# Two-Hidden-Layer Feedforward Networks are Universal Approximators: a Constructive Approach

**Rocio Gonzalez-Diaz**
Dept. of Applied Mathematics I
University of Seville
rogodi@us.es

**Miguel A. Gutiérrez-Naranjo**
Dept. of Computer Science and Artificial Intelligence
University of Seville
magutier@us.es

**Eduardo Paluzo-Hidalgo**
Dept. of Applied Mathematics I
University of Seville
epaluzo@us.es

July 29, 2019

## ABSTRACT

It is well known that Artificial Neural Networks are universal approximators. The classical result proves that, given a continuous function on a compact set on an $n$-dimensional space, then there exists a one-hidden-layer feedforward network which approximates the function. Such result proves the existence, but it does not provide a method for finding it. In this paper, a constructive approach to the proof of this property is given for the case of two-hidden-layer feedforward networks. This approach is based on an approximation of continuous functions by simplicial maps. Once a triangulation of the space is given, a concrete architecture and set of weights can be obtained. The quality of the approximation depends on the refinement of the covering of the space by simplicial complexes.

***Keywords*** Universal approximation theorem · Simplicial approximation theorem · Multilayer feedforward network · Simplicial Complexes

## 1 Introduction

One of the first results in the development of neural networks are the *universal approximation theorems* [1, 2]. These classical results show that a multilayer feedforward network with only one hidden layer and non-polynomial activation function (like the sigmoid) can approximate every continuous function over a compact set on $\mathbb{R}^n$. It is well-known that these results have two important drawbacks for their practical use: On the one hand, the width of the hidden layer grows exponentially and, on the other hand, the proof is not constructive, and it does not provide an algorithm for building such network.

Bearing these results in mind, many researchers are paying attention to theoretical aspects of the current success of new neural network architectures and asking for bounds for the depth and width of such networks and the possibility of acting as universal approximators (see, e.g., [3, 4, 5] among many others).

It is out of any doubt that the use of many hidden layers is a big contribution to the success of deep learning architectures [6], but instead of exploring the power of depth, recently several studies have made interesting contributions on the power of width [7, 8, 9]. To sum up, the authors show that there exists continuous functions on compact sets which cannot be approximated by a neural network if the width of the layers is not larger than a bound, regardless the depth of the network.

---

All authors are listed by alphabetical order.

In this paper, we provide an effective method for finding the weights of a multilayer feedforward network with two hidden layers which approximates any continuous mappings between two triangulable sets on metric spaces. Let us remark that the method is constructive and it only depends on the desired level of approximation to the original continuous mapping.

Our approach is based on a strong result from computational topology. Roughly speaking, our result is rooted in two observations: On the one hand, triangulable sets can be modelled by simplicial complexes, and a continuous mapping between two triangulable spaces can be approximated by a mapping between simplicial complexes. On the second hand, a mapping between simplicial complexes can be modelled as a two-hidden-layer feedforward network.

Let us remark that the classical result is valid for all compact set on $\mathbb{R}^n$ and our theorem is valid for triangulable sets of $\mathbb{R}^n$, which is a smaller collection of sets. Nevertheless, as it will be detailed below, compact and triangulable sets differ in a quite technical topological property, and therefore, our result have a practical application in many current problems.

The paper is organized as follows: in Section 2 the preliminary notions about multilayer feedforward networks and simplicial complexes are provided. Then, in Section 3, a concrete architecture that acts equivalently to a given simplicial map is given. In Sections 4 and 5, we extend the already known simplicial approximation theorem and universal approximation theorem, respectively. Finally, conclusions are supplied in 6.

## 2  Background

In this section, some notions about artificial neural networks and simplicial complexes are recalled.

### 2.1  Multilayer feedforward networks

Artificial neural networks are inspired in biological networks of alive neurons in a brain. The number of different architectures, algorithms, and areas of application has recently grown in many directions. In general, a neural network can be formalized as a mapping $\mathcal{N}_{\omega,\Theta} : \mathbb{R}^n \to \mathbb{R}^m$ which depends on a set of weights $\omega$ and a set of parameters $\Theta$ which involves the description of activation functions, layers, synapses between neurons, and whatever other consideration in its architecture. A good introduction to artificial neural networks can be found in [10].

In this paper, we focus on one of the simplest classes of artificial neural networks: multilayer feedforward networks which consist of three or more fully connected layers of nodes (neurons): an input layer, an output layer, and one or more hidden layers. Each node in one layer has an activation function and it is connected with every node in the following layer. The next definition formalizes this idea.

