Analysis in Theory and Applications
Anal. Theory Appl., Vol. **29**, No. 2 (2013), pp. 169-196
DOI: 10.4208/ata.2013.v29.n2.8

# Constructive Approximation by Superposition of Sigmoidal Functions

Danilo Costarelli and Renato Spigler*

*Dipartimento di Matematica, Università "Roma Tre", 1, Largo S. Leonardo Murialdo, 00146 Rome, Italy.*

Received 21 February 2012

**Abstract.** In this paper, a constructive theory is developed for approximating functions of one or more variables by superposition of sigmoidal functions. This is done in the uniform norm as well as in the $L^p$ norm. Results for the simultaneous approximation, with the same order of accuracy, of a function and its derivatives (whenever these exist), are obtained. The relation with neural networks and radial basis functions approximations is discussed. Numerical examples are given for the purpose of illustration.

**Key Words**: Sigmoidal functions, multivariate approximation, $L^p$ approximation, neural networks, radial basis functions.

**AMS Subject Classifications**: 41A25, 41A30

## 1 Introduction

In this paper, a constructive theory for approximating functions of one or several variables by superposition of sigmoidal functions is developed. The starting point is a slight modification of H. Chen, T. Chen, and R. Liu's constructive proof [13,14] of G. Cybenko's non-constructive result [16]. Having theorems with constructive proofs is important in order to obtain practical applications of the theory. Our approximating sums (in one dimension, e.g.) take on the form

$$G_N(x) := \sum_{k=0}^{N} \alpha_k \sigma(g_k(x)), \quad x \in \mathbb{R},$$

where $\sigma$ is a bounded sigmoidal function, i.e., $\lim_{x\to-\infty}\sigma(x)=0$ and $\lim_{x\to\infty}\sigma(x)=1$, $\alpha_k \in \mathbb{R}$, and $g_k : \mathbb{R} \to \mathbb{R}$, $k=0,1,\cdots,N$. Here, $g_k(x) := w(x-x_k)$, where $x_k$'s are suitable real

*Corresponding author. *Email addresses:* costarel@mat.uniroma3.it (D. Costarelli), spigler@mat.uniroma3.it (R. Spigler)

 

points and $w$ is a positive scaling parameter depending on $N$. Note that $G_N$ is *not* the partial sum of a (convergent) series. Continuity of $\sigma$ is not required in general. When the sigmoidal function is also continuous, our results can be considered as *density* results.

A noteworthy feature of our theory is that, approximating uniformly a given smooth function, *any* of its *derivatives* can also be approximated with the *same* order of *accuracy*. Moreover, the *same coefficients* in all sums of sigmoidal functions suffice to represent them along with the function itself.

The multivariate theory for the kind of approximation developed here is new, and extends the univariate theory proposed in [13, 14]. Also new is the extension of both, the one-dimensional and the multidimensional theory, to $L^p$ of the Chen-Chen-Liu's (or Cybenko's) approximation. In several dimensions this kind of approximation also differs from the case where the inner products in $\mathbb{R}^n$ link the variables inside the argument of $\sigma$ (see, e.g., [1, 8, 16, 20]). This happens in the neural networks approximation, when the activation functions of the networks are sigmoidal functions. However, ours can also be considered as an approximation method based on radial basis functions neural networks (see Theorem 5.1), [24, 26, 31, 32].

On the face of it, our approximation formula is only $\mathcal{O}(1/N)$ accurate ($N$ denoting the number of the superposed sigmoidal functions in $G_N$). This is due to the fact that the coefficients $\alpha_k$ in $G_N$ look merely like first-order finite differences (see Theorem 2.1). On the other hand, the functions that can be approximated are allowed to be non-smooth.

Here is the plan of the paper. In Section 2, we introduce some preliminary notation and review the main result established by H. Chen, T. Chen, and R. Liu. In Section 3, a constructive approximation theorem in $L^p$ is derived, while in Section 4, simultaneous approximations for a given function (in certain classes) and its derivatives are obtained. The subject of Section 5 is the constructive multivariate approximation, which generalizes the previous theory, valid for functions of one variable. In Sections 6 and 7 a few applications based on specific sigmoidal functions and numerical examples are presented. These are rather simple, but serve the purpose of illustration for our approximation method.

## 2 Notation and preliminary results

In the following, we denote by $C[a,b]$ the set of all continuous functions $f:[a,b]\to\mathbb{R}$ on the bounded closed nonempty interval $[a,b]$, equipped with 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 $f$ such that $f\in C^n(a',b')$ for some open real interval $(a',b')$ such that $[a,b]\subset(a',b')$. Furthermore, $Q\subset\mathbb{R}^2$ will denote the square $Q:=[a,b]\times[c,d]$, with $|b-a|=|d-c|$, and $\|(x,y)\|_2:=(x^2+y^2)^{1/2}$, $(x,y)\in\mathbb{R}^2$, the Euclidean norm in $\mathbb{R}^2$.

Let us recall the definition of a *sigmoidal* function.

**Definition 2.1.** A function $\sigma:\mathbb{R}\to\mathbb{R}$ is called "a sigmoidal function" whenever

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

Sometimes, boundedness, continuity and/or monotonicity may be prescribed in addition.

In general, any continuous function can be approximated uniformly on compact sets, with an arbitrary accuracy, by a linear finite combination of bounded sigmoidal functions like

$$G_N(x) := \sum_{k=0}^{N} \alpha_k \sigma(g_k(x)), \tag{2.1}$$

where $\alpha_k \in \mathbb{R}$, $k=0,1,\cdots,N$, and the $g_k(x)$'s are suitable functions.

The previous result for (2.1) was first established by G. Cybenko [16] for the approximation of continuous multivariate functions on the hypercube $[0,1]^n \subset \mathbb{R}^n$, under the assumption of continuity for the sigmoidal function $\sigma$, where $g_k$'s are of the form

$$g_k(x_1,x_2,\cdots,x_n) = \sum_{i=1}^{n} y_i^k \cdot x_i + \theta_k, \quad (x_1,x_2,\cdots,x_n) \in \mathbb{R}^n,$$

$y_i^k$, $\theta_k \in \mathbb{R}$, $i=1,2,\cdots,n$, $k=0,1,\cdots,N$. These sums arise from an application to the theory of artificial neural networks, the sigmoidal function represents the activation function of the network [20]. The proof given by G. Cybenko of his theorem, however, is nonconstructive.

Now we give a constructive proof of Cybenko's approximation theorem in $C[a,b]$. This is similar to that given in [14], but some little changes have been made in order to let better understand our proof below, in Section 5, concerning the constructive multivariate theory. In particular, we prove such theorem using sums of the form in (2.1), where $g_k(x) := w(x-x_k)$, for some $w>0$ and $x_k \in \mathbb{R}$, $k=0,1,\cdots,N$.

The following statement that we display as a lemma can be obtained as an immediate consequence of Definition 2.1.

**Lemma 2.1.** *Let $x_0,x_1,\cdots,x_N \in \mathbb{R}$, $N \in \mathbb{N}^+$, be fixed. For every $\varepsilon$, $h>0$, there exists $\overline{w} := \overline{w}(\varepsilon,h) > 0$ such that for every $w \geq \overline{w}$ and $k=0,1,\cdots,N$, we have*

1. *$|\sigma(w(x-x_k)) - 1| < \varepsilon$, for every $x \in \mathbb{R}$ such that $x - x_k \geq h$;*
2. *$|\sigma(w(x-x_k))| < \varepsilon$, for every $x \in \mathbb{R}$ such that $x - x_k \leq -h$.*

Now we are able to prove the following:

**Theorem 2.1.** *Let $\sigma$ be a bounded sigmoidal function, and let $f \in C[a,b]$ be fixed. For every $\varepsilon > 0$, there exist $N \in \mathbb{N}^+$ and $w > 0$ (depending on $N$), such that if*

$$(G_N f)(x) := \sum_{k=1}^{N} [f(x_k) - f(x_{k-1})] \sigma(w(x-x_k)) + f(x_0) \sigma(w(x-x_{-1})) \tag{2.2}$$

*for $x \in [a,b]$, $h := (b-a)/N$, and $x_k := a+kh$, $k=-1,0,1,\cdots,N$, then*

$$\|G_N f - f\|_\infty < \varepsilon.$$

We stress that *continuity* of $\sigma$ is *not* required, and its boundedness suffices.

*Proof.* Let $\varepsilon>0$ be fixed. Since $f$ is uniformly continuous, for $\eta:=\varepsilon/(\|f\|_\infty+2\|\sigma\|_\infty+2)$ there exists $\delta>0$ such that $|f(x)-f(y)|<\eta$ for every $x$, $y\in[a,b]$, with $|x-y|<\delta$. We fix $N\in\mathbb{N}^+$, $N>3$, such that $h:=(b-a)/N<\delta/2$ and $1/N<\eta$. Moreover, we fix $w\geq\overline{w}(1/N,h)\equiv\overline{w}(1/N)>0$, where $\overline{w}(1/N)$ is obtained by Lemma 2.1 with $\frac{1}{N}$, $h>0$ and with $x_k=a+hk$, $k=-1,0,1,\cdots,N$. Now, consider $G_Nf$ defined in (2.2) with $w$. Let $x\in[a,b]$ be fixed and for some $i\in\{1,\cdots,N\}$, such that $x\in[x_{i-1},x_i]$. Set

$$L_i(x):=f(a)+[f(x_2)-f(x_1)]\sigma(w(x-x_2))+[f(x_1)-f(x_0)]\sigma(w(x-x_1))$$

for $i=1,2$, and

$$\begin{aligned}L_i(x):=&\sum_{k=1}^{i-2}[f(x_k)-f(x_{k-1})]+f(a)+[f(x_{i-1})-f(x_{i-2})]\sigma(w(x-x_{i-1}))\\&+[f(x_i)-f(x_{i-1})]\sigma(w(x-x_i))\end{aligned}$$

for $i\geq 3$. In any case,

$$|(G_Nf)(x)-f(x)|\leq|(G_Nf)(x)-L_i(x)|+|L_i(x)-f(x)|=:I_1+I_2.$$

We now estimate $I_1$ and $I_2$ only for $i\geq3$, since similar estimates can be obtained also for $i=1,2$. As $x-x_k\geq h$ for $k=-1,0,1,\cdots,i-2$ and $x-x_k\leq -h$ for $k=i+1,\cdots,N$, by conditions 1 and 2 of Lemma 2.1, it follows that

$$\begin{aligned}I_1\leq&\sum_{k=1}^{i-2}|f(x_k)-f(x_{k-1})|\cdot|\sigma(w(x-x_k))-1|+|f(a)|\cdot|\sigma(w(x-x_{-1}))-1|\\&+\sum_{k=i+1}^{N}|f(x_k)-f(x_{k-1})|\cdot|\sigma(w(x-x_k))|\\<&\sum_{k=1}^{N}\eta\frac{1}{N}+\frac{1}{N}|f(a)|\leq(1+\|f\|_\infty)\eta.\end{aligned}$$

It may be observed that here in the above one may use $|f(a)|$ in place of $\|f\|_\infty$. Using the identity

$$\sum_{k=1}^{i-2}[f(x_k)-f(x_{k-1})]+f(a)=f(x_{i-2}),$$

and as $|x_{i-2}-x|\leq|x_{i-2}-x_{i-1}|+|x_{i-1}-x|\leq 2h<\delta$, we can estimate $I_2$:

$$\begin{aligned}I_2=&|f(x_{i-2})+[f(x_{i-1})-f(x_{i-2})]\sigma(w(x-x_{i-1}))\\&+[f(x_i)-f(x_{i-1})]\sigma(w(x-x_i))-f(x)|\\\leq&|f(x_{i-1})-f(x_{i-2})|\cdot|\sigma(w(x-x_{i-1}))|+|f(x_i)-f(x_{i-1})|\cdot|\sigma(w(x-x_i))|\\&+|f(x_{i-2})-f(x)|<(2\|\sigma\|_\infty+1)\eta.\end{aligned}$$

From such estimates for $I_1$ and $I_2$, it follows that

$$|(G_N f)(x)-f(x)|\leq I_1+I_2<(\|f\|_\infty+2\|\sigma\|_\infty+2)\eta=\varepsilon.$$

Therefore, as $x\in[a,b]$ is arbitrary, we conclude that $\|G_N f-f\|_\infty<\varepsilon$. □

Note that when $\sigma$ is continuous, Theorem 2.1 can be viewed as a *density result* in $C[a,b]$ for the set of all functions of the form (2.1) with respect to the uniform norm. In addition, as a consequence of Lemma 2.1, we also observe that $\|G_N f-f\|_\infty<\varepsilon$ for every $w\geq\overline{w}$, where $\overline{w}=\overline{w}(1/N)>0$ is chosen as in Theorem 2.1.

