Annali di Matematica (2015) 194:289–306
DOI 10.1007/s10231-013-0378-y

Approximation by series of sigmoidal functions with applications to neural networks

Danilo Costarelli · **Renato Spigler**

Received: 6 March 2013 / Accepted: 28 August 2013 / Published online: 11 September 2013
© Fondazione Annali di Matematica Pura ed Applicata and Springer-Verlag Berlin Heidelberg 2013

Abstract In this paper, we develop a constructive theory for approximating absolutely continuous functions by series of certain sigmoidal functions. Estimates for the approximation error are also derived. The relation with neural networks approximation is discussed. The connection between sigmoidal functions and the scaling functions of r-regular multiresolution approximations are investigated. In this setting, we show that the approximation error for C^1-functions decreases as 2^{-j}, as $j \to +\infty$. Examples with sigmoidal functions of several kinds, such as logistic, hyperbolic tangent, and Gompertz functions, are given.

Keywords Sigmoidal functions · Neural networks approximation · Order of approximation · Truncation error · Multiresolution approximation · Wavelet-scaling functions

Mathematics Subject Classification (2010) 41A25 · 41A30 · 42C40 · 47A58

1 Introduction

In this paper, we develop a new theory, for approximating uniformly functions in some class by series of sigmoidal functions, i.e., functions $\sigma : \mathbb{R} \to \mathbb{R}$ such that $\lim_{x\to-\infty} \sigma(x) = 0$ and $\lim_{x\to+\infty} \sigma(x) = 1$. The idea is to start from appropriate real-valued functions, ϕ, normalized so that $\int_{\mathbb{R}} \phi(t)\,dt = 1$, and to construct sigmoidal functions having the integral form $\sigma_\phi(x) := \int_{-\infty}^{x} \phi(t)\,dt,\ x \in \mathbb{R}$. In this way, we can define the operators

$$(S_w^{\sigma_\phi} f)(x) := \sum_{k\in\mathbb{Z}} \left[\int_a^b \phi(wy-k) f'(y)\,dy \right] \sigma_\phi(wx-k) + f(a), \tag{I}$$

D. Costarelli · R. Spigler (✉)
Dipartimento di Matematica e Fisica, Sezione di Matematica, Università "Roma Tre",
1, Largo S. Leonardo Murialdo, 00146 Rome, Italy
e-mail: spigler@mat.uniroma3.it

D. Costarelli
e-mail: costarel@mat.uniroma3.it

$x \in [a, b]$, where f is an absolutely continuous function on $[a, b] \subset \mathbb{R}$, and $w > 0$ (note that (I) becomes trivial for constants f).

We can show that, the family $(S_w^{\sigma_\phi} f)_{w>0}$ converges to f uniformly on $[a, b]$. Moreover, we derive estimates for the *approximation error* and the *truncation error* of the series.

A remarkable result is obtained when ϕ is the real-valued wavelet-scaling function associated with an r-regular multiresolution approximation of $L^2(\mathbb{R})$, constructed by a suitable procedure, see [11,17,29,30]. In this setting, we replace the weights w with 2^j, $j \in \mathbb{N}^+$, as it seems more natural in view of the relation that ϕ has with the multiresolution approximation. Also in this case, we can show that the family of the operators $(S_j^{\sigma_\phi} f)_{j\in\mathbb{N}^+}$, converges to f as $j \to +\infty$, uniformly on $[a, b]$. Approximating C^1-functions, we obtain an approximation error decreasing to zero as 2^{-j} when $j \to +\infty$.

The approximation procedures based on sigmoidal functions find applications, for instance, in the theory of neural networks (NNs). NNs arise as a practical technique, successfully adopted to model a number of real-world problems, are often used in Approximation Theory as "universal approximators" and have the form

$$\sum_{k=1}^{N} \alpha_k \, \sigma(x \cdot w_k - \theta_k), \quad x,\ w_k \in \mathbb{R}^n, \quad \alpha_k,\ \theta_k \in \mathbb{R}, \tag{II}$$

where $x \cdot w_k := \sum_{i=1}^{n} x_i w_{k_i}$ denotes the inner product in $\mathbb{R}^n$, the w_k's are the weights, the θ_k's are threshold values, and σ is a sigmoidal activation function.

A theory for approximating functions by NNs, defined by (II), was developed by Cybenko in [16], and its feasibility was established by nonconstructive arguments. Often, σ is either the well-known logistic function, or the sigmoidal function generated by the hyperbolic tangent, see [1,2,8]. The theory of NNs is mainly multivariate in nature, but useful constructive approximation results have been obtained also for univariate functions, see, e.g., [1,2,9, 14,19,22,33]. Basic results on NNs were established by Li, Lenze, Mhaskar, Micchelli and Pinkus in [23,26,27,31,32,34]. For results concerning the order of approximation, see [3,10, 13,15,20,24,25]. One-dimensional NNs also play a role in numerical analysis. For instance, they have been used to solve ordinary differential equations [28], or to solve Fredholm or Volterra integral equations of the second kind [7,12]. In this context, available constructive approximation algorithms are fundamental.

The theory for approximating certain functions by series of sigmoidal functions proposed in this paper can be exploited to obtain some kind of NNs approximation. Such an approach is completely new and allows us to obtain a *constructive* approximation algorithm based on a new class of sigmoidal functions.

Such a theory, in the present form, however, does not cover the important cases of NNs activated by either logistic, hyperbolic tangent or Gompertz sigmoidal functions. Therefore, in Sect. 5, we propose an extension of the theory previously developed, which includes such cases, also providing estimates for the approximation errors for functions belonging to the Lipschitz class.

2 Approximation by series of sigmoidal functions

In what follows, we denote by $C[a, b]$ and $AC[a, b]$ the sets of all continuous and absolutely continuous functions, $f : [a, b] \to \mathbb{R}$, on the bounded closed nonempty interval $[a, b]$, respectively; $\| \cdot \|_\infty$ is the usual sup norm $\|f\|_\infty := \max_{x\in[a,b]} |f(x)|$. Moreover,

$\widehat{C}^n[a, b]$, $n \in \mathbb{N}^+$, will denote the set of all functions $f \in C^n(a', b')$, for some open real interval (a', b'), such that $[a, b] \subset (a', b')$.

Let us introduce the class of functions we will work with.

Definition 2.1 The function $\phi : \mathbb{R} \to \mathbb{R}_0^+$ is said to belong to the class Φ, if it satisfies the following conditions:

($\varphi 1$) ϕ is continuous on $\mathbb{R}$ and there exists $C > 0$ such that

$$\phi(x) \leq C(1 + |x|)^{-\alpha},$$

for every $x \in \mathbb{R}$, and for some $\alpha \geq 2$;

($\varphi 2$) $\sum_{k\in\mathbb{Z}} \phi(x - k) = 1$, for every $x \in \mathbb{R}$.

Remark 2.2 The condition ($\varphi 2$) is equivalent to

$$\widehat{\phi}(k) := \begin{cases} 0, & k \in \mathbb{Z} \setminus \{0\}, \\ 1, & k = 0, \end{cases}$$

where $\widehat{\phi}(v) := \int_{\mathbb{R}} \phi(t)\, e^{-ivt}\, dt$, $v \in \mathbb{R}$, is the Fourier transform of ϕ; see [6]. In particular, it turns out that $\widehat{\phi}(0) = \int_{\mathbb{R}} \phi(t)\, dt = 1$.