**Definition 1.** *A multilayer feedforward network defined on a real-valued $n$-dimensional space is a mapping $\mathcal{N} : \mathbb{R}^n \to \mathbb{R}^m$ such that, for each $x \in \mathbb{R}^n$, $\mathcal{N}(x)$ is computed as the composition of functions*

$$\mathcal{N}(x) = f_{k+1} \circ f_k \circ \cdots \circ f_1(x)$$

*where $k \in \mathbb{Z}$ is the number of hidden layers, $k \geq 1$, $f_i : \mathbb{R}^{d_{i-1}} \to \mathbb{R}^{d_i}$, $1 \leq i \leq k+1$, is defined as $f_i(y) = \phi_i(W^{(i)}; y; b_i)$ being $W^{(i)} = W^{d_i}_{d_{i-1}}$ a real-valued $d_{i-1} \times d_i$ matrix, $b_i \in \mathbb{R}^{d_i}$ the bias term, and $\phi_i$ a bounded, continuous, and non-constant function (called activation function). Notice that $d_0 = n$, $d_{k+1} = m$, $d_i \in \mathbb{Z}$, $1 \leq i \leq k$, is called the width of the $i$-th hidden layer and $\max\{d_0, \ldots, d_{k+1}\}$ is the width of the multilayer feedforward network.*

Next, we re-write one of the most important theoretical results of multilayer feedforward networks adapted to our notation.

**Theorem 1** (Universal Approximation Theorem, [2])**.** *Let $A$ be any compact subset of $\mathbb{R}^n$. The space of real-valued continuous functions on $A$ is denoted by $C(A)$. Then, given any $\epsilon > 0$ and any function $g \in C(A)$, there exists a one-hidden-layer feedforward network $\mathcal{N} : \mathbb{R}^n \to \mathbb{R}$ defined as $\mathcal{N}(x) = f_2 \circ f_1(x)$ with $f_1(y) = \phi_1(W^{(1)}; y; b_1)$ and $f_2(y) = W^{(2)}y$, such that $\mathcal{N}$ is an approximation of the function $g$, that is,*

$$|\mathcal{N}(x) - g(x)| < \epsilon$$

*for all $x \in A$.*

The known proofs given to this theorem are non constructive (see [11, 2, 1]), i.e., given $g$ and $\epsilon$, the theorem claims that a neural network $\mathcal{N}$ exists fulfilling the conditions, but there is no an algorithm to build it in the general case. In this paper, we will provide a constructive approach to this theorem but using a multilayer feedforward network with two hidden layers, instead of one. The most important restriction is that our result is valid for triangulable sets, instead of compact ones, but, as pointed out above, it covers most of the real-world problems.

## 2.2 Simplicial Complexes

The main result in this paper is based on a strong result in computational topology known as the *Simplicial Approximation Theorem*. Next, we recall such theorem and some basics on the theory of simplicial complexes. We have followed the notation from [12, 13].

Simplicial complexes are a data structure widely used to represent topological spaces. They are a versatile mathematical objects to represent a decomposition of a topological space into simple pieces. For a further comprehension of the field [14, 15, 12, 16, 13] can be consulted. Next, we recall the formal definition of the small pieces, called simplices, that are used to construct our main data structure: simplicial complexes.

**Definition 2.** *Let $v_0, v_1, \ldots, v_i$ be $i + 1$ affinely independent points in $\mathbb{R}^n$ with $(i \leq n)$. An $i$-simplex, $\sigma = (v_0, v_1, \ldots, v_i)$, is the convex set*

$$\Big\{ x \in \mathbb{R}^n \mid x = \sum_{j=0}^{i} \lambda_j v_j \text{ with } \lambda_j \geq 0 \text{ and } \sum_{j=0}^{i} \lambda_j = 1 \Big\}.$$

*The points $v_0, v_1, \ldots, v_i$ are the vertices of $\sigma$. Let us denote by $\dim \sigma$ the dimension of $\sigma$ which in this case is $i$. We say that $\sigma'$ is a face of $\sigma$ if $\sigma'$ is an $i'$-simplex, with $i' \leq i$, and whose vertices are also vertices of $\sigma$.*

When several simplices are joint together, a more complex structure called simplicial complex is built. The following definition exhibits the way simplices can be glued together to obtain such a simplicial complex.

**Definition 3.** *A simplicial complex $K$ is a finite collection of simplices such that:*

   *1. $\sigma \in K$ and $\sigma' \subseteq \sigma$ implies $\sigma' \in K$;*

   *2. $\sigma, \mu \in K$ implies $\sigma \cap \mu$ is either empty or a face of both.*

*The underlying space of $K$, denoted by $|K|$ is the union of its simplices together with the topology inherited from the ambient Euclidean space where the simplices are placed. A subcomplex $L$ of $K$ is a simplicial complex such that $L \subseteq K$. The subcomplex of $K$ consisting of all simplices of $K$ of dimension $j$ or less is called the $j$-skeleton of $K$ and it is denoted by $K^{(j)}$. Let us remark that the 0-skeleton of $K$ is its vertex set.*

**Definition 4.** *Let $K$ be a simplicial complex. A simplex $\sigma \in K$ is maximal if it is not a face of any other simplex in $K$. A simplex $\sigma'$ is adjacent to $\sigma$ if $\sigma$ and $\sigma'$ share a face in $K$.*

Next, we recall the idea of star of a simplex $\sigma$ in a simplicial complex $K$. Intuitively, it is the set of simplices in $K$ which have $\sigma$ as a face.

