Paper published in the: Mediterranean Journal of Mathematics, Vol. 21 (2024) Article number 199, DOI: 10.1007/s00009-024-02752-8.

# Asymptotic analysis of neural network operators employing the Hardy-Littlewood maximal inequality

**Danilo Costarelli*** **Michele Piconi**
Department of Mathematics and Computer Science
University of Perugia
1, Via Vanvitelli, 06123 Perugia, Italy
danilo.costarelli@unipg.it - michele.piconi@unipg.it

**Abstract**

The main aim of the present paper is to provide a full asymptotic analysis of a family of neural network (NN) operators based on suitable density functions within the $L^p$-setting, and in the space of continuous functions. Two approaches are pursued: the first employs the celebrated Hardy-Littlewood (HL) maximal inequality, while the second adopts a constructive, fully moment-based, method. A crucial step in the proof of the previous results is provided by achieving asymptotic estimates for the NN operators in the cases of functions belonging to Sobolev spaces. By means of the previously mentioned first approach, we are able to establish sharp estimates that can not be applied with $p = 1$, since in that case the HL maximal inequality fails. This justifies resorting to the second complementary approach, which is revealed to be very useful to cover the remaining case. The asymptotic analysis is finally completed by deducing the corresponding qualitative order of approximation for functions within suitable Lipschitz classes. At the end of the paper, several examples of density functions are also presented and discussed in relation to the previous results. Finally, we recall that NN operators based on the well-known ReLU or RePUs functions are also included in the present theory.

AMS 2010 Mathematics Subject Classification: 47A58, 47A63, 47A57, 41A25, 41A05

Key words and phrases: neural network operators, asymptotic analysis, Hardy-Littlewood maximal inequality, Sobolev spaces, sigmoidal functions, ReLU, RePUs

## 1 Introduction

It is well-known that artificial neural networks (NNs) were initially conceptualized in the early 1900s with the primary objective of creating a simplistic model mirroring the main workings of the human brain. The foundational idea revolved around a network architecture structured in one or more layers composed of many nodes, where each one emulates the behaviour of a biological neuron.

The pervasive influence of neural networks extends across several domains, making them one of the most extensively studied topics, as evidenced by recent references like [42, 43] and several others. Their widespread applicability spans fields such as artificial intelligence, machine learning, as well as more deterministic and rigorous disciplines, like the mathematics, as exemplified, e.g., by [38, 21, 30, 29, 26, 39].

One of the main abilities of the neural network model is the possibility to be learned by the implementation of training algorithms; formally this can be viewed as an approximation process. The latter remark justified all the studies performed in the last forty years in relation to this subject, that is, the study of NN-type approximation methods.

*Corresponding author

Among the possible considered approaches, we can find the widely studied NN operators, see, e.g., the following set of papers [6, 7, 8, 2, 19, 24, 4, 12, 27, 34, 3, 5, 28, 40, 11]), where some of them are very recent.

Among the most studied versions of the NN operators, we can find that one introduced and studied in [20], where the authors defined a so-called Kantorovich version of these NN operators, which is revealed to be very suitable in order to reconstruct multivariate not-necessarily continuous data (i.e., in the $L^p$-setting, $1 \leq p < +\infty$), other than the continuous ones. The main idea behind the definition of such operators was that their coefficients must be defined by means of integral averages of the type $\int_{\frac{k}{n}}^{\frac{k+1}{n}} f(u)du$, where $f$ is a locally integrable target function.

More precisely, the Kantorovich neural network (NN) operators assume the following form

$$(K_n f)(x) \;=\; \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \left[ n \int_{k/n}^{(k+1)/n} f\,(u)\; du \right] \phi_\sigma\,(nx-k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \phi_\sigma(nx-k)},$$

where $n \in \mathbb{N}$, $f : [a,b] \to \mathbb{R}$, and $x \in [a,b]$. In the above definition, $\phi_\sigma(\cdot)$ denotes a density function generated by a suitable finite linear combination of sigmoidal activation functions, as extensively shown in the following section, while the symbols $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denote respectively the ceiling and the integer part of a given number.

In the last years, several studies have been carried out for Kantorovich NN operators (see [10, 36]), including for their nonlinear max-product variants (see, e.g., [17]).

We recall that sigmoidal functions have been considered as activation of NN-type models since the beginning of this topic; currently, other kinds of activation functions are also taken into account, such as the rectified linear unit (ReLU) and the rectified power units (RePUs) functions. At the end of the paper, other than examples of sigmoidal functions for which the theory of Kantorovich NN can be applied, we claim that also ReLU and RePUs functions can be included in the present theory.

In the present paper, we provide a fully asymptotic analysis for the above operators $K_n$ in the $L^p$-setting and in the space of continuous functions, in order to quantify their approximation performances in terms of a specific rate of convergence.

Such analysis will be performed thanks to the use of the well-known modulus of smoothness $\omega(f,\delta)_p$, $1 \leq p \leq +\infty$, according to the convergence results previously recalled.

To reach the above aim, the path to go through is the following: first asymptotic estimates must be derived in a regular subspace of $L^p$, that is the usual Sobolev space $W^{1,p}$, then, using a constructive density theorem of Sendov and Popov ([35]) based on the well-known Steklov functions ([14]), we are able to extend such quantitative estimates to the whole $L^p$-space.

The central point in the above proof is the application of the celebrated Hardy-Littlewood (HL) maximal inequality, which involves the so-called HL maximal function; in this way we provide sharp estimates in the case of $p$-norm, with $1 < p \leq +\infty$, since it is well-known that the HL maximal inequality is not valid, in general, when $p = 1$.

Motivated by the fact that the above result can not be used to treat the case $p = 1$, we also propose a different strategy of proof that allows us to cover this gap.

Hence, we propose an additional and complementary proof of the above quantitative analysis based on classical techniques, i.e., on a constructive moment-type approach. Here, we provide an asymptotic theorem in the case of functions belonging to the Sobolev space $W^{2,p}$, $1 \leq p \leq +\infty$, and requiring a slightly stronger assumption on the sigmoidal function $\sigma$ generating $\phi_\sigma$. Finally, the asymptotic theorem is extended to the whole $L^p$-spaces resorting again to the previously mentioned constructive density theorem of Sendov and Popov.

Note that, the estimates achieved by the second proposed approach for the cases $p > 1$ are less sharp than those established by the first proof employing the HL maximal inequality; hence the second proof can be considered meaningful only to face the case $p = 1$.

For the sake of completeness, we complete the analysis of the order of approximation by providing also the qualitative versions of the above estimates, when functions belonging to suitable Lipschitz classes are considered.

This study on the order of approximation in $L^p$-spaces, coupled with the analysis of convergence, provides the theoretical background for many applications problems (see, e.g., [44]). In fact, it is well-known that approximation and quantitative results involving not necessarily continuous functions, that model the majority of real world data, turn out to be useful in several applications across different fields, including Signal and Image Processing and Machine Learning.

## 2 Notations and preliminaries

Let $I \subseteq \mathbb{R}$ be a given compact. We denote by $L^p(I)$, $1 \leq p < +\infty$, the space of all Lebesgue measurable functions $f : I \to \mathbb{R}$ such that $\int_I |f(u)|^p du < +\infty$, endowed with the norm

$$\|f\|_p := \left( \int_I |f(u)|^p du \right)^{\frac{1}{p}}.$$

If $p = +\infty$, let $L^\infty(I)$ be the space of all essentially bounded functions, provided with the norm $\|f\|_\infty := \operatorname{ess\,sup}_{u \in \mathbb{R}} |f(u)|$. Moreover, by $C(I)$ we denote its subspace of all continuous functions on $I$, where $\|\cdot\|_\infty$ is actually the max-norm. Furthermore, we may recall the definition of *Sobolev spaces*, namely

$$W^{k,p}(I) := \left\{ f \in L^p(I) : \, f^{(n)} \in L^p(I), \, 1 \leq n \leq k, \, n \in \mathbb{N} \right\}, \tag{1}$$

$1 \leq p \leq +\infty$, $k \in \mathbb{N}$, where the derivatives $f^{(n)}$ are given in the distributional (or weak) sense. Equivalently, the Sobolev spaces $W^{k,p}(I)$ can be also defined as the set of functions for which the distributional (or weak) derivatives $f^{(k-1)} \in AC(I)$ and $f^{(k)} \in L^p(I)$, where $AC(I)$ denotes the space of absolutely continuous functions on $I$ (see [23]). Finally, by $C^k(I)$ we denote the space of all $k$-times continuously differentiable functions on $I$.
In order to get quantitative estimates to study the order of approximation, it is needed to introduce the notion of the *modulus of smoothness* of a given function $f$, that arises from the definition of finite differences of order $k \in \mathbb{N}$ of $f$, namely

$$\Delta_h^k(f, x) := \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} f(x + jh), \qquad x \in I, \tag{2}$$

where $f : I \to \mathbb{R}$. Obviously, in (2) and in where it is necessary we consider $f$ as extended on the whole real line as a periodic function, with period equal to the length of the compact $I$. This leads to the definition of the *k-th modulus of smoothness* in Lebesgue spaces, given by

$$\omega_k(f, \delta)_p := \sup_{|h| \leq \delta} \|\Delta_h^k(f, \cdot)\|_p, \qquad \delta > 0, \tag{3}$$

where $f \in L^p(I)$ if $1 \leq p < +\infty$, or $f \in C(I)$ if $p = +\infty$. We remark that if $k = 1$, it reduces to the classical *first-order modulus of smoothness*, that can be briefly denoted by $\omega(f, \delta)_p$.

