Supplementary Material: Tropical Geometry of Deep Neural Networks

A. Illustration of Our Neural Network

Figure A.1 summarizes the architecture and notations of the feedforward neural network discussed in this paper.

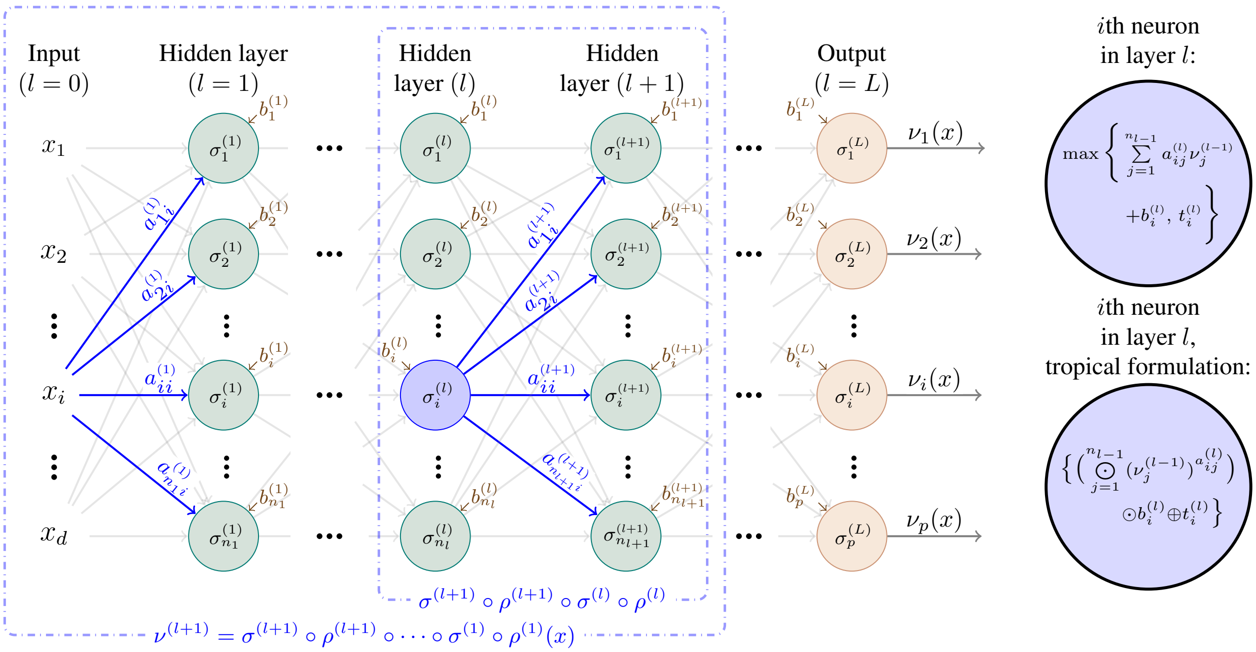


Figure A.1. General form of an ReLU feedforward neural network $\nu : \mathbb{R}^d \to \mathbb{R}^p$ with L layers.

B. Tropical Power

As in Section 2, we write $x^a = x^{\odot a}$; aside from this slight abuse of notation, $\oplus$ and $\odot$ denote tropical sum and product, $+$ and $\cdot$ denote standard sum and product in all other contexts. Tropical power evidently has the following properties:

- For $x, y \in \mathbb{R}$ and $a \in \mathbb{R}$, $a \geq 0$,
$$(x \oplus y)^a = x^a \oplus y^a \quad \text{and} \quad (x \odot y)^a = x^a \odot y^a.$$
If a is allowed negative values, then we lose the first property. In general $(x \oplus y)^a \neq x^a \oplus y^a$ for $a < 0$.
- For $x \in \mathbb{R}$,
$$x^0 = 0.$$
- For $x \in \mathbb{R}$ and $a, b \in \mathbb{N}$,
$$(x^a)^b = x^{a \cdot b}.$$
- For $x \in \mathbb{R}$ and $a, b \in \mathbb{Z}$,
$$x^a \odot x^b = x^{a+b}.$$
- For $x \in \mathbb{R}$ and $a, b \in \mathbb{Z}$,
$$x^a \oplus x^b = x^a \odot (x^{a-b} \oplus 0) = x^a \odot (0 \oplus x^{a-b}).$$

C. Examples

C.1. Examples of Tropical Curves and Dual Subdivision of Newton Polygon

Let $f \in \mathrm{Pol}(2,1) = \mathbb{T}[x_1, x_2]$, i.e., a bivariate tropical polynomial. It follows from our discussions in Section 3 that the tropical hypersurface $\mathcal{T}(f)$ is a planar graph dual to the dual subdivision $\delta(f)$ in the following sense:

(i) Each two-dimensional face in $\delta(f)$ corresponds to a vertex in $\mathcal{T}(f)$.

(ii) Each one-dimensional edge of a face in $\delta(f)$ corresponds to an edge in $\mathcal{T}(f)$. In particular, an edge from the Newton polygon $\Delta(f)$ corresponds to an unbounded edge in $\mathcal{T}(f)$ while other edges correspond to bounded edges.

Figure 2 illustrates how we may find the dual subdivision for the tropical polynomial $f(x_1, x_2) = 1 \odot x_1^2 \oplus 1 \odot x_2^2 \oplus 2 \odot x_1x_2 \oplus 2 \odot x_1 \oplus 2 \odot x_2 \oplus 2$. First, find the convex hull

$$\mathcal{P}(f) = \mathrm{Conv}\{(2,0,1), (0,2,1), (1,1,2), (1,0,2), (0,1,2), (0,0,2)\}.$$

Then, by projecting the upper envelope of $\mathcal{P}(f)$ to $\mathbb{R}^2$, we obtain $\delta(f)$, the dual subdivision of the Newton polygon.

