VILNIUS
TECH
Vilnius Gediminas
Technical University

Density Results by Deep Neural Network Operators with Integer Weights

Danilo Costarelli

Department of Mathematics and Computer Science, University of Perugia
Via Vanvitelli, 06123 Perugia, Italy
E-mail(*corresp.*): `danilo.costarelli@unipg.it`

Abstract. In the present paper, a new family of multi-layers (deep) neural network (NN) operators is introduced. Density results have been established in the space of continuous functions on $[-1, 1]$, with respect to the uniform norm. First, the case of the operators with two-layers is considered in detail, then the definition and the corresponding density results have been extended to the general case of multi-layers operators. All the above definitions allow us to prove approximation results by a constructive approach, in the sense that, for any given f all the weights, the thresholds, and the coefficients of the deep NN operators can be explicitly determined. Finally, examples of activation functions have been provided, together with graphical examples. The main motivation of this work resides in the aim to provide the corresponding multi-layers version of the well-known (shallow) NN operators, according to what is done in the applications with the construction of deep neural models.

Keywords: deep neural networks, neural network operators, density results, ReLU activation function, RePUs activation functions, sigmoidal functions.

AMS Subject Classification: 41A25; 41A05.

1 Introduction

The study of deep neural networks (NNs) currently represents one of the most studied topic (see, e.g., [25, 39, 40]), in view of the possible implications regarding several application aspects ([26, 28]), including artificial intelligence and machine learning. A very complete overview on the topic can be found, e.g., in [35].

In particular, it is known since the end of the 80's that among the main tasks that can be performed by NNs there is the functions approximation. For

this regard, in the literature can be found a wide number of articles, in particular for what concerns the approximation by one-layer (shallow) NNs. One of the pioneering work in this sense is due to Cybenko [20], who established a uniform approximation result in case of NNs activated by sigmoidal function. Cybenko's approximation theorem was proved by resorting to non-constructive arguments, and exploiting the well-known Hahn-Banach theorem of functional analysis. More precisely, he proved that the vector space containing one-layer NNs activated by continuous and non-decreasing sigmoidal functions is dense with respect to the uniform topology in the space of continuous functions defined on the multivariate set $[0,1]^d := [0,1] \times \cdots \times [0,1] \subset \mathbb{R}^d$, $d \geq 1$. In other words, he claims that a continuous function of several variables can be uniformly approximated by any degree of accuracy by the superposition of certain (univariate) sigmoidal functions using a sufficiently high number of artificial neurons. The latter fact is known in the literature as the so-called "universal approximation property" in the spirit of the well-known Kolmogorov approximation theorem (see, e.g., [27]).

Later on, several authors studied the above problem; the peculiarity of many of them was that they are faced by non-constructive techniques. Among them we can quote, e.g., [5, 31, 32, 33].

However, especially for the applications, it can be important to have at disposal constructive approximation algorithms; this has motivated several research in this direction, that mainly concerned shallow NNs (see, e.g., [11, 29] for some old papers, or, e.g., [8, 9, 21] for some more recent ones).

For instance, the theory of neural network (NN) operators introduced with the work of Cardaliaguet and Euvrard [10] is one of the possible ways to approach the above problem. The main limit of the theory proposed in [10] was that it could be applied only in case of bell-shaped activation functions with compact support. This limit has been overcome in [8], where NN operators (of one-variable) activated by the logistic function have been considered. Subsequently, in [16, 17] the theory of the NN operators has been formulated (for functions of one and several variables) for a wide class of sigmoidal activation functions (including, of course, the logistic function) satisfying suitable assumptions.

Even if the theory of NN operators have been studied in depth by several authors, only the one-layer case has been considered, see, e.g., [1, 2, 3, 4, 15, 19, 24, 37].

For the above reason, in the present paper we introduce a family of multi-layers (deep) NN operators, and we provide constructive approximation results. More precisely, we provide density theorems for the family of multi-layers (deep) NN operators in the space of continuous functions on the interval $[-1, 1]$.

The main motivation of this work resides in the aim to provide the corresponding multi-layers version of the well-known (shallow) NN operators, according to what is done in the applications with the construction of deep neural models.

From the mathematical point of view, the family of operators here introduced consists in the evaluation of the approximation performances of the nested application of families of positive linear operators.

This is done exploiting the "density approach", as typically occurs with NN type approximation, see, e.g., [22, 23]; due to this property one can usually claim that NNs are universal approximators.

The proposed approach can be defined "constructive" in the sense that for any given f all the weights, the thresholds, and the coefficients of the NN operators can be explicitly determined.

For the sake of clarity, in the paper we first propose the theory in case of two-layers operators (see Section 3), and then we propose its multi-layers generalization (see Section 4).

The above results have been proved in case of sigmoidal activation functions σ; all the required assumptions on σ, together with some preliminary considerations have been recalled in Section 2. However, other than sigmoidal activation functions, we also show that (see Section 5) deep NN operators based upon the well-known ReLU (Rectified Linear Unit, [38]) activation function, and also on RePUs (rectified powers units functions, [30]) are included in the present theory. Finally, graphical examples have been provided with the purpose of illustration.

