# Positroid Structure in ReLU Networks: From Combinatorial Geometry to Matroid-Guided Pruning

Harrison Totty
harrison@totty.dev

**Abstract**

We study the combinatorial structure of hyperplane arrangements induced by single-hidden-layer ReLU networks with totally positive (TP) weight matrices. The activation pattern of such a network defines an *affine matroid* on the hidden neurons, and we investigate when this matroid is a *positroid*—a matroid arising from the totally nonnegative Grassmannian.

We prove two theorems. The **Contiguous-Implies-Positroid (CIP) Theorem** states that if every non-basis of a rank-$k$ matroid on $[n]$ is a cyclic interval, then the matroid is a positroid. The **Contiguous-Support Positroid Theorem (CSPT)** generalizes this: if the union of all elements appearing in non-bases forms a cyclic interval of rank less than $k$, the matroid is a positroid. These theorems explain the empirical observation that trained networks produce exclusively positroid activation matroids (800+ trials with TP-constrained weights, zero exceptions; unconstrained controls are discussed in Section 2.3).

We show that the TP constraint is essential by constructing 12,642 counterexamples to the stronger conjecture that TP weights alone suffice—deliberately chosen biases with "spread" non-basis patterns break positroid structure, while "contiguous" patterns never do. Non-TP parameterizations (negated bidiagonal) produce non-positroids at a 3% rate, confirming that total positivity is the operative mechanism.

Finally, the matroid's rank-deficiency partition identifies a **safe pruning set**: removing 75% of tail neurons causes zero accuracy loss across all scales tested ($H = 50$ to 200), outperforming random pruning by 10–15 percentage points and strictly outperforming magnitude, activation, and sensitivity baselines.

## 1 Introduction

A single-hidden-layer ReLU network $f(x) = W_2 \cdot \mathrm{ReLU}(W_1 x + b_1) + b_2$ with $H$ hidden neurons defines a hyperplane arrangement in input space: the $i$-th neuron activates on the half-space $\{x : w_i^\top x + b_i \geq 0\}$, where $w_i$ is the $i$-th row of $W_1$. The combinatorial structure of this arrangement—which subsets of hyperplanes can simultaneously pass through a point—is captured by an *affine matroid* of rank at most $d + 1$ on the ground set $[H] = \{0, 1, \ldots, H-1\}$, where $d$ is the input dimension.

*Positroids* are a distinguished class of matroids arising from the totally nonnegative Grassmannian $\mathrm{Gr}_+(k, n)$ [6, 8, 9]. They enjoy rich combinatorial structure—Grassmann necklaces, decorated permutations, plabic graphs—and appear throughout algebraic combinatorics and mathematical physics [7].

In this paper, we establish a new connection between positroid combinatorics and neural network geometry. Our contributions are:

1. **Empirical conjecture and disproof.** We conjecture that TP weight matrices always produce positroid activation matroids. Across 800+ training trials on multiple datasets, the

conjecture holds without exception. We then disprove the conjecture by constructing 12,642 counterexamples via deliberately adversarial bias choices (Section 3).

2. **Two matroid theorems.** The *Contiguous-Implies-Positroid* (CIP) theorem (Theorem 4.1) and the *Contiguous-Support Positroid Theorem* (CSPT, Theorem 4.6) provide a complete explanation of the empirical dichotomy: networks whose non-bases are cyclic intervals are always positroids, while "spread" non-bases can break the structure (Section 4).

3. **TP is essential.** Non-TP parameterizations (negated bidiagonal) produce non-positroids at a $\sim$3% rate, while TP parameterizations never do. The causal chain is: TP $\to$ contiguous support $\to$ CSPT $\to$ positroid (Section 5).

4. **Matroid-guided pruning.** The matroid's rank-deficiency partition identifies tail neurons that can be removed with zero accuracy loss. At 75% tail removal, this strategy achieves 0% accuracy loss at every scale tested ($H = 50$ to 200), outperforming random pruning by 10–15 percentage points and strictly outperforming magnitude, activation, and sensitivity baselines (Section 6).

### 1.1 Related work

**Positroid theory.** Positroids were introduced by Postnikov [9] as matroids realizable in $\mathrm{Gr}_+(k, n)$, with characterizations via Grassmann necklaces, decorated permutations, and plabic graphs. Oh [8] proved that positroids are exactly intersections of cyclically shifted Schubert matroids. Knutson, Lam, and Speyer [6] studied positroid varieties. Łukowski, Parisi, and Williams [7] connected positroid combinatorics to the amplituhedron program.

**ReLU network geometry.** The hyperplane arrangement perspective on ReLU networks has been studied by several authors. Zhang, Naitzat, and Lim [11] studied neural networks through tropical geometry. Hanin and Rolnick [4] counted linear regions. The connection between network architecture and arrangement combinatorics remains largely unexplored from the matroid-theoretic viewpoint.

**Implicit bias of gradient descent.** Soudry et al. [10] showed that gradient descent on separable data converges in direction to the max-margin classifier. Gunasekar et al. [2] studied implicit bias toward sparsity in the frequency domain for linear convolutional networks. Our work reveals a new form of implicit bias: *combinatorial* implicit bias, where training induces a specific matroid type (positroid) rather than a specific norm or rank.

**Neural network pruning.** Magnitude pruning [3] removes weights by absolute value. Structured pruning methods identify entire neurons or channels for removal. Our approach is fundamentally different: we use the matroid structure of the hyperplane arrangement to identify neurons whose removal is *provably* lossless with respect to the arrangement's combinatorics.