C.2. Polytopes of a Two-Layer Neural Network

We illustrate our discussions in Section 6.2 with a two-layer example. Let $\nu : \mathbb{R}^2 \to \mathbb{R}$ be with $n_0 = 2$ input nodes, $n_1 = 5$ nodes in the first layer, and $n_2 = 1$ nodes in the output:

$$\omega = \nu^{(1)}(x) = \max\left\{ \begin{bmatrix} -1 & 1 \\ 1 & -3 \\ 1 & 2 \\ -4 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ 2 \\ 0 \\ -2 \end{bmatrix}, 0 \right\},$$

$$\nu^{(2)}(\omega) = \max\{\omega_1 + 2\omega_2 + \omega_3 - \omega_4 - 3\omega_5,\, 0\}.$$

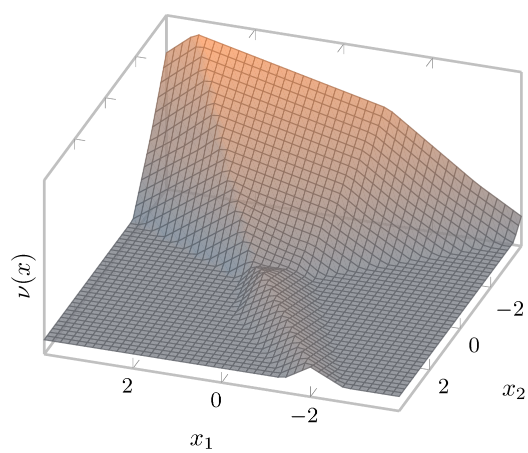
We first express $\nu^{(1)}$ and $\nu^{(2)}$ as tropical rational maps,

$$\nu^{(1)} = F^{(1)} \oslash G^{(1)}, \quad \nu^{(2)} = f^{(2)} \oslash g^{(2)},$$

where

$$y := F^{(1)}(x) = H^{(1)}(x) \oplus G^{(1)}(x),$$

$$z := G^{(1)}(x) = \begin{bmatrix} x_1 \\ x_2^3 \\ 0 \\ x_1^4 \\ 0 \end{bmatrix}, \qquad H^{(1)}(x) = \begin{bmatrix} 1 \odot x_2 \\ (-1) \odot x_1 \\ 2 \odot x_1x_2^2 \\ x_2 \\ (-2) \odot x_1^3x_2^2 \end{bmatrix},$$



and

$$\begin{aligned} f^{(2)}(x) &= g^{(2)}(x) \oplus h^{(2)}(x), \\ g^{(2)}(x) &= y_4 \odot y_5^3 \odot z_1 \odot z_2^2 \odot z_3 \\ &= (x_2 \oplus x_1^4) \odot ((-2) \odot x_1^3x_2^2 \oplus 0)^3 \odot x_1 \odot (x_2^3)^2, \\ h^{(2)}(x) &= y_1 \odot y_2^2 \odot y_3 \odot z_4 \odot z_5^3 \\ &= (1 \odot x_2 \oplus x_1) \odot ((-1) \odot x_1 \oplus x_2^3)^2 \odot (2 \odot x_1x_2^2 \oplus 0) \odot x_1^4. \end{aligned}$$

We will write $F^{(1)} = (f_1^{(1)}, \ldots, f_5^{(1)})$ and likewise for $G^{(1)}$ and $H^{(1)}$. The monomials occurring in $g_j^{(1)}(x)$ and $h_j^{(1)}(x)$ are all of the form $cx_1^{a_1}x_2^{a_2}$. Therefore $\mathcal{P}(g_j^{(1)})$ and $\mathcal{P}(h_j^{(1)})$, $j = 1, \ldots, 5$, are points in $\mathbb{R}^3$.

Since $F^{(1)} = G^{(1)} \oplus H^{(1)}$, $\mathcal{P}(f_j^{(1)})$ is a convex hull of two points, and thus a line segment in $\mathbb{R}^3$. The Newton polygons associated with $f_j^{(1)}$, equal to their dual subdivisions in this case, are obtained by projecting these line segments back to the plane spanned by a_1, a_2, as shown on the left in Figure C.1.

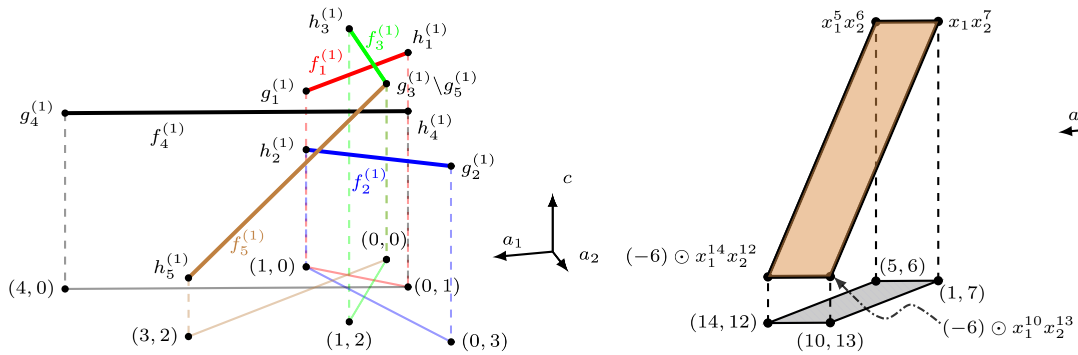


Figure C.1. Left: $\mathcal{P}(f_j^{(1)})$ and dual subdivision of $\Delta(f_j^{(1)}), j = 1, \ldots, 5$. Right: $\mathcal{P}(g^{(2)})$ and dual subdivision of $\Delta(g^{(2)})$. In both figures, dual subdivisions have been translated along the $-c$ direction (downwards) and separated from the polytopes for visibility.

