A survey on approximation by means of neural network operators

Danilo Costarelli[1], **Renato Spigler**[2], **Gianluca Vinti**[3]

(1,3) Department of Mathematics and Computer Science,
University of Perugia,
1, Via Vanvitelli, 06123 Perugia, Italy

(2) Department of Mathematics and Physics,
Roma Tre University,
1, Largo San L. Murialdo, 00146 Roma, Italy

danilo.costarelli@unipg.it - spigler@mat.uniroma3.it
gianluca.vinti@unipg.it

Abstract

In the present survey, we recall the main convergence results concerning the theory of neural network (NN) operators. Pointwise and uniform approximation results have been proved for the classical (linear) NN operators, as well as, for their corresponding max-product (nonlinear) version, when continuous functions defined on bounded domains are approximated. In order to approximate also not necessarily continuous functions, a Kantorovich-type version of the above NN operators has been studied in an L^p-setting. Finally, several examples of sigmoidal activation functions for the aforementioned operators have been provided.

AMS 2010 Mathematics Subject Classification: 41A25, 41A05, 41A30, 47A58

Key words and phrases: sigmoidal functions; neural networks operators; uniform approximation; max-product operators; Kantorovich type operators; L^p approximation.

1 Introduction

In the present survey paper, we describe in detail a new approach to study neural network (NN) approximation, by exploiting the Operator Theory, in particular the study of the NN operators. The theory of NN operators has been introduced in last years as a constructive approach for approximating functions by a neural process. The latter topic, other than the "classical" theory of artificial NNs, is strictly related to the Approximation Theory. Indeed, many papers have been written concerning applications of NNs to this subject, see, e.g., [43, 63, 64, 13, 46, 52, 45, 18, 50, 47, 56]. The first author who studied approximations by NNs activated by sigmoidal functions was G. Cybenko [43], who exploited the Hahn-Banach theorem to prove a density result by means of the aforementioned NNs. Estimates for the order of approximation, in various setting have been proved in [11, 48, 58, 59]. Moreover, a subject which was deeply studied was the "best approximation problem" by NNs in suitable functions spaces, see, e.g., [44, 53, 54, 55]. The latter study is based on techniques from Functional Analysis.

For any bounded function $f : \mathcal{R} \to \mathbb{R}$, where $\mathcal{R} := [a_1, b_1] \times ... \times [a_s, b_s] \subset \mathbb{R}^s$, the classical definition of an NN operator is given by:

$$F_n(f, \underline{x}) \; := \; \frac{\displaystyle\sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor} \cdots \sum_{k_s=\lceil na_s \rceil}^{\lfloor nb_s \rfloor} f\left(\underline{k}/n\right) \Psi_\sigma(n\underline{x} - \underline{k})}{\displaystyle\sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor} \cdots \sum_{k_s=\lceil na_s \rceil}^{\lfloor nb_s \rfloor} \Psi_\sigma(n\underline{x} - \underline{k})}, \qquad \text{(I)}$$

for every $\underline{x} \in \mathcal{R}$, and $\underline{k} = (k_1, k_2, \ldots, k_s)$. Here, $\lfloor \cdot \rfloor$ stands for the "integral part" of a given real number, while $\lceil \cdot \rceil$ stands for the "ceiling", while $\Psi_\sigma(\underline{x})$ is a multivariate density function, generated by the product of s univariate functions $\phi_\sigma(x)$, where each of them is constructed by using a suitable finite linear combination of *sigmoidal* functions σ.

Originally, the main difficulty arising to study approximations by NNs was to be able to construct concretely the NNs which approximate a given function, f, defined on a fixed bounded set of $\mathbb{R}^s$. Such a problem was solved, e.g., for functions of one variable by several authors (see, e.g., [62, 20, 65]), who provided constructive proofs, for instance, in the space of continuous functions. Many more difficulties arise when dealing with functions of several variables. Some results have been obtained by (quite non standard) approaches based on convolutions, and resorting to the theory of "ridge functions", see, e.g., [21].

A solution can be proposed for the theoretical problems described above, by an approach based on operators of the form (I). Indeed, in the classical NN operators, the coefficients, the weights and the thresholds needed in order to generate a NN which approximates a function f are known in their analytical form. For this reason, we can say that the approximation algorithm provided by (I) can be called "constructive".

The theory mentioned above, was introduced by Anastassiou in [1], and was inspired by a paper of Cardaliaguet and Euvrard [19]. The latter topic was widely investigated in a number of articles, recently collected in the monograph [2]. The approximation results mainly proved in [2], for functions of one as well as several variables, involve NN operators activated by logistic and hyperbolic tangent sigmoidal functions, i.e., $\Psi_\sigma(\underline{x})$ is generated by $\sigma_\ell(x) := (1+e^{-x})^{-1}$, and $\sigma_h(x) := (1/2)(\tanh x + 1)$, respectively. The results proved in [2] have been successfully improved in the papers [30, 31], for what concerns the *order* of approximation as well as the *kind* of activation functions adopted for the NN operators. More precisely, these results have been extended in order to include a larger class of sigmoidal activation functions. Some convergence results in the specific cases of NN operators activated by σ_ℓ and by σ_h were also obtained in [14, 15].

The results described above, are proved only in the case of continuous functions, which is the most natural setting for the classical operators in (I), due to their *pointwise* nature. In [32], a Kantorovich version of the previous operators has been introduced and studied, in order to approximate by NN-type operators also functions that are not necessarily continuous. Such operators are of the form

$$K_n(f,\underline{x}) \ := \ \frac{\displaystyle\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor-1} \cdots \sum_{k_s=\lceil na_s\rceil}^{\lfloor nb_s\rfloor-1} \left[n^s \int_{R^n_{\underline{k}}} f\,(\underline{u})\ d\underline{u} \right] \Psi_\sigma(n\underline{x}-\underline{k})}{\displaystyle\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor-1} \cdots \sum_{k_s=\lceil na_s\rceil}^{\lfloor nb_s\rfloor-1} \Psi_\sigma(n\underline{x}-\underline{k})}, \qquad \text{(II)}$$

for every $\underline{x} \in \mathcal{R}$. Here, the symbols $R^n_{\underline{k}}$ denote the sets

$$R^n_{\underline{k}} \ := \ \left[\frac{k_1}{n}, \frac{k_1+1}{n}\right] \times \cdots \times \left[\frac{k_s}{n}, \frac{k_s+1}{n}\right] \ \subset \ \mathbb{R}^s, \qquad \text{(III)}$$

in which we compute the mean values $n^s \int_{R^n_{\underline{k}}} f\,(\underline{u})\ \underline{u}$ of a given locally integrable function f. For the operators in (II), convergence results in $L^p(\mathcal{R})$, $1 \leq p < +\infty$, have been obtained, other than pointwise and uniform approximation results as well. Note the essential feature that in (II) *averages* of f replace pointwise values.

Another useful approach for the study of NN operators, has been introduced in [3] as further extension of [19], by using the *max-product* approach. The max-product approach was first introduced by Coroianu and Gal in some papers (see, e.g., [24, 25, 26]), and consists essentially in replacing the symbol $\sum$ (for operators defined by finite sum or series) with the sup-operator $\bigvee$. This simple modification transforms *linear* into *nonlinear* operators, thus allowing to improve the order of approximation that can be achieved when continuous functions are approximated. In [3], the authors considered operators of the NN-type, with centered bell-shaped density functions having compact support, and which satisfy some other additional conditions. In this paper we recall the results proved in [40] (see also [39] for the theory in a one-dimensional setting), where the max-product NN operators, $F_n^{(M)}$, are introduced as an extension of the operators in (I), and we extend the results of [3] avoiding the assumption of compact support on the bell-shaped density functions ϕ_σ and Ψ_σ. The *max-product NN operators* are of the form

$$F_n^{(M)}(f, \underline{x}) \ := \ \frac{\bigvee_{\underline{k} \in \mathcal{J}_n} f\,(\underline{k}/n)\, \Psi_\sigma(n\underline{x} - \underline{k})}{\bigvee_{\underline{k} \in \mathcal{J}_n} \Psi_\sigma(n\underline{x} - \underline{k})}, \qquad \underline{x} \in \mathcal{R}, \qquad \text{(IV)}$$

where $\mathcal{J}_n$ denotes the set of indexes $\underline{k} \in \mathbb{Z}^s$, such that $k_i = \lceil na_i \rceil$, $\lceil na_i \rceil + 1$, $\lceil na_i \rceil + 2$, ..., $\lfloor nb_i \rfloor$, $i = 1, ..., s$, for every fixed $n \in \mathbb{N}^+$ sufficiently large, where f is a bounded real-valued function. By the family of the nonlinear operators in (IV) it is possible to approximate (pointwise and uniformly) functions more efficiently than its corresponding linear counterpart in (I).

