

Article

# Inverse Result of Approximation for the Max-Product Neural Network Operators of the Kantorovich Type and Their Saturation Order

Marco Cantarini <sup>1</sup>, Lucian Coroianu <sup>2,\*</sup>, Danilo Costarelli <sup>3</sup>, Sorin G. Gal <sup>2</sup> and Gianluca Vinti <sup>3</sup>

<sup>1</sup> Department of Industrial Engineering and Mathematical Sciences, Marche Polytechnic University, 60121 Ancona, Italy; m.cantarini@univpm.it

<sup>2</sup> Department of Mathematics and Computer Science, University of Oradea, 410087 Oradea, Romania; galso@uoradea.ro

<sup>3</sup> Department of Mathematics and Computer Science, University of Perugia, 06123 Perugia, Italy; danilo.costarelli@unipg.it (D.C.); gianluca.vinti@unipg.it (G.V.)

\* Correspondence: lcoroianu@uoradea.ro

**Abstract:** In this paper, we consider the max-product neural network operators of the Kantorovich type based on certain linear combinations of sigmoidal and ReLU activation functions. In general, it is well-known that max-product type operators have applications in problems related to probability and fuzzy theory, involving both real and interval/set valued functions. In particular, here we face inverse approximation problems for the above family of sub-linear operators. We first establish their saturation order for a certain class of functions; i.e., we show that if a continuous and non-decreasing function  $f$  can be approximated by a rate of convergence higher than  $1/n$ , as  $n$  goes to  $+\infty$ , then  $f$  must be a constant. Furthermore, we prove a local inverse theorem of approximation; i.e., assuming that  $f$  can be approximated with a rate of convergence of  $1/n$ , then  $f$  turns out to be a Lipschitz continuous function.

**Keywords:** sigmoidal functions; ReLU function; neural network operators; saturation result; local inverse theorem



**Citation:** Cantarini, M.; Coroianu, L.; Costarelli, D.; Gal, S.G.; Vinti, G.

Inverse Result of Approximation for the Max-Product Neural Network Operators of the Kantorovich Type and Their Saturation Order.

*Mathematics* **2022**, *10*, 63. <https://doi.org/10.3390/math10010063>

Academic Editor: Xiangmin Jiao

Received: 6 December 2021

Accepted: 23 December 2021

Published: 25 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The introduction of the max-product version of families of linear approximation operators is due to Bede, Coroianu and Gal (see, e.g., [1,2]) and it led to a new branch of approximation theory. The new theory of max-product operators has been deeply studied, and recently the above-mentioned authors summarized their results in a complete monograph [3].

In general, the max-product version of a sequence/net of linear operators is a family of nonlinear (more precisely sub-linear) operators with better approximation properties of their original version: in many cases, the order of convergence is faster than their linear counterparts [4–6]. Further, the above operators can also be useful, for instance, in the applications of probability and fuzzy theory involving both real and interval/set valued functions (see, e.g., [7,8]).

In the present paper, we study the max-product form of the neural network (NN) operators of the Kantorovich type  $K_n^{(M)}$ , first introduced in [9] and here recalled in Definition 3 of Section 2. In general, the NN operators (see [10]) are strictly related to the theory of artificial neural networks, which has been introduced in order to provide a very simple model for the human brain, which is able to reproduce all its main abilities [11–14].

Each basic element composing a neural network is called an artificial neuron; its behavior is regulated by suitable activation functions, which must represent the two possible states of the biological neuron: the activation and the quiet phases [15]. From the

mathematical point of view, functions which better represent the latter fact are those with sigmoidal shape (see, e.g., [16]).

For the above reasons, in the present paper, we mainly consider the operators  $K_n^{(M)}$  activated by suitable sigmoidal functions. Very useful examples of sigmoidal functions (in view of their importance in learning algorithms, [17]) are, e.g., the logistic and the hyperbolic tangent functions [18].

However, independently of its biological meaning, in some recent papers, a new unbounded activation function has also been introduced and deeply investigated. This function is the so-called rectified linear unit (ReLU) function (see, e.g., [19]), and it is simply defined by the positive part of  $x$ , for every  $x \in \mathbb{R}$ . The ReLU activation is revealed to be very suitable for training deep (i.e., multi-layer) neural networks, in view of the very simple form that is assumed by its derivative (whenever it exists).

Here, we show that the above operators can also be based on a certain finite linear combination of ReLU activation function, and in this case, the approximation properties of  $K_n^{(M)}$  are also preserved.

Problems of interpolation, or more in general, of approximation, are related to the topic of training a neural network by sample values belonging to a certain training set: this explains the interest in studying approximation results by means of NN operators in various contexts [15,20–24].

Indeed, as can be seen from the references therein, results in this sense have been studied deeply in terms of various aspects, such as the convergence and the order of approximation.

In this paper, we deal with the problem of the saturation order and of inverse results of approximation.

In general, the problem of establishing the saturation order for a family of operators  $L_n$  (see [25–28]),  $n \in \mathbb{N}$ , consists in determining a class of functions  $\mathcal{D}$ , a certain subclass  $\mathcal{E}$  of trivial functions of  $\mathcal{D}$ , and a positive non-increasing function  $\varphi(n)$ ,  $n \in \mathbb{N}$ , such that there exists  $g \in \mathcal{D} \setminus \mathcal{E}$  with  $\|L_n g - g\| = \mathcal{O}(\varphi(n))$ , as  $n \rightarrow +\infty$ , and with the property that, for any  $f \in \mathcal{D}$  with:

$$\|L_n f - f\| = o(\varphi(n)), \quad n \rightarrow +\infty,$$

it turns out that  $f \in \mathcal{E}$ , and vice versa. Here,  $\|\cdot\|$  denotes any suitable norm on  $\mathcal{D}$ . In this case,  $\varphi(n)$  is said to be the *saturation order* of the approximation process  $L_n$ , and it represents the best possible order of approximation that can be achieved on  $\mathcal{D}$  by the above approximation operators.