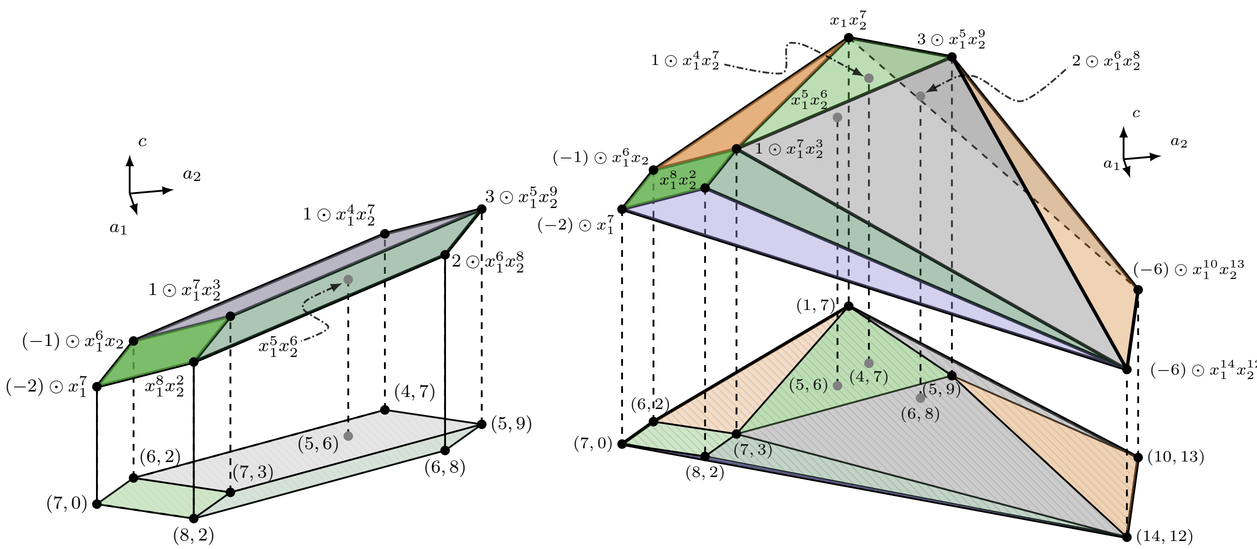


Figure C.2. Left: The polytope associated with $h^{(2)}$ and its dual subdivision. Right: $\mathcal{P}(f^{(2)})$ and dual subdivision of $\Delta(f^{(2)})$. In both figures, dual subdivisions have been translated along the $-c$ direction (downwards) and separated from the polytopes for visibility.

The line segments $\mathcal{P}(f_j^{(1)})$, $j = 1, \ldots, 5$, and points $\mathcal{P}(g_j^{(1)})$, $j = 1, \ldots, 5$, serve as building blocks for $\mathcal{P}(h^{(2)})$ and $\mathcal{P}(g^{(2)})$, which are constructed as weighted Minkowski sums:

$$\mathcal{P}(h^{(2)}) = \mathcal{P}(f_4^{(1)}) + 3\mathcal{P}(f_5^{(1)}) + \mathcal{P}(g_1^{(1)}) + 2\mathcal{P}(g_2^{(1)}) + \mathcal{P}(g_3^{(1)}),$$
$$\mathcal{P}(g^{(2)}) = \mathcal{P}(f_1^{(1)}) + 2\mathcal{P}(f_2^{(1)}) + \mathcal{P}(f_3^{(1)}) + \mathcal{P}(g_4^{(1)}) + 3\mathcal{P}(g_5^{(1)}).$$

$\mathcal{P}(g^{(2)})$ and the dual subdivision of its Newton polygon are shown on the right in Figure C.1. $\mathcal{P}(h^{(2)})$ and the dual subdivision of its Newton polygon are shown on the left in Figure C.2. $\mathcal{P}(f^{(2)})$ is the convex hull of the union of $\mathcal{P}(g^{(2)})$ and $\mathcal{P}(h^{(2)})$. The dual subdivision of its Newton polygon is obtained by projecting the upper faces of $\mathcal{P}(f^{(2)})$ to the plane spanned by a_1, a_2. These are shown on the right in Figure C.2.

D. Proofs

D.1. Proof of Corollary 3.4

Proof. Let V_1 and V_2 be the sets of vertices on the upper and lower envelopes of P respectively. By Theorem 3.3, P has

$$n_1 := 2\sum_{j=0}^{d} \binom{m-1}{j}$$

vertices in total. By construction, we have $|V_1 \cup V_2| = n_1$. It is well-known that zonotopes are centrally symmetric and so there are equal number of vertices on the upper and lower envelopes, i.e., $|V_1| = |V_2|$. Let $P' := \pi(P)$ be the projection of P into $\mathbb{R}^d$. Since the projected vertices are assumed to be in general positions, P' must be a d-dimensional zonotope generated by m nonparallel line segments. Hence, by Theorem 3.3 again, P' has

$$n_2 := 2\sum_{j=0}^{d-1}\binom{m-1}{j}$$

vertices. For any vertex $v \in P$, $\pi(v)$ is a vertex of P' if and only if v belongs to both the upper and lower envelopes, i.e., $v \in V_1 \cap V_2$. Therefore the number of vertices on P' equals $|V_1 \cap V_2|$. By construction, we have $|V_1 \cap V_2| = n_2$. Consequently the number of vertices on the upper envelope is

$$|V_1| = \frac{1}{2}(|V_1 \cup V_2| - |V_1 \cap V_2|) + |V_1 \cap V_2| = \frac{1}{2}(n_1 - n_2) + n_2 = \sum_{j=0}^{d}\binom{m}{j}. \qquad \square$$