# 3 Constructive approximation in $L^p[a,b]$

Theorem 2.1 is extended also to $L^p$-functions in [16] and then in [14], but by non-constructive methods. Results in $L^p$ were also given in [10,11]. Note that a theory similar to that given above, but developed in the framework of $L^p$ spaces, is only meaningful on real *bounded* intervals, since, in general terms like $\alpha_k\sigma(g_k(x))$ with $\alpha_k\neq 0$ do *not* belong to $L^p(I)$ whenever $I$ is unbounded (for instance, in case $I=\mathbb{R}$). In this section, we give a constructive proof of an approximation theorem for functions $f\in L^p[a,b]$, $1\leq p<\infty$, by linear combinations of sigmoidal functions as in (2.1). The following theorem is trivial, but we prefer to state it explicitly:

**Theorem 3.1.** *Let $\sigma$ be a bounded sigmoidal function, and let $1\leq p<\infty$ be fixed. For every $f\in C[a,b]$ and $\varepsilon>0$, there exist $N\in\mathbb{N}^+$ and $w>0$ (depending on $N$), such that the function $G_N f$ defined in* (2.2) *with $w$, is such that*

$$\|G_N f-f\|_{L^p[a,b]}<\varepsilon.$$

*Proof.* Let $f\in C[a,b]$ and $\varepsilon>0$ be fixed. By Theorem 2.1, for $\eta:=\varepsilon/(b-a)^{1/p}$, there exist $N\in\mathbb{N}^+$, $N>3$, and $w>0$ depending on $N$ such that $\|G_N f-f\|_\infty<\eta$. Therefore,

$$\|G_N f-f\|_{L^p[a,b]}=\left(\int_a^b|(G_N f)(x)-f(x)|^p\,dx\right)^{1/p}<\left(\int_a^b\eta^p\,dx\right)^{1/p}=\varepsilon.$$

The proof is complete. □

Now we can prove the following approximation result in $L^p[a,b]$.

**Theorem 3.2.** *Let $\sigma$ be a bounded sigmoidal function and let $f\in L^p[a,b]$, $1\leq p<\infty$, be fixed. Then, for every $\varepsilon>0$ there exists $N\in\mathbb{N}^+$ and a linear combination $G_N$ of sigmoidal functions as in* (2.1), *such that*

$$\|G_N-f\|_{L^p[a,b]}<\varepsilon.$$

*Proof.* The proof is constructive. Define the function $\widetilde{f}:\mathbb{R}\to\mathbb{R}$ as

$$\widetilde{f}(x):=\begin{cases} f(x), & x\in[a,b],\\ 0, & \text{otherwise}.\end{cases} \tag{3.1}$$

Note that $\widetilde{f}\in L^p(\mathbb{R})$ and $\widetilde{f}=f$ on $[a,b]$. Let $\{\rho_n\}_{n\in\mathbb{N}^+}$, $\rho_n:\mathbb{R}\to\mathbb{R}$, be a sequence of mollifiers, i.e., $\rho_n\in C_c^\infty(\mathbb{R})$, $\operatorname{supp}\rho_n\subseteq[-1/n,1/n]$, $\int_{\mathbb{R}}\rho_n(x)dx=1$, and $\rho_n(x)\geq 0$ for every $x\in\mathbb{R}$, $n\in\mathbb{N}^+$. Define the family $\{f_n\}_{n\in\mathbb{N}^+}$ by

$$f_n(x):=(\rho_n*\widetilde{f})(x)=\int_{\mathbb{R}}\rho_n(x-y)\widetilde{f}(y)dy,\quad x\in\mathbb{R}, \tag{3.2}$$

where $*$ denotes the convolution. By the general property of sequences of mollifiers and of convolution products [5], it turns out that $f_n=\rho_n*\widetilde{f}\in C(\mathbb{R})$ for every $n\in\mathbb{N}^+$, and $f_n\to\widetilde{f}$ in $L^p(\mathbb{R})$ as $n\to\infty$. Let $\varepsilon>0$ be fixed. Then, there exists $\overline{n}\in\mathbb{N}^+$ such that

$$\|f_n-f\|_{L^p[a,b]}=\left\|f_n-\widetilde{f}\right\|_{L^p[a,b]}\leq\left\|f_n-\widetilde{f}\right\|_{L^p(\mathbb{R})}<\frac{\varepsilon}{2},$$

for every $n\geq\overline{n}$. Let now $n\geq\overline{n}$ be fixed. Since $f_n\in C(\mathbb{R})\subset C[a,b]$, as a consequence of Theorem 3.1, for $\varepsilon/2$ there exist $N\in\mathbb{N}^+$, $N>3$, and $w>0$ (depending on $N$), such that the function $G_Nf_n$ defined in (2.2) is such that

$$\|G_Nf_n-f_n\|_{L^p[a,b]}=\left\|G_N(\rho_n*\widetilde{f})-(\rho_n*\widetilde{f})\right\|_{L^p[a,b]}<\frac{\varepsilon}{2}.$$

Hence, we can conclude that

$$\|G_Nf_n-f\|_{L^p[a,b]}\leq\|G_Nf_n-f_n\|_{L^p[a,b]}+\|f_n-f\|_{L^p[a,b]}<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$

Setting $G_N(x):=(G_Nf_n)(x)$ completes the proof. □

We stress that the proof of Theorem 3.2 is *constructive*, and by choosing a specific sequence of mollifiers a precise analytic form of the sums $G_Nf_n$ approximating $f\in L^p[a,b]$ ($f_n$ defined in (3.2)) would be obtained.

One could construct a sequence of mollifiers starting from a single function $\rho\in C_c^\infty(\mathbb{R})$ such that $\operatorname{supp}\rho\subset[-1,1]$ and $\rho\geq 0$ on $\mathbb{R}$. For instance, choosing

$$\rho(x):=\begin{cases} e^{1/(|x|^2-1)}, & |x|<1,\\ 0, & |x|\geq 1,\end{cases} \tag{3.3}$$

we obtain a sequence of mollifiers by setting

$$\rho_n(x):=Cn\rho(nx),\quad x\in\mathbb{R}, \tag{3.4}$$

where $C := (\int_{\mathbb{R}} \rho(x)\,dx)^{-1}$ [5,24]. In this case, the approximating sums take the form

$$G_N(\rho_n * \widetilde{f})(x) = \sum_{k=1}^{N} \alpha_k \sigma(w(x-x_k)) + \alpha_0 \sigma(w(x-x_{-1})), \quad x \in [a,b],$$

where $N$, $w>0$, $x_k = a+kh$, $k=-1,0,1,\cdots,N$, for $h=(b-a)/N$, and the coefficients are

$$\alpha_k = \int_{\mathbb{R}} \rho_n(x_k - y)\widetilde{f}(y)\,dy - \int_{\mathbb{R}} \rho_n(x_{k-1}-y)\widetilde{f}(y)\,dy,$$

for $k=1,\cdots,N$, while $\alpha_0 = \int_{\mathbb{R}} \rho_n(a-y)\widetilde{f}(y)\,dy$ for $n$ sufficiently large, $\widetilde{f}$ being defined in (3.1).

# 4 Simultaneous approximation of functions and their derivatives

The theory presented in Section 2 allows to approximate any given function $f \in C[a,b]$ by means of functions like $G_N f$ defined in (2.2), and hence if $f$ is sufficiently smooth, say e.g., $f \in \widehat{C}^1[a,b]$, then its derivative $f'$ could also be approximated by means of $G_N f'$. In particular, we can choose the same number $N \in \mathbb{N}^+$ to construct simultaneous approximations of $f$ and $f'$, using $G_N f$ and $G_N f'$. In this way, we can also choose the same value for $w>0$. However, $2(N+1)$ coefficients, namely the values $f(x_k)$ and $f'(x_k)$ for $k=0,1,\cdots,N$, will be needed in this case.

Similarly, we could construct simultaneous approximations to $f, f', \cdots, f^{(n)}$ by means of approximants like $G_N f, G_N f', \cdots, G_N f^{(n)}$, respectively, for every given $f \in \widehat{C}^n[a,b]$ requiring $(n+1)(N+1)$ coefficients.

In view of a number of applications, it could be useful to approximate $f$ *along with* some of its derivatives by superposing sigmoidal functions but using *only* the $N+1$ coefficients involved in the approximation formula for $f$.

**Remark 4.1.** Let $f \in \widehat{C}^1[a,b]$ be fixed and $\sigma \in C^1(\mathbb{R})$ be a bounded sigmoidal function. Moreover, let $\varepsilon>0$ such that $\|G_N f - f\|_\infty < \varepsilon$ for suitable values $N$ and $w>0$. Consider $(G_N f)'$, the first derivative of $G_N f$, which is of the form

$$(G_N f)'(x) = \sum_{k=1}^{N} w\,[f(x_k) - f(x_{k-1})]\,\sigma'(w(x-x_k)) + w f(x_0)\sigma'(w(x-x_{-1})) \tag{4.1}$$

for $x \in [a,b]$. First of all, note that in general $(G_N f)'$ is *not* a sum of sigmoidal functions, but its coefficients, which are the same appearing in the sum $G_N f$, multiplied by the scaling term $w>0$, are known. Besides, $\sigma'(w(x-x_k))$, $k=-1,1,\cdots,N$, are also known. Hence, $f'$ could be uniformly approximated by $(G_N f)'$, but in general the condition $\|(G_N f)' - f'\|_\infty < \varepsilon$ will not be fulfilled. A similar problem would arise if we try to approximate $f'$ by a sum of sigmoidal functions like $G_M[(G_N f)']$.

Consider for instance $f(x)=x$ on $[0,1]$ and $\sigma(x)=(1+e^{-x})^{-1}$. Then, $f'(x)=1$ and $\sigma'(x)=e^{-x}/(1+e^{-x})^2$, while the sum $G_N f$ takes the form

$$(G_N f)(x)=h\sum_{k=1}^{N}\sigma(w_0(x-x_k)),\quad x\in[0,1],$$

where $h=1/N$ and $x_k=k/N$, $k=1,\cdots,N$. Assume that $\|G_N f-f\|_\infty<\varepsilon$ for $\varepsilon>0$, $N\in\mathbb{N}^+$, and a fixed value of $w_0>0$ depending on $N$. As noted in Section 2, we also have $\|G_N f-f\|_\infty<\varepsilon$ if we replace $w_0$ in $G_N f$ by $w\geq w_0$. Moreover, consider

$$(G_N f)'(x)=wh\sum_{k=1}^{N}\sigma'(w(x-x_k)),\quad x\in[0,1].$$

Let now $\overline{x}\in[0,1]$ be fixed. Then,

$$(G_N f)'(\overline{x})=wh\sum_{k=1}^{N}\sigma'(w(\overline{x}-x_k))=:wC,$$

where $C=C(w)$ is such that $0<C\leq 1$ for every $w>0$. Since $\lim_{w\to+\infty}(wC-1)=+\infty$, a condition like

$$\left|(G_N f)'(\overline{x})-f'(\overline{x})\right|=|wC-1|<\varepsilon$$

cannot be satisfied, in general, for $w>0$ sufficiently large.