For any fixed $\phi \in \Phi$, the function $K_\phi : \mathbb{R}^2 \to \mathbb{R}_0^+$, defined by

$$K_\phi(x, y) := \sum_{k\in\mathbb{Z}} \phi(x - k)\, \phi(y - k), \quad (x, y) \in \mathbb{R}^2, \tag{1}$$

will be called the *kernel* associated to ϕ. Clearly, it follows from condition ($\varphi 2$) and by Remark 2.2 that

$$\int_{\mathbb{R}} K_\phi(x, y)\, dy = 1, \quad \text{for every } x \in \mathbb{R}. \tag{2}$$

Moreover, using ($\varphi 1$), it is easy to see that

$$K_\phi(x, y) \leq L\, (1 + |x - y|)^{-\alpha}, \quad \text{for every } x,\ y \in \mathbb{R}, \tag{3}$$

for some positive constant L. Under the previous assumptions on K_ϕ, the following lemma, which will turn out to be useful later, could be established. Its proof is classical and can be found in [30].

Lemma 2.3 *Let $(T_w)_{w>0}$ be the family of operators defined explicitly by*

$$(T_w f)(x) := w \int_{\mathbb{R}} K(wx, wy)\, f(y)\, dy, \quad x \in \mathbb{R},$$

for $f : \mathbb{R} \to \mathbb{R}$ (or $\mathbb{C}$), and where the kernel $K : \mathbb{R}^2 \to \mathbb{R}$ (or $\mathbb{C}$) meets the conditions (2) and (3). Then, for any uniformly continuous and bounded function f, we have

$$\lim_{w\to+\infty} \|T_w f - f\|_\infty = 0.$$

Moreover, for every $f \in L^p(\mathbb{R})$, $1 \leq p < +\infty$, it results

$$\lim_{w\to+\infty} \|T_w f - f\|_p = 0.$$

Recall now the following

Definition 2.4 A function $\sigma : \mathbb{R} \to \mathbb{R}$ is called a "sigmoidal function", whenever $\lim_{x\to-\infty} \sigma(x) = 0$ and $\lim_{x\to+\infty} \sigma(x) = 1$.

Sometimes, boundedness, continuity and/or monotonicity are prescribed in addition. Let now $\phi \in \Phi$ be fixed and define the function $\sigma_\phi : \mathbb{R} \to \mathbb{R}_0^+$ as

$$\sigma_\phi(x) := \int_{-\infty}^{x} \phi(t)\, dt, \quad x \in \mathbb{R}. \tag{4}$$

Clearly, from condition (φ2) and Remark 2.2, such a function σ_ϕ is a sigmoidal function. We can now give the following

Definition 2.5 For every fixed function $\phi \in \Phi$, we define the family of operators $(S_w^{\sigma_\phi})_{w>0}$ by

$$(S_w^{\sigma_\phi} f)(x) := \sum_{k\in\mathbb{Z}} \left[\int_a^b \phi(wy-k)\, f'(y)\, dy \right] \sigma_\phi(wx-k) + f(a), \quad x \in [a,b],$$

for every $f \in AC[a,b]$ and $w > 0$. We call $S_w^{\sigma_\phi} f$ the "series of sigmoidal functions for f, based on ϕ", for the given value of $w > 0$.

Clearly, when f is a constant function, the Definition 2.5 becomes trivial. Now, we can prove the following

Theorem 2.6 *Let $\phi \in \Phi$ be fixed. For any given $f \in AC[a,b]$, the family $(S_w^{\sigma_\phi} f)_{w>0}$ converges uniformly to f on $[a,b]$, i.e.,*

$$\lim_{w\to\infty} \| S_w^{\sigma_\phi} f - f \|_\infty = 0.$$

Moreover, if $f \in \widehat{C}^1[a,b]$, we have

$$\| S_w^{\sigma_\phi} f - f \|_\infty \leq \widetilde{C} w^{-1},$$

for some positive constant $\widetilde{C}$ and for every $w > 0$.

Proof Since $f \in AC[a,b]$, $f(x) = \int_a^x f'(z)\, dz + f(a)$ for every $x \in [a,b]$. Then, setting $\widetilde{f}'(z) = f'(z)$ for $z \in [a,b]$ and $\widetilde{f}'(z) = 0$ for $z \notin [a,b]$, we obtain

$$\begin{aligned} |(S_w^{\sigma_\phi} f)(x) - f(x)| &= \left| \sum_{k\in\mathbb{Z}} \left[\int_a^b \phi(wy-k)\, f'(y)\, dy \right] \sigma_\phi(wx-k) - \int_a^x f'(z)\, dz \right| \\ &= \left| \sum_{k\in\mathbb{Z}} \left[\int_{\mathbb{R}} \phi(wy-k)\, \widetilde{f}'(y)\, dy \right] \int_{-\infty}^{wx-k} \phi(t)\, dt - \int_{-\infty}^{x} \widetilde{f}'(z)\, dz \right|. \end{aligned}$$

Changing variable, by setting $t = wz - k$, we get

$$
\begin{aligned}
&|(S_w^{\sigma_\phi} f)(x) - f(x)| \\
&\leq \int_{-\infty}^{x} \left| \sum_{k\in\mathbb{Z}} \left[w \int_{\mathbb{R}} \phi(wy-k)\, \widetilde{f}'(y)\, dy \right] \phi(wz-k) - \widetilde{f}'(z) \right| dz \\
&= \int_{-\infty}^{x} \left| w \int_{\mathbb{R}} K_\phi(wz, wy)\, \widetilde{f}'(y)\, dy - \widetilde{f}'(z) \right| dz \\
&\leq \int_{-\infty}^{+\infty} \left| w \int_{\mathbb{R}} K_\phi(wz, wy)\, \widetilde{f}'(y)\, dy - \widetilde{f}'(z) \right| dz.
\end{aligned} \tag{5}
$$

Being $\widetilde{f}' \in L^1(\mathbb{R})$, we obtain by Lemma 2.3 and inequality (5)

$$
\lim_{w\to+\infty} \|S_w^{\sigma_\phi} f - f\|_\infty \leq \lim_{w\to+\infty} \|T_w \widetilde{f}' - \widetilde{f}'\|_1 = 0,
$$

which completes the proof of the first part of the theorem.

Consider now $f \in \widehat{C}^1[a, b]$. Note that, by conditions ($\varphi 2$) and (2), we have

$$
w \int_{\mathbb{R}} K_\phi(wz, wy)\, dy = 1, \qquad \text{for every } z \in \mathbb{R} \text{ and } w > 0.
$$

Then, again from inequality (5), we obtain

$$
\begin{aligned}
&|(S_w^{\sigma_\phi} f)(x) - f(x)| \\
&\leq \int_{\mathbb{R}} \left| w \int_{\mathbb{R}} K_\phi(wz, wy)\, \widetilde{f}'(y)\, dy - \widetilde{f}'(z) w \int_{\mathbb{R}} K_\phi(wz, wy)\, dy \right| dz \\
&\leq w \int_{\mathbb{R}} \int_{\mathbb{R}} K_\phi(wz, wy)\, |\widetilde{f}'(y) - \widetilde{f}'(z)|\, dy\, dz \\
&\leq 2w \|f'\|_\infty \int_{\mathbb{R}} \int_{\mathbb{R}} K_\phi(wz, wy)\, dy\, dz.
\end{aligned} \tag{6}
$$

Changing the variables z and y in the last integral in (6) with z_1/w and y_1/w, respectively, we obtain, in view of condition (3),

$$
\begin{aligned}
\|S_w^{\sigma_\phi} f - f\|_\infty &\leq 2w^{-1} \|f'\|_\infty \int_{\mathbb{R}} \int_{\mathbb{R}} K_\phi(z_1, y_1)\, dy_1\, dz_1 \\
&\leq 2w^{-1} \|f'\|_\infty L \int_{\mathbb{R}} \int_{\mathbb{R}} (1 + |z_1 - y_1|)^{-\alpha}\, dy_1\, dz_1 =: \widetilde{C} w^{-1},
\end{aligned}
$$

for every $w > 0$, for some $\widetilde{C} > 0$, and where $\alpha \geq 2$ is the constant of condition ($\varphi 1$). This completes the proof of the second part of the theorem. □

Examples of functions $\phi \in \Phi$ will be given in the next sections.

