Asymptotic Expansion for Neural Network Operators of the Kantorovich Type and High Order of Approximation

Marco Cantarini, Danilo Costarelli⬤ and Gianluca Vinti

Abstract. In this paper, we study the rate of pointwise approximation for the neural network operators of the Kantorovich type. This result is obtained proving a certain asymptotic expansion for the above operators and then by establishing a Voronovskaja type formula. A central role in the above resuts is played by the truncated algebraic moments of the density functions generated by suitable sigmoidal functions. Furthermore, to improve the rate of convergence, we consider finite linear combinations of the above neural network type operators, and also in the latter case, we obtain a Voronovskaja type theorem. Finally, concrete examples of sigmoidal activation functions have been deeply discussed, together with the case of rectified linear unit (ReLu) activation function, very used in connection with deep neural networks.

Mathematics Subject Classification. 41A25, 41A05, 41A30, 47A58.

Keywords. Sigmoidal functions, neural networks operators, asymptotic formula, Voronovskaja formula, high-order convergence, truncated algebraic moment.

1. Introduction

The theory of the neural network (NN) operators arise since 1992 with the pioneer work of Cardaliaguet and Euvrard [15], and then in the next years, they have been largely studied by several authors under different aspects, see, e.g., [12,13,32,33]; the main advantage of the above theory lies on its connection with the artificial neural networks and their applications to Approximation Theory. For a complete overview concerning applications of neural networks and learning theory, see, e.g., the very complete monograph of Cucker and Zhou [29], and also [9,37,42,43,46–49,53,56].

A generalization to the L^p-setting of the above NN operators has been proved in [21]; here, the authors showed that any multivariate L^p-data can be approximated in the L^p-norm by means of the NN operators of the

Kantorovich-type. However, even if their natural settings are the L^p-spaces, the above operators converge also pointwise and uniformly when continuous functions are considered.

Such operators are called "of Kantorovich-type", since their coefficients are defined by means of suitable averages of the functions f that is approximated.

Usually, the activation functions of the NN operators are a sigmodal function ([30]); the reason is that a sigmoidal curve allows to simulate the two states of the biological neurons, that are, the activated and the non-activated ones, see, e.g., [9,38,39,44].

Recently, also a new unbounded activation functions has been investigated, that is the so-called ReLu (rectified linear unit) activation function, defined by the positive part of x, for every $x \in \mathbb{R}$. The ReLu activation function has a linear grow as x goes to $+\infty$ and it revealed to be very suitable in order to train deep (i.e., the multi-layers) neural networks, see, e.g., [31,51].

Due to its definition, it is easy to show that the ReLu activation function can be used to express the well-known ramp (or cut) sigmoidal function (see [36]); furthermore, it is also well known that the ramp function can be considered as the activation function in the NN operators. Then, the above relation implies that also the ReLu activation function can be used in NN operators.

The main purpose of this paper is to study the pointwise rate of approximation for the NN operators of the Kantorovich type. The above aim is pursuit by adopting the following strategy: first of all, an asymptotic expansion for the above operators is established, and then, a Voronovskaja type theorem is proved.

The asymptotic formula allows to expand the above approximation operators, when they are evaluated for a sufficiently smooth functions f, in terms of derivatives of the approximated function f evaluated at suitable nodes. A central role in the asymptotic expansion is played by the truncated algebraic moments of the density functions $\phi_\sigma(x)$; the density function is generated by suitable sigmoidal functions and it can be considered as the kernel of the operators.

The asymptotic formula can be used to prove a Voronovskaja-type theorem (see, e.g., [6–8]), from which we establish the rate of pointwise approximation of the operators under investigation.

In general, Voronovskaja formulas are widely studied in Approximation Theory (see, e.g., [4,10,34,54]), both from the qualitative and quantitative point of view (see, e.g., [1,50]), also for their relations with the saturation phenomenon for families of linear operators (see [27,35]).

Furthermore, to improve the above convergence results and to obtain a family of NN operators with an higher order of approximation, here, we also considered suitable families of finite linear combination of the above NN operators of the Kantorovich type. Also for these families of operators, we derive asymptotic and Voronovskaja type formulas, following the same steps above described.

Finally, we discuss in details several examples of sigmoidal functions for which the above results hold. In particular, we consider the cases of the logistic and the hyperbolic tangent sigmoidal functions, which are of crucial importance in the theory of learning artificial neural networks, and the case of the sigmoidal functions generated by the central B-splines, which are useful to construct high-order convergence NN operators.

2. Notations and Preliminary Results

In this paper, we denote by $C(I)$ the space of all functions $f : I \to \mathbb{R}$ which are continuous on $I := [-1, 1]$ and we denote by $C^r(I)$, $r \in \mathbb{N}^+$ the space of functions of $C(I)$ which have continuous derivative $f^{(s)}$ on I, for every $1 \leq s \leq r$, $s \in \mathbb{N}^+$. The above spaces will be considered endowed by the usual max-norm $\|\cdot\|_\infty$. Now, denoting by $\Phi : \mathbb{R} \to \mathbb{R}$ a given function, we can recall the following definitions.

Definition 1. Let $h \in \mathbb{N}$ be fixed. We define the *truncated algebraic moment of Φ order h*, by:

$$m_h^n(\Phi, u) := \sum_{k=-n}^{n-1} \Phi(u - k)(k - u)^h, \quad u \in \mathbb{R}, \tag{1}$$

for every $n \in \mathbb{N}^+$.

Definition 2. Let $h \geq 0$ be fixed. We call *discrete absolute moment of Φ of order h* what follows:

$$M_h(\Phi) := \sup_{u \in \mathbb{R}} \sum_{k \in \mathbb{Z}} |\Phi(u - k)| \, |u - k|^h.$$

The discrete absolute moments of a given function are widely used tools very useful to establish the convergence of families of linear operators, see, e.g., [6,7,55].

It is well known that, if Φ is bounded on $\mathbb{R}$ and has a sufficiently fast decay to 0 as $x \to \pm\infty$, then $M_h(\phi) < +\infty$, for suitable h, see, e.g., [19].

Now, we recall the definition of sigmoidal function (see, e.g., [26,30]).

