermore we have

$$(N^*) \int_D f d\mu \geq 0, \tag{7}$$

$$f \geq g \text{ implies } (N^*) \int_D f d\mu \geq (N^*) \int_D g d\mu, \tag{8}$$

where $f, g : \Omega \to [0, \infty]$ are measurable.

Let $a \leq f(\omega) \leq b$ for almost every $\omega \in E$, then

$$a\mu(E) \leq (N^*) \int_E f d\mu \leq b\mu(E);$$

$$(N^*) \int_E 1 d\mu = \mu(E);$$

$f > 0$ almost everywhere and $(N^*) \int_E f d\mu = 0$ imply $\mu(E) = 0$;

$(N^*) \int_\Omega f d\mu = 0$ if and only $f = 0$ almost everywhere;

$(N^*) \int_\Omega f d\mu < \infty$ implies that

$$\overline{N}(f) := \{\omega \in \Omega | f(\omega) \neq 0\} \text{ has } \sigma\text{-finite measure;} \tag{9}$$

$$(N^*) \int_D (f+g) d\mu \leq (N^*) \int_D f d\mu + (N^*) \int_D g d\mu;$$

and

$$\left| (N^*) \int_D f d\mu - (N^*) \int_D g d\mu \right| \leq (N^*) \int_D |f - g| d\mu. \tag{10}$$

From now on in this work we assume that $\mu : \mathcal{F} \to [0, +\infty)$.

2.2. **On Multi-composite Activation Functions.** We mention:

Definition 2.1. Let $i = 1, 2, ...,$ and $h_i : \mathbb{R} \to [-1, 1]$ be general sigmoid activation functions, such that they are strictly increasing, $h_i(0) = 0$, $h_i(-x) = -h_i(x)$, $x \in \mathbb{R}$, $h_i(+\infty) = 1$, $h_i(-\infty) = -1$. Also h_i is strictly convex over $(-\infty, 0]$ and strictly concave over $[0, +\infty)$, with $h_i^{(2)} \in C(\mathbb{R})$. Examples $\tanh x$, $erf(x)$, normalized $\arctan x$, normalize Gudermanian(x), etc.

Notice here $0 < h_i(1) \leq 1$, $i = 1, 2, ...$ Any composition $h_1 \circ h_2 \circ h_3 \circ ... \circ h_\lambda$ is meant to be $h_1|_{[-1,1]} \circ h_2|_{[-1,1]} \circ h_3|_{[-1,1]} \circ ... \circ h_{\lambda-1}|_{[-1,1]} \circ h_\lambda$, $\lambda \in \mathbb{N}$, and for convenience, we denote it by $G_\lambda := h_1 \circ h_2 \circ h_3 \circ ... \circ h_\lambda$. We have for any $\lambda \in \mathbb{N} : 0 < h_\lambda(1) \leq 1$, hence $0 < h_{\lambda-1}(h_\lambda(1)) \leq h_{\lambda-1}(1) \leq 1$, and $0 < h_{\lambda-2}(h_{\lambda-1}(h_\lambda(1))) \leq h_{\lambda-2}(h_{\lambda-1}(1)) \leq h_{\lambda-2}(1) \leq 1$.

Inductively we derive that $0 < G_\lambda(1) \leq 1$, $\forall\ \lambda \in \mathbb{N}$.

Clearly, it is $G_\lambda(0) = 0$ and G_λ is strictly increasing over $\mathbb{R}$. Furthermore it holds

$$G_\lambda(-x) = h_1(h_2(h_3(...(h_{\lambda-1}(h_\lambda(-x)))))) = h_1(h_2(h_3(...(h_{\lambda-1}(-h_\lambda(x))))))$$
$$= ... = -h_1(h_2(h_3(...(h_{\lambda-1}(h_\lambda(x)))))) = -G_\lambda(x), \quad x \in \mathbb{R}.$$

Clearly it holds $G_\lambda^{(2)} \in C(\mathbb{R})$.

We notice that

$$G_{\lambda}(+\infty) = h_1(h_2(h_3(...(h_{\lambda-1}(h_{\lambda}(+\infty)))))) =$$

$$h_1(h_2(h_3(...(h_{\lambda-1}(1))))) = G_{\lambda-1}(1), \tag{11}$$

and

$$G_{\lambda}(-\infty) = h_1(h_2(h_3(...(h_{\lambda-1}(h_{\lambda}(-\infty)))))) = h_1(h_2(h_3(...(h_{\lambda-1}(-1)))))$$

$$= -h_1(h_2(h_3(...(h_{\lambda-1}(1))))) = -G_{\lambda-1}(1). \tag{12}$$

Consequently, it holds

$$-G_{\lambda-1}(1) \leq G_{\lambda}(x) \leq G_{\lambda-1}(1), \quad \forall\, x \in \mathbb{R}. \tag{13}$$

Thus, $y = \pm G_{\lambda-1}(1)$ are asymptotes of $G_{\lambda}(x)$, any $\lambda \in \mathbb{N}$.

