MODERN MATHEMATICAL METHODS
2 (2024), No. 2, pp. 90-102
https://modernmathmeth.com/
ISSN 3023 - 5294

Research Article

A Jackson-type estimate in terms of the τ-modulus for neural network operators in L^p-spaces

LORENZO BOCCALI, DANILO COSTARELLI*, AND GIANLUCA VINTI

ABSTRACT. In this paper, we study the order of approximation with respect to the L^p-norm for the (shallow) neural network (NN) operators. We establish a Jackson-type estimate for the considered family of discrete approximation operators using the averaged modulus of smoothness introduced by Sendov and Popov, also known by the name of τ-modulus, in the case of bounded and measurable functions on the interval $[-1, 1]$. The results here proved, improve those given by Costarelli (J. Approx. Theory 294:105944, 2023), obtaining a sharper approximation. In order to provide quantitative estimates in this context, we first establish an estimate in the case of functions belonging to Sobolev spaces. In the case $1 < p < +\infty$, a crucial role is played by the so-called Hardy-Littlewood maximal function. The case of $p = 1$ is covered in case of density functions with compact support.

Keywords: Neural network operators, averaged moduli of smoothness, Jackson-type estimates, sigmoidal functions, Hardy-Littlewood maximal function.

2020 Mathematics Subject Classification: 41A25, 41A05.

1. INTRODUCTION

The theory of feed-forward artificial neural networks (NNs) has been widely studied since the first half of the 1990s in view of their wide use in many applied fields such as artificial intelligence [15, 16, 19] and neuroscience. A crucial first impulse to that theory came from the introduction of an oversimplified model of the human brain, thanks to the contribution of McCulloch and Pitts [7] (1943). Actually, simplifying the real functioning of the most complex organ of the human body, McCulloch and Pitts view the brain as a collection of neurons represented by dots that communicate with each other through connections modeLled by lines in accordance with "threshold logic": if the magnitude of the incoming impulse exceeds a certain threshold value, the individual neuron sends its charge in response to the neurons to which it is connected; otherwise, it maintains its quiet status.

In recent decades, many authors have shown that neural networks can be successfully used for function approximation, finding interesting applications in Approximation Theory, as well in [1, 3, 4, 5, 18, 20, 22]. One of the purposes was to obtain constructive approximation algorithms based on NNs. In this direction, the pioneer result is the approximation theorem established by Cybenko [13] for single layer (shallow) NNs defined by the superposition of sigmoidal activation functions. Although this result is elegant from a mathematical point of view, since its proof is based on well-known theorems of functional analysis, such as the famous Hahn-Banach theorem, at the same time it has a strong non-constructive character that severely limits its use in the application domain.

Received: 10.06.2024; Accepted: 09.08.2024; Published Online: 14.08.2024
*Corresponding author: Danilo Costarelli; danilo.costarelli@unipg.it

In order to solve this problem, a constructive approximation approach was faced by Cardaliaguet and Euvrard in [6], where uniform approximation processes through NN operators activated by bell-shaped functions, exclusively with compact support, have been studied. The definition of Cardaliaguet-Euvrard NN operators has been improved in [10], where the authors introduced the following family of (shallow) NN operators activated by a general class of sigmoidal functions satisfying suitable assumptions:

$$(F_n f)(x) = \frac{\sum_{k=-n}^{n} f\left(\frac{k}{n}\right) \phi_\sigma(nx-k)}{\sum_{k=-n}^{n} \phi_\sigma(nx-k)}, \qquad x \in I := [-1,1]. \tag{I}$$

The approximation capabilities of the above operators have been extensively investigated in relation to the approximation of continuous functions, establishing a pointwise and uniform convergence theorem and qualitative and quantitative estimates about the rate of convergence with respect to the usual sup norm (see, e.g., [8, 10]).
Recently, the problem of studying the approximation properties of the operators defined in (I) in L^p-spaces and, thus in fact, their power in approximating not-necessarily continuous functions has been faced in [9]. Here, quantitative estimates for the aliasing error with respect to the L^p-norm for the operators F_n have been provided in terms of the well-known averaged modulus of smoothness introduced by Sendov and Popov [23], known by the name of τ-modulus. The use of τ-modulus is necessitated by the intrinsic pointwise nature of the operators considered, which depend strongly on the single values of the function to be approximated. Indeed, unlike the usual L^p-modulus of smoothness, which is not able to consider increments of f on sets with null-measure, the averaged modulus of smoothness allows us to estimate the approximation error for several families of discrete pointwise operators in case of not-necessarily continuous functions (see [23, Chapter 4]). Moreover, the above context forces us to work in an unaccustomed setting of L^p-setting, in which we do not identify functions coinciding almost everywhere; that is, we are not dealing with equivalence classes of functions, as usual. In fact, we separate each measurable and bounded function f on I, uniquely defined by the individual values assumed at each point $x \in I$, from its own equivalence class, as is often the case in the literature when similar problems of approximating L^p functions are faced (see, e.g., [2]).
In this frame, the goal of this paper is to show that, by resorting to the τ-modulus, we can improve the results proved in [9], establishing Jackson-type estimates with respect to the L^p-norm for the above NN operators. To reach the desired result, after recalling in Section 2 the definition of the involved operators and all the auxiliary results used in this paper, we will establish an estimate in the case of functions belonging to Sobolev spaces. In this proof, for $1 < p < +\infty$, a crucial role is played by the well-known Hardy-Littlewood maximal function. Instead, in order to get an analogous result also in the case $p = 1$, we consider activation functions with compact support. As a direct consequence of these results, exploiting the basic properties of the averaged modulus of smoothness, we are able to obtain Jackson-type estimations for the approximation of L^p-functions and thus a convergence theorem (already proved in [9]) can be deduced. Finally, we recall several examples of sigmoidal activation functions for which the above theory can be applied.

