

Sharp Rank Laws for Maxout Switch Density

Abstract

We resolve the linear-rank conjecture of Goujon, Etemadi, and Unser [10] for the expected density of active-candidate switches in random Maxout units. Their quadratic bound counts all pairwise candidate intersections; our analysis counts only intersections visible on the upper envelope. A multiplicity-aware area formula and a sharp visibility inequality give a rank-linear bound along every deterministic finite-length rectifiable path. For a rank- K unit in the original iid-coordinate model, with weight-coordinate variance σ_w^2 and independent bias density bounded by B , we prove $\mathbb{E}\lambda \leq (K/\sqrt{3})B\sigma_w$, where λ is the switch count divided by path length. The fixed-rank coefficient $K/\sqrt{3}$ is sharp in the supremal sense, and the optimal universal constant multiplying $(K-1)B\sigma_w$ is $2/\sqrt{3}$. The argument extends to weight-dependent conditional bias-density envelopes and independent heterogeneous candidates. For iid copies of a jointly Gaussian weight-bias vector, we derive an exact spherical-length formula whenever score variance is nonzero along the path. Its coefficient is $c_K = \mathbb{E} \max_{j \leq K} G_j / \sqrt{\pi} \sim \sqrt{2 \log K / \pi}$, where G_j are iid standard normals. This identifies a sharp contrast between worst-case linear rank dependence and Gaussian square-root-logarithmic growth. The single-unit laws also yield explicit layerwise switch-density bounds for networks with independent layers, linking the resolved conjecture to pathwise complexity in deep random networks.

Note. This paper was generated entirely by AI, including MiMo, using an automated research pipeline developed by Chenghua Liu and Hanyu Li.

1 Introduction

Maxout units take the maximum of finitely many affine functions [9]. They are basic convex piecewise-affine maps, and their active candidates determine the linear regions of Maxout networks; this viewpoint is closely related to the affine-spline description of deep networks [3, 4]. Much of the expressivity literature therefore counts regions or activation patterns globally or in the worst case [12, 17, 18, 23]. A complementary line of work measures how many region boundaries a one-dimensional probe encounters. This pathwise viewpoint is both dimensionally economical and well suited to random initialization [10, 11, 19].

For a rank- K unit, the direct argument examines every one of the $\binom{K}{2}$ equality hyperplanes. Proposition 6 of Goujon et al. [10] consequently gives a quadratic rank factor and is followed by the conjecture that one power of K should disappear. The obstruction to the all-pairs count is simple but quantitative: most pairwise intersections lie below a third candidate and never change the Maxout output. At a common height y , the probability that the other $K-2$ candidates lie below the pair is $F(y)^{K-2}$, where F is the common score CDF. Our main lemma shows that integrating this visibility factor under a bounded conditional-density envelope saves the sharp factor $1/(K-1)$.

Our contributions follow two complementary proof routes. For independent candidates, a multiplicity-aware area formula counts only pair crossings visible on the upper envelope. In the iid case, a sharp visibility inequality turns this identity into a rank-linear bound in a tilted Gini mean-difference path functional. Under the original iid-coordinate assumptions, the sharp supremal coefficient is $K/\sqrt{3}$ at fixed rank, and the sharp constant in the uniform $(K-1)$ -normalization is $2/\sqrt{3}$. The same argument yields an aggregate candidate-additive bound for heterogeneous

candidates. It requires no density assumption on the weights, while retaining conditional bias anti-concentration.

For iid Gaussian weight–bias vectors whose score variance is nonzero along the probe, a marked spherical Crofton argument gives an exact c_K -times-spherical-length law. We give finite-rank bounds, a second-order asymptotic, and a planar Gaussian-polytope interpretation of c_K . Both single-unit results feed into a layerwise switch-and-length recursion; for Gaussian layers, an exact active-weight second moment sharpens the generic length factor.

Visibility markings have classical antecedents in random convex-hull geometry [8, 20], and the deep recursion builds on existing composition and curve-length arguments [10, 25]. The new contributions are the sharp bounded-density visibility inequality, its distribution-free constants, the Gaussian upper-envelope coefficient, and their consequences for Maxout networks.

2 Setup and main results

This section formalizes the pathwise switch count and states the distribution-free and Gaussian results used later. We begin with the iid model, record heterogeneous and random-probe extensions, and then state the exact Gaussian law.