**Definition 5.** *Let $K$ be a simplicial complex and $\sigma$ a face of $K$. The star of $\sigma$ in $K$, denoted by $\mathrm{st}(\sigma)$, is defined as:*

$$\mathrm{st}(\sigma) = \big\{ \mu \in K \mid \exists\, \xi \in K \text{ such that } \sigma \subseteq \xi \text{ and } \mu \subseteq \xi \big\}.$$

Simplicial complexes are combinatorial data structures used to "model" topological spaces. A way to obtain a refined model from an existing one is to subdivide it in small pieces such that the result is topologically equivalent to the former.

**Definition 6.** *Let $K$ and $K'$ be simplicial complexes. It is said that $K'$ is a subdivision of $K$ if:*

   *1. $|K| = |K'|$;*

   *2. $\sigma' \in K'$ implies that there exists $\sigma \in K$ such that $\sigma \subseteq \sigma'$.*

The barycentric subdivision is a concrete example of subdivision of simplicial complexes. Let us recall that the barycenter of an $i$-simplex $\sigma = (v_0 \ldots v_i)$ is

$$b(\sigma) = \sum_{j=0}^{i} \frac{1}{j+1} v_j.$$

By using this definition, the idea of barycentric subdivision of a simplicial complex arises in a natural way.

**Definition 7.** *Let $K$ be a simplicial complex. The barycentric subdivision of the 0-skeleton of $K$ is defined as the set of vertices of $K$, that is, $\mathrm{Sd}\, K^{(0)} = K^{(0)}$. Assuming we have $\mathrm{Sd}\, K^{(i-1)}$, which denotes the barycentric subdivision of the $(i-1)$-skeleton of $K$, $\mathrm{Sd}\, K^{(i)}$ is built by adding the barycenter of every $i$-simplex as a new vertex and connecting it to the simplices that subdivide the boundary of such $i$-simplex. The iterated application of barycentric subdivisions is denoted by $\mathrm{Sd}^t K$ where $t$ is the number of iterations.*
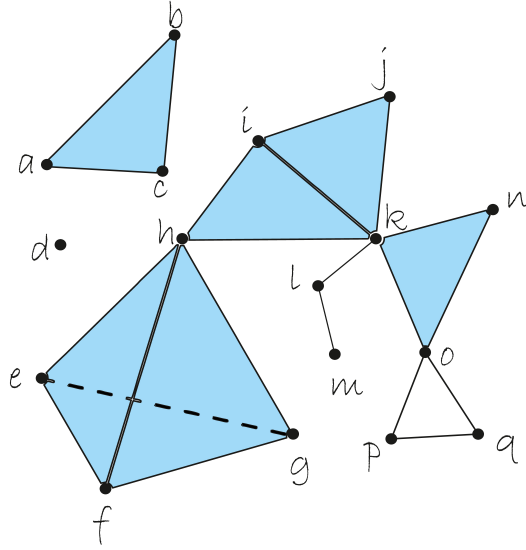
Figure 1: Example of simplicial complex. $(d)$ is a 0-simplex, $(o, p)$ is a 1-simplex , $(a, b, c)$ is a 2-simplex, and $(e, f, g, h)$ is a 2-simplex. $(a, b)$ is a face of the simplex $(a, b, c)$. The simplices $(e, f, g, h)$,$(a, b, c)$, $(i, h, k)$, $(i, j, k)$, $(o, p)$, $(o, q)$, $(p, q)$, and $(d)$ are the maximal simplices of the simplicial complex.
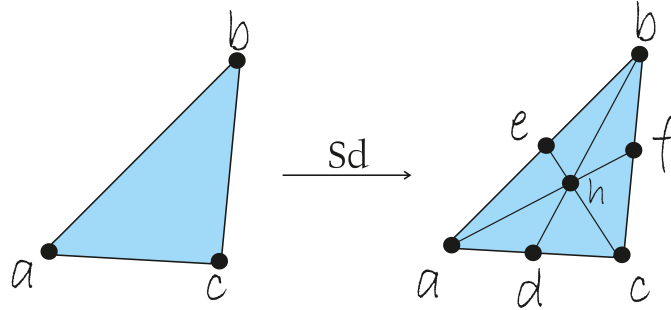


Figure 2: On the left the 2-simplex $(a, b, c)$ is shown, and on the right its first barycentric subdivision is plotted.

Let us see now how the "size" of the simplices of a simplicial complex can be measured.