D.2. Proof of Proposition 5.1

Proof. Writing $A = A_+ - A_-$, we have

$$\begin{aligned}
\rho^{(l+1)}(x) &= \big(A_+ - A_-\big)\big(F^{(l)}(x) - G^{(l)}(x)\big) + b \\
&= \big(A_+F^{(l)}(x) + A_-G^{(l)}(x) + b\big) - \big(A_+G^{(l)}(x) + A_-F^{(l)}(x)\big) \\
&= H^{(l+1)}(x) - G^{(l+1)}(x), \\
\nu^{(l+1)}(x) &= \max\big\{\rho^{(l+1)}(y),\, t\big\} \\
&= \max\big\{H^{(l+1)}(x) - G^{(l+1)}(x),\, t\big\} \\
&= \max\big\{H^{(l+1)}(x),\, G^{(l+1)}(x) + t\big\} - G^{(l+1)}(x) \\
&= F^{(l+1)}(x) - G^{(l+1)}(x).
\end{aligned} \qquad \square$$

D.3. Proof of Theorem 5.4

Proof. It remains to establish the "only if" part. We will write $\sigma_t(x) := \max\{x, t\}$. Any tropical monomial $b_i x^{\alpha_i}$ is clearly such a neural network as

$$b_i x^{\alpha_i} = (\sigma_{-\infty} \circ \rho_i)(x) = \max\{\alpha_i^{\intercal} x + b_i, -\infty\}.$$

If two tropical polynomials p and q are represented as neural networks with l_p and l_q layers respectively,

$$\begin{aligned}
p(x) &= \big(\sigma_{-\infty} \circ \rho_p^{(l_p)} \circ \sigma_0 \circ \ldots \sigma_0 \circ \rho_p^{(1)}\big)(x), \\
q(x) &= \big(\sigma_{-\infty} \circ \rho_q^{(l_q)} \circ \sigma_0 \circ \ldots \sigma_0 \circ \rho_q^{(1)}\big)(x),
\end{aligned}$$

then $(p \oplus q)(x) = \max\{p(x), q(x)\}$ can also be written as a neural network with $\max\{l_p, l_q\} + 1$ layers:

$$(p \oplus q)(x) = \sigma_{-\infty}\big([\sigma_0 \circ \rho_1](y(x)) + [\sigma_0 \circ \rho_2](y(x)) - [\sigma_0 \circ \rho_3](y(x))\big),$$

where $y : \mathbb{R}^d \to \mathbb{R}^2$ is given by $y(x) = (p(x), q(x))$ and $\rho_i : \mathbb{R}^2 \to \mathbb{R}$, $i = 1, 2, 3$, are linear functions defined by

$$\rho_1(y) = y_1 - y_2, \quad \rho_2(y) = y_2, \quad \rho_3(y) = -y_2.$$

Thus, by induction, any tropical polynomial can be written as a neural network with ReLU activation. Observe also that if a tropical polynomial is the tropical sum of r monomials, then it can be written as a neural network with no more than $\lceil \log_2 r \rceil + 1$ layers.

Next we consider a tropical rational function $(p \oslash q)(x) = p(x) - q(x)$ where p and q are tropical polynomials. Under the same assumptions, we can represent $p \oslash q$ as

$$(p \oslash q)(x) = \sigma_{-\infty}\big([\sigma_0 \circ \rho_4](y(x)) - [\sigma_0 \circ \rho_5](y(x)) + [\sigma_0 \circ \rho_6](y(x)) - [\sigma_0 \circ \rho_7](y(x))\big)$$

where $\rho_i : \mathbb{R}^2 \to \mathbb{R}^2$, $i = 4, 5, 6, 7$, are linear functions defined by

$$\rho_4(y) = y_1, \quad \rho_5(y) = -y_1, \quad \rho_6(y) = -y_2, \quad \rho_7(y) = y_2.$$

Therefore $p \oslash q$ is also a neural network with at most $\max\{l_p, l_q\} + 1$ layers.

Finally, if f and g are tropical polynomials that are respectively tropical sums of r_f and r_g monomials, then the discussions above show that $(f \oslash g)(x) = f(x) - g(x)$ is a neural network with at most $\max\{\lceil \log_2 r_f \rceil, \lceil \log_2 r_g \rceil\} + 2$ layers. □

D.4. Proof of Proposition 5.5

Proof. It remains to establish the "if" part. Let $\mathbb{R}^d$ be divided into N polyhedral region on each of which ν restricts to a linear function

$$\ell_i(x) = a_i^\intercal x + b_i, \quad a_i \in \mathbb{Z}^d, \quad b_i \in \mathbb{R}, \quad i = 1, \dots, L,$$

i.e., for any $x \in \mathbb{R}^d$, $\nu(x) = \ell_i(x)$ for some $i \in \{1, \dots, L\}$. It follows from (Tarela & Martinez, 1999) that we can find N subsets of $\{1, \dots, L\}$, denoted by S_j, $j = 1, \dots, N$, so that ν has a representation

$$\nu(x) = \max_{j=1,\dots,N} \min_{i \in S_j} \ell_i.$$

It is clear that each ℓ_i is a tropical rational function. Now for any tropical rational functions p and q,

$$\min\{p, q\} = -\max\{-p, -q\} = 0 \oslash [(0 \oslash p) \oplus (0 \oslash q)] = [p \odot q] \oslash [p \oplus q].$$