3 Application to neural networks

Here, we give some applications of the theory developed in the previous sections to NNs of the form (II). Below, we will study NNs of the type in (II) in a univariate setting and activated by the sigmoidal functions generated by (4). We will denote by Φ_C the subset of Φ of functions having a compact support.

Let $\phi \in \Phi_C$ be fixed, and let $M_1,\ M_2 > 0$ such that $\operatorname{supp}\phi \subseteq [-M_1, M_2]$. In this case, we have for any $f \in AC[a,b]$ and $w > 0$,

$$\int_a^b \phi(wy-k) f'(y)\, dy = 0,$$

for every $k < wa - M_2$ and $k > wb + M_1$, $k \in \mathbb{Z}$, since for these values of k, $[wa - k, wb - k] \cap [-M_1, M_2] = \emptyset$. Then, the series appearing in the definition of the operator $S_w^{\sigma_\phi} f$ reduces to a finite sum, i.e.,

$$(S_w^{\sigma_\phi} f)(x) = \sum_{k=\lfloor wa-M_2 \rfloor}^{\lceil wb+M_1 \rceil} \left[\int_a^b \phi(wy-k)\, f'(y)\, dy \right] \sigma_\phi(wx-k) + f(a), \tag{7}$$

for every $x \in [a,b]$, where the functions $\lceil x \rceil$ and $\lfloor x \rfloor$ denote the upper and the lower integer part of $x \in \mathbb{R}$, respectively. Now, we introduce the following modification in definition 2.5 for the case $\phi \in \Phi_C$. For any $f \in AC[a,b]$, set

$$(G_w^{\sigma_\phi} f)(x) := \sum_{k=\lfloor wa-M_2 \rfloor}^{\lceil wb+M_1 \rceil} \left[\int_a^b \phi(wy-k)\, f'(y)\, dy \right] \sigma_\phi(wx-k) + f(a)\, \sigma_\phi(w(x-a+1)),$$

for every $x \in [a,b]$ and $w > 0$. The $G_w^{\sigma_\phi} f$'s are a kind of NNs. They approximate f, uniformly on $[a,b]$, as $w \to +\infty$. The proof of this claim follows from the same arguments made in Theorem 2.6, taking into account that

$$\sup_{x\in[a,b]} |f(a)||1-\sigma_\phi(w(x-a+1))| \le |f(a)||1-\sigma_\phi(w)| \ = \ 0, \tag{8}$$

for $w > 0$ sufficiently large. Indeed, by the definition of σ_ϕ, for every $w > M_2$ we have

$$\sigma_\phi(w) = \int_{-\infty}^{w} \phi(x)\, dx = \int_{\mathbb{R}} \phi(x)\, dx = 1. \tag{9}$$

Moreover, again by Theorem 2.6, if $f \in \widehat{C}^1[a,b]$ we obtain the convergence rate given by $\|G_w^{\sigma_\phi} f - f\|_\infty \le \widetilde{C} w^{-1}$, for some positive constants $\widetilde{C}$ and for every sufficiently large $w > 0$.

Our work provides a *unified* approach for NNs approximations. In addition, our proofs are *constructive* in nature and allow us to determine explicitly the form of the NN. In particular, we show that the set of NNs $G_w^{\sigma_\phi} f$ is dense in the set $AC[a,b]$, with respect to the uniform norm.

Now, we show that we can obtain NNs also starting from functions $\phi \in \Phi$ which are not necessarily compactly supported. Let first prove the following

Lemma 3.1 *The series* $\sum_{k\in\mathbb{Z}} \phi(wx-k)$ *converges uniformly on the compact subsets of* $\mathbb{R}$, *for every fixed* $w > 0$.

In particular, we have for every $[a,b] \subset \mathbb{R}$

$$\sup_{x\in[a,b]} \sum_{|k|>N} \phi(wx-k) \le \overline{C}\left\{(N-wb+1)^{-(\alpha-1)} + (N+wa+1)^{-(\alpha-1)}\right\},$$

for some $\overline{C} > 0$*, for every* $N > w\max\{|a|,|b|\}$, $N \in \mathbb{N}^+$*, where* $\alpha \ge 2$ *is the constant of condition* $(\varphi 1)$.

Proof Let $[a,b] \subset \mathbb{R}$ be fixed. By condition $(\varphi 1)$ and for $N > w\max\{|a|,|b|\}$ we have

$$\sup_{x\in[a,b]} \sum_{|k|>N} \phi(wx-k) \le C \sup_{x\in[a,b]} \sum_{|k|>N} (1+|wx-k|)^{-\alpha}$$

$$= C\left\{\sup_{x\in[a,b]} \sum_{k>N} (1+|wx-k|)^{-\alpha} + \sup_{x\in[a,b]} \sum_{k>N} (1+|wx+k|)^{-\alpha}\right\}$$

$$\le C\left\{\sum_{k>N}(1+k-wb)^{-\alpha} + \sum_{k>N}(1+wa+k)^{-\alpha}\right\} \le C\left\{\int_N^{+\infty}(1+x-wb)^{-\alpha}dx\right.$$

$$\left. + \int_N^{+\infty}(1+wa+x)^{-\alpha}dx\right\} =: \overline{C}\left\{(N-wb+1)^{-(\alpha-1)} + (N+wa+1)^{-(\alpha-1)}\right\}.$$

The proof then follows. □

We can now establish the following

Theorem 3.2 *(i) For any* $f \in AC[a,b]$*, we denote by*

$$(G^{\sigma_\phi}_{N,w} f)(x) := \sum_{k=-N}^{N}\left[\int_a^b \phi(wy-k)\, f'(y)\, dy\right]\sigma_\phi(wx-k)$$
$$+ f(a)\,\sigma_\phi(w(x-a+1)), \tag{10}$$

for $x \in [a,b]$, $w > 0$*, and* $N \in \mathbb{N}^+$*. Then, for every* $\varepsilon > 0$ *there exist* $w > 0$ *and* $N \in \mathbb{N}^+$ *such that*

$$\|G^{\sigma_\phi}_{N,w} f - f\|_\infty < \varepsilon.$$

(ii) Moreover, for any $f \in \widehat{C}^1[a,b]$ *we have*

$$\|G^{\sigma_\phi}_{N,w} f - f\|_\infty \le C_1\left\{(N-wb+1)^{-(\alpha-1)} + (N+wa+1)^{-(\alpha-1)}\right\}$$
$$+ C_2\, w^{-1} + C_3\, w^{-(\alpha-1)},$$

for some constants $C_1, C_2, C_3 > 0$*, and for every* $w > 0$ *with* $N > w\max\{|a|,|b|\}$, $N \in \mathbb{N}^+$*, where* $\alpha \ge 2$ *is the constant appearing in condition* $(\varphi 1)$.