**Definition 8.** *Let $K$ be a simplicial complex. The diameter of a simplex $\sigma$ in $K$ is defined as*

$$\delta(\sigma) = \max\{\|x - y\| \mid x, y \in \sigma\}$$

*and the mesh of $K$ is defined as*

$$m(K) = \max\{\delta(\sigma) \mid \sigma \in K\}$$

**Theorem 2** ([13, p. 86]). *Given a simplicial complex $K$, and $\epsilon > 0$, there exists an integer $t > 0$ such that $m(\mathrm{Sd}^t K) \le \epsilon$.*

Next, we recall one of the main tools used in this paper. A simplicial complex can be seen as mathematical model of a region of an $n$-dimensional space. By using subdivisions, more and more tuned simplicial complexes can be obtained. Let us now think about maps between simplicial complexes. These maps can be considered as extensions of simpler maps defined from vertices to vertices of two simplicial complexes. Interestingly, such maps can be considered as approximations of continuous functions defined on the underlying topological spaces that the simplicial complexes are modeling. Let us formalize these notions.

4

**Definition 9.** *Let $K$ and $L$ be two simplicial complexes. A vertex map is a function $\varphi : K^{(0)} \to L^{(0)}$ with the property that the vertices of every simplex in $K$ map to vertices of a simplex in $L$.*

A vertex map $\varphi$ can be extended to a continuous map $\varphi_c : |K| \to |L|$ in the following way.

**Definition 10.** *Let $K$ and $L$ be two simplicial complexes and let $\varphi : K^{(0)} \to L^{(0)}$ be a vertex map. The simplicial map $\varphi_c$ induced by $\varphi$ is defined as follows. Let $x \in |K|$. Then there exists an $i$-simplex $\sigma = (v_0, \ldots, v_i))$ in $K$ and numbers $\lambda_j \geq 0$ such that*

$$\sum_{j=0}^{i} \lambda_j = 1 \text{ and } x = \sum_{j=0}^{i} \lambda_j v_j.$$

*Then*

$$\varphi_c(x) = \sum_{j=0}^{i} \lambda_j \varphi(v_j).$$

Any vertex map $\varphi$ induces a simplicial map $\varphi_c$, but, if we want that such map is an approximation between the underlying spaces of both simplicial complexes, a restriction about the star of each simplex needs to be added.

**Definition 11.** *Let $K$ and $L$ be simplicial complexes and let $g : |K| \to |L|$ be a continuous map. A simplicial map $\varphi_c : |K| \to |L|$ induced by a vertex map $\varphi : K^{(0)} \to L^{(0)}$ is a simplicial approximation if*

$$g(|\operatorname{st}(v)|) \subset |\operatorname{st}(\varphi(v))|$$

*for each vertex $v$ of $K$.*

Furthermore, there exists strong results on simplicial complexes. Concretely, we focus our attention in the Simplicial Approximation Theorem which ensures the existence of simplicial maps that approximate continuous maps arbitrarily closely.

**Theorem 3** (Simplicial Approximation Theorem [12, p. 56]). *If $g : |K| \to |L|$ is continuous, then there is a sufficiently large integer $t > 0$ such that $\varphi_c : |Sd^t K| \to |L|$ is a simplicial approximation of $g$.*

Theorems 1 and 3 are key results in their respective fields. Our aim is to take Theorem 3 as a pillar for obtaining a result close to Theorem 1. Roughly speaking, the idea is to consider a simplicial approximation between two simplicial complexes. On the one hand, such simplicial approach is characterized via a vertex map which can be expressed as a neural network. On the other hand, the simplicial approximation can be chosen in such way that it approximates the continuous map between the underlying spaces of the simplicial complexes. Since the parameters of the neural network can be effectively obtained from the vertex map, this method provides a constructive way to find a neural network which approximates a continuous function on the underlying spaces of two simplicial complexes.

Following that aim, let us formalize the relationship between topological spaces and simplicial complexes.

**Definition 12.** *A triangulation of a topological space $X$ consists of a simplicial complex $K$ and a homeomorphism $\tau : X \to |K|$.*

The spaces that can be triangulated by simplicial complexes (see [16, Theorem A.7, p. 525]) are every compact, locally contractible that can be embedded in $\mathbb{R}^n$ for some $n$. Let us remark that the Universal Approximation Theorem (Theorem 1) is valid for any compact subset of $\mathbb{R}^n$, regardless if they are locally contractible or not. Not all compact sets in a metric space are locally contractible (see [17, Chapter 17.7, p. 426]). Nevertheless, non-locally contractible spaces are very odd in $\mathbb{R}^n$ and this technical topological property has no practical application in real-world problems. Therefore, we can say that the results proved on triangulable spaces are true on a large amount of current neural network problems.

Next, we can extend the definition of a mesh of a simplicial complex in the following way.

**Definition 13.** *Let $X$ be a triangulable topological space and $(K, \tau)$ a triangulation of $X$. The mesh of $X$ induced by $(K, \tau)$ is defined as*

$$\tilde{m}_{(K,\tau)}(X) = \max\{\tilde{\delta}(\sigma) \mid \sigma \in K\}$$

*where $\tilde{\delta}(\sigma) = \max\left\{||x - y|| \text{ such that } x = \tau^{-1}(a), \ y = \tau^{-1}(b) \text{ with } a, b \in \sigma\right\}$ is the extended diameter of a simplex.*

## 3   Multilayer Feedforward Networks and Simplicial Maps

In this section, we will show that such simplicial maps can be modelled via multilayer feedforward networks in a straightforward way.