In case of the max-product NN operators of the Kantorovich type, according to the studies given in [9,29], we expect that for a certain subclass  $\mathcal{D}$  of  $C([0, 1])$  (endowed with the usual max-norm), with  $\varphi(n) = 1/n$ ,  $n \in \mathbb{N}$ , then  $f$  is constant over  $[0, 1]$ . Hence, we also have that the trivial class of functions  $\mathcal{E}$  is given by constant functions. Indeed, one of the main results that we establish in the present paper is exactly the proof of the above claim.

Further, since in [9] it has been proved that, in the space  $\text{Lip}([0, 1])$ , the order of approximation is exactly  $1/n$ , as  $n$  goes to  $+\infty$ , it is natural to ask if also the converse implication holds.

In this paper, we proved exactly a local version of such an inverse approximation theorem, i.e., if the relation:

$$\left\| K_n^{(M)}(f, \cdot) - f(\cdot) \right\|_\infty = O(1/n), \quad n \rightarrow +\infty,$$

holds, with  $f$  continuous and non-decreasing, then  $f$  belongs to  $\text{Lip}([a, b])$ , for every sub-interval  $[a, b] \subset [0, 1]$ .

## 2. Preliminaries

By  $C([0, 1])$ , we will denote the space of continuous functions  $f : [0, 1] \rightarrow \mathbb{R}$ , while by  $C_+([0, 1])$ , we will indicate the subspace of  $C([0, 1])$  of the non-negative valued functions.

Furthermore, we will denote by  $\text{Lip}([0, 1])$  the subspace of  $C([0, 1])$  of the Lipschitz continuous functions on  $[0, 1]$ , i.e., the space of functions  $f$  for which there exists a positive constant  $L$  such that

$$|f(x) - f(y)| \leq L|x - y|, \quad x, y \in [0, 1].$$

Finally, we also denote by  $\|\cdot\|_\infty$  the classical max-norm. Obviously, all the above notations can be given by replacing the interval  $[0, 1]$  with any bounded or unbounded interval  $I \subset \mathbb{R}$ .

We now recall the definition of a sigmoidal function introduced by Cybenko [30].

**Definition 1.** Let  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  be a measurable function. We call  $\sigma$  a sigmoidal function if

$$\lim_{x \rightarrow -\infty} \sigma(x) = 0, \quad \text{and} \quad \lim_{x \rightarrow +\infty} \sigma(x) = 1.$$

In what follows, we consider non-decreasing sigmoidal functions  $\sigma$  that satisfy the following conditions:

- (Σ1)  $\sigma(x) - 1/2$  is an odd function;
- (Σ2)  $\sigma \in C^2(\mathbb{R})$  is concave for  $x \geq 0$ ;
- (Σ3)  $\sigma(x) = O(|x|^{-\alpha-1})$  as  $x \rightarrow -\infty$  for some  $\alpha > 0$ .

Notice that assumptions (Σ*i*),  $i = 1, 2, 3$  are satisfied by the main examples of sigmoidal functions known in the literature.

For instance, examples of sigmoidal activation functions are given by  $\sigma_\ell(x) := (1 + e^{-x})^{-1}$ ,  $x \in \mathbb{R}$  (i.e., the so-called logistic function), and by  $\sigma_h(x) := (\tanh(x) + 1)/2$ ,  $x \in \mathbb{R}$  (i.e., the so-called hyperbolic tangent activation function). Note that both  $\sigma_\ell(x)$  and  $\sigma_h(x)$  satisfy (Σ3) for all  $\alpha > 0$  in view of their exponential decay at  $x \rightarrow -\infty$ .

Further, we can also recall the definition of the sigmoidal functions that can be generated by the well-known central B-splines of order  $n \in \mathbb{N}^+$

$$M_n(x) := \frac{1}{(n-1)!} \sum_{i=0}^n (-1)^i \binom{n}{i} \left(\frac{n}{2} + x - i\right)_+^{n-1}, \quad x \in \mathbb{R},$$

where the function  $(x)_+ := \max\{x, 0\}$ . Hence, we can define the sigmoidal function  $\sigma_{M_n}$  generated by  $M_n$  as follows:

$$\sigma_{M_n}(x) := \int_{-\infty}^x M_n(t) dt, \quad x \in \mathbb{R}.$$

Obviously,  $\sigma_{M_n}$  satisfies assumption (Σ1) for every  $n \geq 1$ . Further, since  $M_n$  have compact supports, and the supports are contained in the intervals  $\left[-\frac{n}{2}, \frac{n}{2}\right]$ , it turns out that  $\sigma_{M_n}$  also satisfies (Σ3) for every  $\alpha > 0$ . Further, assumption (Σ2) is satisfied for  $n \geq 1$ .

Now, considering (from now on) a function  $\sigma$  that satisfies the above assumptions, we can recall the definition of the density (kernel) function  $\phi_\sigma$ , that is:

$$\phi_\sigma(x) := \frac{1}{2}[\sigma(x+1) - \sigma(x-1)], \quad x \in \mathbb{R}.$$

For the function  $\phi_\sigma$ , the following lemma can be proved.

