# Polytope neural networks*

Juan L. Valerdi†

## 1 Introduction

A major challenge in the theory of neural networks is to precisely characterize the functions they can represent [2]. This topic differs from universal approximation theorems [1], which aim to guarantee the existence of neural networks that approximate functions well. Although it is well known that feedforward neural networks with ReLU activation are continuous piecewise linear (CPWL) functions [2, 3], the minimum number of layers required to represent any CPWL function remains an open question.

A potential way to solve this problem is through the concept of depth of a polytope given by neural networks.

**Definition 1.** *The collection of polytope neural networks with depth $m$ is defined as*

$$\Delta(m) = \left\{ \sum_{i=1}^{p} conv\{P_i, Q_i\} \;\middle|\; P_i, Q_i \in \Delta(m-1) \right\},$$

*where the sum corresponds to Minkowski sum and $conv\{P_i, Q_i\}$ means the convex hull of $P_i \cup Q_i$. The base set $\Delta(0)$ represents the polytopes consisting of a single point.*

**Definition 2.** *A polytope $P$ is said to have (minimal) depth $m$, denoted as $d(P) = m$, if $P \in \Delta(m)$ and $P \notin \Delta(m-1)$.*

Neural networks are traditionally named after their building object or operation. For example, ReLU neural networks use ReLU activation, and convolutional neural networks [3] are based on convolution kernels. In a similar manner, the naming of polytope neural networks is derived from their underlying object.

The connection between ReLU and polytope neural networks can be found through tropical geometry [10]. Any ReLU network can be decomposed into the difference of two convex CPWL functions, which can be mapped to polytopes via Newton polytopes. In particular, understanding the functions representable by ReLU neural networks of a given depth is equivalent to studying which polytopes can be constructed at that depth, as defined in Definition 1.

The open question for ReLU networks reduces to whether the function $\max\{x_1, x_2, \ldots, x_n, 0\}$ can be represented with minimal depth $\lceil \log_2(n+1) \rceil$. This question can be rephrased in the language of polytopes as follows.

**Conjecture 3** (Hertrich et al. [6])**.** *Let $S$ be an $n$-simplex, then $d(S+P) = \lceil \log_2(n+1) \rceil$, for any polytope $P$ with $d(P) < \lceil \log_2(n+1) \rceil$.*

*The full version of this work can be found in [8] and will be published elsewhere.

†Email: j.valerdi11@gmail.com

Our understanding of the sets $\Delta(m)$, beyond the case of $m = 1$, which corresponds to the set of zonotopes, remains limited. The conjecture is known to be true for $n = 2$ and $n = 3$ [2, 7]. However, to this date, the only contribution addressing Conjecture 3 for any $n$ has been made by Haase et al. [5], who have proven it for lattice polytopes. Their approach involved relating depth with subdivision and volume properties of Minkowski sums and convex hulls.

The goal of this work is to advance our knowledge of polytope neural networks relevant to Conjecture 3. We show basic depth properties from Minkowski sums, convex hulls, number of vertices, faces, affine transformations, and indecomposable polytopes. More significantly, key findings include depth characterization of polygons; identification of polytopes with an increasing number of vertices, exhibiting small depth and others with arbitrary large depth; and most importantly, depth computation for simplices.

**Acknowledgements.** I extend my gratitude to Ansgar Freyer for providing the proof of Theorem 12 for $n = 4$, which was expanded to the general case with minor adjustments. I also thank Francisco Santos for his hospitality during my visits to the University of Cantabria, and for valuable discussions on this work, including presentation enhancements and the ideation and proof of Theorem 14.

## 2 Basic properties

To develop the main results in Section 3, it is necessary to establish some basic depth properties for polytopes. We assume $\mathbb{R}^n$ as the ambient space throughout.

We begin by computing depth bounds for Minkowski sums and convex hulls, which are the fundamental operations in Definition 1.

**Proposition 4.** *Let $P_1, P_2$ be polytopes with $d(P_i) \leq m_i$. Then, $d(P_1 + P_2) \leq \max\{m_1, m_2\}$ and $d(conv\{P_1, P_2\}) \leq \max\{m_1, m_2\} + 1$.*

*Proof.* If $d(P_i) \leq m_i$, then $P_i \in \Delta(\max\{m_1, m_2\})$. This implies $\text{conv}\{P_1, P_2\} \in \Delta(\max\{m_1, m_2\} + 1)$ by definition, and therefore $d(\text{conv}\{P_1, P_2\}) \leq \max\{m_1, m_2\} + 1$.