2. Preliminary Notions

We begin this section by recalling the definition of the neural network (NN) operators considered in this paper. A measurable function $\sigma : \mathbb{R} \to \mathbb{R}$ is called a sigmoidal function if

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

Using this definition, we can define the density function ϕ_σ associated with the operators, that we will consider below, that is:

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

Here, we consider σ as any non-decreasing sigmoidal function that satisfies the following conditions:
$(\Sigma 1)$ $\sigma(x)-1/2$ 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 $\alpha>1$.

The above assumptions on the sigmoidal function σ have been first introduced in [10] in order to obtain constructive approximation results for (shallows) NN operators that extend the theory developed in [6].

Definition 2.1. *Let σ be a sigmoidal function satisfying $(\Sigma 1)-(\Sigma 3)$. The corresponding family of (shallow) NN operators is defined by:*

$$(F_n f)(x):=\frac{\sum_{k=-n}^{n} f\left(\frac{k}{n}\right)\phi_\sigma(nx-k)}{\sum_{k=-n}^{n}\phi_\sigma(nx-k)}, \quad x\in I:=[-1,1],\ n\in\mathbb{N},$$

where $f:I\to\mathbb{R}$ is bounded.

Note that here we consider for simplicity the case of functions defined in $I=[-1,1]$, but the following result can also be proved for any general interval $[a,b]$. We summarize some well-known properties of ϕ_σ established in [10] in the following lemma.

Lemma 2.1. (i) *$\phi_\sigma(x)\geq 0$ for every $x\in\mathbb{R}$, with $\phi_\sigma(1)>0$, and moreover $\lim_{x\to\pm\infty}\phi_\sigma(x)=0$;*
(ii) *The function $\phi_\sigma(x)$ is even;*
(iii) *The function $\phi_\sigma(x)$ is non-decreasing for $x<0$ and non-increasing for $x\geq 0$;*
(iv) *Let α be the positive constant of condition $(\Sigma 3)$. Then:*

$$\phi_\sigma(x)=\mathcal{O}(|x|^{-\alpha}), \text{ as } x\to\pm\infty.$$

Thus, we have $\phi_\sigma\in L^1(\mathbb{R})$;
(v) *For every $x\in\mathbb{R}$,*

$$\sum_{k\in\mathbb{Z}}\phi_\sigma(x-k)=1, \textit{ and } \|\phi_\sigma\|_1=\int_{\mathbb{R}}\phi_\sigma(x)\,dx=1;$$

(vi) *Let $x\in I$ and $n\in\mathbb{N}$ be fixed. Then:*

$$\sum_{k=-n}^{n}\phi_\sigma(nx-k)\geq\phi_\sigma(1)>0. \tag{2.1}$$

The inequality (2.1) is of crucial importance to have that $F_n f$ is well-defined, since it guarantees that F_n has always a non-zero denominator. Moreover, for f bounded we also have:

$$|(F_n f)(x)| \leq \frac{\displaystyle\sum_{k=-n}^{n} \left| f\left(\frac{k}{n}\right)\right| \phi_\sigma(nx-k)}{\displaystyle\sum_{k=-n}^{n} \phi_\sigma(nx-k)} \leq \|f\|_\infty < +\infty, \; x \in I, \; n \in \mathbb{N},$$

where $\|\cdot\|_\infty$ denotes the usual max-norm on I. We now recall the following useful notion of the discrete absolute moment of order $\nu \geq 0$ of ϕ_σ (see, e.g., [10]), i.e.,

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

It is well-known, from the assumptions $(\Sigma 1)-(\Sigma 3)$ on σ (see [12, Lemma 2.6]), that it turns out that:

$$M_\nu(\phi_\sigma) < +\infty, \quad 0 \leq \nu < \alpha - 1,$$

where α is the constant of condition $(\Sigma 3)$. The latter result follows by (iv) of Lemma 2.1 and the boundedness of the density function ϕ_σ.
From now on, we consider the space $L^p(I)$, where we do not identify functions coinciding almost everywhere since our operators are sensitive to single function values and they could map different functions of the same equivalence class to different classes.
Now, we recall the following useful estimate (see [9, Lemma 4.1]) for the L^p-norm of the family of NN operators F_n in case of bounded and measurable functions.

Lemma 2.2. *Let σ be a sigmoidal function satisfying $(\Sigma 1)-(\Sigma 3)$. For every bounded function $f: I \to \mathbb{R}$, there holds:*

$$\|F_n f\|_p \leq \frac{\|f\|_{l^p(\Sigma_n)}}{\phi_\sigma(1)^{1/p}}, \quad 1 \leq p < +\infty,$$

where:

$$\Sigma_n := \left\{\frac{k}{n}, \; k = -n, ..., n\right\},$$

and

$$\|f\|_{l^p(\Sigma_n)} := \left\{ \sum_{k=-n}^{n} \left| f\left(\frac{k}{n}\right)\right|^p n^{-1} \right\}^{1/p},$$

denotes a discrete l^p norm of the function f on Σ_n, $n \in \mathbb{N}$.