- Lemma 1.** (i)  $\phi_\sigma(x) \geq 0$  for every  $x \in \mathbb{R}$ , with  $\phi_\sigma(2) > 0$ , and  $\lim_{x \rightarrow \pm\infty} \phi_\sigma(x) = 0$ ;  
 (ii)  $\phi_\sigma(x)$  is an even function;  
 (iii)  $\phi_\sigma(x)$  is non-decreasing for  $x < 0$  and non-increasing for  $x \geq 0$ ;  
 (iv)  $\phi_\sigma(x) = O(|x|^{-\alpha-1})$  as  $x \rightarrow \pm\infty$  where  $\alpha$  is the positive constant of condition (Σ3).

For a proof of conditions (i)–(iv), see [31].

Now, we introduce the following notation used in the literature (see, e.g., [3]) in order to define the so-called max-product type operators.

**Definition 2.** Let  $K_1, K_2$  be two integers with  $K_1 \leq K_2$  and let  $A_k \in \mathbb{R}, k = K_1, \dots, K_2$ . Then we define

$$\bigvee_{k=K_1}^{K_2} A_k := \max\{A_k, k = K_1, \dots, K_2\}.$$

Now, we recall the following lemma that will be useful in order to show that the family of operators investigated in this paper are well-defined.

**Lemma 2.** ([29])  $\bigvee_{k=0}^{n-1} \phi_\sigma(nx - k) \geq \phi_\sigma(2) > 0$  for every  $x \in [0, 1]$ .

The definition of the max-product NN operators of the Kantorovich type can now be recalled.

**Definition 3.** Let  $f : [0, 1] \rightarrow \mathbb{R}$  be a bounded and locally integrable function and let  $n \in \mathbb{N}^+$ . The max-product NN operators of the Kantorovich type activated by  $\sigma$  are defined by:

$$K_n^{(M)}(f, x) := \frac{\bigvee_{k=0}^{n-1} \phi_\sigma(nx - k)n \int_{k/n}^{(k+1)/n} f(u)du}{\bigvee_{k=0}^{n-1} \phi_\sigma(nx - k)}, \quad x \in [0, 1].$$

Clearly, in view of the properties established in Lemma 2, it turns out that  $K_n^{(M)}(f, x)$  are well-defined, and, moreover, it is quite simple to observe that  $\|K_n^{(M)}(f, \cdot)\|_\infty \leq \|f\|_\infty < +\infty$ .

Concerning the assumptions  $(\Sigma 1)$ ,  $(\Sigma 2)$ , and  $(\Sigma 3)$  assumed above, we can observe that condition  $(\Sigma 2)$  could be avoided, requiring that the sigmoidal function  $\sigma$  is such that  $\sigma(3) > \sigma(1)$  and condition (iii) of Lemma 1 is fulfilled by  $\phi_\sigma$ . The main advantage that can be achieved by the latter fact is that one could apply all the approximation results established below also to discontinuous and non-smooth sigmoidal functions.

An example of continuous but non-smooth sigmoidal function (given according to the above remark) is the so-called ramp function  $\sigma_R(x)$  (see [12]) defined as follows:

$$\sigma_R(x) := \begin{cases} 0, & x < -\frac{3}{2} \\ \frac{x}{3} + \frac{1}{2}, & -\frac{3}{2} \leq x \leq \frac{3}{2} \\ 1, & x > 1. \end{cases}$$

In particular,  $\sigma_R$  satisfies condition  $(\Sigma 3)$  for all  $\alpha > 0$ , and  $\phi_{\sigma_R}$  turns out to be a function with compact support; moreover,  $\sigma_R(3) > \sigma_R(1)$ .

Note that (see [27]) the sigmoidal function  $\sigma_{M_1}(3 \cdot)$  coincides with the ramp function  $\sigma_R$ ; now recalling the definition of the well-known rectified linear unit (ReLU) activation function (see, e.g., [19,32]):

$$\psi_{ReLU}(x) := (x)_+, \quad x \in \mathbb{R},$$

it turns out that:

$$\sigma_{M_1}(3x) := \psi_{ReLU}(3x + 1/2) - \psi_{ReLU}(3x - 1/2), \quad x \in \mathbb{R}.$$

Thus, the density function  $\phi_{\sigma_{M_1}(3 \cdot)}$  can be expressed in terms of ReLU activation function as follows:

$$\phi_{\sigma_R}(x) = \phi_{\sigma_{M_1}(3 \cdot)}(x) = \psi_{ReLU}(3x + 1) - 2\psi_{ReLU}(3x) + \psi_{ReLU}(3x - 1), \quad x \in \mathbb{R}.$$

As a consequence of the latter relation, the NN operators  $K_n^{(M)}$  activated by  $\sigma_{M_1}(3 \cdot)$  can be considered as an NN activated by the above linear combination of ReLU activation functions. Recently, it has been proved that  $\psi_{ReLU}$  is very suitable in order to train deep (i.e., multi-layer) neural networks; see, e.g., [33,34]. For more details concerning  $\psi_{ReLU}$ , see also [19,35].