Also by definition, consider the decomposition

$$P_i = \sum_{j=1}^{q_i} \text{conv}\{Q_{j,i}, R_{j,i}\},$$

where $Q_{j,i}, R_{j,i} \in \Delta(\max\{m_1, m_2\} - 1)$ for all $i = 1, 2$ and $j = 1, \ldots, q_i$. Consequently, $d(P_1 + P_2) \leq \max\{m_1, m_2\}$ as

$$P_1 + P_2 = \sum_{j=1}^{q_1} \text{conv}\{Q_{j,1}, R_{j,1}\} + \sum_{j=1}^{q_2} \text{conv}\{Q_{j,2}, R_{j,2}\} \in \Delta(\max\{m_1, m_2\}). \qquad \square$$

Now, using Proposition 4 we can bound the depth of a polytope by its vertices.

**Proposition 5.** *If a polytope $P$ is given by its vertices $P = conv\{x_1, \ldots, x_p\}$, then $d(P) \leq \lceil \log_2 p \rceil$.*

*Proof.* By definition, $d(\{x_1\}) = 0$ and $d(\text{conv}\{x_1, x_2\}) = 1$. Supposing the statement is true up to $p - 1$, consider a polytope $P = \text{conv}\{x_1, \ldots, x_p\}$ and decompose it as

$$P = \text{conv}\{\text{conv}\{x_1, \ldots, x_k\}, \text{conv}\{x_{k+1}, \ldots, x_p\}\},$$

where $k$ is the largest integer power of 2 such that $k < p$. Using the induction hypothesis, we obtain that $d(\text{conv}\{x_1, \ldots, x_k\}) \leq \log_2 k$ and $d(\text{conv}\{x_{k+1}, \ldots, x_p\}) \leq \lceil \log_2(p-k) \rceil$. Therefore, by Proposition 4, we conclude $d(P) \leq \log_2 k + 1 = \lceil \log_2 p \rceil$. $\square$

Other basic properties concerns the depth of a polytope in relation to its faces and affine transformations.

**Proposition 6.** *Any face $F \neq \emptyset$ of a polytope $P$ satisfies $d(F) \leq d(P)$.*

*Proof.* For $d(P) = 0$, there is nothing to prove. If $d(P) = 1$, then $P$ is a zonotope, and any face $F$ is also a zonotope; therefore, $d(F) \leq 1$. For the sake of induction, suppose the statement is true up to depth $m - 1$ and consider $d(P) = m$. By definition,

$$P = \sum_{i=1}^{q} \mathrm{conv}\{P_i, Q_i\}, \quad P_i, Q_i \in \Delta(m-1).$$

A face $F$ of $P$ is then expressed as

$$F = \sum_{i=1}^{q} \mathrm{conv}\{F_i, G_i\},$$

where $F_i, G_i$ are faces of $P_i, Q_i$ respectively. By the induction hypothesis, $d(F_i) \leq m - 1$ and $d(G_i) \leq m - 1$, and consequently $F_i, G_i \in \Delta(m-1)$ for all $i$. Therefore, $F \in \Delta(m)$ and $d(F) \leq d(P)$. □

**Proposition 7.** *Let $P$ be a polytope in $\mathbb{R}^n$ and $\varphi : \mathbb{R}^n \to A$ be an affine transformation, where $A$ is an affine subspace of $\mathbb{R}^d$. Then, $d(\varphi(P)) \leq d(P)$, with equality holding if $\varphi$ is invertible.*

*Proof.* Let $\varphi(x) = Mx + c$, where $M \in \mathbb{R}^{d\times n}$ and $c \in \mathbb{R}^d$. For the case $d(P) = 0$ consider $P = \{a\}$, then $\varphi(P) = \{Ma + c\}$, which implies $d(\varphi(P)) = 0$. For the purpose of induction, assume that the statement, in the general case, is true up to $m - 1$. Let $d(P) = m$ and express it as

$$P = \sum_{i=1}^{p} \mathrm{conv}\{P_i, Q_i\}, \quad P_i, Q_i \in \Delta(m-1).$$

Then,

$$\varphi(P) = \varphi\left(\sum_{i=1}^{p} \mathrm{conv}\{P_i, Q_i\}\right) = M \sum_{i=1}^{p} \mathrm{conv}\{P_i, Q_i\} + c = \sum_{i=1}^{p} \mathrm{conv}\{MP_i, MQ_i\} + \{c\}.$$

Utilizing the induction hypothesis and Proposition 4, we deduce that $d(\varphi(P)) \leq m$. In the case of $\varphi$ being invertible, we get

$$d(P) = d(\varphi^{-1}(\varphi(P))) \leq d(\varphi(P)) \leq d(P).$$ □

A class of polytopes in which computing their depth may be easier is that of indecomposable polytopes. Two polytopes, $P$ and $Q$, are said to be *positively homothetic*, if $P = \lambda Q + w$ for some $\lambda > 0$ and $w \in \mathbb{R}^n$. A polytope $P$ is said to be *indecomposable* if any decomposition $P = \sum_{i=1}^{k} P_i$ is only possible when $P_i$ is positively homothetic to $P$ for all $i = 1, \dots, k$.