In order to establish quantitative estimates for the rate of convergence for the above NN operators with respect to the L^p-norm, we will use the so-called averaged modulus of smoothness, also known by the name of τ-modulus, introduced in [23] by the Bulgarian school of the mathematician Sendov. However, before introducing its definition, we need to recall the notion of the well-known local modulus of smoothness [23] for bounded and measurable functions defined on I (see also [2]).
From now on, we denote by $M(I)$ the set of all bounded and measurable functions $f: I \to \mathbb{R}$.
Let $f \in M(I)$ be fixed. The local modulus of smoothness of order $r \in \mathbb{N}$ $(r \geq 1)$ of the function f at the point $x \in I$ is defined, for $0 < \delta \leq 2/r$, by:

$$\omega_r(f, x; \delta) := \sup\left\{ |(\Delta_h^r f)(t)| : \; t, \, t + rh \in \left[x - \frac{r\delta}{2}, x + \frac{r\delta}{2}\right] \cap I \right\},$$

where

$$(\Delta_h^r f)(t) := \sum_{j=0}^{r} \binom{r}{j} (-1)^{r-j} f(t+jh)$$

is the classical finite forward difference of order r of f with increment h at the point t. It is possible to prove (see [23, Theorem 1.3]) that the local modulus of smoothness $\omega_r(f,x;\delta)$ of $f \in M(I)$, considered as a function of $x \in I$, belongs to $M(I)$. Now, from the boundedness of $\omega_r(f,x;\delta)$ as a function of $x \in I$, it follows that it also belongs to $L^p(I)$, $1 \leq p < +\infty$. This fact justifies the introduction of the averaged modulus of smoothness (τ-modulus) of $f \in M(I)$ as the L^p-norm of the above local modulus of smoothness. Namely,

Definition 2.2. *Let $f \in M(I)$ be fixed. The averaged modulus of smoothness of order $r \in \mathbb{N}$ $(r \geq 1)$ (or τ-modulus) of the function f is defined by:*

$$\tau_r(f,\delta)_p := \|\omega_r(f,\cdot;\delta)\|_p = \left\{ \int_{-1}^{1} [\omega_r(f,x;\delta)]^p \, dx \right\}^{1/p}, \qquad 1 \leq p < +\infty,$$

for $0 < \delta \leq 2/r$.

The main advantage of this tool over the usual L^p-modulus of smoothness, which is unable to consider the increments of a given function f on subsets of I with null measure, is that, by representing a mean of $\omega_r(f,x;\delta)$, it allows the increments of f to be evaluated at any fixed point x in I.

In analogy to the usual moduli of smoothness, the τ-modulus has the following useful properties established in [23]:

(a) monotonicity:

$$\tau_r(f,\delta')_p \leq \tau_r(f,\delta'')_p, \qquad \text{for} \quad \delta' \leq \delta'';$$

(b) sub-additivity:

$$\tau_r(f+g,\delta)_p \leq \tau_r(f,\delta)_p + \tau_r(g,\delta)_p, \qquad \delta > 0;$$

(c) estimate of the high order modulus by means of lower order one:

$$\tau_r(f,\delta)_p \leq 2\,\tau_{r-1}\left(f, \frac{r}{r-1}\,\delta\right)_p, \qquad \delta > 0;$$

(d) estimate of the modulus of order r of f by means of the modulus of order $r-1$ of f' (whenever it exists):

$$\tau_r(f,\delta)_p \leq \delta\,\tau_{r-1}\left(f', \frac{r}{r-1}\,\delta\right)_p, \qquad \delta > 0;$$

(e) the inequality:

$$\tau_r(f,n\delta)_p \leq (2n)^{r+1}\tau_r(f,\delta)_p, \quad n \in \mathbb{N}, \; \delta > 0.$$

For more details concerning $\tau_r(f,\delta)_p$, see [23] again. In the last part of this section, we recall some well-known results concerning the τ-modulus that will be very useful in the next section. We begin with the following lemma. For a proof, see [9, Lemma 4.3].

Lemma 2.3. *Let $r \geq 1$ be an integer. The following estimate:*

$$\left\{ \sum_{k=-n}^{n} [\omega_r(f,k/n,2h)]^p \, n^{-1} \right\}^{1/p} \leq 2^{(1/p)+2(r+1)}\,\tau_r\left(f, h + \frac{1}{(n+1)\,r}\right)_p, \qquad 1 \leq p < +\infty,$$

holds, for every $0 < h \leq 2/r$.

For what concerns the behaviour for $\delta \to 0^+$, $\tau_r(f,\delta)_p$ converges to zero under suitable assumptions upon the function f. Indeed, we can state the following (see [2, Proposition 6 and Remark 7]):

Proposition 2.1. *Let $1 \le p < +\infty$. For every $f \in M(I)$ that is Riemann integrable, there holds:*

$$\lim_{\delta \to 0^+} \tau_r(f,\delta)_p = 0, \quad r \in \mathbb{N},\ r \ge 1.$$

In the particular case $r = 1$, $\lim_{\delta\to 0^+} \tau_1(f,\delta)_p = 0$ is a necessary and sufficient condition for $f \in M(I)$ to be Riemann integrable.