2 Preliminary results and notations

In this section, we recall some preliminary notion and results that will be useful in the following. By the symbol $C(\mathcal{R})$, we will denote the space of all continuous functions $f : \mathcal{R} \to \mathbb{R}$, equipped with the usual sup-norm $\| \cdot \|_\infty$. Here, $\mathcal{R} \subset \mathbb{R}^s$, denotes the multivariate set $\mathcal{R} := [a_1, b_1] \times ... \times [a_s, b_s]$. Moreover, by $C_+(\mathcal{R})$, we will denote the subspace of $C(\mathcal{R})$ containing all the nonnegative functions. Finally, by $L^p(\mathcal{R})$, $1 \leq p < +\infty$, we will define the usual L^p-space equipped with the usual $\| \cdot \|_p$ norm.

Now, we recall the following

Definition 2.1. *A measurable function $\sigma : \mathbb{R} \to \mathbb{R}$ is called a sigmoidal function, if* $\lim_{x \to -\infty} \sigma(x) = 0$ *and* $\lim_{x \to +\infty} \sigma(x) = 1$.

From now on, we consider non-decreasing sigmoidal functions which satisfy all the following assumptions:

(Σ1) $\sigma(x) - 1/2$ is an odd function;

(Σ2) $\sigma \in C^2(\mathbb{R})$, and is concave for $x \geq 0$;

(Σ3) $\sigma(x) = \mathcal{O}(|x|^{-1-\alpha})$ as $x \to -\infty$, for some $\alpha > 0$.

Now, we define the following univariate and multivariate density functions:

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

and,

$$\Psi_\sigma(\underline{x}) \ := \ \phi_\sigma(x_1) \cdot \phi_\sigma(x_2) \cdot ... \cdot \phi_\sigma(x_s), \qquad \underline{x} := (x_1, ..., x_s) \in \mathbb{R}^s. \qquad (2)$$

In the following lemma, we summarize a number of important properties enjoyed by $\phi_\sigma(x)$ and $\Psi_\sigma(\underline{x})$, established in [30, 31] and used below.

Lemma 2.2. *(i)* $\phi_\sigma(x) \geq 0$ *for every* $x \in \mathbb{R}$ *and* $\lim_{x\to\pm\infty} \phi_\sigma(x) = 0$*;*

(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) $\phi_\sigma(x) = \mathcal{O}(|x|^{-1-\alpha})$*, as* $x \to \pm\infty$*, and* $\Phi_\sigma(\underline{x}) = \mathcal{O}(\|\underline{x}\|_2^{-1-\alpha})$*, as* $\|x\|_2 \to +\infty$*, where by* $\|\cdot\|_2$ *denotes the usual Euclidean norm of* $\mathbb{R}^s$*,* $\|\underline{x}\|_2 := (x_1^2 + ... + x_s^2)^{1/2}$*, with* $\underline{x} \in \mathbb{R}^s$*;*

(v) for every $\underline{x} \in \mathbb{R}^s$*,* $\sum_{\underline{k}\in\mathbb{Z}^s} \Psi_\sigma(\underline{x} - \underline{k}) = 1$*, and* $\Psi_\sigma \in L^1(\mathbb{R}^s)$*;*

(vi) the series $\sum_{\underline{k}\in\mathbb{Z}^s} \Psi_\sigma(\underline{x} - \underline{k})$ *converges uniformly on compact subsets of* $\mathbb{R}^s$*;*

(vii) for every $\gamma > 0$*, we have*

$$\lim_{n\to+\infty} \sum_{\|\underline{x}-\underline{k}\|_2 > \gamma n} \Psi_\sigma(\underline{x} - \underline{k}) = 0,$$

uniformly with respect to $\underline{x} \in \mathbb{R}^s$*.*

Remark 2.3. (a) The function $\phi_\sigma(x)$ is a "centered bell shaped function", according to the definition given in [19] by Cardaliaguet and Euvrard.

(b) From condition *(iii)* of Lemma 2.2, and the assumptions made on σ, it is easy to deduce that $\phi_\sigma(0) \geq \phi_\sigma(x)$ for every $x \in \mathbb{R}$, and $\phi_\sigma(0) \leq 1/2$. Consequently, we have $\Psi_\sigma(\underline{x}) \leq \Psi_\sigma(0) \leq 2^{-s}$, for every $\underline{x} \in \mathbb{R}^s$.

(c) Note that, condition ($\Sigma 2$) can be weakened by assuming that σ satisfies (iii) of Lemma 2.2, i.e., requiring that the corresponding ϕ_σ satisfies the following condition:

$$\phi_\sigma(x) \text{ is non-decreasing for } x < 0 \text{ and non-increasing for } x \geq 0. \qquad (3)$$

In this case, the present theory still holds and the main advantage that can be achieved is that, we can apply it to *non-smooth* sigmoidal functions as well. For more details, see [30, 31].

3 The classical neural network operators

In this section, we recall definition and main results concerning the classical NN operators.

Recall first that $\lfloor\cdot\rfloor$ stands for the "integral part" of a given real number x, i.e.,

$$\lfloor x \rfloor \ := \ \max\{k : \text{ such that } k \in \mathbb{Z}, \text{ and } k \leq x\},$$

while $\lceil\cdot\rceil$ stands for the "ceiling", i.e.,

$$\lceil x \rceil \ := \ \min\{k : \text{ such that } k \in \mathbb{Z}, \text{ and } k \geq x\}.$$

Here we consider sigmoidal activation functions, σ, which satisfy the assumptions made in the previous section, together with the technical condition

$$\sigma(2) \ > \ \sigma(0). \tag{4}$$

The following inequality,

$$\sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor} \cdots \sum_{k_s=\lceil na_s \rceil}^{\lfloor nb_s \rfloor} \Psi_\sigma(n\underline{x} - \underline{k})$$

$$= \prod_{i=1}^{s} \sum_{k_i=\lceil na_i \rceil}^{\lfloor nb_i \rfloor} \phi_\sigma(nx_i - k_i) \ \geq \ [\phi_\sigma(1)]^s \ > \ 0, \tag{5}$$

can be established for every $\underline{x} \in \mathcal{R}$, where $\underline{k} := (k_1, ..., k_s) \in \mathbb{Z}^s$, see [31].

Definition 3.1. *Let $f : \mathcal{R} \to \mathbb{R}$ be a bounded function, and $n \in \mathbb{N}^+$ such that $\lceil na_i \rceil \leq \lfloor nb_i \rfloor$ for every $i = 1, ..., s$. The linear multivariate NN operators $F_n(f, \underline{x})$, activated by the sigmoidal function σ, and acting on f, are defined by*

$$F_n(f, \underline{x}) \ := \ \frac{\displaystyle\sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor} \cdots \sum_{k_s=\lceil na_s \rceil}^{\lfloor nb_s \rfloor} f\left(\underline{k}/n\right) \Psi_\sigma(n\underline{x} - \underline{k})}{\displaystyle\sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor} \cdots \sum_{k_s=\lceil na_s \rceil}^{\lfloor nb_s \rfloor} \Psi_\sigma(n\underline{x} - \underline{k})},$$

for every $\underline{x} \in \mathcal{R}$ and $\underline{k}/n := (k_1/n, ..., k_s/n)$.

For $n \in \mathbb{N}^+$ sufficiently large, we always obtain $\lceil na_i \rceil \leq \lfloor nb_i \rfloor$, $i = 1, ..., s$. Moreover, $a_i \leq \frac{k_i}{n} \leq b_i$ if and only if $\lceil na_i \rceil \leq k_i \leq \lfloor nb_i \rfloor$, and since f is bounded, $F_n(f, \underline{x})$ turns out to be well defined for all $\underline{x} \in \mathcal{R}$. Note that $F_n(1, \underline{x}) = 1$, for every n sufficiently large. We now prove the following pointwise and uniform convergence results for the classical NN operators [31].