Next we act over $(-\infty, 0]$: let $\lambda, \mu \geq 0 : \lambda + \mu = 1$. Then by convexity of h_2 there we have

$$h_2(\lambda x + \mu y) \leq \lambda h_2(x) + \mu h_2(y), \quad \forall\, x, y \in (-\infty, 0];$$

and

$$h_1(h_2(\lambda x + \mu y)) \leq h_1(\lambda h_2(x) + \mu h_2(y)) \leq$$

$$\lambda(h_1 \circ h_2)(x) + \mu(h_1 \circ h_2)(y), \quad \forall x,\ y \in (-\infty, 0]. \tag{14}$$

So that $h_1 \circ h_2$ is convex over $(-\infty, 0]$.

Now we work on $[0, +\infty)$: let $\lambda, \mu \geq 0 : \lambda + \mu = 1$. Then by concavity of h_2 there we have

$$h_2(\lambda x + \mu y) \geq \lambda h_2(x) + \mu h_2(y), \quad \forall\, x, y \in [0, +\infty);$$

and

$$h_1(h_2(\lambda x + \mu y)) \geq h_1(\lambda h_2(x) + \mu h_2(y)) \geq$$

$$\lambda(h_1 \circ h_2)(x) + \mu(h_1 \circ h_2)(y), \quad \forall x,\ y \in [0, +\infty). \tag{15}$$

Thus, $h_1 \circ h_2$ is concave over $[0, +\infty)$.

Therefore $G_2 = h_1 \circ h_2$ is a general sigmoid activation function with asymptotes $y = \pm h_1(1)$, and fulfilling the rest of conditions of Definition 2.1.

Arguing as above $h_2 \circ h_3 : \mathbb{R} \to [-1, 1]$, fulfills Definition 2.1 and $h_1 \circ h_2 \circ h_3$ does the same with asymptotes $h = \pm h_1(h_2(1))$.

Inductively, we prove that G_{λ} fulfills Definition 2.1 with asymptotes $y = \pm G_{\lambda-1}(1)$.

We have proved the following:

Theorem 2.1. *Let $\lambda \in \mathbb{N}$. Then $G_{\lambda} := h_1 \circ h_2 \circ h_3 \circ ... \circ h_{\lambda}$ fulfills all the properties of Definition 2.1 with asymptotes $y = \pm G_{\lambda-1}(1)$. That is G_{λ} is a multi-composite general sigmoid activation function from $\mathbb{R} \to [-1, 1]$.*

Corollary 2.2. $\frac{G_{\lambda}}{G_{\lambda-1}(1)}$ *fulfills Definition 2.1 with asymottes $y = \pm 1$.*

We call

$$\widetilde{G}_{\lambda} := \frac{G_{\lambda}}{G_{\lambda-1}(1)}. \tag{16}$$

Remark 2.1. Next we consider the function

$$T_{\lambda}(x) := \frac{1}{4}\left(\widetilde{G}_{\lambda}(x+1) - \widetilde{G}_{\lambda}(x-1)\right) > 0, \ \forall\, x \in \mathbb{R}, \ \lambda \in \mathbb{N}. \tag{17}$$

We observe that

$$T_{\lambda}(-x) = \frac{1}{4}\left(\widetilde{G}_{\lambda}(-x+1) - \widetilde{G}_{\lambda}(-x-1)\right) =$$
$$\frac{1}{4}\left(\widetilde{G}_{\lambda}(-(x-1)) - \widetilde{G}_{\lambda}(-(x+1))\right) = \frac{1}{4}\left(-\widetilde{G}_{\lambda}(x-1) + \widetilde{G}_{\lambda}(x+1)\right) = \tag{18}$$
$$\frac{1}{4}\left(\widetilde{G}_{\lambda}(x+1) - \widetilde{G}_{\lambda}(x-1)\right) = T_{\lambda}(x).$$

That is T_{λ} is an even function,

$$T_{\lambda}(-x) = T_{\lambda}(x), \quad \forall\, x \in \mathbb{R}, \ \lambda \in \mathbb{N}. \tag{19}$$

We see that

$$T_{\lambda}(0) = \frac{\widetilde{G}_{\lambda}(1)}{2}. \tag{20}$$

Let $x > 1$, we have that

$$T_{\lambda}'(x) = \frac{1}{4}\left(\widetilde{G}_{\lambda}{}'(x+1) - \widetilde{G}_{\lambda}{}'(x-1)\right) < 0,$$

by $\widetilde{G}_{\lambda}'$ being strictly decreasing over $[0, +\infty)$.