Proof (i) Let $\varepsilon > 0$ be fixed. For every $x \in [a,b]$ we have

$$|(G^{\sigma_\phi}_{N,w} f)(x) - f(x)| \le |(G^{\sigma_\phi}_{N,w} f)(x) - (S^{\sigma_\phi}_w f)(x)| + |(S^{\sigma_\phi}_w f)(x) - f(x)|$$

$$\le \sum_{|k|>N}\left[\int_a^b \phi(wy-k)\,|f'(y)|\,dy\right]\sigma_\phi(wx-k) + |f(a)||1-\sigma_\phi(w(x-a+1))|$$
$$+ \|S^{\sigma_\phi}_w f - f\|_\infty =: S_1 + S_2 + S_3. \tag{11}$$

Proceeding as in (8) and using ($\varphi 1$), we can write

$$S_2 \leq |f(a)||1-\sigma_\phi(w)| = |f(a)| \int_w^{+\infty} \phi(x)\,dx$$

$$\leq |f(a)| C \int_w^{+\infty} (1+x)^{-\alpha} dx =: \underline{C}\,(1+w)^{-(\alpha-1)}, \qquad (12)$$

where $\alpha \geq 2$ is the constant appearing in condition ($\varphi 1$), and $\underline{C} > 0$, then $S_2 < \varepsilon$ for $w > 0$ sufficiently large. Moreover, we obtain from Theorem 2.6 that $S_3 < \varepsilon$ for $w > 0$ sufficiently large. Finally, we can estimate S_1. Being $\|\sigma_\phi\|_\infty \leq 1$, we obtain for S_1

$$S_1 \;\leq\; \|\sigma_\phi\|_\infty \sum_{|k|>N} \left[\int_a^b \phi(wy-k)\,|f'(y)|\,dy \right]$$

$$\leq \left[\sup_{y\in[a,b]} \sum_{|k|>N} \phi(wy-k) \right] \int_a^b |f'(y)|\,dy. \qquad (13)$$

We have by Lemma 3.1, for every fixed and sufficiently large $w > 0$,

$$\sup_{y\in[a,b]} \sum_{|k|>N} \phi(wy-k) \leq \overline{C} \left\{ (N-wb+1)^{-(\alpha-1)} + (N+wa+1)^{-(\alpha-1)} \right\}, \qquad (14)$$

for some constant $\overline{C} > 0$ and for every $N > w \max\{|a|, |b|\}$ with $N \in \mathbb{N}^+$. Then, for N sufficiently large, we obtain $S_1 < \varepsilon$. This completes the proof of (i).

(ii) For any $f \in \widehat{C}^1[a,b]$, Theorem 2.6 shows that $S_3 \leq \widetilde{C} w^{-1}$ uniformly with respect to $x \in [a,b]$, for every $w > 0$. Moreover, we obtain by (12) and (14)

$$S_1 + S_2 + S_3 \leq \overline{C} \left[\int_a^b |f'(y)|\,dy \right] \left\{ (N-wb+1)^{-(\alpha-1)} \right.$$

$$\left. + (N+wa+1)^{-(\alpha-1)} \right\} + \widetilde{C}\,w^{-1} + \underline{C}\,(1+w)^{-(\alpha-1)}$$

$$\leq C_1 \left\{ (N-wb+1)^{-(\alpha-1)} + (N+wa+1)^{-(\alpha-1)} \right\}$$

$$+ C_2\,w^{-1} + C_3\,w^{-(\alpha-1)},$$

uniformly with respect to $x \in [a,b]$, for some constants $C_1,\ C_2,\ C_3 > 0$, and for $w > 0$ sufficiently large, with $N > w \max\{|a|,\ |b|\}$. □

Remark 3.3 Setting $C_3 = 0$ in Theorem 3.2 (ii), we also obtain an estimate for the *truncation error* for the series of sigmoidal functions introduced in Sect. 2. Note that, when the weight, w, increases, we need a higher number of neurons, N, which depends on w.

We now construct few examples of sigmoidal functions, σ_ϕ, providing first some examples of functions $\phi \in \Phi_C$ satisfying all hypotheses of our theory. Recall that the "central B-splines" of order $n \in \mathbb{N}^+$, are defined as

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

where $(x)_+ := \max\{x, 0\}$ is the positive part of $x \in \mathbb{R}$ [5]. The Fourier transform of M_n is given by

$$\widehat{M_n}(v) := \text{sinc}^n\left(\frac{v}{2\pi}\right), \quad v \in \mathbb{R},$$

where the *sinc* function is defined by

$$\text{sinc}(x) := \begin{cases} \dfrac{\sin(\pi x)}{\pi x}, & x \in \mathbb{R} \setminus \{0\}, \\ 1, & x = 0. \end{cases}$$

The M_n's are bounded and continuous on $\mathbb{R}$ for all $n \in \mathbb{N}^+$, and are compactly supported on $[-n/2, n/2]$. This implies that $M_n \in L^1(\mathbb{R})$ and satisfies condition ($\varphi 1$) for every $\alpha \geq 2$. Finally, condition ($\varphi 2$) holds, in view of Remark 2.2, hence, $M_n \in \Phi_C$ for every $n \in \mathbb{N}^+$. Therefore, we can construct explicitly the NNs $G_w^{\sigma_{M_n}} f$, $n \in \mathbb{N}^+$.

As an example of function $\phi \in \Phi$ which is not compactly supported, consider the continuous function

$$F(x) := \frac{1}{2\pi}\text{sinc}^2\left(\frac{x}{2\pi}\right), \qquad x \in \mathbb{R}.$$

Clearly, $F(x) = \mathcal{O}(x^{-2-\varepsilon})$ as $x \to \pm\infty$, $\varepsilon > 0$, hence, F satisfies condition ($\varphi 1$) with $\alpha = 2$, see [5]. Moreover, its Fourier transform is

$$\widehat{F}(v) := \begin{cases} 1 - |v|, & |v| \leq 1, \\ 0, & |v| > 1, \end{cases}$$

(see [5] again). By Remark 2.2, F satisfies also condition ($\varphi 2$), and then $F \in \Phi$.

Remark 3.4 Note that the theory developed in this section cannot be applied to the case of NNs activated by the logistic functions, $\sigma_\ell(x) := (1 + e^{-x})^{-1}$ (see [4,21], e.g., for applications to Demography and Economics), or to the hyperbolic tangent sigmoidal functions, $\sigma_h(x) := \frac{1}{2} + \frac{1}{2}\tanh(x) = \frac{1}{2} + \frac{e^{2x}-1}{2(e^{2x}+1)}$, [1,2,8]. In fact, σ_ℓ and σ_h can be generated by (4) from $\phi_\ell(x) := e^{-x}(1 + e^{-x})^{-2}$ and $\phi_h(x) := 2e^{2x}(e^{2x} + 1)^{-2}$, respectively. However, $\widehat{\phi_\ell}(v) = \pi v/\sinh(\pi v)$ and $\widehat{\phi_h}(v) = \pi v/(2\sinh(\pi v/2))$, respectively, which do not meet the condition in Remark 2.2, i.e., do not satisfy condition ($\varphi 2$). In Sect. 5 below, an extension of the theory developed above is proposed, which allows to use NNs activated by σ_ℓ or σ_h.

4 Sigmoidal functions and multiresolution approximation

In this section, we will show a connection between the theory of multiresolution approximation and our theory for approximating functions by series of sigmoidal functions. We first recall some basic facts concerning the multiresolution approximation. For the detailed theory, see [11,17,29,30,36]. We start recalling the following

Definition 4.1 A multiresolution approximation of $L^2(\mathbb{R})$ is an increasing sequence, V_j, $j \in \mathbb{Z}$, of linear closed subspaces of $L^2(\mathbb{R})$, enjoying the following properties:

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\}, \qquad \bigcup_{j \in \mathbb{Z}} V_j \text{ is dense in } L^2(\mathbb{R}); \tag{15}$$

for all $f \in L^2(\mathbb{R})$ and all $j \in \mathbb{Z}$,

$$f(x) \in V_j \Longleftrightarrow f(2x) \in V_{j+1}; \tag{16}$$

for all $f \in L^2(\mathbb{R})$ and all $k \in \mathbb{Z}$,

$$f(x) \in V_0 \Longleftrightarrow f(x-k) \in V_0; \tag{17}$$

there exists a function, $h(x) \in V_0$, such that the sequence

$$(h(x-k))_{k\in\mathbb{Z}} \text{ is a Riesz basis of } V_0. \tag{18}$$

Recall that a sequence of functions $(h_k)_{k\in\mathbb{Z}}$ is a Riesz basis of an Hilbert space, $H \subseteq L^2(\mathbb{R})$, if there exist two constants, C_1 and C_2, with $C_1 > C_2 > 0$, such that, for every sequence of real or complex numbers $(a_k)_{k\in\mathbb{Z}} \in l^2(\mathbb{Z})$, it turns out that

$$C_2 \left(\sum_{k\in\mathbb{Z}} |a_k|^2 \right)^{1/2} \le \left\| \sum_{k\in\mathbb{Z}} a_k h_k \right\|_{L^2(\mathbb{R})} \le C_1 \left(\sum_{k\in\mathbb{Z}} |a_k|^2 \right)^{1/2},$$

and the vector space of finite linear combinations of h_k, is dense in H.