From now on, we denote by $W_p^r(I)$ the usual Sobolev spaces, i.e., the subspaces of $L^p(I)$ of all functions $f : I \to \mathbb{R}$ with an absolutely continuous $(r-1)$-th derivative and with the r-th derivative belonging to $L^p(I)$, with $1 \le p < +\infty$, and $r \ge 1$. Now, let us recall another preliminary density result that establishes a connection between the τ-modulus $\tau_r(f,\delta)_p$ and the so-called Steklov functions (see, e.g., [23, 25]), that is:

Theorem 2.1 ([23, Theorem 2.5′, p. 34])**.** *Let f be a bounded function belonging to $L^p(I)$, $1 \le p < +\infty$. For every integer $r \ge 1$ and every $0 < h \le 2$, there exists a Steklov-type function $f_{r,h} \in L^p(I)$, defined by:*

$$f_{r,h}(x) := (-h)^{-r} \int_0^h \cdots \int_0^h \sum_{m=1}^{r} (-1)^{r-m+1} \binom{r}{m} f\left(x + \frac{m}{r}(t_1 + \ldots + t_r)\right) dt_1 \ldots dt_r,$$

where, here, f has to be considered extended on the whole $\mathbb{R}$, as a periodic function of period 2, satisfying the following properties:

(i) $|f(x) - f_{r,h}(x)| \le \omega_r(f, x; 2h)$, $x \in I$;

(ii) $\|f - f_{r,h}\|_p \le \tau_r(f, 2h)_p$;

(iii) $f_{r,h} \in W_p^r(I)$ *and for its s-th derivative, the following inequality holds:*

$$\|f_{r,h}^{(s)}\|_p \ \le\ c(r)\, h^{-s} \tau_s(f,h)_p, \qquad s = 1, \ldots, r,$$

where the constant $c(r)$ is dependent only on r.

Finally, at the end of this section, we recall the following useful and well-known inequality. To do this, we introduce the well-known Hardy-Littlewood maximal function, defined by:

$$M(f;x) := \sup_{t \in I, t \ne x} \frac{1}{t-x} \int_x^t |f(u)|\, du.$$

It is known that the L^p-norm of the Hardy-Littlewood maximal function can be estimated as follows (see [24]):

$$\|M(f;\cdot)\|_p \le C_p \|f\|_p, \quad f \in L^p(I), \quad 1 < p < +\infty. \tag{2.2}$$

3. Main Results

We begin this section with the following estimate in case of functions belonging to the Sobolev spaces $W_p^1(I)$, $1 < p < +\infty$.

Theorem 3.2. *Let σ be a sigmoidal function satisfying $(\Sigma 3)$ with $\alpha > p+1$, $1 < p < +\infty$. For every $f \in W_p^1(I)$, it turns out that:*

$$\|F_n f - f\|_p \le \left(\frac{M_p(\phi_\sigma)}{\phi_\sigma(1)}\right)^{1/p} n^{-1} C_p \|f'\|_p, \quad n \in \mathbb{N},$$

where $M_p(\phi_\sigma) < +\infty$ and C_p is the constant arising from inequality (2.2).

Proof. According to the definition of Sobolev spaces given in the previous section, since $f \in W^1_p(I)$, we can write the following first order Taylor formula with integral remainder (see [14, p. 37]):

$$f(u) = f(x) + \int_x^u f'(t)\,dt, \quad x, u \in I;$$

hence, for every fixed $x \in I$ and integer $n \geq 1$, by the definition of $F_n f$, we obtain:

$$(3.3)\quad \begin{aligned}(F_n f)(x) &= \frac{\sum_{k=-n}^{n}\left[f(x) + \int_x^{k/n} f'(t)\,dt\right]\phi_\sigma(nx-k)}{\sum_{k=-n}^{n}\phi_\sigma(nx-k)} \\ &= f(x) + \frac{\sum_{k=-n}^{n}\left[\int_x^{k/n} f'(t)\,dt\right]\phi_\sigma(nx-k)}{\sum_{k=-n}^{n}\phi_\sigma(nx-k)}.\end{aligned}$$

Now, using the above expression for the operator F_n, the discrete Jensen inequality (see, e.g., [11]) with the convexity of $|\cdot|^p$ and inequality (2.1), we can write what follows:

$$\begin{aligned}\int_{-1}^{1}|(F_n f)(x) - f(x)|^p dx &= \int_{-1}^{1}\left|\frac{\sum_{k=-n}^{n}\left[\int_x^{k/n} f'(t)dt\right]\phi_\sigma(nx-k)}{\sum_{k=-n}^{n}\phi_\sigma(nx-k)}\right|^p dx \\ &\leq \int_{-1}^{1}\frac{\sum_{k=-n}^{n}\left|\int_x^{k/n} f'(t)dt\right|^p\phi_\sigma(nx-k)}{\sum_{k=-n}^{n}\phi_\sigma(nx-k)}dx \\ (3.4)\quad &\leq \frac{1}{\phi_\sigma(1)}\int_{-1}^{1}\sum_{k=-n}^{n}\left|\int_x^{k/n} f'(t)\,dt\right|^p\phi_\sigma(nx-k)dx \\ &\leq \frac{1}{\phi_\sigma(1)}\int_{-1}^{1}\sum_{k=-n}^{n}\left|\frac{k/n - x}{k/n - x}\int_x^{k/n}|f'(t)|dt\right|^p\phi_\sigma(nx-k)dx.\end{aligned}$$