**Proposition 8.** *If $P$ is an indecomposable polytope, then there exist polytopes $P_1, P_2$ such that $P =$ conv$\{P_1, P_2\}$ and $d(P) = \max\{d(P_1), d(P_2)\} + 1$.*

*Proof.* By definition, there exist $P_i, Q_i \in \Delta(d(P) - 1), i = 1, \dots, k$, such that

$$P = \sum_{i=1}^{k} \mathrm{conv}\{P_i, Q_i\},$$

where an index $j$ necessarily satisfies $\max\{d(P_j), d(Q_j)\} + 1 = d(P)$. As $P$ is indecomposable, there exist $\lambda_j > 0$ and $w_j \in \mathbb{R}^n$ such that $P = \lambda_j \mathrm{conv}\{P_j, Q_j\} + w_j = \mathrm{conv}\{\lambda_j P_j + w_j, \lambda_j Q_j + w_j\}$, and by Proposition 7,

$$d(P) = \max\{d(P_j), d(Q_j)\} + 1 = \max\{d(\lambda_j P_j + w_j), d(\lambda_j Q_j + w_j\}) + 1.$$ □

## 3 Main results

We first present a full depth characterization for polygons.

**Theorem 9.** *Any polygon $P$ satisfies $d(P) \leq 2$.*

*Proof.* Let $P$ be a polygon. If $P$ is a zonotope, then $d(P) = 1$; whereas, if $P$ is a triangle, then $d(P) = 2$ due to Proposition 5 and the fact that $P$ is not a zonotope. Suppose that $P$ is neither a zonotope nor a triangle; then, it can be decomposed as $P = \sum_{i=1}^{k} P_i$, where $P_i$ is a zonotope or a triangle for all $i = 1, \ldots, k$ [4]. Therefore, $d(P) \leq 2$ by Proposition 4. $\square$

From Theorem 9, we deduce that a polygon can have depth 0 if it consists of a single point, depth 1 if it is a zonotope, or depth 2 otherwise.

We continue with zonotopes and (bi)pyramids, as example of polytopes which can have large number of vertices and small depth.

**Proposition 10.** *Any $n$-(bi)pyramid, $n \geq 3$, with a zonotope base has depth 2.*

*Proof.* A 3-(bi)pyramid $P$ includes triangular facets, therefore it is not a zonotope, and thus $d(P) \geq 2$. Assuming that up to $n-1$, (bi)pyramids has depth greater than or equal to 2, let's consider a facet $F$ of an $n$-(bi)pyramid $P$ containing an/the apex. Since $F$ is a pyramid of dimension $n-1$, then $d(F) \geq 2$ based on the induction hypothesis. Consequently, $d(P) \geq d(F) \geq 2$ by Proposition 6.

Now, consider $P$ an arbitrary $n$-(bi)pyramid with a zonotope base $Z$ and apex (or apices) $A$. Then, $2 \leq d(P) = d(\mathrm{conv}\{Z, \mathrm{conv}\ A\}) \leq 2$ according to Proposition 4. $\square$

**Theorem 11.** *Let $v_p = 2\sum_{i=0}^{n-1} \binom{p-1}{i}$ for $p \geq n$. For each $p$ satisfying this condition, there exist polytopes with $v_p$ vertices and depth 1, and also with $v_p + 1$ vertices and depth 2.*

*Proof.* Let $g_i = [\mathbf{0}, b_i]$, where $i = 1, \ldots, p$, represent line segments with $b_1, \ldots, b_p$ denoting points in $\mathbb{R}^n$ in general position. The zonotope $Z = \sum_{i=1}^{p} g_i$ has depth 1 and has $v_p$ vertices given the generators are in general position [9]. Lifting $Z$ to $\mathbb{R}^{n+1}$ by adding 0 to the new coordinate allows the construction of a pyramid $P$ with $Z$ as its base. Therefore, $d(P) = 2$ by Proposition 10. $\square$

In Theorem 11, we constructed two families of polytopes, zonotopes and pyramids, which exhibit an increasing number of vertices and possess depths of 1 and 2, respectively. This indicates that depth bounds from Proposition 5 may be far from the true depth of a polytope. However, this bound based on vertices cannot be further refined, as it is tight for simplices.