Definition 3. Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a measurable function. We call σ a sigmoidal function if:

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

In what follows, we consider non-decreasing sigmoidal functions σ which satisfy the following conditions:

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

We now recall the definition of the following density function:

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

The following lemma summarize some useful properties of the function ϕ_σ.

Lemma 4. (i) $\phi_\sigma(x) \geq 0$ for every $x \in \mathbb{R}$, with $\phi_\sigma(2) > 0$, and moreover, $\lim_{x \to \pm\infty} \phi_\sigma(x) = 0$;

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

(iii) $\phi_\sigma(x)$ is non-decreasing for $x < 0$ and non-increasing for $x \geq 0$;

(iv) $\phi_\sigma(x) = O\left(|x|^{-1-\alpha} \right)$ as $x \to \pm\infty$ where α is the positive constant of condition $(\Sigma 3)$;

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

$$m_0^n(\phi_\sigma, nx) = \sum_{k=-n}^{n-1} \phi_\sigma(nx - k) \geq \phi_\sigma(2) > 0.$$

Furthermore, if σ is a sigmoidal function which satisfies $(\Sigma 3)$ for $\alpha > r$, for some $r \in \mathbb{N}^+$, then:

(vi) $M_h(\phi_\sigma) < +\infty$, for every $0 \leq h \leq r$;

(vii) for every fixed $\gamma > 0$, it turns out that:

$$\lim_{n \to +\infty} \sum_{|u-k| > \gamma n} \phi_\sigma(u - k) |u - k|^h = 0,$$

uniformly with respect to $u \in \mathbb{R}$, for every $0 \leq h \leq r$;

(viii) the sequences $(m_h^n(\phi_\sigma, u)_n)_{n \in \mathbb{N}^+}$ of the truncated algebraic moments of order h are equibounded on $\mathbb{R}$, for every $h = 0, 1, \ldots, r$.

Remark 5. Note that, if we remove condition $(\Sigma 2)$ on σ, and we assume directly that ϕ_σ satisfies assertion (iii) of Lemma 4, the theory still holds, see, e.g., [25]. The consequence of the above observation is that, we can apply the above theory to C^2 as well as to *non-smooth* sigmoidal functions, such that the corresponding ϕ_σ satisfies (iii) of Lemma 4, and $\phi_\sigma(2) > 0$.

For a proof of the properties (i) - (v) see [24]; for the properties (vi) - $(viii)$, see Lemma 2.6 of [28].

Now, we recall the definition of the NN operators of the Kantorovich type.

Definition 6. Let $f : I \to \mathbb{R}$ be a locally integrable function. The Kantorovich-type NN operators $K_n(f, \cdot)$, activated by a sigmoidal function σ, and acting on f, are defined by:

$$K_n(f, x) := \frac{\displaystyle\sum_{k=-n}^{n-1} \phi_\sigma(nx - k) \, n \int_{k/n}^{(k+1)/n} f(u) \, du}{\displaystyle\sum_{k=-n}^{n-1} \phi_\sigma(nx - k)}, \ x \in I. \tag{2}$$

Clearly, for every $n \in \mathbb{N}^+$ and $k \in \mathbb{Z}$ such that $-n \le k \le n-1$, it turns out that $-1 \le \frac{k}{n} < \frac{k+1}{n} \le 1$; the latter inequality, together with condition (v) of Lemma 4, implies that the operators are well defined, e.g., in case of locally integrable and bounded functions on I. Furthermore, it turns out that:

$$|K_n(f, x)| \le \|f\|_\infty,$$

for all $x \in I$.

Remark 7. It is well known that (see, e.g., [21]) the family $K_n(f, \cdot)$ converges pointwise at each point of continuity of any bounded f. The convergence turns out to be uniform on I if f is continuous on the whole I.

3. Asymptotic Formulas for the NN Operators of the Kantorovich Type

The main aim of this section is to study the order of pointwise approximation for the operators K_n by means of Voronovskaja type theorem.

From now on, when we refer to a sigmoidal function σ, we always consider a sigmoidal function σ satisfying conditions $(\Sigma 1)$, $(\Sigma 2)$, and $(\Sigma 3)$ introduced in Sect. 2.

We begin establishing the following asymptotic expansion for the operators K_n.

Theorem 8. *Let σ be a sigmoidal function witch satisfies assumption $(\Sigma 3)$ with $\alpha > r$, for some $r \in \mathbb{N}^+$. Moreover, let $f \in C^r(I)$ be fixed. Then, the following asymptotic formula holds:*

$$K_n(f, x) = f(x) + \sum_{\nu=1}^{r} \frac{f^{(\nu)}(x)}{\nu! n^\nu} \sum_{h=0}^{\nu} \binom{\nu}{h} \frac{m_h^n(\phi_\sigma, nx)}{(\nu - h + 1) m_0^n(\phi_\sigma, nx)} + o(n^{-r}),$$

$$n \to \infty,$$

for every $x \in I$, where $m_0^n(\phi_\sigma, nx) > 0$, for every $x \in I$, and $n \in \mathbb{N}^+$.

Proof. From the local Taylor formula (for details, see, e.g., [8, p. 283]), we have:

$$f(u) = \sum_{\nu=0}^{r} \frac{f^{(\nu)}(x)}{\nu!} (u - x)^\nu + h(u - x)(u - x)^r$$

for every $x, u \in I$, where $h(y)$ is a suitable bounded[1] function, such that $h(y) \to 0$ as $y \to 0$. Then, for every fixed $x \in I$, we can write what follows:

$$K_n(f, x) = \frac{\displaystyle\sum_{k=-n}^{n-1} \phi_\sigma(nx - k) \, n \int_{k/n}^{(k+1)/n} f(u) \, du}{\displaystyle\sum_{k=-n}^{n-1} \phi_\sigma(nx - k)}$$

---

[1] Note that the boundedness of h derives from the boundedness of f and its derivatives.