## 3 The Kantorovich NN operators

A 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 always consider non-decreasing sigmoidal functions $\sigma$, with $\sigma(1) < 1$, satisfying the following assumptions:

$(S1)$ $\sigma(x) - 1/2$ is an odd function;

$(S2)$ $\sigma \in C^2(\mathbb{R})$ and it is also concave on $[0, +\infty)$;

$(S3)$ $\sigma(x) = \mathcal{O}(|x|^{-\alpha-1})$ as $x \to -\infty$, for some $\alpha > 0$,

according to the general theory developed in [18].
We recall that, in general, given $f, g : \mathbb{R} \to \mathbb{R}$, we write $f(x) = \mathcal{O}(g(x))$, as $x \to +\infty$ (or $x \to -\infty$), if there exist $K, c > 0$ such that $|f(x)| \le K\,|g(x)|$, for every $x > c$ (or $x < -c$, respectevely).
The density function $\phi_\sigma$ generated by $\sigma$ is defined as follows:

$$\phi_\sigma(x) := \frac{1}{2}[\sigma(x+1) - \sigma(x-1)], \qquad x \in \mathbb{R}. \tag{4}$$

**Definition 3.1.** *[[18]] Let $\sigma$ be a sigmoidal function assumed as above. Let $n \in \mathbb{N}$ such that $\lceil na \rceil \le \lfloor nb \rfloor - 1$, where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denote the "ceiling" and the "integer part" of a given number and $a$, $b$ are real numbers. We define the Kantorovich NN operators, by:*

$$(K_n f)(x) = \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \left[ n \int_{k/n}^{(k+1)/n} f(u)\, du \right] \phi_\sigma(nx - k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \phi_\sigma(nx - k)}, \qquad x \in I := [a, b], \tag{5}$$

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

From now on, by the symbol $I$ we denote (as in Definition 3.1) the interval $[a, b]$.
The definition of $\phi_\sigma$ has been first introduced in [7, 8] in the cases of the logistic and the hyperbolic tangent functions (such examples are also discussed in Section 4.3), while it has been extended in [18] to every sigmoidal function satisfying the previous conditions $(S_i)$, $i = 1, 2, 3$.

In the following lemma we can find some well-known properties of $\phi_\sigma$ established in [18] and [20], that can be useful in order to study the approximation properties of the operators $K_n$.

**Lemma 3.2.** *$(i)$ $\phi_\sigma(x) \ge 0$ for every $x \in \mathbb{R}$, with $\phi_\sigma(2) > 0$, and moreover $\lim_{x\to\pm\infty} \phi_\sigma(x) = 0$;*

*$(ii)$ The function $\phi_\sigma(x)$ is even;*

*$(iii)$ The function $\phi_\sigma(x)$:*

$$\text{is non-decreasing for } x < 0 \text{ and non-increasing for } x \ge 0; \tag{6}$$

*$(iv)$ Let $\alpha$ be the positive constant of condition $(S3)$. Then:*

$$\phi_\sigma(x) = \mathcal{O}(|x|^{-\alpha-1}), \quad as \quad x \to \pm\infty. \tag{7}$$

*Hence, it turns out that $\phi_\sigma \in L^1(\mathbb{R})$;*

*$(v)$ For every $x \in \mathbb{R}$,*

$$\sum_{k\in\mathbb{Z}} \phi_\sigma(x - k) = 1, \tag{8}$$

*and*

$$\|\phi_\sigma\|_1 = \int_{\mathbb{R}} \phi_\sigma(x)\, dx = 1; \tag{9}$$

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

$$\sum_{k=-n}^{n-1} \phi_\sigma(nx - k) \ge \phi_\sigma(2) > 0. \tag{10}$$

**Remark 3.3.** Note that, the assumed technical condition $\sigma(1) < 1$ is necessary to prove that $\phi_\sigma(2) > 0$ (see (i) of Lemma 3.2). Indeed, if we suppose by contradiction that $\sigma(3) = \sigma(1)$, by the Lagrange theorem we get $\sigma'(x) = 0$ for every $x \in [1, 3]$ and hence $\sigma(x) = \sigma(1)$, for every $x \in [1, 3]$. In particular, being $\sigma'(x) \geq 0$ and non-decreasing for $x \geq 0$ (from $(S2)$), it turns out that $\sigma'(x) = 0$ for every $x \geq 1$ and hence $\sigma(x) = \sigma(1) < 1$, for every $x \geq 1$. Therefore, we reach a contradiction, in view of the fact that $\lim_{x\to+\infty} \sigma(x) = 1$.
Obviously, if $\sigma$ is strictly increasing, the technical condition $\sigma(1) < 1$ can be omitted.

Notice that, $K_n$ are well-defined, e.g., for $f \in L^\infty(I)$; this is a consequence of (10) and of the following estimate:

$$|(K_n f)(x)| \;\leq\; \|f\|_\infty \;<\; +\infty, \quad x \in I, \quad n \in \mathbb{N}. \tag{11}$$

We now recall the useful notion of the *discrete absolute moment of order* $\nu \geq 0$ of $\phi_\sigma$ ([18]), i.e.,

$$M_\nu(\phi_\sigma) \;:=\; \sup_{u\in\mathbb{R}} \sum_{k\in\mathbb{Z}} \phi_\sigma(u-k)\,|u-k|^\nu; \tag{12}$$

in particular, by (8), one can easily observe that $M_0(\phi_\sigma) = 1$. Under the above assumptions on $\sigma$ there holds:

$$M_\nu(\phi_\sigma) \;<\; +\infty, \qquad 0 \leq \nu < \alpha, \tag{13}$$

(see, e.g., [15, 16]). The NN Kantorovich operators are well-defined not only in $L^\infty(I)$, but in every Lebesgue space $L^p(I)$, with $1 \leq p < +\infty$. Furthermore, a norm convergence theorem in this context has been established (see, e.g., [20]).

**Theorem 3.4** ([20])**.** *Let $f \in L^p(I)$, with $1 \leq p < +\infty$. Thus, the following inequality holds*

*(i)* $\|K_n f\|_p \;\leq\; \phi_\sigma(2)^{-1/p} \cdot \|f\|_p$, *i.e., Kantorovich NN operators turn out to be linear and bounded from $L^p(I)$ into itself.*

*Moreover, a convergence theorem with respect to the $L^p$-norm holds, that is*

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

Finally, we highlight that also a uniform convergence theorem for Kantorovich NN operators has been provided in the space $C(I)$ (see [20]).

# 4 Asymptotic analysis in $L^p$-spaces

Here, we study the asymptotic behaviour of the Kantorovich NN operators when functions belonging to Sobolev spaces are considered. This leads to establish suitable quantitative estimates in terms of the aforementioned modulus of smoothness; this represents the main goal of the present paper. Such result is useful to give a measure of the approximation error of the Kantorovich NN operators with respect to the $L^p$-convergence. To achieve this, we present two complementary approaches, that will be presented below in details.

## 4.1 A first approach employing the Hardy-Littlewood maximal function

Firstly, we have to recall an important tool of Functional Analysis, that plays a crucial role in the following. It is the well-known Hardy–Littlewood maximal function (HL maximal function, [1]), that is defined (in its uncentered version) as

$$\mathcal{M}f(x) := \sup_{\substack{u\in[a,b]\\ u\neq x}} \frac{1}{|x-u|} \int_x^u |f(t)|\,dt, \qquad x \in [a, b],$$

for a locally integrable function $f : [a, b] \to \mathbb{R}$. The celebrated theorem of Hardy, Littlewood and Wiener asserts that $\mathcal{M}f$ is bounded on $L^p(\mathbb{R})$ for $1 < p \leq +\infty$ for every $f \in L^p(\mathbb{R})$, namely

$$\|\mathcal{M}f\|_p \leq C_p \|f\|_p, \tag{14}$$

where the constant $C_p$ depends only on $p$ (Theorem I.1 of [37]). Furthermore, it is well-known that the corresponding result for $p = 1$ fails.
In this section, we will consider two classes of sigmoidal functions satisfying the decay condition $(S3)$ for the following values of $\alpha > 0$ and $1 < p \leq +\infty$. In particular, if $1 < p < +\infty$, we define

$$\mathcal{D}(\alpha, p) := \{\sigma \mid \sigma \text{ is a sigmoidal function satisfying } (S3) \text{ for } \alpha > p\},$$

while if $p = +\infty$, we introduce

$$\mathcal{D}(\alpha, +\infty) := \{\sigma \mid \sigma \text{ is a sigmoidal function satisfying } (S3) \text{ for } \alpha > 1\}.$$

Now, we are able to present a first asymptotic estimate for the Kantorovich NN operators.