In the following theorem, we will compute a two-hidden-layer feedforward network to a simplicial map $\varphi_c : |K| \to |L|$ where all the simplices of $K$ and $L$ are maximal simplices with maximal dimension. This is not an important constraint in our case, since our final aim is to design a multilayer feedforward network that approximates a continuous function between triangulable spaces.

**Theorem 4.** *Given a simplicial map $\varphi_c : |K| \to |L|$ where $K$ and $L$ are composed by maximal simplices with maximal dimension. A two-hidden layer feedforward network $\mathcal{N}_\varphi$ such that $\varphi_c(x) = \mathcal{N}_\varphi(x)$ for all $x \in |K|$ can be explicitly defined.*

*Proof.* Let us assume $|K| \subset \mathbb{R}^n$ and $|L| \subset \mathbb{R}^m$. Besides, let us consider that $K$ is composed by $k$ maximal $n$-simplices $\{\sigma_1, \ldots \sigma_k\}$ where $\sigma_i = [v_0^i, \ldots, v_n^i]$ for all $i$, and $L$ by $\ell$ maximal $m$-simplices $\{\mu_1, \ldots, \mu_\ell\}$ where $\mu_i = [u_0^j, \ldots, u_m^j]$ for all $j$. Let us consider a multilayer feedforward network $\mathcal{N}_\varphi$ with the following architecture: (1) An input layer composed by $d_0 = n$ neurons; (2) a first hidden layer composed by $d_1 = k \cdot (n + 1)$ neurons where ; (3) a second hidden layer composed by $d_2 = \ell \cdot (m + 1)$ neurons; and (4) an output layer with $d_3 = m$ neurons. Then, $\mathcal{N}_\varphi = f_3 \circ f_2 \circ f_1$ being $f_i(y) = \phi_i(W^{(i)}; y; b_i)$, $i = 1, 2, 3$. The idea is to codify the simplicial complexes involved in the mapping in the hidden layers of the multilayer feedforward network. Firstly, a point $x$ in $\mathbb{R}^n$ is transformed into a $k \cdot (n + 1)$ vector. This vector can be seen as the juxtaposition of $k$ vectors of dimension $n + 1$, one for each of the $k$ simplices in $K$. Each vector of dimension $k$ represents the barycentric coordinates of $x$ with respect to the corresponding simplex. The matrix of the weights to obtain the output of the first hidden layer, $W^{(1)}$, and the corresponding bias term $b_1$ can be obtained from the equations for getting the barycentric coordinates as follows:

$$W^{(1)} = \begin{pmatrix} W_1^{(1)} \\ \vdots \\ W_k^{(1)} \end{pmatrix} \in \mathcal{M}_{k(n+1) \times n},$$

where $W_i^{(1)} \in \mathcal{M}_{(n+1) \times n}$ is:

$$\begin{pmatrix} v_0^i & \cdots & v_n^i \\ 1 & \cdots & 1 \end{pmatrix}^{-1} = \left( W_i^{(1)} \mid B_i \right),$$

and $B_i \in \mathbb{R}^{n+1}$ is:

$$b_1 = \begin{pmatrix} B_1 \\ \vdots \\ B_k \end{pmatrix};$$

The matrix of weights between the first and the second hidden layer encodes the vertex map between vertices. Since $K$ is encoded in the first hidden layer with $k \cdot (n + 1)$ neurons and $L$ is represented via $\ell \cdot (m + 1)$ neurons, the matrix is composed by 0's and 1's. The value 1 represents that the corresponding vertices are related via the vertex map, and 0 represents that they do not. Then, this matrix, $W^{(2)}$, can be defined as follows:

$$W^{(2)} = \left( W_{s_1, s_2}^{(2)} \right) \in \mathcal{M}_{\ell(m+1) \times k(n+1)},$$

being $s_1 = j(r + 1)$ and $s_2 = i(t + 1)$ for $i = 1, \ldots, k$; $j = 1, \ldots, \ell$; $t = 0, \ldots, n$; and $r = 0, \ldots, m$. Finally,

$$W_{s_1, s_2}^{(2)} = \begin{cases} 1 \text{ if } \varphi(v_t^i) = w_r^j, \\ 0 \text{ otherwise}; \end{cases}$$

and $b_2 = 0$.