$$= \frac{\sum_{k=-n}^{n-1} \phi_\sigma (nx-k) \, n \int_{k/n}^{(k+1)/n} \left[\sum_{\nu=0}^{r} \frac{f^{(\nu)}(x)}{\nu!} (u-x)^\nu \right] du}{\sum_{k=-n}^{n-1} \phi_\sigma (nx-k)}$$

$$+ \frac{\sum_{k=-n}^{n-1} \phi_\sigma (nx-k) \, n \int_{k/n}^{(k+1)/n} \left[h(u-x)(u-x)^r \right] du}{\sum_{k=-n}^{n-1} \phi_\sigma (nx-k)}$$

$$=: I_1 + I_2.$$

Let us analyze the term I_1. Using the binomial theorem and recalling that $\sum_{k=-n}^{n-1} \phi_\sigma (nx-k) = m_0^n (\phi_\sigma, nx)$, we have that:

$$I_1 = \left[\sum_{k=-n}^{n-1} \phi_\sigma (nx-k) \, n \sum_{\nu=0}^{r} \frac{f^{(\nu)}(x)}{\nu!} \int_{k/n}^{(k+1)/n} (u-x)^\nu \, du \right] \frac{1}{m_0^n (\phi_\sigma, nx)}$$

$$= \left[\sum_{k=-n}^{n-1} \phi_\sigma (nx-k) \, n \sum_{\nu=0}^{r} \frac{f^{(\nu)}(x)}{\nu!} \right.$$
$$\left. \int_{k/n}^{(k+1)/n} (u - k/n + k/n - x)^\nu \, du \right] \frac{1}{m_0^n (\phi_\sigma, nx)}$$

$$= \left[\sum_{k=-n}^{n-1} \phi_\sigma (nx-k) \, n \sum_{\nu=0}^{r} \frac{f^{(\nu)}(x)}{\nu!} \sum_{h=0}^{\nu} \binom{\nu}{h} (k/n - x)^h \right.$$
$$\left. \int_{k/n}^{(k+1)/n} (u - k/n)^{\nu-h} \, du \right] \frac{1}{m_0^n (\phi_\sigma, nx)},$$

so

$$I_1 = \frac{1}{m_0^n (\phi_\sigma, nx)} \sum_{k=-n}^{n-1} \phi_\sigma (nx-k) \sum_{\nu=0}^{r} \frac{f^{(\nu)}(x)}{\nu!} \sum_{h=0}^{\nu} \binom{\nu}{h} \frac{(k/n - x)^h}{(\nu - h + 1) n^{\nu-h}}$$

$$= \frac{1}{m_0^n (\phi_\sigma, nx)} \sum_{\nu=0}^{r} \frac{f^{(\nu)}(x)}{\nu! n^\nu} \sum_{h=0}^{\nu} \binom{\nu}{h} \frac{1}{\nu - h + 1} \sum_{k=-n}^{n-1} (k - nx)^h \phi_\sigma (nx-k)$$

$$= \sum_{\nu=0}^{r} \frac{f^{(\nu)}(x)}{\nu! n^\nu} \sum_{h=0}^{\nu} \binom{\nu}{h} \frac{m_h^n (\phi_\sigma, nx)}{(\nu - h + 1) m_0^n (\phi_\sigma, nx)}.$$

Now, let us analyze the term I_2. Let $\varepsilon > 0$ be fixed. Since $h(y) \to 0$ as $y \to 0$, then there exists a $\gamma > 0$, such that $|h(y)| < \varepsilon$ if $|y| \le \gamma$. Hence, we can write:

$$I_2 = \frac{1}{m_0^n (\phi_\sigma, nx)} \sum_{\substack{k=-n \\ |nx-k| < \gamma n/2}}^{n-1} \phi_\sigma (nx-k) \, n \int_{k/n}^{(k+1)/n} h(u-x)(u-x)^r \, du$$

$$+ \frac{1}{m_0^n(\phi_\sigma, nx)} \sum_{\substack{k=-n \\ |nx-k| \geq \gamma n/2}}^{n-1} \phi_\sigma(nx-k) \, n \int_{k/n}^{(k+1)/n} h(u-x)(u-x)^r \, du$$

$$= I_{2,1} + I_{2,2}.$$

As concerns $I_{2,1}$, we observe that, if $u \in [k/n, (k+1)/n]$, we have:

$$|u - x| \leq |u - k/n| + |x - k/n| \leq 1/n + \gamma/2 < \gamma$$

for every sufficiently large n, and hence:

$$|I_{2,1}| \leq \frac{\varepsilon}{m_0^n(\phi_\sigma, nx)} \sum_{\substack{k=-n \\ |nx-k| < \gamma n/2}}^{n-1} \phi_\sigma(nx-k) \, n \int_{k/n}^{(k+1)/n} |u-x|^r \, du$$

$$\leq \frac{\varepsilon}{m_0^n(\phi_\sigma, nx)} \sum_{\substack{k=-n \\ |nx-k| < \gamma n/2}}^{n-1} \phi_\sigma(nx-k) \, n$$

$$\int_{k/n}^{(k+1)/n} \sum_{m=0}^{r} \binom{r}{m} (u-k/n)^{r-m} |k/n - x|^m \, du$$

$$\leq \frac{\varepsilon}{m_0^n(\phi_\sigma, nx) \, n^r} \sum_{k=-n}^{n-1} \sum_{m=0}^{r} \binom{r}{m} \frac{|k-nx|^m}{r-m+1} \phi_\sigma(nx-k)$$

$$\leq \frac{\varepsilon}{m_0^n(\phi_\sigma, nx) \, n^r} \sum_{m=0}^{r} \binom{r}{m} \frac{M_m(\phi_\sigma)}{r-m+1}.$$