**Theorem 4.1.** *Let $1 < p \leq +\infty$ and $\sigma \in \mathcal{D}(\alpha, p)$. Thus, for every function $f \in W^{1,p}(I)$, there holds*

$$\|K_n f - f\|_p \leq \mu_p \cdot \frac{\|f'\|_p}{n} < +\infty,$$

*for $n \in \mathbb{N}$, with*

$$\mu_p := C_p^{\frac{1}{p}} \left( \frac{2^{p-1}}{\phi_\sigma(2)} \right)^{\frac{1}{p}} \left[ \frac{1}{p+1} + M_p(\phi_\sigma) \right]^{\frac{1}{p}},$$

*if $1 < p < +\infty$, and*

$$\mu_\infty := C_\infty \frac{1 + 2M_1(\phi_\sigma)}{2\phi_\sigma(2)},$$

*where the constants $C_p$ are those arising from (14).*

*Proof.* Let $f \in W^{1,p}(I)$, with $p > 1$, $x \in I$ and $n \in \mathbb{N}$ be fixed. Firstly, we suppose $1 < p < +\infty$. We now expand the function $f$ according to the first-order Taylor formula in Sobolev spaces with integral remainder (see, e.g., equation (5.6) of [23], p. 37), that is

$$f(u) = f(x) + \int_x^u f'(t)dt, \qquad x, u \in I. \tag{15}$$

Thus, by (5) and (15), we obtain

$$\begin{aligned}
(K_n f)(x) &= \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \left\{ n \int_{k/n}^{(k+1)/n} \left[ f(x) + \int_x^u f'(t)dt \right] du \right\} \phi_\sigma(nx - k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \phi_\sigma(nx - k)} \\
&= f(x) + \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \left\{ n \int_{k/n}^{(k+1)/n} \left[ \int_x^u f'(t)dt \right] du \right\} \phi_\sigma(nx - k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \phi_\sigma(nx - k)} \\
&=: f(x) + R_{1,n}.
\end{aligned}$$

Therefore, the problem reduces to estimate the $L^p$-norm of the remainder $R_{1,n}$. To this aim, we use Jensen inequality twice, property (vi) of Lemma 3.2 and the boundedness of the Hardy-Littlewood maximal function (14) as follows

$$
\begin{aligned}
\|R_{1,n}\|_p^p &= \int_a^b \left| \frac{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left\{ n \int_{k/n}^{(k+1)/n} \left[ \int_x^u f'(t)dt \right] du \right\} \phi_\sigma(nx-k)}{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \phi_\sigma(nx-k)} \right|^p dx \\
&\leq \int_a^b \frac{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left| n \int_{k/n}^{(k+1)/n} \left[ \int_x^u |f'(t)|\, dt \right] du \right|^p \phi_\sigma(nx-k)}{\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \phi_\sigma(nx-k)} dx \\
&\leq \frac{1}{\phi_\sigma(2)} \int_a^b \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left| n \int_{k/n}^{(k+1)/n} \left[ \int_x^u |f'(t)|\, dt \right] du \right|^p \phi_\sigma(nx-k)\, dx \\
&\leq \frac{1}{\phi_\sigma(2)} \int_a^b \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left\{ n \int_{k/n}^{(k+1)/n} \left| \int_x^u |f'(t)|\, dt \right|^p du \right\} \phi_\sigma(nx-k)\, dx \\
&\leq \frac{1}{\phi_\sigma(2)} \int_a^b \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left\{ n \int_{k/n}^{(k+1)/n} |u-x|^p \left|\mathcal{M}f'(x)\right|^p du \right\} \phi_\sigma(nx-k)\, dx. \\
&= \frac{1}{\phi_\sigma(2)} \int_a^b \left|\mathcal{M}f'(x)\right|^p \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left\{ n \int_{k/n}^{(k+1)/n} |u-x|^p\, du \right\} \phi_\sigma(nx-k)\, dx.
\end{aligned}
$$

Now, we note that for any $\gamma \geq 1$, using the convexity of $|\cdot|^\gamma$, the following general estimate holds

$$
\begin{aligned}
n \int_{\frac{k}{n}}^{\frac{k+1}{n}} |u-x|^\gamma du &\leq n \int_{\frac{k}{n}}^{\frac{k+1}{n}} \left| \frac{2}{2}\left(u - \frac{k}{n}\right) + \frac{2}{2}\left(\frac{k}{n} - x\right) \right|^\gamma du \\
&\leq 2^{\gamma-1} n \left[ \int_{\frac{k}{n}}^{\frac{k+1}{n}} \left(u - \frac{k}{n}\right)^\gamma du + \left|\frac{k}{n} - x\right|^\gamma \frac{1}{n} \right] \qquad (16) \\
&= 2^{\gamma-1} \left[ \frac{1}{\gamma+1} \left(\frac{1}{n}\right)^\gamma + |k-nx|^\gamma \frac{1}{n^\gamma} \right] \\
&= \frac{2^{\gamma-1}}{n^\gamma} \left[ \frac{1}{\gamma+1} + |k-nx|^\gamma \right].
\end{aligned}
$$

Hence, using (16) with $\gamma = p$, we obtain

$$
\begin{aligned}
\|R_{1,n}\|_p^p &\leq \frac{2^{p-1}}{\phi_\sigma(2)} \int_a^b \frac{|\mathcal{M}f'(x)|^p}{n^p} \left[ \frac{M_0(\phi_\sigma)}{p+1} + M_p(\phi_\sigma) \right] dx \\
&= \frac{2^{p-1}}{\phi_\sigma(2)} \left[ \frac{1}{p+1} + M_p(\phi_\sigma) \right] \frac{\|\mathcal{M}f'\|_p^p}{n^p} \\
&\leq \frac{2^{p-1}}{\phi_\sigma(2)} \left[ \frac{1}{p+1} + M_p(\phi_\sigma) \right] \frac{C_p \|f'\|_p^p}{n^p} < +\infty,
\end{aligned}
$$

where $C_p$ is the constant arising from (14) and the moment $M_p(\phi_\sigma)$ is finite being $\sigma \in \mathcal{D}(\alpha, p)$ and using (13).
The case $p = +\infty$ is analogous. In particular, by using (5), (15), property (vi) of Lemma 3.2, (16) with $\gamma = 1$, and (14), we get

$$
\begin{aligned}
|R_{1,n}(x)| &\leq \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \left\{ n \int_{k/n}^{(k+1)/n} \left| \int_x^u |f'(t)| \, dt \right| du \right\} \phi_\sigma(nx - k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \phi_\sigma(nx - k)} \\
&\leq \frac{1}{\phi_\sigma(2)} \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \left\{ n \int_{k/n}^{(k+1)/n} |u - x| \, |\mathcal{M}f'(x)| \, du \right\} \phi_\sigma(nx - k) \\
&\leq \frac{\|\mathcal{M}f'\|_\infty}{\phi_\sigma(2)} \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \left\{ n \int_{k/n}^{(k+1)/n} |u - x| \, du \right\} \phi_\sigma(nx - k) \\
&\leq \frac{\|\mathcal{M}f'\|_\infty}{n\phi_\sigma(2)} \left[ \frac{1}{2} + M_1(\phi_\sigma) \right] \\
&\leq \frac{C_\infty}{n\phi_\sigma(2)} \left[ \frac{1}{2} + M_1(\phi_\sigma) \right] \|f'\|_\infty < +\infty,
\end{aligned}
$$

where $C_\infty$ is the constant arising from (14) and $M_1(\phi_\sigma) < +\infty$ being $\sigma \in \mathcal{D}(\alpha, \infty)$ and using again (13). This completes the proof. $\square$

Now, in order to establish the theorem concerning quantitative estimates based in terms of the modulus of smoothness, we use the following useful result, that can be found, e.g., in the following book of B. Sendov and V.A. Popov of 1988.

**Theorem 4.2** ([35]). *Let $f$ be a function belonging to $L^p(I)$, $1 \leq p < +\infty$ (or $f \in C(I)$ with the norm $\|\cdot\|_\infty$). For every integer $k > 0$ and every $h$ such that $0 < h \leq (b-a)/k$, there exists a function $f_{k,h} \in L^p(\mathbb{R})$ if $1 \leq p < +\infty$ (or $f_{h,k} \in C(I)$) satisfying*

*(i) $\|f - f_{k,h}\|_p \leq c_1(k)\omega_k(f, h)_p$ (with $p = +\infty$ when $f \in C(I)$);*

*(ii) $f_{h,k} \in W^{k,p}(I)$ if $1 \leq p < +\infty$ (or $f_{h,k} \in C^k(I)$) and*

$$\left\| f_{k,h}^{(s)} \right\|_p \leq c_2(k) h^{-s} \omega_s(f, h)_p, \qquad s = 1, \ldots, k,$$

*(with $p = +\infty$ when $f \in C(I)$) where $c_1(k)$ and $c_2(k)$ are suitable constants depending only on $k$.*

**Remark 4.3.** The proof of Theorem 4.2 is constructive and functions $f_{k,h}$ can be taken as the well-known *Steklov functions*. For further details, please see pp. 31-34 of [35] or [14].

