Probl. Anal. Issues Anal. Vol. ..., No ..., 2026
DOI: 10.15393/j3.art.2026.19630

UDC 517.51, 517.98

G. A. Anastassiou

MULTIVARIATE CASE OF COMPOSITION OF ACTIVATION FUNCTIONS AND THE REDUCTION TO FINITE DOMAIN

Abstract. This work deals with the determination of the rate of pointwise, uniform and L_p convergences to the unit operator of the "multivariate normalized cusp neural network operators". The multivariate cusp is a compact support activation function, which derives by the composition of two general activation functions having as domain the whole real line. These convergences are given via the multivariate modulus of continuity of the engaged function or its partial derivatives in the form of Jackson type multivariate inequalities.

The composition of activation functions aims to more flexible and powerful neural networks, introducing for the first time the reduction of infinite domains to the one multivariate domain of compact support.

Key words: *neural network approximation, multivariate cusp activation function, multivariate modulus of continuity, reduction of domain*

2020 Mathematical Subject Classification: *41A17, 41A25, 41A30, 41A36.*

1. Introduction. From AI and computer science we have the following: In essence, composing activation functions in neural networks offers the advantage of potentially tailoring the network's ability to learn and model complex, non-linear relationships in data. Here's a breakdown of the potential benefits:

1. Enhanced Capacity for Complex Modeling:

- Diversification of Non-linearity: Different activation functions have different characteristics. For example, ReLU introduces sparsity, while Sigmoid squashes values into a range. By composing them, the network potentially can learn a wider variety of non-linear transformations and capture more intricate patterns in the data.

2. Improved Training Dynamics:

- Mitigating Gradient Problems: Activation functions influence gradient flow during training. Using different activation functions can potentially help address issues like vanishing or exploding gradients, which hinder learning in deep networks.

- Faster Convergence: Certain activation functions, like ReLU, can accelerate the convergence of the training process compared to others like Sigmoid or Tanh. Combining different functions can potentially lead to faster training and competitive performance.

3. Enhanced Generalization and Robustness:

- Better Generalization: By learning richer representations of the data through diverse activation functions, the network's ability to generalize well to unseen data improves, reducing the risk of overfitting.

- Increased Robustness: Networks with carefully chosen activation functions can handle variations in input data more effectively, adapting to noise , missing data, or unexpected perturbations.

4. Adaptation to Input Characteristics:

- Handling Diverse Data: Different activation functions can be suited to different data characteristics. For instance, tanh can be useful when dealing with data containing both positive and negative values.

5. Potential for Architectural Interpretability:

- Insight into Learning: By using distinct activation functions, different parts of the network might become responsible for capturing specific features, which can potentially offer insights into how the model learns.

In summary, composing activation functions potentially allows for a more flexible and powerful neural network capable of:

- Learning more complex patterns.

- Faster and more stable training.

- Better generalization to new data.

- Greater adaptability to diverse data.

Attention: While composing activation functions can offer benefits, it's important to choose them judiciously and with consideration for the specific problem at hand, as some combinations might not be beneficial or could even lead to unwanted behaviors like exploding gradients. Empirical testing and validation are crucial when exploring different activation function compositions.

The author was greatly inspired and motivated by [3] and was the pioneer of quantitative neural network approximation, see [1], and since then he has published numerous of papers and books, e.g. see [2].

In this article we continue this trend.

In mathematical neural network approximation AMS Mathscinet lists no articles related to composition of activation functions. So this is seminal work.

By using composition of activation functions we achieve the first extensive part of this introduction and most notably this composition leads to a multivariate activation function of compact support, though the initial activation functions had an infinite domain, the whole real line.

Now the resulting activation function is a multivariate open cusp of compact support $[-1,1]^d$, $d \in \mathbb{N}$. Our involved activation functions are very general, and the constructed multivariate neural network operators resemble the squashing operators in [1], [2], and so do the produced quantitative results.

As a result our produced convergence multivariate inequalities look much simpler and nicer.

Of great inspiration are the articles [4], [5], [6]. References [7], [8], [9] are foundational.

2. Foundamentals. Let $i = 1, 2$, and $h_i \colon \mathbb{R} \to [-1,1]$ 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 $h_i^{(2)} \in C(\mathbb{R})$.