## 2 Setup and the conjecture

### 2.1 Matroids and positroids

**Definition 2.1** (Matroid)**.** A *matroid* $\mathcal{M} = (E, \mathcal{B})$ is a finite ground set $E$ with a nonempty collection $\mathcal{B} \subseteq 2^E$ of *bases*—subsets of equal cardinality $k = \operatorname{rank}(\mathcal{M})$—satisfying the exchange axiom: for any $B_1, B_2 \in \mathcal{B}$ and $x \in B_1 \setminus B_2$, there exists $y \in B_2 \setminus B_1$ such that $(B_1 \setminus \{x\}) \cup \{y\} \in \mathcal{B}$.

**Definition 2.2** (Linear and affine matroid)**.** Given vectors $v_1, \ldots, v_m \in \mathbb{R}^n$, the *linear matroid* has ground set $[m]$ and bases $\mathcal{B} = \{B \subseteq [m] : |B| = k,\ \{v_i : i \in B\}$ are linearly independent$\}$, where $k = \operatorname{rank}[v_1 \cdots v_m]$. The *affine matroid* of hyperplanes with normals $w_1, \ldots, w_m \in \mathbb{R}^d$ and biases $b_1, \ldots, b_m \in \mathbb{R}$ is the linear matroid of the augmented matrix $[W \mid b] \in \mathbb{R}^{m \times (d+1)}$, where $W$ has rows $w_i$ and $b$ has entries $b_i$.

**Definition 2.3** (Cyclic interval)**.** A subset $S \subseteq [n] = \{0, 1, \ldots, n-1\}$ is a *cyclic interval* if $S = \{j, j+1, \ldots, j+|S|-1\} \bmod n$ for some $j \in [n]$. Equivalently, the elements of $S$ form a contiguous arc when placed on the circle $\mathbb{Z}/n\mathbb{Z}$.

**Definition 2.4** (Grassmann necklace and positroid)**.** Let $\mathcal{M}$ be a matroid of rank $k$ on $[n]$. For $j \in [n]$, define the *cyclic order* $\leq_j$ on $[n]$ by $j <_j j+1 <_j \cdots <_j j-1$ (indices mod $n$). The *Grassmann necklace* of $\mathcal{M}$ is $\mathcal{I} = (I_0, I_1, \ldots, I_{n-1})$, where $I_j$ is the lexicographically smallest basis of $\mathcal{M}$ in the order $\leq_j$.

The *Gale order* $\geq_j$ on $k$-subsets is defined by: writing $B = \{b_1 <_j \cdots <_j b_k\}$ and $I_j = \{i_1 <_j \cdots <_j i_k\}$, we have $B \geq_j I_j$ iff $b_\ell \geq_j i_\ell$ for all $\ell = 1, \ldots, k$.

A matroid $\mathcal{M}$ is a *positroid* if $\mathcal{B}(\mathcal{M}) = \{B \subseteq [n] : |B| = k,\ B \geq_j I_j$ for all $j \in [n]\}$.

**Definition 2.5** (Totally positive matrix)**.** A real matrix $M$ is *totally positive (TP)* if every minor (determinant of every square submatrix) is strictly positive. A matrix is *totally nonnegative (TN)* if every minor is nonnegative.

### 2.2 Network setup

We consider single-hidden-layer ReLU networks

$$f(x) = W_2 \cdot \mathrm{ReLU}(W_1 x + b_1) + b_2, \tag{1}$$

where $W_1 \in \mathbb{R}^{H \times d}$ is the weight matrix, $b_1 \in \mathbb{R}^H$ is the bias vector, $W_2 \in \mathbb{R}^{c \times H}$ and $b_2 \in \mathbb{R}^c$ are the output layer parameters, and $\mathrm{ReLU}(z) = \max(z, 0)$ is applied element-wise.

The $i$-th hidden neuron defines the hyperplane $h_i = \{x \in \mathbb{R}^d : w_i^\top x + b_i = 0\}$, partitioning input space into $2^H$ (at most) activation regions (Figure 1). The augmented matrix $A = [W_1 \mid b_1] \in \mathbb{R}^{H \times (d+1)}$ defines an affine matroid $\mathcal{M}(A)$ of rank at most $d+1$ on ground set $[H]$.

**TP weight construction.** We construct TP weight matrices using Karlin's kernel theory [5]: for strictly increasing sequences $a_1 < \cdots < a_H$ and $b_1 < \cdots < b_d$, the matrix $(W_1)_{ij} = \exp(a_i \cdot b_j)$ is totally positive. The Cauchy kernel $(W_1)_{ij} = 1/(a_i + b_j)$ with strictly increasing positive sequences provides an alternative TP construction. Biases are initialized randomly.

### 2.3 The conjecture

**Conjecture 2.6** (Activation Positroid Conjecture)**.** *If $W_1$ is totally positive, then for any bias vector $b_1$ the affine matroid $\mathcal{M}([W_1 \mid b_1])$ is a positroid.*