The output of the second hidden layer can be seen as the juxtaposition of $\ell$ vectors of dimension $m + 1$, one vector for each simplex in the simplicial complex $L$. Each of the vectors represents the barycentric coordinates of $\varphi_c(x)$ with respect to the corresponding simplex. For each simplex in $L$, the corresponding barycentric coordinates are a vector of $m + 1$ components whose sum is 1, but all coordinates are greater than or equal to zero if $\varphi_c(x)$ belongs to that simplex. Only those coordinates are considered, and, in the next step, $\phi_3(W^{(3)}; y; b_3)$ transforms this vector and the outputs the Cartesian coordinates of $\varphi_c(x)$. Then:

$$W^{(3)} = \begin{pmatrix} W_1^{(3)} & \cdots & W_\ell^{(3)} \end{pmatrix} \in \mathcal{M}_{m \times \ell(m+1)},$$

being $W_j^{(3)} = \begin{pmatrix} u_0^j & \cdots & u_m^j \end{pmatrix}; b_3 = 0$.

And, finally, the following functions are defined. For the first two layers:

$$\phi_i(W^{(i)}; y; b_i) = W^{(i)}y + b_i \text{ for } i = 1, 2.$$

In the case of the third layer, $\phi_3$ is defined as follows:

$$\phi_3(W^{(3)}; y; b_3) = \frac{\sum_{j=1}^\ell z_j \psi(y_j)}{\sum_{j=1}^\ell y_j}$$

with $z_j = W_j^{(3)} \cdot y_j$ , and $y = \begin{pmatrix} y_1 \\ \vdots \\ y_\ell \end{pmatrix} \in \mathcal{M}^{\ell \cdot (m+1)}$ where $y_j \in \mathbb{R}^{m+1}$ and $\psi(y_j) = 1$ if all the coordinates of $y_j$ are greater than or equal to 0, and $\psi(y_j) = 0$ otherwise.

$\square$

This proposition establishes that two-hidden-layer feedforward network can act equivalently to simplicial maps. Besides, the architecture and the specific computation of the parameters is shown.

## 4   Simplicial Approximation Theorem extension

In this section, we provide an extension of the simplicial approximation theorem, and an explicit algorithm to compute a simplicial approximation *as close as we want to* a given continuous function $g : |K| \to |L|$ between the underlying spaces of, respectively, two simplicial complexes $K$ and $L$. The first observation is that the Simplicial Approximation Theorem (Theorem 3) refers to any continuous mapping. Continuity is a property of functions in topological spaces, not necessarily metric spaces. The next result introduces the concept of metric into the simplicial approximations.

**Proposition 1.** *Given $\epsilon > 0$ and a continuous function $g : |K| \to |L|$ between the underlying spaces of two simplicial complexes $K$ and $L$, there exists $t_1, t_2 > 0$ such that $\varphi_c : |\mathrm{Sd}^{t_1} K| \to |\mathrm{Sd}^{t_2} L|$ is a simplicial approximation of $g$ and $||g - \varphi_c|| \le \epsilon$.*

*Proof.* By Theorem 2, there exists $t_2$ such that $m(\mathrm{Sd}^{t_2} L) \le \epsilon$. Then, by Theorem 3, there exist $t_1$ such that $\varphi_c : |\mathrm{Sd}^{t_1} K| \to |\mathrm{Sd}^{t_2} L|$ is a simplicial approximation of $g$:

$$
\begin{array}{ccc}
|K| & \xrightarrow{\ g\ } & |L| \\
\uparrow & & \uparrow \\
K & & L \\
\downarrow {\scriptstyle \mathrm{Sd}} & & \downarrow {\scriptstyle \mathrm{Sd}} \\
\mathrm{Sd}^{t_1} K & & \mathrm{Sd}^{t_2} L \\
\downarrow & & \downarrow \\
|\mathrm{Sd}^{t_1} K| & \xrightarrow{\ \varphi_c\ } & |\mathrm{Sd}^{t_2} L|
\end{array}
$$

Besides, $||g - \varphi_c|| \le \epsilon$ because $m(\mathrm{Sd}^{t_2} L) \le \epsilon$.

$\square$

Algorithm 1 is inspired in the proof of the Simplicial Approximation Theorem in [12, p.56] and computes a vertex map $\varphi : (Sd^t K)^{(0)} \to (SdL)^{(0)}$ from which the simplicial approximation $\varphi_c : |Sd^t K| \to |L|$ of a given function $g : |K| \to |L|$ between the underlying spaces of, respectively, two simplicial complexes $K$ and $L$ can be defined.

---

**Algorithm 1:** Computing simplicial approximations

---

**Input:** A continuous function $g : |K| \to |L|$ between the underlying spaces of two simplicial complexes $K$ and $L$, and an integer $t$ where $Sd^t k$ satisfies the *star condition*: $g(|st(v)|) \subseteq |st(w)|$ for all $v \in (Sd^t k)^{(0)}$ and for some $w \in L^{(0)}$.

**Output:** A simplicial approximation $\varphi_c$ of $g$.

**foreach** *point* $v \in (\mathrm{Sd}^t K)^{(0)}$ **do**

    ⌊ Let $\varphi(v) = w$ that exists by the star condition.