Now, we are ready to state one of the main theorem of the present section, for the case $p > 1$, providing the desired quantitative estimates in terms of the modulus of smoothness for the Kantorovich NN operators.

**Theorem 4.4.** *Let $1 < p < +\infty$ and $\sigma \in \mathcal{D}(\alpha, p)$. Thus, for every function $f \in L^p(I)$, there holds*

$$\|K_n f - f\|_p \leq \lambda_p \cdot \omega\left(f, \frac{1}{n}\right)_p,$$

*while if $f \in C(I)$ and $\sigma \in \mathcal{D}(\alpha, +\infty)$, we have*

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

*for a sufficiently large $n \in \mathbb{N}$, where $\lambda_p > 0$, $1 < p \leq +\infty$, are suitable constants.*

*Proof.* Let $f \in L^p(I)$, with $1 < p < +\infty$, or $f \in C(I)$, be fixed. By Theorem 4.2 (i), for every $0 < h \leq b-a$, there exists $f_{1,h} \in W^{1,p}(I)$ (or $f_{1,h} \in C^1(I)$) such that

$$\|f - f_{1,h}\|_p \leq c_1(1)\omega(f,h)_p, \tag{17}$$

(with $p = +\infty$ here and also in the estimates below, when $f \in C(I)$).
Moreover, since $f - f_{1,h} \in L^p(I)$ (or $f - f_{1,h} \in C(I)$) we deduce from (i) of Theorem 3.4 and (11) that

$$\|K_n(f - f_{1,h})\|_p \leq A_p \|f - f_{1,h}\|_p, \tag{18}$$

where

$$A_p := \phi(2)^{-\frac{1}{p}}, \text{ if } 1 < p < +\infty, \text{ and } A_\infty := 1, \text{ if } p = +\infty. \tag{19}$$

Then, by the linearity of the operators $K_n$ and in view of (17) and (18), we have

$$\begin{aligned}
\|K_n f - f\|_p &\leq \|K_n f - K_n f_{1,h}\|_p + \|K_n f_{1,h} - f_{1,h}\|_p + \|f_{1,h} - f\|_p \\
&= \|K_n(f - f_{1,h})\|_p + \|K_n f_{1,h} - f_{1,h}\|_p + \|f_{1,h} - f\|_p \\
&\leq A_p\|f_{1,h} - f\|_p + \|K_n f_{1,h} - f_{1,h}\|_p + \|f_{1,h} - f\|_p \\
&= (A_p + 1)\,\|f_{1,h} - f\|_p + \|K_n f_{1,h} - f_{1,h}\|_p \\
&\leq c_1(1)\,(A_p + 1)\,\omega(f,h)_p + \|K_n f_{1,h} - f_{1,h}\|_p.
\end{aligned}$$

Now, by using Theorem 4.1 applied to $f_{1,h} \in W^{1,p}(I)$ (or $f_{1,h} \in C^1(I)$) and also Theorem 4.2 (ii), we can write

$$\|K_n f_{1,h} - f_{1,h}\|_p \leq \frac{\mu_p}{n} \cdot \|f'_{1,h}\|_p \leq \frac{\mu_p}{n} \cdot c_2(1) h^{-1} \omega(f,h)_p,$$

where $\mu_p$ is the constant arising from Theorem 4.1.
In summary, the following estimate holds

$$\|K_n f - f\|_p \leq c_1(1)\,(A_p + 1)\,\omega(f,h)_p + \frac{\mu_p}{n} \cdot c_2(1) h^{-1} \omega(f,h)_p,$$

with $1 < p \leq +\infty$. Now, considering $h = \frac{1}{n} \leq b - a$, with $n \in \mathbb{N}$, we finally get

$$\begin{aligned}
\|K_n f - f\|_p &\leq [c_1(1)\,(A_p + 1) + c_2(1)\mu_p] \cdot \omega\left(f, \frac{1}{n}\right)_p \\
&=: \lambda_p \cdot \omega\left(f, \frac{1}{n}\right)_p,
\end{aligned}$$

for every sufficiently large $n \in \mathbb{N}$. This completes the proof. □

For an exhaustive study of the order of approximation, we also provide a qualitative result thanks to Theorem 4.4.
For this purpose, we have to recall the definition of *Lipschitz classes* in terms of the first-order modulus of smoothness, which are defined as

$$Lip(\nu, p) = \left\{ f \in L^p(I) : \omega(f,h)_p = \mathcal{O}(h^\nu), \text{ as } h \to 0^+ \right\}, \tag{20}$$

with $0 < \nu \leq 1$, $1 \leq p < +\infty$, while

$$Lip(\nu, +\infty) = \left\{ f \in C(I) : \omega(f,h)_\infty = \mathcal{O}(h^\nu), \text{ as } h \to 0^+ \right\}, \tag{21}$$

see, e.g., [23]).
Therefore, we can directly deduce a qualitative estimate for the order of approximation when $p > 1$, as a consequence of Theorem 4.4.

**Corollary 4.5.** *Let $1 < p \leq +\infty$ and $\sigma \in \mathcal{D}(\alpha, p)$. Thus, for every function $f \in Lip(\nu, p)$, with $0 < \nu \leq 1$, there holds*

$$\|K_n f - f\|_p \leq C \lambda_p n^{-\nu},$$

*for every sufficiently large $n \in \mathbb{N}$, where the positive constants $\lambda_p$ and $C$ arise from Theorem 4.4 and definition (20), respectively.*

### 4.2 A complementary approach

Now, we want to establish a quantitative estimate also for the case $p = 1$, since we already know from Theorem 3.4 (ii) that the convergence also holds in this case. Therefore, to remove the restriction $p > 1$ from Theorem 4.1, we need to avoid the use in the proof of the HL maximal function, and we must adopt a strategy in which we have to expand the involved function by the Taylor formula with integral remainder up to the second order. Moreover, we also have to assume a slightly stronger version of condition $(S3)$ with respect to the previous cases (see Theorem 4.1 and Theorem 4.4). In fact, in this section we will consider classes of sigmoidal functions having a suitable order of decay according to condition $(S3)$ for $\alpha > 0$, as follows:

$$\widetilde{\mathcal{D}}(\alpha, p) := \{\sigma \mid \sigma \text{ is a sigmoidal function satisfying } (S3) \text{ for } \alpha > 2p - 1\},$$

when $1 \leq p < +\infty$, and

$$\widetilde{\mathcal{D}}(\alpha, +\infty) := \{\sigma \mid \sigma \text{ is a sigmoidal function satisfying } (S3) \text{ for } \alpha > 2\},$$

when $p = +\infty$.

We can prove the following.

**Theorem 4.6.** *Let $1 \leq p < +\infty$ and $\sigma \in \widetilde{\mathcal{D}}(\alpha, p)$. Thus, for every function $f \in W^{2,p}(I)$, there holds*

$$\|K_n f - f\|_p \leq \mu_{1,p} \cdot \frac{\|f'\|_p}{n} + \mu_{2,p} \cdot \frac{\|f''\|_p}{n^{\frac{2p-1}{p}}} < +\infty,$$

*for $n \in \mathbb{N}$, where*

$$\mu_{1,p} := \left(\frac{4^{p-1}}{\phi(2)}\right)^{\frac{1}{p}} \left[\frac{1}{p+1} + M_p(\phi_\sigma)\right]^{\frac{1}{p}},$$

*and*

$$\mu_{2,p} := (b-a)^{\frac{1}{p}} \left(\frac{8^{p-1}}{\phi(2)}\right)^{\frac{1}{p}} \left[\frac{1}{2p} + M_{2p-1}\right]^{\frac{1}{p}}.$$

*Furthermore, if $\sigma \in \widetilde{\mathcal{D}}(\alpha, +\infty)$, for every function $f \in W^{2,\infty}(I)$, there holds*

$$\|K_n f - f\|_\infty \leq \mu_{1,\infty} \cdot \frac{\|f'\|_\infty}{n} + \mu_{2,\infty} \cdot \frac{\|f''\|_\infty}{n^2} < +\infty,$$

*for $n \in \mathbb{N}$, where*

$$\mu_{1,\infty} := \frac{2M_1(\phi_\sigma) + 1}{2\phi_\sigma(2)},$$

*and*

$$\mu_{2,\infty} := \frac{1 + 3M_2(\phi_\sigma)}{3\phi_\sigma(2)}.$$

*Proof.* Let $f \in W^{2,p}(I)$, with $1 \leq p < +\infty$, $x \in I$ and $n \in \mathbb{N}$ be fixed. First, we expand the function $f$ according to the second-order Taylor formula in Sobolev spaces with integral remainder (see again [23]), that is

$$f(u) = f(x) + f'(x)(u - x) + \int_x^u f''(t)(u - t)dt, \qquad x, u \in I. \tag{22}$$