Let now $0 < x < 1$, then $1 - x > 0$ and $0 < 1 - x < 1 + x$. It holds $\widetilde{G}_{\lambda}'(x-1) = \widetilde{G}_{\lambda}'(1-x) > \widetilde{G}_{\lambda}'(x+1)$, so that again $T_{\lambda}'(x) < 0$. Consequently T_{λ} is strictly decreasing on $(0, +\infty)$.

Clearly, T_{λ} is strictly increasing on $(-\infty, 0)$, and $T_{\lambda}'(0) = 0$.

Observe that

$$\lim_{x \to +\infty} T_{\lambda}(x) = \frac{1}{4}\left(\widetilde{G}_{\lambda}(+\infty) - \widetilde{G}_{\lambda}(+\infty)\right) = 0, \tag{21}$$

and

$$\lim_{x \to -\infty} T_{\lambda}(x) = \frac{1}{4}\left(\widetilde{G}_{\lambda}(-\infty) - \widetilde{G}_{\lambda}(-\infty)\right) = 0. \tag{22}$$

That is the x-axis is the horizontal asymptote on T_{λ}.

As a result T_{λ} is a bell shaped symmetric function with maximum

$$T_{\lambda}(0) = \frac{\widetilde{G}_{\lambda}(1)}{2}. \tag{23}$$

We need

Theorem 2.3. *It holds*

$$\sum_{i=-\infty}^{\infty} T_{\lambda}(x - i) = 1, \ \forall\, x \in \mathbb{R}. \tag{24}$$

Proof. As similar to [5], p. 286 is omitted. □

Theorem 2.4. *We have that*

$$\int_{-\infty}^{\infty} T_{\lambda}(x)\, dx = 1. \tag{25}$$

Proof. As similar to [5], p. 287, it is omitted. □

So that $T_\lambda(x)$ can serve as a density function in general.
We need

Theorem 2.5. *([7]) Let $0 < \alpha < 1$, and $n \in \mathbb{N}$ with $n^{1-\alpha} > 2$. Then*

$$\sum_{\begin{cases} k=-\infty \\ : |nx-k| \geq n^{1-\alpha} \end{cases}}^{\infty} T_\lambda(nx-k) < \left(1 - \widetilde{G}_\lambda\left(n^{1-\alpha} - 2\right)\right), \tag{26}$$

and

$$\lim_{n\to+\infty} \left(1 - \widetilde{G}_\lambda\left(n^{1-\alpha} - 2\right)\right) = 0. \tag{27}$$

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

Theorem 2.6. *Let $x \in [a,b] \subset \mathbb{R}$ and $n \in \mathbb{N}$ so that $\lceil na \rceil \leq \lfloor nb \rfloor$. It holds*

$$\frac{1}{\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} T_\lambda(nx-k)} < \frac{1}{T_\lambda(1)}, \quad \forall\, x \in [a,b]. \tag{28}$$

Proof. As similar to [5], p. 289 is omitted. □

Remark 2.2. We have that

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

for at least some $x \in [a,b]$.
See [5], p. 290, same reasoning.

Note 2.7. 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 (24))

$$\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} T_\lambda(nx-k) \leq 1. \tag{30}$$

We make

Remark 2.3. We define

$$\widehat{Z}(x_1,...,x_N) := \widehat{Z}(x) := \prod_{i=1}^{N} T_\lambda(x_i), \;\; x = (x_1,...,x_N) \in \mathbb{R}^N, \;\; N \in \mathbb{N}. \tag{31}$$

It has the properties:

(i)

$$\widehat{Z}(x) > 0, \;\; \forall x \in \mathbb{R}^N, \tag{32}$$

(ii)

$$\sum_{k=-\infty}^{\infty} \widehat{Z}(x-k) := \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} ... \sum_{k_N=-\infty}^{\infty} \widehat{Z}(x_1-k_1,...,x_N-k_N) =$$

$$\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}...\sum_{k_N=-\infty}^{\infty}\prod_{i=1}^{N}T_{\lambda}(x_i-k_i)=\prod_{i=1}^{N}\left(\sum_{k_i=-\infty}^{\infty}T_{\lambda}(x_i-k_i)\right)\overset{(24)}{=}1.$$

Hence

$$\sum_{k=-\infty}^{\infty}\widehat{Z}(x-k)=1. \tag{33}$$

That is

(iii)

$$\sum_{k=-\infty}^{\infty}\widehat{Z}(nx-k)=1,\quad \forall x\in\mathbb{R}^N;\ n\in\mathbb{N}. \tag{34}$$

And

(iv)

$$\int_{\mathbb{R}^N}\widehat{Z}(x)\,dx=\int_{\mathbb{R}^N}\left(\prod_{i=1}^{N}T_{\lambda}(x_i)\right)dx_1...dx_N=\prod_{i=1}^{N}\left(\int_{-\infty}^{\infty}T_{\lambda}(x_i)\,dx_i\right)\overset{(25)}{=}1, \tag{35}$$

thus

$$\int_{\mathbb{R}^N}\widehat{Z}(x)\,dx=1, \tag{36}$$

that is $\widehat{Z}$ is a multivariate density function.