Definition 4.2 A multiresolution approximation, $V_j,\ j \in \mathbb{Z}$, is called r-regular ($r \in \mathbb{N}^+$), if the function h in (18) is such that $h \in C^r(\mathbb{R})$ and

$$|h^{(i)}(x)| \le C_m (1+|x|)^{-m}, \quad x \in \mathbb{R}, \tag{19}$$

for each integer $m \in \mathbb{N}^+$ and for every positive index $i \le r$.

For every r-regular multiresolution approximation $V_j,\ j \in \mathbb{Z}$, we can define the function $\phi \in L^2(\mathbb{R})$, called *scaling function*, as

$$\widehat{\phi}(v) := \widehat{h}(v) \left(\sum_{k\in\mathbb{Z}} |\widehat{h}(v+2\pi k)|^2 \right)^{-1/2}, \quad v \in \mathbb{R}. \tag{20}$$

In [30, Ch. 2], it is proved that $\sum_{k\in\mathbb{Z}} |\widehat{h}(v+2\pi k)|^2 \ge c > 0$, hence, ϕ is well-defined. Moreover, by the regularity of h, we have, as a consequence of the Sobolev's embedding theorem, that $\sum_{k\in\mathbb{Z}} |\widehat{h}(v+2\pi k)|^2$ is a $C^\infty(\mathbb{R})$ function. Furthermore, the family $(\phi(x-k))_{k\in\mathbb{Z}}$ turns out to be an orthonormal basis of V_0, [17,30], and from (16) and (17), we obtain by a simple change of scale that $(2^{j/2}\phi(2^j x-k))_{k\in\mathbb{Z}}$ forms an orthonormal basis of V_j.

Now, by smoothness and periodicity of $\left(\sum_{k\in\mathbb{Z}} |\widehat{h}(v+2\pi k)|^2\right)^{-1/2}$, the latter can be written by means of its Fourier series $\sum_{k\in\mathbb{Z}} \alpha_k e^{ikv}$, where the coefficients α_k decrease rapidly. We thus obtain $\widehat{\phi}(v) = \left(\sum_{k\in\mathbb{Z}} \alpha_k e^{ikv}\right) \widehat{h}(v)$ which gives $\phi(x) = \sum_{k\in\mathbb{Z}} \alpha_k\, h(x+k)$, and then it follows that the scaling function ϕ satisfies the estimates in (19). In particular, we have

$$|\phi(x)| \le \widetilde{C}_\alpha\, (1+|x|)^{-\alpha}, \qquad x \in \mathbb{R}, \tag{21}$$

for some $\widetilde{C}_\alpha > 0$, for every integer $\alpha \in \mathbb{N}^+$, i.e., ϕ satisfies condition ($\varphi 1$) for every $\alpha \in \mathbb{N}^+$.

Let now E_j be the orthogonal projection of $L^2(\mathbb{R})$ onto V_j, given by

$$(E_j f)(x) := \sum_{k\in\mathbb{Z}} \left[2^j \int_{\mathbb{R}} f(y)\, \overline{\phi}(2^j y - k)\, dy \right] \phi(2^j x - k), \quad f \in L^2(\mathbb{R}), \tag{22}$$

where $\overline{\phi}$ is the complex conjugate of ϕ. Let define $E(x, y) := \sum_{k\in\mathbb{Z}} \overline{\phi}(y-k)\, \phi(x-k)$, the kernel of the projection operator E_0, hence, $2^j\, E(2^j x, 2^j y),\ j \in \mathbb{Z}$ will be the kernel of the projection operator E_j.

Again in [30], it is proved the following remarkable property for the kernel, E,

$$\int_{\mathbb{R}} E(x,y)\, y^{\alpha}\, dy = x^{\alpha}, \quad \text{for every } x \in \mathbb{R}, \tag{23}$$

for every integer $\alpha \in \mathbb{N}$ and $\alpha \leq r$. From (23) with $\alpha = 0$, the integral property

$$\int_{\mathbb{R}} E(x,y)\, dy = 1, \quad \text{for every } x \in \mathbb{R},$$

follows. Moreover, since ϕ satisfies (21), it is easy to see that

$$|E(x,y)| \leq \overline{C}_{\alpha}(1+|x-y|)^{-\alpha}, \quad \forall\, (x,y) \in \mathbb{R}^2, \text{ and } \quad \forall\, \alpha \in \mathbb{N}^+,$$

where $\overline{C}_{\alpha} > 0$. Hence, E is a bivariate kernel satisfying conditions (2) and (3). Then, by Lemma 2.3, we infer that $\|E_j f - f\|_p \to 0$ as $j \to +\infty$, for every $f \in L^p(\mathbb{R})$ and $1 \leq p < \infty$. Moreover, exploiting the properties of the projection operators E_j, the quantity

$$\Sigma(x,v) := \sum_{k\in\mathbb{Z}} e^{i2\pi kx}\, \widehat{\phi}(v+2k\pi)\, \overline{\widehat{\phi}}(v), \quad x,\, v \in \mathbb{R},$$

can be defined, which satisfies the condition $\Sigma(x,0) = 1$, for every $x \in \mathbb{R}$, [30]. This yields

$$\sum_{k\in\mathbb{Z}} e^{i2\pi kx} \widehat{\phi}(2k\pi)\, \overline{\widehat{\phi}}(0) = 1. \tag{24}$$

Now, we can adjust the scaling function ϕ merely multiplying $\widehat{\phi}$ by a suitable constant of modulus 1 so that $\widehat{\phi}(0) = \int_{\mathbb{R}} \phi(t)\, dt = 1$, while preserving all the other properties, [30]. By the regularity of ϕ, the Poisson summation formula holds, and from (24), we obtain

$$1 = \sum_{k\in\mathbb{Z}} e^{i2\pi kx}\, \widehat{\phi}(2k\pi) = \sum_{k\in\mathbb{Z}} \phi(x+k) = \sum_{k\in\mathbb{Z}} \phi(x-k), \quad x \in \mathbb{R},$$

i.e., the scaling function ϕ satisfies condition ($\varphi 2$). Using (4), we can now consider the function σ_{ϕ} constructed by the scaling function ϕ. Clearly, if ϕ is *real valued*, σ_{ϕ} turns out to be a sigmoidal function. Then, we have the following

Theorem 4.3 *Let ϕ be a real-valued* scaling *function like that constructed above, associated with an r-* regular multiresolution *approximation of $L^2(\mathbb{R})$.*

(i) Then, for any $f \in AC[a,b]$, the sequence of operators $(S_j^{\sigma_{\phi}} f)_{j\in\mathbb{N}^+}$, defined by

$$(S_j^{\sigma_{\phi}} f)(x) := \sum_{k\in\mathbb{Z}} \left[\int_a^b \phi(2^j y - k)\, f'(y)\, dy \right] \sigma_{\phi}(2^j x - k) + f(a),$$

for every $x \in [a,b]$, converges uniformly to f on $[a,b]$.
In particular, if $f \in \widehat{C}^1[a,b]$, we have

$$\|S_j^{\sigma_{\phi}} f - f\|_{\infty} \leq C\, 2^{-j},$$

for some positive constant C and for every positive integer j.