2 Preliminaries

A measurable function $\sigma : \mathbb{R} \to \mathbb{R}$ is called a *sigmoidal function* if:

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

From now on, we always consider non-decreasing sigmoidal functions σ, satisfying the following assumptions:

($\Sigma 1$) $\sigma(x) - 1/2$ is an odd function;
($\Sigma 2$) $\sigma \in C^2(\mathbb{R})$ is concave for $x \geq 0$;
($\Sigma 3$) $\sigma(x) = \mathcal{O}(|x|^{-\alpha})$ as $x \to -\infty$, for some $\alpha > 0$.

Now, we can recall the definition of the density function (see, e.g., [16]):

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

Note that, based on the above properties, it is not difficult to see that $0 \leq \sigma'(x) \leq \sigma'(0)$, $x \in \mathbb{R}$, then it turns out that σ (and obviously also ϕ_σ) is Lipschitz continuous on $\mathbb{R}$.

Further, it is well-known (see [16]) that under the above assumptions the function ϕ_σ turns out to be non-negative, even, and it satisfies the following property:

$$\phi_\sigma(x) \text{ is non-decreasing for } x < 0 \text{ and non-increasing for } x \geq 0, \qquad (2.1)$$

with $\phi_\sigma(1) > 0$, and $\phi_\sigma(x) = \mathcal{O}(|x|^{-\alpha})$, as $x \to \pm\infty$, where α is the positive constant of condition ($\Sigma 3$). Hence, it turns out that $\phi_\sigma \in L^1(\mathbb{R})$ provided that $\alpha > 1$.

Moreover, based on the above assumptions, it is not difficult to see that (see [16] again):

$$1 \geq \sum_{k=-n}^{n} \phi_\sigma(nx - k) \geq \phi_\sigma(1) > 0, \qquad (2.2)$$

for every $x \in [-1, 1]$ and $n \in \mathbb{N}_+$.

Remark 1. Note that, if we remove condition $(\Sigma 2)$ on σ, and we assume directly that ϕ_σ satisfies (2.1) together with $\phi_\sigma(1) > 0$, the theory still holds, see e.g. [16]. The consequence of the above observation is that, we can apply the above theory to C^2 as well as to *non-smooth* sigmoidal functions, such that the corresponding ϕ_σ satisfies (2.1) and $\phi_\sigma(1) > 0$.

Now, we can define the discrete absolute moments of ϕ_σ, as follows:

$$M_\nu(\phi_\sigma) := \sup_{u \in \mathbb{R}} \sum_{k \in \mathbb{Z}} |\phi_\sigma(u - k)| \, |u - k|^\nu,$$

with $\nu \geq 0$.

It is well-known (see, e.g., [7]) that under the above assumptions, and assuming in addition that the parameter $\alpha > 1$, it turns out that:

$$M_\nu(\phi_\sigma) < +\infty, \qquad 0 \leq \nu < \alpha - 1. \tag{2.3}$$

Now, we recall the (shallow) neural network (NN) operators, introduced in [16].

Let σ be a sigmoidal function assumed as above. We define the (shallow) NN operators, by:

$$(F_n f)(x) = \frac{\sum\limits_{k=-n}^{n} f(k/n) \, \phi_\sigma(nx - k)}{\sum\limits_{k=-n}^{n} \phi_\sigma(nx - k)}, \qquad x \in I := [-1, 1],$$

$n \in \mathbb{N}_+$, where $f : I \to \mathbb{R}$ is bounded.

Notice that, F_n are well-defined; this is a consequence of the previous properties of the function ϕ_σ. Further, it is well-known ([16]) the following convergence theorem holds.

Theorem 1. *Let σ be a sigmoidal function satisfying assumption $(\Sigma 3)$ with $\alpha > 1$, and let $f \in C(I)$ be fixed. Then:*

$$\lim_{n \to +\infty} \|F_n f - f\|_\infty = 0.$$

3   Two-layers NN operators: approximation results

We now introduce the following family of deep NN operators.

DEFINITION 1. Let σ be a sigmoidal function, assumed as in Section 2. We introduce the two-layers (deep) neural network operators as follows:

$$(D^2_{(n_1,n_2)}f)(x) := \frac{\sum\limits_{k=-n_1}^{n_1} f\left(\frac{k}{n_1}\right) \phi_\sigma\left(n_1 \frac{\sum\limits_{j=-n_2}^{n_2} \frac{j}{n_2} \phi_\sigma(n_2 x - j)}{\sum_{j=-n_2}^{n_2} \phi_\sigma(n_2 x - j)} - k\right)}{\sum\limits_{k=-n_1}^{n_1} \phi_\sigma(n_1 x - k)}, \qquad x \in I,$$

with $f : I \to \mathbb{R}$, and $n_1, n_2 \in \mathbb{N}_+$.