For $K \geq 2$, let

$$M_K(x) := \max_{1 \leq j \leq K} (W_j^\top x + \beta_j), \quad x \in \mathbb{R}^d, \quad (2.1)$$

where the candidate pairs (W_j, β_j) are independent and identically distributed unless stated otherwise. We call K , the number of affine candidates, the Maxout rank. The case $K = 1$ is affine and has no active-candidate switches.

Let $\gamma : [0, L] \rightarrow \mathbb{R}^d$ be a deterministic rectifiable path of positive finite length $L = \ell(\gamma) = \text{Len}(\gamma)$, parameterized by arclength. Thus γ is 1-Lipschitz and its derivative $u(s) := \dot{\gamma}(s)$ exists with $\|u(s)\|_2 = 1$ for almost every s . Self-intersections and retracing are retained with parameter multiplicity. Continuous piecewise-linear (CPWL) paths are a special case.

For distinct $a, b \in \mathbb{R}^d$, let $[a, b]$ also denote the arclength parametrization $s \mapsto a + s(b-a)/\|b-a\|_2$, $0 \leq s \leq \|b-a\|_2$. Thus $\text{kt}_{M_K}^{[a,b]}$ and $\lambda_{M_K}^{[a,b]}$ refer to this affine inclusion path. For $a = b$, its length and switch count are defined to be zero.

Definition 2.1 (Active-candidate switches and path-averaged density). Put $f_j(s) = W_j^\top \gamma(s) + \beta_j$. An active switch is a parameter value $s \in (0, L)$ at which exactly two candidates tie at the maximum and their difference changes sign across s : in some neighborhood of s , the difference is nonzero on either side and has opposite signs on the two sides. The active-switch count $\text{kt}_{M_K}^\gamma \in \mathbb{N}_0 \cup \{+\infty\}$ is the number of such parameter values. Endpoints are not counted, while self-intersections and retracing are counted at each distinct parameter value. Intersections of two inactive candidates are not counted. The path-averaged switch density is

$$\lambda_{M_K}^\gamma := \frac{\text{kt}_{M_K}^\gamma}{\ell(\gamma)}. \quad (2.2)$$

Under the hypotheses of Lemma 3.1, the relevant equality set is almost surely finite, and every equality is interior, involves a unique pair, and is transverse. Hence the definition equals the number of changes of the unique maximizing index. For a CPWL probe it agrees almost surely with the projection-region change count in Definition 6 of Goujon et al. [10].

Throughout, an active switch changes the active candidate within one unit, a pathwise breakpoint is a parameter value at which an output curve ceases to be locally affine, and a global activation region is a cell in input space. These objects need not coincide.

Proposition 6 of Goujon et al. [10] counts all candidate equality hyperplanes and gives the distribution-free estimate $\sqrt{2}\binom{K}{2} \|\rho_\beta\|_\infty \sigma_w$, and the Gaussian estimate $\sqrt{2}\binom{K}{2} \sigma_w / (\pi \sigma_\beta)$. The authors conjecture immediately after their equation (31) that one factor of the rank should be removable.

Open problem 2.2 (Rank-linear Maxout knot density). Suppose the coordinates of W are iid, their common law has a density and finite variance σ_w^2 , and the independent bias β has a bounded density ρ_β . Is there a universal constant C such that, for every K, d , every admissible law, and every deterministic finite-length CPWL path γ ,

$$\mathbb{E} \lambda_{M_K}^\gamma \leq C(K-1) \sigma_w \|\rho_\beta\|_\infty? \quad (2.3)$$

What is the optimal rank dependence and constant, and what sharper statement holds for Gaussian parameters?

Our weighted iid result allows the density envelope to depend on the weight. Let μ be the law of W , and suppose that, for μ -almost every w , a regular conditional law of β given $W = w$ is absolutely continuous. Choose a jointly measurable conditional density. After modifying it on a $\mu(dw) db$ -null set, and defining it arbitrarily on a μ -null set of weights, we take the representative satisfying

$$\mathbb{P}[\beta \in db \mid W = w] = g(b \mid w) db, \quad 0 \leq g(b \mid w) \leq b_0(w), \quad (2.4)$$

for every b and μ -almost every w , where $b_0 : \mathbb{R}^d \rightarrow \mathbb{R}$ is measurable and nonnegative, and

$$0 < \bar{B} := \mathbb{E} b_0(W) < \infty, \quad \frac{d\tilde{\mu}}{d\mu}(w) := \frac{b_0(w)}{\bar{B}}. \quad (2.5)$$

Write $\widetilde{W}, \widetilde{W}'$ for independent draws from $\tilde{\mu}$, and $\widetilde{\Sigma} = \text{Cov}(\widetilde{W})$ when it exists. The envelope is imposed on the conditional density, not merely on the marginal density of β .