(ii) *Denote by* $S^{\sigma_\phi}_{N,j} f$, $N \in \mathbb{N}^+$, *the truncated series* $S^{\sigma_\phi}_j f$, *i.e.,*

$$(S^{\sigma_\phi}_{N,j} f)(x) := \sum_{k=-N}^{N} \left[\int_a^b \phi(2^j y - k)\, f'(y)\, dy \right] \sigma_\phi(2^j x - k) + f(a).$$

Then, for every $f \in \widehat{C}^1[a,b]$, *we have*

$$\|S^{\sigma_\phi}_{N,j} f - f\|_\infty \le C_1\, 2^{-j} + C_{2,\alpha} \left\{ (N - 2^j b + 1)^{-(\alpha-1)} + (N + 2^j a + 1)^{-(\alpha-1)} \right\},$$

for some positive constants C_1 *and* $C_{2,\alpha}$, *for every* $j \in \mathbb{N}^+$, *and* $N > 2^j \max\{|a|, |b|\}$, *where* $\alpha \in \mathbb{N}^+$ *is an arbitrary integer.*

The proof of Theorem 4.3 (i) follows as the proof of Theorem 2.6, taking into account that, the sequence $(E_j f)_{j \in \mathbb{Z}}$, $f \in L^1(\mathbb{R})$, converges to f in $L^1(\mathbb{R})$. Moreover, the proof of Theorem 4.3 (ii) follows, as the proof of Theorem 3.2 (ii), using condition (21) and Lemma 3.1, where we have 2^j in place of w.

Remark 4.4 Note that, in the special setting of r-regular multiresolution approximations, we are able to prove that the real-valued scaling functions ϕ, constructed above, are such that $\phi \in \Phi$. Moreover, condition (16) in definition 4.1 allows us to consider the *weights* in the basis $(2^{j/2}\phi(2^j x - k))_{k \in \mathbb{Z}}$, and then in the series $S^{\sigma_\phi}_j f$, as 2^j, i.e., the weights increase exponentially with respect to j. Then, the error of approximation of C^1-functions decreases as 2^{-j}. Moreover, conditions (19) and (21) are crucial to prove that the *truncation error* also decrease rapidly.

Examples of r-regular multiresolution analysis satisfying the conditions above can be given, assuming h to be generated by spline wavelets of order $r + 1$. These are defined by

$$h_r(x) := \frac{1}{r!} \sum_{i=0}^{r+1} (-1)^i \binom{r+1}{i} (x - i)^r_+, \quad x \in \mathbb{R}, \tag{25}$$

which can be viewed just as shifted central B-spline M_n. Generally speaking, the definition of h_n is given in terms of convolution, i.e., h_n can be defined as the convolution of $r + 1$ characteristic functions of the interval $[0, 1)$, see [35]. Note that, also the central B-spline can be defined similarly, in terms of convolutions of the characteristic functions of the interval $[-1/2, 1/2)$, see [5]. The Fourier transform of h_r can be easily obtained by

$$\widehat{h}_r(v) := e^{-iv(r+1)/2}\, \mathrm{sinc}^{r+1}\left(\frac{v}{2\pi}\right), \quad v \in \mathbb{R}.$$

The scaling function ϕ associated with the spline wavelet multiresolution approximation can be obtained using (20) and the normalization procedure described above, see [18,30,35].

5 An extension of the theory for neural networks approximation

The theory developed in the previous sections, concerning the approximation by means of series of sigmoidal functions based on σ_ϕ is beset by the technical difficulty of checking that ϕ satisfies condition $(\varphi 2)$. To this purpose, we could use the condition given in Remark 2.2. However, this does not simplify the problem. In fact, evaluating the Fourier transform of a given function is often a difficult task. Moreover, as noticed in Remark 3.4, the sigmoidal functions most used for NN approximation do not satisfy ($\varphi 2$). Below, we propose an extension of the theory developed in the previous sections, aiming at obtaining approximations with NNs activated by sigmoidal functions σ_ϕ, without assuming that condition ($\varphi 2$) be satisfied by ϕ.

Through this section, we consider functions $\phi : \mathbb{R} \to \mathbb{R}_0^+$, with $\int_{\mathbb{R}} \phi(t)\, dt = 1$ and satisfying condition ($\varphi 1$) with $\alpha > 2$. Moreover, we set

$$\psi_\phi(t) := \sigma_\phi(t+1) - \sigma_\phi(t) > 0, \quad t \in \mathbb{R},$$

and assume in addition that ψ_ϕ satisfies:

$$(\Psi 1) \qquad \psi_\phi(t) \le A(1+|t|)^{-\alpha},$$

for every $t \in \mathbb{R}$ and some $A > 0$. We denote by $\mathcal{T}$ the set of all functions ϕ satisfying such conditions. We can now prove the following

Lemma 5.1 *For any given $\phi \in \mathcal{T}$, the relation*

$$\sum_{k \in \mathbb{Z}} \psi_\phi(x-k) = 1, \quad x \in \mathbb{R}$$

holds.

Proof Let $x \in \mathbb{R}$ be fixed. Then,

$$\sum_{k=-N}^{N} \psi_\phi(x-k) = \sum_{k=-N}^{N} [\sigma_\phi(x-k+1) - \sigma_\phi(x-k)] = \sigma_\phi(x+N+1) - \sigma_\phi(x-N),$$

since the sum is telescopic. Passing to the limit for $N \to +\infty$, we obtain immediately

$$\sum_{k=-\infty}^{+\infty} \psi_\phi(x-k) = \lim_{N \to +\infty} [\sigma_\phi(x+N+1) - \sigma_\phi(x-N)] = 1.$$

□

Let now introduce the bivariate kernel

$$K_{\phi,\psi}(x,y) := \sum_{k \in \mathbb{Z}} \psi_\phi(x-k)\, \phi(y-k), \quad (x,y) \in \mathbb{R}^2.$$

As made in Sect. 2 for the kernel K_ϕ, we can show, using Lemma 5.1 and conditions ($\varphi 1$) and ($\Psi 1$), that $K_{\phi,\psi}$ satisfy both, (2) and (3). Now, for any given $\phi \in \mathcal{T}$, we consider the family of operators $(F_w^\phi)_{w>0}$, defined by

$$(F_w^\phi f)(x) := \sum_{k \in \mathbb{Z}} w \left[\int_{\mathbb{R}} \phi(wy-k) f(y)\, dy \right] \psi_\phi(wx-k)$$
$$:= w \int_{\mathbb{R}} K_{\phi,\psi}(wx, wy) f(y)\, dy, \quad x \in \mathbb{R},$$

for every bounded $f : \mathbb{R} \to \mathbb{R}$, $w > 0$.

Remark 5.2 Note that, by Lemma 2.3, for every uniformly continuous and bounded function f, the family of operators $(F_w^\phi f)_{w>0}$ converges uniformly to f on $\mathbb{R}$, as $w \to +\infty$.

To study the order of approximation for the operators above, we define the Lipschitz class of the Zygmund type we will work with. Let us define

$$\mathrm{Lip}(\nu) := \left\{ f : \mathbb{R} \to \mathbb{R} : \| f(\cdot) - f(\cdot + t)\|_\infty = \mathcal{O}(|t|^\nu) \text{ as } t \to 0 \right\},$$

for every $0 < \nu \leq 1$. We can now prove the following lemma concerning the order of approximation of $(F_w^{\sigma_\phi} f)_{w>0}$ to $f(x)$:

Lemma 5.3 *Let $f \in Lip(\nu)$, $0 < \nu \leq 1$, be a fixed bounded function. Then, there exist $C_1 > 0$ and $C_2 > 0$ such that*

$$\sup_{x\in\mathbb{R}} |(F_w^\phi f)(x) - f(x)| \leq C_1\, w^{-\nu} + C_2\, w^{-(\alpha-1)},$$

for every sufficiently large $w > 0$.