We now consider $I_{2,2}$. By the boundedness of $h(y)$ and arguing as in $I_{2,1}$, we get:

$$|I_{2,2}| \leq \frac{\|h\|_\infty}{m_0^n(\phi_\sigma, nx)} \sum_{\substack{k=-n \\ |nx-k| \geq \gamma n/2}}^{n-1} \phi_\sigma(nx-k) \, n \int_{k/n}^{(k+1)/n} |u-x|^r \, du$$

$$\leq \frac{\|h\|_\infty}{m_0^n(\phi_\sigma, nx) \, n^r} \sum_{m=0}^{r} \binom{r}{m} \frac{1}{r-m+1} \sum_{\substack{k=-n \\ |nx-k| \geq \gamma n/2}}^{n-1} \phi_\sigma(nx-k) |k-nx|^m$$

$$\leq \frac{\|h\|_\infty}{m_0^n(\phi_\sigma, nx) \, n^r} \frac{2^{r+1}-1}{r+1} \max_{m=0,\dots,r} \sum_{|nx-k| \geq \gamma n/2} \phi_\sigma(nx-k) |k-nx|^m$$

$$\leq \frac{\|h\|_\infty}{m_0^n(\phi_\sigma, nx) \, n^r} \frac{2^{r+1}-1}{r+1} \varepsilon$$

for n sufficiently large, by the property (vii) of Lemma 4, and the trivial identity:

$$\sum_{m=0}^{r} \binom{r}{m} \frac{x^{r-m+1}}{r-m+1} = \frac{(1+x)^{r+1} - 1}{r+1}.$$

This concludes the proof. $\qquad\square$

As a consequence of the previous result, we can prove the following Voronovskaja type theorem.

Theorem 9. *Let σ be a sigmoidal function which satisfies assumption* $(\Sigma 3)$ *with $\alpha > 1$. Suppose that, for every $x \in \mathbb{R}$:*

$$m_h^n(\phi_\sigma, x) = m_h + o\left(n^{-\theta_h}\right), \; n \to +\infty, \; h = 0, 1, \tag{3}$$

where $m_h \in \mathbb{R}$, $\theta_h > 0$, $h = 0, 1$, and the "o-term" tends to zero uniformly with respect to $x \in \mathbb{R}$. Then, for any $f \in C^1(I)$, we have:

$$\lim_{n \to +\infty} n\left[K_n(f, x) - f(x)\right] = f'(x)\left(\frac{1}{2} + \frac{m_1}{m_0}\right),$$

where $m_0 > 0$ in view of Lemma 4 (v).

Proof. From the asymptotic formula of Theorem 8, we have:

$$n\left[K_n(f, x) - f(x)\right] = n\left[\frac{f'(x)}{n}\left(\frac{1}{2} + \frac{m_1 + o\left(n^{-\theta_1}\right)}{m_0 + o\left(n^{-\theta_0}\right)}\right) + o\left(n^{-1}\right)\right],$$

and the result follows taking the limit as $n \to +\infty$. $\square$

Remark 10. The previous theorem shows that the order of pointwise approximation by means of the Kantorovich NN operators is at least of order $O\left(n^{-1}\right)$ when $n \to +\infty$ and $f \in C^1(I)$.

Moreover, we can observe that, under the assumption of Theorem 9, it is not hard to show that, if $f \in C^1(I)$, we have:

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

as $n \to +\infty$.

Thus, it seems clear that the best possible order of uniform approximation that can be achieved by the operators K_n is $O(1/n)$, as $n \to +\infty$. This shows that, in general, a better order of approximation cannot be obtained; indeed, in the main examples of sigmoidal functions (see Section 5 below) $m_1 = 0$, then the limit of Theorem 9 gives the exact order of approximation. The order is also less rapid than that achieved in case of the classical NN operators, (see [28]), that is $O(1/n^2)$, as $n \to +\infty$.

4. High-Order Convergence of NN Operators of the Kantorovich Type

To construct NN operators of the Kantorovich type with higher order of approximation than K_n, we adopt the following strategy.

Let $r \in \mathbb{N}^+$, $r > 1$ and $\alpha_j \in \mathbb{R} \setminus \{0\}$, $j = 1, \ldots, r$ fixed, such that:

$$\sum_{j=1}^r \alpha_j = 1. \tag{4}$$

Let us define:

$$\mathcal{K}_n^r(f, x) := \sum_{j=1}^r \alpha_j K_{nj}(f, x), \; x \in I,$$

where f is any bounded locally integrable function. It is clear that condition (4) is such that $\mathcal{K}_n^r$ preserves the approximation properties of K_n.

Obviously, also in this section, when we refer to σ, we always consider sigmoidal functions satisfying (σi), $i = 1, 2, 3$.

Now, we can prove the following asymptotic and Voronovskaja type formula for $\mathcal{K}_n^r$.

Theorem 11. *Let σ be a sigmoidal function which satisfies assumption $(\Sigma 3)$ with $\alpha > r > 1$, for some $r \in \mathbb{N}^+$. Moreover, let $f \in C^r(I)$ be fixed. Then:*

$$\mathcal{K}_n^r(f, x) = f(x) + \sum_{\nu=1}^r \frac{f^{(\nu)}(x)}{\nu! n^\nu} \sum_{h=0}^\nu \binom{\nu}{h} \frac{1}{\nu - h + 1} \sum_{j=1}^r \frac{\alpha_j}{j^\nu} \frac{m_h^{nj}(\phi_\sigma, njx)}{m_0^{nj}(\phi_\sigma, njx)}$$
$$+ o(n^{-r}), \tag{5}$$

as $n \to +\infty$ and for every $x \in I$. Furthermore, if we suppose that:

$$m_h^n(\phi_\sigma, x) = m_h, \ h = 0, \dots, r \tag{6}$$

for $n \in \mathbb{N}^+$ sufficiently large, where $m_h \in \mathbb{R}$, such that $\alpha_j, \ j = 1, \dots, r$ satisfies the following linear algebraic system:

$$\sum_{j=1}^r \alpha_j = 1, \qquad\qquad \sum_{j=1}^r \frac{\alpha_j}{j^\nu} = 0, \qquad \nu = 1, \dots, r-1 \tag{7}$$

then

$$\lim_{n \to +\infty} n^r \left[\mathcal{K}_n^r(f, x) - f(x) \right] = \frac{f^{(r)}(x)}{r!} \sum_{h=0}^r \binom{r}{h} \frac{1}{r - h + 1} \frac{m_h}{m_0} \sum_{j=1}^r \frac{\alpha_j}{j^r}. \tag{8}$$