The tilt in (2.5) charges each weight value according to its conditional bias-concentration envelope. The first main theorem shows that the resulting mean-difference functional controls visible switches with linear rank dependence.

Theorem 2.3 (Sharp weighted iid visibility bound). *Under (2.4)–(2.5), every deterministic finite-length rectifiable path satisfies*

$$\mathbb{E} \text{kt}_{M_K}^\gamma \leq \frac{K\bar{B}}{2} \int_0^L \mathbb{E} \left| u(s)^\top (\widetilde{W} - \widetilde{W}') \right| ds. \quad (2.6)$$

If $\widetilde{W}$ has finite covariance, then

$$\mathbb{E} \text{kt}_{M_K}^\gamma \leq \frac{K\bar{B}}{\sqrt{3}} \int_0^L \sqrt{u(s)^\top \widetilde{\Sigma} u(s)} ds. \quad (2.7)$$

In particular, if $b_0(w) \equiv B$ and $\Sigma = \text{Cov}(W)$, then

$$\mathbb{E} \lambda_{M_K}^\gamma \leq \frac{KB}{\sqrt{3}} \sqrt{\lambda_{\max}(\Sigma)}. \quad (2.8)$$

Remark 2.4 (Minimal moment requirement). Finite tilted covariance is used only in (2.7). The Gini mean-difference estimate (2.6) needs only the displayed directional first moments and may apply when the original or tilted law has infinite variance; without them, it is understood as an inequality in $[0, +\infty]$. The constant-envelope case $b_0 \equiv B$ recovers the original weight law.

We now specialize the weighted theorem to the original iid-coordinate class and separate the sharp fixed-rank coefficient from the constant uniform in rank.

Corollary 2.5 (Solution of the open problem and sharp constants). *Under the assumptions of open problem 2.2, put $B = \|\rho_\beta\|_\infty$. Then*

$$\mathbb{E}\lambda_{M_K}^\gamma \leq \frac{K}{\sqrt{3}}B\sigma_w \leq \frac{2}{\sqrt{3}}(K-1)B\sigma_w. \quad (2.9)$$

For each fixed $K \geq 2$ and $d \geq 1$, the first coefficient is sharp in the supremal sense:

$$\sup \frac{\mathbb{E}\lambda_{M_K}^\gamma}{B\sigma_w} = \frac{K}{\sqrt{3}}, \quad (2.10)$$

where the supremum ranges over admissible laws on $\mathbb{R}^d$ with $B\sigma_w > 0$ and positive-length deterministic rectifiable paths. Hence the sharp distribution-free rank order is $\Theta(K)$, and the optimal universal constant in (2.3) is

$$C_* = \frac{2}{\sqrt{3}}. \quad (2.11)$$

The iid assumption is not needed to replace an all-pairs count by a candidate-additive bound. The next result isolates the appropriate aggregate for heterogeneous candidates.

Theorem 2.6 (Independent heterogeneous candidates). *Let the candidate pairs (W_i, β_i) , $1 \leq i \leq K$, be mutually independent but not necessarily identically distributed, and let P_i be the law of W_i . Suppose $\beta_i \mid W_i = w$ admits a jointly measurable density $g_i(b \mid w)$ satisfying $0 \leq g_i(b \mid w) \leq B_i < \infty$ for every b and for P_i -almost every w , and $\mathbb{E}\|W_i\| < \infty$. Then*

$$\mathbb{E} \text{kt}_{M_K}^\gamma \leq \int_0^L \inf_{c \in \mathbb{R}} \sum_{i=1}^K B_i \mathbb{E} \left| u(s)^\top W_i - c \right| ds. \quad (2.12)$$

If W_i has mean μ_i and covariance Σ_i , then the integrand is at most

$$\inf_{c \in \mathbb{R}} \sum_{i=1}^K B_i \sqrt{u(s)^\top \Sigma_i u(s) + (u(s)^\top \mu_i - c)^2}. \quad (2.13)$$

Here c is a common one-dimensional reference location for the projected slopes. In particular, a common mean removes the location term. For $K = 2$, the pointwise integrand bound in (2.12) is sharp in the shrinking-segment supremal sense.

Two further consequences make the pathwise nature of the iid estimate explicit. They are useful when the probe itself is random or when several Maxout units are evaluated in parallel.

