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 σ 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 σ 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 σ (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 α_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

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