Proof. We simply observe that, proceeding as in Theorem 8, we can write:

$$\mathcal{K}_n^r(f, x) = \sum_{j=1}^r \alpha_j K_{nj}(f, x)$$

$$= \sum_{j=1}^r \alpha_j \left(\frac{\sum_{k=-nj}^{nj-1} \phi_\sigma(njx - k) \, nj \int_{k/(nj)}^{(k+1)/(nj)} f(u) \, du}{\sum_{k=-nj}^{nj-1} \phi_\sigma(njx - k)} \right)$$

$$= \sum_{j=1}^r \alpha_j \left(f(x) + \sum_{\nu=1}^r \frac{f^{(\nu)}(x)}{\nu! n^\nu j^\nu} \sum_{h=0}^\nu \binom{\nu}{h} \frac{1}{\nu - h + 1} \frac{m_h^{nj}(\phi_\sigma, njx)}{m_0^{nj}(\phi_\sigma, njx)} \right)$$

$$+ \sum_{j=1}^r \alpha_j \left(\frac{\sum_{k=-nj}^{nj-1} \phi_\sigma(njx - k) \, nj \int_{k/(nj)}^{(k+1)/(nj)} \left[h(u - x)(u - x)^r \right] du}{m_0^{nj}(\phi_\sigma, njx)} \right);$$

hence, the asymptotic formula (5) follows arguing as for $I_{2,1}$ and $I_{2,2}$, in the proof of Theorem 8. Now, to prove the second part of the theorem, it is

sufficient to observe that, under the additional assumptions on the truncated algebraic moments, we have:

$$\mathcal{K}_n^r(f,x) - f(x) = \sum_{\nu=1}^{r} \frac{f^{(\nu)}(x)}{\nu! n^\nu} \sum_{h=0}^{\nu} \binom{\nu}{h} \frac{1}{\nu - h + 1} \frac{m_h}{m_0} \sum_{j=1}^{r} \frac{\alpha_j}{j^\nu} + o\left(n^{-r}\right)$$

$$= \frac{f^{(r)}(x)}{r! n^r} \sum_{h=0}^{r} \binom{r}{h} \frac{1}{r - h + 1} \frac{m_h}{m_0} \sum_{j=1}^{r} \frac{\alpha_j}{j^r} + o\left(n^{-r}\right)$$

for $n \in \mathbb{N}^+$ sufficiently large, $x \in I$. Then:

$$n^r \left[\mathcal{K}_n^r(f,x) - f(x)\right] = \frac{f^{(r)}(x)}{r!} \sum_{h=0}^{r} \binom{r}{h} \frac{1}{r - h + 1} \frac{m_h}{m_0} \sum_{j=1}^{r} \frac{\alpha_j}{j^r} + n^r o\left(n^{-r}\right)$$

$$(9)$$

and passing to the limit as $n \to +\infty$, the proof of the Voronovskaja formula (8) follows. $\qquad \square$

    The previous theorem shows that the order of pointwise approximation by the NN operators $\mathcal{K}_n^r(f,x)$ is at least $O\left(n^{-r}\right)$ as $n \to +\infty$ when functions that belong to $C^r(I)$, $r \in \mathbb{N}$ are approximated. Again, it is quite simple to observe that, under the assumption of Theorem 11 and for $f \in C^r(I)$, we have:

$$\|\mathcal{K}_n^r(f,\cdot) - f(\cdot)\|_\infty = O\left(n^{-r}\right)$$

as $n \to +\infty$.

5. Applications to Special Cases

In this section, we discuss the results proved above for some concrete cases of sigmoidal functions.

5.1. Applications with the Logistic Function

First of all, we consider the case of the Kantorovich NN type operators activated by the well-known logistic function (see, e.g., [20,40,41] and Fig. 1):

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

NN operators activated by logistic functions have been widely studied, e.g., in [12,32].

    Obviously, σ_ℓ is a smooth function and it satisfies all the assumptions (Σi), $i = 1, 2, 3$. Furthermore, by its exponential decay to zero as $x \to -\infty$, condition $(\Sigma 3)$ is satisfied for every $\alpha > 0$. Hence, it turns out that $M_h(\phi_{\sigma_\ell}) < +\infty$ for every $h \geq 0$ in view of Lemma 4, (vi). The above function is very useful in the theory of artificial neural network, since it is used as activation function in neuronal models when training algorithms would applied, see, e.g., [14,16,45,48,52].

    Now, to apply the results proved in the previous sections, the truncated algebraic moments of the function $\phi_{\sigma_\ell}(x)$ must be computed. In general, it is possible to compute the truncated algebraic moments of a given function
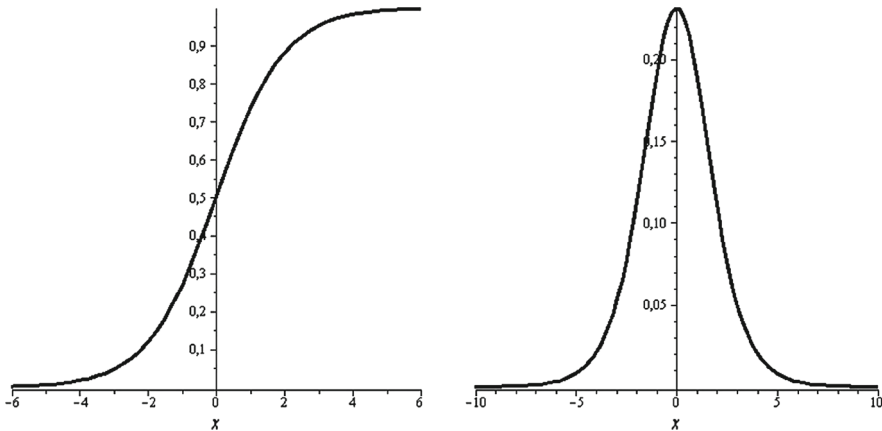
Figure 1. The logistic function $\sigma_\ell(x)$ (left) and the corresponding density function $\phi_{\sigma_\ell}(x)$ (right)

exploiting a well-known result of Fourier analysis, i.e., the so-called Poisson summation formula (see, e.g., [11]). In particular, since $(\Sigma 3)$ holds for every $\alpha > 0$, by Lemma 4 (iv), we have that $\phi_{\sigma_\ell} \in L^1(\mathbb{R})$, and then, the usual L^1-Fourier transform can be used to apply the above mentioned Poisson summation formula. In particular, following the procedure given in [28], it turns out that ϕ_{σ_ℓ} satisfies assumption (3) with $m_0 = 1$, $m_1 = 0$ and $\theta_0, \theta_1 > 1$. Then, from Theorem 9, we can obtain what follows.