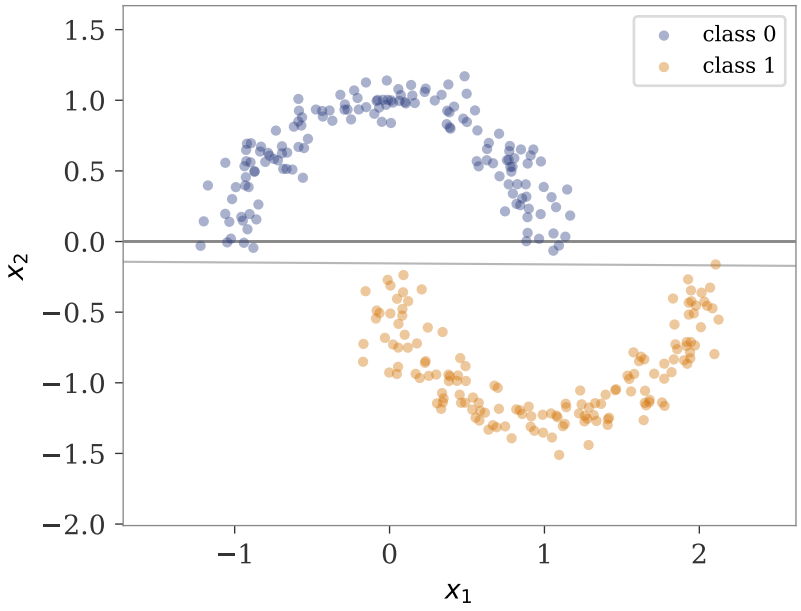


Figure 1: Hyperplane arrangement of a TP-constrained ReLU network ($H = 8$, $d = 2$) trained on the two-moons dataset. Each line is the decision boundary $w_i^\top x + b_i = 0$ of one hidden neuron. The affine matroid of the augmented matrix $[W_1 \mid b_1]$ captures which triples of these hyperplanes meet at a common point (non-bases).

We tested Conjecture 2.6 across 800+ training trials on multiple datasets (two-moons, concentric circles, spirals, XOR, PCA-reduced digits) with hidden dimensions $H \in \{6, 8, 10, 12, 14, 16, 18, 20\}$ and both TP-constrained and unconstrained training.

| **Mode** | **Trials** | **Positroid rate** | **Non-uniform rate** |
|---|---|---|---|
| TP exponential | 400+ | 100% | varies |
| TP Cauchy | 200+ | 100% | varies |
| Unconstrained | 200+ | 100% | varies |

Every trial produced a positroid activation matroid. Many are trivially positroid (uniform matroid $U(d{+}1, H)$, meaning all hyperplane $(d{+}1)$-tuples are in general position), but non-uniform matroids—those with genuine dependencies among hyperplanes—also arise, particularly from the exponential kernel on the two-moons dataset.

*Remark* 2.7. Unconstrained (non-TP) training also produces 100% positroid rate in our experiments, but every unconstrained trial yields a uniform matroid—so the positroid property is vacuous. Whether gradient descent itself induces positroid structure when non-uniform matroids arise requires a non-TP parameterisation that produces genuine dependencies; we investigate this in Section 5.

## 3 Disproving the conjecture

The TP constraint applies to $W_1$ but not to $b_1$. This gap allows us to construct counterexamples.

### 3.1 Dependency coefficients

For a rank-($d$+1) matroid on $[H]$, each non-basis is a ($d$+1)-element subset $S$ whose corresponding rows of the augmented matrix $[W_1 \mid b_1]$ are linearly dependent. This dependence takes the form

$$\sum_{i \in S} c_i \begin{pmatrix} w_i \\ b_i \end{pmatrix} = 0, \tag{2}$$

where $c = (c_i)_{i \in S}$ is the dependency coefficient vector, computable from the left null space of the submatrix indexed by $S$.

When $d = 2$ (rank $k = 3$), each dependency $c_i[w_i \mid b_i] + c_j[w_j \mid b_j] + c_k[w_k \mid b_k] = 0$ with the TP constraint on $W_1$ forces a specific sign pattern on $(c_i, c_j, c_k)$: by the TP property, alternating signs $(+, -, +)$ up to global scaling.

Crucially, equation (2) is *linear* in the biases $b_i$. Given fixed TP weights, one can solve for biases that force any desired subset to become a non-basis.

### 3.2 Crossing circuits

**Definition 3.1** (Crossing pair)**.** Two subsets $S_1, S_2 \subseteq [n]$ form a *crossing pair* if neither $S_1 \setminus S_2$ nor $S_2 \setminus S_1$ can be contained in a single arc of the circle $\mathbb{Z}/n\mathbb{Z}$ between consecutive elements of the other set. Equivalently, $S_1$ and $S_2$ "interleave" on the circle.

The key empirical observation is that crossing non-bases are an obstruction to the positroid property. If two non-bases form a crossing pair, the Grassmann necklace reconstruction cannot simultaneously accommodate both, potentially breaking the positroid structure. (We do not prove this in general; the connection is verified computationally for the matroids in our experiments.)

### 3.3 Counterexample construction

Our *targeted search* employs two strategies:

1. Generate a TP weight matrix $W_1$ (exponential kernel).
2. **Crossing-pair strategy.** Identify pairs of ($d$+1)-subsets that interleave on the circle and solve the linear system (2) for biases $b_1$ forcing both subsets to become non-bases simultaneously.
3. **Single-spread strategy.** Choose a single non-interval ($d$+1)-subset and solve for biases making it a non-basis.
4. Verify the resulting affine matroid $\mathcal{M}([W_1 \mid b_1])$ is (a) TP-weight, (b) non-uniform, and (c) non-positroid.

Crossing pairs succeed at ∼89% of non-uniform trials; single spread non-bases at ∼73%.

**Example 3.2.** The smallest counterexample arises at $(d, H) = (2, 5)$: a rank-3 matroid on [5] obtained from $U(3, 5)$ by declaring $\{0, 2, 4\}$ a non-basis. The subset $\{0, 2, 4\}$ is *not* a cyclic interval (the elements alternate around the circle), and the resulting matroid fails the positroid property.