Here denote $x=(x_1,...,x_N)$, $\|x\|_{\infty}:=\max\{|x_1|,...,|x_N|\}$, $x\in\mathbb{R}^N$, also set $\infty:=(\infty,...,\infty)$, $-\infty:=(-\infty,...,-\infty)$ upon the multivariate context, $0<\beta<1$,

(v) We have

$$\sum_{\substack{k=-\infty \\ \left\|\frac{k}{n}-x\right\|_{\infty}>\frac{1}{n^{\beta}}}}^{\infty}\widehat{Z}(nx-k)=\sum_{\substack{k_1=-\infty \\ \left\|\frac{k}{n}-x\right\|_{\infty}>\frac{1}{n^{\beta}}}}^{\infty}...\sum_{k_N=-\infty}^{\infty}\left(\prod_{i=1}^{N}T_{\lambda}(nx_i-k_i)\right)=$$

$$\prod_{i=1}^{N}\left(\sum_{\substack{k_i=-\infty \\ \left\|\frac{k}{n}-x\right\|_{\infty}>\frac{1}{n^{\beta}}}}^{\infty}T_{\lambda}(nx_i-k_i)\right)\leq\quad(\text{for some } r\in\{1,...,N\})$$

$$\left(\prod_{\substack{i=1 \\ i\neq r}}^{N}\left(\sum_{k_i=-\infty}^{\infty}T_{\lambda}(nx_i-k_i)\right)\right)\left(\sum_{\substack{k_r=-\infty \\ \left|\frac{k_r}{n}-x_r\right|>\frac{1}{n^{\beta}}}}^{\infty}T_{\lambda}(nx_r-k_r)\right)= \tag{37}$$

$$\sum_{\substack{k_r=-\infty \\ \left|\frac{k_r}{n}-x_r\right|>\frac{1}{n^{\beta}}}}^{\infty}T_{\lambda}(nx_r-k_r)=\sum_{\substack{k_r=-\infty \\ |nx_r-k_r|>n^{1-\beta}}}^{\infty}T_{\lambda}(nx_r-k_r)\overset{(26)}{<}$$

$$1-\widetilde{G}_{\lambda}\left(n^{1-\beta}-2\right).$$

That is

$$\sum_{\substack{k=-\infty \\ \left\|\frac{k}{n}-x\right\|_{\infty}>\frac{1}{n^{\beta}}}}^{\infty} \widehat{Z}(nx-k)<1-\widetilde{G}_{\lambda}\left(n^{1-\beta}-2\right). \tag{38}$$

$0<\beta<1$, $n\in\mathbb{N}: n^{1-\beta}>2$, $\forall\, x\in\prod_{i=1}^{N}[a_i,b_i]$.

Denote by

$$\varepsilon_N(\beta,n):=1-\widetilde{G}_{\lambda}\left(n^{1-\beta}-2\right), \quad \text{any } N\in\mathbb{N}, \tag{39}$$

$0<\beta<1$.

For $f\in C_B^{+}\left(\mathbb{R}^N\right)$ (continuous and bounded functions from $\mathbb{R}^N$ into $\mathbb{R}_+$), we define the first modulus of continuity

$$\omega_1(f,\delta):=\sup_{\substack{x,y\in\mathbb{R}^N \\ \|x-y\|_{\infty}\leq h}}|f(x)-f(y)|, \quad h>0. \tag{40}$$

Given that $f\in C_U^{+}\left(\mathbb{R}^N\right)$ (uniformly continuous from $\mathbb{R}^N$ into $\mathbb{R}_+$, same definition for ω_1), we have that

$$\lim_{h\to 0}\omega_1(f,h)=0. \tag{41}$$

When $N=1$, ω_1 is defined as in (40) with $\|\cdot\|_{\infty}$ collapsing to $|\cdot|$ and has the property (41).

3. Main Results

We need

Definition 3.1. Let $\mathcal{L}$ be the Lebesgue σ-algebra on $\mathbb{R}^N$, $N\in\mathbb{N}$, and the maxitive measure $\mu:\mathcal{L}\to[0,+\infty)$, such that for any $A\in\mathcal{L}$ with $A\neq\varnothing$, we get $\mu(A)>0$.