Corollary 12. *Let $\sigma_\ell(x)$ be the logistic function, and $f \in C^1(I)$ be fixed. Then, for every $x \in I$, we have:*

$$\lim_{n \to +\infty} n\left[K_n^{\sigma_\ell}(f, x) - f(x)\right] = \frac{f'(x)}{2}.$$

Moreover, $\phi_{\sigma_\ell}(x)$ does not satisfy assumption (6) of Theorem 11, which, hence, cannot be used to construct the operators $\mathcal{K}_n^r$.

5.2. Applications with Hyperbolic Tangent Sigmoidal Function

In this section, we study the case of the Kantorovich NN type operators activated by the well-known hyperbolic tangent sigmoidal function (see, e.g., [25]):

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

The graph of $\sigma_h(x)$ together with its density function $\phi_h(x)$ is plotted in Fig. 2.

The NN operators activated by the hyperbolic tangent sigmoidal functions have been widely studied, see, e.g., [22,23,33]. Again, we can observe that σ_h is a smooth function that satisfies (Σi), $i = 1, 2, 3$, it tends to zero exponentially as $x \to -\infty$, and so $(\Sigma 3)$ holds for all $\alpha > 0$. Now, computing again the truncated algebraic moments of ϕ_{σ_h} using the Poisson summation
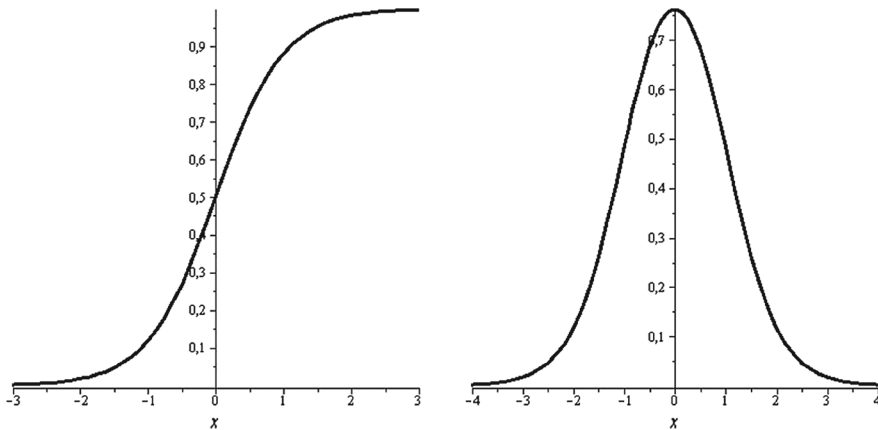
Figure 2. The hyperbolic tangent sigmoidal function$\sigma_h(x)$
(left) and the corresponding density function $\phi_{\sigma_h}(x)$ (right)

formula (as in the case of logistic function), it turns out that $m_0 = 1$, $m_1 = 0$, and $\theta_0, \theta_1 > 1$ (see [28] again). Hence, we can obtain the following corollary.

Corollary 13. *Let $\sigma_h(x)$ be the hyperbolic tangent sigmoidal function, and $f \in C^1(I)$ be fixed. Then, for every $x \in I$, we have:*

$$\lim_{n \to +\infty} n\left[K_n^{\sigma_h}(f,x) - f(x)\right] = \frac{f'(x)}{2}.$$

Also in this case, assumption (6) of Theorem 11 is not satisfied.

To provide examples of sigmoidal functions useful to construct high-order NN type operators, we can consider what follows.

5.3. Applications with Sigmoidal Functions Generated by B-splines

In [22], the well-known central B-splines of order $d \geq 1$, defined by (see, e.g., [5,17,19,27,31]):

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

have been used to introduce a class of sigmoidal functions. Here, $(x)_+ := \max\{x, 0\}$ denotes the "positive part" of $x \in \mathbb{R}$. The sigmoidal function $\sigma_{M_d}(x)$ associated with the central B-spline M_d, is defined by the following formula:

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

Consequently, the corresponding density function has the following expression:

$$\phi_{\sigma_{M_d}}(x) := \frac{1}{2} \int_{x-1}^{x+1} M_d(t)\, dt, \ x \in \mathbb{R}. \tag{10}$$

By simple computations it is easy to prove that, also for σ_{M_d}, assumptions (Σi), $i = 1, 2, 3$ are satisfied. In particular, since the central B-splines have compact support, then $(\Sigma 3)$ is satisfied by $\sigma_{M_d}(x)$ for every $\alpha > 0$. Exploiting the relation (10), it is easy to obtain that:

$$m_h^n\left(\phi_{\sigma_{M_d}}, x\right) = \frac{1}{2}\sum_{j=0}^{h}\binom{h}{j}\frac{\gamma_j^d}{h-j+1}\left[1 - (-1)^{h-j+1}\right],$$

where

$$\gamma_j^d := m_j^n\left(M_d, x\right) = i^{-j}\frac{d^j}{dx^j}\left(\frac{\sin\left(\frac{x}{2}\right)^d}{\left(\frac{x}{2}\right)^d}\right)(0), \ j \in \mathbb{N}$$

with i the imaginary unit, then, for n sufficiently large, we obtain that $m_0^n\left(\phi_{\sigma_{M_d}}, x\right) = 1$ and $m_1^n\left(\phi_{\sigma_{M_d}}, x\right) = 0$. Hence:

Corollary 14. *Let $\sigma_{M_d}(x)$, $d \in \mathbb{N}^+$, be the sigmoidal function generated by the central B-spline $M_d(x)$, and $f \in C^1(I)$ be fixed. Then, for every $x \in I$, we have:*

$$\lim_{n\to+\infty} n\left[K_n^{\sigma_{M_d}}(f, x) - f(x)\right] = \frac{f'(x)}{2}.$$

Remark 15. As showed in Remark 5, the previous results still hold also in case of the well-known (not differentiable) ramp function:

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

since the corresponding density function satisfies the condition (iii) of Lemma 4.