We next present two approaches for calculating the depth of simplices. The first approach leverages the face structure and indecomposability of simplices, while the second approach results from a more general finding regarding polytopes containing complete subgraphs.

**Theorem 12.** *Any $n$-simplex has minimal depth $\lceil \log_2(n+1) \rceil$.*

*Proof.* We know that 2-simplices have depth 2. Let's make the assumption that, for $k = 3, \ldots, n-1$, $k$-simplices have depth $\lceil \log_2(k+1) \rceil$ and consider an $n$-simplex $P$. Given that $P$ is indecomposable [4], we can employ Proposition 8 to get a pair of polytopes $P_1, P_2$ such that $P = \mathrm{conv}\{P_1, P_2\}$ and $\max\{d(P_1), d(P_2)\} = d(P) - 1$.

Without loss of generality, one of the $P_i$, let's say $P_1$, contains at least $q = \lceil \frac{n+1}{2} \rceil$ points that are vertices of $P$. Consider $F = \mathrm{conv}\{x_1, \ldots, x_q\}$, where $x_i, i = 1, \ldots, q$ are vertices of $P$ contained in $P_1$. Then, $F$ is a $(q-1)$-simplex and a face of $P$. Let $H$ be a supporting hyperplane of $P$ associated with $F$. From

$$F = H \cap F \subset H \cap P_1 \subset H \cap P = F,$$

we deduce that $F$ is also a face of $P_1$. By the induction hypothesis,

$$d(F) = \left\lceil \log_2 \left\lceil \frac{n+1}{2} \right\rceil \right\rceil = \lceil \log_2(n+1) \rceil - 1$$

Referring to Proposition 5, Proposition 6, and Proposition 8, we derive that

$$\lceil \log_2(n+1) \rceil - 1 \leq d(P_1) \leq \max\{d(P_1), d(P_2)\} = d(P) - 1 \leq \lceil \log_2(n+1) \rceil - 1,$$

thus concluding that $d(P) = \lceil \log_2(n+1) \rceil$. □

For the second approach we will compute the depth of 2-neighbourly polytopes, for which we need the following result.

**Lemma 13.** *If the graph of a polytope $G(P)$ contains a complete subgraph with $p \geq 3$ vertices, and $P$ can be decomposed as $P = \sum_{i=1}^{k} P_i$, then at least one of $G(P_j)$ also contains a complete subgraph with $p$ vertices.*

*Proof.* Consider that $u, v, w$ are vertices of $P$ in the complete subgraph of $G(P)$ with $p \geq 3$ vertices. Given that any vertex of $P$ can be uniquely represented as the sum of vertices of $P_i, i = 1, \ldots, k$, let $u_i, v_i, w_i$ be those vertices for $P_i$ that represent $u, v, w$ respectively. Therefore, we can express the edges $[u, v], [u, w], [v, w]$ as

$$[u, v] = \sum_{i=1}^{k} [u_i, v_i], \quad [u, w] = \sum_{i=1}^{k} [u_i, w_i], \quad [v, w] = \sum_{i=1}^{k} [v_i, w_i].$$

The edges $[u_i, v_i], [u_i, w_i], [v_i, w_i]$ are parallel to $[u, v], [u, w], [v, w]$ respectively, and because $u, v, w$ form a triangle in $G(P)$, it follows that their ratios of edge lengths satisfies

$$\frac{|u_i - v_i|}{|u - v|} = \frac{|u_i - w_i|}{|u - w|} = \frac{|v_i - w_i|}{|v - w|}.$$

This implies there exists an index $j$ for which these ratios are nonzero, implying that vertices $u_j, v_j, w_j$ form a triangle in $G(P_j)$. Extending this reasoning to any other vertex $z$ in the complete subgraph, by applying the same logic with vertices $u, v, z$, it is deduced that $u_j, v_j, z_j$ also form a triangle in $G(P_j)$, and this pattern continues with other vertices. □

**Theorem 14.** *If the graph of a polytope $G(P)$ contains a complete subgraph with $p \geq 3$ vertices, then $d(P) \geq \lceil \log_2 p \rceil$.*

*Proof.* Suppose a subgraph of $G(P)$ is complete and contains $p = 3$ or $p = 4$ vertices. If we assume $d(P) = 1$, then $P = \sum_{i=1}^{k} P_i$, where each $P_i$ is a segment. This contradicts Lemma 13, which implies that at least one $P_i$ must include $p$ vertices. Therefore, we conclude $d(P) \geq 2$.