For $f\in C_B^{+}\left(\mathbb{R}^N\right)$, we define the multi-activated multivariate Kantorovich-Shilkret type neural network operators for any $x\in\mathbb{R}^N$:

$$\Delta_n^{\mu}(f,x)=\Delta_n^{\mu}(f,x_1,...,x_N):=$$

$$\sum_{k=-\infty}^{\infty}\left(\frac{(N^*)\int_{\left[0,\frac{1}{N}\right]^N} f\left(t+\frac{k}{n}\right)d\mu(t)}{\mu\left(\left[0,\frac{1}{n}\right]^N\right)}\right)\widehat{Z}(nx-k)=$$

$$\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}\cdots\sum_{k_N=-\infty}^{\infty}\left(\frac{(N^*)\int_0^{\frac{1}{n}}\cdots\int_0^{\frac{1}{n}} f\left(t_1+\frac{k_1}{n},t_2+\frac{k_2}{n},...,t_N+\frac{k_N}{n}\right)d\mu(t_1,...,t_N)}{\mu\left(\left[0,\frac{1}{n}\right]^N\right)}\right) \tag{42}$$

$$\cdot\left(\prod_{i=1}^{N}T_{\lambda}(nx_i-k_i)\right),$$

where $x=(x_1,...,x_N)\in\mathbb{R}^N$, $k=(k_1,...,k_N)$, $t=(t_1,...,t_N)$, $n\in\mathbb{N}$.

Clearly here $\mu\left(\left[0,\frac{1}{n}\right]^N\right)>0$, $\forall\ n\in\mathbb{N}$.

Above we notice that

$$\left\|\Delta_n^{\mu}(f)\right\|_{\infty}\leq\|f\|_{\infty}, \tag{43}$$

so that $\Delta_n^{\mu}(f,x)$ is well-defined.

We make

Remark 3.1. Let $t\in\left[0,\frac{1}{n}\right]^N$ and $x\in\mathbb{R}^N$, then

$$f\left(t+\frac{k}{n}\right)=f\left(t+\frac{k}{n}\right)-f(x)+f(x)\leq\left|f\left(t+\frac{k}{n}\right)-f(x)\right|+f(x), \tag{44}$$

hence

$$(N^*)\int_{\left[0,\frac{1}{n}\right]^N}f\left(t+\frac{k}{n}\right)d\mu(t)\leq$$

$$(N^*)\int_{\left[0,\frac{1}{n}\right]^N}\left|f\left(t+\frac{k}{n}\right)-f(x)\right|d\mu(t)+f(x)\,\mu\left(\left[0,\frac{1}{n}\right]^N\right). \tag{45}$$

That is

$$(N^*)\int_{\left[0,\frac{1}{n}\right]^N}f\left(t+\frac{k}{n}\right)d\mu(t)-f(x)\,\mu\left(\left[0,\frac{1}{n}\right]^N\right)\leq \tag{46}$$

$$(N^*)\int_{\left[0,\frac{1}{n}\right]^N}\left|f\left(t+\frac{k}{n}\right)-f(x)\right|d\mu(t).$$

Similarly, we have

$$f(x)=f(x)-f\left(t+\frac{k}{n}\right)+f\left(t+\frac{k}{n}\right)\leq\left|f\left(t+\frac{k}{n}\right)-f(x)\right|+f\left(t+\frac{k}{n}\right),$$

hence

$$(N^*)\int_{\left[0,\frac{1}{n}\right]^N}f(x)\,d\mu(t)\leq(N^*)\int_{\left[0,\frac{1}{n}\right]^N}\left|f\left(t+\frac{k}{n}\right)-f(x)\right|d\mu(t)$$

$$+(N^*)\int_{\left[0,\frac{1}{n}\right]^N}f\left(t+\frac{k}{n}\right)d\mu(t).$$

That is

$$f(x)\,\mu\left(\left[0,\frac{1}{n}\right]^N\right)-(N^*)\int_{\left[0,\frac{1}{n}\right]^N}f\left(t+\frac{k}{n}\right)d\mu(t)\leq \tag{47}$$

$$(N^*)\int_{\left[0,\frac{1}{n}\right]^N}\left|f\left(t+\frac{k}{n}\right)-f(x)\right|d\mu(t).$$

By (46) and (47) we derive

$$\left|(N^*)\int_{\left[0,\frac{1}{n}\right]^N}f\left(t+\frac{k}{n}\right)d\mu(t)-f(x)\,\mu\left(\left[0,\frac{1}{n}\right]^N\right)\right|\leq$$

$$(N^*)\int_{\left[0,\frac{1}{n}\right]^N}\left|f\left(t+\frac{k}{n}\right)-f(x)\right|d\mu(t). \tag{48}$$

In particular it holds

$$\left| \frac{(N^*) \int_{\left[0,\frac{1}{n}\right]^N} f\left(t+\frac{k}{n}\right) d\mu(t)}{\mu\left(\left[0,\frac{1}{n}\right]^N\right)} - f(x) \right| \leq$$

$$\frac{(N^*) \int_{\left[0,\frac{1}{n}\right]^N} \left| f\left(t+\frac{k}{n}\right) - f(x) \right| d\mu(t)}{\mu\left(\left[0,\frac{1}{n}\right]^N\right)}. \tag{49}$$

We give the following approximation result.