Note that (see [18, 26]) the sigmoidal function $\sigma_{M_1}(3\cdot)$ coincides with the ramp function σ_R. Then, if we recall the definition of the well-known ReLu activation function (see, e.g., [36]):

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

it turns out that:

$$\sigma_{M_1}(3x) := \psi_{ReLu}(3x + 1/2) - \psi_{ReLu}(3x - 1/2),$$

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

$$\phi_{\sigma_R}(x) = \phi_{\sigma_{M_1}(3\cdot)}(x) = \psi_{ReLu}(3x + 1) - 2\psi_{ReLu}(3x) + \psi_{ReLu}(3x - 1).$$

As a consequence of the above relation, the NN operators $K_n^{\sigma_{M_1}(3\cdot)}$ can be considered as an NN activated by the above linear combination of ReLu activation function. For more details concerning the usefulness of ψ_{ReLu}, see, e.g., [2, 3].

As showed above, the truncated algebraic moments of $\phi_{\sigma_{M_d}}$ are exactly equal to suitable constants. Then, the sigmoidal functions σ_{M_d} can be used to generate the high-order convergence operators $\mathcal{K}_n^r$. Here, we consider the linear combination of $\alpha_j^{(1)}$, $j = 1, 2$, and of $\alpha_j^{(2)}$, $j = 1, 2, 3$, which satisfy respectively:

$$\sum_{j=1}^{2} \alpha_j^{(1)} = 1, \qquad \sum_{j=1}^{2} \frac{\alpha_j^{(1)}}{j} = 0$$

and

$$\sum_{j=1}^{3} \alpha_j^{(2)} = 1, \qquad \sum_{j=1}^{3} \frac{\alpha_j^{(2)}}{j} = 0, \qquad \sum_{j=1}^{3} \frac{\alpha_j^{(2)}}{j^2} = 0.$$

In this way, we obtain the following operators:

$$\mathcal{K}_n^{2,d}(f, x) := -K_n^{\sigma_{M_d}}(f, x) + 2K_{2n}^{\sigma_{M_d}}(f, x)$$
$$\mathcal{K}_n^{3,d}(f, x) := \frac{1}{2} K_n^{\sigma_{M_d}}(f, x) - 4K_{2n}^{\sigma_{M_d}}(f, x) + \frac{9}{2} K_{3n}^{\sigma_{M_d}}(f, x).$$

Thus, we obtain:

Corollary 16. *Let $\sigma_{M_d}(x)$, $d \in \mathbb{N}^+$, be the sigmoidal function generated by the central B-spline $M_d(x)$, and let $x \in I$ be fixed.*

For any $f \in C^2(I)$, there holds:

$$\lim_{n \to +\infty} n^2 \left[\mathcal{K}_n^{2,d}(f, x) - f(x) \right] = -\frac{f^{(2)}(x)}{4} \left(\frac{2}{3} + \gamma_2^d \right),$$

while for any $f \in C^3(I)$, we have:

$$\lim_{n \to +\infty} n^3 \left[\mathcal{K}_n^{3,d}(f, x) - f(x) \right] = \frac{3f^{(3)}(x)}{72} \left(\frac{1}{2} + \gamma_2^d \right).$$

Clearly, an analogous of Corollary 16 can be reformulated also for a finite linear combination of Kantorovich NN type operators K_n with $r > 3$ to achieve a faster convergence.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Adell, J.A., Cardenas-Morales, D.: Quantitative generalized Voronovskaja's formulae for Bernstein polynomials. J. Approx. Theory **231**, 41–52 (2018)

[2] Agarap, A.F.: Deep Learning using Rectified Linear Units (ReLU). arXiv:1803.08375 (2018)

[3] Agostinelli, F., Hoffman, M., Sadowski, P., Baldi, P.: Learning Activation Functions to Improve Deep Neural Networks. arXiv:1412.6830v3 (2015)

[4] Aral, A., Acar, T., Rasa, I.: The new forms of Voronovskaya's theorem in weighted spaces. Positivity **20**(1), 25–40 (2016)

[5] Asdrubali, F., Baldinelli, G., Bianchi, F., Costarelli, D., Rotili, A., Seracini, M., Vinti, G.: Detection of thermal bridges from thermographic images by means of image processing approximation algorithms. Appl. Math. Comput. **317**, 160–171 (2018)

[6] Bardaro, C., Mantellini, I.: Voronovskaja formulae for Kantorovich generalized sampling series. Int. J. Pure Appl. Math. **62**(3), 247–262 (2010)

[7] Bardaro, C., Mantellini, I.: Approximation properties for linear combinations of moment type operators. Comput. Math. Appl. **62**(5), 2304–2313 (2011)

[8] Bardaro, C., Mantellini, I.: Asymptotic formulae for linear combinations of generalized sampling operators. Z. Anal. Ihre Anwend. **32**(3), 279–298 (2013)

[9] Barron, A.R., Klusowski, J.M.: Uniform approximation by neural networks activated by first and second order ridge splines. arXiv preprint arXiv:1607.07819 (2016)

[10] Boccuto, A., Bukhvalov, A.V., Sambucini, A.R.: Some inequalities in classical spaces with mixed norms. Positivity **6**, 393–411 (2002)

[11] Butzer, P.L., Nessel, R.J.: Fourier Analysis and Approximation I. Academic Press, New York (1971)

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

[13] Cao, F., Chen, Z.: Scattered data approximation by neural networks operators. Neurocomputing **190**, 237–242 (2016)

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

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

[16] Chui, C. K., Mhaskar, H. N.: Deep nets for local manifold learning. arXiv preprint arXiv:1607.07110 (2016)

[17] Coroianu, L., Gal, S.G.: Approximation by truncated max-product operators of Kantorovich-type based on generalized (φ, ψ)-kernels. Math. Methods Appl. Sci. **41**(17), 7971–7984 (2018)

[18] Costarelli, D.: Approximate solutions of Volterra integral equations by an interpolation method based on ramp functions. Comput. Appl. Math. **38**(4) (article 159) (2019)