---

**Theorem 5.** *Given a continuous function $g : |K| \to |L|$ and $\epsilon > 0$, a two-hidden-layer feedforward network $\mathcal{N}$ such that $||\mathcal{N} - g|| \leq \epsilon$ can be explicitly defined.*

*Proof.* By proposition 1, there exists a simplicial approximation $\varphi_c$ of $g$ such that $||g - \varphi_c|| \leq \epsilon$. Then, by Theorem 4 there exists $\mathcal{N}_\varphi$ such that $\varphi_c = \mathcal{N}_\varphi$.

$\square$

## 5 Universal Approximation Theorem Extension

The results of the previous section can be extended to triangulable spaces. In this case, the approach is constructive if the homeomorphisms of the triangulations are known. In the previous sections, we have proved that a continuous function between triangulable sets can be approximated by using the simplicial approximating theorem. In this section, using the new version of the simplicial approximation theorem (Proposition 3) we can give a constructive version of the Universal Approximation Theorem that approximates any continuous function (under some conditions) arbitrarily close.

**Proposition 2.** *Let $(K, \tau)$ be a triangulation of a space $X$. For all $\epsilon > 0$ there exists $n$ and $\gamma > 0$ such that if $m(\mathrm{Sd}^n K) \leq \gamma$ then $\tilde{m}_{(\mathrm{Sd}^n K, \tau_K)}(X) \leq \epsilon$.*

*Proof.* Consider $x, y_0 \in |K|$. If $x$ and $y_0$ belong to $\sigma_0 \in K$. Then, $||x - y_0|| \leq \delta(\sigma_0)$ and $||\tau(x) - \tau(y_0)|| \leq \tilde{\delta}(\sigma_0)$. Otherwise, we repeat the reasoning with $SdK$. Now, $x$ and $y_0$ can belong to the same simplex in $SdK$ or not. If they belong to the same simplex, write $y_1 = y_0$. If not, take a new point $y_1$ such that $x, y_1 \in \sigma_1$ and $\sigma_1 \in \mathrm{Sd}\,\sigma_0$. Therefore, $||x - y_1|| \leq \delta(\sigma_1) \leq \delta(\sigma_0)$. Besides, $||\tau(x) - \tau(y_1)|| \leq \tilde{\delta}(\sigma_1) \leq \tilde{\delta}(\sigma_0)$. This can be iterated: $||x - y_t|| \leq \delta(\sigma_n) \leq \cdots \leq \delta(\sigma_1) \leq \delta(\sigma_0)$ and $||\tau(x) - \tau(y_t)|| \leq \tilde{\delta}(\sigma_t) \leq \cdots \leq \tilde{\delta}(\sigma_1) \leq \tilde{\delta}(\sigma_0)$. By this, we have defined a sequence $\{y_i\}_{i=0}^t$ that converges to $x$. Therefore, given $\epsilon > 0$, there exists $n$ such that $||\tau(x) - \tau(y_n)|| \leq \epsilon$. Let us suppose, without loss of generality, that $\tilde{\delta}(\sigma_n) = \tilde{m}_{(\mathrm{Sd}^n K, \tau)}(X)$. Then, we can consider $\gamma = m(\mathrm{Sd}^n K)$.

$\square$

**Corollary 1.** *Given $\epsilon > 0$ and a triangulation $(K, \tau)$ of $X$, there exists $t$ such that $\tilde{m}(\mathrm{Sd}^t K) \leq \epsilon$.*

*Proof.* By Theorem 2 there exists $t'$ such that $m(\mathrm{Sd}^{t'} K) \leq \gamma$. Then, by Proposition 2, there exists $t$ such that $\tilde{m}_{(\mathrm{Sd}^t K, X)} \leq \epsilon$.

$\square$

Finally, we reach the main result of this section. Given a continuous map $g$ between two triangulable spaces $X$ and $Y$, we can obtain two simplicial complexes $K$ and $L$ associated to them, and a simplicial approximation $\varphi_c$ between them which approximates $g$.

**Proposition 3.** *Let $X$ and $Y$ be two triangulable topological spaces, $g : X \to Y$ a continuous map, and $\epsilon > 0$. Then, there exists two triangulations $(K, \tau_K)$ and $(L, \tau_L)$ of $X$ and $Y$, respectively, and a simplicial approximation $\varphi_c : |Sd^{t_1} K| \to |Sd^{t_2} L|$ such that $||g - \tau_L^{-1} \circ \varphi_c \circ \tau_K|| \leq \epsilon$.*