Theorem 3.1. *Let $f \in C_B^+\left(\mathbb{R}^N\right)$, $0 < \beta < 1$, $x \in \mathbb{R}^N$; $N, n \in \mathbb{N}$ with $n^{1-\beta} > 2$. Then*

i)

$$\sup_{\mu} \left| \Delta_n^{\mu}(f, x) - f(x) \right| \leq \omega_1\left(f, \frac{1}{n} + \frac{1}{n^{\beta}}\right) + 2\|f\|_{\infty} \varepsilon_N(\beta, n) =: \rho_n, \tag{50}$$

and

ii)

$$\sup_{\mu} \left\| \Delta_n^{\mu}(f) - f \right\|_{\infty} \leq \rho_n. \tag{51}$$

Given that $f \in \left(C_U^+\left(\mathbb{R}^N\right) \cap C_B^+\left(\mathbb{R}^N\right)\right)$, we obtain $\lim_{n\to\infty} \Delta_n^{\mu}(f) = f$, uniformly. Above $\varepsilon_N(\beta, n)$ is as in (39).

Proof. We observe that

$$\left| \Delta_n^{\mu}(f, x) - f(x) \right| =$$

$$\left| \sum_{k=-\infty}^{\infty} \left(\frac{(N^*) \int_{\left[0,\frac{1}{n}\right]^N} f\left(t+\frac{k}{n}\right) d\mu(t)}{\mu\left(\left[0,\frac{1}{n}\right]^N\right)} \right) \widehat{Z}(nx-k) - \sum_{k=-\infty}^{\infty} f(x) \widehat{Z}(nx-k) \right| =$$

$$\left| \sum_{k=-\infty}^{\infty} \left(\left(\frac{(N^*) \int_{\left[0,\frac{1}{n}\right]^N} f\left(t+\frac{k}{n}\right) d\mu(t)}{\mu\left(\left[0,\frac{1}{n}\right]^N\right)} \right) - f(x) \right) \widehat{Z}(nx-k) \right| \leq$$

$$\sum_{k=-\infty}^{\infty} \left| \left(\frac{(N^*) \int_{\left[0,\frac{1}{n}\right]^N} f\left(t+\frac{k}{n}\right) d\mu(t)}{\mu\left(\left[0,\frac{1}{n}\right]^N\right)} \right) - f(x) \right| \widehat{Z}(nx-k) \overset{(49)}{\leq}$$

$$\sum_{k=-\infty}^{\infty} \left(\frac{(N^*) \int_{\left[0,\frac{1}{n}\right]^N} \left| f\left(t+\frac{k}{n}\right) - f(x) \right| d\mu(t)}{\mu\left(\left[0,\frac{1}{n}\right]^N\right)} \right) \widehat{Z}(nx-k) =$$

$$\sum_{\left\{\substack{k=-\infty \\ : \left\|\frac{k}{n}-x\right\|_{\infty} \leq \frac{1}{n^{\beta}}}\right.}^{\infty} \left(\frac{(N^*) \int_{\left[0,\frac{1}{n}\right]^N} \left| f\left(t+\frac{k}{n}\right) - f(x) \right| d\mu(t)}{\mu\left(\left[0,\frac{1}{n}\right]^N\right)} \right) \widehat{Z}(nx-k) + \tag{52}$$

$$\sum_{\substack{k=-\infty \\ : \left\|\frac{k}{n}-x\right\|_{\infty} > \frac{1}{n^{\beta}}}}^{\infty} \left(\frac{(N^{*}) \int_{\left[0,\frac{1}{n}\right]^{N}} \left| f\left(t+\frac{k}{n}\right) - f(x) \right| d\mu(t)}{\mu\left(\left[0,\frac{1}{n}\right]^{N}\right)} \right) \widehat{Z}(nx-k) \leq$$

$$\sum_{\substack{k=-\infty \\ : \left\|\frac{k}{n}-x\right\|_{\infty} \leq \frac{1}{n^{\beta}}}}^{\infty} \left(\frac{(N^{*}) \int_{\left[0,\frac{1}{n}\right]^{N}} \omega_{1}\left(f, \|t\|_{\infty} + \left\|\frac{k}{n}-x\right\|_{\infty}\right) d\mu(t)}{\mu\left(\left[0,\frac{1}{n}\right]^{N}\right)} \right) \widehat{Z}(nx-k)$$