We now introduce some notation. For any fixed $f\in C[a,b]$, and for any given uniform partition $\{x_0,x_1,\cdots,x_N\}$ of the interval $[a,b]$, with $x_0=a$, $x_N=b$ and $h=x_k-x_{k-1}$, $k=1,\cdots,N$, we define

$$\Delta_k^j f:=\frac{1}{h^j}\sum_{\nu=0}^{j}\binom{j}{\nu}(-1)^\nu f(x_{k+j-\nu}),\tag{4.2}$$

for $j\in\mathbb{N}$, $j\leq N$, and $k=0,1,\cdots,N-j$. We can establish the following:

**Theorem 4.1** (Simultaneous approximation of $f$ and its derivatives)**.** *Let $\sigma$ be a bounded sigmoidal function and let $f\in\widehat{C}^{n+1}[a,b]$, $n\in\mathbb{N}^+$, be fixed. For every $\varepsilon>0$, there exist $N\in\mathbb{N}^+$ and $w>0$ (depending on $N$), such that for every $j=1,\cdots,n$, by defining*

$$(G_N^j f)(x):=\sum_{k=1}^{N-j}\left(\Delta_k^j f-\Delta_{k-1}^j f\right)\sigma(w(x-x_k))+\Delta_0^j f\cdot\sigma(w(x-x_{-1})),\tag{4.3}$$

*for $x\in[a,b]$, $h:=(b-a)/N$ and $x_k:=a+kh$, $k=-1,0,1,\cdots,N$, we have*

$$\left\|G_N^j f-f^{(j)}\right\|_\infty<\varepsilon.$$

*Proof.* Let $j=1,\cdots,n$ and $\varepsilon>0$ be fixed. Since $f^{(j)}\in\widehat{C}^1[a,b]$ is uniformly continuous, for

$$\eta:=\Big(4C_j\|\sigma\|_\infty+2\|\sigma\|_\infty+3C_j+\Big\|f^{(j)}\Big\|_\infty+2\Big)^{-1}\varepsilon,$$

where $C_j=C_j(f^{(j+1)},a,b)$ is a fixed constant that will be determined later, there exists $\delta>0$ such that for every $x$, $y\in[a,b]$ with $|x-y|<\delta$, we have $\big|f^{(j)}(x)-f^{(j)}(y)\big|<\eta$. We choose $N\in\mathbb{N}^+$, $N>j+3$ sufficiently large so that $h:=(b-a)/N<\delta/\max\{2,j\}$ and $1/N<\eta$. Moreover, we fix

$$w\geq\overline{w}(1/N,h)\equiv\overline{w}(1/N)>0,$$

where $\overline{w}(1/N)$ is obtained by using Lemma 2.1 with $1/N$, $h>0$ and with the points $x_k=a+hk$, $k=-1,0,1,\cdots,N$. Now consider the sum $G^j_N f$ defined in (4.3) with $w$, and let $x\in[a,b]$ be fixed. Then, there exists $i=1,\cdots,N$ such that $x\in[x_{i-1},x_i]$. Set

$$L_i(x):=\Delta^j_0 f+\Big(\Delta^j_2 f-\Delta^j_1 f\Big)\sigma(w(x-x_2))+\Big(\Delta^j_1 f-\Delta^j_0 f\Big)\sigma(w(x-x_1))$$

for $i=1,2$,

$$\begin{aligned}L_i(x):=&\sum_{k=1}^{i-2}\Big(\Delta^j_k f-\Delta^j_{k-1}f\Big)+\Delta^j_0 f+\Big(\Delta^j_{i-1}f-\Delta^j_{i-2}f\Big)\sigma(w(x-x_{i-1}))\\&+\Big(\Delta^j_i f-\Delta^j_{i-1}f\Big)\sigma(w(x-x_i))\end{aligned}$$

for $i=3,\cdots,N-j$,

$$L_i(x):=\sum_{k=1}^{N-j-1}\Big(\Delta^j_k f-\Delta^j_{k-1}f\Big)+\Delta^j_0 f+\Big(\Delta^j_{N-j}f-\Delta^j_{N-j-1}f\Big)\sigma(w(x-x_{N-j}))$$

for $i=N-j+1$, and

$$L_i(x):=\sum_{k=1}^{N-j}\Big(\Delta^j_k f-\Delta^j_{k-1}f\Big)+\Delta^j_0 f$$

for $i=N-j+2,\cdots,N$. We now write

$$\Big|(G^j_N f)(x)-f^{(j)}(x)\Big|\leq\Big|(G^j_N f)(x)-L_i(x)\Big|+\Big|L_i(x)-f^{(j)}(x)\Big|=:J_1+J_2$$

and start to estimate $J_1$. We confine only to $i=3,\cdots,N-j$, since the other cases can be obtained similarly.

As $x-x_k \geq h$ for $k=-1,0,1,\cdots,i-2$ and $x-x_k \leq -h$ for $k=i+1,\cdots,N-j$, it follows by the conditions 1 and 2 of Lemma 2.1 that

$$\begin{aligned}
J_1 \leq& \sum_{k=1}^{i-2} \left|\Delta_k^j f - \Delta_{k-1}^j f\right| |\sigma(w(x-x_k))-1| + \left|\Delta_0^j f\right| |\sigma(w(x-x_{-1}))-1| \\
&+ \sum_{k=i+1}^{N-j} \left|\Delta_k^j f - \Delta_{k-1}^j f\right| |\sigma(w(x-x_k))| \\
<& \frac{1}{N}\sum_{k=1}^{N-j} \left|\Delta_k^j f - \Delta_{k-1}^j f\right| + \frac{1}{N}\left|\Delta_0^j f\right| \\
\leq& \frac{1}{N}\sum_{k=1}^{N-j} \left|\Delta_k^j f - f^{(j)}(x_k)\right| + \frac{1}{N}\sum_{k=1}^{N-j}\left|f^{(j)}(x_k)-f^{(j)}(x_{k-1})\right| \\
&+ \frac{1}{N}\sum_{k=1}^{N-j}\left|f^{(j)}(x_{k-1}) - \Delta_{k-1}^j f\right| + \frac{1}{N}\left|\Delta_0^j f - f^{(j)}(x_0)\right| + \frac{1}{N}\left|f^{(j)}(x_0)\right|.
\end{aligned}$$

We now observe that, for every $k=0,1,\cdots,N-j$, the terms $\Delta_k^j f$ provide an approximation to $f^{(j)}(x_k)$, obtained by forward finite differences. It is well known that there exists a positive constant, $\widetilde{C}_j>0$, depending only on $f^{(j+1)}$, such that

$$\left|\Delta_k^j f - f^{(j)}(x_k)\right| \leq \widetilde{C}_j h = \widetilde{C}_j \frac{(b-a)}{N} =: \frac{C_j}{N}, \tag{4.4}$$

for every $k=0,1,\cdots,N-j$, where $C_j=C_j(f^{(j+1)},a,b)$. Then, using (4.4) the uniform continuity of $f^{(j)}$ and the previous inequality for $J_1$, we obtain

$$J_1 \leq 2\sum_{k=1}^{N-j+1}\frac{C_j}{N^2} + \frac{1}{N}\sum_{k=1}^{N-j}\eta + \frac{1}{N}\left\|f^{(j)}\right\|_\infty < 2C_j\eta + \eta + \left\|f^{(j)}\right\|_\infty \eta.$$

Finally, we estimate $J_2$ separately in four cases.

*Case* 1: $i=1,2$. Being $|x_0-x| \leq 2h \leq \max\{2,j\}\, h < \delta$, we have

$$\begin{aligned}
J_2 \leq& \left(\left|\Delta_2^j f - f^{(j)}(x_2)\right| + \left|f^{(j)}(x_2)-f^{(j)}(x_1)\right| + \left|f^{(j)}(x_1)-\Delta_1^j f\right|\right)\|\sigma\|_\infty \\
&+ \left(\left|\Delta_1^j f - f^{(j)}(x_1)\right| + \left|f^{(j)}(x_1)-f^{(j)}(x_0)\right| + \left|f^{(j)}(x_0)-\Delta_0^j f\right|\right)\|\sigma\|_\infty \\
&+ \left|\Delta_0^j f - f^{(j)}(x_0)\right| + \left|f^{(j)}(x_0)-f^{(j)}(x)\right| \\
\leq& \left(4C_j\eta + 2\eta\right)\|\sigma\|_\infty + C_j\eta + \eta.
\end{aligned}$$

*Case* 2: $i=3,\cdots,N-j$. We obtain as above

$$J_2 < \left(4C_j\eta + 2\eta\right)\|\sigma\|_\infty + C_j\eta + \eta.$$

*Case* 3: $i=N-j+1$.

$$\begin{aligned}J_2 \le &\Big(\Big|\Delta^j_{N-j}f - f^{(j)}(x_{N-j})\Big| + \Big|f^{(j)}(x_{N-j}) - f^{(j)}(x_{N-j-1})\Big| \\ &+ \Big|f^{(j)}(x_{N-j-1}) - \Delta^j_{N-j-1}f\Big|\Big)\,\|\sigma\|_\infty + \Big|\Delta^j_{N-j-1}f - f^{(j)}(x_{N-j-1})\Big| \\ &+ \Big|f^{(j)}(x_{N-j-1}) - f^{(j)}(x)\Big| \\ < &\,(2C_j\eta+\eta)\,\|\sigma\|_\infty + C_j\eta + \eta.\end{aligned}$$

*Case* 4: $i=N-j+2,\cdots,N-j$.

$$J_2 \le \Big|\Delta^j_{N-j}f - f^{(j)}(x_{N-j})\Big| + \Big|f^{(j)}(x_{N-j}) - f^{(j)}(x)\Big| \le C_j\eta + \eta,$$

since $|x_{N-j}-x| \le jh \le \max\{2,j\}\,h < \delta$. Thus,

$$\begin{aligned}\Big|(G^j_N f)(x) - f^{(j)}(x)\Big| &\le J_1 + J_2 \\ &< 2C_j\eta + \eta + \Big\|f^{(j)}\Big\|_\infty \eta \; + C_j\eta + \eta + (4C_j\eta + 2\eta)\,\|\sigma\|_\infty \\ &= \Big(4C_j\|\sigma\|_\infty + 2\|\sigma\|_\infty + 3C_j + \Big\|f^{(j)}\Big\|_\infty + 2\Big)\,\eta = \varepsilon.\end{aligned}$$

In particular, we have $|(G^j_N f)(x) - f^{(j)}(x)| < \varepsilon$ for every $x\in[a,b]$, hence $\|G^j_N f - f^{(j)}\|_\infty < \varepsilon$. This completes the proof. □

For example, if we want to approximate $f'$, as $f\in\widehat{C}^2[a,b]$, by a superposition of sigmoidal functions with the same coefficients used for approximating $f$, we can consider the sum $G^1_N f$, which takes the form

$$\begin{aligned}(G^1_N f)(x) = &\sum_{k=1}^{N-1} \frac{f(x_{k+1}) - 2f(x_k) + f(x_{k-1})}{h}\sigma(w(x-x_k)) \\ &+ \frac{f(x_1)-f(x_0)}{h}\sigma(w(x-x_{-1})), \quad x\in\mathbb{R}.\end{aligned}$$

At this point, we want to estimate the *error* made approximating a given function as well as its derivatives, by a superposition of sigmoidal functions.