Thus, by (5) and (22), we obtain

$$(K_n f)(x) = \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \left\{ n \int_{k/n}^{(k+1)/n} \left[ f(x) + f'(x)(u - x) + \int_x^u f''(t)(u - t)dt \right] du \right\} \phi_\sigma(nx - k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \phi_\sigma(nx - k)}$$

$$= f(x) + f'(x) \, \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \left\{ n \int_{k/n}^{(k+1)/n} (u-x)\, du \right\} \phi_\sigma (nx-k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \phi_\sigma(nx-k)}$$

$$+ \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \left\{ n \int_{k/n}^{(k+1)/n} \left[ \int_x^u f''(t)(u-t)dt \right] du \right\} \phi_\sigma (nx-k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \phi_\sigma(nx-k)}.$$

Now, exploiting the convexity of the function $|\cdot|^p$ when $p \geq 1$, we have

$$\|K_n f - f\|_p^p \leq 2^{p-1} \int_a^b \left| f'(x) \right|^p \left| \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \left\{ n \int_{k/n}^{(k+1)/n} (u-x)\, du \right\} \phi_\sigma (nx-k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \phi_\sigma(nx-k)} \right|^p dx$$

$$+ 2^{p-1} \int_a^b \left| \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \left\{ n \int_{k/n}^{(k+1)/n} \left[ \int_x^u f''(t)(u-t)dt \right] du \right\} \phi_\sigma (nx-k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \phi_\sigma(nx-k)} \right|^p dx$$

$$=: 2^{p-1} \left( I_{1,n} + I_{2,n} \right).$$

We now focus on $I_{1,n}$. By using Jensen inequality twice, property (vi) of Lemma 3.2, the convexity of $|\cdot|^p$ and (16) with $\gamma = p$, we get

$$I_{1,n} = \int_a^b \left| f'(x) \right|^p \left| \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \left\{ n \int_{k/n}^{(k+1)/n} (u-x)\, du \right\} \phi_\sigma (nx-k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \phi_\sigma(nx-k)} \right|^p dx$$

$$\leq \int_a^b \left| f'(x) \right|^p \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \left| n \int_{k/n}^{(k+1)/n} |u-x|\, du \right|^p \phi_\sigma (nx-k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \phi_\sigma(nx-k)} dx$$

$$\leq \int_a^b \left| f'(x) \right|^p \frac{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \left\{ n \int_{k/n}^{(k+1)/n} |u-x|^p\, du \right\} \phi_\sigma (nx-k)}{\displaystyle\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor -1} \phi_\sigma(nx-k)} dx$$

$$\leq \frac{2^{p-1}}{\phi(2)} \int_a^b \frac{|f'(x)|^p}{n^p} \left[ \frac{1}{p+1} + M_p(\phi_\sigma) \right] dx$$

$$\leq \frac{2^{p-1}}{\phi(2)} \left[ \frac{1}{p+1} + M_p(\phi_\sigma) \right] \frac{\|f'\|_p^p}{n^p} < +\infty,$$

where $M_p(\phi_\sigma) < +\infty$ being $\sigma \in \widetilde{\mathcal{D}}(\alpha, p)$ and taking into account (13). Moreover, $\|f'\|_p < +\infty$ since $f \in W^{2,p}(I)$.
Now, we estimate $I_{2,n}$. Here, we use Jensen inequality three times (the first two times as in the estimate of $I_{1,n}$) together with property (vi) of Lemma 3.2 and the convexity of $|\cdot|^p$. We have

$$\begin{aligned}
I_{2,n} &\leq \int_a^b \frac{\displaystyle\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left| n \int_{k/n}^{(k+1)/n} \left[ \int_x^u f''(t)(u-t)dt \right] du \right|^p \phi_\sigma(nx-k)}{\displaystyle\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \phi_\sigma(nx-k)} dx \\
&\leq \frac{1}{\phi(2)} \int_a^b \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left| n \int_{k/n}^{(k+1)/n} |u-x| \left[ \int_x^u |f''(t)|\, dt \right] du \right|^p \phi_\sigma(nx-k)\, dx \\
&\leq \frac{1}{\phi(2)} \int_a^b \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left\{ n \int_{k/n}^{(k+1)/n} |u-x|^p \left| \int_x^u |f''(t)|\, dt \right|^p du \right\} \phi_\sigma(nx-k)\, dx \\
&= \frac{1}{\phi(2)} \int_a^b \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left\{ n \int_{k/n}^{(k+1)/n} |u-x|^{2p} \left| \frac{1}{|u-x|} \int_x^u |f''(t)|\, dt \right|^p du \right\} \phi_\sigma(nx-k)\, dx \\
&\leq \frac{1}{\phi(2)} \int_a^b \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left\{ n \int_{k/n}^{(k+1)/n} |u-x|^{2p-1} \left[ \int_x^u |f''(t)|^p\, dt \right] du \right\} \phi_\sigma(nx-k)\, dx \\
&\leq \frac{\|f''\|_p^p}{\phi(2)} \int_a^b \sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left\{ n \int_{k/n}^{(k+1)/n} |u-x|^{2p-1}\, du \right\} \phi_\sigma(nx-k)\, dx.
\end{aligned}$$

Now, considering the estimate given in (16) with $\gamma = 2p-1$, we obtain

$$I_{2,n} \leq \frac{2^{2p-2}}{\phi(2)} \frac{\|f''\|_p^p}{n^{2p-1}} \left[ \frac{1}{2p} + M_{2p-1}(\phi_\sigma) \right] (b-a) < +\infty,$$

where $M_{2p-1}(\phi_\sigma) < +\infty$ since condition $(S3)$ holds for $\alpha > 2p-1$, being $\sigma \in \widetilde{\mathcal{D}}(\alpha, p)$. Rearranging all the above estimates, we finally obtain the desired inequality for $1 \leq p < +\infty$.
Now, we want to study the case $p = +\infty$. Proceeding as above, we obtain

$$\begin{aligned}
|(K_n f)(x) - f(x)| &\leq |f'(x)| \left| \frac{\displaystyle\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left\{ n \int_{k/n}^{(k+1)/n} (u-x)\, du \right\} \phi_\sigma(nx-k)}{\displaystyle\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \phi_\sigma(nx-k)} \right| \\
&\quad + \left| \frac{\displaystyle\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \left\{ n \int_{k/n}^{(k+1)/n} \left[ \int_x^u f''(t)(u-t)dt \right] du \right\} \phi_\sigma(nx-k)}{\displaystyle\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \phi_\sigma(nx-k)} \right| \\
&=: J_{1,n} + J_{2,n}.
\end{aligned}$$

By using property (vi) of Lemma 3.2 and (16) with $\gamma = 1$, we immediately get

$$J_{1,n} \leq \frac{|f'(x)|}{\phi_\sigma(2)} \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \left\{ n \int_{k/n}^{(k+1)/n} |u - x| \, du \right\} \phi_\sigma(nx - k)$$
$$\leq \frac{\|f'\|_\infty}{2\phi(2)} \frac{2M_1(\phi_\sigma) + 1}{n} < +\infty,$$

being $\sigma \in \widetilde{\mathcal{D}}(\alpha, +\infty)$. Moreover, applying Jensen inequality three times, property (vi) of Lemma 3.2 and (16) with $\gamma = 2$, we obtain

$$J_{2,n} \leq \frac{1}{\phi(2)} \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \left\{ n \int_{k/n}^{(k+1)/n} |u - x| \left[ \int_x^u |f''(t)| \, dt \right] du \right\} \phi_\sigma(nx - k)$$

$$\leq \frac{\|f''\|_\infty}{\phi(2)} \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \left\{ n \int_{k/n}^{(k+1)/n} |u - x|^2 du \right\} \phi_\sigma(nx - k)$$
$$\leq \frac{2\|f''\|_\infty}{3\phi(2)} \frac{1 + 3M_2(\phi_\sigma)}{n^2} < +\infty,$$

being $\sigma \in \widetilde{\mathcal{D}}(\alpha, p)$. Now, the proof easily follows passing to the supremum with respect to $x \in I$. □

Herein, we give the main theorem including also the case $p = 1$.