Note that, the operators $D^2_{(n_1,n_2)}f$ turns out to be well-defined, e.g., for bounded functions. Indeed, using (2.2), we have:

$$|(D^2_{(n_1,n_2)}f)(x)| \leq \frac{1}{\phi_\sigma(1)} \left| \sum_{k=-n_1}^{n_1} f\left(\frac{k}{n_1}\right) \phi_\sigma\left(n_1 \frac{\sum_{j=-n_2}^{n_2} \frac{j}{n_2}\phi_\sigma(n_2x-j)}{\sum_{j=-n_2}^{n_2}\phi_\sigma(n_2x-j)} - k\right)\right|$$

$$\leq \frac{\|f\|_\infty}{\phi_\sigma(1)} \sum_{k=-n_1}^{n_1} \phi_\sigma\left(n_1 \frac{\sum_{j=-n_2}^{n_2} \frac{j}{n_2}\phi_\sigma(n_2x-j)}{\sum_{j=-n_2}^{n_2}\phi_\sigma(n_2x-j)} - k\right)$$

$$\leq \frac{\|f\|_\infty}{\phi_\sigma(1)} \sup_{u\in\mathbb{R}} \sum_{k\in\mathbb{Z}}\phi_\sigma(u-k) = \frac{\|f\|_\infty}{\phi_\sigma(1)} M_0(\phi_\sigma) < +\infty,$$

for every $x \in I$, and $n_1, n_2 \in \mathbb{N}_+$, where $M_0(\phi_\sigma) < +\infty$ in view of what is observed in (2.3).

Remark 2. We can observe that the two-layers NN operators are in fact deep NNs, organized in two layers. This can be simply seen observing that, in fact, the above operators are constructed by a mathematical composition of two, one-layer NNs. In practice, two simple NNs have been nested to get a more complex architecture composed of $(2n_1+1) + (2n_2+1)$ artificial neurons.

Since the structure of the considered operators is slightly different from classical NNs, below we can provide a comparison of $D^2_{(n_1,n_2)}f$ (for any fixed f) with classical feed-forward neural network models. Indeed, one and two layer classical feed-forward NNs activated by a sigmoidal function σ are defined (from the mathematical point of view) by:

$$N_1(x) := \sum_{k_1=0}^{n_1} a_{k_1}\sigma(w_{k_1}x - \theta_{k_1}), \quad N_2(x) := \sum_{k_2=0}^{n_2} a_{k_2}\sigma(w_{k_2}N_1(x) - \theta_{k_2}),$$

respectively. In the above definitions, w_{k_1} and w_{k_2} are the weights, θ_{k_1} and θ_{k_2} are thresholds, and finally a_{k_1} and a_{k_2} are the coefficients of the networks. Comparing N_2 with $D^2_{(n_1,n_2)}f$ we can observe what follows.

(i) In the inner layer of $D^2_{(n_1,n_2)}f$ (with $2n_2+1$ neurons, that can be compared with N_1), we have the network:

$$\sum_{j=-n_2}^{n_2} \frac{j}{n_2}\phi_\sigma(n_2x-j) \Big/ \sum_{j=-n_2}^{n_2}\phi_\sigma(n_2x-j),$$

where the integer parameter n_2 represents the (constant) weights, the parameter j (with $j = -n_2, ..., n_2$) provide the threshold values (or bias), while the values:

$$\frac{s}{n_2 \sum_{j=-n_2}^{n_2}\phi_\sigma(n_2x-j)}, \quad s = -n_2, ..., n_2\,{}^1, \tag{3.1}$$

are the coefficients of the inner network.

---

[1] Here we denoted the index at the numerator with the parameter s (instead of j as in Definition 1) in order to avoid confusion with the parameter j at the denominator.

(ii) Concerning the second layer of $D^2_{(n_1,n_2)}f$ with $2n_1 + 1$ neurons, we have that the integer parameter n_1 represents the weights, k (with $k = -n_1, ..., n_1$) are the threshold values, and finally:

$$\frac{f(s/n_1)}{\sum_{k=-n_1}^{n_1} \phi_\sigma(n_1 x - k)}, \quad s = -n_1, ..., n_1{}^2, \tag{3.2}$$

are the coefficients.

(iii) The activation functions of the operators $D^2_{(n_1,n_2)}f$ are defined using sigmoidal functions, as happens in N_1 and N_2.

We can note that the denominators in the definition of the above operators (that is crucial from the mathematical point of view) can be seen (see Equations (3.1) and (3.2)) as a part of the coefficients of both layers of $D^2_{(n_1,n_2)}$. This can be interpreted as the presence of a further connection among the various neurons of each layer. More precisely, the above denominators show that the input layer has a double (direct) connection with both the involved layers of the network.

Obviously, in view of the above comparison we can finally observe that the NN operators $D^2_{(n_1,n_2)}$ can not be viewed as *usual* two-layer NNs, since coefficients stated in (3.1) and (3.2) are not constants, as instead happens for N_1 and N_2.

Now, we can prove the following density theorem for the family of two-layers NN operators into the space $C(I)$ of all real valued continuous functions on I, with respect to the usual norm $\| \cdot \|_\infty$.