Concerning the order of approximation for Lipschitz continuous functions and functions of bounded variation, see [6, 7, 9, 12, 17, 18, 22, 23]. There, it was proved that the error of approximating $f$ by linear combinations of $N$ sigmoidal functions is of order $\mathcal{O}(1/N)$ (for $N$ sufficiently large). Here we can prove that *the same* error estimate can be established for approximating the $j$th derivative of $f$ with sums of the form $G^j_N f$. In fact,

**Theorem 4.2.** *Let $\sigma$ be a bounded sigmoidal function, $f\in\widehat{C}^{n+1}[a,b]$, $n\in\mathbb{N}^+$, and $j=1,\cdots,n$ be fixed. For every $N\in\mathbb{N}^+$, $N>j+3$, there exists $\overline{w}>0$ (depending on $N$) such that for every $w\geq\overline{w}$ and $G_N^j f$ defined in* (4.3) *with $w$, we have*

$$\left\|G_N^j f-f^{(j)}\right\|_\infty<\frac{1}{N}\Big[L_j(b-a)\left(2\|\sigma\|_\infty+1+\max\{2,j\}\right)+\widetilde{C}_j(b-a)\left(4\|\sigma\|_\infty+3\right)+\left\|f^{(j)}\right\|_\infty\Big],$$

*where $\widetilde{C}_j>0$ is a constant depending only on $f^{(j+1)}$ and $L_j>0$ is the Lipschitz constant for $f^{(j)}$.*

*Proof.* Let $j=1,\cdots,n$ and $N\in\mathbb{N}^+$, $N>j+3$, be fixed. Set $h:=(b-a)/N$ and $x_k:=a+hk$, for $k=-1,0,1,\cdots,N$. Moreover, let $\overline{w}=\overline{w}(1/N,h)=\overline{w}(1/N)>0$ as shown in Lemma 2.1 with $1/N$, $h>0$ and with the points $x_k=a+hk$, $k=-1,0,1,\cdots,N$. Consider now $G_N^j f$ defined in (4.3) for $w\geq\overline{w}$, and let $x\in[a,b]$ be fixed. Then, there exists $i\in\{1,\cdots,N\}$ such that $x\in[x_{i-1},x_i]$. We set as in Theorem 4.1,

$$L_i(x):=\Delta_0^j f+\left(\Delta_2^j f-\Delta_1^j f\right)\sigma(w(x-x_2))+\left(\Delta_1^j f-\Delta_0^j f\right)\sigma(w(x-x_1))$$

for $i=1,2$,

$$L_i(x):=\sum_{k=1}^{i-2}\left(\Delta_k^j f-\Delta_{k-1}^j f\right)+\Delta_0^j f+\left(\Delta_{i-1}^j f-\Delta_{i-2}^j f\right)\sigma(w(x-x_{i-1}))+\left(\Delta_i^j f-\Delta_{i-1}^j f\right)\sigma(w(x-x_i))$$

for $i=3,\cdots,N-j$,

$$L_i(x):=\sum_{k=1}^{N-j-1}\left(\Delta_k^j f-\Delta_{k-1}^j f\right)+\Delta_0^j f+\left(\Delta_{N-j}^j f-\Delta_{N-j-1}^j f\right)\sigma(w(x-x_{N-j}))$$

for $i=N-j+1$, and

$$L_i(x):=\sum_{k=1}^{N-j}\left(\Delta_k^j f-\Delta_{k-1}^j f\right)+\Delta_0^j f$$

for $i=N-j+2,\cdots,N$. We then write

$$\left|(G_N^j f)(x)-f^{(j)}(x)\right|\leq\left|(G_N^j f)(x)-L_i(x)\right|+\left|L_i(x)-f^{(j)}(x)\right|=:J_1+J_2$$

and estimate $J_1$ (only for $i=3,\cdots,N-j$, the other cases are similar), by using the same argument as that in Theorem 4.1. We obtain

$$J_1<\frac{1}{N}\sum_{k=1}^{N-j}\left|\Delta_k^j f-f^{(j)}(x_k)\right|+\frac{1}{N}\sum_{k=1}^{N-j}\left|f^{(j)}(x_k)-f^{(j)}(x_{k-1})\right|+\frac{1}{N}\sum_{k=1}^{N-j}\left|f^{(j)}(x_{k-1})-\Delta_{k-1}^j f\right|+\frac{1}{N}\left|\Delta_0^j f-f^{(j)}(x_0)\right|+\frac{1}{N}\left|f^{(j)}(x_0)\right|.$$

By the inequality (4.4) and as $f^{(j)}\in\widehat{C}^1[a,b]$ is Lipschitz continuous, say, $L_j>0$ is the Lipschitz constant for $f^{(j)}$, we have

$$J_1<\frac{1}{N}\left(2\widetilde{C}_j(b-a)+L_j(b-a)+\left\|f^{(j)}\right\|_\infty\right),$$

where $\widetilde{C}_j=\widetilde{C}_j(f^{(j+1)})>0$. We now estimate $J_2$ in four different cases.

*Case* 1: $i=1,2$.

$$\begin{aligned}J_2\le&\left(\left|\Delta_2^j f-f^{(j)}(x_2)\right|+\left|f^{(j)}(x_2)-f^{(j)}(x_1)\right|+\left|f^{(j)}(x_1)-\Delta_1^j f\right|\right)\|\sigma\|_\infty\\&+\left(\left|\Delta_1^j f-f^{(j)}(x_1)\right|+\left|f^{(j)}(x_1)-f^{(j)}(x_0)\right|+\left|f^{(j)}(x_0)-\Delta_0^j f\right|\right)\|\sigma\|_\infty\\&+\left|\Delta_0^j f-f^{(j)}(x_0)\right|+\left|f^{(j)}(x_0)-f^{(j)}(x)\right|\\\le&\left(4\frac{\widetilde{C}_j(b-a)}{N}+2L_j\frac{(b-a)}{N}\right)\|\sigma\|_\infty+\frac{\widetilde{C}_j(b-a)}{N}+2L_j\frac{(b-a)}{N}.\end{aligned}$$

*Case* 2: $i=3,\cdots,N-j$. Proceeding as above,

$$J_2\le\left(4\frac{\widetilde{C}_j(b-a)}{N}+2L_j\frac{(b-a)}{N}\right)\|\sigma\|_\infty+\frac{\widetilde{C}_j(b-a)}{N}+2L_j\frac{(b-a)}{N}.$$

*Case* 3: $i=N-j+1$. Similarly,

$$J_2\le\left(2\frac{\widetilde{C}_j(b-a)}{N}+L_j\frac{(b-a)}{N}\right)\|\sigma\|_\infty+\frac{\widetilde{C}_j(b-a)}{N}+2L_j\frac{(b-a)}{N}.$$

*Case* 4: $i=N-j+2,\cdots,N-j$.

$$\begin{aligned}J_2&\le\left|\Delta_{N-j}^j f-f^{(j)}(x_{N-j})\right|+\left|f^{(j)}(x_{N-j})-f^{(j)}(x)\right|\\&\le\frac{\widetilde{C}_j(b-a)}{N}+L_j|x_{N-j}-x|\\&\le\frac{\widetilde{C}_j(b-a)}{N}+jL_j\frac{(b-a)}{N}.\end{aligned}$$

Then,

$$\begin{aligned}\left|(G_N^j f)(x)-f^{(j)}(x)\right|\le J_1+J_2<\frac{1}{N}\Big[&L_j(b-a)(2\|\sigma\|_\infty+1+\max\{2,j\})\\&+\widetilde{C}_j(b-a)(4\|\sigma\|_\infty+3)+\left\|f^{(j)}\right\|_\infty\Big].\end{aligned}$$

In particular, the estimates above hold for every $x \in [a,b]$. Therefore,

$$\left\| G_N^j f - f^{(j)} \right\|_\infty < \frac{1}{N} \Big[ L_j (b-a)(2\|\sigma\|_\infty + 1 + \max\{2,j\}) \\ + \widetilde{C}_j (b-a)(4\|\sigma\|_\infty + 3) + \left\| f^{(j)} \right\|_\infty \Big].$$

The proof is complete. □

For further results concerning the order of approximation, see [3,28].

## 5 Constructive multivariate approximation

In this section, we propose a multivariate extension of the constructive theory developed in Sections 2 and 3. For simplicity, the proof will be given only for functions of two variables, as for higher dimensions the extension is straightforward.

We first establish the following lemma, obtained as an easy consequence of Definition 2.1. This represents a generalization of Lemma 2.1.

**Lemma 5.1.** *Let $(x_0,y_0),(x_1,y_1),\cdots,(x_N,y_N) \in \mathbb{R}^2$, for some fixed $N \in \mathbb{N}^+$. For every $\varepsilon$ and $h>0$, there exists $\overline{w} := \overline{w}(\varepsilon,h) > 0$ such that for every $w \geq \overline{w}$ and $k=0,1,\cdots,N$, we have*

1. $|\sigma(w\|(x,y)-(x_k,y_k)\|_2) - 1| < \varepsilon$;
2. $|\sigma(-w\|(x,y)-(x_k,y_k)\|_2)| < \varepsilon$,

*for every $(x,y) \in \mathbb{R}^2$ such that $\|(x,y)-(x_k,y_k)\|_2 \geq h$.*

Note that here the function $\sigma(\|(x,y)\|_2)$ is actually a *radial basis* function (RBF). We can now prove the following theorem, where we stress that no continuity assumption on $\sigma$ is made.

**Theorem 5.1.** *Let $\sigma$ be a bounded sigmoidal function, and let $f \in C(Q)$, where $Q := [a,b] \times [c,d] \subset \mathbb{R}^2$ with $b-a=d-c=l$ fixed. For every $\varepsilon > 0$, there exist $N \in \mathbb{N}^+$ and $w > 0$ (depending on $N$) such that if*

$$(\widetilde{G}_N f)(x,y) := \sum_{i=1}^N \sum_{j=1}^N [f(x_i,y_j) - f(x_i,y_{j-1})]\, \sigma\Big( w\chi_{ij}(x,y) \left\| (x,y) - (t_{x_i}, t_{y_j}) \right\|_2 \Big) \\ + \sum_{i=1}^N f(x_i,y_0)\, \sigma\big( w\chi_{i0}(x,y) \, \|(x,y) - (t_{x_i}, t_{y_0})\|_2 \big), \tag{5.1}$$

*where $(x,y) \in Q$, $h := l/N$, $x_i := a+hi$, $y_j := c+hj$, for $i,j = -1,0,1,\cdots,N$ and $t_{x_i} := (x_{i-1}+x_i)/2$, $t_{y_j} := (y_{j-1}+y_j)/2$ for $i,j=0,1,\cdots,N$, and moreover*

$$\chi_{ij}(x,y) := \begin{cases} +1, & \text{if } x \in (x_{i-1}, x_i] \text{ and } y \geq y_j, \\ -1, & \text{otherwise}, \end{cases}$$

*for* $i=2,\cdots,N$ *and* $j=0,\cdots,N$*, while*

$$\chi_{1j}(x,y):=\begin{cases} +1, & \text{if } x\in[x_0,x_1] \text{ and } y\geq y_j,\\ -1, & \text{otherwise,}\end{cases}$$