$$+2\|f\|_{\infty} \left(\sum_{\substack{k=-\infty \\ : \left\|\frac{k}{n}-x\right\|_{\infty} > \frac{1}{n^{\beta}}}}^{\infty} \widehat{Z}(nx-k) \right) \text{ (by (38))}$$

$$\leq \omega_{1}\left(f, \frac{1}{n} + \frac{1}{n^{\beta}}\right) + 2\|f\|_{\infty} \varepsilon_{N}(\beta, n), \tag{53}$$

proving the claim. □

Additionally we give

Definition 3.2. Denote by $C_{B}^{+}\left(\mathbb{R}^{N}, \mathbb{C}\right) = \{f : \mathbb{R}^{N} \to \mathbb{C} | f = f_{1} + if_{2}$, where $f_{1}, f_{2} \in C_{B}^{+}\left(\mathbb{R}^{N}\right)$, $N \in \mathbb{N}\}$. We set for $f \in C_{B}^{+}\left(\mathbb{R}^{N}, \mathbb{C}\right)$ that

$$\Delta_{n}^{\mu}(f, x) := \Delta_{n}^{\mu}(f_{1}, x) + i\, \Delta_{n}^{\mu}(f_{2}, x), \tag{54}$$

$\forall\, n \in \mathbb{N}$, $x \in \mathbb{R}^{N}$, $i = \sqrt{-1}$.

Theorem 3.2. *Let $f \in C_{B}^{+}\left(\mathbb{R}^{N}, \mathbb{C}\right)$, $f = f_{1} + if_{2}$, $N \in \mathbb{N}$, $0 < \beta < 1$, $x \in \mathbb{R}^{N}$; $n \in \mathbb{N}$ with $n^{1-\beta} > 2$. Then*

i)

$$\sup_{\mu} \left| \Delta_{n}^{\mu}(f, x) - f(x) \right| \leq \left[\omega_{1}\left(f_{1}, \frac{1}{n} + \frac{1}{n^{\beta}}\right) + \omega_{1}\left(f_{2}, \frac{1}{n} + \frac{1}{n^{\beta}}\right) \right]$$
$$+2\left(\|f_{1}\|_{\infty} + \|f_{2}\|_{\infty}\right) \varepsilon_{N}(\beta, n) =: J_{n}, \tag{55}$$

and

ii)

$$\sup_{\mu} \left\| \Delta_{n}^{\mu}(f) - f \right\| \leq J_{n}. \tag{56}$$

Proof.

$$\left| \Delta_{n}^{\mu}(f, x) - f(x) \right| = \left| \Delta_{n}^{\mu}(f_{1}, x) + i\, \Delta_{n}^{\mu}(f_{2}, x) - f_{1}(x) - if_{2}(x) \right| =$$
$$\left| \left(\Delta_{n}^{\mu}(f_{1}, x) - f_{1}(x) \right) + i\left(\Delta_{n}^{\mu}(f_{2}, x) - f_{2}(x) \right) \right| \leq$$
$$\left| \Delta_{n}^{\mu}(f_{1}, x) - f_{1}(x) \right| + \left| \Delta_{n}^{\mu}(f_{2}, x) - f_{2}(x) \right| \overset{(50)}{\leq} \tag{57}$$
$$\left(\omega_{1}\left(f_{1}, \frac{1}{n} + \frac{1}{n^{\beta}}\right) + 2\|f_{1}\|_{\infty} \varepsilon_{N}(\beta, n) \right) +$$

$$\left(\omega_1\left(f_2, \frac{1}{n}+\frac{1}{n^\beta}\right)+2\left\|f_2\right\|_\infty \varepsilon_N(\beta, n)\right)=$$
$$\left[\omega_1\left(f_1, \frac{1}{n}+\frac{1}{n^\beta}\right)+\omega_1\left(f_2, \frac{1}{n}+\frac{1}{n^\beta}\right)\right]+$$
$$2\left(\left\|f_1\right\|_\infty+\left\|f_2\right\|_\infty\right) \varepsilon_N(\beta, n). \tag{58}$$

proving the claim. □

We finish with an L_{p_1}, $p_1 \geq 1$, estimate.

Theorem 3.3. *Let $f \in C_B^+\left(\mathbb{R}^N, \mathbb{C}\right)$, $f=f_1+if_2$, $N \in \mathbb{N}$, $0<\beta<1$, $n \in \mathbb{N}$ with $n^{1-\beta}>2$, and $p_1 \geq 1$. Then*

$$\left\|\Delta_n^\mu(f)-f\right\|_{p_1, \overline{\Lambda}} \leq J_n\left|\overline{\Lambda}\right|^{\frac{1}{p_1}}, \tag{59}$$

where $\left|\overline{\Lambda}\right|<\infty$, is the Lebesgue measure of compact $\overline{\Lambda} \subset \mathbb{R}^N$, and J_n as in (55).