Recalling the definition of the Hardy-Littlewood maximal function, the notion of the discrete absolute moment of order p of ϕ_σ, and inequality (2.2), we finally obtain:

$$\begin{aligned}\int_{-1}^{1}|(F_n f)(x) - f(x)|^p dx &\leq \frac{1}{\phi_\sigma(1)}\int_{-1}^{1}\sum_{k=-n}^{n}\left|\left(\frac{k}{n} - x\right)M(f';x)\right|^p\phi_\sigma(nx-k)dx \\ &= \frac{1}{\phi_\sigma(1)}n^{-p}\int_{-1}^{1}|M(f';x)|^p\sum_{k=-n}^{n}|nx-k|^p\phi_\sigma(nx-k)dx \\ &\leq \frac{M_p(\phi_\sigma)}{\phi_\sigma(1)}n^{-p}\|M(f';\cdot)\|_p^p\end{aligned}$$

$$\leq \frac{M_p(\phi_\sigma)}{\phi_\sigma(1)} n^{-p} C_p^p \|f'\|_p^p < +\infty,$$

where $M_p(\phi_\sigma) < +\infty$, since condition $(\Sigma 3)$ is satisfied for $\alpha > p+1$ and $\|f'\|_p^p < +\infty$ since $f \in W_p^1(I)$. This completes the proof. □

Note that the estimate established in Theorem 3.2 holds for every $1 < p < +\infty$ but does not cover the case $p = 1$. This is due to the fact that inequality (2.2), which is shown to be crucial in the proof of the previous theorem, does not hold in the case $p = 1$. However, if we suppose that the density function ϕ_σ has compact support, we are able to establish the following result for the operators $F_n f$ with $f \in W_p^1(I)$, $1 \leq p < +\infty$. Namely, in the latter case, we also achieve an estimate for $p = 1$.

Theorem 3.3. *Let σ be a sigmoidal function such that supp $\phi_\sigma \subseteq [-T, T]$, $T > 0$. Further, let $f \in W_p^1(I)$, $1 \leq p < +\infty$, be fixed. Then:*

$$\|F_n f - f\|_p \leq \left(\frac{2T M_{p-1}(\phi_\sigma)}{\phi_\sigma(1)} \right)^{1/p} n^{-1} \|f'\|_p,$$

for $n \in \mathbb{N}$, where $M_{p-1}(\phi_\sigma) < +\infty$, since ϕ_σ is bounded and with compact support.

Proof. Let $n \in \mathbb{N}$ and $f \in W_p^1(I)$, $1 \leq p < +\infty$, be fixed. Repeating the same computations of (3.3) and (3.4), we obtain

$$\int_{-1}^{1} |(F_n f)(x) - f(x)|^p \, dx \leq \frac{1}{\phi_\sigma(1)} \int_{-1}^{1} \sum_{k=-n}^{n} \left| \frac{|k/n - x|}{|k/n - x|} \int_x^{k/n} |f'(t)| \, dt \right|^p \phi_\sigma(nx - k) \, dx$$

$$= \frac{1}{\phi_\sigma(1)} \int_{-1}^{1} \sum_{k=-n}^{n} \left| \frac{k}{n} - x \right|^p \left| \frac{1}{|k/n - x|} \int_x^{k/n} |f'(t)| \, dt \right|^p \phi_\sigma(nx - k) \, dx.$$

Now, we formally extend f' on the whole $\mathbb{R}$, as a periodic function of period 2. Thus, recalling that *supp* $\phi_\sigma \subseteq [-T, T]$, and using the continuous Jensen inequality, we can write what follows:

$$\int_{-1}^{1} |(F_n f)(x) - f(x)|^p dx$$

$$\leq \frac{1}{\phi_\sigma(1)} \int_{-1}^{1} \sum_{k=-n}^{n} \left| \frac{k}{n} - x \right|^{p-1} \left| \int_x^{k/n} |f'(t)|^p \, dt \right| \phi_\sigma(nx - k) dx$$

$$\leq \frac{1}{\phi_\sigma(1)} \int_{-1}^{1} \sum_{\substack{k=-n \\ |nx-k| \leq T}}^{n} \left| \frac{k}{n} - x \right|^{p-1} \left| \int_0^{k/n - x} |f'(y + x)|^p dy \right| \phi_\sigma(nx - k) dx$$

$$\leq \frac{1}{\phi_\sigma(1)} \int_{-1}^{1} \sum_{\substack{k=-n \\ |nx-k| \leq T}}^{n} \left| \frac{k}{n} - x \right|^{p-1} \left[\int_{|y| \leq |k/n - x|} |f'(y + x)|^p \, dy \right] \phi_\sigma(nx - k) dx$$

$$\leq \frac{1}{\phi_\sigma(1)} \int_{-1}^{1} \sum_{\substack{k=-n \\ |nx-k| \leq T}}^{n} \left| \frac{k}{n} - x \right|^{p-1} \left[\int_{|y| \leq T/n} |f'(y + x)|^p \, dy \right] \phi_\sigma(nx - k) dx.$$