Clearly, $h_1 \circ h_2 = h_1\big|_{(-1,1)} \circ h_2$ is strictly increasing and $(h_1 \circ h_2)(0) = 0$, and

$$h_1 \circ h_2(-x) = h_1(h_2(-x)) = h_1(-h_2(x)) = -h_1(h_2(x)) = -(h_1 \circ h_2)(x),$$

that is

$$(h_1 \circ h_2)(-x) = -(h_1 \circ h_2)(x), \ \ \forall\, x \in \mathbb{R}.$$

Furthermore

$$(h_1 \circ h_2)(+\infty) = h_1(h_2(+\infty)) = h_1(1),$$
$$(h_1 \circ h_2)(-\infty) = h_1(h_2(-\infty)) = h_1(-1).$$

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

$$h_2(\lambda x + \mu y) \leqslant \lambda h_2(x) + \mu h_2(y), \quad x, y \in \mathbb{R}_-;$$
$$h_1(h_2(\lambda x + \mu y)) \leqslant h_1(\lambda h_2(x) + \mu h_2(y)) \leqslant \lambda h_1(h_2(x)) + \mu h_1(h_2(y)),$$

i.e.

$$(h_1 \circ h_2)(\lambda x + \mu y) \leqslant \lambda (h_1 \circ h_2)(x) + \mu (h_1 \circ h_2)(y),$$

$x, y \in \mathbb{R}_-$.

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

Similarly, over $[0, +\infty)$ we get: let $\lambda, \mu \geqslant 0$: $\lambda + \mu = 1$. Then, by concavity of h_2, we have

$$h_2(\lambda x + \mu y) \geqslant \lambda h_2(x) + \mu h_2(y), \quad x, y \in \mathbb{R}_+;$$

$$h_1(h_2(\lambda x + \mu y)) \geqslant h_1(\lambda h_2(x) + \mu h_2(y)) \geqslant \lambda h_1(h_2(x)) + \mu h_1(h_2(y)).$$

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

Also, it is

$$(h_1(h_2(x)))'' = h_1''(h_2(x))(h_2'(x))^2 + h_1'(h_2(x))h_2''(x) \in C(\mathbb{R}), \quad x \in \mathbb{R}.$$

So $h_1 \circ h_2$ is, generally speaking, a sigmoid activation function.

Next, we consider the function

$$\psi_{1,2}(x) := \frac{1}{4}(h_1 \circ h_2(x+1) - h_1 \circ h_2(x-1)) > 0, \quad \forall\, x \in \mathbb{R}.$$

We observe that

$$\psi_{1,2}(-x) = \frac{1}{4}(h_1(h_2(-x+1)) - h_1(h_2(-x-1))) =$$
$$\frac{1}{4}(h_1(h_2(-(x-1))) - h_1(h_2(-(x+1)))) =$$
$$\frac{1}{4}(h_1(-h_2(x-1)) - h_1(-h_2(x+1))) =$$
$$\frac{1}{4}(-h_1(h_2(x-1)) + h_1(h_2(x+1))) =$$

$$\frac{1}{4}\left(h_1h_2\left(x+1\right)-h_1h_2\left(x-1\right)\right)=\psi_{1,2}\left(x\right),$$

that is

$$\psi_{1,2}\left(-x\right)=\psi_{1,2}\left(x\right),\ \ \forall\ x\in\mathbb{R}.$$

So $\psi_{1,2}$ can serve as a density function in general.

So we have $h_2\colon\mathbb{R}\to(-1,1)$, $h_1|_{(-1,1)}\colon(-1,1)\to(-1,1)$, and the strictly increasing function $H:=h_1|_{(-1,1)}\circ h_2\colon\mathbb{R}\to(-1,1)$, with the graph of H containing an arc of finite length, such that $H\left(0\right)=0$, starting at $(-1,h_1(h_2\left(-1\right)))$ and terminating at $(1,h_1(h_2(1)))$. We call this arc also H. In particular, H is negative and convex over $(-1,0]$, and it is positive and concave over $[0,1)$.

So it has compact support $[-1,1]$ and it is like a squashing function, see [2, Ch. 1, p. 8].

We will work from now on with $|H|$, which has as a graph a cusp joining the points $(-1,|h_1(h_2\left(-1\right))|)$, $(0,0)$, $(1,h_1(h_2\left(1\right)))$ and with compact support, again, $[-1,1]$. The points $(-1,|h_1\left(h_2\left(-1\right)\right)|)$, $(1,h_1(h_2\left(1\right)))$ belong to the graph of $|H|$ and $(0,0)$ as well.