### 3. The Saturation Order

It is well known that, if  $f \in C_+([0, 1])$  the family  $(K_n^{(M)}(f, \cdot))$  converges uniformly to  $f$  (see [9]). Moreover, we also know that the following quantitative estimates:

$$\|K_n^{(M)}(f, \cdot) - f(\cdot)\|_\infty \leq M\omega\left(f, \frac{1}{n}\right), \tag{1}$$

as  $n \rightarrow +\infty$ , there holds if condition  $(\Sigma 3)$  is satisfied for  $\alpha \geq 1, M > 0$  and where:

$$\omega\left(f, \frac{1}{n}\right) := \sup\{|f(x) - f(y)| : x, y \in [0, 1], \text{ with } |x - y| \leq 1/n\}$$

denotes the usual modulus of continuity of the function  $f \in C_+([0, 1])$  (see, e.g., [36]). From the latter result, it turns out that, if the function  $f$  belongs to  $\text{Lip}([0, 1])$ , then the order of uniform approximation is  $1/n$ , as  $n \rightarrow +\infty$ .

In this section, we study the saturation order for the NN operator of the Kantorovich type activated by sigmoidal functions; i.e., we show that  $1/n$  is the best possible order of approximation that can be achieved for non-decreasing functions that belong to  $C_+([0, 1])$ .

In order to reach our main purpose, we need some preliminary lemmas.

**Lemma 3.** For any  $j \in \{0, \dots, n - 1\}, n \in \mathbb{N}^+$ , we have:

$$\bigvee_{k=0}^{n-1} \phi_\sigma(nx - k) = \phi_\sigma(nx - j)$$

for every  $x \in \left[\frac{j}{n} - \frac{1}{2n}, \frac{j}{n} + \frac{1}{2n}\right] \cap [0, 1]$ . Further, if  $x \in \left(1 - \frac{1}{2n}, 1\right]$  we have:

$$\bigvee_{k=0}^{n-1} \phi_\sigma(nx - k) = \phi_\sigma(nx - n + 1).$$

**Proof.** Let  $j \in \{0, \dots, n - 1\}$  be fixed and  $x \in \left[\frac{j}{n} - \frac{1}{2n}, \frac{j}{n} + \frac{1}{2n}\right] \cap [0, 1]$ . Observing that

$$|nx - j| = n|x - j/n| \leq 1/2 \tag{2}$$

and, since  $\phi_\sigma(x)$  is even and non-increasing for  $x \geq 0$ , we get:

$$\phi_\sigma(nx - j) = \phi_\sigma(|nx - j|) \geq \phi_\sigma(1/2) \geq \phi_\sigma(2) > 0.$$

Similarly, if  $k \in \{0, \dots, n - 1\}$  and  $k \neq j$ , we have

$$|nx - k| = n|x - k/n| \geq 1/2 \tag{3}$$

and so, using the previous properties of  $\phi_\sigma(x)$  we have:

$$\phi_\sigma(nx - k) = \phi_\sigma(|nx - k|) \leq \phi_\sigma(1/2).$$

If  $x \in \left(1 - \frac{1}{2n}, 1\right]$  we note that

$$|nx - n + 1| = n|x - (n - 1)/n| \leq 1$$

and, if  $k \neq n - 1$

$$|nx - k| = n|x - k/n| > 1$$

and the claim follows, arguing as in (2) and (3), respectively.  $\square$

**Lemma 4.** Let  $I \subseteq \mathbb{R}$  be a bounded or unbounded interval and  $f \in C(I)$ . Suppose in addition that there exists an absolute positive constant  $C$  with the property that for every  $\varepsilon > 0$  there exists  $n(\varepsilon) \in \mathbb{N}^+$  such that for any  $n \in \mathbb{N}, n \geq n(\varepsilon)$  and  $j \in \mathbb{Z}$ , with  $\frac{j}{n}, \frac{j+1}{n} \in I$ , we have

$$\left| f\left(\frac{j+1}{n}\right) - f\left(\frac{j}{n}\right) \right| \leq \frac{C\varepsilon}{n}. \tag{4}$$

Then  $f$  is a constant function.

**Proof.** Let us choose arbitrary  $x_0, y_0 \in I, x_0 < y_0$  and  $\varepsilon > 0$ . The continuity of  $f$  implies the existence of  $n_0(\varepsilon) \in \mathbb{N}^+$ , such that for any  $x, y \in I, |x - x_0| \leq \frac{1}{n_0(\varepsilon)}, |y - y_0| \leq \frac{1}{n_0(\varepsilon)}$ , we have

$$|f(x) - f(x_0)| \leq \varepsilon, |f(y) - f(y_0)| \leq \varepsilon.$$

We now fix  $n_1 = \max\left\{n(\varepsilon), n_0(\varepsilon), \frac{1}{y_0 - x_0}\right\}$ , where  $n(\varepsilon)$  is the constant arising from the assumptions, and let us choose arbitrary  $n \in \mathbb{N}$  such that  $n \geq n_1$ . Since  $\frac{1}{n_1} \leq y_0 - x_0$ , it follows that there exists  $k \in \mathbb{Z}$  and  $l \in \mathbb{N}$  such that

$$\frac{k-1}{n} \leq x_0 \leq \frac{k}{n} \leq \frac{k+1}{n} \leq \dots \leq \frac{k+l}{n} \leq y_0 \leq \frac{k+l+1}{n}.$$

Applying successively the triangle inequality we get

$$\begin{aligned} |f(x_0) - f(y_0)| &\leq \left| f(x_0) - f\left(\frac{k}{n}\right) \right| + \left| f\left(\frac{k}{n}\right) - f\left(\frac{k+1}{n}\right) \right| \\ &\quad + \dots + \left| f\left(\frac{k+l-1}{n}\right) - f\left(\frac{k+l}{n}\right) \right| \\ &\quad + \left| f\left(\frac{k+l}{n}\right) - f(y_0) \right|. \end{aligned}$$