Now, we can use the Fubini-Tonelli theorem in order to interchange the integrals in the above computations, obtaining:

$$\begin{aligned}
&\int_{-1}^{1} |(F_n f)(x) - f(x)|^p dx \\
&\leq \frac{1}{\phi_\sigma(1)} \int_{-1}^{1} \sum_{\substack{k=-n \\ |nx-k|\leq T}}^{n} \left|\frac{k}{n} - x\right|^{p-1} \left[\int_{|y|\leq T/n} |f'(y+x)|^p \, dy\right] \phi_\sigma(nx-k) dx \\
&= \frac{1}{\phi_\sigma(1)} n^{1-p} \int_{|y|\leq T/n} dy \int_{-1}^{1} |f'(y+x)|^p \sum_{\substack{k=-n \\ |nx-k|\leq T}}^{n} |nx-k|^{p-1} \phi_\sigma(nx-k) dx \\
&\leq \frac{M_{p-1}(\phi_\sigma)}{\phi_\sigma(1)} n^{1-p} \int_{|y|\leq T/n} \|f'(y+\cdot)\|_p^p dy,
\end{aligned}$$

where $M_{p-1}(\phi_\sigma) < +\infty$ denotes the discrete absolute moment of order $(p-1)$ of ϕ_σ. Finally, observing that:

$$\|f'(y+\cdot)\|_p^p = \|f'\|_p^p$$

for every $y \in [-T/n, T/n]$, we finally obtain:

$$\|F_n f - f\|_p^p \leq \frac{2T M_{p-1}(\phi_\sigma)}{\phi_\sigma(1)} n^{-p} \|f'\|_p^p.$$

This completes the proof. □

Now, we can prove the main theorem of this paper, i.e., a quantitative Jackson-type estimate in terms of the first order τ-modulus for the approximation error in the L^p-norm by the operators F_n.

Theorem 3.4. *Let σ be a sigmoidal function satisfying $(\Sigma 3)$ with $\alpha > p+1$, $1 < p < +\infty$. For any bounded $f \in L^p(I)$, it turns out that:*

$$\|F_n f - f\|_p \leq K \tau_1\left(f, \frac{1}{n}\right)_p,$$

for every $n \in \mathbb{N}$, where:

$$K := \frac{2^{(1/p)+8}}{\phi_\sigma(1)^{1/p}} + 16 + C_p c(1) \left(\frac{M_p(\phi_\sigma)}{\phi_\sigma(1)}\right)^{1/p},$$

C_p is the constant arising from inequality (2.2) and $c(1)$ is given in (iii) of Theorem 2.1.

Proof. Let $n \in \mathbb{N}$ and $0 < h \leq 1$ be fixed. Since Theorem 2.1 with $r = 1$ holds, we can construct for the function f the corresponding Steklov type function $f_{1,h}$. Thus, we can write what follows:

$$\|F_n f - f\|_p \leq \|F_n f - F_n f_{1,h}\|_p + \|F_n f_{1,h} - f_{1,h}\|_p + \|f_{1,h} - f\|_p.$$

By using (ii) of Theorem 2.1 (with $r = 1$), we obtain:

$$\|f_{1,h} - f\|_p \leq \tau_1(f, 2h)_p. \tag{3.5}$$

Further, since $f_{1,h} \in W_p^1(I)$, exploiting Theorem 3.2 and (iii) of Theorem 2.1 (with $r = 1$), we immediately get that:

$$\|F_n f_{1,h} - f_{1,h}\|_p \leq \left(\frac{M_p(\phi_\sigma)}{\phi_\sigma(1)}\right)^{1/p} n^{-1} C_p \|f'_{1,h}\|_p \leq \bar{K} n^{-1} h^{-1} \tau_1(f,h)_p,$$

where:

$$\bar{K} := C_p c(1) \left(\frac{M_p(\phi_\sigma)}{\phi_\sigma(1)} \right)^{1/p},$$

for suitable positive constants C_p and $c(1)$.
Finally, exploiting the linearity of the operators F_n, using Lemma 2.2, again Theorem 2.1 (i) (with $r = 1$), Lemma 2.3 with $r = 1$ and the fact that $\tau_1(f, \cdot)_p$ is non-decreasing, respectively, we can estimate the quantity $\|F_n f - F_n f_{1,h}\|_p$ as follows:

$$\begin{aligned}
\|F_n f - F_n f_{1,h}\|_p = \|F_n(f - f_{1,h})\|_p &\leq \frac{1}{\phi_\sigma(1)^{1/p}} \|f - f_{1,h}\|_{l^p(\Sigma_n)} \\
&= \frac{1}{\phi_\sigma(1)^{1/p}} \left\{ \sum_{k=-n}^{n} \left| f\left(\frac{k}{n}\right) - f_{1,h}\left(\frac{k}{n}\right) \right|^p n^{-1} \right\}^{1/p} \\
&\leq \frac{1}{\phi_\sigma(1)^{1/p}} \left\{ \sum_{k=-n}^{n} \left[\omega_1\left(f, \frac{k}{n}, 2h\right) \right]^p n^{-1} \right\}^{1/p} \\
&\leq \frac{2^{(1/p)+4}}{\phi_\sigma(1)^{1/p}} \tau_1\left(f, h + \frac{1}{n+1}\right)_p \\
&\leq \frac{2^{(1/p)+4}}{\phi_\sigma(1)^{1/p}} \tau_1\left(f, h + \frac{1}{n}\right)_p. \qquad (3.6)
\end{aligned}$$

Now, we set $h = 1/n \leq 1$. Thus, rearranging all the above estimates, recalling the property (e) of the first order τ-modulus and that $\alpha > p + 1$, we finally obtain:

$$\begin{aligned}
\|F_n f - f\|_p &\leq \left(\frac{2^{(1/p)+4}}{\phi_\sigma(1)^{1/p}} + 1 \right) \tau_1\left(f, \frac{2}{n}\right)_p + \bar{K}\tau_1\left(f, \frac{1}{n}\right)_p \\
&\leq \left(\frac{2^{(1/p)+8}}{\phi_\sigma(1)^{1/p}} + 16 + \bar{K} \right) \tau_1\left(f, \frac{1}{n}\right)_p.
\end{aligned}$$

□

The quantitative estimate established in the previous theorem does not hold if $p = 1$. However, under the same assumptions as in Theorem 3.3, we can obtain an analogous result that covers the latter case. Indeed, we can establish the following.