For the sake of induction, let's assume that the result holds for all cases up to $p - 1$. Now, consider that $G(P)$ includes a complete subgraph consisting of $p$ vertices. By definition, we can express $P$ as

$$P = \sum_{i=1}^{k} \operatorname{conv}\{P_i, Q_i\}, \quad \text{where } d(P_i), d(Q_i) \leq d(P) - 1.$$

According to Lemma 13, there exists an index $j$ for which $G(\operatorname{conv}\{P_j, Q_j\})$ also contains a complete subgraph $K$ with $p$ vertices. Without loss of generality, we can assume that $P_j$ contains at least $\lceil \frac{p}{2} \rceil$ vertices of $K$, and consequently the complete subgraph induced by those vertices. Using the induction hypothesis we obtain

$$d(P) - 1 \geq d(P_j) \geq \left\lceil \log_2 \left\lceil \frac{p}{2} \right\rceil \right\rceil = \lceil \log_2 p \rceil - 1,$$

from which it follows $d(P) \geq \lceil \log_2 p \rceil$. □

**Corollary 15.** *Any 2-neighbourly polytope $P$ with $p$ vertices satisfies $d(P) = \lceil \log_2 p \rceil$.*

*Proof.* It is a direct consequence of Theorem 14 and Proposition 5. □

**Corollary 16.** *Any $n$-simplex has depth $\lceil \log_2(n+1) \rceil$.*

Another important consequence of Theorem 14 is that allows to find a family of polytopes with the same dimension and increasingly large depth.

**Corollary 17.** *For every $p > n \geq 4$ the cyclic $n$-polytope with $p$ vertices has depth $\lceil \log_2(p+1) \rceil$.*

## 4 Concluding remarks

Knowing that $n$-simplices has depth $\lceil \log_2(n+1) \rceil$ reveals one part of Conjecture 3, and together with Proposition 4, we have obtained an upper depth bound for the conjecture. However, a tight lower bound is still needed to prove it.

In ReLU neural networks, from which Conjecture 3 originated, it has been proven that, for CPWL functions $f$ and $g$, if their depth satisfy $d(f) < d(g)$, then $d(f+g) = d(g)$ [8]. If this result also holds true in polytope neural networks, it could solve the conjecture. However, the existing proof for CPWL functions is inapplicable to polytopes, as it requires the inverse for the sum operation.

Another interesting contrast between polytope and ReLU networks is found in Corollary 17, where cyclic $n$-polytopes, for $n \geq 4$, have arbitrary large depth. Instead, for a fixed domain $\mathbb{R}^n$, all CPWL functions can be computed by ReLU neural networks with a depth of $\lceil \log_2(n+1) \rceil$. For polytopes, this contrast is also seen with Theorem 9, where polygons are shown to have a maximum depth of 2.

## References


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

[2] Ronald DeVore, Boris Hanin, and Guergana Petrova. Neural network approximation. *Acta Numerica*, 30:327–444, 2021.

[3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning.* MIT Press, 2016.

[4] Branko Grünbaum. *Convex Polytopes.* Springer, 2003.

[5] Christian Alexander Haase, Christoph Hertrich, and Georg Loho. Lower bounds on the depth of integral relu neural networks via lattice polytopes. In *International Conference on Learning Representations*, 2023.

[6] Christoph Hertrich, Amitabh Basu, Marco Di Summa, and Martin Skutella. Towards lower bounds on the depth of relu neural networks. *SIAM Journal on Discrete Mathematics*, 37(2):997–1029, 2023.

[7] Anirbit Mukherjee and Amitabh Basu. Lower bounds over boolean inputs for deep neural networks with relu gates. *arXiv preprint arXiv:1711.03073*, 2017.

[8] Juan L. Valerdi. On minimal depth in neural networks. *arXiv preprint arXiv:2402.15315*, 2024.

[9] Thomas Zaslavsky. *Facing up to arrangements: Face-count formulas for partitions of space by hyperplanes: Face-count formulas for partitions of space by hyperplanes*, volume 154. American Mathematical Soc., 1975.

[10] Liwen Zhang, Gregory Naitzat, and Lek-Heng Lim. Tropical geometry of deep neural networks. In *International Conference on Machine Learning*, pages 5824–5832. PMLR, 2018.