*for* $j=0,\cdots,N$*, then*

$$\left\|\widetilde{G}_N f-f\right\|_\infty<\varepsilon.$$

*Proof.* Let $\varepsilon>0$ be fixed. Since $f$ is uniformly continuous, for $\eta:=\varepsilon/(\|f\|_\infty+\|\sigma\|_\infty+2)$, there exists $\delta>0$ such that $|f(x,y)-f(z,t)|<\eta$ for every $(x,y)$, $(z,t)\in Q$ with $\|(x,y)-(z,t)\|_2<\delta$. Now, choose $N\in\mathbb{N}^+$ such that $h:=l/N<\delta/\sqrt{2}$ and $1/N<\eta$. Moreover, fix $w\geq\overline{w}(1/N^2,h/2)=\overline{w}(1/N^2)>0$, where $\overline{w}(1/N^2)$ is obtained from Lemma 5.1 with $1/N^2$, $h/2>0$ and with the points $(t_{x_i},t_{y_j})$, $i,j=0,1,\cdots,N$. Consider $\widetilde{G}_N f$ defined in (5.1) for $w$, and let $(x,y)\in Q$ be fixed. Thus, there exist $k,\mu=1,\cdots,N$, such that $(x,y)\in(x_{k-1},x_k]\times[y_{\mu-1},y_\mu]$ provided $k\geq 2$ or $(x,y)\in[x_0,x_1]\times[y_{\mu-1},y_\mu]$ otherwise. Set

$$\mathcal{L}_{k\mu}(x,y):=f(x_k,y_0)+[f(x_k,y_1)-f(x_k,y_0)]\,\sigma(w\chi_{k\mu}(x,y)\,\|(x,y)-(t_{x_k},t_{y_1})\|_2),$$

if $\mu=1$, and

$$\begin{aligned}\mathcal{L}_{k\mu}(x,y):=&\sum_{j=1}^{\mu-1}\left[f(x_k,y_j)-f(x_k,y_{j-1})\right]+f(x_k,y_0)\\ &+\left[f(x_k,y_\mu)-f(x_k,y_{\mu-1})\right]\sigma(w\chi_{k\mu}(x,y)\,\|(x,y)-(t_{x_k},t_{y_\mu})\|_2),\end{aligned}$$

if $\mu\geq 2$. In both cases, let write

$$\begin{aligned}&\left|(\widetilde{G}_N f)(x,y)-f(x,y)\right|\\ \leq&\left|(\widetilde{G}_N f)(x,y)-\mathcal{L}_{k\mu}(x,y)\right|+\left|\mathcal{L}_{k\mu}(x,y)-f(x,y)\right|=:H_1+H_2.\end{aligned}$$

We first estimate $H_1$ when $\mu\geq 2$. We have

$$\begin{aligned}H_1\leq&\sum_{\substack{i=1\\ i\neq k}}^{N}\sum_{j=1}^{N}|f(x_i,y_j)-f(x_i,y_{j-1})|\left|\sigma(w\chi_{ij}(x,y)\left\|(x,y)-(t_{x_i},t_{y_j})\right\|_2)\right|\\ &+\sum_{\substack{i=1\\ i\neq k}}^{N}|f(x_i,y_0)|\left|\sigma(w\chi_{i0}(x,y)\,\|(x,y)-(t_{x_i},t_{y_0})\|_2)\right|\\ &+\sum_{j=1}^{\mu-1}\left|f(x_k,y_j)-f(x_k,y_{j-1})\right|\left|\sigma(w\chi_{kj}(x,y)\left\|(x,y)-(t_{x_k},t_{y_j})\right\|_2)-1\right|\\ &+\sum_{j=\mu+1}^{N}\left|f(x_k,y_j)-f(x_k,y_{j-1})\right|\left|\sigma(w\chi_{kj}(x,y)\left\|(x,y)-(t_{x_k},t_{y_j})\right\|_2)\right|\\ &+|f(x_k,y_0)|\left|\sigma(w\chi_{k0}(x,y)\,\|(x,y)-(t_{x_k},t_{y_0})\|_2)-1\right|.\end{aligned}$$

As $\|(x_i,y_j)-(x_i,y_{j-1})\|_2<\delta$ for every $i,j=1,\cdots,N$, $\|(x,y)-(t_{x_i},t_{y_j})\|_2\geq h/2$ for every $(t_{x_i},t_{y_j})\neq(t_{x_k},t_{y_\mu})$, by the conditions 1 and 2 of Lemma 5.1 and the definition of $\chi_{ij}$, we obtain

$$\begin{aligned}H_1&<\frac{1}{N^2}\sum_{\substack{i=1\\i\neq k}}^{N}\sum_{j=1}^{N}\eta+\frac{1}{N^2}\sum_{\substack{i=1\\i\neq k}}^{N}\|f\|_\infty+\frac{1}{N^2}\sum_{j=1}^{\mu-1}\eta+\frac{1}{N^2}\sum_{j=\mu+1}^{N}\eta+\frac{1}{N^2}\|f\|_\infty\\&\leq\eta+\frac{1}{N}\|f\|_\infty<(1+\|f\|_\infty)\eta.\end{aligned}$$

Note that the same estimate for $H_1$ also holds if $\mu=1$. Finally, if we note that

$$\sum_{j=1}^{\mu-1}\left[f(x_k,y_j)-f(x_k,y_{j-1})\right]+f(x_k,y_0)=f(x_k,y_{\mu-1})$$

and that $\|(x,y)-(x_k,y_{\mu-1})\|_2\leq\sqrt{2}h<\delta$, we obtain

$$H_2<\left|f(x_k,y_{\mu-1})-f(x,y)\right|+\eta\,\|\sigma\|_\infty=(\|\sigma\|_\infty+1)\eta, \tag{5.2}$$

((5.2) holds also in case $\mu=1$). We conclude that

$$\left|(\widetilde{G}_Nf)(x,y)-f(x,y)\right|\leq H_1+H_2<(\|\sigma\|_\infty+\|f\|_\infty+2)\eta=\varepsilon,$$

and since $(x,y)\in Q$ is arbitrary, it follows that $\|\widetilde{G}_Nf-f\|_\infty<\varepsilon$. □

**Remark 5.1.** We remark that the results in Theorem 5.1 differ substantially from those (apparently similar) established by G. Cybenko in [16] and by B. Lenze in [21]. In particular, others than Cybenko's theory, ours is *constructive*, while Lenze's result is a different argument of the sigmoidal functions, aiming at describing some special kind of neural networks. Our results lead to *RBF neural networks* [25,26,29–32] (see $(\widetilde{G}_Nf)(x,y)$ in (5.1)), and reduce, essentially, to the one-dimensional case of Section 2 (where the nodes $x_k$ should be replaced by the midpoints of the $k$th subinterval, as chosen in [13,14]).

**Remark 5.2.** Theorem 5.1 holds also when instead of $\widetilde{G}_Nf$, $N\in\mathbb{N}^+$, we have sums of the form

$$\begin{aligned}(\overline{G}_Nf)(x,y):=&\sum_{j=1}^{N}\sum_{i=1}^{N}\left[f(x_i,y_j)-f(x_{i-1},y_j)\right]\sigma\Big(w\widetilde{\chi}_{ij}(x,y)\left\|(x,y)-(t_{x_i},t_{y_j})\right\|_2\Big)\\&+\sum_{j=1}^{N}f(x_0,y_j)\sigma\Big(w\widetilde{\chi}_{0j}(x,y)\left\|(x,y)-(t_{x_0},t_{y_j})\right\|_2\Big),\end{aligned} \tag{5.3}$$

where $(x,y)\in Q:=[a,b]\times[c,d]$, $h:=l/N$ with $l=b-a=d-c$, $x_i:=a+hi$, $y_j:=c+hj$ for $i,j=-1,0,1,\cdots,N$ and $t_{x_i}:=(x_{i-1}+x_i)/2$, $t_{y_j}:=(y_{j-1}+y_j)/2$ for $i,j=0,1,\cdots,N$, and moreover

$$\widetilde{\chi}_{ij}(x,y):=\begin{cases}+1, & \text{for } x\geq x_i \text{ and } y\in(y_{j-1},y_j],\\-1, & \text{otherwise,}\end{cases}$$

for $i=0,\cdots,N$ and $j=2,\cdots,N$, while

$$\widetilde{\chi}_{i1}(x,y):=\begin{cases} +1, & \text{for } x\geq x_j \text{ and } y\in[y_0,y_1],\\ -1, & \text{otherwise,}\end{cases}$$

for $i=0,\cdots,N$, when $f\in C(Q)$. The proof is similar to that of Theorem 5.1, and again Eq. (5.3) contains RBF neural networks.

By simple modifications in the proof of Theorem 5.1, the following estimate for the approximation error, i.e., for $\|\widetilde{G}_N f-f\|_\infty$, can be obtained.

**Theorem 5.2.** *Let $\sigma$ be a bounded sigmoidal function, and $f\in C(Q)$ be an Hölder-continuous function of order $\alpha$, $0<\alpha\leq 1$ with Hölder constant $L>0$. Then, for every $N\in\mathbb{N}^+$, $N>2$, there exists $\overline{w}>0$ (depending on $N$), such that for every $w\geq\overline{w}$, $\widetilde{G}_N f$ defined in (5.1) with $w$, is such that*

$$\left\|\widetilde{G}_N f-f\right\|_\infty<\frac{1}{N^\alpha}\left[L2^{\alpha/2+1}(b-a)^\alpha+2^{\alpha/2}(b-a)^\alpha\|\sigma\|_\infty+\|f\|_\infty\right].$$

*Proof.* Let $N\in\mathbb{N}^+$, $N>2$, be fixed. Set $h:=(b-a)/N=(d-c)/N$, $x_i:=a+hi$, $y_j:=c+hj$ for $i,j=-1,0,1,\cdots,N$ and $t_{x_i}:=(x_{i-1}+x_i)/2$, $t_{y_j}:=(y_{j-1}+y_j)/2$ for $i,j=0,1,\cdots,N$. Moreover, let $\overline{w}=\overline{w}(1/N^2,h/2)=\overline{w}(1/N^2)>0$ obtained by using Lemma 5.1 with $1/N^2$, $h/2>0$ and with the points $(t_{x_i},t_{y_j})$, $i,j=0,1,\cdots,N$. Consider $\widetilde{G}_N f$ defined in (5.1) for $w\geq\overline{w}$ and let $(x,y)\in Q$ be fixed. Adopting the same notation and following the same steps as in the proof of Theorem 5.1, we obtain

$$\begin{aligned}&\left|(\widetilde{G}_N f)(x,y)-f(x,y)\right|\\ \leq&\left|(\widetilde{G}_N f)(x,y)-\mathcal{L}_{k\mu}(x,y)\right|+\left|\mathcal{L}_{k\mu}(x,y)-f(x,y)\right|\\ =:&H_1+H_2.\end{aligned}$$

Now, we can estimate $H_1$ and $H_2$ as in Theorem 5.1, obtaining

$$\begin{aligned}H_1&<\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N|f(x_i,y_j)-f(x_i,y_{j-1})|+\frac{1}{N^2}\sum_{i=1}^N\|f\|_\infty\\ &\leq L2^{\alpha/2}\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\frac{(b-a)^\alpha}{N^\alpha}+\frac{1}{N}\|f\|_\infty\\ &\leq L2^{\alpha/2}\frac{(b-a)^\alpha}{N^\alpha}+\frac{1}{N}\|f\|_\infty.\end{aligned}$$

Moreover,

$$H_2<L2^{\alpha/2}\frac{(b-a)^\alpha}{N^\alpha}+2^{\alpha/2}\frac{(b-a)^\alpha}{N^\alpha}\|\sigma\|_\infty.$$

Hence,

$$\begin{aligned}\left|(\widetilde{G}_N f)(x,y)-f(x,y)\right| &\le H_1+H_2\\ &< L2^{\alpha/2+1}\frac{(b-a)^\alpha}{N^\alpha}+2^{\alpha/2}\frac{(b-a)^\alpha}{N^\alpha}\|\sigma\|_\infty+\frac{1}{N}\|f\|_\infty\\ &\le\frac{1}{N^\alpha}\left[L2^{\alpha/2+1}(b-a)^\alpha+2^{\alpha/2}(b-a)^\alpha\|\sigma\|_\infty+\|f\|_\infty\right],\end{aligned}$$

and since $(x,y)\in Q$ is arbitrary, it follows that

$$\left\|\widetilde{G}_N f-f\right\|_\infty<\frac{1}{N^\alpha}\left[L2^{\alpha/2+1}(b-a)^\alpha+2^{\alpha/2}(b-a)^\alpha\|\sigma\|_\infty+\|f\|_\infty\right].$$

The proof is complete. □

Aiming at building a constructive theory also in $L^p(Q)$, $1\le p<\infty$, we prove the following theorem, which parallels that established in Section 3 for the univariate case.