Proof Let $x \in \mathbb{R}$ be fixed. Since $f \in \mathrm{Lip}(\nu)$, there exist $M > 0$ and $\gamma > 0$ such that

$$\| f(\cdot) - f(\cdot + t)\|_\infty \leq M\, |t|^\nu,$$

for every $|t| \leq \gamma$. Moreover, we infer from condition (2)

$$w \int_{\mathbb{R}} K_{\phi,\psi}(wx, wy)\, dy = 1, \quad x \in \mathbb{R}, \tag{26}$$

and then we can write

$$\begin{aligned} |(F_w^\phi f)(x) - f(x)| &\leq w \int_{\mathbb{R}} K_{\phi,\psi}(wx, wy)\, |f(y) - f(x)|\, dy \\ &= \left[\int_{|y-x|\leq\gamma} + \int_{|y-x|>\gamma} \right] w\, K_{\phi,\psi}(wx, wy)\, |f(y) - f(x)|\, dy =: J_1 + J_2. \end{aligned}$$

Let first estimate J_1. From (3) and (26), by the change of variable $y = (t/w) + x$, and being $f \in \mathrm{Lip}(\nu)$, we obtain for $w > 0$ sufficiently large

$$\begin{aligned} J_1 &= \int_{w^{-1}|t|\leq\gamma} K_{\phi,\psi}(wx, t + wx)\, |f\,(x + t/w) - f(x)|\, dt \\ &\leq M \left[\int_{|t|\leq w\,\gamma} K_{\phi,\psi}(wx, t + wx) \left| \frac{t}{w} \right|^\nu dt \right] \leq \widetilde{L} w^{-\nu} \int_{\mathbb{R}} (1 + |t|)^{-\alpha}\, |t|^\nu\, dt, \end{aligned}$$

where $\widetilde{L} > 0$ is a suitable constant. Now, since $\alpha > 2$, we have $\widetilde{L} \int_{\mathbb{R}} (1 + |t|)^{-\alpha}\, |t|^\nu\, dt =: C_1 < +\infty$, then $J_1 \leq C_1\, w^{-\nu}$, for $w > 0$ sufficiently large. Moreover, setting $t = wy$ and using again condition (3), we have

$$\begin{aligned} J_2 &= \int_{|t-wx|>w\gamma} K_{\phi,\psi}(wx, t) |f(t/w) - f(x)|\, dt \\ &\leq 2\|f\|_\infty \int_{|t-wx|>w\gamma} K_{\phi,\psi}(wx, t)\, dt \leq \overline{L} \int_{|t-wx|>w\gamma} (1 + |t - wx|)^{-\alpha}\, dt, \end{aligned}$$

where $\overline{L}$ is a suitable positive constant. Changing now the variable t into z, setting $z = t - wx$ in the last integral, we obtain

$$J_2 \leq \overline{L} \int_{|z|>w\gamma} (1+|z|)^{-\alpha}\, dz \leq C_2\, w^{-(\alpha-1)},$$

for every $w > 0$. This completes the proof. ◻

We can now prove the following

Theorem 5.4 *Let $\phi \in \mathcal{T}$ be fixed. Define the NNs*

$$(N^{\phi}_{N,w} f)(x) := \sum_{k=-N}^{N} w \left[\int_{\mathbb{R}} \phi(wy-k)\, f(y)\, dy\right] \psi_\phi(wx-k), \quad x \in \mathbb{R},$$

where $w > 0,\ N \in \mathbb{N}^+$, and $f : \mathbb{R} \to \mathbb{R}$ is a bounded function on $\mathbb{R}$.

(i) Let $f \in C[a,b]$ be fixed. Then, for every $\varepsilon > 0$, there exist $w > 0$ and $N > w \max\{|a|,\ |b|\}$, such that

$$\|N^{\phi}_{N,w} \widetilde{f} - f\|_\infty = \sup_{x\in[a,b]} |(N^{\phi}_{N,w} \widetilde{f})(x) - f(x)| < \varepsilon,$$

where $\widetilde{f}$ is a continuous extensions of f such that $\widetilde{f}$ has compact support and $\widetilde{f} = f$ on $[a,b]$.

(ii) Let $f \in Lip(\nu),\ 0 < \nu \leq 1$, and $[a,b] \subset \mathbb{R}$ be fixed. Then, we have

$$\|N^{\phi}_{N,w} f - f\|_\infty = \sup_{x\in[a,b]} |(N^{\phi}_{N,w} f)(x) - f(x)|$$
$$\leq C_1\, w^{-\nu} + C_2\, w^{-(\alpha-1)} + C_3 \left\{(N - wb + 1)^{-(\alpha-1)} + (N + wa + 1)^{-(\alpha-1)}\right\},$$

for every sufficiently large $w > 0$ and $N > w \max\{|a|, |b|\}$, for some positive constants $C_1,\ C_2$, and C_3.

Proof (i) Suppose for the sake of simplicity that $\|f\|_\infty = \|\widetilde{f}\|_\infty$, and note that $\widetilde{f}$ is uniformly continuous. Let now $\varepsilon > 0$ and $x \in [a,b]$ be fixed. We can write

$$|(N^{\phi}_{N,w} \widetilde{f})(x) - f(x)| \leq |f(x) - (F^{\phi}_{w} \widetilde{f})(x)| + |(F^{\phi}_{w} \widetilde{f})(x) - (N^{\phi}_{N,w} \widetilde{f})(x)|$$
$$=: I_1 + I_2.$$

By Remark 5.2 we have $I_1 < \varepsilon$ for $w > 0$ sufficiently large. Moreover,

$$I_2 \leq \sum_{|k|>N} w \left[\int_{\mathbb{R}} \phi(wy-k)\, |\widetilde{f}(y)|\, dy\right] \psi_\phi(wx-k).$$

Hence, $w \int_{\mathbb{R}} \phi(wy-k)\,dy = 1$, and since $(\Psi 1)$ holds, we obtain for ψ_ϕ the same estimate given in Lemma 3.1 for ϕ, then for every fixed sufficiently large $w > 0$ we have

$$\begin{aligned} I_2 &\leq \|f\|_\infty \sup_{x\in[a,b]} \sum_{|k|>N} \left[w \int_{\mathbb{R}} \phi(wy-k)\,dy \right] \psi_\phi(wx-k) \\ &= \|f\|_\infty \left[\sup_{x\in[a,b]} \sum_{|k|>N} \psi_\phi(wx-k) \right] \\ &< \|f\|_\infty \widetilde{C} \left\{ (N-wb+1)^{-(\alpha-1)} + (N+wa+1)^{-(\alpha-1)} \right\} < \varepsilon, \qquad (27) \end{aligned}$$

for some positive constant $\widetilde{C}$, $N \in \mathbb{N}^+$, $N > w \max\{|a|, |b|\}$, and then, (i) is proved being $\varepsilon > 0$ arbitrary.

(ii) Let now $f \in \mathrm{Lip}(\nu)$ be a fixed. We have by Lemma 5.3

$$I_1 \leq C_1\, w^{-\nu} + C_2\, w^{-(\alpha-1)},$$

for every sufficiently large $w > 0$ and for some positive constants C_1 and C_2. Moreover, we obtain from (27)

$$I_2 \leq C_3 \left\{ (N-wb+1)^{-(\alpha-1)} + (N+wa+1)^{-(\alpha-1)} \right\},$$

for a suitable constant $C_3 > 0$. Then, the second part of the theorem is proved.