*Proof.* By Corollary 1, there exists $t_2$ such that $\tilde{m}_{(\mathrm{Sd}^{t_2} \tilde{L}, \tau_L)}(Y) \leq \epsilon$. Next, by Theorem 3, there exists $t_1$ and $\tilde{\varphi}$ such that $\tilde{\varphi} : \mathrm{Sd}^{t_1} \tilde{K} \to \mathrm{Sd}^{t_2} \tilde{L}$ is a simplicial approximation of $\tau_{\tilde{L}} \circ g \circ \tau_{\tilde{K}}^{-1}$. Taking into account that $|\mathrm{Sd}^{t_1} \tilde{K}| = |\tilde{K}|$ and $|\mathrm{Sd}^{t_2} \tilde{L}| = |\tilde{L}|$, we can define the

simplicial map $\tilde{\varphi}_c : |\tilde{K}| \to |\tilde{L}|$ induced by $\tilde{\varphi}$. Finally, we can define $\varphi_c = \tau_{\tilde{L}}^{-1} \circ \tilde{\varphi}_c \circ \tau_{\tilde{K}}$, $K = \mathrm{Sd}^{t_1} \tilde{K}$, and $L = \mathrm{Sd}^{t_2} \tilde{L}$. Since $\tilde{m}_{(L,\tau_L)}(Y) \leq \epsilon$ then $||g - \tau_L^{-1} \circ \varphi_c \circ \tau_K|| \leq \epsilon$

$$
\begin{array}{ccc}
X & \xrightarrow{\ \ g\ \ } & Y \\
\downarrow{\scriptstyle \tau_{\tilde{K}}} & & \downarrow{\scriptstyle \tau_{\tilde{L}}} \\
|\tilde{K}| & \xrightarrow{\ \ \tilde{\varphi}_c\ \ } & |\tilde{L}| \\
\uparrow & & \uparrow \\
\tilde{K} & & \tilde{L} \\
\downarrow{\scriptstyle Sd} & & \downarrow{\scriptstyle Sd} \\
Sd^{t_1}\tilde{K} = K & \xrightarrow{\ \ \tilde{\varphi}\ \ } & Sd^{t_2}\tilde{L} = L
\end{array}
$$

□

**Theorem 6.** *Given a continuous function $g : X \to Y$ and two triangulations $(K, \tau_K)$ and $(L, \tau_L)$ of $X$ and $Y$, respectively. A two-hidden-layer feedforward network $\mathcal{N}$ that approximates $g$ arbitrarily close can be explicitly defined.*

*Proof.* By Proposition 3, there exists a simplicial approximation $\varphi_c$ of $g$ that approximates $g$ arbitrarily close. Besides, by Theorem 4, there exists $\mathcal{N}$ such that $\mathcal{N} = \varphi_c$ in all the domain. Therefore, $\mathcal{N}$ approximates $g$ arbitrarily close.

□

## 6 Conclusions

In this paper, we provide an effective method for building a multilayer feedforward network which approximates a continuous function between triangulable spaces. The main contribution of the paper is that the weights can be exactly and effectively computed without any training process. On the disadvantages, two can be considered from a technical point of view: Firstly, the homeomorfisms between the triangulable spaces and the simplicial complexes can be hard to find, and secondly, the classic theorem for approximation of neural networks is valid for compact sets, and our result is only valid for triangulable sets. Nonetheless, as pointed above, most of the real-world problems are covered by our result and, therefore the approximation to the continuous function via neural network can be effectively built.

## References

[1] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control Signal and Systems*, 2(4):303–314, 1989.

[2] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, March 1991.

[3] Shiyu Liang and R. Srikant. Why deep neural networks? *CoRR*, abs/1610.04161, 2016.

[4] Itay Safran and Ohad Shamir. Depth-width tradeoffs in approximating natural functions with neural networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2979–2987, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.

[5] Matus Telgarsky. Benefits of depth in neural networks. *CoRR*, abs/1602.04485, 2016.

[6] Shizhao Sun, Wei Chen, Liwei Wang, Xiaoguang Liu, and Tie-Yan Liu. On the depth of deep neural networks: A theoretical view. In Dale Schuurmans and Michael P. Wellman, editors, *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.*, pages 2066–2072. AAAI Press, 2016.

[7] Boris Hanin and Mark Sellke. Approximating continuous functions by relu nets of minimal width, 2017.

[8] Zhou Lu, Hongming Pu, Feicheng Wang, Zhiqiang Hu, and Liwei Wang. The expressive power of neural networks: A view from the width. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 6232–6240, 2017.

[9] Quynh Nguyen, Mahesh Chandra Mukkamala, and Matthias Hein. Neural networks should be wide enough to learn disconnected decision regions. *CoRR*, abs/1803.00094, 2018.

[10] Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.

[11] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:356–366, 1989.

[12] Herbert Edelsbrunner and John Harer. *Computational Topology - an Introduction*. American Mathematical Society, 2010.

[13] James R. Munkres. *Elements of algebraic topology*. Addison-Wesley, 1984.

[14] R. Ayala, E. Domínguez, and A. Quintero. *Elementos de la teoría de homología clásica*. Ciencias (Universidad de Sevilla). Secretariado de Publicaciones, Universidad de Sevilla, 2002.

[15] J.D. Boissonnat, F. Chazal, and M. Yvinec. *Geometric and Topological Inference*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2018.

[16] Allen Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.

[17] R. Geoghegan. *Topological Methods in Group Theory*. Graduate Texts in Mathematics. Springer New York, 2010.