Since $p \odot q$ and $p \oplus q$ are both tropical rational functions, so is their tropical quotient. By induction, $\min_{i \in S_j} \ell_i$ is a tropical rational function for any $j = 1, \dots, N$, and therefore so is their tropical sum ν. □

D.5. Proof of Proposition 5.6

Proof. For a one-layer neural network $\nu(x) = \max\{Ax + b, t\} = (\nu_1(x), \dots, \nu_p(x))$ with $A \in \mathbb{R}^{p \times d}$, $b \in \mathbb{R}^p$, $x \in \mathbb{R}^d$, $t \in (\mathbb{R} \cup \{-\infty\})^p$, we have

$$\nu_k(x) = \left(b_k \odot \bigodot_{j=1}^{d} x_j^{a_{kj}} \right) \oplus t_k = \left(b_k \odot \bigodot_{j=1}^{d} x_j^{a_{kj}} \right) \oplus \left(t_k \odot \bigodot_{j=1}^{d} x_j^{0} \right), \qquad k = 1, \dots, p.$$

So for any $k = 1, \dots, p$, if we write $\bar{b}_1 = b_k$, $\bar{b}_2 = t_k$, $\bar{a}_{1j} = a_{kj}$, $\bar{a}_{2j} = 0$, $j = 1, \dots, d$, then

$$\nu_k(x) = \bigoplus_{i=1}^{2} \bar{b}_i \bigodot_{j=1}^{d} x_j^{\bar{a}_{ij}}$$

is clearly a tropical signomial function. Therefore ν is a tropical signomial map. The result for arbitrary number of layers then follows from using the same recurrence as in the proof in Section D.2, except that now the entries in the weight matrix are allowed to take real values, and the maps $H^{(l)}(x)$, $G^{(l)}(x)$, $F^{(l)}(x)$ are tropical signomial maps. Hence every layer can be written as a tropical rational signomial map $\nu^{(l)} = F^{(l)} \oslash G^{(l)}$. □

D.6. Proof of Proposition 6.1

We prove a slightly more general result.

Proposition D.1 (Level sets)**.** *Let $f \oslash g \in \operatorname{Rat}(d, 1) = \mathbb{T}(x_1, \dots, x_d)$.*

(i) *Given a constant $c > 0$, the* level set

$$\mathcal{B} := \{x \in \mathbb{R}^d : f(x) \oslash g(x) = c\}$$

divides $\mathbb{R}^d$ into at most $\mathcal{N}(f)$ connected polyhedral regions where $f(x) \oslash g(x) > c$, and at most $\mathcal{N}(g)$ such regions where $f(x) \oslash g(x) < c$.

(ii) *If $c \in \mathbb{R}$ is such that there is no tropical monomial in $f(x)$ that differs from any tropical monomial in $g(x)$ by c, then the level set $\mathcal{B}$ is contained in a tropical hypersurface,*

$$\mathcal{B} \subseteq \mathcal{T}(\max\{f(x),\, g(x)+c\}) = \mathcal{T}(c \odot g \oplus f).$$

Proof. We show that the bounds on the numbers of connected positive (i.e., above c) and negative (i.e., below c) regions are as we claimed in (i). The tropical hypersurface of f divides $\mathbb{R}^d$ into $\mathcal{N}(f)$ convex regions $C_1, \dots, C_{\mathcal{N}(f)}$ such that f is linear on each C_i. As g is piecewise linear and convex over $\mathbb{R}^d$, $f \oslash g = f - g$ is piecewise linear and concave on each C_i. Since the level set $\{x : f(x) - g(x) = c\}$ and the superlevel set $\{x : f(x) - g(x) \geq c\}$ must be convex by the concavity of $f - g$, there is at most one positive region in each C_i. Therefore the total number of connected positive regions cannot exceed $\mathcal{N}(f)$. Likewise, the tropical hypersurface of g divides $\mathbb{R}^d$ into $\mathcal{N}(g)$ convex regions on each of which $f \oslash g$ is convex. The same argument shows that the number of connected negative regions does not exceed $\mathcal{N}(g)$.

We next address (ii). Upon rearranging terms, the level set becomes

$$\mathcal{B} = \big\{x \in \mathbb{R}^d : f(x) = g(x) + c\big\}.$$

Since $f(x)$ and $g(x) + c$ are both tropical polynomial, we have

$$f(x) = b_1 x^{\alpha_1} \oplus \cdots \oplus b_r x^{\alpha_r},$$
$$g(x) + c = c_1 x^{\beta_1} \oplus \cdots \oplus c_s x^{\beta_s},$$

with appropriate multiindices $\alpha_1, \dots, \alpha_r$, $\beta_1, \dots, \beta_s$, and real coefficients $b_1, \dots, b_r$, $c_1, \dots, c_s$. By the assumption on the monomials, we have that $x_0 \in \mathcal{B}$ only if there exist i, j so that $\alpha_i \neq \beta_j$ and $b_i x_0^{\alpha_i} = c_j x_0^{\beta_j}$. This completes the proof since if we combine the monomials of $f(x)$ and $g(x) + c$ by (tropical) summing them into a single tropical polynomial, $\max\{f(x),\, g(x)+c\}$, the above implies that on the level set, the value of the combined tropical polynomial is attained by at least two monomials and therefore $x_0 \in \mathcal{T}(\max\{f(x),\, g(x)+c\})$. ☐

Proposition 6.1 follows immediately from Proposition D.1 since the decision boundary $\{x \in \mathbb{R}^d : \nu(x) = s^{-1}(c)\}$ is a level set of the tropical rational function ν.