**Theorem 5.3.** *Let $\sigma$ be a bounded sigmoidal function, and $1\le p<\infty$ be fixed. For any $f\in C(Q)$, $Q\subset\mathbb{R}^2$ and $\varepsilon>0$, there exist $N\in\mathbb{N}^+$ and $w>0$ (depending on $N$), such that $\widetilde{G}_N f$ defined in* (5.1) *with $w$, is such that*

$$\left\|\widetilde{G}_N f-f\right\|_{L^p(Q)}<\varepsilon.$$

*Proof.* Let $f\in C(Q)$ and $\varepsilon>0$ be fixed. By Theorem 5.1, for $\eta:=\varepsilon/|Q|^{1/p}$ ($|Q|$ denoting the Lebesgue measure of $Q$), there exist $N\in\mathbb{N}^+$ and $w>0$ depending on $N$ such that $\|\widetilde{G}_N f-f\|_\infty<\eta$. Therefore,

$$\begin{aligned}\left\|\widetilde{G}_N f-f\right\|_{L^p(Q)}&=\left(\int_Q\left|(\widetilde{G}_N f)(x,y)-f(x,y)\right|^p dxdy\right)^{1/p}\\ &<\left(\int_Q\eta^p dxdy\right)^{1/p}=\varepsilon.\end{aligned}$$

The proof is complete. □

We are now able to prove an approximation theorem in $L^p(Q)$.

**Theorem 5.4.** *Let $\sigma$ be a bounded sigmoidal function and let $f\in L^p(Q)$, $1\le p<\infty$, $Q\subset\mathbb{R}^2$ be fixed. Then, for every $\varepsilon>0$, there exists $N\in\mathbb{N}^+$ and $\widetilde{G}_N$, that is a linear combination of sigmoidal functions based on $\sigma$, having the bivariate form introduced in Theorem* 5.1, *such that*

$$\left\|\widetilde{G}_N-f\right\|_{L^p(Q)}<\varepsilon.$$

*Proof.* Define the function $\widetilde{f}:\mathbb{R}^2\to\mathbb{R}$ by

$$\widetilde{f}(x,y):=\begin{cases} f(x,y), & \text{for } (x,y)\in Q,\\ 0, & \text{otherwise.}\end{cases} \tag{5.4}$$

Now, $\widetilde{f}\in L^p(\mathbb{R})$ and $\widetilde{f}=f$ on $Q$. Let $\{\rho_n\}_{n\in\mathbb{N}^+}$, $\rho_n:\mathbb{R}^2\to\mathbb{R}$, be a sequence of (bivariate) mollifiers ($\rho_n$ enjoys the same properties listed in Theorem 3.2 for the univariate case). Define the family of functions $\{f_n\}_{n\in\mathbb{N}}$ by

$$f_n(x,y)=(\rho_n*\widetilde{f})(x,y):=\int_{\mathbb{R}^2}\rho_n(x-z,y-t)\widetilde{f}(z,t)\,dzdt,$$

where $(x,y)\in\mathbb{R}^2$ and $*$ denotes, as usual, the convolution product. By general properties of the sequences of mollifiers and of convolution [5], it turns out that $f_n=\rho_n*\widetilde{f}\in C(\mathbb{R}^2)$ for every $n\in\mathbb{N}^+$, and $f_n\to\widetilde{f}$ in $L^p(\mathbb{R}^2)$ as $n\to\infty$. Let $\varepsilon>0$ be fixed. Then, there exist $\overline{n}\in\mathbb{N}$ such that

$$\|f_n-f\|_{L^p(Q)}=\left\|f_n-\widetilde{f}\right\|_{L^p(Q)}\le\left\|f_n-\widetilde{f}\right\|_{L^p(\mathbb{R}^2)}<\frac{\varepsilon}{2}$$

for every $n\ge\overline{n}$. Let now $n\ge\overline{n}$ be fixed. Since $f_n\in C(\mathbb{R}^2)\subset C(Q)$, it follows that, as a consequence of Theorem 5.3, correspondingly to $\varepsilon/2$ there exist $N\in\mathbb{N}^+$ and $w>0$, depending on $N$, such that $\widetilde{G}_N f_n$ defined in (5.1) with $w$, is such that

$$\left\|\widetilde{G}_N f_n-f_n\right\|_{L^p(Q)}=\left\|\widetilde{G}_N(\rho_n*\widetilde{f})-(\rho_n*\widetilde{f})\right\|_{L^p(Q)}<\frac{\varepsilon}{2}.$$

Therefore, we obtain by the previous estimates

$$\left\|\widetilde{G}_N f_n-f\right\|_{L^p(Q)}\le\left\|\widetilde{G}_N f_n-f_n\right\|_{L^p(Q)}+\|f_n-f\|_{L^p(Q)}<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$

Setting $\widetilde{G}_N(x,y):=(\widetilde{G}_N f_n)(x,y)$ completes the proof. □

Examples of sequences of bivariate mollifiers (or, more generally, of multivariate mollifiers) are given in, e.g., [5, 24]. In particular, if we consider $\widetilde{\rho}(x_1,\cdots,x_n)=\rho(\|(x_1,\cdots,x_n)\|_2)$, $n\in\mathbb{N}^+$, the natural extension to the multivariate case of the function $\rho$ defined in (3.3), we can build the sequence of mollifiers $\widetilde{\rho}_k(x_1,\cdots,x_n):=Ck^n\widetilde{\rho}(kx_1,\cdots,kx_n)$, where

$$C:=\left(\int_{\mathbb{R}^n}\widetilde{\rho}(x_1,\cdots,x_n)\,dx_1\cdots dx_n\right)^{-1}.$$

# 6 Applications based on specific sigmoidal functions

A first example of sigmoidal function is the so-called logistic-function, which was already used in Remark 4.1, and is defined as $\sigma(x):=(1+e^{-x})^{-1}$, $x\in\mathbb{R}$. Logistic functions are largely used in many fields, such as Biology, Physics, Biomathematics, Statistics, Economics, and Demography (see, e.g., [4, 19]), and indeed they were first introduced in the 19th Century as a model to describe population growth. Clearly, $\sigma(x)$ is a bounded sigmoidal function with $0<\sigma(x)<1$ for every $x\in\mathbb{R}$. Using the logistic function above, we can see that, if $x_0,x_1,\cdots,x_M\in\mathbb{R}$, $M\in\mathbb{N}^+$, for every $N\in\mathbb{N}^+$, $N>2$, and $h>0$, there exists $\overline{w}:=\frac{1}{h}\log(N-1)>0$ such that for every $w>\overline{w}$ and $k=0,1,\cdots,M$, we have

1. $|\sigma(w(x-x_k))-1|<\frac{1}{N}$, for every $x\in\mathbb{R}$ such that $x-x_k\geq h$;
2. $|\sigma(w(x-x_k))|<\frac{1}{N}$, for every $x\in\mathbb{R}$ such that $x-x_k\leq -h$.

In fact, let $N\in\mathbb{N}^+$, $N>2$, be fixed. Then, as $0<\sigma(x)<1$, for every $x\in\mathbb{R}$,

$$|\sigma(x)-1|=1-\frac{1}{1+e^{-x}}<\frac{1}{N}\quad\text{for } x>\log(N-1),$$
$$|\sigma(x)|=\frac{1}{1+e^{-x}}<\frac{1}{N}\quad\text{for } x<-\log(N-1).$$

Therefore, for every $w>\overline{w}:=\frac{1}{h}\log(N-1)$ and for every $x\in\mathbb{R}$ with $x-x_k\geq h$, $k=0,\cdots,M$, we have $w(x-x_k)>\overline{w}h=\log(N-1)$, hence $|\sigma(w(x-x_k))-1|<\frac{1}{N}$. Similarly, for every $w>\overline{w}$ and $x\in\mathbb{R}$ with $x-x_k\leq -h$, we have $|\sigma(w(x-x_k))|<\frac{1}{N}$.

The previous inequalities provide an estimate for $w>0$ in case of approximations made by finite linear combination of logistic function. Consequently, by Theorems 2.1 and 4.1 we can prove the following:

**Corollary 6.1.** *Let $\sigma(x)=(1+e^{-x})^{-1}$ and $f\in\widehat{C}^{n+1}[a,b]$, $n\in\mathbb{N}^+$, be fixed. Denote by*

$$\begin{aligned}(G_Nf)(x):=&\sum_{k=1}^{N}[f(x_k)-f(x_{k-1})]\sigma\left(\left(\frac{N}{(b-a)}\log(N-1)+\delta\right)(x-x_k)\right)\\&+f(x_0)\sigma\left(\left(\frac{N}{(b-a)}\log(N-1)+\delta\right)(x-x_{-1})\right),\\(G^j_Nf)(x):=&\sum_{k=1}^{N-j}\left(\Delta^j_kf-\Delta^j_{k-1}f\right)\sigma\left(\left(\frac{N}{(b-a)}\log(N-1)+\delta\right)(x-x_k)\right)\\&+\Delta^j_0f\sigma\left(\left(\frac{N}{(b-a)}\log(N-1)+\delta\right)(x-x_{-1})\right),\end{aligned}$$

*$x\in[a,b]$, $N\in\mathbb{N}^+$, $j=1,\cdots,n$, $\delta>0$, and $x_k=a+k\frac{(b-a)}{N}$, $k=-1,0,1,\cdots,N$. Then, for every $\varepsilon>0$ there exists $N\in\mathbb{N}^+$, $N>n+3$, such that*

*(i)* $\|G_Nf-f\|_\infty<\varepsilon$;

*(ii)* $\|G^j_Nf-f^{(j)}\|_\infty<\varepsilon$, *for every $j=1,\cdots,n$.*

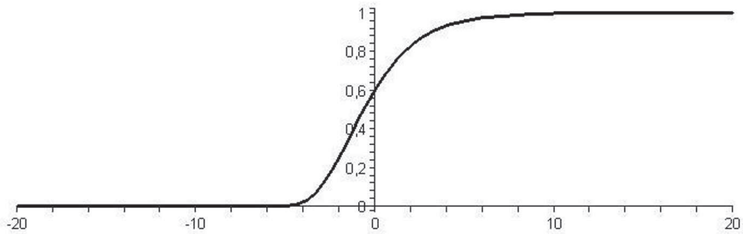


Figure 1: The sigmoidal Gompertz function for $\alpha=0.5$ and $\beta=0.5$.

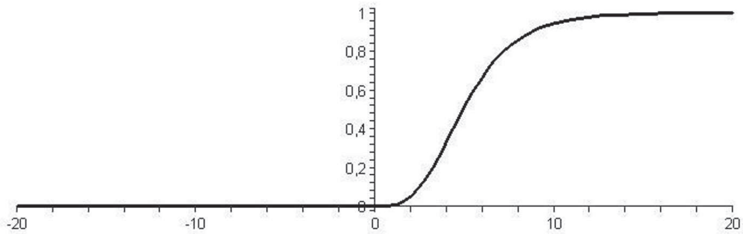


Figure 2: The sigmoidal Gompertz function for $\alpha=8$ and $\beta=0.5$.