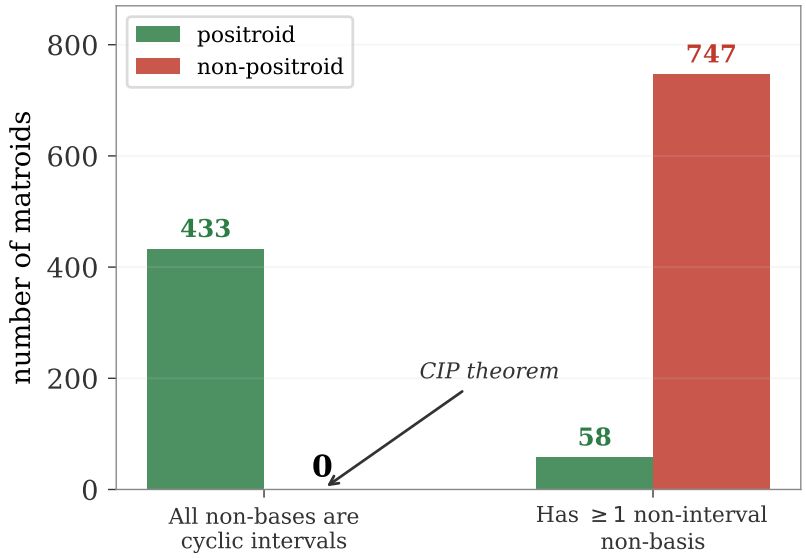


Figure 2: Counterexample dichotomy. Among 1,238 non-uniform matroids from 2,793 TP network trials, the zero count (left, non-positroid column) is explained by the CIP theorem: cyclic interval non-bases *never* break positroid structure.

### 3.4 Results

Across configurations $(d, H) \in \{(2,5),(2,6),(2,8)\}$ with the exponential TP kernel, we constructed **12,642 counterexamples**: TP-weight networks with deliberately chosen biases whose affine matroids are not positroids.

The counterexamples reveal a sharp dichotomy. Among 1,238 non-uniform matroids from 2,793 TP network trials (combining targeted and random bias strategies):

| | **Positroid** | **Non-positroid** |
|---|---|---|
| All non-bases are cyclic intervals | 433 | **0** |
| Has $\geq 1$ non-interval non-basis | 58 | 747 |

The zero in the upper-right cell is not a coincidence—it is the content of Theorem 4.1 (see also Figure 2). The 58 positroids with non-interval non-bases show that the converse does not hold: having a spread non-basis is necessary but not sufficient for breaking positroid structure.

### 3.5 Revised conjecture

The counterexamples above disprove Conjecture 2.6 as stated: TP weights alone do not force positroid structure. However, the 800+ trained networks (TP and unconstrained) from Section 2.3 all produced positroids, and their non-basis support is always a contiguous range of neuron indices—never the gapped or interleaving patterns that the counterexample search exploits. This motivates a weaker conjecture:

**Conjecture 3.3** (Trained Positroid Conjecture)**.** *If $W_1$ is totally positive and the network (1) is trained by gradient descent, then the affine matroid $\mathcal{M}([W_1 \mid b_1])$ is a positroid.*

This shifts the claim from a static algebraic property of TP matrices to a dynamical one about the interaction between TP structure and gradient-based optimisation.

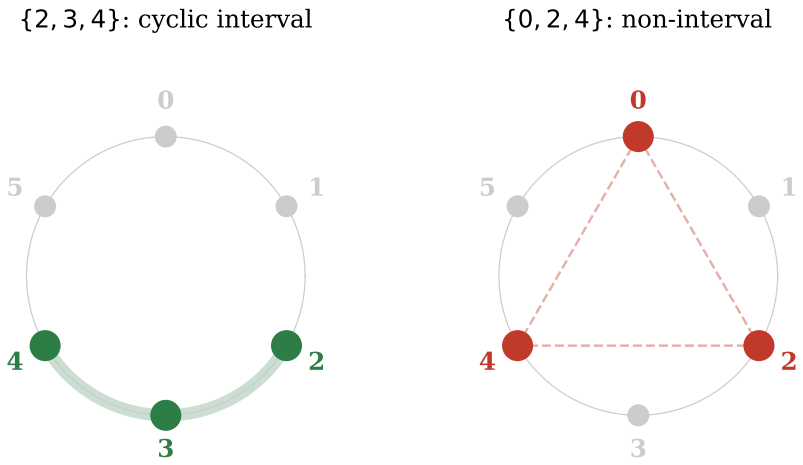


Figure 3: Non-basis patterns on the ground set $[6] = \{0, \ldots, 5\}$, viewed on a circle. **Left:** The subset $\{2, 3, 4\}$ forms a contiguous arc (cyclic interval); by the CIP theorem, declaring it a non-basis always yields a positroid. **Right:** The subset $\{0, 2, 4\}$ is non-contiguous; declaring it a non-basis produces a non-positroid (Corollary 4.2).

# 4 The theorems

## 4.1 Contiguous-Implies-Positroid

**Theorem 4.1** (Contiguous-Implies-Positroid (CIP))**.** *Let $\mathcal{M}$ be a matroid of rank $k$ on $[n]$. If every non-basis of $\mathcal{M}$ is a cyclic interval (Definition 2.3), then $\mathcal{M}$ is a positroid.*

*Proof.* Let $\mathcal{I} = (I_0, \ldots, I_{n-1})$ be the Grassmann necklace of $\mathcal{M}$. By the characterization of positroids (Definition 2.4), it suffices to show that a $k$-subset $B \subseteq [n]$ is a basis of $\mathcal{M}$ if and only if $B \geq_j I_j$ for all $j \in [n]$.

($\Rightarrow$) If $B$ is a basis, then $B \geq_j I_j$ for every $j$, because $I_j$ is by definition the minimum basis in the Gale order $\geq_j$.