Theorem 3.2. *Let $f : \mathcal{R} \to \mathbb{R}$ be bounded. Then,*

$$\lim_{n \to +\infty} F_n(f, \underline{x}) \ = \ f(\underline{x})$$

at each point $\underline{x} \in \mathcal{R}$ of continuity of f. Moreover, if $f \in C(\mathcal{R})$, then

$$\lim_{n \to +\infty} \sup_{\underline{x} \in \mathcal{R}} |F_n(f, \underline{x}) - f(\underline{x})| \ = \ \lim_{n \to +\infty} \|F_n(f, \cdot) - f(\cdot)\|_\infty \ = \ 0.$$

Proof. Let $\underline{x} \in \mathcal{R}$ be a point where f is continuous. By means of (5) we can write

$$|F_n(f,\underline{x}) - f(\underline{x})| = \frac{\left| \sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor} \dots \sum_{k_s=\lceil na_s \rceil}^{\lfloor nb_s \rfloor} [f(\underline{k}/n) - f(\underline{x})] \ \Psi_\sigma(n\underline{x} - \underline{k}) \right|}{\sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor} \dots \sum_{k_s=\lceil na_s \rceil}^{\lfloor nb_s \rfloor} \Psi_\sigma(n\underline{x} - \underline{k})}$$

$$\leq \frac{1}{[\phi_\sigma(1)]^s} \sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor} \dots \sum_{k_s=\lceil na_s \rceil}^{\lfloor nb_s \rfloor} |f(\underline{k}/n) - f(\underline{x})| \, \Psi_\sigma(n\underline{x} - \underline{k}),$$

for every $n \in \mathbb{N}^+$ sufficiently large. Let now $\varepsilon > 0$ be fixed. From the continuity of f at $\underline{x}$, there exists $\gamma > 0$ such that $|f(\underline{y}) - f(\underline{x})| < \varepsilon$ for every $\underline{y} \in \mathcal{R}$ with $\|\underline{y} - \underline{x}\|_2 < \gamma$. Hence

$$|F_n(f,\underline{x}) - f(\underline{x})| \leq \frac{1}{[\phi_\sigma(1)]^s} \left\{ \sum_{\|n\underline{x}-\underline{k}\|_2 \leq n\gamma} |f(\underline{k}/n) - f(\underline{x})| \, \Psi_\sigma(n\underline{x} - \underline{k}) \right.$$

$$\left. + \sum_{\|n\underline{x}-\underline{k}\|_2 > n\gamma} |f(\underline{k}/n) - f(\underline{x})| \, \Psi_\sigma(n\underline{x} - \underline{k}) \right\} =: \frac{1}{[\phi_\sigma(1)]^s} (I_1 + I_2) \, .$$

Let first estimate I_1. Using Lemma 2.2 (v), the continuity of f at $\underline{x}$, and that $\|n\underline{x} - \underline{k}\|_2 \leq n\gamma$, i.e., $\|\underline{k}/n - \underline{x}\|_2 \leq \gamma$, we obtain

$$I_1 \ < \ \varepsilon \sum_{\|n\underline{x}-\underline{k}\|_2 \leq n\gamma} \Psi_\sigma(n\underline{x} - \underline{k}) \ \leq \ \varepsilon.$$

Furthermore, by the boundedness of f and Lemma 2.2 (vii), we have, for n sufficiently large,

$$I_2 \ \leq \ 2\,\|f\|_\infty \sum_{\|n\underline{x}-\underline{k}\|_2 > n\gamma} \Psi_\sigma(n\underline{x} - \underline{k}) \ < \ 2\,\|f\|_\infty \, \varepsilon,$$

uniformly with respect to $\underline{x} \in \mathbb{R}^s$. The proof of the first part of the theorem then follows by the arbitrariness of ε. When $f \in C(\mathcal{R})$, the second part of the theorem follows similarly, replacing $\gamma > 0$ with the parameter of the uniform continuity of f in $\mathcal{R}$. □

4 The max-product neural network operators

In this section, we recall definition and main results concerning the max-product NN operators. The max-product NN operators represent neuro-processing models in which the global behavior of the network is mainly determined by one of the artificial neuron of the network.

Also here, we consider sigmoidal activation functions, σ, which satisfy the technical condition $\sigma(2) > \sigma(0)$, introduced in Section 3.

In order to study the aforementioned family of nonlinear operators, we first introduce some notation. We define

$$\bigvee_{k\in J} A_k \ := \ \sup\{A_k : \ k \in J\},$$

where J is any set of indices. Clearly, if the cardinality of J is finite, the sup in the previous definition reduces to the maximum value. In the next lemma, we will show that, by using the max-product operator $\bigvee$ in place of the symbol $\sum$, properties similar to those showed in Section 2 can be established, see, e.g., [40].

Lemma 4.1. *(i) For any fixed $\underline{x} \in \mathcal{R}$, the following holds:*

$$\bigvee_{\underline{k}\,\in\,\mathcal{J}_n} \Psi_\sigma(n\underline{x} - \underline{k}) \ \geq \ [\phi_\sigma(1)]^s > 0,$$

where $\mathcal{J}_n$ denotes the set of indices $\underline{k} \in \mathbb{Z}^s$, such that $k_i = \lceil na_i \rceil$, $\lceil na_i \rceil + 1$, $\lceil na_i \rceil + 2$, ..., $\lfloor nb_i \rfloor$, $i = 1, ..., s$, for every $n \in \mathbb{N}^+$ sufficiently large.

(ii) For every $\gamma > 0$, we have

$$\bigvee_{\substack{\underline{k}\,\in\,\mathbb{Z}^s\,: \\ \|\underline{x}-\underline{k}\|_2 > n\gamma}} \Psi_\sigma(\underline{x} - \underline{k}) \ = \ \mathcal{O}(n^{-1-\alpha}), \qquad as \qquad n \to +\infty,$$

uniformly with respect to $\underline{x} \in \mathbb{R}^s$.

Remark 4.2. Note that, in order to prove the convergence results for the max-product NN operators defined below, it suffices to assume a weakened version of assumption $(\Sigma 3)$ on σ. Indeed, by requiring that $\sigma(x) = \mathcal{O}(|x|^{-\alpha})$, as $x \to -\infty$, for $\alpha > 0$, it turns out that

$$\bigvee_{\substack{\underline{k}\,\in\,\mathbb{Z}^s\,: \\ \|\underline{x}-\underline{k}\|_2 > n\gamma}} \Psi_\sigma(\underline{x} - \underline{k}) \ = \ \mathcal{O}(n^{-\alpha}), \qquad as \qquad n \to +\infty,$$

uniformly with respect to $\underline{x} \in \mathbb{R}^s$, with $\gamma > 0$, see [39, 40]. The above property can be used in order to prove a pointwise and uniform convergence theorem for the max-product NN operators.

Now, we recall the following

Definition 4.3. *Let $f : \mathcal{R} \to \mathbb{R}$ be a bounded function, and $n \in \mathbb{N}^+$ such that $\lceil na_i \rceil \leq \lfloor nb_i \rfloor$, $i = 1, ..., s$. The multivariate max-product neural network (NN) operators activated by σ are defined by*

$$F_n^{(M)}(f,\, \underline{x}) \ := \ \frac{\displaystyle\bigvee_{\underline{k} \in \mathcal{J}_n} f\left(\underline{k}/n\right) \Psi_\sigma(n\underline{x} - \underline{k})}{\displaystyle\bigvee_{\underline{k} \in \mathcal{J}_n} \Psi_\sigma(n\underline{x} - \underline{k})}, \qquad \underline{x} \in \mathcal{R},$$

where $\mathcal{J}_n$ denotes the set of indexes $\underline{k} \in \mathbb{Z}^s$, such that $k_i = \lceil na_i \rceil, \lceil na_i \rceil + 1, \lceil na_i \rceil + 2, ..., \lfloor nb_i \rfloor$, $i = 1, ..., s$, for every $n \in \mathbb{N}^+$ sufficiently large.

In general, when $n \in \mathbb{N}^+$ is sufficiently large, it turns out that every component $\lceil na_i \rceil \leq \lfloor nb_i \rfloor$, $i = 1, ..., s$. Further, in view of Lemma 4.1 (i), and since f is bounded, we have that the above operators are well-defined. Moreover, denoting by $\mathbf{1}$ the unitary constant function on $\mathcal{R}$, it is easy to prove the following useful property

$$F_n^{(M)}(\mathbf{1},\, \underline{x}) \ = \ 1, \qquad \underline{x} \in \mathcal{R}.$$

Other important properties of the operators $F_n^{(M)}$ can be proved in the following

Theorem 4.4. *Let $f,\, g : \mathcal{R} \to \mathbb{R}$ be bounded functions. The following properties hold for sufficiently large $n \in \mathbb{N}^+$:*

(i) If $\sigma(x)$ is continuous on $\mathbb{R}$, then $F_n^{(M)}(f, \cdot)$ is continuous on $\mathcal{R}$.

(ii) If $f(\underline{x}) \leq g(\underline{x})$, for each $\underline{x} \in \mathcal{R}$, we have $F_n^{(M)}(f, \underline{x}) \leq F_n^{(M)}(g, \underline{x})$, for every $\underline{x} \in \mathcal{R}$.