Corollary 2.7 (Random probes and parallel Maxout units). *In the constant-envelope case of Theorem 2.3, let Γ be an almost surely rectifiable path of finite positive length, represented by a jointly measurable arclength parametrization and independent of all candidate parameters. Write $u_\Gamma(s) := \dot{\Gamma}(s)$ for its almost-everywhere tangent. Suppose the expectation below is finite. Then*

$$\mathbb{E} \text{kt}_{M_K}^\Gamma \leq \frac{KB}{\sqrt{3}} \mathbb{E}_\Gamma \int_0^{\ell(\Gamma)} \sqrt{u_\Gamma(s)^\top \Sigma u_\Gamma(s)} ds. \quad (2.14)$$

Moreover, let $\mathbf{M} = (M_{K_1}^{(1)}, \dots, M_{K_q}^{(q)})$ be a collection of Maxout units on $\mathbb{R}^d$. Within unit r , suppose the candidates are iid and satisfy Theorem 2.3 with constants B_r, Σ_r ; no independence between different units is required. If γ is any deterministic positive-length rectifiable path and $\text{kt}_{\mathbf{M}}^\gamma$ counts changes of the joint active-index tuple, then

$$\mathbb{E} \text{kt}_{\mathbf{M}}^\gamma \leq \frac{1}{\sqrt{3}} \int_0^L \sum_{r=1}^q K_r B_r \sqrt{u(s)^\top \Sigma_r u(s)} ds. \quad (2.15)$$

Proof. Condition on the random path, use independence, and apply Theorem 2.3; Tonelli's theorem gives (2.14). For (2.15), every change of the joint active-index tuple is a change in at least one component, so $\text{kt}_{\mathbf{M}}^\gamma \leq \sum_r \text{kt}_{M_{K_r}^{(r)}}^\gamma$. Apply Theorem 2.3 to each marginal unit and use linearity of expectation. $\square$

The Gaussian model admits a smaller, exactly identifiable rank factor. Let ϕ and Φ denote the standard-normal density and CDF, and let $G_1, \dots, G_K$ be iid standard normal variables. Put $M_K^G := \max_{j \leq K} G_j$. We write $\mathbb{S}^d = \{z \in \mathbb{R}^{d+1} : \|z\|_2 = 1\}$, and $\text{Len}_{\mathbb{S}^d}$ for parameterized spherical arclength, including retracing multiplicity. Define

$$c_K := \frac{\mathbb{E} \max_{1 \leq j \leq K} G_j}{\sqrt{\pi}}. \quad (2.16)$$

Theorem 2.8 (Jointly Gaussian spherical-length law). *Let $\Theta = (W, \beta) \in \mathbb{R}^{d+1}$ be Gaussian with arbitrary mean and covariance $\Omega \succeq 0$, and let the K candidates be iid copies. Put $q_x = (x, 1)$, and assume*

$$q_{\gamma(s)}^\top \Omega q_{\gamma(s)} > 0, \quad 0 \leq s \leq L. \quad (2.17)$$

For every x with $q_x^\top \Omega q_x > 0$, define

$$\Psi_\Omega(x) := \frac{\Omega^{1/2} q_x}{\sqrt{q_x^\top \Omega q_x}} \in \mathbb{S}^d, \quad (2.18)$$

where $\Omega^{1/2}$ is the unique symmetric positive-semidefinite principal square root. Condition (2.17) places the whole probe in this domain. Then γ obeys

$$\boxed{\mathbb{E} \text{kt}_{M_K}^\gamma = c_K \text{Len}_{\mathbb{S}^d}(\Psi_\Omega \circ \gamma)}. \quad (2.19)$$

The spherical length counts self-intersections and retracing with parameter multiplicity. The formula permits correlation between weights and biases and singular Ω ; only the nonvanishing condition (2.17) is required along the probe.

Corollary 2.9 (Independent anisotropic Gaussian parameters). *Suppose $W \sim N(\mu_w, \Sigma)$ and $\beta \sim N(\mu_\beta, \sigma_\beta^2)$ independently, with $\sigma_\beta > 0$ and possibly singular $\Sigma \succeq 0$. Define*

$$\Psi_\Sigma(x) := \frac{(\Sigma^{1/2} x, \sigma_\beta)}{\sqrt{\sigma_\beta^2 + x^\top \Sigma x}}. \quad (2.20)$$

Then (2.19) holds with Ψ_Ω replaced by Ψ_Σ , and

$$\mathbb{E} \lambda_{M_K}^\gamma \leq c_K \frac{\sqrt{\lambda_{\max}(\Sigma)}}{\sigma_\beta}. \quad (2.21)$$