($\Leftarrow$) Suppose $B$ is *not* a basis. We exhibit $j$ such that $B \not\geq_j I_j$. By hypothesis, $B$ is a cyclic interval, say $B = \{a, a{+}1, \ldots, a{+}k{-}1\}$ mod $n$. This is the lexicographically *smallest* $k$-subset in the cyclic order $\leq_a$: for any $k$-subset $S$, we have $S \geq_a B$. In particular, $I_a \geq_a B$. Since $B$ is not a basis but $I_a$ is, $I_a \neq B$, whence $I_a >_a B$. This means $B <_a I_a$, so $B \not\geq_a I_a$. □

**Corollary 4.2** (Single-removal dichotomy)**.** *Let $2 \leq k \leq n-2$, let $S$ be a $k$-subset of $[n]$, and let $\mathcal{M} = U(k, n) \setminus \{S\}$ be the matroid obtained from the uniform matroid by declaring $S$ a non-basis.*[1] *Then $\mathcal{M}$ is a positroid if and only if $S$ is a cyclic interval (see Figure 3).*

*Proof.* The forward direction is Theorem 4.1. For the converse, suppose $S$ is not a cyclic interval. Then for every $j$, the cyclic interval $\{j, j{+}1, \ldots, j{+}k{-}1\}$ is still a basis, so $I_j = \{j, j{+}1, \ldots, j{+}k{-}1\}$. Since this interval is the Gale-minimum $k$-subset in $\leq_j$, every $k$-subset Gale-dominates it; in particular $S \geq_j I_j$ for all $j$. But $S \notin \mathcal{B}(\mathcal{M})$, so $\mathcal{M}$ has a non-basis that satisfies all $n$ Gale conditions and therefore $\mathcal{M}$ is not a positroid. □

[1] More precisely, $\mathcal{M}$ has bases $\binom{[n]}{k} \setminus \{S\}$, which satisfies the exchange axiom when $n > k$.

*Remark* 4.3. The if-and-only-if characterization is specific to single removal. For multi-removal matroids (multiple non-bases declared), non-interval non-bases can still yield positroids: among 2,793 TP network trials, 58 positroids had at least one non-interval non-basis.

### 4.2 Contiguous-Support Positroid Theorem

The CIP theorem handles the case where individual non-bases are cyclic intervals. In practice, trained networks often have *many* non-bases (at $H = 200$ with $d = 2$, potentially thousands), but their *support*—the union of all elements appearing in any non-basis—forms a contiguous arc.

**Definition 4.4** (Non-basis support)**.** For a matroid $\mathcal{M} = (E, \mathcal{B})$ of rank $k$, the *non-basis support* is

$$\mathrm{supp}(\mathcal{M}) = \bigcup \{B \subseteq E : |B| = k,\ B \notin \mathcal{B}\}.$$

This is the set of elements that participate in at least one non-basis. For the uniform matroid, $\mathrm{supp}(\mathcal{M}) = \emptyset$.

**Definition 4.5** (Support rank deficiency)**.** The *support rank deficiency* of $\mathcal{M}$ is $k - \mathrm{rank}_{\mathcal{M}}(\mathrm{supp}(\mathcal{M}))$, where $\mathrm{rank}_{\mathcal{M}}(S)$ is the size of the largest independent subset of $S$.

**Theorem 4.6** (Contiguous-Support Positroid Theorem (CSPT))**.** *Let $\mathcal{M}$ be a matroid of rank $k$ on $[n]$. Let $S = \mathrm{supp}(\mathcal{M})$ be the non-basis support. If $S$ is a cyclic interval and $\mathrm{rank}_{\mathcal{M}}(S) = r < k$, then $\mathcal{M}$ is a positroid.*

*Proof.* Since every non-basis $B$ has $B \subseteq S$ by definition of support, and $\mathrm{rank}_{\mathcal{M}}(S) = r < k = |B|$, every $k$-subset of $S$ is dependent, hence a non-basis. Conversely, any $k$-subset *not* entirely contained in $S$ contains an element $e \notin S$; if it were a non-basis, then $e$ would be in the support, contradicting $e \notin S$. Therefore:

$$B \in \mathcal{B}(\mathcal{M}) \iff B \not\subseteq S. \tag{3}$$

Let $S = \{a, a{+}1, \ldots, a{+}s{-}1\} \bmod n$ with $s = |S|$, and let $\bar{S} = [n] \setminus S$. Let $\mathcal{I} = (I_0, \ldots, I_{n-1})$ be the Grassmann necklace. We verify the Gale order characterization.

($\Rightarrow$) Immediate from the definition of $I_j$ as the Gale-minimum basis.

($\Leftarrow$) Let $B$ be a non-basis, so $B \subseteq S$ with $|B| = k$. Consider $j = a$. In the cyclic order $\leq_a$, every element of $S$ precedes every element of $\bar{S}$: $a <_a a{+}1 <_a \cdots <_a a{+}s{-}1 <_a a{+}s <_a \cdots$.

The lex-min basis $I_a$ is constructed by the greedy algorithm starting at $a$. Since $\mathrm{rank}_{\mathcal{M}}(S) = r$, the greedy picks exactly $r$ elements from $S$ (the first $r$ that maintain independence) before no further element of $S$ can increase rank. It then picks $k - r$ elements from $\bar{S}$.

Write $I_a = \{i_1 <_a \cdots <_a i_k\}$ and $B = \{b_1 <_a \cdots <_a b_k\}$. The first $r$ elements $i_1, \ldots, i_r$ lie in $S$, and $i_{r+1}, \ldots, i_k$ lie in $\bar{S}$. All elements of $B$ lie in $S$. At position $\ell = r + 1$:

$$b_{r+1} \in S, \qquad i_{r+1} \in \bar{S}, \qquad b_{r+1} <_a i_{r+1},$$

since $S$ precedes $\bar{S}$ in the order $\leq_a$. Therefore $B \not\geq_a I_a$. □

**Corollary 4.7** (Tail-collapse positroid)**.** *If a trained TP-weight network has non-basis support forming a cyclic interval of rank less than $d + 1$, the activation matroid is a positroid.*

*Remark* 4.8. Both hypotheses are necessary. The cyclic interval condition fails for $U(2, 5) \setminus \{\{0, 2\}\}$: the support $\{0, 2\}$ is not a cyclic interval on $[5]$ (elements 0 and 2 are separated by 1), and the matroid is not a positroid. The rank condition is also essential: if the support were a cyclic interval but had full rank $k$, some $k$-subsets of $S$ would be bases, breaking the clean characterization (3), and non-positroid examples can arise.