D.7. Proof of Theorem 6.3

The linear regions of a tropical polynomial map $F \in \mathrm{Pol}(d, m)$ are all convex but this is not necessarily the case for a tropical rational map $F \in \mathrm{Rat}(d, n)$. Take for example a bivariate real-valued function $f(x, y)$ whose graph in $\mathbb{R}^3$ is a pyramid with base $\{(x, y) \in \mathbb{R}^2 : x, y \in [-1, 1]\}$ and zero everywhere else, then the linear region where f vanishes is $\mathbb{R}^2 \setminus \{(x, y) \in \mathbb{R}^2 : x, y \in [-1, 1]\}$, which is nonconvex. The nonconvexity invalidates certain geometric arguments that only apply in the convex setting. Nevertheless there is a way to subdivide each of the nonconvex linear regions into convex ones to get ourselves back into the convex setting. We will start with the number of *convex* linear regions for tropical rational maps although later we will deduce the required results for the number of linear regions (without imposing convexity).

We first extend the notion of tropical hypersurface to tropical rational maps: Given a tropical rational map $F \in \mathrm{Rat}(d, m)$, we define $\mathcal{T}(F)$ to be the boundaries between adjacent linear regions. When $F = (f_1, \dots, f_m) \in \mathrm{Pol}(d, m)$, i.e., a tropical polynomial map, this set is exactly the union of tropical hypersurfaces $\mathcal{T}(f_i)$, $i = 1, \dots, m$. Therefore this definition of $\mathcal{T}(F)$ extends Definition 3.1.

For a tropical rational map F, we will examine the smallest number of convex regions that form a refinement of $\mathcal{T}(F)$. For brevity, we will call this the *convex degree* of F; for consistency, the number of linear regions of F we will call its *linear degree*. We define convex degree formally below. We will write $F|_C$ to mean the restriction of map F to $C \subseteq \mathbb{R}^d$.

Definition D.1. *The* convex degree *of a tropical rational map $F \in \mathrm{Rat}(d, n)$ is the minimum division of $\mathbb{R}^d$ into convex regions over which F is linear, i.e.*

$$\mathcal{N}_c(F) \coloneqq \min\big\{n : C_1 \cup \cdots \cup C_n = \mathbb{R}^d,\ C_i \text{ convex, } F|_{C_i} \text{ linear}\big\}.$$

Note that $C_1, \dots, C_{\mathcal{N}_c(F)}$ either divide $\mathbb{R}^d$ into the same regions as $\mathcal{T}(F)$ or form a refinement.

For $m \le d$, we will denote by $\mathcal{N}_c(F \mid m)$ the maximum convex degree obtained by restricting F to an m-dimensional affine subspace in $\mathbb{R}^d$, i.e.,

$$\mathcal{N}_c(F \mid m) := \max\big\{\mathcal{N}_c(F|_\Omega) : \Omega \subseteq \mathbb{R}^d \text{ is an } m\text{-dimensional affine space}\big\}.$$

For any $F \in \mathrm{Rat}(d, n)$, there is at least one tropical polynomial map that subdivides $\mathcal{T}(F)$, and so convex degree is well-defined (e.g., if $F = (p_1 \oslash q_1, \dots, p_n \oslash q_n) \in \mathrm{Rat}(d, n)$, then we may choose $P = (p_1, \dots, p_n, q_1, \dots, q_n) \in \mathrm{Pol}(d, 2n)$). Since the linear regions of a tropical polynomial map are always convex, we have $\mathcal{N}(F) = \mathcal{N}_c(F)$ for any $F \in \mathrm{Pol}(d, n)$.

Let $F = (f_1, \dots, f_n) \in \mathrm{Rat}(d, n)$ and $\alpha = (a_1, \dots, a_n) \in \mathbb{Z}^n$. Consider the tropical rational function[1]

$$F^\alpha := \alpha^\intercal F = a_1 f_1 + \cdots + a_n f_n = \bigodot_{j=1}^{n} f_j^{a_j} \in \mathrm{Rat}(d, 1).$$

For some α, F^α may have fewer linear regions than F, e.g, $\alpha = (0, \dots, 0)$. As such, we need the following notion.

Definition D.2. *$\alpha = (a_1, \dots, a_n) \in \mathbb{Z}^n$ is said to be a* general exponent *of $F \in \mathrm{Rat}(d, n)$ if the linear regions of F^α and the linear regions of F are identical.*

We show that general exponent always exists for any $F \in \mathrm{Rat}(d, n)$ and may be chosen to have all entries nonnegative.

Lemma D.2. *Let $F \in \mathrm{Rat}(d, n)$. Then*

(i) *$\mathcal{N}(F^\alpha) = \mathcal{N}(F)$ if and only if α is a general exponent;*

(ii) *F has a general exponent $\alpha \in \mathbb{N}^n$.*

Proof. It follows from the definition of tropical hypersuface that $\mathcal{T}(F^\alpha)$ and $\mathcal{T}(F)$ comprise respectively the points $x \in \mathbb{R}^d$ at which F^α and F are not differentiable. Hence $\mathcal{T}(F^\alpha) \subseteq \mathcal{T}(F)$, which implies that $\mathcal{N}(F^\alpha) < \mathcal{N}(F)$ unless $\mathcal{T}(F^\alpha) = \mathcal{T}(F)$. This concludes (i).