(iii) $F_n^{(M)}(f + g,\, \underline{x}) \leq F_n^{(M)}(f, \underline{x}) + F_n^{(M)}(g, \underline{x})$, $\underline{x} \in \mathcal{R}$, i.e., the $F_n^{(M)}$'s are sub-additive (or sub-linear) operators.

(iv) $\left| F_n^{(M)}(f, \underline{x}) - F_n^{(M)}(g, \underline{x}) \right| \leq F_n^{(M)}(|f - g|,\, \underline{x})$, $\underline{x} \in \mathcal{R}$.

(v) The $F_n^{(M)}$'s are positive homogeneous operators, i.e., for each $\lambda > 0$, it turns out that $F_n^{(M)}(\lambda f, \underline{x}) = \lambda F^{(M)}(f, \underline{x})$, $\underline{x} \in \mathcal{R}$.

The proof of all such properties (listed in Theorem 4.4), follows easily by the properties of $\bigvee$. For more details, see [39, 40].

We are now able to prove the following approximation results.

Theorem 4.5. *Let $f : \mathcal{R} \to \mathbb{R}_0^+$ be a bounded function, and let $\underline{x} \in \mathcal{R}$ be a point of continuity for f. Then*

$$\lim_{n \to +\infty} F_n^{(M)}(f, \underline{x}) \ = \ f(\underline{x}).$$

Moreover, if $f \in C_+(\mathcal{R})$ and bounded, it turns out that

$$\lim_{n\to+\infty} \|F_n^{(M)}(f,\cdot) - f(\cdot)\|_\infty \;=\; 0.$$

Proof. We will prove the first part of theorem only, since the second one follows by similar arguments and exploiting the uniform continuity of the function f.

Let now $\underline{x} \in \mathcal{R}$ be a fixed point of continuity for f. By using the properties stated in Theorem 4.4, we can write

$$\begin{aligned}
|F_n^{(M)}(f,\underline{x}) - f(\underline{x})| \;&\leq\; |F_n^{(M)}(f,\underline{x}) - f(\underline{x})F_n^{(M)}(\mathbf{1},\underline{x})| + |f(\underline{x})F_n^{(M)}(\mathbf{1},\underline{x}) - f(\underline{x})| \\
&=\; |F_n^{(M)}(f,\underline{x}) - f(\underline{x})F_n^{(M)}(\mathbf{1},\underline{x})| \;+\; f(\underline{x})|F_n^{(M)}(\mathbf{1},\underline{x}) - 1| \\
&=\; |F_n^{(M)}(f,\underline{x}) - F_n^{(M)}(f_{\underline{x}},\, \underline{x})| \;\leq\; F_n^{(M)}(|f - f_{\underline{x}}|,\, \underline{x}),
\end{aligned}$$

for every $n \in \mathbb{N}^+$ sufficiently large, where by $f_{\underline{x}}$ we denote the following auxiliary function:

$$f_{\underline{x}}(\underline{t}) \;:=\; f(\underline{x}), \qquad \underline{t} \in \mathcal{R}.$$

Let now $\varepsilon > 0$ be fixed, we denote by $\gamma > 0$ the corresponding parameter related to the continuity of f at $\underline{x}$. Thus, by Lemma 4.1 (i), we can obtain

$$F_n^{(M)}(|f - f_{\underline{x}}|,\, \underline{x}) \;\leq\; \frac{1}{[\phi_\sigma(1)]^s} \left[\bigvee_{\underline{k}\in\mathcal{J}_n} |f\,(\underline{k}/n) - f(\underline{x})|\; \Psi_\sigma(n\underline{x} - \underline{k}) \right]$$

$$= \frac{1}{[\phi_\sigma(1)]^s} \max \left\{ \bigvee_{\substack{\underline{k}\in\mathcal{J}_n \\ \|n\underline{x} - \underline{k}\|_2 \leq n\gamma}} |f\,(\underline{k}/n) - f(\underline{x})|\; \Psi_\sigma(n\underline{x} - \underline{k}); \right.$$

$$\left. \bigvee_{\substack{\underline{k}\in\mathcal{J}_n \\ \|n\underline{x} - \underline{k}\|_2 > n\gamma}} |f\,(\underline{k}/n) - f(\underline{x})|\; \Psi_\sigma(n\underline{x} - \underline{k}) \right\} \;=:\; \frac{1}{[\phi_\sigma(1)]^s} \max\{I_1,\, I_2\}.$$

Now, concerning I_1, if $\|n\underline{x} - \underline{k}\|_2 \leq n\gamma$ or equivalently $\|\underline{x} - (\underline{k}/n)\|_2 \leq \gamma$, and by the continuity of f at $\underline{x}$, we have

$$I_1 \;\;\leq\;\; \varepsilon \cdot \bigvee_{\substack{\underline{k}\in\mathcal{J}_n \\ \|n\underline{x} - \underline{k}\|_2 \leq n\gamma}} \Psi_\sigma(n\underline{x} - \underline{k}) \;\;\leq\;\; \varepsilon \cdot [\phi_\sigma(0)]^s,$$

in view of Remark 2.3. As for I_2, we can use the boundedness of f and Lemma 4.1 (ii), to show that

$$I_2 \;\;\leq\;\; 2\,\|f\|_\infty \cdot \bigvee_{\|n\underline{x} - \underline{k}\|_2 > n\gamma} \Psi_\sigma(n\underline{x} - \underline{k}) \;\;\leq\;\; \varepsilon,$$

for $n \in \mathbb{N}^+$ sufficiently large. Finally, the proof follows by the arbitrariness of ε. □

Remark 4.6. Note that, the convergence results proved in Theorem 4.5 can be extended to every bounded function $f : \mathcal{R} \to \mathbb{R}$, by the following procedure. Let $c \in \mathbb{R}$ be a constant, such that $c \leq \inf_{\underline{x} \in \mathcal{R}} f(\underline{x})$. In this case, we have $f(\underline{x}) - c \geq 0$, for every $\underline{x} \in \mathcal{R}$, then $F_n^{(M)}(f - c, \cdot)$ converges to $f - c$, as $n \to +\infty$, pointwise, at any point of continuity of f, or uniformly, if f is continuous on the whole $\mathcal{R}$. For the above reasons, it is easy to see that the family $F_n^{(M)}(f - c, \cdot) + c$ converges to f, as $n \to +\infty$, (see [27] for the same consideration in case of Bernstein max-product operators).

The multivariate NN operators above of the max-product type, allow us to obtain a *constructive*, *nonlinear* approximation formula, which allow us to achieve *more accurate* approximations than the corresponding linear counterparts studied in Section 3.

More precisely, the estimates for the *order of approximation* that can be achieved using the operators studied in Section 3, are very similar to those established in [40], and the theoretical order of approximation is the same in both cases, i.e., it is proportional to $\omega(f, n^{-1})$, when the assumption $(\Sigma 3)$ is satisfied for $\alpha \geq 1$. Here $\omega(f, \delta)$ denotes the modulus of continuity of f, $\delta > 0$, as customary. However, the quality of the approximation turns out to be globally better using the NN operators of the max-product type, since all the constants obtained in the aforementioned estimates are sharper compared to those obtained in [31]. For more details, see also [39].

Remark 4.7. The theory of NN operators of the max-product type activated by sigmoidal functions, introduced and analyzed in [39, 40], arise as an extension of the results proved in [3]. In particular, in [3] the authors studied max-product NN operators in one dimension, similar to those introduced by Cardaliaguet and Euvrard in [19], and defined as

$$C_{n,\alpha}^{(M)}(f)(x) := \frac{\bigvee_{k=-n^2}^{n^2} b\left(n^{1-\alpha}\left(x - \frac{k}{n}\right)\right) f\left(\frac{k}{n}\right)}{\bigvee_{k=-n^2}^{n^2} b\left(n^{1-\alpha}\left(x - \frac{k}{n}\right)\right)}, \qquad x \in \mathbb{R},$$

for any uniformly continuous and bounded function $f : \mathbb{R} \to \mathbb{R}$, where $b(x)$ is a certain centered bell-shaped function, and $0 < \alpha < 1$.

In order to obtain approximation results, the authors make on the function $b(x)$ some quite restrictive assumptions, such as: $b(x)$ continuous, $supp\, b \subseteq [-T, T]$, $T > 0$, and, moreover,

$$m_1(T - x) \;\leq\; b(x) \;\leq\; M_1(T - x), \qquad x \in [0, T], \tag{6}$$

$$m_2(T+x) \;\leq\; b(x) \;\leq\; M_2(T+x), \qquad x \in [-T, 0], \tag{7}$$

for some positive constants m_1, m_2, M_1, and M_2. One of the advantages we obtained in this section by the present approach, and by means of the operators $F_n^{(M)}$, is that, in case of the univariate theory, the function $\Psi_\sigma(\underline{x})$ reduces to $\phi_\sigma(x)$, which is a centered bell-shaped function (as noted in Remark 2.3(a)), not necessarily continuous, whose support can be also unbounded, and thus assumptions (6) and (7) can be dropped.