Theorem 2. *Let σ be a sigmoidal function satisfying assumption $(\Sigma 3)$ with $\alpha > 1$, and let $f \in C(I)$ be fixed. For any $\varepsilon > 0$ there exist $n_1, n_2 \in \mathbb{N}_+$ such that:*

$$\|D^2_{(n_1,n_2)}f - f\|_\infty < \varepsilon.$$

Proof. Let $\varepsilon > 0$. Then, using Theorem 1 we know that, corresponding to $\varepsilon/2$ there exists $n_1 \in \mathbb{N}_+$ sufficiently large such that:

$$\|F_{n_1}f - f\|_\infty < \varepsilon/2. \tag{3.3}$$

Now, we can also choose $n_2 \in \mathbb{N}_+$ such that:

$$\|F_{n_2}\theta - \theta\|_\infty < \frac{\varepsilon}{2\|f\|_\infty n_1(2n_1 + 1)|\sigma'(0)|} \phi_\sigma(1),$$

where the function $\theta(x) := x$, $x \in I$. Now, for every fixed $x \in I$ we have:

$$|(D^2_{(n_1,n_2)}f)(x) - f(x)|$$
$$\leq |(D^2_{(n_1,n_2)}f)(x) - (F_{n_1}f)(x)| + |(F_{n_1}f)(x) - f(x)| =: I_1 + I_2.$$

_____

[2] Here, we used the parameter s at the numerator (instead of k), for the same reason as before.

Concerning I_1, recalling that ϕ_σ is Lipschitz continuous with Lipschitz constant $|\sigma'(0)|$, we can write what follows:

$$
I_1 = \frac{\left| \displaystyle\sum_{k=-n_1}^{n_1} f\left(\frac{k}{n_1}\right) \left[\phi_\sigma\left(n_1 \frac{\sum_{j=-n_2}^{n_2} \frac{j}{n_2}\phi_\sigma(n_2 x - j)}{\sum_{j=-n_2}^{n_2} \phi_\sigma(n_2 x - j)} - k\right) - \phi_\sigma(n_1 x - k) \right] \right|}{\sum_{k=-n_1}^{n_1} \phi_\sigma(n_1 x - k)}
$$

$$
\leq \frac{\|f\|_\infty}{\phi_\sigma(1)} \sum_{k=-n_1}^{n_1} \left| \phi_\sigma\left(n_1 \frac{\sum_{j=-n_2}^{n_2} \frac{j}{n_2}\phi_\sigma(n_2 x - j)}{\sum_{j=-n_2}^{n_2} \phi_\sigma(n_2 x - j)} - k\right) - \phi_\sigma(n_1 x - k) \right|
$$

$$
\leq \frac{\|f\|_\infty}{\phi_\sigma(1)} |\sigma'(0)| \sum_{k=-n_1}^{n_1} n_1 \left| \frac{\sum_{j=-n_2}^{n_2} \frac{j}{n_2}\phi_\sigma(n_2 x - j)}{\sum_{j=-n_2}^{n_2} \phi_\sigma(n_2 x - j)} - x \right|
$$

$$
= \frac{\|f\|_\infty}{\phi_\sigma(1)} |\sigma'(0)| \, n_1 \, (2\,n_1 + 1) \, |(F_{n_2}\theta)(x) - \theta(x)| \;<\; \varepsilon/2.
$$

Hence the proof follows by observing that also $I_2 < \varepsilon/2$ in view of (3.3). $\quad\square$

Remark 3. Note that the density approach is one of the most common when one deals with NN type approximation. In this sense, we can refer, e.g., to the Cybenko approximation theorem [20] (inspired by the well-known Kolmogorov representation theorem).

4 Multi-layers NN operators

In Section 3, we introduced and studied a family of deep NN operators with two layers. In this section, the above definition and the corresponding density results will be extended to the more general context of multi-layers NN operators. The idea is to extend the above definition proceeding by induction on the number of layers.

DEFINITION 2. Let σ be a sigmoidal function, assumed as in Section 2. We define the m-layers (deep) NN operators, $m \in \mathbb{N}$, $m \geq 2$, as follows:

$$
(D^m_{\underline{n}_{[m]}} f)(x) = \frac{\displaystyle\sum_{k=-n_1}^{n_1} f\left(\frac{k}{n_1}\right) \phi_\sigma\left(n_1 \, (D^{m-1}_{\underline{n}_{[m-1]}}\theta)(x) - k\right)}{\displaystyle\sum_{k=-n_1}^{n_1} \phi_\sigma(n_1 x - k)}, \qquad x \in I,
$$

where $f : I \to \mathbb{R}$, $\theta(x) := x$, $x \in I$, $\underline{n}_{[m]} \in \mathbb{N}_+^m$, with $\underline{n}_{[m]} = (n_1, n_2, ..., n_m)$, $\underline{n}_{[m-1]} \in \mathbb{N}_+^{m-1}$, with $\underline{n}_{[m-1]} = (n_2, n_3, ..., n_m)$, and:

$$
(D^1_{\underline{n}_{[1]}}\theta)(x) := (F_{n_m}\theta)(x),
$$

$$(D^2_{\underline{n}_{[2]}}\theta)(x) \; := \; \frac{\displaystyle\sum_{k=-n_{m-1}}^{n_{m-1}} \frac{k}{n_{m-1}} \phi_\sigma(n_{m-1}(D^1_{\underline{n}_{[1]}}\theta)(x) - k)}{\displaystyle\sum_{k=-n_{m-1}}^{n_{m-1}} \phi_\sigma(n_{m-1}x - k)},$$

$$\cdots$$

$$(D^m_{n_{[m-1]}}\theta)(x) \; := \; \frac{\displaystyle\sum_{k=-n_2}^{n_2} \frac{k}{n_2} \phi_\sigma(n_2(D^{m-1}_{n_{[m-2]}}\theta)(x) - k)}{\displaystyle\sum_{k=-n_2}^{n_2} \phi_\sigma(n_2 x - k)}, \qquad m \in \mathbb{N}.$$

Clearly, for $m = 2$ the above deep NN operators coincide with those considered in Section 3.