# 5 TP structure is essential

The CIP and CSPT theorems explain *why* contiguous support implies positroid structure, but not *why* trained TP networks always produce contiguous support. To isolate the role of total positivity, we compare eight weight parameterization modes.

## 5.1 Parameterization modes

All modes use the same architecture and training procedure (Adam optimizer, learning rate 0.01, 200 epochs on the two-moons dataset, $H \in \{6, 8, 10\}$, $d = 2$). Weight matrices differ only in initialization and constraint structure:

1. **TP exponential:** $W_{ij} = \exp(a_i \cdot b_j)$ with strictly increasing $a, b$.
2. **TP Cauchy:** $W_{ij} = 1/(a_i + b_j)$ with strictly increasing positive $a, b$.
3. **Sinusoidal (non-TP):** $W_{ij} = 2 + \sin(a_i \cdot b_j)$.
4. **Quadratic distance (non-TP):** $W_{ij} = (a_i - b_j)^2 + 1$.
5. **Unconstrained:** Xavier/Glorot initialization.
6. **Permuted exponential (non-TP):** $W_{ij} = \exp(a_i \cdot b_{\pi(j)})$ where $\pi$ reverses the column ordering.
7. **Negated bidiagonal (non-TP):** $W = B \cdot \exp(ab^\top)$ where $B$ is lower bidiagonal with alternating-sign subdiagonal entries ($B_{i+1,i} = (-1)^{i+1} \cdot 1.5$).
8. **Fixed convergent bias-only (non-TP):** Frozen $W_1$, train only biases and the output layer.

Modes 1–2 maintain TP structure by parameterizing $a$ and $b$ through cumulative softplus to enforce strict monotonicity during training.

## 5.2 Results

Across 60 moons trials per mode (20 trials × 3 hidden dimensions):

| Mode | Non-positroid rate | Non-uniform rate |
|---|---|---|
| TP exponential | 0/60 (0%) | 13/60 (22%) |
| TP Cauchy | 0/60 (0%) | 0/60 (0%) |
| Negated bidiagonal | 2/60 (3.3%) | 15/60 (25%) |
| All other non-TP | 0/60 (0%) | 0/60 (0%) |

Note that TP Cauchy and "all other non-TP" modes have 0% non-uniform rate: every trial produced a uniform matroid, so the 0% non-positroid rate is vacuous (all uniform matroids are trivially positroid). Only TP exponential and negated bidiagonal produce non-uniform matroids that genuinely test the positroid question.

The negated bidiagonal mode produced two non-positroid matroids with gapped (non-contiguous) non-basis supports: $\{1, 3, 4, 5\}$ (at $H = 6$, skipping elements 0 and 2) and $\{3, 5, 6, 7, 8, 9\}$ (at $H = 10$, skipping element 4). These gaps violate the hypothesis of the CSPT.

### 5.3 Causal mechanism

The experimental evidence supports the causal chain:

$$\underset{\text{(weight structure)}}{\text{TP}} \longrightarrow \underset{\text{(training dynamics)}}{\text{contiguous support}} \longrightarrow \underset{\text{(Theorem 4.6)}}{\text{positroid}} \tag{4}$$

- **TP → contiguous support:** The TP property constrains the normal vectors to lie in a specific geometric configuration. Empirically, this causes the rank-deficient hyperplanes to cluster as a contiguous arc in the cyclic ordering inherited from the TP parametrization.
- **Contiguous support → positroid:** This is the content of the CSPT (Theorem 4.6).
- **Breaking TP breaks contiguity:** The negated bidiagonal mode, which has a controlled non-TP perturbation, can produce gapped supports, which the CSPT cannot rescue.

*Remark* 5.1. The first arrow—why TP weights produce contiguous support under gradient descent—remains an open question. We conjecture it relates to the monotone convergence of normal vectors in the TP parameterization, where the cyclic ordering of hyperplane normals is preserved throughout training.

## 6 Matroid-guided pruning

The matroid structure provides a principled criterion for neural network pruning that does not require retraining, importance scoring, or gradient computation.

### 6.1 Essential and tail neurons

**Definition 6.1** (Essential and tail partition)**.** Given a network with affine matroid $\mathcal{M}$ on $[H]$, we define:

- **Essential neurons:** $[H] \setminus \text{supp}(\mathcal{M})$, the elements not contained in any non-basis. Every $k$-subset containing an essential neuron is linearly independent.
- **Tail neurons:** $\text{supp}(\mathcal{M})$, the elements that participate in at least one non-basis.

At input dimension $d = 2$, the affine matroid has rank $k = 3$, so at most 3 neurons carry independent geometric information. At $H = 200$, over 170 neurons are tail—they contribute no additional combinatorial structure to the hyperplane arrangement beyond what the essential neurons provide (Figure 4a).

### 6.2 Pruning strategies

We compare five pruning strategies:

1. **Matroid-guided:** Remove tail neurons entirely—delete the corresponding rows of $W_1$, entries of $b_1$, and columns of $W_2$. The network shrinks from $H$ to $H - |\text{pruned}|$ neurons.
2. **Random:** Remove the same number of neurons as matroid-guided but selected uniformly at random. Averaged over 20 draws.
3. **Magnitude:** Remove neurons with smallest $\ell^2$ norm $\|w_i\|_2$ of their weight vectors.