When $\lambda_{\max}(\Sigma) > 0$, this coefficient is sharp in the precise sense that

$$\sup_{\gamma} \frac{\mathbb{E} \lambda_{M_K}^{\gamma}}{\sqrt{\lambda_{\max}(\Sigma)}/\sigma_{\beta}} = c_K, \quad (2.22)$$

where the supremum is over deterministic positive-length rectifiable paths. Moreover,

$$c_K \sim \sqrt{\frac{2 \log K}{\pi}}. \quad (2.23)$$

In the isotropic case $\Sigma = \sigma_w^2 I_d$,

$$\mathbb{E} \lambda_{M_K}^{\gamma} \leq \frac{K-1}{\pi} \frac{\sigma_w}{\sigma_{\beta}}, \quad (2.24)$$

and $1/\pi$ is the optimal universal constant in this rank-linear form, uniformly over $K \geq 2$.

The proofs have two independent branches. The distribution-free results use a multiplicity area formula followed by a sharp visibility inequality, whereas the Gaussian identity uses a marked spherical Crofton argument. Both single-unit laws then enter the layerwise recursion.

3 Rectifiable visible-crossing formula

We derive the marked-crossing identity underlying all distribution-free bounds. We first treat independent heterogeneous candidates; the iid formula then follows by symmetry.

For candidate i , let P_i be the law of W_i . We use a jointly measurable conditional density representative satisfying $0 \leq g_i(b | w) \leq b_i(w)$ for every b and P_i -almost every w , where $\mathbb{E} b_i(W_i) < \infty$. For almost every $s \in (0, L)$, set

$$A_{i,s} := W_i^{\top} u(s), \quad S_{i,s} := W_i^{\top} \gamma(s), \quad Y_{i,s} := S_{i,s} + \beta_i, \quad (3.1)$$

and let $F_{i,s}$ be the CDF of $Y_{i,s}$. Define the finite measure

$$\nu_{i,s,y}(E) := \mathbb{E} \left[\mathbf{1}_{\{A_{i,s} \in E\}} g_i(y - S_{i,s} | W_i) \right] \quad (3.2)$$

and its mass $h_{i,s}(y) := \nu_{i,s,y}(\mathbb{R})$. The envelope gives

$$\nu_{i,s,y}(\mathbb{R}) \leq \mathbb{E} b_i(W_i) < \infty. \quad (3.3)$$

Thus $dF_{i,s}(y) = h_{i,s}(y) dy$. For $i \neq j$, put

$$R_{ij,s}(y) := \iint_{\mathbb{R}^2} |a - a'| \nu_{i,s,y}(da) \nu_{j,s,y}(da'). \quad (3.4)$$

This is a nonnegative product integral, possibly equal to $+\infty$. All these quantities admit jointly measurable versions because they are integrals of nonnegative jointly measurable functions. Empty products and $F(y)^0$ are defined as 1, including when $F(y) = 0$.

The next identity is the counting interface for all distribution-free bounds: it integrates a pair crossing only when every other candidate lies below its common height.

Lemma 3.1 (Exact rectifiable visible-crossing identity). *For mutually independent candidates, suppose $g_i(b \mid w) \leq b_i(w) < \infty$ for every b and $\mathbf{P}_i$ -almost every w , and $\mathbb{E}b_i(W_i) < \infty$. Then*

$$\mathbb{E} \text{kt}_{M_K}^\gamma = \sum_{1 \leq i < j \leq K} \int_0^L \int_{\mathbb{R}} \left(\prod_{k \neq i,j} F_{k,s}(y) \right) R_{ij,s}(y) dy ds. \quad (3.5)$$

The equality holds in $[0, \infty]$. Under the assumptions of Theorem 2.6, both sides are finite. In the weighted iid case they are finite whenever the directional first-moment integral in (2.6) is finite, in particular when $\tilde{\Sigma}$ is finite. In the iid case, writing $F_s = F_{i,s}$, $\nu_{s,y} = \nu_{i,s,y}$, and

$$R_s(y) := \iint |a - a'| \nu_{s,y}(da) \nu_{s,y}(da'), \quad (3.6)$$

the formula becomes

$$\mathbb{E} \text{kt}_{M_K}^\gamma = \binom{K}{2} \int_0^L \int_{\mathbb{R}} F_s(y)^{K-2} R_s(y) dy ds. \quad (3.7)$$