In particular, the request that $supp\, b \subseteq [-T, T]$ does not allow to choose $b(x) = \phi_\sigma(x)$, being σ either the logistic or the hyperbolic tangent sigmoidal functions, which are very important instances of sigmoidal functions within the theory of the NNs, see, e.g., [43, 52, 50].

Finally, we can observe that the *order* of approximation achieved by the operators $C_{n,\alpha}^{(M)}$ (see Theorem 3.2 and Corollary 3.3. in [3]) is shown by

$$|f(x) - C_{n,\alpha}^{(M)}(f)(x)| \;\leq\; C\, \omega(f, n^{\alpha-1}), \qquad x \in \mathbb{R}, \tag{8}$$

$0 < \alpha < 1$, $C > 0$, $n \in \mathbb{N}^+$. Comparing the estimate in (8) with the result showed in [39] concerning the order of approximation achieved by $F_n^{(M)}$, it is clear that the rate of approximation achieved by the max-product NN operators, $F_n^{(M)}$, is better than that attained by the operators $C_{n,\alpha}^{(M)}$.

For the sake of completeness, we can observe that, by means of the operators $C_{n,\alpha}^{(M)}$, we can approximate positive functions defined on the whole real line, while using $F_n^{(M)}$, we deal with functions defined on compact sets. In fact, in order to obtain operators useful to approximate functions defined on the full real line, the theory developed in this section has been extended in [39], obtaining the same advantages described above.

5 The neural network operators of the Kantorovich type

The theoretical results proved in previous sections are given for continuous functions only. This is due to the pointwise nature of both, the classical and max-product NN operators. In order to establish approximation results also valid for not necessarily continuous functions, we resort to an "averaged version" of the previous NN operators, i.e., to a Kantorovich-type version of the F_n's, see [32].

In this section, we introduce and analyze this kind of operators. Here, we consider sigmoidal activation functions, σ, which satisfy the technical condition

$$\sigma(3) \;>\; \sigma(1). \tag{9}$$

In this case, we first prove that

$$\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor-1} \dots \sum_{k_s=\lceil na_s\rceil}^{\lfloor nb_s\rfloor-1} \Psi_\sigma(n\underline{x}-\underline{k})$$

$$= \prod_{i=1}^{s} \sum_{k_i=\lceil na_i\rceil}^{\lfloor nb_i\rfloor-1} \phi_\sigma(nx_i-k_i) \;\geq\; [\phi_\sigma(2)]^s \;>\; 0, \qquad (10)$$

for every $\underline{x} \in \mathcal{R}$, where $\underline{k} := (k_1, ..., k_s) \in \mathbb{Z}^s$, see [32]. We then define the operators that will be studied in this section.

Definition 5.1. Let $f : \mathcal{R} \to \mathbb{R}$ be a locally integrable function, and $n \in \mathbb{N}^+$ such that $\lceil na_i\rceil \leq \lfloor nb_i\rfloor - 1$ for every $i = 1, ..., s$. The "linear multivariate Kantorovich-type NN operators", $K_n(f, \cdot)$, activated by the sigmoidal function σ and acting on f, are defined by

$$K_n(f,\underline{x}) \;:=\; \frac{\displaystyle\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor-1} \cdots \sum_{k_s=\lceil na_s\rceil}^{\lfloor nb_s\rfloor-1} \left[n^s \int_{R^n_{\underline{k}}} f(\underline{u})\, d\underline{u} \right] \Psi_\sigma(n\underline{x}-\underline{k})}{\displaystyle\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor-1} \cdots \sum_{k_s=\lceil na_s\rceil}^{\lfloor nb_s\rfloor-1} \Psi_\sigma(n\underline{x}-\underline{k})},$$

where $R^n_{\underline{k}}$ are the sets defined by

$$R^n_{\underline{k}} \;:=\; \left[\frac{k_1}{n}, \frac{k_1+1}{n}\right] \times \cdots \times \left[\frac{k_s}{n}, \frac{k_s+1}{n}\right], \qquad (11)$$

for every $\underline{k} = (k_1, ..., k_s) \in \mathbb{Z}^s$, $n \in \mathbb{N}^+$, and $\underline{x} \in \mathcal{R} \subset \mathbb{R}^s$.

For $n \in \mathbb{N}^+$ sufficiently large, we always obtain $\lceil na_i\rceil \leq \lfloor nb_i\rfloor - 1$, $i = 1, ..., s$. Moreover, if $\lceil na_i\rceil \leq k_i \leq \lfloor nb_i\rfloor - 1$, then $a_i \leq \frac{k_i}{n} \leq b_i - \frac{1}{n}$ and $a_i \leq \frac{k_i+1}{n} \leq b_i$.

Note that, the operators K_n are well-defined, for instance, when applied to bounded functions. Indeed, when $f \in L^\infty(\mathcal{R})$, it is easy to see that $|K_n(f,\underline{x})| \leq \|f\|_\infty$, for all $\underline{x} \in \mathcal{R}$. Moreover, we can observe that $K_n(1,\underline{x}) = 1$, for every $\underline{x} \in \mathcal{R}$ and n sufficiently large. First of all, we recall the following theorem which shows that, even for $(K_n)_{n\in\mathbb{N}^+}$, pointwise and uniform results for continuous functions, can be proved see [32].

Theorem 5.2. *Let $f : \mathcal{R} \to \mathbb{R}$ be bounded. Then,*

$$\lim_{n\to+\infty} K_n(f,\underline{x}) = f(\underline{x}),$$

at each point $\underline{x} \in \mathcal{R}$ where f is continuous. Moreover, if $f \in C(\mathcal{R})$, then

$$\lim_{n\to+\infty} \sup_{\underline{x}\in\mathcal{R}} |K_n(f,\underline{x}) - f(\underline{x})| = \lim_{n\to+\infty} \|K_n(f,\cdot) - f(\cdot)\|_\infty = 0.$$

We omit the proof of Theorem 5.2, since it can be obtained by a procedure similar to that followed for the proofs of the pointwise and uniform convergence results of Section 3 and Section 4. As an easy consequence of Theorem 5.2, we have the following

Theorem 5.3. *For every $f \in C(\mathcal{R})$, we have*

$$\lim_{n\to+\infty} \|K(f,\cdot) - f(\cdot)\|_p = 0,$$

where $\|\cdot\|_p$ denotes the usual $L^p(\mathcal{R})$ norm, with $1 \leq p < +\infty$.

In order to establish the convergence in L^p of the family of the above operators, we first prove the following

Theorem 5.4. *The inequality*

$$\|K_n(f,\cdot)\|_p \ \leq \ \frac{\|\Psi_\sigma\|_1^{1/p}}{[\phi(2)]^{s/p}}\,\|f\|_p$$

holds for every $f \in L^p(\mathcal{R})$, $1 \leq p < +\infty$, where $\|\cdot\|_p$ is the usual $L^p(\mathcal{R})$ norm, and $\|\Psi_\sigma\|_1 < +\infty$ (Ψ_σ belongs to $L^1(\mathbb{R}^s)$).