Theorem 3.5. *Let σ be a sigmoidal function such that supp $\phi_\sigma \subseteq [-T, T]$, $T > 0$. Further, let $f \in L^p(I)$ be a bounded function. Then:*

$$\|F_n f - f\|_p \leq K\tau_1\left(f, \frac{1}{n}\right)_p,$$

for $n \in \mathbb{N}$ and $1 \leq p < +\infty$, where:

$$K := \frac{2^{(1/p)+8}}{\phi_\sigma(1)^{1/p}} + 16 + c(1) \left(\frac{2T M_{p-1}(\phi_\sigma)}{\phi_\sigma(1)} \right)^{1/p}$$

and $c(1)$ is the constant arising from Theorem 2.1 (iii).

The proof of Theorem 3.5 is the same of Theorem 3.4, where we used the inequality achieved in Theorem 3.3 in place of that one of Theorem 3.2.
Obviously, using Proposition 2.1 together with Theorem 3.4 and Theorem 3.5, we can deduce the L^p-convergence of F_n (see [9]).

Remark 3.1. *In [9], quantitative estimates for the approximation error in the L^p-norm for the operators F_n have already been achieved by using the τ-modulus. In particular, in [9, Theorem 4.4], the author proved that, if the sigmoidal function σ satisfies $(\Sigma 3)$ with $\alpha > 2p$, $1 \leq p < +\infty$, then for any $f \in M(I)$, there exist two suitable positive absolute constants K_1, K_2 such that:*

$$\|F_n f - f\|_p \leq K_1 \tau_1 \left(f, \frac{1}{n^{1-1/2p}}\right)_p + K_2 \tau_2 \left(f, \frac{1}{n^{1-1/2p}}\right)_p , \tag{3.7}$$

for $n \in \mathbb{N}$. Comparing the above results with those established in this paper, it is clear that the Jackson-type estimates for the NN operators F_n established in Theorem 3.4 and Theorem 3.5 are better than the one given in (3.7), and this explains the improvement with respect to the results proved in [9]. We stress that, unlike the quantitative estimate in (3.7), to achieve the order of approximation established in this paper, we adopted a different strategy of proof, in which a crucial role is played by the celebrated Hardy-Littlewood maximal inequality (see (2.2)).

For what concerns possible examples of sigmoidal activation functions to which the above theory can be applied, we recall the classical cases of the logistic and the hyperbolic tangent sigmoidal functions (see, e.g., [4, 5]). It is well-known that, due to their exponential decay to zero as $x \to -\infty$, both satisfy condition $(\Sigma 3)$ for every $\alpha > 0$. This means that for the NN operators activated by these smooth sigmoidal functions, both Theorem 3.2 and Theorem 3.4 hold.
On the other hand, we can also consider non-smooth (continuous) sigmoidal functions. Indeed, as it is known (see, e.g., [10]), if we replace condition $(\Sigma 2)$ by directly assuming the condition (iii) of Lemma 2.1, together with $\phi_\sigma(1) > 0$, all the approximation results for F_n of this paper still hold. A remarkable class of continuous (non-smooth) sigmoidal functions can be generated from suitable finite linear combinations of shifted Rectified Power Units (RePUs), that are, in fact, powers of the well-known Rectified Linear Unit (ReLU) activation function (see, e.g., [17, 21, 26]), widely used in applications of neural network type approximation.

Acknowledgments

The authors are members of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM), of the Gruppo UMI (Unione Matematica Italiana) T.A.A. (Teoria dell'Approssimazione e Applicazioni), and of the network RITA (Research ITalian network on Approximation).

Funding

The first author has been partially supported by: (1) 2024 GNAMPA-INdAM project "DYNAMIChE: DYNAmical Methods: Inverse problems, Chaos and Evolution", CUP E53C23001670001. The second and the third authors have been partially supported within the (2) 2022 GNAMPA-INdAM Project "Enhancement e segmentazione di immagini mediante operatori di tipo campionamento e metodi variazionali per lo studio di applicazioni biomediche", (3) "National Innovation Ecosystem grant ECS00000041 - VITALITY", funded by the European Union - NextGenerationEU under the Italian Ministry of University and Research (MUR). The second author has been partially supported by (4) 2023 GNAMPA-INdAM Project "Approssimazione costruttiva e astratta mediante operatori di tipo sampling e loro applicazioni", CUP E53C22001930001, (5) 2024 GNAMPA-INdAM Project "Tecniche di approssimazione in spazi funzionali con applicazioni a problemi di diffusione", CUP E53C23001670001, (6) PRIN 2022: "AI- and DIP-Enhanced DAta Augmentation for Remote Sensing of Soil Moisture and Forest Biomass (AIDA)" funded by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, under the Italian Ministry of University and Research (MUR) (Project Code: 20229FX3B9, CUP: J53D23000660001), and (7) PRIN 2022 PNRR: "RETINA: REmote sensing daTa INversion with multivariate functional modeling for essential climAte variables characterization" funded by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, under the MUR (Project Code: P20229SH29, CUP: J53D23015950001), while the third author by (8) PRIN 2022: "EXPANSION - EXPlainable AI through high eNergy physicS for medical Imaging in ONcology" funded by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, under the MUR (Project Code: 2022Z3BT9F, CUP: J53D23002530006) and by University Research Fund 2024 of the University of Perugia.