**Theorem 4.7.** *Let $1 \leq p < +\infty$ and $\sigma \in \widetilde{\mathcal{D}}(\alpha, p)$. Thus, for every function $f \in L^p(I)$, there holds*

$$\|K_n f - f\|_p \leq \lambda_{1,p} \cdot \omega\left(f, \frac{1}{n^{1-\frac{1}{2p}}}\right)_p + \lambda_{2,p} \cdot \omega_2\left(f, \frac{1}{n^{1-\frac{1}{2p}}}\right)_p,$$

*for every sufficiently large $n \in \mathbb{N}$, where $\lambda_{1,p}$ and $\lambda_{2,p}$ are two suitable positive constants. Moreover, if $f \in C(I)$ and $\sigma \in \widetilde{\mathcal{D}}(\alpha, +\infty)$, there holds*

$$\|K_n f - f\|_\infty \leq \lambda_{1,\infty} \cdot \omega\left(f, \frac{1}{n}\right)_\infty + \lambda_{2,\infty} \cdot \omega_2\left(f, \frac{1}{n}\right)_\infty,$$

*for every sufficiently large $n \in \mathbb{N}$, where $\lambda_{1,\infty}$ and $\lambda_{2,\infty}$ are other suitable positive constants.*

*Proof.* Let $f \in L^p(I)$ be fixed, with $1 \leq p < +\infty$. Arguing as in the first part of the proof of Theorem 4.4, for every $0 < h \leq \frac{b-a}{2}$, there exists $f_{2,h} \in W^{2,p}(I)$ such that

$$\|K_n f - f\|_p \leq (A_p + 1) \|f_{2,h} - f\|_p + \|K_n f_{2,h} - f_{2,h}\|_p$$
$$\leq c_1(2) (A_p + 1) \, \omega(f, h)_p + \|K_n f_{2,h} - f_{2,h}\|_p,$$

where constants $A_p$ are defined in (19). We first consider the case of $1 \leq p < +\infty$. Here, by using Theorem 4.6 applied to $f_{2,h} \in W^{2,p}(I)$ and then Theorem 4.2 (ii) with $k = 1, 2$, we can write

$$\|K_n f_{2,h} - f_{2,h}\|_p \leq \mu_{1,p} \frac{\|f'_{2,h}\|_p}{n} + \mu_{2,p} \frac{\|f''_{2,h}\|_p}{n^{\frac{2p-1}{p}}}$$
$$\leq \mu_{1,p} \frac{c_2(1) h^{-1}}{n} \cdot \omega(f, h)_p + \mu_{2,p} \frac{c_2(2) h^{-2}}{n^{\frac{2p-1}{p}}} \cdot \omega_2(f, h)_p,$$

where constants $\mu_{1,p}$ and $\mu_{2,p}$ arise from Theorem 4.6.
In summary, we have

$$\|K_n f - f\|_p \leq \left( c_1(2) (A_p + 1) + \mu_{1,p} \frac{c_2(1) h^{-1}}{n} \right) \cdot \omega(f, h)_p + \mu_{2,p} \frac{c_2(2) h^{-2}}{n^{\frac{2p-1}{p}}} \cdot \omega_2(f, h)_p.$$

Now, setting $h = n^{-1+\frac{1}{2p}} \leq \frac{b-a}{2}$, with $n \in \mathbb{N}$, we finally get

$$\begin{aligned}\|K_n f - f\|_p &\leq (c_1(2)\,(A_p+1) + c_2(1)\mu_{1,p}) \cdot \omega\left(f, \frac{1}{n^{1-\frac{1}{2p}}}\right)_p + c_2(2)\mu_{2,p} \cdot \omega_2\left(f, \frac{1}{n^{1-\frac{1}{2p}}}\right)_p \\ &=: \lambda_{1,p} \cdot \omega\left(f, \frac{1}{n^{1-\frac{1}{2p}}}\right)_p + \lambda_{2,p} \cdot \omega_2\left(f, \frac{1}{n^{1-\frac{1}{2p}}}\right)_p,\end{aligned}$$

for every sufficiently large $n \in \mathbb{N}$.
The case $f \in C(I)$ is analogous. Here, by Theorem 4.2 (i) and (ii) and Theorem 4.6, we similarly obtain

$$\|K_n f - f\|_\infty \leq \left(2c_1(2) + \mu_{1,\infty}\frac{c_2(1)h^{-1}}{n}\right) \cdot \omega(f,h)_\infty + \mu_{2,\infty}\frac{c_2(2)h^{-2}}{n^2} \cdot \omega_2(f,h)_\infty.$$

where constants $\mu_{1,\infty}$ and $\mu_{2,\infty}$ arise again from Theorem 4.6. In this case, we put $h = n^{-1} \leq \frac{b-a}{2}$, with $n \in \mathbb{N}$ in order to get

$$\begin{aligned}\|K_n f - f\|_\infty &\leq (2c_1(2) + \mu_{1,\infty}c_2(1)) \cdot \omega\left(f, \frac{1}{n}\right)_\infty + c_2(2)\mu_{2,\infty} \cdot \omega_2\left(f, \frac{1}{n}\right)_\infty \\ &=: \lambda_{1,\infty} \cdot \omega\left(f, \frac{1}{n}\right)_\infty + \lambda_{2,\infty} \cdot \omega_2\left(f, \frac{1}{n}\right)_\infty,\end{aligned}$$

for every sufficiently large $n \in \mathbb{N}$. This completes the proof. □

Now, we want to deduce a qualitative version of the above estimates as made in Subsection 4.1. To this aim, accordingly to what we have done at the end of the previous subsection, if $p = 1$ we recall that

$$Lip(1,1) = BV(I),$$

where $BV(I)$ denotes the space of bounded variation functions on $I$. In the following, we also use the well-known inequality

$$\omega_r(f,\delta)_p \leq 2^{r-k}\omega_k(f,\delta)_p, \qquad \delta > 0, \tag{23}$$

where $k$, $r$ are positive integers such that $1 \leq k < r$ and $1 \leq p \leq +\infty$ (see, e.g., [23]).

**Corollary 4.8.** *Let $1 \leq p \leq +\infty$ and $\sigma \in \widetilde{\mathcal{D}}(\alpha, p)$. Thus, for every $f \in Lip(\nu, p)$, with $0 < \nu \leq 1$ and $1 \leq p < +\infty$, there holds*

$$\|K_n f - f\|_p \leq \Lambda_p \cdot n^{\nu\left(\frac{1}{2p}-1\right)},$$

*for every sufficiently large $n \in \mathbb{N}$, where the positive constant $\Lambda_p := C\,(\lambda_{1,p} + 2\lambda_{2,p})$ arises from Theorem 4.7, (20) and (23). Moreover, if $f \in Lip(\nu, +\infty)$, with $0 < \nu \leq 1$, there holds*

$$\|K_n f - f\|_\infty \leq \Lambda_\infty \cdot n^{-\nu},$$

*for every sufficiently large $n \in \mathbb{N}$, where the positive constant $\Lambda_\infty := \overline{C}\,(\lambda_{1,\infty} + 2\lambda_{2,\infty})$ arises from Theorem 4.7, (20) and (23).*

**Remark 4.9.** Note that, by Corollary 4.8, we achieve estimates also for the cases $1 < p \leq +\infty$, but they turn out to be worse than those established in Corollary 4.5. Hence, the main usefulness of Corollary 4.8 resides in the estimates that we achieve for the case $p = 1$, that was not covered by Corollary 4.5.

Based on the latter remark, when we deal with $L^p$-estimations with $p > 1$, we will always refer to Corollary 4.5 only.

## 5 Some examples

We point out that in literature there are several kinds of examples of sigmoidal functions satisfying the assumptions here required. Among these, we may recall the well-known *logistic function* defined by

$$\sigma_l(x) := (1 + e^{-x})^{-1}, \qquad x \in \mathbb{R},$$

and the *hyperbolic tangent function*, given by

$$\sigma_h(x) := (\tanh x + 1)/2, \qquad x \in \mathbb{R}.$$

Both of them easily satisfy condition $(S3)$ for every $\alpha > 0$, in view of their exponential decay to zero, as $x \to -\infty$, together with $(S_1)$ and $(S_2)$.

**Remark 5.1.** In the presented results, we consider sigmoidal functions satisfying conditions $(Si)$, for $i = 1, 2, 3$. Notably, the developed theory remains applicable even when assumption $(S2)$ concerning $\sigma$ is omitted and replaced by assuming that $\phi_\sigma$ satisfies condition (6) along with $\phi_\sigma(2) > 0$. This means that the presented theory is versatile and can be applied to a broad range of sigmoidal functions, including those that may not necessarily belong to $C^2(\mathbb{R})$.

An example of non-smooth sigmoidal function is given by the *ramp function* $\sigma_r$, defined as

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

Obviously, condition $(S3)$ is again satisfied for every $\alpha > 0$ and the corresponding $\phi_{\sigma_r}$ is a compactly supported function. For a graphical representation of the above examples, the reader can see Fig. 1.