Typically H has a steeper slope than of h_2, but it is flatter and closer to the x-axis than h_2 is, e.g. $\tanh\left(\tanh x\right)$ has asymptotes ± 0.76, while $\tanh x$ has asymptotes ± 1, notice that $\tanh\left(1\right)=0.76$. Clearly H has applications in spiking neural networks.

Our multivariate activation function here will be $F\colon\mathbb{R}^d\to\mathbb{R}$, $d\in\mathbb{N}$: $F\left(x\right):=\prod_{i=1}^{d}|H\left(x_i\right)|$, where $x:=\left(x_1,\ldots,x_d\right)\in\mathbb{R}^d$, which has compact support $[-1,1]^d$.

Clearly, here F is a multivariate d-dimensional open cusp function. In particular, for all $i=1,\ldots,d$,

$$t\to\left|H\left(x_1\right)\right|\ldots\left|H\left(x_{i-1}\right)\right|\left|H\left(t\right)\right|\left|H\left(x_{i+1}\right)\right|\ldots\left|H\left(x_d\right)\right|$$

is an open cusp function going through $(0,0)$, etc.

Let $f\colon\mathbb{R}^d\to\mathbb{R}$ be either a uniformly continuous function or a continuous and bounded function.

We call

$$\omega_1\left(f,h\right):=\sup_{\substack{\text{all } x,y\in\mathbb{R}^d:\\ \|x-y\|_\infty\leqslant h}}\left|f\left(x\right)-f\left(y\right)\right|,\tag{1}$$

where $h > 0$, the multivariate first modulus of continuity of f, where $\|\cdot\|_\infty$ denotes the maximum norm in $\mathbb{R}^d$, $d \in \mathbb{N}$.

In this article, we study the pointwise, uniform, and L_p, $p \geqslant 1$, convergences with rates over $\mathbb{R}^d$, to the unit operator of the "multivariate normalized cusp neural network operators"

$$B_n(f)(x) := \frac{\sum_{k_1=-n^2}^{n^2} \cdots \sum_{k_d=-n^2}^{n^2} f\left(\frac{k_1}{n}, \ldots, \frac{k_d}{n}\right) F\left(n^{1-\alpha}\left(x_1 - \frac{k_1}{n}\right), \ldots, n^{1-\alpha}\left(x_d - \frac{k_d}{n}\right)\right)}{\Theta(x)}, \tag{2}$$

where

$$\Theta(x) := \sum_{k_1=-n^2}^{n^2} \cdots \sum_{k_d=-n^2}^{n^2} F\left(n^{1-\alpha}\left(x_1 - \frac{k_1}{n}\right), \ldots, n^{1-\alpha}\left(x_d - \frac{k_d}{n}\right)\right), \tag{3}$$

$0 < \alpha < 1$ and $n \in \mathbb{N}$, $x \in \mathbb{R}^d$.

One can rewrite

$$B_n(f)(x) := \frac{\sum_{k_1=-n^2}^{n^2} \cdots \sum_{k_d=-n^2}^{n^2} f\left(\frac{k_1}{n}, \ldots, \frac{k_d}{n}\right) \left(\prod_{i=1}^{d} \left|H\left(n^{1-\alpha}\left(x_i - \frac{k_i}{n}\right)\right)\right|\right)}{\sum_{k_1=-n^2}^{n^2} \cdots \sum_{k_d=-n^2}^{n^2} \left(\prod_{i=1}^{d} \left|H\left(n^{1-\alpha}\left(x_i - \frac{k_i}{n}\right)\right)\right|\right)}. \tag{4}$$

Clearly B_n is a positive linear operator.

The terms in the ratio of sums (2) can be nonnegative and the expression is well-defined if and only if

$$0 < \left|n^{1-\alpha}\left(x_i - \frac{k_i}{n}\right)\right| \leqslant 1, \tag{5}$$

i.e. $0 < \left|x_i - \frac{k_i}{n}\right| \leqslant \frac{1}{n^{1-\alpha}}$, all $i = 1, \ldots, d$, if and only if

$$nx_i - n^\alpha \leqslant k_i \leqslant nx_i + n^\alpha, \quad x_i \neq \frac{k_i}{n},$$

all $i = 1, \ldots, d$.