Proof. For each pair, we apply the multiplicity area formula, attach the visibility mark from the remaining candidates, and then exclude degenerate intersections. Fix $i < j$, fixed weights w_i, w_j , and $D_R = (0, L) \times [-R, R]$. The map

$$\Phi_{w_i, w_j}(s, y) := (y - w_i^\top \gamma(s), y - w_j^\top \gamma(s)) \quad (3.8)$$

is Lipschitz. Its two-dimensional Jacobian exists for almost every (s, y) and equals

$$J\Phi_{w_i, w_j}(s, y) = \left| (w_i - w_j)^\top u(s) \right|. \quad (3.9)$$

For every nonnegative measurable mark a , the multiplicity area formula [7, Theorem 3.2.3(2), p. 243] gives

$$\int_{D_R} a(s, y) J\Phi_{w_i, w_j}(s, y) ds dy = \int_{\mathbb{R}^2} \sum_{\substack{(s, y) \in D_R: \\ \Phi_{w_i, w_j}(s, y) = (\xi_i, \xi_j)}} a(s, y) d\xi_i d\xi_j. \quad (3.10)$$

Thus a crossing time and common height are mapped to the two required biases, with every parameter preimage counted. Weight the bias integral by $g_i(\xi_i \mid w_i)g_j(\xi_j \mid w_j)$, apply the area formula, and then average over w_i, w_j . Using (3.9), the contribution at (s, y) is precisely $R_{ij,s}(y)$. Letting $R \uparrow \infty$ is justified by monotone convergence.

No conditioning on a probability-zero equality event is involved. For fixed (s, y) , independence says directly that all remaining candidates are strictly below y with probability $\prod_{k \neq i,j} F_{k,s}(y)$; strict and weak inequalities agree because each $Y_{k,s}$ has a density. Multiplying by this mark in the area formula gives the (i, j) summand of (3.5).

We finally dispose of degeneracies in four steps. First, the product of the null set where γ is not differentiable with $[-R, R]$ is a two-dimensional null set, and its Lipschitz image under (3.8) is null. Second, applying (3.10) to the zero-Jacobian set shows that almost every bias pair has no preimage there. Conditional absolute continuity of (β_i, β_j) therefore makes all remaining equalities transverse. Third, conditional on fixed weights, the expected number of pair equalities is at most

$$\min\{b_i(w_i), b_j(w_j)\} \int_0^L \left| (w_i - w_j)^\top u(s) \right| ds < \infty,$$

so its equality set is almost surely finite. Fourth, conditional on the full parameters of the pair and this finite set, an independent third candidate has an absolutely continuous value at every equality time; a finite union then excludes triple ties. The same argument excludes endpoint ties.

Every remaining visible transverse equality changes the upper-envelope maximizer, and every active switch belongs to a unique pair. Summing over pairs proves the lemma. Tonelli's theorem justifies all nonnegative integrations. This is an area-formula proof of a marked Kac–Rice identity; see also Azaïs and Wschebor [2] for the probabilistic framework. $\square$

4 Sharp visibility inequalities

The power $F_s(y)^{K-2}$ in (3.7) removes the inactive intersections. We first quantify this saving, then convert the remaining Gini mean difference into a covariance bound and derive the heterogeneous estimate. No symmetry, log-concavity, or tail assumption is needed for the visibility step.

Lemma 4.1 (Weighted distribution-free visibility estimate). *Fix a parameter s where $u(s)$ exists, suppress s from the notation, and let $n \geq 0$ be an integer. Under (2.4)–(2.5), put $\tilde{A} = u(s)^\top \tilde{W}$, and let $\tilde{A}'$ be an independent copy. Then*

$$\int_{\mathbb{R}} F(y)^n R(y) dy \leq \frac{\bar{B}}{n+1} \mathbb{E} |\tilde{A} - \tilde{A}'|. \quad (4.1)$$

Proof. We split the projected-slope mass at a threshold, bound the mass on opposite sides, and integrate those bounds through the layer-cake identity. Write $A = u(s)^\top W$ and $S = W^\top \gamma(s)$, and let h be the density of $S + \beta$. For a threshold $r \in \mathbb{R}$, define

$$\tilde{p}_r := \mathbb{P}[\tilde{A} > r], \quad (4.2)$$

$$q_r^+(y) := \mathbb{E}[\mathbf{1}_{\{A > r\}} g(y - S | W)], \quad (4.3)$$

$$q_r^-(y) := \mathbb{E}[\mathbf{1}_{\{A \leq r\}} g(y - S | W)]. \quad (4.4)$$

Then $q_r^+ + q_r^- = h$, and the pointwise density bound gives

$$q_r^+(y) \leq \mathbb{E}[\mathbf{1}_{\{A > r\}} b_0(W)] = \bar{B} \tilde{p}_r, \quad q_r^-(y) \leq \bar{B}(1 - \tilde{p}_r). \quad (4.5)$$