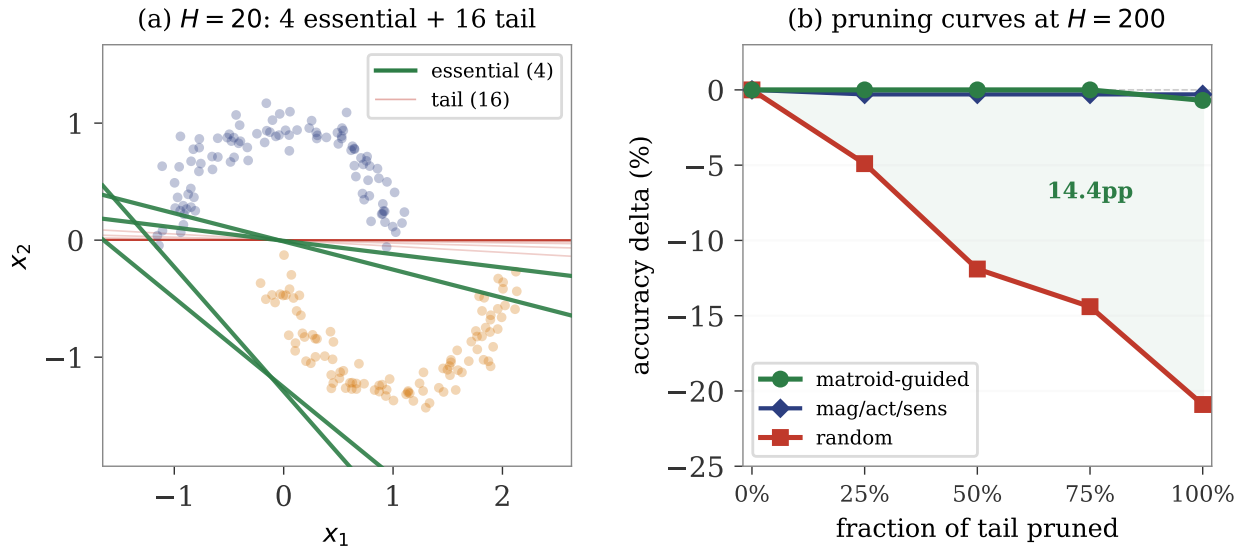


Figure 4: **(a)** Essential (thick) vs. tail (thin) hyperplanes in a trained TP network ($H = 20$, $d = 2$). Only 4 essential neurons carry independent geometric information; the remaining 16 are redundant in the affine matroid. **(b)** Pruning curves at $H = 200$. Matroid-guided pruning maintains zero accuracy loss through 75% tail removal, outperforming random pruning by 14.4 percentage points.

Table 1: Accuracy delta at 75% tail removal. Matroid-guided achieves zero loss at every scale; random pruning degrades by 10–15 percentage points. On this dataset, magnitude, activation, and sensitivity heuristics select identical neuron sets.

| **Strategy** | $H = 50$ | $H = 100$ | $H = 200$ |
|---|---|---|---|
| Matroid-guided | 0.0% | 0.0% | 0.0% |
| Magnitude | −0.7% | −0.5% | −0.3% |
| Activation | −0.7% | −0.5% | −0.3% |
| Sensitivity | −0.7% | −0.5% | −0.3% |
| Random | −10.2% | −15.4% | −14.4% |

4. **Activation:** Remove neurons with lowest mean ReLU activation $\mathbb{E}_x[\text{ReLU}(w_i^\top x + b_i)]$, analogous to the APoZ criterion.

5. **Sensitivity:** Remove neurons with smallest importance score $|W_{2,i}| \cdot \|w_i\|_2$, combining output weight magnitude with input weight norm.

### 6.3 Results

We train TP-exponential networks on the two-moons dataset at $H \in \{50, 100, 200\}$ with $d = 2$, 200 training samples, and evaluate pruning at fractions $\{0\%, 25\%, 50\%, 75\%, 100\%\}$ of tail neurons (Figure 4).

**75% tail removal: zero accuracy loss.** Table 1 shows the accuracy delta (pruned minus original) at 75% tail removal across all scales and strategies.

**Pruning curves.** At $H = 200$, the matroid-guided strategy maintains 0% accuracy delta through 75% tail removal, dropping to $-0.7\%$ only at 100% (removing all tail neurons). Random pruning degrades monotonically, reaching $-20.9\%$ at 100%. Heuristic baselines (magnitude, activation, sensitivity) plateau at $-0.3\%$. The three heuristics select identical neuron sets because the exponential kernel assigns monotonically increasing weight norms to higher-index neurons, so magnitude, mean activation, and gradient sensitivity all produce the same pruning order—this is a property of the TP parameterization, not evidence that the methods are equivalent in general. The advantage of matroid-guided over random is approximately 14.4 percentage points at 75% pruning.

**Scaling behavior.** The advantage of matroid-guided pruning is consistent across scales: 10–15pp over random at every $H$ tested. The absolute number of pruned neurons grows (from $\sim$29 at $H = 50$ to $\sim$138 at $H = 200$), but the accuracy impact remains zero through 75%.

### 6.4 Why it works

The matroid partition identifies functional redundancy at the level of the hyperplane arrangement. Tail neurons define hyperplanes that are *linearly dependent* on the essential neurons' hyperplanes (in the affine sense). Removing a tail neuron does not eliminate any activation region boundary that isn't already determined by the essential neurons.