Further, we can observe that the multi-layers NN operators are well-defined for every $m \geq 2$, for any bounded function f. Indeed, it is easy to see that:

$$|(D^m_{\underline{n}_{[m]}}f)(x)| \leq \frac{\|f\|_\infty}{\phi_\sigma(1)} \sum_{k=-n_1}^{n_1} \phi_\sigma\left(n_1\left(D^m_{\underline{n}_{[m-1]}}\theta\right)(x) - k\right)$$

$$\leq \frac{\|f\|_\infty}{\phi_\sigma(1)} M_0(\phi_\sigma) < +\infty,$$

since $(D^m_{\underline{n}_{[m-1]}}\theta)(x) \in \mathbb{R}$, for every $x \in I$ and $\underline{n}_{[m]} \in \mathbb{N}^m_+$. Now, we are able to prove the following density result for the family of multi-layers (deep) NN operators in the space $C(I)$ with respect to the uniform norm.

Theorem 3. *Let σ be a sigmoidal function satisfying assumption (Σ3) with $\alpha > 1$. Further, let $m \in \mathbb{N}$, $m \geq 2$ and $f \in C(I)$ be fixed. Then, for every $\varepsilon > 0$ there exists $\underline{n}_{[m]} \in \mathbb{N}^m_+$ such that:*

$$\|D^m_{\underline{n}_{[m]}}f - f\|_\infty < \varepsilon.$$

Proof. We proceed by induction on $m \geq 2$. In the case $m = 2$ the density result immediately follows by Theorem 2. We now suppose that the thesis holds for $m - 1 \geq 2$. Let now $f \in C(I)$ and $\varepsilon > 0$ be fixed. In view of Theorem 1, we can choose $\widetilde{n} \in \mathbb{N}_+$ such that:

$$\|F_{\widetilde{n}}f - f\|_\infty < \varepsilon/2.$$

Further, by the inductive assumption, in correspondence to $\varepsilon/2$ there exists $\underline{n}_{[m-1]} \in \mathbb{N}^{m-1}_+$, with $\underline{n}_{[m-1]} = (n_1, n_2, ..., n_{m-1})$, such that:

$$\|D^{m-1}_{\underline{n}_{[m-1]}}\theta - \theta\|_\infty \; < \; \frac{\varepsilon\,\phi_\sigma(1)}{2\,\|f\|_\infty\,|\sigma'(0)|\widetilde{n}(2\widetilde{n}+1)}.$$

Now, we set $\underline{n}_{[m]} = (\widetilde{n}, n_1, n_2, ..., n_{m-1}) \in \mathbb{N}^m_+$. Thus, in the case of m-layers NN operators we can write what follows:

$$\|D^m_{\underline{n}_{[m]}}f - f\|_\infty \leq \|D^m_{\underline{n}_{[m]}}f - F_{\widetilde{n}}f\|_\infty + \|F_{\widetilde{n}}f - f\|_\infty < \|D^m_{\underline{n}_{[m]}}f - D^{m-1}_{\underline{n}_{[m-1]}}f\|_\infty + \varepsilon/2.$$

Let now $x \in I$ be fixed. Using the fact that ϕ_σ is Lipschitz continuous with Lipschitz constant $|\sigma'(0)|$, we immediately have:

$$\left| \frac{\sum_{k=-\widetilde{n}}^{\widetilde{n}} f\left(\frac{k}{\widetilde{n}}\right) \phi_\sigma\left(\widetilde{n}\left(D_{\underline{n}_{[m-1]}}^{m-1}\theta\right)(x) - k\right)}{\sum_{k=-\widetilde{n}}^{\widetilde{n}} \phi_\sigma(\widetilde{n}x - k)} - \frac{\sum_{k=-\widetilde{n}}^{\widetilde{n}} f\left(\frac{k}{\widetilde{n}}\right) \phi_\sigma(\widetilde{n}\,x - k)}{\sum_{k=-\widetilde{n}}^{\widetilde{n}} \phi_\sigma(\widetilde{n}x - k)} \right|$$

$$\leq \frac{|\sigma'(0)|\,\|f\|_\infty}{\phi_\sigma(1)} \sum_{k=-\widetilde{n}}^{\widetilde{n}} \widetilde{n}\,\left|\left(D_{\underline{n}_{[m-1]}}^{m-1}\theta\right)(x) - x\right| < \varepsilon/2.$$

This completes the proof. □

5 Activation functions and graphical examples

As a first example, we study in detail the case of the deep NN operators activated by the well-known logistic function:

$$\sigma_\ell(x) := \left(1 + e^{-x}\right)^{-1}, \qquad x \in \mathbb{R}.$$

It is well-known (see, e.g., [16, 19]) that the logistic function is a Lipschitz continuous function satisfying ($\Sigma 1$), ($\Sigma 2$) and ($\Sigma 3$). In particular, due to its exponential decay to zero as $x \to -\infty$, σ_ℓ fulfills condition ($\Sigma 3$) for every $\alpha > 0$.