To have the order

$$-n^2 \leqslant nx_i - n^\alpha \leqslant k_i \leqslant nx + n^\alpha \leqslant n^2, \tag{6}$$

we need

$$n \geqslant 1 + |x_i|, \quad x_i \neq \frac{k_i}{n}, \quad \text{all } i = 1, \ldots, d.$$

Thus, (6) holds when we take

$$n \geqslant \max_{i \in \{1,\ldots,d\}} (1 + |x_i|), \quad x_i \neq \frac{k_i}{n}, \quad \text{all } i = 1, \ldots, d. \tag{7}$$

When $x_i \in [-1, 1]$, $x_i \neq \frac{k_i}{n}$, it is enough to assume $n \geqslant 2$, which implies (6), and $x_i \neq \frac{k_i}{n}$, all $i = 1, \ldots, d$.

But the isolated case of $x_i = \frac{k_i}{n}$, for some $i = 1, \ldots, d$, contributes nothing and can be ignored.

Thus, without loss of generality, we may always assume that $x_i \neq \frac{k_i}{n}$, all $i = 1, \ldots, d$.

Hence, for $x_i \in [-1, 1]$ and $n \geqslant 2$, the estimate (6) holds for all $i = 1, \ldots, d$.

Consider the closed interval $J_i := [nx_i - n^{\alpha}, nx_i + n^{\alpha}]$, $i = 1, \ldots, d$, $n \in \mathbb{N}$.

The length of J_i is $2n^{\alpha}$. By Proposition 1 of [1], we get that the [cardinality of $k_i \in \mathbb{Z}$ that belong to J_i]$:= card\,(k_i) \geqslant \max(2n^{\alpha} - 1, 0)$, any $i \in \{1, \ldots, d\}$.

In order to have $card\,(k_i) \geqslant 1$, we need $2n^{\alpha} - 1 \geqslant 1$ if and only if $n \geqslant 1$, any $i \in \{1, \ldots, d\}$.

Therefore, a sufficient condition in order to get the order (6) along with the interval J_i to contain at least one integer for all $i = 1, \ldots, d$ is that

$$n \geqslant \max_{i \in \{1,\ldots,d\}} \{1 + |x_i|\}. \tag{8}$$

Clealry, as $n \to +\infty$ we get that $card\,(k_i) \to +\infty$, all $i = 1, \ldots, d$. Also notice that $card\,(k_i)$ equals to the cardinality of integers in $[\lceil nx_i - n^{\alpha} \rceil, [nx_i + n^{\alpha} t]]$, for all $i = 1, \ldots, d$. Here, $[\cdot]$ denotes the integral part of the number, while $\lceil \cdot \rceil$ denotes its ceiling.

Henceforth, we assume (8). Consequently, it holds

$$B_n(f)(x) = \frac{\sum_{k_1 = \lceil nx_1 - n^{\alpha} \rceil}^{[nx_1 + n^{\alpha}]} \cdots \sum_{k_d = \lceil nx_d - n^{\alpha} \rceil}^{[nx_d + n^{\alpha}]} f\left(\frac{k_1}{n}, \ldots, \frac{k_d}{n}\right) F\left(n^{1-\alpha}\left(x_1 - \frac{k_1}{n}\right), \ldots, n^{1-\alpha}\left(x_d - \frac{k_d}{n}\right)\right)}{\Theta(x)}, \tag{9}$$

where

$$\Theta(x) = \sum_{k_1=\lceil nx_1-n^\alpha\rceil}^{[nx_1+n^\alpha]} \cdots \sum_{k_d=\lceil nx_d-n^\alpha\rceil}^{[nx_d+n^\alpha]} F\Big(n^{1-\alpha}\Big(x_1-\frac{k_1}{n}\Big),\ldots,n^{1-\alpha}\Big(x_d-\frac{k_d}{n}\Big)\Big), \tag{10}$$

where $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $0 < \alpha < 1$, $n \in \mathbb{N}$.

So, $\left|n^{1-\alpha}\left(x_i - \frac{k_i}{n}\right)\right| \leqslant 1$, $i = 1, \ldots, d$, if and only if

$$\left\|x - \frac{\vec{k}}{n}\right\|_\infty \leqslant \frac{1}{n^{1-\alpha}}, \tag{11}$$

where $\vec{k} := (k_1, \ldots, k_d)$.

Notice that $B_n(1) = 1$, $n \in \mathbb{N}$.