For (ii), we need to show that there always exists an $\alpha \in \mathbb{N}^n$ such that F^α divides its domain $\mathbb{R}^d$ into the same set of linear regions as F. In other words, for every pair of adjacent linear regions of F, the $(d-1)$-dimensional face in $\mathcal{T}(F)$ that separates them is also present in $\mathcal{T}(F^\alpha)$ and so $\mathcal{T}(F^\alpha) \supseteq \mathcal{T}(F)$.

Let L and M be adjacent linear regions of F. The differentials of $F|_L$ and $F|_M$ must have integer coordinates, i.e., $dF|_L, dF|_M \in \mathbb{Z}^{n\times d}$. Since L and M are distinct linear regions, we must have $dF|_L \neq dF|_M$ (or otherwise L and M can be merged into a single linear region). Note that the differentials of $F^\alpha|_L$ and $F^\alpha|_M$ are given by $\alpha^\intercal dF|_L$ and $\alpha^\intercal dF|_M$.

To ensure the $(d-1)$-dimensional face separating L and M still exists in $\mathcal{T}(F^\alpha)$, we need to choose α so that $\alpha^\intercal dF|_L \neq \alpha^\intercal dF|_M$. Observe that the solution to $(dF|_L - dF|_M)^\intercal \alpha = 0$ is contained in a one-dimensional subspace of $\mathbb{R}^n$.

Let $\mathcal{A}(F)$ be the collection of all pairs of adjacent linear regions of F. Since the set of α that degenerates two adjacent linear regions into a single one, i.e.,

$$\mathcal{S} := \bigcup_{(L,M)\in\mathcal{A}(F)} \big\{\alpha \in \mathbb{N}^n : (dF|_L - dF|_M)^\intercal \alpha = 0)\big\},$$

is contained in a union of a finite number of hyperplanes in $\mathbb{R}^n$, $\mathcal{S}$ cannot cover the entire lattice of nonnegative integers $\mathbb{N}^n$. Therefore the set $\mathbb{N}^n \cap (\mathbb{R}^n \setminus \mathcal{S})$ is nonempty and any of its element is a general exponent for F. ☐

Lemma D.2 shows that we may study the linear degree of a tropical rational map by studying that of a tropical rational function, for which the results in Section 3.1 apply.

We are now ready to prove a key result on the convex degree of composition of tropical rational maps.

Theorem D.3. *Let $F = (f_1, \dots, f_m) \in \mathrm{Rat}(n, m)$ and $G \in \mathrm{Rat}(d, n)$. Define $H = (h_1, \dots, h_m) \in \mathrm{Rat}(d, m)$ by*

$$h_i := f_i \circ G, \qquad i = 1, \dots, m.$$

Then

$$\mathcal{N}(H) \le \mathcal{N}_c(H) \le \mathcal{N}_c(F \mid d) \cdot \mathcal{N}_c(G).$$

[1] This is in the sense of a tropical power but we stay consistent to our slight abuse of notation and write F^α instead of $F^{\odot\alpha}$.

Proof. Only the upper bound requires a proof. Let $k = \mathcal{N}_c(G)$. By the definition of $\mathcal{N}_c(G)$, there exist convex sets $C_1, \ldots, C_k \subseteq \mathbb{R}^d$ whose union is $\mathbb{R}^d$ and on each of which G is linear. So $G|_{C_i}$ is some affine function ρ_i. For any i,

$$\mathcal{N}_c(F \circ \rho_i) \leq \mathcal{N}_c(F \mid d),$$

by the definition of $\mathcal{N}_c(F \mid d)$. Since $F \circ G = F \circ \rho_i$ on C_i, we have

$$\mathcal{N}_c(F \circ G) \leq \sum_{i=1}^{k} \mathcal{N}_c(F \circ \rho_i).$$

Hence

$$\mathcal{N}_c(F \circ G) \leq \sum_{i=1}^{k} \mathcal{N}_c(F \circ \rho_i) \leq \sum_{i=1}^{k} \mathcal{N}_c(F \mid d) = \mathcal{N}_c(F \mid d) \cdot \mathcal{N}_c(G). \qquad \square$$

We now apply our observations on tropical rational functions to neural networks. The next lemma follows directly from Corollary 3.4.

Lemma D.4. *Let $\sigma^{(l)} \circ \rho^{(l)} : \mathbb{R}^{n_{l-1}} \to \mathbb{R}^{n_l}$ where $\sigma^{(l)}$ and $\rho^{(l)}$ are the affine transformation and activation of the lth layer of a neural network. If $d \leq n_l$, then*

$$\mathcal{N}_c(\sigma^{(l)} \circ \rho^{(l)} \mid d) \leq \sum_{i=0}^{d} \binom{n_l}{i}.$$

Proof. $\mathcal{N}_c(\sigma^{(l)} \circ \rho^{(l)} \mid d)$ is the maximum convex degree of a tropical rational map $F = (f_1, \ldots, f_{n_l}) : \mathbb{R}^d \to \mathbb{R}^{n_l}$ of the form

$$f_i(x) \coloneqq \sigma_i^{(l)} \circ \rho^{(l)} \circ (b_1 \odot x^{\alpha_1}, \ldots, b_{n_{l-1}} \odot x^{\alpha_{n_{l-1}}}), \qquad i = 1, \ldots, n_l.$$

For a general affine transformation $\rho^{(l)}$,

$$\rho^{(l)}(b_1 \odot x^{\alpha_1}, \ldots, b_{n_{l-1}} \odot x^{\alpha_{n_{l-1}}}) = \big(b'_1 \odot x^{\alpha'_1}, \ldots, b'_{n_l} \odot x^{\alpha'_{n_l}}\big) =: G(x)$$

for some $\alpha'_1, \ldots, \alpha'_{n_l}$ and $b'_1, \ldots, b'_{n_l}$, and we denote this map by $G : \mathbb{R}^d \to \mathbb{R}^{n_l}$. So $f_i = \sigma_i^{(l)} \circ G$. By Theorem D.3, we have $\mathcal{N}_c(\sigma^{(l)} \circ \rho^{(l)} \mid d) = \mathcal{N}_c(\sigma^{(l)} \mid d) \cdot \mathcal{N}_c(G) = \mathcal{N}_c(\sigma^{(l)} \mid d)$; note that $\mathcal{N}_c(G) = 1$ as G is a linear function.