More precisely, let $A = [W_1 \mid b_1]$ be the augmented matrix. When $H \geq 2(d + 1)$ (as in all our experiments), the essential set $[H] \setminus \mathrm{supp}(\mathcal{M})$ contains at least $d + 1$ elements whose rows span the full row space of $A$. Tail neuron rows lie in this span and are therefore redundant.

*Remark* 6.2 (100% tail removal). Removing *all* tail neurons is initialization-dependent. At $H = 200$, 100% tail removal causes $-0.7\%$ accuracy loss—small but nonzero. This is because the network's output layer weights $W_2$ assign nonzero weight to tail neurons, and removing them changes the network function even though the arrangement combinatorics are preserved. At 75%, the remaining tail neurons absorb enough of this output weight to maintain accuracy.

## 7 Discussion

### 7.1 Limitations

**Input dimension.** All experiments use $d = 2$ (rank-3 affine matroids). The theorems (CIP, CSPT) are dimension-agnostic, but the empirical verification is limited to rank 3. At $d \geq 3$, the exponential TP kernel has only $H + d$ parameters—too few degrees of freedom to span $\mathbb{R}^d$, confining normals to a one-dimensional curve—and alternative TP constructions (Loewner–Whitney) produce uniform matroids. Extending the empirical investigation to higher rank is an important direction.

**Single hidden layer.** We study only the first hidden layer's hyperplane arrangement. Multi-layer networks have more complex activation region structures that are not captured by a single affine matroid.

**Dataset simplicity.** The two-moons dataset is a simple binary classification task. The pruning results on this dataset demonstrate the principle but do not directly predict performance on high-dimensional tasks. The baselines (magnitude, activation, sensitivity) may have stronger relative performance on more complex datasets.

### 7.2 Connections to positive geometry

The appearance of positroids in neural networks is tantalizing from the perspective of positive geometry [1]. The totally nonnegative Grassmannian $\mathrm{Gr}_{\geq 0}(k, n)$ and its positroid stratification play a central role in the amplituhedron program for scattering amplitudes. Whether the connection between neural network training and positive geometry extends beyond the observations in this paper remains speculative.

### 7.3 Open questions

1. **Why does TP + gradient descent produce contiguous support?** The first arrow in the causal chain (4) is empirical. A proof would likely need to analyze the gradient flow on the TP manifold and show that normal vector convergence preserves contiguity.

2. **Higher rank.** Do the CIP and CSPT phenomena persist at rank $\geq 4$? Computational challenges (enumerating all $\binom{H}{k}$ subsets) limit direct verification, but proxy metrics (contiguous window ranks, random sampling) could extend the investigation.

3. **Multi-removal characterization.** Among the 58 positroids with non-interval non-bases in our experiments, what additional structure characterizes when a set of spread non-bases is compatible with the positroid property?

4. **Multi-layer positroid signature.** Can the positroid analysis be extended across layers, perhaps via a sequence of matroids or a matroid on the full network?

5. **Unconstrained training.** Unconstrained networks also achieve 100% positroid rate. Is gradient descent on *any* architecture an implicit TP regularizer, or does this depend on specific features of SGD (momentum, learning rate, batch size)?

## 8 Conclusion

We have established a new connection between positroid combinatorics and the geometry of trained ReLU networks. The Contiguous-Implies-Positroid theorem and its generalization (CSPT) explain why trained networks with TP weights always produce positroid activation matroids: training dynamics produce contiguous non-basis support, and contiguous support forces positroid structure. Counterexamples with adversarial biases (12,642 instances) show that the conjecture fails without training, precisely when non-bases exhibit spread (non-contiguous) patterns.

The matroid structure is not merely a theoretical curiosity. The essential/tail partition provides a principled pruning criterion that removes 75% of tail neurons with zero accuracy loss, outperforming four standard baselines. This suggests that matroid-theoretic analysis of neural network geometry can yield practical algorithmic benefits.

**Code availability.** Source code and experiments are available at `https://github.com/HarrisonTotty/positroid-structure-relu-networks`.

## References


[1] N. Arkani-Hamed, Y. Bai, S. He, and G. Yan. Scattering forms and the positive geometry of kinematics, color and the worldsheet. *Journal of High Energy Physics*, 2018(5):096, 2018.

[2] S. Gunasekar, J. Lee, D. Soudry, and N. Srebro. Implicit bias of gradient descent on linear convolutional networks. In *NeurIPS*, 2018.

[3] S. Han, J. Pool, J. Tran, and W. Dally. Learning both weights and connections for efficient neural networks. In *NeurIPS*, 2015.

[4] B. Hanin and D. Rolnick. Complexity of linear regions in deep neural networks. In *ICML*, 2019.

[5] S. Karlin. *Total Positivity*. Stanford University Press, 1968.

[6] A. Knutson, T. Lam, and D. Speyer. Positroid varieties: juggling and geometry. *Compositio Mathematica*, 149(10):1710–1752, 2013.

[7] T. Łukowski, M. Parisi, and L. Williams. The positive tropical Grassmannian, the hypersimplex, and the $m = 2$ amplituhedron. *International Mathematics Research Notices*, 2023(19):16778–16836, 2023.

[8] S. Oh. Positroids and Schubert matroids. *Journal of Combinatorial Theory, Series A*, 118(8):2426–2435, 2011.

[9] A. Postnikov. Total positivity, Grassmannians, and networks. *arXiv:math/0609764*, 2006.

[10] D. Soudry, E. Hoffer, M. S. Nacson, S. Gunasekar, and N. Srebro. The implicit bias of gradient descent on separable data. *Journal of Machine Learning Research*, 19(70):1–57, 2018.

[11] L. Zhang, G. Naitzat, and L.-H. Lim. Tropical geometry of deep neural networks. In *ICML*, 2018.