Proof. For every $f \in L^p(\mathcal{R})$, $1 \leq p < +\infty$, we have

$$\|K_n(f,\cdot)\|_p = \left(\int_{\mathcal{R}} |K_n(f,\underline{x})|^p \ d\underline{x}\right)^{1/p}$$

$$= \left(\int_{\mathcal{R}} \left| \frac{\displaystyle\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor-1} \cdots \sum_{k_s=\lceil na_s\rceil}^{\lfloor nb_s\rfloor-1} \left[n^s \int_{R^n_{\underline{k}}} f\left(\underline{u}\right)\, d\underline{u}\right] \Psi_\sigma(n\underline{x}-\underline{k})}{\displaystyle\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor-1} \cdots \sum_{k_s=\lceil na_s\rceil}^{\lfloor nb_s\rfloor-1} \Psi_\sigma(n\underline{x}-\underline{k})} \right|^p d\underline{x} \right)^{1/p}.$$

Being $|\cdot|^p$ convex, we infer from Jensen's inequality (see, e.g., [34]) and (10), that

$$\|K_n(f,\cdot)\|_p \leq \left(\int_{\mathcal{R}} \frac{\displaystyle\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor-1} \cdots \sum_{k_s=\lceil na_s\rceil}^{\lfloor nb_s\rfloor-1} \Psi_\sigma(n\underline{x}-\underline{k}) \left| n^s \int_{R^n_{\underline{k}}} f\left(\underline{u}\right)\, d\underline{u}\right|^p}{\displaystyle\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor-1} \cdots \sum_{k_s=\lceil na_s\rceil}^{\lfloor nb_s\rfloor-1} \Psi_\sigma(n\underline{x}-\underline{k})}\, d\underline{x}\right)^{1/p}$$

$$\leq \frac{1}{[\phi(2)]^{s/p}} \left(\int_{\mathcal{R}} \sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor-1} \cdots \sum_{k_s=\lceil na_s\rceil}^{\lfloor nb_s\rfloor-1} \Psi_\sigma(n\underline{x}-\underline{k}) \left| n^s \int_{R^n_{\underline{k}}} f\left(\underline{u}\right)\, d\underline{u}\right|^p d\underline{x}\right)^{1/p}$$

$$\leq \frac{1}{[\phi(2)]^{s/p}} \left(\sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor - 1} \cdots \sum_{k_s=\lceil na_s \rceil}^{\lfloor nb_s \rfloor - 1} \int_{\mathbb{R}^s} \Psi_\sigma(n\underline{x} - \underline{k}) \left| n^s \int_{R^n_{\underline{k}}} f\left(\underline{u}\right) \, d\underline{u} \right|^p d\underline{x} \right)^{1/p}.$$

Changing variables, setting $\underline{x} = (\underline{t} + \underline{k})/n$, we obtain

$$\|K_n(f, \cdot)\|_p \leq \frac{1}{[\phi(2)]^{s/p}} \left(\sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor - 1} \cdots \sum_{k_s=\lceil na_s \rceil}^{\lfloor nb_s \rfloor - 1} \frac{1}{n^s} \int_{\mathbb{R}^s} \Psi_\sigma(\underline{t}) \left| n^s \int_{R^n_{\underline{k}}} f\left(\underline{u}\right) \, d\underline{u} \right|^p d\underline{t} \right)^{1/p}$$

$$= \frac{1}{[\phi(2)]^{s/p}} \left(\frac{1}{n^s} \int_{\mathbb{R}^s} \Psi_\sigma(\underline{t}) \sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor - 1} \cdots \sum_{k_s=\lceil na_s \rceil}^{\lfloor nb_s \rfloor - 1} \left| n^s \int_{R^n_{\underline{k}}} f\left(\underline{u}\right) \, d\underline{u} \right|^p d\underline{t} \right)^{1/p}.$$

Using again Jensen's inequality (see e.g. [34]), we obtain

$$\|K_n(f, \cdot)\|_p \leq \frac{1}{[\phi(2)]^{s/p}} \left(\frac{1}{n^s} \int_{\mathbb{R}^s} \Psi_\sigma(\underline{t}) \, d\underline{t} \sum_{k_1=\lceil na_1 \rceil}^{\lfloor nb_1 \rfloor - 1} \cdots \sum_{k_s=\lceil na_s \rceil}^{\lfloor nb_s \rfloor - 1} n^s \int_{R^n_{\underline{k}}} \left| f\left(\underline{u}\right) \right|^p \, d\underline{u} \right)^{1/p}$$

$$\leq \frac{1}{[\phi(2)]^{s/p}} \left(\int_{\mathbb{R}^s} \Psi_\sigma(\underline{t}) \, d\underline{t} \int_{\mathcal{R}} \left| f\left(\underline{u}\right) \right|^p \, d\underline{u} \right)^{1/p} = \frac{1}{[\phi(2)]^{s/p}} \|f\|_p \left(\int_{\mathbb{R}^s} \Psi_\sigma(\underline{t}) \, d\underline{t} \right)^{1/p}.$$

Thus, we can finally write:

$$\|K_n(f, \cdot)\|_p \;\leq\; \frac{\|\Psi_\sigma\|_1^{1/p}}{[\phi(2)]^{s/p}} \, \|f\|_p.$$

□

Remark 5.5. Note that, in case of continuous sigmoidal functions σ (recall that the present theory can also be applied to discontinuous sigmoidal functions, in view of Remark 2.3), the first part of condition (v) of Lemma 2.2, in the univariate case, $s = 1$, is equivalent to

$$\widehat{\phi_\sigma}(2\pi k) = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0, \end{cases} \tag{12}$$

with $k \in \mathbb{Z}$, being $\widehat{\phi_\sigma}(v) := \int_{\mathbb{R}} \phi_\sigma(u) e^{-iuv} \, du$ the Fourier transform of ϕ_σ, see, e.g., [12, 10]. Hence, it turns out that $\widehat{\phi_\sigma}(0) = \int_{\mathbb{R}} \phi_\sigma(y) \, dy = 1$. In the present case, by the definition of the multivariate density function $\Psi_\sigma(\underline{x})$, it is easy to see that

$$\|\Psi_\sigma\|_1 = \int_{\mathbb{R}^s} \Psi_\sigma(\underline{x}) \, d\underline{x} \;=\; \int_{\mathbb{R}} \phi_\sigma(x_1) \, dx_1 \cdot \ldots \cdot \int_{\mathbb{R}} \phi_\sigma(x_s) \, dx_s \;=\; 1.$$

Then, the inequality of Theorem 5.4 reduces to

$$\|K_n(f, \cdot)\|_p \;\leq\; \phi_\sigma(2)^{-s/p} \cdot \|f(\cdot)\|_p, \tag{13}$$

for every $f \in L^p(\mathcal{R})$, $1 \leq p < +\infty$.

From Theorem 5.4, we conclude that the operators $K_n : L^p(\mathcal{R}) \to L^p(\mathcal{R})$, $1 \leq p < +\infty$, are continuous.

We are now able to prove the following convergence result.

Theorem 5.6. *For every $f \in L^p(\mathcal{R})$, $1 \leq p < +\infty$, we have*

$$\lim_{n\to+\infty} \|K_n(f,\cdot) - f\|_p = 0.$$

Proof. Let be $f \in L^p(\mathcal{R})$ and $\varepsilon > 0$ be fixed. Since the space $C(\mathcal{R})$ is dense in $L^p(\mathcal{R})$ with respect the norm $\|\cdot\|_p$, there exists $g \in C(\mathcal{R})$ such that $\|f(\cdot) - g(\cdot)\|_p < (\phi_\sigma(2)^{-s/p} + 1)^{-1}\varepsilon/2$. Then, by Theorem 5.4,

$$\|K_n(f,\cdot) - f(\cdot)\|_p \leq \|K_n(f,\cdot) - K_n(g,\cdot)\|_p + \|K_n(g,\cdot) - g(\cdot)\|_p + \|g(\cdot) - f(\cdot)\|_p$$

$$\leq \frac{1}{[\phi_\sigma(2)]^{s/p}}\|g(\cdot) - f(\cdot)\|_p + \|K_n(g,\cdot) - g(\cdot)\|_p + \|g(\cdot) - f(\cdot)\|_p$$

$$\leq \left(\frac{1}{[\phi_\sigma(2)]^{s/p}} + 1\right)\|g(\cdot) - f(\cdot)\|_p + \|K_n(g,\cdot) - g(\cdot)\|_p < \frac{\varepsilon}{2} + \|K_n(g,\cdot) - g(\cdot)\|_p.$$

Finally, by Theorem 5.3,

$$\|K_n(f,\cdot) - f(\cdot)\|_p \ < \ \frac{\varepsilon}{2} \ + \ \frac{\varepsilon}{2} \ = \ \varepsilon,$$

for $n \in \mathbb{N}^+$ sufficiently large, the proof follows by the arbitrariness of ε. □

An extension of the results proved in this section to functions belonging to Orlicz spaces, has been obtained in [41]. Moreover, a max-product version of the Kantorovich NN operators has also been obtained in [42].

6 Examples of sigmoidal activation functions

In this section we provide some examples of sigmoidal activation functions which satisfy the assumptions needed in the above theory of the NN operators.

Examples of smooth sigmoidal functions are provided by the well-known *logistic function*, $\sigma_\ell(x) = (1+e^{-x})^{-1}$, $x \in \mathbb{R}$ [14, 49, 57], and by the *hyperbolic tangent* sigmoidal function, $\sigma_h(x) := (1/2)(\tanh x + 1)$, $x \in \mathbb{R}$ [2, 16]. In particular, σ_ℓ and σ_h satisfy condition ($\Sigma 3$) for all $\alpha > 0$. Moreover, we observe that both, σ_ℓ and σ_h, satisfy the technical assumptions given in (4) and (9), thus, they can be used as activation functions for all the NN operators studied in the present survey.