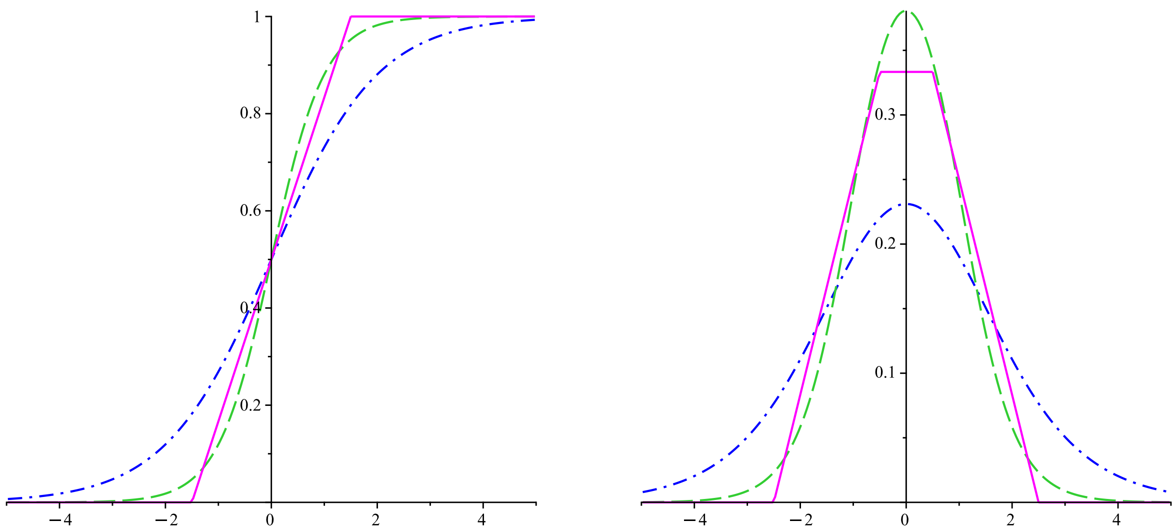


Figure 1: Plots of $\sigma_l$ (blue dashed-dotted line), $\sigma_h$ (green dashed line) and $\sigma_r$ (violet solid line) on the left and the corresponding density functions on the right.

Other non-smooth sigmoidal functions can be generated by the well-known *central B-splines of order* $\beta \in \mathbb{N}$, given by

$$\beta_n(x) := \frac{1}{(n-1)!} \sum_{j=0}^{n} (-1)^j \binom{n}{j} \left( \frac{n}{2} + x - j \right)_+^{n-1}, \qquad x \in \mathbb{R}, \tag{24}$$

where $(\cdot)_+$ denotes the positive part, i.e., $(x)_+ := \max\{x, 0\}$ (see, e.g., [13]). The latter lead to define

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

whose support turns out to be contained in $[k, +\infty)$, where $k > 0$ is a suitable constant (see, e.g, Fig. 2). Therefore, condition $(S3)$ is trivially satisfied for every $\alpha > 0$.

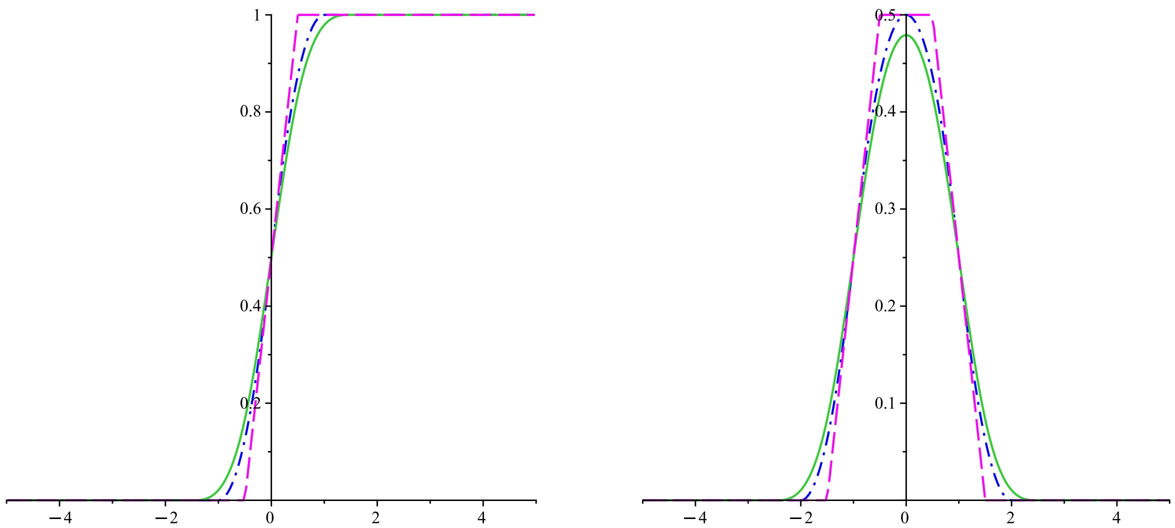


Figure 2: Plots of $\sigma_{\beta_1}$ (violet dashed line), $\sigma_{\beta_2}$ (blue dashed-dotted line) and $\sigma_{\beta_3}$ (green solid line) on the left and the corresponding density functions on the right.

Now we recall a further sigmoidal function depending on a parameter $\gamma > 0$, defined by

$$\sigma_\gamma(x) := \begin{cases} \frac{2^{-\gamma}}{4}|x|^{-\gamma}, & x \leq -\frac{1}{2} \\ \frac{1}{2}x + \frac{1}{2}, & -\frac{1}{2} < x \leq \frac{1}{2} \\ 1 - \frac{2^{-\gamma}}{4}|x|^{-\gamma}, & x > \frac{1}{2}, \end{cases}$$

and the corresponding density function

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

(see, e.g., [9]). If we denote the family of NN Kantorovich operators based on $\phi_{\sigma_\gamma}$ by $K_n^{\phi_\gamma}$, we now explain under which conditions the main results presented here hold.
We start with the case $L^1(I)$, for which the theory developed holds requiring $\gamma > 2$. In fact, this implies that $\sigma_\gamma \in \widetilde{\mathcal{D}}(\alpha, 1)$ and we can apply Theorem 4.7. To extend this to the entire $L^p$-setting when $1 < p < +\infty$ or to $C(I)$, it is necessary to require $\gamma > p+1$ or $\gamma > 2$, respectively (see Fig. 3). In fact, under this assumption, one can see that $\sigma_\gamma \in \mathcal{D}(\alpha, p)$ for every $1 < p \leq +\infty$ and so it is possible to employ Theorem 4.4. Otherwise, if we use Theorem 4.7, the order of decay should be higher, i.e., $\gamma > 2p$, in the $L^p$-case with $1 < p < +\infty$. In this sense, we may observe that Theorem 4.4 allows us to require slightly weaker assumptions on the involved sigmoidal activation functions and this makes the HL-based approach more convenient.
From the above considerations, we can deduce by Theorem 4.4, Theorem 4.7, Corollary 4.5 and Corollary 4.8, the following comprehensive result.

**Corollary 5.2.** *Let $f \in L^1(I)$ and $\gamma > 2$, it turns out that*

$$\|K_n^{\phi_\gamma} f - f\|_1 \leq \lambda_{1,1}\omega\left(f, \frac{1}{\sqrt{n}}\right) + \lambda_{2,1}\omega_2\left(f, \frac{1}{\sqrt{n}}\right),$$

*for every sufficiently large value of $n \in \mathbb{N}$, where $\lambda_{1,1}$ and $\lambda_{2,1}$ are suitable positive constants. In particular, if $f \in Lip(1,1) = BV(I)$ there holds*

$$\|K_n^{\phi_\gamma} f - f\|_1 = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),$$

*as $n \to +\infty$. Furthermore, if $f \in L^p(I)$ with $1 < p < +\infty$ and $\gamma > p+1$, or if $f \in C(I)$ (with $p = +\infty$ below) and $\gamma > 2$, it turns out that*

$$\|K_n^{\phi_\gamma} f - f\|_p \leq \lambda_p \, \omega\left(f, \frac{1}{n}\right)_p,$$

*for every sufficiently large value of $n \in \mathbb{N}$, where $\lambda_p$ is a suitable positive constant. In particular, if $f \in Lip(1,p)$ with $1 < p \leq +\infty$, there holds*

$$\|K_n^{\phi_\gamma} f - f\|_p = \mathcal{O}\left(\frac{1}{n}\right),$$

*as $n \to +\infty$.*

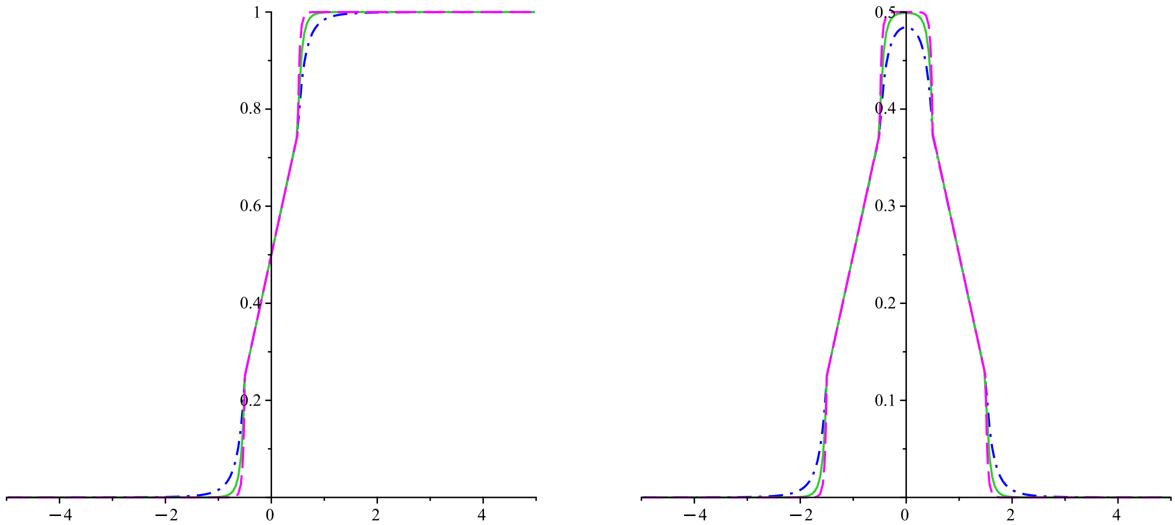


Figure 3: Plots of $\sigma_\gamma$ for $\gamma = 4$ (blue dashed-dotted line), $\gamma = 8$ (green solid line) and $\gamma = 16$ (violet dashed line) on the left and the corresponding density functions $\phi_\gamma$ on the right.