References

[1] G. A. Anastassiou, L. Coroianu and S. G. Gal: *Approximation by a nonlinear Cardaliaguet-Euvrard neural network operator of max-product kind*, J. Comput. Anal. Appl., **12** (2) (2010), 396–406.
[2] C. Bardaro, P. L. Butzer, R. L. Stens and G. Vinti: *Approximation error of the Whittaker cardinal series in terms of an averaged modulus of smoothness covering discontinuous signals*, J. Math. Anal. Appl., **316** (2006), 269–306.
[3] M. Cantarini, L. Coroianu, D. Costarelli, S. G. Gal and G. Vinti: *Inverse result of approximation for the max-product neural network operators of the Kantorovich type and their saturation order*, Mathematics, **10** (63) (2022), 1–11.
[4] F. Cao, Z. Chen: *The approximation operators with sigmoidal functions*, Comput. Math. Appl., **58** (4) (2009), 758–765.
[5] 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.
[6] P. Cardaliaguet, G. Euvrard: *Approximation of a function and its derivative with a neural network*, Neural Netw., **5** (2) (1992), 207–220.
[7] S. Chakraverty, D. M. Sahoo and N. R. Mahato: *McCulloch-Pitts neural network model*, In: Concepts of Soft Computing: Fuzzy and ANN with Programming, Springer, Singapore (2019).
[8] D. Costarelli: *Density results by deep neural network operators with integer weights*, Math. Model. Anal., **27** (4) (2022), 547–560.
[9] D. Costarelli: *Approximation error for neural network operators by an averaged modulus of smoothness*, J. Approx. Theory, **294** (2023), 105944.
[10] D. Costarelli, R. Spigler: *Approximation results for neural network operators activated by sigmoidal functions*, Neural Netw., **44** (2013), 101–106.
[11] D. Costarelli, R. Spigler: *How sharp is the Jensen inequality?*, J. Inequal. Appl., 2015:69 (2015), 1–10.
[12] D. Costarelli, G. Vinti: *Voronovskaja type theorems and high order convergence neural network operators with sigmoidal functions*, Mediterr. J. Math., **17** (2020), Article ID: 77.
[13] G. Cybenko: *Approximation by superpositions of a sigmoidal function*, Math. Control Signals Systems, **2** (1989), 303–314.
[14] R. A. DeVore, G. G. Lorentz: *Constructive approximation*, Springer Science & Business Media, Vol. 303, (1993).
[15] C. Fang, J. D. Lee, P. Yang and T. Zhang: *Modeling from features: A mean-field framework for over-parameterized deep neural networks*, arXiv preprint (2020), https://arxiv.org/abs/2007.01452 (Accessed 23 December 2020).
[16] D. Gotleyb, G. Lo Sciuto, C. Napoli, R. Shikler, E. Tramontana and M. Wozniak: *Characterization and modeling of organic solar cells by using radial basis neural networks*, In: Artificial Intelligence and Soft Computing, (2016), 91–103.
[17] I. Gühring, M. Raslan: *Approximation rates for neural networks with encodable weights in smoothness spaces*, Neural Netw., **134** (2021), 107–130.
[18] P. C. Kainen, V. Kurková and A. Vogt: *Approximative compactness of linear combinations of characteristic functions*, J. Approx. Theory, **257** (2020), Article ID: 105435.
[19] V. Kurková, M. Sanguineti: *Classification by Sparse Neural Networks*, IEEE Trans. on Neural Netw. Learning Syst., **30** (9) (2019), 2746–2754.
[20] S. Langer: *Approximating smooth functions by deep neural networks with sigmoid activation function*, J. Multivariate Anal., **182** (2021), 104696.
[21] B. Li, S. Tang and H. Yu: *Better Approximations of High Dimensional Smooth Functions by Deep Neural Networks with Rectified Power Units*, arXiv:1903.05858v4, (2019).
[22] A. Mishra, P. Chandra, U. Ghose and S. S. Sodhi: *Bi-modal derivative adaptive activation function sigmoidal feedforward artificial neural networks*, Appl. Soft Comput., **61** (2017), 983–994.
[23] B. Sendov, V. A. Popov: *The averaged moduli of smoothness: applications in Numerical Analysis and Approximation*, Volume in Pure and Applied Mathematics, Wiley - Interscience Series of Texts, Monographs, and Tracts, Chichester - New York - Brisbane - Toronto - Singapore (1988).
[24] E. M. Stein: *Singular Integrals*, Princeton, New Jersey (1970).
[25] A. F. Timan: *Theory of approximation of functions of a real variable*, MacMillan, New York (1965).
[26] D. Yarotsky: *Universal Approximations of Invariant Maps by Neural Networks*, Constr. Approx., (2021), https://doi.org/10.1007/s00365-021-09546-1.

LORENZO BOCCALI
UNIVERSITY OF FLORENCE
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE "ULISSE DINI"
VIALE MORGAGNI 67/A, 50134, FLORENCE, ITALY
Email address: lorenzo.boccali@unifi.it

DANILO COSTARELLI
UNIVERSITY OF PERUGIA
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
VIA VANVITELLI 1, 06123, PERUGIA, ITALY
Email address: danilo.costarelli@unipg.it

GIANLUCA VINTI
UNIVERSITY OF PERUGIA
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
VIA VANVITELLI 1, 06123, PERUGIA, ITALY
Email address: gianluca.vinti@unipg.it