Based on the above considerations, we deduce that in case of the logistic function, the density results established in the previous section hold, hence we can formulate what follows.

Corollary 1. Let σ_ℓ be the logistic function, $m \in \mathbb{N}$, $m \geq 2$, and $f \in C(I)$ be fixed. Then, denoting by $D_{\underline{n}_{[m]}}^{\sigma_\ell,m}$ the m-layers NN operators activated by σ_ℓ, for every $\varepsilon > 0$ there exists $\underline{n}_{[m]} \in \mathbb{N}_+^m$ such that:

$$\|D_{\underline{n}_{[m]}}^{\sigma_\ell,m} f - f\|_\infty < \varepsilon.$$

As a second example, we consider the sigmoidal function $\sigma_{M_d}(x)$ (introduced in [18]) that are associated to the central B-spline ([6, 12]):

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

The sigmoidal function $\sigma_{M_d}(x)$ ([18]) are defined by the following integral-formula:

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

and the corresponding density functions assume the following expression:

$$\phi_{\sigma_{M_d}}(x) := \frac{1}{2}[\sigma_{M_d}(x+1) - \sigma_{M_d}(x-1)]$$
$$= \frac{1}{2} \int_{x-1}^{x+1} M_s(t)\,dt = \frac{1}{2} \int_{-1}^{+1} M_s(y+x)\,dy, \qquad x \in \mathbb{R}.$$

In [18] it has been proved that $\sigma_{M_d}(x)$ satisfies all the assumptions required in Section 2. In particular, since the central B-spline has compact support, the $\sigma_{M_d}(x) = 0$, for every $x < -T$, where $T > 0$ is a suitable constant. This shows that assumption $(\Sigma 3)$ is satisfied by $\sigma_{M_d}(x)$, for every $\alpha > 0$.

As a consequence, the following corollary can be formulated.

Corollary 2. Let σ_{M_d}, $d \in \mathbb{N}$, $m \in \mathbb{N}$, $m \geq 2$, and $f \in C(I)$ be fixed. Then, denoting by $D^{\sigma_{M_d}, m}_{\underline{n}_{[m]}}$ the m-layers NN operators activated by σ_{M_d}, for every $\varepsilon > 0$ there exists $\underline{n}_{[m]} \in \mathbb{N}^m_+$ such that:

$$\|D^{\sigma_{M_d}, m}_{\underline{n}_{[m]}} f - f\|_\infty \ < \ \varepsilon.$$

Finally, we also observe that, in case of $d = 1$:

$$\sigma_{M_1}(x) \ =: \ \sigma_R(x) \ = \ \begin{cases} 0, & x < -1/2, \\ x + (1/2), & -1/2 \leq x \leq 1/2, \\ 1, & x > 1/2. \end{cases}$$

The function $\sigma_{M_1}(x)$ is also known with the name of the ramp function, see e.g., [9,13,14]. Now, if we recall the definition of the well-known ReLu activation function (see, e.g., [21]): $\psi_{ReLu}(x) := (x)_+$, $x \in \mathbb{R}$, it turns out that:

$$\sigma_{M_1}(x) \ := \ \psi_{ReLu}(x + 1/2) - \psi_{ReLu}(x - 1/2),$$

then, the corresponding density function can be expressed in term of ReLu activation function:

$$\phi_{\sigma_R}(x) \ = \phi_{\sigma_{M_1}}(x) = \frac{1}{2} \left[\psi_{ReLu}(x + 3/2) \right.$$
$$\left. - \psi_{ReLu}(x + 1/2) - \psi_{ReLu}(x - 1/2) + \psi(x - 3/2) \right].$$

As a consequence of the above relation, the deep NN operators $D^{m, \sigma_{M_1}}_n$ can be considered as a deep NN activated by the above linear combination of ReLu activation function. For more details concerning the usefulness of ψ_{ReLu}, see, e.g., [38].

Note that, central B-splines, and consequently the corresponding σ_{M_d}, can be easily expressed in terms of powers of ψ_{ReLU}. In the theory of NNs, powers of the ReLU function are known with the name of rectified power unit functions (RePUs). Hence, reasoning in a similar way to the case of ReLu function, it is clear that the above results hold also in case of deep NN operators activated by suitable combinations of RePUs activation functions (see, e.g., [30]).

Finally, we provide the following graphical examples with the main purpose to illustrate the approximation performances of the deep NN operators. For instance, here we consider the continuous function $f(x) = x^2 - 1 + \sin x$, on the interval $[-1, 1]$; approximation of f by means of the 2-layers and 3-layers NN operators activated by the logistic function have been considered in Figure 1.