3. Main results. We now present our first main result.

Theorem 1. *Let $x \in \mathbb{R}^d$, $d \in \mathbb{N}$, $x = (x_1, \ldots, x_d)$, and $n \in \mathbb{N}$ such that $n \geqslant \max_{i\in\{1,\ldots,d\}} \{1 + |x_i|\}$, and $f\colon \mathbb{R}^d \to \mathbb{R}$ be either a uniformly continuous function or a bounded function. Then*

$$|B_n(f)(x) - f(x)| \leqslant \omega_1\Big(f, \frac{1}{n^{1-\alpha}}\Big). \tag{12}$$

Inequality (12) is attained by constant functions, and when f is uniformly continuous, it gives the pointwise convergence with rates, as $n \to +\infty$, of $B_n(f)(x) \to f(x)$.

When $n \geqslant 2$, we obtain

$$\|B_n(f) - f\|_{\infty,[-1,1]^d} \leqslant \omega_1\Big(f, \frac{1}{n^{1-\alpha}}\Big). \tag{13}$$

If f is uniformly continuous, we get that $\lim_{n\to+\infty} B_n(f) = f$, uniformly over $[-1,1]^d$.

Proof. We observe that

$$|B_n(f)(x) - f(x)| \overset{(9)}{=}$$

$$\left|\frac{\sum_{k_1=\lceil nx_1-n^\alpha\rceil}^{[nx_1+n^\alpha]} \cdots \sum_{k_d=\lceil nx_d-n^\alpha\rceil}^{[nx_d+n^\alpha]} f\Big(\frac{k_1}{n},\ldots\frac{k_d}{n}\Big) F\Big(n^{1-\alpha}\Big(x_1-\frac{k_1}{n}\Big),\ldots,n^{1-\alpha}\Big(x_d-\frac{k_d}{n}\Big)\Big)}{\Theta(x)} - f(x)\right| =$$

$$\left|\frac{\sum\limits_{k_1=\lceil nx_1-n^{\alpha}\rceil}^{[nx_1+n^{\alpha}]}\cdots\sum\limits_{k_d=\lceil nx_d-n^{\alpha}\rceil}^{[nx_d+n^{\alpha}]}\left(f\left(\frac{\vec{k}}{n}\right)-f\left(x\right)\right)F\left(n^{1-\alpha}\left(x-\frac{\vec{k}}{n}\right)\right)}{\Theta\left(x\right)}\right|\leqslant$$

$$\frac{\sum\limits_{\vec{k}=\lceil nx-n^{\alpha}\rceil}^{[nx+n^{\alpha}]}\left|f\left(\frac{\vec{k}}{n}\right)-f\left(x\right)\right|F\left(n^{1-\alpha}\left(x-\frac{\vec{k}}{n}\right)\right)}{\Theta\left(x\right)}\leqslant$$

$$\frac{\sum\limits_{\vec{k}=\lceil nx-n^{\alpha}\rceil}^{[nx+n^{\alpha}]}\omega_1\left(f,\left\|x-\frac{\vec{k}}{n}\right\|_{\infty}\right)F\left(n^{1-\alpha}\left(x-\frac{\vec{k}}{n}\right)\right)}{\Theta\left(x\right)}\leqslant$$

$$\omega_1\left(f,\frac{1}{n^{1-\alpha}}\right).$$

□

Our second main result follows:

Theorem 2. *Let $x\in\mathbb{R}^d$, $f\in C^N\left(\mathbb{R}^d\right)$, $N\in\mathbb{N}$, such that all of its partial derivatives $f_{\widetilde{\alpha}}$ of order N, $\widetilde{\alpha}\colon\left|\widetilde{\alpha}\right|=N$, are uniformly continuous or continuous and bounded. Here $n\in\mathbb{N}\colon n\geqslant\max\limits_{i\in\{1,\ldots,d\}}\left\{1+\left|x_i\right|\right\}$. Then*

$$\left|\left(B_n\left(f\right)\right)\left(x\right)-f\left(x\right)\right|\leqslant\left\{\sum_{j=1}^{N}\frac{1}{j!n^{j(1-\alpha)}}\left(\left(\sum_{i=1}^{d}\left|\frac{\partial}{\partial x_i}\right|\right)^j f\left(x\right)\right)\right\}+\frac{d^N}{N!n^{N(1-\alpha)}}\max_{\widetilde{\alpha}\colon\left|\widetilde{\alpha}\right|=N}\omega_1\left(f_{\widetilde{\alpha}},\frac{1}{n^{1-\alpha}}\right).\tag{15}$$

Inequality (15) is attained by constant functions and gives us the pointwise convergence of $B_n\left(f\right)\to f$ over $\mathbb{R}^d$, as $n\to+\infty$.