Since F is absolutely continuous with density h ,

$$T := \int_{\mathbb{R}} F(y)^n h(y) dy = \left[\frac{F(y)^{n+1}}{n+1} \right]_{-\infty}^{\infty} = \frac{1}{n+1}. \quad (4.6)$$

Put

$$U_r := \int_{\mathbb{R}} F(y)^n q_r^+(y) dy, \quad H_r := \int_{\mathbb{R}} F(y)^n q_r^+(y) q_r^-(y) dy.$$

Using the first and second bounds in (4.5), respectively, gives

$$H_r \leq \bar{B} \tilde{p}_r (T - U_r), \quad H_r \leq \bar{B}(1 - \tilde{p}_r) U_r. \quad (4.7)$$

If $U_r \leq \tilde{p}_r T$, use the second estimate; otherwise use the first. In both cases,

$$H_r \leq \frac{\bar{B}}{n+1} \tilde{p}_r (1 - \tilde{p}_r). \quad (4.8)$$

The layer-cake identity

$$|a - a'| = \int_{\mathbb{R}} (\mathbf{1}_{\{a \leq r < a'\}} + \mathbf{1}_{\{a' \leq r < a\}}) dr$$

and Tonelli's theorem yield

$$\begin{aligned} \int_{\mathbb{R}} F(y)^n R(y) dy &= 2 \int_{\mathbb{R}} H_r dr \\ &\leq \frac{2\overline{B}}{n+1} \int_{\mathbb{R}} \tilde{p}_r (1 - \tilde{p}_r) dr \\ &= \frac{\overline{B}}{n+1} \mathbb{E} |\tilde{A} - \tilde{A}'|, \end{aligned} \tag{4.9}$$

where the last equality applies the same layer-cake formula to the common tilted law of $\tilde{A}, \tilde{A}'$. Integrating first over $r \in [-R, R]$ and then taking $R \uparrow \infty$ proves the inequality as an extended-valued statement when the first moment is infinite. $\square$

The factor $1/(n+1)$ is pointwise sharp at the base point. Indeed, in the constant-envelope case, if $S = 0$ almost surely and β is independent of W and uniform on an interval of length $1/B$, then $R(y) = \rho_\beta(y)^2 \mathbb{E} |A - A'|$ and $\int F(y)^n \rho_\beta(y)^2 dy = B/(n+1)$, so equality holds in (4.1), where A' is an independent copy of A . At $n = K - 2$, this sharp factor cancels one rank factor before the candidate pairs are summed.

The visibility estimate leaves a Gini mean difference of the projected tilted weights. The next sharp inequality converts that term into the covariance bound in Theorem 2.3.

Lemma 4.2 (Sharp Gini–standard-deviation inequality). *For every square-integrable real random variable X and an independent copy X' ,*

$$\mathbb{E} |X - X'| \leq \frac{2}{\sqrt{3}} \sqrt{\text{Var}(X)}. \tag{4.10}$$

The constant is sharp, with equality for every nondegenerate uniform law.

Proof. Let Q be a quantile function of X and let $m = \mathbb{E}X$. The standard quantile identity for the Gini mean difference, followed by Cauchy–Schwarz, gives

$$\begin{aligned} \mathbb{E} |X - X'| &= 2 \int_0^1 (2s - 1) Q(s) ds \\ &= 2 \int_0^1 (2s - 1) (Q(s) - m) ds \\ &\leq 2 \left(\int_0^1 (2s - 1)^2 ds \right)^{1/2} \left(\int_0^1 (Q(s) - m)^2 ds \right)^{1/2} \\ &= \frac{2}{\sqrt{3}} \sqrt{\text{Var}(X)}. \end{aligned}$$

An affine quantile function, equivalently a uniform law up to location and scale, attains equality. See also La Haye and Zizler [15]. $\square$

Proof of Theorem 2.6. Fix s and abbreviate $A_i = u(s)^\top W_i$. For $c \in \mathbb{R}$, define

$$m_{i,s}(y; c) := \int_{\mathbb{R}} |a - c| \nu_{i,s,y}(da).$$

The triangle inequality gives

$$R_{ij,s}(y) \leq m_{i,s}(y; c) h_{j,s}(y) + h_{i,s}(y) m_{j,s}(y; c). \quad (4.11)$$