Proof. By integrating (55), etc. □

References

[1] G.A. Anastassiou, *Rate of convergence of some neural network operators to the unit-univariate case*, J. Math. Anal. Appli. 212 (1997), 237-262.
[2] G.A. Anastassiou, *Quantitative Approximations*, Chapman & Hall/CRC, Boca Raton, New York, 2001.
[3] G.A. Anastassiou, *Intelligent Systems: Approximation by Artificial Neural Networks*, Intelligent Systems Reference Library, Vol. 19, Springer, Heidelberg, 2011.
[4] G.A. Anastassiou, *Intelligent Systems II: Complete Approximation by Neural Network Operators*, Springer, Heidelberg, New York, 2016.
[5] G.A. Anastassiou, *Intelligent Computations: Abstract Fractional Calculus, Inequalities, Approximations*, Springer, Heidelberg, New York, 2018.
[6] G.A. Anastassiou, *Parametrized, Deformed and General Neural Networks*, Springer, Heidelberg, New York, 2023.
[7] G.A. Anastassiou, *General multi-composite sigmoid relied Banach space valued univariate neural network approximation*, submitted, 2025.
[8] Z. Chen and F. Cao, *The approximation operators with sigmoidal functions*, Computers and Mathematics with Applications, 58 (2009), 758-765.
[9] Niel Shilkret, *Maxitive measure and integration*, Indagationes Mathematicae, 33 (1971), 109-116.
[10] D. Costarelli, R. Spigler, *Approximation results for neural network operators activated by sigmoidal functions*, Neural Networks 44 (2013), 101-106.
[11] D. Costarelli, R. Spigler, *Multivariate neural network operators with sigmoidal activation functions*, Neural Networks 48 (2013), 72-77.
[12] I.S. Haykin, *Neural Networks: A Comprehensive Foundation* (2 ed.), Prentice Hall, New York, 1998.
[13] W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bulletin of Mathematical Biophysics, 7 (1943), 115-133.
[14] T.M. Mitchell, *Machine Learning*, WCB-McGraw-Hill, New York, 1997.
[15] Yu, Dansheng; Cao, Feilong, *Construction and approximation rate for feed-forward neural network operators with sigmoidal functions*, J. Comput. Appl. Math. 453 (2025), paper No. 116150.
[16] S. Cen, B. Jin, Q. Quan, Z. Zhou, *Hybrid neural-network FEM approximation of diffusion coefficient in elliptic and parabolic problems.* IMA J. Numer. Anal. 44 (2024), no. 5, 3059-3093.

[17] Coroianu Lucian; Costarelli, Danillo; Natale, Mariarosaria; Pantiş, Alexandra; *The approximation capabilities of Durrmeyer-type neural network operators*. J. Appl. Math. Comput. 70 (2024), no. 5, 4581-4599.

[18] Warin, Xavier; *The GroupMax neural network approximation of convex functions*. IEEE Trans. Neural Netw. Learn. Syst. 35 (2024), no. 8, 11608-11612.

[19] Fabra, Arnau; Guasch, Oriol; Baiges, Joan; Codina, Ramon; *Approximation of acoustic black holes with finite element mixed formulations and artificial neural network correction terms*. Finite Elem. Anal. Des. 241 (2024), paper No. 104236.

[20] Grohs, Philipp; Voigtlaender, Felix. *Proof of the theory-to-practice gap in deep learning via sampling complexity bounds for neural network approximation spaces*. Found. Comput. Math. 24 (2024), no. 4, 1085-1143.

[21] Basteri, Andrea; Trevisan, Dario; *Quantitative Gaussian approximation of randomly initialized deep neural networks*. Mach. Learn. 113 (2024), no. 9, 6373-6393.

[22] De Ryck, T.; Mishra, S. *Error analysis for deep neural network approximations of parametric hyperbolic conservation laws*. Math. Comp. 93 (2024), no. 350, 2643-2677.

[23] Liu, Jie; Zhang, Baoji; Lai, Yuyang; Fang, Liqiau. *Hull form optimization research based on multi-precision back-propagation neural network approximation model*. Internal. J. Numer. Methods Fluid 96 (2024), no. 8, 1445-1460.

[24] Yoo, Jihahm; Kim, Jaywon; Gim, Minjung; Lee, Haesung. *Error estimates of physics-informed neural networks for initial value problems*. J. Korean Soc. Ind. Appl. Math. 28 (2024), no. 1, 33-58.

Received 30 January 2026

Department of Mathematical Sciences, University of Memphis, Memphis, TN 38152, U.S.A.

Email address: ganastss@memphis.edu