By relation (4), we have

$$\left| f\left(\frac{k+p}{n}\right) - f\left(\frac{k+p+1}{n}\right) \right| \leq \frac{C\varepsilon}{n}, p = 0, 1, \dots, l-1,$$

and since  $\max\left\{\left|x_0 - \frac{k}{n}\right|, \left|y_0 - \frac{k+l}{n}\right|\right\} \leq \frac{1}{n} \leq \frac{1}{n_0(\varepsilon)}$ , we get

$$|f(x_0) - f(y_0)| \leq \frac{1C\varepsilon}{n} + 2\varepsilon.$$

On the other hand, we observe that  $\frac{k+l}{n} - \frac{k}{n} \leq y_0 - x_0$ , which implies that  $l \leq n(y_0 - x_0)$ . Thus, we obtain

$$|f(x_0) - f(y_0)| \leq C\varepsilon(y_0 - x_0) + 2\varepsilon.$$

Now, since  $\varepsilon > 0$  has been chosen arbitrarily, passing to the infimum for  $\varepsilon > 0$  in the previous inequality, we deduce that  $f(x_0) = f(y_0)$ . By the arbitrariness of  $x_0$  and  $y_0$ , it turns out that  $f$  is a constant function on the whole  $I$ .  $\square$

**Lemma 5.** Let  $I \subseteq \mathbb{R}$  be a bounded or unbounded interval and  $f \in C(I)$  be a non-decreasing function with the property that for any couple  $a, b, a < b$ , of inner points of  $I$ , and for every  $\varepsilon > 0$  there exists  $n(a, b, \varepsilon) \in \mathbb{N}^+$  such that for any  $n \in \mathbb{N}, n \geq n(a, b, \varepsilon)$  and  $j \in \mathbb{Z}$ , such that  $\frac{j+1}{n}, \frac{2j+1}{2n} \in [a, b]$ , we have

$$f\left(\frac{j+1}{n}\right) - f\left(\frac{2j+1}{2n}\right) \leq \frac{\varepsilon}{n}. \tag{5}$$

Then  $f$  is a constant function over  $I$ .

**Proof.** Let  $n \in \mathbb{N}, n \geq n(a, b, \varepsilon)$ , such that  $\frac{1}{n} \leq b - a$ , and let us choose arbitrary  $j \in \mathbb{Z}$  such that  $a \leq \frac{j}{n} \leq \frac{j+1}{n} \leq b$ . We observe that for any  $k \in \mathbb{N}^+$ , we have

$$\frac{j}{n} < \frac{2^k j + 1}{2^k n} < \frac{j+1}{n}.$$

Therefore, applying successively relation (5) we obtain

$$\begin{aligned} f\left(\frac{j+1}{n}\right) - f\left(\frac{2j+1}{2n}\right) &\leq \frac{\varepsilon}{n}, \\ f\left(\frac{2j+1}{2n}\right) - f\left(\frac{4j+1}{4n}\right) &\leq \frac{\varepsilon}{2n}, (j := 2j, n := 2n) \\ f\left(\frac{4j+1}{4n}\right) - f\left(\frac{8j+1}{8n}\right) &\leq \frac{\varepsilon}{4n}, (j := 4j, n := 4n) \\ &\vdots \\ f\left(\frac{2^k j + 1}{2^k n}\right) - f\left(\frac{2^{k+1} j + 1}{2^{k+1} n}\right) &\leq \frac{\varepsilon}{2^k n}, (j := 2^k j, n := 2^k n). \end{aligned}$$

Taking, respectively, the sums of all the terms in the first and second parts of the previous inequalities, we obtain

$$f\left(\frac{j+1}{n}\right) - f\left(\frac{2^{k+1} j + 1}{2^{k+1} n}\right) \leq \frac{\varepsilon}{n} \cdot \left(1 + \frac{1}{2} + \dots + \frac{1}{2^k}\right) \leq \frac{2\varepsilon}{n}$$

Since

$$\lim_{k \rightarrow \infty} f\left(\frac{2^{k+1} j + 1}{2^{k+1} n}\right) = f\left(\frac{j}{n}\right),$$

it follows that

$$f\left(\frac{j+1}{n}\right) - f\left(\frac{j}{n}\right) \leq \frac{2\varepsilon}{n}$$

and since  $f$  is non-decreasing, we get

$$\left|f\left(\frac{j+1}{n}\right) - f\left(\frac{j}{n}\right)\right| \leq \frac{2\varepsilon}{n}.$$

By Lemma 4 it follows that  $f$  is constant in  $[a, b]$ . Since  $a$  and  $b$  are two arbitrary inner points of  $I$  and  $f$  is continuous, it easily results that  $f$  is constant in  $I$ .  $\square$

Note that Lemma 4 and Lemma 5 can also be extrapolated from the monograph [3]. Now we can prove the main theorem of this section.

**Theorem 1.** Let  $f \in C_+([0, 1])$  be a non-decreasing function such that

$$\left\| K_n^{(M)}(f, \cdot) - f(\cdot) \right\|_\infty = o(n^{-1})$$

as  $n \rightarrow \infty$ . Then  $f$  is constant over  $[0, 1]$ .