An example of non-smooth sigmoidal function is also provided by the *ramp function*, $\sigma_{R_1}(x)$, defined by

$$\sigma_{R_1}(x) \ := \ \begin{cases} 0, & x < -1/2 \\ x + 1/2, & -1/2 \leq x \leq 1/2 \\ 1, & x > 1/2, \end{cases}$$

see e.g., [20, 15]. Here, condition (Σ3) is satisfied for every $\alpha > 0$, and the corresponding function $\phi_{\sigma_{R_1}}(x)$ has compact support. Furthermore, we can note that $\sigma_{R_1}(x)$ satisfies the technical condition (4), while it does not satisfy (9). For the latter reason, $\sigma_{R_1}(x)$ can be used as activation function for the operators F_n and $F_n^{(M)}$, but it cannot be used to define the Kantorovich - type NN operators. A modified version of $\sigma_{R_1}(x)$, for which (9) holds, is given by the "modified ramp function"

$$\sigma_{R_2}(x) \ := \ \begin{cases} 0, & x < -3/2 \\ x/3 + 1/2 & -3/2 \leq x \leq 3/2 \\ 1, & x > 3/2. \end{cases}$$

Also in the case above, all the assumptions of Section 2 are satisfied.

Other examples of sigmoidal functions satisfying the assumptions of the previous theory can be constructed starting from the well-known central B-spline of order $n \in \mathbb{N}^+$ [12],

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

by a simple procedure, described in [33]. Here, the function $(x)_+ := \max\{x, 0\}$ is the positive part of $x \in \mathbb{R}$. We then define the sigmoidal function $\sigma_{M_n}(x)$ by

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

Note that, $\sigma_{M_1}(x)$ coincides exactly with the ramp function, $\sigma_{R_1}(x)$. An application of such sigmoidal functions has been obtained in [28, 29, 17], where the authors established some interpolation results by means of a family of NN operators activated by $\sigma_{M_n}(x)$. This fact is useful from the point of view of the theory of artificial NNs, and is relate to the problem of the "training of an NN", see, e.g., [67, 51, 61, 28, 29, 60].

We now introduce the following step function (denoted by $\sigma_3(x)$), as an example of a discontinuous sigmoidal function which satisfies all assumptions of Section 2, namely

$$\sigma_3(x) \ := \ \begin{cases} 0, & x < -1 \\ 1/2, & -1 \leq x \leq 1 \\ 1, & x > 1. \end{cases}$$

In this case, $\sigma_3(x)$ satisfies both (4) and (9).

Finally, we can choose as activation functions for the above NN operators, the family of functions

$$\sigma_\gamma(x) \ := \ \begin{cases} \dfrac{1}{|x|^\gamma + 2}, & x < -2^{1/\gamma} \\[2ex] 2^{-(1/\gamma)-2}x + (1/2), & -2^{1/\gamma} \le x \le 2^{1/\gamma} \\[2ex] \dfrac{x^\gamma+1}{x^\gamma+2}, & x > 2^{1/\gamma}, \end{cases} \tag{14}$$

where $0 < \gamma \le 1$. Through simple calculations, it can be proved that σ_γ satisfies the inequality

$$\sigma'_\gamma(x) \ \le \ \phi_{\sigma_\gamma}(x) \ \le \ \sigma'_\gamma(x) + \frac{1}{4}\,\sigma''(x+1), \qquad \text{for every} \qquad x < -2^{1/\gamma} - 1,$$

where $\phi_{\sigma_\gamma}(x) := \frac{1}{2}[\sigma_\gamma(x+1) - \sigma_\gamma(x-1)]$, and observing that $\sigma'_\gamma(x)$ decays asymptotically as $|x|^{-\gamma-1}$, for $x \to -\infty$, and $\sigma''_\gamma(x)$ decays asymptotically as $|x|^{-\gamma-2}$, for $x \to -\infty$, it turns out that $\phi_{\sigma_\gamma}(x) = \mathcal{O}(|x|^{-\gamma-1})$, as $x \to \pm\infty$ (since ϕ_{σ_γ} is an even function), then it belongs to $L^1(\mathbb{R})$. Clearly, also $\Psi_{\sigma_\gamma} \in L^1(\mathbb{R}^s)$, and together with ϕ_{σ_γ}, they satisfy all the properties listed in Lemma 2.2, required to prove the convergence results displayed in the present paper.

7 Final remarks and conclusions

In this paper, we recall some convergence results concerning the theory of NN operators. By means of such operators, we are able to obtain some constructive approximation processes based on a kind of NNs. The theory is developed for functions defined only on bounded domains of $\mathbb{R}^s$. However, the theory can be extended in such a way that even functions defined on the whole real line can be approximated in this way, see [30, 31].

The main difference between the operators studied in Section 3 and Section 4 (with the Kantorovich-type operators), is that the classical and the max-product NN operators, due to their definition, are suited to approximate continuous functions, while the Kantorovich-type NN operators allows us to reconstruct also not necessarily continuous functions (e.g., functions belonging to L^p-spaces). In addition, the latter could be useful in a number of applications to Image Processing.

The operators F_n, $F_n^{(M)}$, and K_n are strictly related to the theory of *sampling operators* (see, e.g., [10, 7, 8, 9, 22, 23]). In fact, the operators studied in this paper, allow us to reconstruct a given analog signal by means of discrete families of its sample values. More specifically, the NN operators above allow us to reconstruct *duration limited* signals (see, e.g., [4, 5, 6, 32, 35, 36, 37, 38, 39]), and are characterized by kernel functions constructed by using sigmoidal functions. Note that, max-product sampling operators defined by kernels with compact support, Fejér type kernels, and *sinc* kernels have been studied in [24, 25, 26] by Coroianu and Gal.

Acknowledgments

The first and the third author of the paper are members of the Gruppo Nazionale per l'Analisi Matematica, la Probabilitá e le loro Applicazioni (GNAMPA), the second author is member of the Gruppo Nazionale per la Fisica Matematica (GNFM), of the Istituto Nazionale di Alta Matematica (INdAM). The authors were supported, in part, by the "Department of Mathematics and Computer Science", University of Perugia, Italy, and by the "Department of Mathematics and Physics" (Section of Mathematics) of the Roma Tre University, Italy. The first author was partially supported by the GNAMPA-INdAM Project "Metodi di approssimazione e applicazioni al Signal e Image Processing", project number U2015/000396, 12/03/2015.

References

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

[2] G.A. Anastassiou, Intelligent Systems: Approximation by Artificial Neural Networks, Intelligent Systems Reference Library 19, Springer-Verlag, Berlin, 2011.

[3] G.A. Anastassiou, L. Coroianu, S.G. Gal, Approximation by a nonlinear Cardaliaguet-Euvrard neural network operator of max-product kind, J. Comp. Anal. Appl. 12 (2) (2010), 396-406.

[4] L. Angeloni, G. Vinti, Convergence and rate of approximation for linear integral operators in BV^{ϕ}-spaces in multidimensional setting, Journal of Mathematical Analysis and Applications 349 (2009), 317-334.

[5] L. Angeloni, G. Vinti, Approximation with respect to Goffman-Serrin variation by means of non-convolution type integral operators, Numerical Functional Analysis and Optimization 31 (2010), 519-548.

[6] L. Angeloni, G. Vinti, Approximation in variation by homothetic operators in multidimensional setting, Differential Integral Equations 26 (5-6) (2013), 655-674.

[7] C. Bardaro, P.L. Butzer, R.L. Stens, G. Vinti, Kantorovich-Type generalized sampling series in the setting of Orlicz spaces, Sampling Theory in Signal and Image Processing, 6 (1) (2007), 29-52.

[8] C. Bardaro, H. Karsli, G. Vinti, On pointwise convergence of linear integral operators with homogeneous kernels, Integral Transforms and Special Functions 19 (6) (2008), 429-439.

[9] C. Bardaro, H. Karsli, G. Vinti, Nonlinear integral operators with homogeneous kernels: pointwise approximation theorems, Applicable Analysis, 90 (2011), 463-474.

[10] C. Bardaro, J. Musielak, G. Vinti, Nonlinear Integral Operators and Applications, New York, Berlin: De Gruyter Series in Nonlinear Analysis and Applications 9, 2003.

[11] A.R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Trans. Inform. Theory 39 (3) (1993) 930-945.

[12] P.L. Butzer and R.J. Nessel, Fourier Analysis and Approximation I, Academic Press, New York-London, 1971.

[13] E.J. Candés, Harmonic analysis of neural networks, Applied and Computational Harmonic Analysis 6 (1999) 197-218.