[19] Costarelli, D., Minotti, A.M., Vinti, G.: Approximation of discontinuous signals by sampling Kantorovich series. J. Math. Anal. Appl. **450**(2), 1083–1103 (2017)

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

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

[22] Costarelli, D., Spigler, R.: Approximation by series of sigmoidal functions with applications to neural networks. Ann. Mat. Pura Appl. **194**(1), 289–306 (2015)

[23] Costarelli, D., Vinti, G.: Pointwise and uniform approximation by multivariate neural network operators of the max-product type. Neural Netw. **81**, 81–90 (2016)

[24] Costarelli, D., Vinti, G.: Convergence for a family of neural network operators in Orlicz spaces. Math. Nachr. **290**(2–3), 226–235 (2017)

[25] Costarelli, D., Vinti, G.: Convergence results for a family of Kantorovich max-product neural network operators in a multivariate setting. Math. Slov. **67**(6), 1469–1480 (2017)

[26] Costarelli, D., Vinti, G.: Saturation classes for max-product neural network operators activated by sigmoidal functions. Results Math. **72**(3), 1555–1569 (2017)

[27] Costarelli, D., Vinti, G.: Inverse results of approximation and the saturation order for the sampling Kantorovich series. J. Approx. Theory **242**, 64–82 (2019)

[28] Costarelli, D., Vinti, G.: Voronovskaja formulas for high order convergence neural network operator with sigmoidal functions. Mediterr. J. Math. **17** (article numb. 77), https://doi.org/10.1007/s00009-020-01513-7 (2020)

[29] Cucker, F., Zhou, D.X.: Learning Theory an Approximation Theory Viewpoint. Cambridge University Press, Cambridge (2007)

[30] Cybenko, G.: Approximation by superpositions of a sigmoidal function. Math. Control Signals Syst. **2**, 303–314 (1989)

[31] Eckle, K., Schmidt-Hieber, J.: A comparison of deep networks with ReLU activation function and linear spline-type methods. Neural Netw. **110**, 232–242 (2019)

[32] Fard, S.P., Zainuddin, Z.: The universal approximation capabilities of cylindrical approximate identity neural networks. Arab. J. Sci. Eng. 1–8 (2016)

[33] Fard, S.P., Zainuddin, Z.: Theoretical analyses of the universal approximation capability of a class of higher order neural networks based on approximate identity. In: Nature-Inspired Computing: Concepts, Methodologies, Tools, and Applications (in print). https://doi.org/10.4018/978-1-5225-0788-8.ch055 (2016)

[34] Finta, Z.: On generalized Voronovskaja theorem for Bernstein polynomials. Carpathian J. Math. **28**(2), 231–238 (2012)

[35] Gavrea, I., Ivan, M.: The Bernstein Voronovskaja-type theorem for positive linear approximation operators. J. Approx. Theory **192**, 291–296 (2015)

[36] Goebbels, S.: On sharpness of error bounds for single hidden layer feedforward neural networks. arXiv:1811.05199 (2018)

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

[38] Guliyev, N.J., Ismailov, V.E.: On the approximation by single hidden layer feedforward neural networks with fixed weights. Neural Netw. **98**, 296–304 (2018)

[39] Guliyev, N.J., Ismailov, V.E.: Approximation capability of two hidden layer feedforward neural networks with fixed weights. Neurocomputing **316**, 262–269 (2018)

[40] Iliev, A., Kyurkchiev, N.: On the Hausdor distance between the Heaviside function and some transmuted activation functions. Math. Model. Appl. **2**(1), 1–5 (2016)

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

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

[43] Kainen, P.C., Kurkovà, V.: An integral upper bound for neural network approximation. Neural Comput. **21**, 2970–2989 (2009)

[44] Kurkovà, V.: Lower bounds on complexity of shallow perceptron networks. Eng. Appl. Neural Netw. Commun. Comput. Inform. Sci. **629**, 283–294 (2016)

[45] Lin, S., Zeng, J., Zhang, X.: Constructive neural network learning. arXiv preprint arXiv:1605.00079 (2016)

[46] Maiorov, V., Meir, R.: On the near optimality of the stochastic approximation of smooth functions by neural networks. Adv. Comput. Math. **13**(1), 79–103 (2000)

[47] Makovoz, Y.: Uniform approximation by neural networks. J. Approx. Theory **95**(2), 215–228 (1998)

[48] Mhaskar, H., Poggio, T.: Deep vs. shallow networks: an approximation theory perspective. Anal. Appl. 14 (6), 829–848 (2016)

[49] Moritani, Y., Ogihara, N.: A hypothetical neural network model for generation of human precision grip. Neural Netw. **110**, 213–224 (2019)

[50] Nasaireh, F., Rasa, I.: Another look at Voronovskaja type formulas. J. Math. Inequal. **12**(1), 95–105 (2018)

[51] Petersen, P., Voigtlaender, F.: Optimal approximation of piecewise smooth functions using deep ReLU neural networks. Neural Netw. **108**, 296–330 (2018)

[52] Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)

[53] Smale, S., Zhou, D.X.: Learning theory estimates via integral operators and their approximations. Construct. Approx. **26**(2), 153–172 (2007)

[54] Ulusoy, G., Acar, T.: q-Voronovskaya type theorems for q-Baskakov operators. Math. Methods Appl. Sci. (2015). https://doi.org/10.1002/mma.3784

[55] Vinti, G., Zampogni, L.: A unifying approach to convergence of linear sampling type operators in Orlicz spaces. Adv. Differ. Equ. **16**(5–6), 573–600 (2011)

[56] Zhang, Y., Wu, J., Cai, Z., Du, B., Yu, P.S.: An unsupervised parameter learning model for RVFL neural network. Neural Netw. **112**, 85–97 (2019)

Marco Cantarini
Marche Polytechinc University
Via delle Brecce Bianche, 12
60131 Ancona
Italy
e-mail: `m.cantarini@univpm.it`

Danilo Costarelli and Gianluca Vinti
Department of Mathematics and Computer Science
University of Perugia
Via Vanvitelli, 1
06123 Perugia
Italy
e-mail: `danilo.costarelli@unipg.it`

Gianluca Vinti
e-mail: `gianluca.vinti@unipg.it`