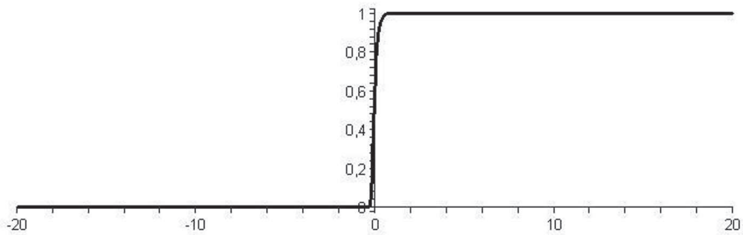


Figure 3: The sigmoidal Gompertz function for $\alpha=0.5$ and $\beta=8$.

Other interesting examples of sigmoidal functions are provided by the class of the Gompertz functions, first introduced by Benjamin Gompertz for the study of his demographic model, which represents a refinement of the Malthus model. The class of sigmoidal Gompertz functions is defined by

$$\sigma_{\alpha\beta}(x):=e^{-\alpha e^{-\beta x}}, \quad x\in\mathbb{R}, \tag{6.1}$$

where $\alpha$, $\beta>0$ represent an effective translation and a scaling term, respectively (see Fig. 1, 2 and 3). Gompertz functions find applications, e.g., in modeling tumor growth [2,15,27]. Note that $\sigma_{\alpha\beta}:\mathbb{R}\to\mathbb{R}$ is a bounded sigmoidal function with $0<\sigma_{\alpha\beta}(x)<1$, for every $x\in\mathbb{R}$. Moreover, if $x_0,x_1,\cdots,x_M\in\mathbb{R}$, and $M\in\mathbb{N}^+$, for every $N\in\mathbb{N}^+$, $N>2$ and $h>0$, there exists

$$\overline{w}:=\frac{1}{h\beta}\max\left\{\left|\log\left(-\frac{1}{\alpha}\log\left(\frac{N-1}{N}\right)\right)\right|,\left|\log\left(\frac{1}{\alpha}\log(N)\right)\right|\right\}$$

such that for every $w>\overline{w}$ and $k=0,1,\cdots,M$, we have

1. $\left|\sigma_{\alpha\beta}(w(x-x_k))-1\right|<\frac{1}{N}$, for every $x\in\mathbb{R}$ such that $x-x_k\geq h$,
2. $\left|\sigma_{\alpha\beta}(w(x-x_k))\right|<\frac{1}{N}$, for every $x\in\mathbb{R}$ such that $x-x_k\leq -h$.

In fact, let $N\in\mathbb{N}^+$, $N>2$ be fixed. Then, as $0<\sigma_{\alpha\beta}(x)<1$ for every $x\in\mathbb{R}$, we have

$$\left|\sigma_{\alpha\beta}(x)-1\right|=1-e^{-\alpha e^{-\beta x}}<\frac{1}{N}\quad\text{for }\ x>-\frac{1}{\beta}\log\left(-\frac{1}{\alpha}\log\left(\frac{N-1}{N}\right)\right),$$
$$\left|\sigma_{\alpha\beta}(x)\right|=e^{-\alpha e^{-\beta x}}<\frac{1}{N}\quad\text{for }\ x<-\frac{1}{\beta}\log\left(\frac{1}{\alpha}\log(N)\right).$$

Set

$$\overline{w}:=\frac{1}{h\beta}\max\left\{\left|\log\left(-\frac{1}{\alpha}\log\left(\frac{N-1}{N}\right)\right)\right|,\left|\log\left(\frac{1}{\alpha}\log(N)\right)\right|\right\}.$$

For every $w>\overline{w}$ and for every $x\in\mathbb{R}$ such that $x-x_k\geq h$, $k=0,\cdots,M$, we have

$$w(x-x_k)>\overline{w}h\geq\left|\frac{1}{\beta}\log\left(-\frac{1}{\alpha}\log\left(\frac{N-1}{N}\right)\right)\right|,$$

then $\left|\sigma_{\alpha\beta}(w(x-x_k))-1\right|<\frac{1}{N}$. Similarly, for every $w>\overline{w}$ and $x\in\mathbb{R}$ such that $x-x_k\leq -h$, we have $\left|\sigma_{\alpha\beta}(w(x-x_k))\right|<\frac{1}{N}$. Consequently, we have from Theorems 2.1 and 4.1 the following:

**Corollary 6.2.** *Let $\sigma_{\alpha\beta}(x)=e^{-\alpha e^{-\beta x}}$, $\alpha$, $\beta>0$, and $f\in\widehat{C}^{n+1}[a,b]$, $n\in\mathbb{N}^+$, be fixed. Define*

$$(G_N f)(x):=\sum_{k=1}^{N}[f(x_k)-f(x_{k-1})]\sigma_{\alpha\beta}(W(x-x_k))+f(x_0)\sigma_{\alpha\beta}(W(x-x_{-1})),$$
$$(G_N^j f)(x):=\sum_{k=1}^{N-j}\left(\Delta_k^j f-\Delta_{k-1}^j f\right)\sigma_{\alpha\beta}(W(x-x_k))+\Delta_0^j f\,\sigma_{\alpha\beta}(W(x-x_{-1})),$$

*where*

$$W:=\frac{N}{(b-a)\beta}\max\left\{\left|\log\left(-\frac{1}{\alpha}\log\left(\frac{N-1}{N}\right)\right)\right|,\left|\log\left(\frac{1}{\alpha}\log(N)\right)\right|\right\}+\delta,$$

*$x\in[a,b]$, $N\in\mathbb{N}^+$, $j=1,\cdots,n$, $\delta>0$, and $x_k=a+k\frac{(b-a)}{N}$, $k=-1,0,1,\cdots,N$. Then, for every $\varepsilon>0$ there exists $N\in\mathbb{N}^+$, $N>n+3$, such that*

*(i)* $\|G_N f-f\|_\infty<\varepsilon$;

*(ii)* $\|G_N^j f-f^{(j)}\|_\infty<\varepsilon$, *for every* $j=1,\cdots,n$.

Corollaries 6.1 and 6.2 can be easily extended to the multivariate case.

Other useful examples of sigmoidal functions are the unit step function (or Heaviside function) $H(x)$ defined by $H(x):=1$ for $x\geq 0$ and $H(x):=0$ for $x<0$, the arctan-related sigmoidal function

$$\sigma_1(x):=\frac{1}{2}+\frac{1}{\pi}\arctan x,\quad x\in\mathbb{R},$$

and that based on the hyperbolic tangent

$$\sigma_2(x) := \frac{1}{2}(1+\tanh x) = \frac{1}{2} + \frac{1}{2}\frac{e^{2x}-1}{e^{2x}+1}, \quad x \in \mathbb{R}.$$

# 7 Numerical examples

In this section, we give some examples to illustrate applications of approximation of functions by means of linear combination of sigmoidal functions.

**Example 7.1.** Consider the function

$$f(x) := (\cos^2 x + 2)\sin x + 2x + \frac{1}{8}x^2 + 4, \quad x \in \mathbb{R}. \tag{7.1}$$

We first construct approximations of $f$ on the interval $[-5,5]$, obtained by the superposition of logistic sigmoidal functions (Figs. 4 and 5), for $N=25$ and $N=50$ and by the Gompertz function $\sigma_{\alpha\beta}$ with $\alpha=8$ and $\beta=0.5$ (Fig. 6), for $N=50$. The choice of the parameter $w>0$ was done according to Corollaries 6.1 and 6.2, i.e., $w=N^2/(b-a)$ in the case of $\sigma(x)=(1+e^{-x})^{-1}$, and $w=N^2/((b-a)\alpha\beta)$, in case of $\sigma_{\alpha\beta}(x)$ ($\alpha=8$, $\beta=0.5$).

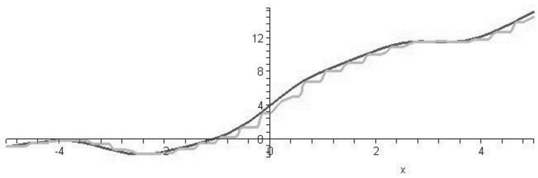

Figure 4: Approximation of $f$ (black) of Example 7.1 by $G_N f$ (grey), $N=25$, defined by the logistic sigmoidal functions ($w=N^2/(b-a)$). Here $\|G_N f - f\|_\infty / \|f\|_\infty \approx 1.08\times 10^{-1}$.

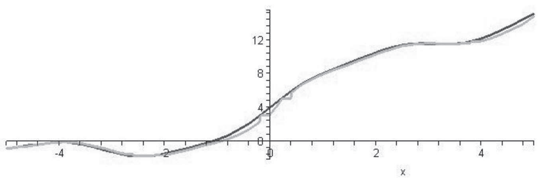

Figure 5: Approximation of $f$ (black) of Example 7.1 by $G_N f$ (grey), $N=50$, defined by the logistic sigmoidal functions ($w=N^2/(b-a)$). Here $\|G_N f - f\|_\infty / \|f\|_\infty \approx 5.88\times 10^{-2}$.

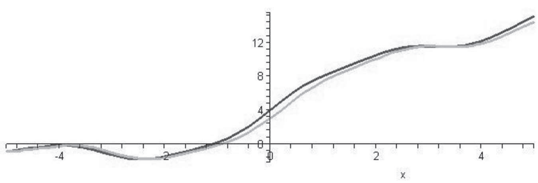

Figure 6: Approximation of $f$ (black) of Example 7.1 by $G_N f$ (grey), $N=50$, defined by the Gompertz sigmoidal function $\sigma_{\alpha\beta}$, $\alpha=8$, $\beta=0.5$ ($w=N^2/((b-a)\alpha\beta)$). Here $\|G_N f - f\|_\infty / \|f\|_\infty \approx 7.27\times 10^{-2}$.

Note that the choice of a *specific* sigmoidal function $\sigma$ affects the quality of the approximation, which turns out to be better in case of $\sigma(x)=(1+e^{-x})^{-1}$. In addition, we have

$$f'(x):=\left(\cos^2 x+2\right)\cos x-2\sin^2 x\cos x+\frac{1}{4}x+2 \quad x\in\mathbb{R},$$

and, as a consequence of Theorem 4.1, we can also approximate $f'$ on the interval $[-5,5]$ by means of the functions $G_N^1 f$, $N\in\mathbb{N}^+$. The graphs of $f'$ and its approximation obtained for $N=25$ with logistic functions, $w=N^2/(b-a)$, are plotted in Fig. 7.

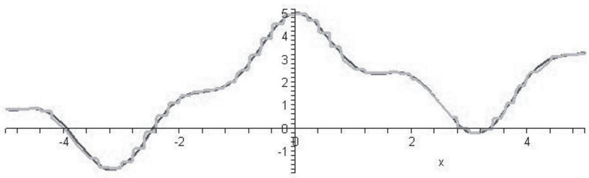


Figure 7: Approximation of $f'$ (black) by $G_N^1 f$ (grey) of Example 7.1, $N=25$ with logistic functions, $\sigma$, with $w=N^2/(b-a)$, $b-a=10$. Here $\|G_N^1 f-f'\|_\infty/\|f'\|_\infty\approx 9.40\times 10^{-2}$.

Note that the approximation obtained for $f'$ is better compared to those for $f$, according to the results concerning the size of the approximation errors in Theorem 4.2. In fact, this states that such an error depends on the sup-norm (on the fixed interval $[a,b]$) of the function being approximated, and here $\|f\|_\infty\approx 15.13$, while $\|f'\|_\infty\approx 5$.