6   Conclusions

The multi-layers (deep) neural network operators introduced and studied in the present paper allows to establish *constructive* approximation results by a family
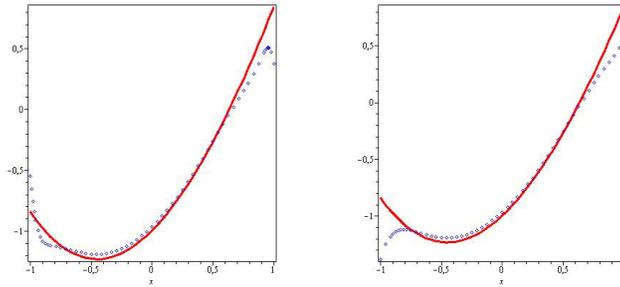
Figure 1. On the left: the plots of the function f (red line) and of the 2-layers NN operators $D^2_{(10,10)}f$ (blue dots) activated by the logistic function. On the right: the plots of the function f (red line) and of the 3-layers NN operators $D^3_{(10,10,10)}f$ (blue dots) activated by the logistic function.

of deep neural networks. The present theory deals with the approximation of functions of one-variable. It is well-known that the theory of artificial neural networks is mainly a multivariate theory; indeed the present results lead the way to the introduction of the corresponding version of deep NN operators for approximating functions of several variables. This will be done in a future work following the strategy depicted in [17].

In recent years, the theory of multi-layer NNs has been deeply studied, see, e.g., [34], in view of its wide importance in both theoretical and applied fields. Concerning very recent approximation results (that can be interpreted as density theorems) for the above tools, one can see, e.g., [36]. In the latter paper, the authors established the order of approximation for deep (two or multi-layer) NNs activated by ReLU functions, in case of multivariate *Hölder* continuous functions. One of the main aspects arising from the above result was that the approximation error depends not only from the number of the considered neurons but also from the depth of the net, i.e., by the number of considered layers. Indeed, increasing the number of the layers in the NNs the accuracy of the approximation improves.

In view of the importance covered by the above topic, in the present paper we decided to introduce a multi-layer (deep) version of the so-called NN operators. Here, only the density problem has been treated, however the considered results are not limited to the case of the ReLU function, but hold also for a class of activation functions including RePUs and sigmoidal functions. Obviously, the role that the additional layers of the NN operators play for the degree of accuracy of the achieved approximations will be the object of a future investigation. Actually, the graphical results shown in the previous section seems to suggest the fact that, also in this case, by increasing the number of layers, the corresponding order of approximation improves. Finally, also the problem of best approximation will be considered in detail in a future study.

Acknowledgements

References

[1] G.A. Anastassiou.   Rate of convergence of some neural network operators to the unit-univariate case.   *J. Math. Anal. Appl.*, **212**(1):237–262, 1997. https://doi.org/10.1006/jmaa.1997.5494.

[2] G.A. Anastassiou. *Intelligent Systems II: Complete Approximation by Neural Network Operators.* Studies in Computational Intelligence book series (SCI, volume 608), Springer, Cham, 2016. https://doi.org/10.1007/978-3-319-20505-2.

[3] S. Bajpeyi and A. Sathish Kumar. Max-product type exponential neural network operators. In *International Conference on Mathematical Analysis and Computing, ICMAC 2019*, Mathematical Analysis and Computing, pp. 561–571, 2019. https://doi.org/10.1007/978-981-33-4646-8_44.

[4] S. Bajpeyi and A. Sathish Kumar. Approximation by exponential sampling type neural network operators. *Analysis and Mathematical Physics*, **11**(108), 2021. https://doi.org/10.1007/s13324-021-00543-y.

[5] A.R. Barron. Universal approximation bounds for superpositions of a sigmoidal function.   *IEEE Transactions on Information Theory*, **39**(3):930–945, 1993. https://doi.org/10.1109/18.256500.

[6] P. Bohra, J. Campos, H. Gupta, S. Aziznejad and M. Unser. Learning activation functions in deep (spline) neural networks. *IEEE Open Journal of Signal Proc.*, **1**:295–309, 2020. https://doi.org/10.1109/OJSP.2020.3039379.

[7] P.L. Butzer and R.J. Nessel. *Fourier Analysis and Approximation I.* Academic Press, New York-London, 1971. https://doi.org/10.1007/978-3-0348-7448-9.

[8] F. Cao and Z. Chen.   The approximation operators with sigmoidal functions.   *Comput. Math. Appl.*, **58**(4):758–765, 2009. https://doi.org/10.1016/j.camwa.2009.05.001.

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

[10] P. Cardaliaguet and G. Euvrard.   Approximation of a function and its derivative with a neural network.   *Neural Netw.*, **5**(2):207–220, 1992. https://doi.org/10.1016/S0893-6080(05)80020-6.

[11] H. Chen, T. Chen and R. Liu. A constructive proof and an extension of Cybenko's approximation theorem. In Connie Page and Raoul LePage(Eds.), *Computing Science and Statistics*, pp. 163–168, New York, NY, 1992. Springer New York. https://doi.org/10.1007/978-1-4612-2856-1_21.

[12] L. Coroianu and S.G. Gal. Approximation by truncated max-product operators of Kantorovich-type based on generalized (φ, ψ)-kernels. *Math. Meth. Appl. Sci.*, **41**(17):7971–7984, 2018. https://doi.org/10.1002/mma.5262.

[13] D. Costarelli. Interpolation by neural network operators activated by ramp functions. *J. Math. Anal. Appl.*, **419**(1):574–582, 2014. https://doi.org/10.1016/j.jmaa.2014.05.013.