**Proof.** Let us choose arbitrary  $a, b \in (0, 1)$ ,  $a < b$ . Further, let  $n \in \mathbb{N}^+$  be sufficiently large such that  $\frac{1}{n} < b - a$  and  $b < 1 - \frac{1}{2n}$ . Now, we fix  $j \in \{0, 1, \dots, n - 2\}$ , such that  $\frac{j+1}{n}, \frac{j}{n} + \frac{1}{2n} \in [a, b]$ . We have

$$K_n^{(M)}\left(f, \frac{j}{n} + \frac{1}{2n}\right) = K_n^{(M)}\left(f, \frac{2j+1}{2n}\right) = \frac{\sum_{k=0}^{n-1} \phi_\sigma\left(n \cdot \frac{2j+1}{2n} - k\right) n^{\frac{(k+1)}{n}} \int_{k/n} f(u) du}{\sum_{k=0}^{n-1} \phi_\sigma\left(n \cdot \frac{2j+1}{2n} - k\right)}.$$

By Lemma 3, it follows that

$$\begin{aligned} \sum_{k=0}^{n-1} \phi_\sigma\left(n \cdot \frac{2j+1}{2n} - k\right) &= \phi_\sigma\left(n \cdot \frac{2j+1}{2n} - j\right) \\ &= \phi_\sigma\left(n \cdot \frac{2j+1}{2n} - (j+1)\right) \end{aligned}$$

which easily implies that

$$K_n^{(M)}\left(f, \frac{2j+1}{2n}\right) \geq n \int_{(j+1)/n}^{(j+2)/n} f(u) du.$$

Moreover, recalling that  $f$  is non-decreasing, it follows that  $K_n^{(M)}\left(f, \frac{2j+1}{2n}\right) \geq f\left(\frac{j+1}{n}\right)$ , and this implies

$$0 \leq f\left(\frac{j+1}{n}\right) - f\left(\frac{2j+1}{2n}\right) \leq \left\| K_n^{(M)}(f, \cdot) - f \right\|_\infty = o(n^{-1}), \quad \text{as } n \rightarrow +\infty.$$

Then, we can prove that for every  $\varepsilon > 0$ , there exists  $n(\varepsilon) \in \mathbb{N}^+$  such that for any  $n \in \mathbb{N}^+$ , with  $n \geq n(\varepsilon)$  and  $j \in \mathbb{Z}$ , such that  $\frac{j+1}{n}, \frac{2j+1}{2n} \in [a, b]$ , we have

$$f\left(\frac{j+1}{n}\right) - f\left(\frac{2j+1}{2n}\right) \leq \frac{\varepsilon}{n},$$

and hence the proof follows by Lemma 5.  $\square$

**Remark 1.** The previous theorem can be easily generalized in the case of functions defined on arbitrary intervals  $[a, b]$ ,  $a, b \in \mathbb{R}$ , instead of  $[0, 1]$ . This is possible by defining the operators  $K_n^{(M)}$  on generic intervals  $[a, b]$  (as made in [9]) and then working with continuous and non-decreasing  $f : [a, b] \rightarrow \mathbb{R}^+$ .

#### 4. Local Inverse Result

The main aim of this section is to prove an inverse theorem of approximation. We will use a strategy similar to that presented in the previous section.



**Lemma 6.** Let  $I \subseteq \mathbb{R}$  be a bounded or unbounded interval and  $f \in C(I)$ . Suppose in addition that there exists an absolute positive constant  $C$  such that for any sufficiently large  $n \in \mathbb{N}$  and every  $j \in \mathbb{Z}$ , with  $\frac{j}{n}, \frac{j+1}{n} \in I$ , we have

$$\left| f\left(\frac{j+1}{n}\right) - f\left(\frac{j}{n}\right) \right| \leq \frac{C}{n}. \tag{6}$$

Then  $f \in Lip(I)$ .

**Proof.** Let us choose arbitrary  $x_0, y_0 \in I, x_0 < y_0$ . Moreover, let  $\varepsilon > 0$  with  $\varepsilon < y_0 - x_0$ . The continuity of  $f$  implies the existence of  $n_0(\varepsilon) \in \mathbb{N}^+$ , such that for any  $x, y \in I, |x - x_0| \leq \frac{1}{n_0(\varepsilon)}, |y - y_0| \leq \frac{1}{n_0(\varepsilon)}$ , we have

$$|f(x) - f(x_0)| \leq \varepsilon, |f(y) - f(y_0)| \leq \varepsilon.$$

We now fix  $n_1 = \max\left\{n_0(\varepsilon), \frac{1}{y_0 - x_0}\right\}$ , and let us choose arbitrary  $n \in \mathbb{N}$  with  $n \geq n_1$ . Proceeding as in the proof in Lemma 4, we get

$$|f(x_0) - f(y_0)| \leq \frac{1C}{n} + 2\varepsilon.$$

and since  $\varepsilon < y_0 - x_0$ , we get

$$|f(x_0) - f(y_0)| \leq (C + 2)(y_0 - x_0).$$

Then the thesis follows by the arbitrariness of  $x_0$  and  $y_0$ .  $\square$

**Lemma 7.** Let  $I \subseteq \mathbb{R}$  be a bounded or unbounded interval and  $f \in C(I)$  be a non-decreasing function with the property that, for any couple  $a, b, a < b$ , of inner points of  $I$ , there exists a  $C > 0$  such that for every sufficiently large  $n \in \mathbb{N}$  and  $j \in \mathbb{Z}$ , with  $\frac{j+1}{n}, \frac{2j+1}{2n} \in [a, b]$ , we have

$$f\left(\frac{j+1}{n}\right) - f\left(\frac{2j+1}{2n}\right) \leq \frac{C}{n}. \tag{7}$$

Then  $f \in Lip([a, b])$ .