Now, we show an application of the constructive theory developed in Section 3 for $L^p$-functions.

**Example 7.2.** Let $g\in L^1(\mathbb{R})$ be defined by

$$g(x):=\begin{cases}\dfrac{4}{x^2-2}, & x<-2,\\ -3, & -2\le x<0,\\ \dfrac{5}{2}, & 0\le x<2,\\ \dfrac{3x+2}{x^3-1}, & x\ge 2.\end{cases} \tag{7.2}$$

We consider approximations of $g$ on the interval $[-5,5]$ by finite linear combinations of sigmoidal functions of the form $G_N(\rho_n*\widetilde{g})$, where $\rho_n$ are the mollifiers defined in (3.4), and $\widetilde{g}$ is the extension of $g$ defined by

$$\widetilde{g}(x):=\begin{cases} g(x), & -5\le x<5,\\ 0, & \text{otherwise}.\end{cases}$$

In Figs. 8 and 9 such approximations obtained by $G_N(\rho_n*\widetilde{g})$ for $n=10$ and, respectively, $N=50$ and $N=100$, by using logistic functions with $w=N^2/(b-a)$, are shown. In Fig. 10,

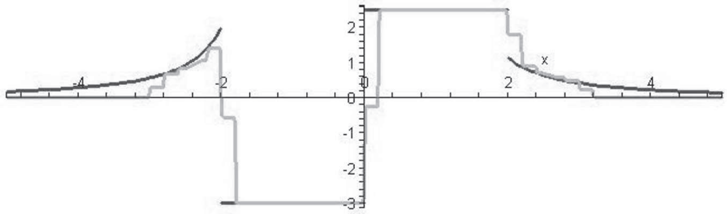

Figure 8: Approximation of the function $g$ (black) of Example 7.2, by $G_N(\rho_n * \widetilde{g})$ (grey), $n=10$, $N=50$, $w=N^2/(b-a)$, $b-a=10$, $\sigma(x)=(1+e^{-x})^{-1}$.

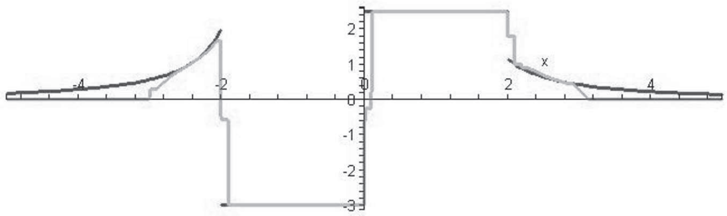

Figure 9: Approximation of the function $g$ (black) of Example 7.2 by $G_N(\rho_n * \widetilde{g})$ (grey), $n=10$, $N=100$, $w=N^2/(b-a)$, $b-a=10$, $\sigma(x)=(1+e^{-x})^{-1}$.

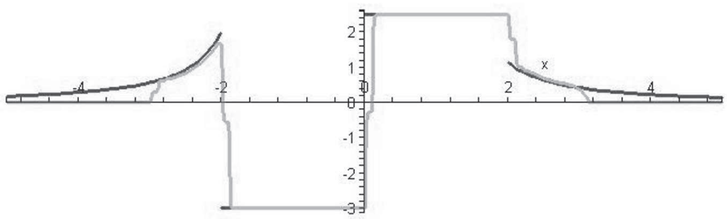

Figure 10: Approximation of the function $g$ (black) of Example 7.2 by $G_N(\rho_n * \widetilde{g})$ (grey), $n=10$, $N=100$, $\sigma_{\alpha\beta}$, $\alpha=8$, $\beta=0.5$ and $w=N^2/((b-a)\alpha\beta)$, $b-a=10$.

the approximation obtained by the Gompertz function $\sigma_{\alpha\beta}(x)$, with $\alpha=8$, $\beta=0.5$, $n=10$, $N=100$ and with $w=N^2/((b-a)\alpha\beta)$, is shown.

We can observe that the approximation improves as $N$ increases, as one would expect. In addition, the error made approximating $g$ with the same values of $N$ is larger than that made in the cases of regular functions, as were $f$ and $f'$ in the previous example.

Finally, here is an example of multivariate approximation.

**Example 7.3.** Consider the function of two variables, $h:[-4,4]\times[-4,4]\to\mathbb{R}$, defined as

$$h(x,y):=\Big(y^4-2y\Big)\sin x - xy^3+\frac{x}{3}+\frac{1}{x^{16}/30+e^{-3|y|}/50+1/100},$$

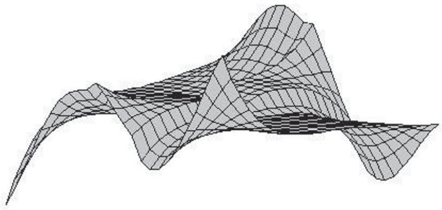

Figure 11: Graph of the bivariate function $h$ of Example 7.3.

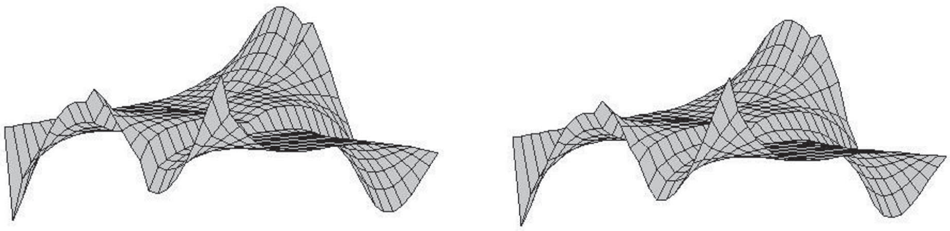

Figure 12: Approximation for the function $h$ in Example 7.3 by $\widetilde{G}_N h$, for $N=25$ and $N=50$, respectively, $w=N^2/8$, $\sigma(x)=(1+e^{-x})^{-1}$.

see Fig. 11. Fig. 12 is obtained by using the functions $\widetilde{G}_N h$ defined in (5.1) for $N=25$ and $N=50$, respectively, where $\sigma$ is the logistic function. Similar approximations can be obtained by using sums $\overline{G}_N h$ like those in (5.3).

# Acknowledgments

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

**References**


[1] Adrianov, A.V., An analog of the Jackson-Nikol'skij theorem on the approximation by superposition of sigmoidal functions, Mathematical Notes [transl. of Matematicheskie Zametki, Russian Academy of Sciences], 59:6 (1996), 661-664.

[2] Albano, G., Giorno, V., On the First Exit Time Problem for a Gompertz-Type Tumor Growth, Lecture Notes in Computer Science, 5717 (2009), 113-120.

[3] Barron, A.R., Universal approximation bounds for superpositions of a sigmoidal function, IEEE Transactions on Information Theory, 39:3(1993), 930-945.

[4] Brauer, F., Castillo Chavez, C., Mathematical Models in Population Biology and Epidemiology, New York: Springer, 2001.

[5] Brezis, H., Functional analysis, Sobolev spaces and partial differential equations. Universitext, New York: Springer, 2011.

[6] Cao, F.L., Chen, Z., The approximation operators with sigmoidal functions, Computer & Mathematics with Applications, 58:4 (2009) 758-765.

[7] Cao, F.L., He, Z.R., Zhang, Y., Interpolation and rates of convergence for a class of neural networks, Applied Math. Modelling, 33:3 (2009) 1441-1456.

[8] Cao, F.L., Xie, T., Xu, Z.B., The estimate for approximation error of neural networks: A constructive approach, Neurocomputing, 71:4-6 (2008), 626-630.

[9] Cao, F.L., Xu, Z.B., The essential order of approximation for neural networks, Sci. China, Ser. F: Information Sciences, 47:1 (2004), 97-112.

[10] Cao, F.L., Xu, Z.B., Simultaneous $L^p$-approximation order for neural networks, Neural Networks, 18:7 (2005), 914-923.

[11] Cao, F.L., Zhang, R., The errors of approximation for feedforward neural networks in the $L^p$ metric, Mathematical and Computer Modelling, 49:7-8 (2009) 1563-1572.

[12] Chen, D., Degree of approximation by superpositions of a sigmoidal function, Approx. Theory Appl., 9:3 (1993), 17-28.

[13] Chen, H., Chen, T., Liu, R., A constructive proof of Cybenko's approximation theorem and its extension, in C. Page and R. LePage (eds.) Computing Science and Statistics, (1990), 163-168.

[14] Chen, H., Chen, T., Liu, R., A constructive proof and an extension of Cybenko's approximation theorem, Computing Science and Statistics, New York: Springer-Verlag, (1992), 163-168.

[15] Chumerina, E.S., Choice of optimal strategy of tumor chemotherapy in Gompertz model, Journal of Computer and Systems Sciences International, 48:2 (2009), 325-331.

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

[17] Gao, B., Xu, Y., Univariant approximation by superpositions of a sigmoidal function, J. Math. Anal. Appl., 178 (1993), 221-226.

[18] Hahm, N., Hong, B., Approximation order to a function in $\overline{C}(\mathbb{R})$ by superposition of a sigmoidal function, Applied Math. Lett., 15 (2002), 591-597.

[19] Hritonenko, N., Yatsenko, Y., Mathematical Modelling in Economics, Ecology and the Environment, Beijing: Science Press, 2006.

[20] Jones, L.K., Constructive approximations for neural networks by sigmoidal functions, Technical Report Series 7, Department of Mathematics, University of Lowell, 1988.

[21] Lenze, B., Constructive multivariate approximation with sigmoidal functions and applications to neural networks, Numerical Methods of Approximation Theory, Birkhaüser Verlag, Basel-Boston-Berlin, (1992), 155-175.

[22] Lewicki, G., Marino, G., Approximation by superpositions of a sigmoidal function, Zeitschrift für Analysis und ihre Anwendungen Journal for Analysis and its Applications, 22:2 (2003), 463-470.

[23] Lewicki, G., Marino, G., Approximation of functions of finite variation by superpositions of a sigmoidal function, Appl. Math. Lett., 17 (2004), 1147-1152.

[24] Li, X., Simultaneous approximations of multivariate functions and their derivatives by neural networks with one hidden layer, Neurocomputing, 12 (1996), 327-343.

[25] Li, X., On Simultaneous approximations by radial basis function neural networks, Appl. Math. Comput., 95 (1998), 75-89.

[26] Li, X., Micchelli, C.A., Approximation by radial basis and neural networks, Numerical Algorithms, 25 (2000), 241-262.

[27] Mantysaari, E.A., Sevon Aimonen, M.L., Stranden, I., Vuori, K., Estimation of non-linear growth models by linearization: a simulation study using a Gompertz function, Genet. Sel. Evol., 38 (2006), 343-358.

[28] Medvedeva, M.V., On sigmoidal functions, Mosc. Univ. Math. Bull., 53:1 (1998), 16-19.

[29] Mhaskar, H.N., Micchelli, C.A., Approximation by superposition of sigmoidal and radial basis functions, Adv. in Appl. Math., 13 (1992), 350-373.
[30] Mhaskar, H.N., Micchelli, C.A., Degree of approximation by neural and translation networks with a single hidden layer, Adv. in Appl. Math., 16 (1995), 151-183.
[31] Park, J., Sandberg, I.W., Universal approximation using radial-basis-function networks, Neural Computation, 3:2 (1991), 246-257.
[32] Powell, M.J.D., The theory of radial basis function approximation, Adv. in Num. Anal. III, Wavelets, Subdivision Algorithms and Radial Basis Functions (W.A. Light Ed.), Oxford: Clarendon Press, (1992), 105-210.