If $n\geqslant 2$ and $f\in C^N\left(\mathbb{R}^d\right)$ has all of its partial derivatives bounded, we derive

$$\left\|B_n\left(f\right)-f\right\|_{\infty,[-1,1]^d}\leqslant\left\{\sum_{j=1}^{N}\frac{1}{j!n^{j(1-\alpha)}}\left(\left(\sum_{i=1}^{d}\left\|\frac{\partial}{\partial x_i}\right\|_{\infty,[-1,1]^d}\right)^j f\left(x\right)\right)\right\}+\frac{d^N}{N!n^{N(1-\alpha)}}\max_{\widetilde{\alpha}\colon\left|\widetilde{\alpha}\right|=N}\omega_1\left(f_{\widetilde{\alpha}},\frac{1}{n^{1-\alpha}}\right).\tag{16}$$

As $n\to+\infty$, we get $B_n\left(f\right)\to f$ uniformly over $[-1,1]^d$.

Proof. Here $x = \vec{x} = (x_1, \ldots, x_d)$. Set

$$g_{\frac{\vec{k}}{n}}(t) := f\Big(x + t\Big(\frac{\vec{k}}{n} - x\Big)\Big), \ \ 0 \leqslant t \leqslant 1. \tag{17}$$

Then

$$g^{(j)}_{\frac{\vec{k}}{n}}(t) = \Big[\Big(\sum_{i=1}^{d}\Big(\frac{k_i}{n} - x_i\Big)\frac{\partial}{\partial x_i}\Big)^j f\Big] t\Big(x_1 + t\Big(\frac{k_1}{n} - x_1\Big), \ldots, x_d + t\Big(\frac{k_d}{n} - x_d\Big)\Big) \tag{18}$$

and $g_{\frac{\vec{k}}{n}}(0) = f(x)$.

By Taylor's formula, we get

$$f\Big(\frac{k_1}{n}, \ldots, \frac{k_d}{n}\Big) = g_{\frac{\vec{k}}{n}}(1) = \sum_{j=0}^{N} \frac{g^{(j)}_{\frac{\vec{k}}{n}}(0)}{j!} + R_N\Big(\frac{\vec{k}}{n}, 0\Big), \tag{19}$$

where

$$R_N\Big(\frac{\vec{k}}{n}, 0\Big) = \int_0^1 \Big(\int_0^{t_1} \ldots \Big(\int_0^{t_{N-1}} \Big(g^{(N)}_{\frac{\vec{k}}{n}}(t_N) - g^{(N)}_{\frac{\vec{k}}{n}}(0)\Big) dt_N\Big) \ldots\Big) dt_1. \tag{20}$$

Here we denote by

$$f_{\widetilde{\alpha}} := \frac{\partial^{\widetilde{\alpha}} f}{\partial x^{\widetilde{\alpha}}}, \ \ \widetilde{\alpha} := (\alpha_1, \ldots, \alpha_d), \quad \alpha_i \in \mathbb{Z}^+,$$

$i = 1, \ldots, d$, such that $|\widetilde{\alpha}| := \sum_{i=1}^{d} \alpha_i = N$. Thus,

$$\frac{f\left(\frac{\vec{k}}{n}\right) F\left(n^{1-\alpha}\left(\vec{x} - \frac{\vec{k}}{n}\right)\right)}{\Theta(\vec{x})} =$$
$$\sum_{j=0}^{N} \frac{g^{(j)}_{\frac{\vec{k}}{n}}(0)}{j!} \frac{F\left(n^{1-\alpha}\left(\vec{x} - \frac{\vec{k}}{n}\right)\right)}{\Theta(\vec{x})} + \frac{F\left(n^{1-\alpha}\left(\vec{x} - \frac{\vec{k}}{n}\right)\right)}{\Theta(\vec{x})} \cdot R_N\Big(\frac{\vec{k}}{n}, 0\Big). \tag{21}$$