**Remark 5.3.** It is important to remark that the presented theory can be also applied to the well-known ReLU activation function (Rectified Linear Unit) (see [22, 33, 41, 31]), that represents a very attractive tool within the NN approximation theory. We may mention also their $k$-th power, denoted by ReLU$^k$. The latter are also known with the name of Rectified Power Units (RePUs) (see [32, 25]). For further details, see also [13].

## 6 Final remarks and conclusions

The neural network operators studied in the present paper have been introduced in order to establish constructive approximation results by a family of neural networks. The present theory deals with the approximation of functions of one-variable and belongs to very active fields of research. It is well-known that the theory of artificial neural networks is mainly a multivariate theory; indeed the corresponding multivariate version of the operators $F_n$ have been introduced and studied in [20]. Following the approach introduced in the present paper, we can extend the results here established also for the multivariate version of the neural network operators given in [20].

## Acknowledgments

The authors would like to thank the anonymous Reviewer for his/her valuable contribution to the improvement of this paper.
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 network RITA (Research ITalian network on Approximation), and of the UMI (Unione Matematica Italiana) group T.A.A. (Teoria dell'Approssimazione e Applicazioni).

## Funding

The author D. Costarelli has been partially supported within the projects: (1) 2023 GNAMPA-INdAM Project "Approssimazione costruttiva e astratta mediante operatori di tipo sampling e loro applicazioni" CUP E53C22001930001, (2) 2024 GNAMPA-INdAM Project "Tecniche di approssimazione in spazi funzionali con applicazioni a problemi di diffusione", CUP E53C23001670001, and (3) "National Innovation Ecosystem grant ECS00000041 - VITALITY", funded by the European Union - NextGenerationEU under the Italian Ministry of University and Research (MUR). The author M. Piconi has been partially supported within the 2024 GNAMPA-INdAM Project: "DYNAMIChE: DYNAmical Methods: Inverse problems, Chaos and Evolution" (CUP: E53C23001670001). Finally, both the authors have been also partially supported by: 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).

## Conflicts of interests/Competing interests

The authors declare that they have not conflict of interest and/or competing interest.

## Availability of data and material and Code availability

Not applicable.

## References

[1] J.M. Aldaz, L. Colzani, J. Pérez Lázaro, Optimal Bounds on the Modulus of Continuity of the Uncentered Hardy–Littlewood Maximal Function, J. Geom. Anal., 22 (2012), 132-167.

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

[3] S. Bajpeyi, Order of approximation for exponential sampling type neural network operators, Res. Math., 78 (2023), Art. No. 99.

[4] S. Bajpeyi, Approximation by exponential sampling type neural network operators, Analysis Math. Phys., 11 (2021), Art. No. 108.

[5] B. Baxhaku, P.N. Agrawal, Neural network operators with hyperbolic tangent functions, Expert Syst. App., 226 (15) (2023), Art. No. 119996.

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

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

[8] 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.

[9] L. Coroianu, D. Costarelli, Best approximation and inverse results for neural network operators, Res. Math., 79 (2024) 193.

[10] L. Coroianu, D. Costarelli, U. Kadak, Quantitative estimates for neural network operators implied by asymptotic behaviour of the sigmoidal activation functions, Mediterranean J. Math., 19 (5) (2022), Art. No. 211.

[11] L. Coroianu, D. Costarelli, M. Natale, A. Pantis, The approximation capabilities of Durrmeyer-type neural network operators, J. Appl. Math. Comput., (2024) 10.1007/s12190-024-02146-9.

[12] D. Costarelli, Density results by deep neural network operators with integer weights, Math. Model. Numer. Anal., 27 (4) (2022), 547–560.

[13] D. Costarelli, Approximation error for neural network operators by an averaged modulus of smoothness, J. Approx. Theory, 294 (2023).

[14] D. Costarelli, Convergence and high order of approximation by Steklov sampling operators, Banach J. Math. Anal. Appl., 18 (2024) 70.

[15] D. Costarelli, M. Piconi, G. Vinti, Quantitative estimates for Durrmeyer-sampling series in Orlicz spaces, Sampl. Theory Signal Process. Data Anal., 21 (3) (2023).

[16] D. Costarelli, M. Piconi, G. Vinti, The multivariate Durrmeyer-sampling type operators in functional spaces, Dolomites Res. Notes Approx., 15 (2023), 128-144.

[17] D. Costarelli, A.R. Sambucini, G. Vinti, Covergence in Orlicz spaces by means of the multivariate max-product neural network operators of the Kantorovich type and applications, Neural Comput. & Appl., 31 (2019), 5069-5078.

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

[19] D. Costarelli, R. Spigler, How sharp is the Jensen inequality ?, J. Inequal. Appl., 69 (2015), 1-10.

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

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

[22] I. Daubechies, R. DeVore, S. Foucart, B. Hanin, G. Petrova, Nonlinear approximation and (Deep) ReLU networks, Construct. Approx., (2021).

[23] R.A. DeVore, G.G. Lorentz, Constructive approximation, Springer Science & Business Media, Vol. 303, (1993).

[24] S. Goebbles, On sharpness of error bounds for univariate approximation by single hidden layer feedforward neural networks, Results Math., 75 (2020), Art. no. 109.

[25] I. Gühring, M. Raslan, Approximation rates for neural networks with encodable weights in smoothness spaces, Neural Netw., 134 (2021), 107-130.

[26] P.C. Kainen, V. Kurková, A. Vogt, Approximative compactness of linear combinations of characteristic functions, J. Approx. Theory, 257 (2020), Art. no. 105435.

[27] U. Kadak, Multivariate fuzzy neural network interpolation operators and applications to image processing, Expert System with Appl., 206 (2022), Art. no. 117771.

[28] U. Kadak, Fractional type multivariate neural network operators, Math. Appl. Sci., 46 (3) (2023), 3045-3065.

[29] M. Kohler, A. Krzyzak, Over-parametrized deep neural networks do not generalize well, Bernoulli, 27 (4) (2021), 2564-2597.

[30] V. Kurková, M. Sanguineti, Classification by sparse neural networks, IEEE Trans. on Neural Netw. Learning Syst., 30 (9) (2019), 2746–2754.

[31] Y. Li, S. Lu, P. Mathé, S.V. Pereverzev, Two-layer networks with the ReLU$^k$ activation function: Barron spaces and derivative approximation, Numer. Math., 156 (1) (2024), 319-344.

[32] B. Li, S. Tang, H. Yu, Better approximations of high dimensional smooth functions by deep neural networks with rectified power units, Commun. Comput. Phys., (2019).

[33] J.A.A. Opshoor, Ch. Schwab, J. Zech, Exponential ReLU DNN expression of holomorphic maps in high dimension, Construct. Approx., 72 (2021).

[34] Y. Qian, D. Yu, Rates of approximation by neural network interpolation operators, Appl. Math. Comput., 418 (2022), Art. no. 126781.

[35] B. Sendov, V.A. Popov, The averaged moduli of smoothness, Pure Appl. Math., (New York), Wiley, Chichester, (1988).

[36] M. Sharma, U. Singh, Some density results by deep Kantorovich type neural network operators, J. Mathematical Anal. Appl., 533 (2) (2024), Art. no. 128009.

[37] E. M. Stein, Singular integrals and differentiability properties of functions, Princeton Univ. Press, Princeton, N. J., (1970).

[38] A.F. Timan, Theory of approximation of functions of a real variable, MacMillan, New York, (1965).

[39] C. Turkun, O. Duman, Modified neural network operators and their convergence properties with summability methods, Rev. Real Acad. Cienc. Exactas Fis. Nat. - A: Mat., 114 (2020), Art. no. 132.

[40] G. Wang, D. Yu, L. Guan, Neural network interpolation operators of multivariate functions, J. Comput. Applied Math., 431 (2023), Art. no. 115266.

[41] D. Yarotsky, Universal approximations of invariant maps by neural networks, Construct. Approx., (2021).

[42] D.X. Zhou, Universality of deep convolutional neural networks, Appl. Comput. Harmonic Anal., 48 (2) (2020), 787-794.

[43] D.X. Zhou, Theory of deep convolutional neural networks: Downsampling, Neural Netw., 124 (2020), 319-327.

[44] R. Zoppoli, M. Sanguineti, G. Gnecco, T. Parisini, Neural Approximations for Optimal Control and Decision, Communications and Control Engineering book series (CCE), Springer Cham (2020) DOI https://doi.org/10.1007/978-3-030-29693-3.