**Proof.** Let  $n \in \mathbb{N}$  be sufficiently large such that  $\frac{1}{n} \leq b - a$ , and let us choose arbitrary  $j \in \mathbb{Z}$  such that  $a \leq \frac{j}{n} \leq \frac{j+1}{n} \leq b$ . Arguing as in Lemma 5, we get

$$f\left(\frac{j+1}{n}\right) - f\left(\frac{2^{k+1}j+1}{2^{k+1}n}\right) \leq \frac{C}{n} \cdot \left(1 + \frac{1}{2} + \dots + \frac{1}{2^k}\right) \leq \frac{2C}{n}, k \in \mathbb{N}^+.$$

Taking the limit as  $k \rightarrow +\infty$ , we obtain

$$f\left(\frac{j+1}{n}\right) - f\left(\frac{j}{n}\right) \leq \frac{2C}{n}$$

and, since  $f$  is non-decreasing,

$$\left| f\left(\frac{j+1}{n}\right) - f\left(\frac{j}{n}\right) \right| \leq \frac{2C}{n}.$$

Hence, by Lemma 6, it follows that  $f \in Lip([a, b])$ .  $\square$

For results similar to that ones of Lemma 6 and Lemma 7 (only in the case of bounded intervals), one can see, e.g., [3].

Now we can finally prove a (local) inverse theorem of approximation.

**Theorem 2.** Let  $f \in C_+([0, 1])$  be a non-decreasing function such that

$$\left\| K_n^{(M)}(f, \cdot) - f(\cdot) \right\|_\infty = O(n^{-1})$$

as  $n \rightarrow \infty$ . Then, for every  $[a, b] \subset (0, 1)$ , it turns out that  $f \in \text{Lip}([a, b])$ .

**Proof.** As in Theorem 1, we choose arbitrary  $a, b \in (0, 1)$ ,  $a < b$ . Let now  $n \in \mathbb{N}^+$  be sufficiently large such that  $\frac{1}{n} < b - a$  and  $b < 1 - \frac{1}{2n}$ . Then, let  $j \in \{0, 1, \dots, n - 2\}$ , such that  $\frac{j+1}{n}, \frac{2j+1}{2n} \in [a, b]$ . Proceeding as in the proof of the above-mentioned theorem, we immediately get

$$0 \leq f\left(\frac{j+1}{n}\right) - f\left(\frac{2j+1}{2n}\right) \leq \left\| K_n^{(M)}(f, \cdot) - f(\cdot) \right\|_\infty = O(n^{-1}), \quad \text{as } n \rightarrow +\infty,$$

and, by Lemma 7, this implies that  $f \in \text{Lip}([a, b])$ .  $\square$

**Author Contributions:** Conceptualization, formal analysis, writing—review and editing M.C.; Conceptualization, formal analysis, writing—review and editing L.C.; Conceptualization, formal analysis, writing—review and editing D.C.; Conceptualization, formal analysis, writing—review and editing S.G.G.; Conceptualization, formal analysis, writing—review and editing G.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** D. Costarelli has been partially supported within the 2020 GNAMPA-INdAM Project “Analisi reale, teoria della misura ed approssimazione per la ricostruzione di immagini”, and G. Vinti has been partially supported by the projects: (1) Ricerca di Base 2017 dell’Università degli Studi di Perugia—“Metodi di teoria degli operatori e di Analisi Reale per problemi di approssimazione ed applicazioni”, (2) Ricerca di Base 2018 dell’Università degli Studi di Perugia—“Metodi di Teoria dell’Approssimazione, Analisi Reale, Analisi Nonlineare e loro Applicazioni”, (3) “Metodi e processi innovativi per lo sviluppo di una banca di immagini mediche per fini diagnostici” funded by the Fondazione Cassa di Risparmio di Perugia (FCRP), 2018, (4) “Metodiche di Imaging non invasivo mediante angiografia OCT sequenziale per lo studio delle Retinopatie degenerative dell’Anziano (M.I.R.A.)”, funded by FCRP, 2019.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The study did not report any data.

**Acknowledgments:** The authors M. Cantarini, D. Costarelli and G. Vinti are members of the Gruppo Nazionale per l’Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM), of the network RITA (Research ITALian network on Approximation), and of the UMI group “Teoria dell’Approssimazione e Applicazioni (T.A.A.)”.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

ReLU Rectified Linear Unit

## References

1. Coroianu, L.; Gal, S.G. Approximation by nonlinear generalized sampling operators of max-product kind. *Sampl. Theory Signal Image Process.* **2010**, *9*, 59–75. [[CrossRef](#)]
2. Coroianu, L.; Gal, S.G. Approximation by max-product sampling operators based on sinc-type kernels. *Sampl. Theory Signal Image Process.* **2011**, *10*, 211–230. [[CrossRef](#)]
3. Bede, B.; Coroianu, L.; Gal, S.G. *Approximation by Max-Product Type Operators*; Springer: Berlin, Germany, 2016. [[CrossRef](#)]