Therefore

$$(B_n(f))(\vec{x}) - f(\vec{x}) =$$
$$\sum_{\vec{k}=\lceil n\vec{x}-n^{\alpha}\rceil}^{[n\vec{x}+n^{\alpha}]} \frac{f\left(\frac{\vec{k}}{n}\right)}{\Theta(\vec{x})} F\left(n^{1-\alpha}\left(\vec{x}-\frac{\vec{k}}{n}\right)\right) - f(\vec{x}) =$$
$$\sum_{j=1}^{N} \frac{1}{j!}\left(\sum_{\vec{k}=\lceil n\vec{x}-n^{\alpha}\rceil}^{[n\vec{x}+n^{\alpha}]} g_{\frac{\vec{k}}{n}}^{(j)}(0)\, \frac{F\left(n^{1-\alpha}\left(\vec{x}-\frac{\vec{k}}{n}\right)\right)}{\Theta(\vec{x})}\right) + R^{*}, \tag{22}$$

where

$$R^{*} := \sum_{\vec{k}=\lceil n\vec{x}-n^{\alpha}\rceil}^{[n\vec{x}+n^{\alpha}]} \frac{F\left(n^{1-\alpha}\left(\vec{x}-\frac{\vec{k}}{n}\right)\right)}{\Theta(\vec{x})} \cdot R_N\left(\frac{\vec{k}}{n}, 0\right). \tag{23}$$

Consequently, we obtain

$$|(B_n(f))(\vec{x}) - f(\vec{x})| \leqslant$$
$$\sum_{j=1}^{N} \frac{1}{j!}\left(\sum_{\vec{k}=\lceil n\vec{x}-n^{\alpha}\rceil}^{[n\vec{x}+n^{\alpha}]} \frac{\left|g_{\frac{\vec{k}}{n}}^{(j)}(0)\right| F\left(n^{1-\alpha}\left(\vec{x}-\frac{\vec{k}}{n}\right)\right)}{\Theta(\vec{x})}\right) + |R^{*}| =: T. \tag{24}$$

Notice that

$$\left|g_{\frac{\vec{k}}{n}}^{(j)}(0)\right| \leqslant \left(\frac{1}{n^{1-\alpha}}\right)^{j}\left(\left(\sum_{i=1}^{d}\left|\frac{\partial}{\partial x_i}\right|\right)^{j} f(\vec{x})\right)$$

and

$$T \leqslant \left\{\sum_{j=1}^{N} \frac{1}{j!}\left(\frac{1}{n^{1-\alpha}}\right)^{j}\left(\left(\sum_{i=1}^{d}\left|\frac{\partial}{\partial x_i}\right|\right)^{j} f(\vec{x})\right)\right\} + |R^{*}|. \tag{25}$$

That is, by (25), we get

$$|(B_n(f))(\vec{x}) - f(\vec{x})| \leqslant$$
$$\left\{\sum_{j=1}^{N} \frac{1}{j!\,n^{j(1-\alpha)}}\left(\left(\sum_{i=1}^{d}\left|\frac{\partial}{\partial x_i}\right|\right)^{j} f(\vec{x})\right)\right\} + \left|R^{*}\right|. \tag{26}$$

Next, we need to estimate $|R^*|$. For that, we observe $(0 \leqslant t_N \leqslant 1)$

$$\left| g_{\frac{\vec{k}}{n}}^{(N)}(t_N) - g_{\frac{\vec{k}}{n}}^{(N)}(0) \right| =$$

$$\left| \left(\sum_{i=1}^{d} \left(\frac{k_i}{n} - x_i \right) \frac{\partial}{\partial x_i} \right)^N f\left(\vec{x} + t_N \left(\frac{\vec{k}}{n} - \vec{x} \right) \right) - \left(\sum_{i=1}^{d} \left(\frac{k_i}{n} - x_i \right) \frac{\partial}{\partial x_i} \right)^N f(\vec{x}) \right|$$

$$\leqslant \frac{d^N}{n^{N(1-\alpha)}} \max_{\tilde{\alpha}:|\tilde{\alpha}|=N} \omega_1 \left(f_{\tilde{\alpha}}, \frac{1}{n^{1-\alpha}} \right). \tag{27}$$

Thus,

$$\left| R_N \left(\frac{\vec{k}}{n}, 0 \right) \right| \leqslant$$

$$\int_0^1 \left(\int_0^{t_1} \ldots \left(\int_0^{t_{N-1}} \left| g_{\frac{\vec{k}}{n}}^{(N)}(t_N) - g_{\frac{\vec{k}}{n}}^{(N)}(0) \right| dt_N \right) \ldots \right) dt_1 \leqslant \tag{28}$$

$$\frac{d^N}{N! n^{N(1-\alpha)}} \max_{\tilde{\alpha}:|\tilde{\alpha}|=N} \omega_1 \left(f_{\tilde{\alpha}}, \frac{1}{n^{1-\alpha}} \right).$$