We have thus reduced the problem to determining a bound on the convex degree of a single layer neural network with n_l nodes $\nu = (\nu_1, \ldots, \nu_{n_l}) : \mathbb{R}^d \to \mathbb{R}^{n_l}$. Let $\gamma = (c_1, \ldots, c_{n_l}) \in \mathbb{N}^{n_l}$ be a nonnegative general exponent for ν. Note that

$$\bigodot_{j=1}^{n_l} \nu_j^{c_j} = \bigodot_{j=1}^{n_l} \Bigg[\bigg(\bigodot_{i=1}^{d} b_i \odot x^{a_{ji}^+}\bigg) \oplus \bigg(\bigodot_{i=1}^{d} x^{a_{ji}^-}\bigg) \odot t_j\Bigg]^{c_j} - \bigodot_{j=1}^{n_l} \bigg(\bigodot_{i=1}^{d} x^{a_{ji}^-}\bigg)^{c_j}.$$

Since the last term is linear in x, we may drop it without affecting the convex degree of the entire expression. It remains to determine an upper bound for the number of linear regions of the tropical polynomial

$$h(x) = \bigodot_{j=1}^{n_l} \Bigg[\bigg(\bigodot_{i=1}^{d} b_i \odot x^{a_{ji}^+}\bigg) \oplus \bigg(\bigodot_{i=1}^{d} x^{a_{ji}^-}\bigg) \odot t_j\Bigg]^{c_j},$$

which we will obtain by counting vertices of the polytope $\mathcal{P}(h)$. By Propositions 3.1 and 3.2 the polytope $\mathcal{P}(h)$ is given by a weighted Minkowski sum

$$\sum_{j=1}^{n_l} c_j \mathcal{P}\Bigg[\bigg(\bigodot_{i=1}^{d} b_i \odot x^{a_{ji}^+}\bigg) \oplus \bigg(\bigodot_{i=1}^{d} x^{a_{ji}^-}\bigg) \odot t_j\Bigg].$$

By Proposition 3.2 again,

$$\mathcal{P}\Bigg[\bigg(\bigodot_{i=1}^{d} b_i \odot x^{a_{ji}^+}\bigg) \oplus \bigg(\bigodot_{i=1}^{d} x^{a_{ji}^-}\bigg) \odot t_j\Bigg] = \operatorname{Conv}\big(\mathcal{V}(\mathcal{P}(f)) \cup \mathcal{V}(\mathcal{P}(g))\big)$$

where

$$f(x) = \bigodot_{i=1}^{d} b_i \odot x^{a_{ji}^{+}} \qquad \text{and} \qquad g(x) = \left(\bigodot_{i=1}^{d} x^{a_{ji}^{-}}\right) \odot t_j$$

are tropical monomials. Therefore $\mathcal{P}(f)$, $\mathcal{P}(g)$ are just points in $\mathbb{R}^{d+1}$ and $\operatorname{Conv}\big(\mathcal{V}(\mathcal{P}(f)) \cup \mathcal{V}(\mathcal{P}(g))\big)$ is a line in $\mathbb{R}^{d+1}$. Hence $\mathcal{P}(h)$ is a Minkowski sum of n_l line segments in $\mathbb{R}^{d+1}$, i.e., a zonotope, and Corollary 3.4 completes the proof. □

Using Lemma D.4, we obtain a bound on the number of linear regions created by one layer of a neural network.

Theorem D.5. *Let $\nu : \mathbb{R}^d \to \mathbb{R}^{n_L}$ be an L-layer neural network satisfying assumptions (a)–(c) with $F^{(l)}$, $G^{(l)}$,$H^{(l)}$, and $\nu^{(l)}$ as defined in Proposition 5.1. Let $n_l \geq d$ for all $l = 1, \dots, L$. Then*

$$\mathcal{N}_c(\nu^{(1)}) = \mathcal{N}(G^{(1)}) = \mathcal{N}(H^{(1)}) = 1, \qquad \mathcal{N}_c(\nu^{(l+1)}) \leq \mathcal{N}_c(\nu^{(l)}) \cdot \sum_{i=0}^{d} \binom{n_{l+1}}{i}.$$

Proof. The $l = 1$ case follows from the fact that $G^{(1)}(x) = A_{-}^{(1)}x$ and $H^{(1)}(x) = A_{+}^{(1)}x + b^{(1)}$ are both linear, which in turn forces $\mathcal{N}_c(\nu^{(1)}) = 1$ as in the proof of Lemma D.4. Since $\nu^{(l)} = (\sigma^{(l)} \circ \rho^{(l)}) \circ \nu^{(l-1)}$, the recursive bound follows from Theorem D.3 and Lemma D.4. □

Theorem 6.3 follows from applying Theorem D.5 recursively.

References

Tarela, J. and Martinez, M. Region configurations for realizability of lattice piecewise-linear models. *Mathematical and Computer Modelling*, 30(11-12):17–27, 1999.