[14] D. Costarelli. Neural network operators: constructive interpolation of multivariate functions. *Neural Netw.*, **67**:28–36, 2015. https://doi.org/10.1016/j.neunet.2015.02.002.

[15] D. Costarelli and A.R. Sambucini. Approximation results in orlicz spaces for sequences of Kantorovich max-product neural network operators. *Results in Mathematics*, **73**(1):15, 2018. https://doi.org/10.1007/s00025-018-0799-4.

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

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

[18] D. Costarelli and R. Spigler. Approximation by series of sigmoidal functions with applications to neural networks. *Annali Mat. Pura Appl.*, **194**(1):289–306, 2015. https://doi.org/10.1007/s10231-013-0378-y.

[19] D. Costarelli and G. Vinti. Saturation classes for max-product neural network operators activated by sigmoidal functions. *Results Math.*, **72**(3):1555–1569, 2017. https://doi.org/10.1007/s00025-017-0692-6.

[20] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems*, **2**:303–314, 1989. https://doi.org/10.1007/BF02551274.

[21] S. Goebbels. On sharpness of error bounds for univariate approximation by single hidden layer feedforward neural networks. *Results Math.*, **75**:109, 2020. https://doi.org/10.1007/s00025-020-01239-8.

[22] K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Netw.*, **4**(2):251–257, 1991. https://doi.org/10.1016/0893-6080(91)90009-T.

[23] K. Hornik, M. Stinchcombe and H. White. Multilayer feedforward networks are universal approximators. *Neural Netw.*, **2**(5):359–366, 1989. https://doi.org/10.1016/0893-6080(89)90020-8.

[24] U. Kadak. Fractional type multivariate neural network operators. *Math. Meth. Appl. Sci.*, 2021. https://doi.org/10.1002/mma.7460.

[25] P.C. Kainen, V. Kurková and A. Vogt. Approximative compactness of linear combinations of characteristic functions. *J. Approx. Theory*, **257**:105435, 2020. https://doi.org/10.1016/j.jat.2020.105435.

[26] M. Kohler and A. Krzyzak. Over-parametrized deep neural networks do not generalize well. *ArXiv preprint arxiv:1912.03925*, 2020.

[27] V. Kurková. Kolmogorov's theorem and multilayer neural networks. *Neural Netw.*, **5**(3):501–506, 1992. https://doi.org/10.1016/0893-6080(92)90012-8.

[28] V. Kurková and M. Sanguineti. Classification by sparse neural networks. *IEEE Trans. on Neural Netw. Learning Syst.*, **30**(9):2746–2754, 2019. https://doi.org/10.1109/TNNLS.2018.2888517.

[29] B. Lenze. *Constructive Multivariate Approximation with Sigmoidal Functions and Applications to Neural Networks*, pp. 155–175. Birkhäuser Basel, Basel, 1992. https://doi.org/10.1007/978-3-0348-8619-2_9.

[30] B. Li, S. Tang and H. Yu. Approximations of high dimensional smooth functions by deep neural networks with rectified power units. *arXiv:1903.05858v4*, 2019.

[31] X. Li. Simultaneous approximations of multivariate functions and their derivatives by neural networks with one hidden layer. *Neurocomputing*, **12**:327–343, 1996. https://doi.org/10.1016/0925-2312(95)00070-4.

[32] X. Li. On simultaneous approximations by radial basis function neural networks. *Appl. Math. Comput.*, **95**:75–89, 1998. https://doi.org/10.1016/S0096-3003(97)10089-3.

[33] H.N. Mhaskar and C.A. Micchelli. Approximation by superposition of sigmoidal and radial basis functions. *Adv. Appl. Math.*, **13**(32):350–373, 1992. https://doi.org/10.1016/0196-8858(92)90016-P.

[34] W. Samek, G. Montavon, S. Lapuschkin, C.J. Anders and K.-R. Müller. Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, **109**(3):247–278, 2021. https://doi.org/10.1109/JPROC.2021.3060483.

[35] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Netw.*, **61**:85–117, 2015. https://doi.org/10.1016/j.neunet.2014.09.003.

[36] Z. Shen, H. Yang and S. Zhang. Optimal approximation rate of ReLU networks in terms of width and depth. *Journal de Mathématiques Pures et Appliquées*, **157**:101–135, 2022. https://doi.org//10.1016/j.matpur.2021.07.009.

[37] C. Turkun and O. Duman. Modified neural network operators and their convergence properties with summability methods. *RACSAM*, **114**:132, 2020. https://doi.org/10.1007/s13398-020-00860-0.

[38] D. Yarotsky. Error bounds for approximations with deep ReLU networks. *Neural Netw.*, **94**:103–114, 2017. https://doi.org/10.1016/j.neunet.2017.07.002.

[39] D.X. Zhou. Theory of deep convolutional neural networks: Downsampling. *Neural Netw.*, **124**:319–327, 2020. https://doi.org/10.1016/j.neunet.2020.01.018.

[40] D.X. Zhou. Universality of deep convolutional neural networks. *Appl. Comput. Harmonic Anal.*, **48**(2):787–794, 2020. https://doi.org/10.1016/j.acha.2019.06.004.