[14] F. Cao, Z. Chen, The approximation operators with sigmoidal functions, Comput. Math. Appl. 58 (4) (2009) 758-765.

[15] F. Cao, Z. Chen, The construction and approximation of a class of neural networks operators with ramp functions, J. Comput. Anal. Appl. 14 (1) (2012) 101-112.

[16] F. Cao, Z. Chen, The construction and approximation of feedforward neural network with hyperbolic tangent function, Applied Mathematics - A Journal of Chinese Universities, 30 (2) (2015), 151-162.

[17] F. Cao, Z. Chen, Scattered data approximation by neural networks operators, in print in: Neurocomputing (2016), 10.1016/j.neucom.2016.01.013.

[18] F. Cao, B. Liu, D.S. Park, Image classification based on effective extreme learning machine, Neurocomputing 102 (2013), 90-97.

[19] P. Cardaliaguet, G. Euvrard, Approximation of a function and its derivative with a neural network, Neural Networks 5 (2) (1992) 207-220.

[20] G. H. L. Cheang, Approximation with neural networks activated by ramp sigmoids, J. Approx. Theory 162 (2010) 1450-1465.

[21] E.W. Cheney, W.A. Light, Y, Xu, Constructive methods of approximation by rigde functions and radial functions, Numerical Algorithms, 4 (2) (1993), 205-223.

[22] F. Cluni, D. Costarelli, A.M. Minotti, G. Vinti, Enhancement of thermographic images as tool for structural analysis in earthquake engineering, NDT & E International, 70 (2015), 60-72.

[23] F. Cluni, D. Costarelli, A.M. Minotti, G. Vinti, Applications of sampling Kantorovich operators to thermographic images for seismic engineering, Journal of Computational Analysis and Applications, 19 (4) (2015) 602-617.

[24] L. Coroianu, S.G. Gal, Approximation by nonlinear generalized sampling operators of max-product kind, Sampling Th. Signal Image Processing 9 (1-3) (2010) 59-75.

[25] L. Coroianu, S.G. Gal, Approximation by max-product sampling operators based on sinc-type kernels, Sampling Th. Signal Image Processing 10 (3) (2011) 211-230.

[26] L. Coroianu, S.G. Gal, Saturation results for the truncated max-product sampling operators based on sinc and Fejér-type kernels, Sampling Th. Signal Image Processing 11 (1) (2012) 113-132.

[27] L. Coroianu, S.G. Gal, Saturation and inverse results for the Bernstein max-product operator, Period. Math. Hungar 69 (2014), 126-133.

[28] D. Costarelli, Interpolation by neural network operators activated by ramp functions, Journal of Mathematical Analysis and Application, 419 (1) (2014) 574-582.

[29] D. Costarelli, Neural network operators: constructive interpolation of multivariate functions, Neural Networks 67 (2015) 28-36.

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

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

[32] D. Costarelli, R. Spigler, Convergence of a family of neural network operators of the Kantorovich type, J. Approx. Theory, 185 (2014) 80-90.

[33] D. Costarelli, R. Spigler, Approximation by series of sigmoidal functions with applications to neural networks, Annali di Matematica Pura ed Applicata, 194 (1) (2015), 289-306.

[34] D. Costarelli, R. Spigler, How sharp is the Jensen inequality?, Journal of Inequalities and Applications, 2015:69 (2015).

[35] D. Costarelli, G. Vinti, Approximation by nonlinear multivariate sampling-Kantorovich type operators and applications to Image Processing, Numerical Functional Analysis and Optimization 34 (8) (2013), 819-844.

[36] D. Costarelli, G. Vinti, Order of approximation for sampling Kantorovich operators, Journal of Integral Equations and Applications 26 (3) (2014) 345-368.

[37] D. Costarelli, G. Vinti, Rate of approximation for multivariate sampling Kantorovich operators on some functions spaces, J. Int. Eq. Appl. 26 (4) (2014) 455-481.

[38] D. Costarelli, G. Vinti, Degree of approximation for nonlinear multivariate sampling Kantorovich operators on some functions spaces, Numerical Functional Analysis and Optimization, 36 (8) (2015) 964-990.

[39] D. Costarelli, G. Vinti, Max-product neural network and quasi interpolation operators activated by sigmoidal functions, submitted (2015).

[40] D. Costarelli, G. Vinti, Pointwise and uniform approximation by multivariate neural network operators of the max-product type, submitted (2015).

[41] D. Costarelli, G. Vinti, Convergence for a family of neural network operators in Orlicz spaces, submitted (2016).

[42] D. Costarelli, G. Vinti, Approximation by max-product neural network operators of Kantorovich type, submitted (2016).

[43] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Systems 2 (1989) 303-314.

[44] F. Girosi, T. Poggio, Networks and best approximation property, Biological Cybernetics 63 (1990) 169-176.

[45] G. Gnecco, M. Sanguineti, On a variational norm tailored to variable-basis approximation schemes, IEEE Trans. Inform. Theory 57 (2011) 549-558.

[46] G. Gripenberg, Approximation by neural network with a bounded number of nodes at each level, J. Approx. Theory 122 (2) (2003) 260-266.

[47] N.J. Guliyev, V. E. Ismailov, A single hidden layer feedforward network with only one neuron in the hidden layer can approximate any univariate function, in print: (2015) arXiv:1601.00013v1.

[48] N. Hahm, B. Hong, Approximation order to a function in $\overline{C}(\mathbb{R})$ by superposition of a sigmoidal function. Applied Math. Lett., 15 (2002) 591-597.

[49] A.I. Iliev, N. Kyurkchiev, S. Markov, On the approximation of the cut and step functions by logistic and Gompertz functions, BIOMATH 4 (2) (2015): Article-ID 1510101.

[50] V.E. Ismailov, On the approximation by neural networks with bounded number of neurons in hidden layers, J. Math. Anal. Appl. 417 (2) (2014) 963-969.

[51] Y. Ito, Independence of unscaled basis functions and finite mappings by neural networks, Math. Sci. 26 (2001) 117-126.

[52] P.C. Kainen, V. Kurková, An integral upper bound for neural network approximation, Neural Comput. 21 (2009) 2970-2989.

[53] P.C. Kainen, V. Kurková, A. Vogt, Best approximation by Heaviside perceptron network, Neural Network 13 (7) (2000), 695-697.

[54] P.C. Kainen, V. Kurková, A. Vogt, Geometry and topology of continuous best and near best approximations, Journal of Approximation Theory 105 (2000) 252-262.

[55] P.C. Kainen, V. Kurková, A. Vogt, Best approximation by linear combinations of characteristic functions of half-spaces, Journal of Approximation Theory 122 (2) (2003) 151-159.

[56] V. Kurková, M. Sanguineti, Model complexities of shallow networks representing highly varying functions, Neurocomputing 171 (2016), 598-604.

[57] N. Kyurkchiev, S. Markov, Sigmoid functions: some approximation and modelling aspects, Some Moduli in Programming Environment Mathematica, LAP (Lambert Acad. Publ.) (2015).

[58] G. Lewicki, G. Marino, Approximation by superpositions of a sigmoidal function, Z. Anal. Anwendungen Journal for Analysis and its Applications, 22 (2) (2003) 463-470.

[59] G. Lewicki, G. Marino, Approximation of functions of finite variation by superpositions of a sigmoidal function. Appl. Math. Lett., 17 (2004) 1147-1152.

[60] S. Lin, Z. Xu, J. Zeng, Error estimate for spherical neural networks interpolation, Neural Processing Letters 42 (2) (2015), 369-379.

[61] B. Llanas, F.J. Sainz, Constructive approximate interpolation by neural networks, J. Comput. Appl. Math. 188 (2006) 283-308.

[62] V. Maiorov, Approximation by neural networks and learning theory, J. Complexity 22 (1) (2006) 102-117.

[63] H.N. Mhaskar, C.A. Micchelli, Approximation by superposition of sigmoidal and radial basis functions, Adv. in Appl. Math., 13 (1992) 350-373.

[64] H.N. Mhaskar, C.A. Micchelli, Degree of approximation by neural and translation networks with a single hidden layer, Adv. in Appl. Math., 16 (1995) 151-183.

[65] Y. Makovoz, Uniform approximation by neural networks, J. Approx. Theory 95 (2) (1998) 215-228.

[66] A. Pinkus, Approximation theory of the MLP model in neural networks, Acta Numer. 8 (1999) 143-195.

[67] E.D. Sontag, Feedforward nets for interpolation and classification, J. Comp. Syst. Sci. 45 (1992) 20-48.

[68] G. Vinti, L. Zampogni, Approximation results for a general class of Kantorovich type operators, Advanced Nonlinear Studies, 14 (2014) 991-1011.