Therefore,

$$|R^*| \leqslant \sum_{\vec{k}=\lceil n\vec{x}-n^{\alpha} \rceil}^{[n\vec{x}+n^{\alpha}]} \frac{F\left(n^{1-\alpha} \left(\vec{x} - \frac{\vec{k}}{n} \right) \right)}{\Theta(\vec{x})} \left| R_N \left(\frac{\vec{k}}{n}, 0 \right) \right| \leqslant \tag{29}$$

$$\frac{d^N}{N! n^{N(1-\alpha)}} \max_{\tilde{\alpha}:|\tilde{\alpha}|=N} \omega_1 \left(f_{\tilde{\alpha}}, \frac{1}{n^{1-\alpha}} \right).$$

By (26) and (29) we get (15). □

It follows:

Corollary 1. *(to Theorem 1.) Let*

$$\Delta := \prod_{i=1}^{d} [-\delta_i, \delta_i] \subset \mathbb{R}^d, \quad \delta_i > 0,$$

and take $n \geqslant \max_{i \in \{1,\ldots,d\}} (1 + \delta_i)$.

Take $p \geqslant 1$. *Then*

$$\|B_n f - f\|_{p,\Delta} \leqslant \omega_1 \left(f, \frac{1}{n^{1-\alpha}} \right) 2^{\frac{d}{p}} \prod_{i=1}^{d} \delta_i^{\frac{1}{p}}, \tag{30}$$

attained by constant functions. If f is uniformly continuous, by (30), we get the L_p convergence of $B_n f$ to f with rates over Δ.

Proof. By (12). □

We finish with

Corollary 2. *(to Theorem 2.) Let*

$$\Delta := \prod_{i=1}^{d} [-\delta_i, \delta_i] \subset \mathbb{R}^d, \quad \delta_i > 0,$$

and take $n \geqslant \max_{i \in \{1,\ldots,d\}} (1 + \delta_i)$; $p \geqslant 1$.

Then

$$\|B_n f - f\|_{p,\Delta} \leqslant \left\{ \sum_{j=1}^{N} \frac{1}{j! n^{j(1-\alpha)}} \left\| \left(\sum_{i=1}^{d} \left| \frac{\partial}{\partial x_i} \right| \right)^j f \right\|_{p,\Delta} \right\} + \frac{d^N}{N! n^{N(1-\alpha)}} \max_{\widetilde{\alpha}: |\widetilde{\alpha}| = N} \omega_1 \left(f_{\widetilde{\alpha}}, \frac{1}{n^{1-\alpha}} \right) 2^{\frac{d}{p}} \prod_{i=1}^{d} \delta_i^{\frac{1}{p}}, \tag{31}$$

attained by constant. By (31) we derive the L_p convergence with rates of $B_n(f)$ to f over Δ.

Proof. By (15). □

References

[1] Anastassiou G. A. *Rate of convergence of some neural network operators to the unit-univariate case.* J. Math. Anal. Appl., 1997, vol. 22, pp. 237 – 262. DOI: https://doi.org/10.1006/jmaa.1997.5494

[2] Anastassiou G. A. *Intelligent systems II: Complete approximation by neural network operators.* 2016, Springer, Heidelberg, New York.

[3] Cardaliaguet P., Euvrard G. *Approximation of a function and its derivative with a neural network.* Neural Networks, 1992, vol. 5, pp. 207 – 220. DOI: https://doi.org/10.1016/S0893-6080(05)80020-6

[4] Chen Z., Cao F. *The approximation operators with sigmoidal functions.* Computers and Mathematics with Applications, 2009, vol. 58, pp. 758 – 765. DOI: https://doi.org/10.1016/j.camwa.2009.05.001

[5] Costarelli D., Spigler R. *Approximation results for neural network operators activated by sigmoidal functions.* Neural Networks, 2013, vol. 44, pp. 101 – 106. DOI: https://doi.org/10.1016/j.neunet.2013.03.015

[6] Costarelli D., Spigler R. *Multivariate neural network operators with sigmoidal activation functions.* Neural Networks, 2013, vol. 48, pp. 72 – 77. DOI: https://doi.org/10.1016/j.neunet.2013.07.009

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

[8] McCulloch W., Pitts W. *A logical calculus of the ideas immanent in nervous activity.* Bulletin of Mathematical Biophysics, 1943, vol. 7, pp. 115 – 133.

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

Received September 24, 2025.
In revised form, November 20, 2025.
Accepted November 20, 2025.
Published online December 7, 2025.

Department of Mathematical Sciences
University of Memphis
Memphis, TN 38152, U.S.A.
E-mail: ganastss@memphis.edu