□

As a first example, we can consider the case of the *logistic* function, σ_ℓ (see, e.g., [8]), generated by $\phi_\ell(x) := e^{-x}(1+e^{-x})^{-2}$. Clearly, conditions $(\varphi 1)$ and $(\Psi 1)$, are fulfilled, since ϕ_ℓ and

$$\psi_\ell(x) := \sigma_\ell(x+1) - \sigma_\ell(x) = \frac{e\,(e-1)\,e^{-x}}{(1+e^{-x-1})(1+e^{-x})},$$

decay exponentially as $x \to \pm\infty$. A second example, is given by the *hyperbolic tangent* sigmoidal function (see, e.g., [1,2]),

$$\sigma_h(x) := \frac{1}{2} + \frac{1}{2}\tanh(x) = \frac{1}{2} + \frac{e^{2x}-1}{2(e^{2x}+1)}.$$

This can be generated by $\phi_h(x) = 2\,e^{2x}\,(e^{2x}+1)^{-2}$, whose associated function ψ_h is

$$\psi_h(x) = \frac{(e^2-1)\,e^{2x}}{(e^{2x+2}+1)(e^{2x}+1)}.$$

It can be easily checked that such a function ϕ_h belongs to $\mathcal{T}$.

Finally, we recall that another remarkable example of sigmoidal function is provided by the class of Gompetz functions, defined by

$$\sigma_{\alpha\beta}(x) := e^{-\alpha\, e^{-\beta x}}, \quad x \in \mathbb{R},$$

for α, $\beta > 0$. Gompertz functions are widely used in such fields as, for instance, demography and in modeling tumor growth.

Remark 5.5 Note that, in closing, in order to approximate functions by the NNs $G^{\sigma_\phi}_{N,w}$, the half of the number of sigmoidal functions needed to approximate functions by the NNs $N^{\phi}_{N,w}$, would now suffice. The theory developed in this section, however, can be applied to important sigmoidal functions for which the theory earlier discussed in Sects. 2 and 3 cannot be applied.

Acknowledgments This work was supported, in part, by the GNAMPA and the GNFM of the Italian INdAM.

References

1. Anastassiou, G.A.: Univariate hyperbolic tangent neural network approximation. Math. Comput. Model. **53**(5–6), 1111–1132 (2011)
2. Anastassiou, G.A.: Multivariate hyperbolic tangent neural network approximation. Comput. Math. Appl. **61**(4), 809–821 (2011)
3. Barron, A.R.: Universal approximation bounds for superpositions of a sigmoidal function. IEEE Trans. Inf. Theory **39**(3), 930–945 (1993)
4. Brauer, F., Castillo Chavez, C.: Mathematical Models in Population Biology and Epidemiology. Springer, New York (2001)
5. Butzer, P.L., Nessel, R.J.: Fourier Analysis and Approximation. Birkhauser Verlag, Bassel and Academic Press, New York (1971)
6. Butzer, P.L., Splettstößer, W., Stens, R.L.: The sampling theorem and linear prediction in signal analysis. Jahresber. Deutsch. Math.-Verein **90**, 1–70 (1988)
7. Buzhabadi, R., Effati, S.: A neural network approach for solving Fredholm integral equations of the second kind. Neural Comput. Appl. **21**, 843–852 (2012)
8. Cao, F., Chen, Z.: The approximation operators with sigmoidal functions. Comput. Math. Appl. **58**(4), 758–765 (2009)
9. Chen, H., Chen, T., Liu, R.: A constructive proof and an extension of Cybenko's approximation theorem. In: Computing Science and Statistics, pp. 163–168. Springer, New York (1992)
10. Chen, D.: Degree of approximation by superpositions of a sigmoidal function. Approx. Theory Appl. **9**(3), 17–28 (1993)
11. Chui, C.K.: An Introduction to Wavelets, Wavelet Analysis and its Applications 1. Academic Press Inc., Boston (1992)
12. Costarelli, D., Spigler, R.: Solving Volterra integral equations of the second kind by sigmoidal functions approximations, to appear in J. Integral Eq. Appl. **25**(2) (2013)
13. Costarelli, D., Spigler, R.: Constructive approximation by superposition of sigmoidal functions. Anal. Theory Appl. **29**(2), 169–196 (2013)
14. Costarelli, D., Spigler, R.: Approximation results for neural network operators activated by sigmoidal functions. Neural Netw. **44**, 101–106 (2013)
15. Costarelli, D., Spigler, R.: Multivariate neural network operators with sigmoidal activation functions. Neural Netw. **48**, 72–77 (2013)
16. Cybenko, G.: Approximation by superpositions of a sigmoidal function. Math. Control Signals Syst. **2**, 303–314 (1989)
17. Daubechies, I.: Ten Lectures on Wavelets, Regional Conference Series in Applied Mathematics 61. Society for Industrial and Applied Mathematics, SIAM, Philadelphia (1992)
18. De Boor, C.: A Practical Guide to Spline, Applied Mathematical Sciences 27. Springer, New York (2001)
19. Gao, B., Xu, Y.: Univariant approximation by superpositions of a sigmoidal function. J. Math. Anal. Appl. **178**, 221–226 (1993)
20. Hahm, N., Hong, B.: Approximation order to a function in $\overline{C}(\mathbb{R})$ by superposition of a sigmoidal function. Appl. Math. Lett. **15**, 591–597 (2002)
21. Hritonenko, N., Yatsenko, Y.: Mathematical Modelling in Economics, Ecology and the Environment. Science Press, Beijing (2006)
22. Jones, L.K.: Constructive approximations for neural networks by sigmoidal functions, Technical Report Series 7. University of Lowell, Dep. of Mathematics (1988)
23. Lenze, B.: Constructive multivariate approximation with sigmoidal functions and applications to neural networks. In: Numer. Methods Approx. Theory, Birkhauser Verlag, Basel-Boston-Berlin, pp. 155–175 (1992)
24. Lewicki, G., Marino, G.: Approximation by superpositions of a sigmoidal function. Z. Anal. Anwendungen J. Anal. Appl. **22**(2), 463–470 (2003)

25. Lewicki, G., Marino, G.: Approximation of functions of finite variation by superpositions of a sigmoidal function. Appl. Math. Lett. **17**, 1147–1152 (2004)
26. Li, X.: Simultaneous approximations of multivariate functions and their derivatives by neural networks with one hidden layer. Neurocomputing **12**, 327–343 (1996)
27. Li, X., Micchelli, C.A.: Approximation by radial bases and neural networks. Numer. Algorithms **25**, 241–262 (2000)
28. Malek, A., Shekari Beidokhti, R.: Numerical solution for high order differential equations using a hybrid neural network—optimization method. Appl. Math. Comput. **183**, 260–271 (2006)
29. Mallat, S.G.: Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. Trans. Am. Math. Soc. **315**, 69–87 (1989)
30. Meyer, Y.: Wavelets and Operators. Cambridge Studies in Advanced Mathematics 37, Cambridge (1992)
31. Mhaskar, H.N., Micchelli, C.A.: Approximation by superposition of sigmoidal and radial basis functions. Adv. Appl. Math. **13**, 350–373 (1992)
32. Mhaskar, H.N., Micchelli, C.A.: Degree of approximation by neural and translation networks with a single hidden layer. Adv. Appl. Math. **16**, 151–183 (1995)
33. Mhaskar, H.N.: Neural networks for optimal approximation of smooth and analytic functions. Neural Comput. **8**, 164–177 (1996)
34. Pinkus, A.: Approximation theory of the MLP model in neural networks. Acta Numer. **8**, 143–195 (1999)
35. Unser, M.: Ten good reasons for using spline wavelets. Wavelets Appl. Signal Image Process. **3169**(5), 422–431 (1997)
36. Xiehua, S.: On the degree of approximation by wavelet expansions. Approx. Theory Appl. **14**(1), 81–90 (1998)