4. Güngör, S.Y.; Ispir, N. Approximation by Bernstein-Chlodowsky operators of max-product kind. *Math. Communic* **2018**, *23*, 205–225.
5. Holhos, A. Weighted Approximation of functions by Meyer-König and Zeller operators of max-product type. *Numer.Funct. Anal. Optim.* **2018**, *39*, 689–703. [[CrossRef](#)]
6. Holhos, A. Weighted approximation of functions by Favard operators of max-product type. *Period. Math. Hungar.* **2018**, *77*, 340–346. [[CrossRef](#)]
7. Gokcer, T.Y.; Duman, O. Approximation by max-min operators: A general theory and its applications. *Fuzzy Sets Syst.* **2020**, *394*, 146–161. [[CrossRef](#)]
8. Gokcer, T.Y.; Duman, O. Regular summability methods in the approximation by max-min operators. *Fuzzy Sets Syst.* **2022**, *426*, 106–120. [[CrossRef](#)]
9. Costarelli, D.; Vinti, G. Approximation by max-product neural network operators of Kantorovich type. *Results Math.* **2016**, *69*, 505–519. [[CrossRef](#)]
10. Cardaliaguet, P.; Euvrard, G. Approximation of a function and its derivative with a neural network. *Neural Netw.* **1992**, *5*, 207–220. [[CrossRef](#)]
11. Cao, F.; Chen, Z. The approximation operators with sigmoidal functions. *Comput. Math. Appl.* **2009**, *58*, 758–765.
12. Cao, F.; Chen, Z. The construction and approximation of a class of neural networks operators with ramp functions. *J. Comput. Anal. Appl.* **2012**, *14*, 101–112.
13. Cao, F.; Chen, Z. Scattered data approximation by neural networks operators. *Neurocomputing* **2016**, *190*, 237–242.
14. Dai, H.; Xie, J.; Chen, W. Event-Triggered Distributed Cooperative Learning Algorithms over Networks via Wavelet Approximation. *Neural Process. Lett.* **2019**, *50*, 669–700. 10.1007-s11063-019-10031-x. [[CrossRef](#)]
15. Ismailov, V.E. On the approximation by neural networks with bounded number of neurons in hidden layers. *J.Math. Anal. Appl.* **2014**, *417*, 963–969. [[CrossRef](#)]
16. Cao, F.; Liu, B.; Park, D.S. Image classification based on effective extreme learning machine. *Neurocomputing* **2013**, *102*, 90–97. [[CrossRef](#)]
17. Agostinelli, F.; Hoffman, M.; Sadowski, P.; Baldi, P. Learning Activation Functions to Improve Deep Neural Networks. *arXiv* **2015**, arXiv:1412.6830v3.
18. Iliev, A.; Kyurkchiev, N.; Markov, S. On the approximation of the cut and step functions by logistic and Gompertz functions. *Biomath* **2015**, *4*, 1510101. [[CrossRef](#)]
19. Yarotsky, D. Error bounds for approximations with deep ReLU networks. *Neural Netw.* **2017**, *94*, 103–114. [[CrossRef](#)]
20. Bajpeyi, S.; Sathish Kumar, A. Approximation by exponential sampling type neural network operators. *Anal. Math. Phys.* **2021**, *11*, 108. [[CrossRef](#)]
21. Cantarini, M.; Costarelli, D.; Vinti, G. Asymptotic expansions for the neural network operators of the Kantorovich type and high order of approximation. *Mediterr. J. Math.* **2021**, *18*, 66. [[CrossRef](#)]
22. Costarelli, D.; Sambucini, A.R. Approximation results in Orlicz spaces for sequences of Kantorovich max-product neural network operators. *Results Math.* **2018**, *73*, 15. [[CrossRef](#)]
23. Cucker, F.; Zhou, D.X. *Learning Theory An Approximation Theory Viewpoint*; Cambridge University Press: Cambridge, UK, 2007.
24. Kadak, U. Fractional type multivariate neural network operators. *Math. Meth. Appl. Sci.* **2021**. [[CrossRef](#)]
25. Coroianu, L.; Gal, S.G. Saturation results for the truncated max-product sampling operators based on sinc and Fejér-type kernels. *Sampl. Theory Signal Image Process.* **2012**, *11*, 113–132. [[CrossRef](#)]
26. Coroianu, L.; Gal, S.G. Saturation and inverse results for the Bernstein max-product operator. *Period. Math.Hung.* **2014**, *69*, 126–133. [[CrossRef](#)]
27. Costarelli, D.; Vinti, G. Saturation classes for max-product neural network operators activated by sigmoidal functions. *Results Math.* **2017**, *72*, 1555–1569. [[CrossRef](#)]
28. Ivanov, K. On a new characteristic of functions. II. Direct and converse theorems for the best algebraic approximation in  $C[-1, 1]$  and  $L^p[-1, 1]$ . *Pliska* **1983**, *5*, 151–163.
29. Costarelli, D.; Vinti, G. Convergence results for a family of Kantorovich max-product neural network operators in a multivariate setting. *Math. Slovaca* **2017**, *67*, 1469–1480. [[CrossRef](#)]
30. Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems* **1989**, *2*, 303–314. [[CrossRef](#)]
31. Costarelli, D.; Vinti, G. Convergence for a family of neural network operators in Orlicz spaces. *Math. Nachrichten* **2017**, *290*, 226–235. [[CrossRef](#)]
32. Goebbels, S. On sharpness of error bounds for univariate single hidden layer feedforward neural networks. *Results Math.* **2020**, *75*, 109. [[CrossRef](#)]
33. Li, Y.; Yuan, Y. Convergence Analysis of Two-layer Neural Networks with ReLU Activation. *arXiv* **2017**, arXiv:1705.09886. Available online: <https://arxiv.org/abs/1705.09886> (accessed on 1 December 2021).
34. Zhang, C.; Woodland, P.C. DNN speaker adaptation using parameterised sigmoid and ReLU hidden activation functions. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016.
35. Agarap, A.F. Deep Learning using Rectified Linear Units (ReLU). *arXiv* **2018**, arXiv:1803.08375.
36. DeVore, R.A.; Lorentz, G.G. *Constructive Approximation*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1992.