Writing $P_{ij,s}(y) = \prod_{k \neq i,j} F_{k,s}(y)$, summing the two oriented terms in (4.11) yields

$$\begin{aligned} \sum_{i < j} P_{ij,s} R_{ij,s} &\leq \sum_{i=1}^K m_{i,s}(y; c) \sum_{j \neq i} h_{j,s}(y) P_{ij,s}(y) \\ &= \sum_{i=1}^K m_{i,s}(y; c) \left(\prod_{k \neq i} F_{k,s}(y) \right)'. \end{aligned} \quad (4.12)$$

The density envelope implies

$$m_{i,s}(y; c) \leq B_i \mathbb{E} |A_i - c|. \quad (4.13)$$

Each product of $K - 1$ CDFs rises from zero to one. Integrating (4.12) over y , using (4.13), and then using Lemma 3.1 gives the integrand in (2.12). The infimum may be restricted to rational c , which also verifies measurability as a function of s .

Cauchy–Schwarz gives

$$\mathbb{E} |A_i - c| \leq \sqrt{\text{Var}(A_i) + (\mathbb{E} A_i - c)^2},$$

proving (2.13). To see local sharpness for $K = 2$, take deterministic projected slopes $a_1 < a_2$ at the origin. Let the two independent biases be uniform on nested intervals of lengths $1/B_1$ and $1/B_2$. Their density overlap integrates to $\min(B_1, B_2)$, so the exact local intensity is $\min(B_1, B_2)(a_2 - a_1)$, equal to $\inf_c \{B_1 |a_1 - c| + B_2 |a_2 - c|\}$. Shrinking a segment around the origin turns the local equality into the claimed supremal sharpness for positive-length probes. $\square$

5 Sharpness and necessity of the assumptions

This section first proves the covariance corollary and its sharp constants. It then gives two counterexamples that delimit the conditional anti-concentration assumption.

Proof of Theorem 2.3. Apply Lemma 4.1 with $n = K - 2$ in (3.7). For almost every s , put $\tilde{A}_s = u(s)^\top \tilde{W}$ and $\tilde{A}'_s = u(s)^\top \tilde{W}'$. Then

$$\begin{aligned} \mathbb{E} \text{kt}_{M_K}^\gamma &\leq \binom{K}{2} \frac{\bar{B}}{K-1} \int_0^L \mathbb{E} |\tilde{A}_s - \tilde{A}'_s| ds \\ &= \frac{K\bar{B}}{2} \int_0^L \mathbb{E} |u(s)^\top (\tilde{W} - \tilde{W}')| ds. \end{aligned} \quad (5.1)$$

By Lemma 4.2,

$$\mathbb{E} |u^\top (\tilde{W} - \tilde{W}')| \leq \frac{2}{\sqrt{3}} \sqrt{u^\top \tilde{\Sigma} u}. \quad (5.2)$$

Integration proves (2.7). In the constant-envelope case, $\bar{B} = B$, $\tilde{\Sigma} = \Sigma$, and $u^\top \Sigma u \leq \lambda_{\max}(\Sigma)$; division by L gives (2.8). $\square$

Under the iid-coordinate assumptions of open problem 2.2, $\Sigma = \sigma_w^2 I_d$. Thus (2.9) follows at once from Theorem 2.3. We now prove that neither of its constants can be improved. There are two distinct equality mechanisms: uniform biases make the visibility estimate exact for any prescribed projected weight law, while a uniform projected weight makes the Gini–variance step exact. The following construction realizes both simultaneously.

Proof of sharpness in Corollary 2.5. We first compute the exact local intensity and then realize it as a limit on positive-length segments. Fix $B, \sigma_w > 0$ and $d \geq 1$. Let $X_1, \dots, X_d, \beta$ be mutually independent, with

$$X_1, \dots, X_d \stackrel{\text{iid}}{\sim} \text{Unif}[-\sqrt{3}\sigma_w, \sqrt{3}\sigma_w], \quad \beta \sim \text{Unif}[-1/(2B), 1/(2B)] \quad (5.3)$$

and set $W = (X_1, \dots, X_d)$. Probe along the first coordinate axis and write $X = X_1$ for the projected slope; let X' be an independent copy. At the origin, the value β and X are independent, and

$$\mathbb{E}|X - X'| = \frac{2\sigma_w}{\sqrt{3}}, \quad \int_{\mathbb{R}} \rho_\beta(y)^2 F_\beta(y)^{K-2} dy = \frac{B}{K-1}. \quad (5.4)$$

Consequently, the exact local intensity furnished by Lemma 3.1 is

$$\binom{K}{2} \frac{2\sigma_w}{\sqrt{3}} \frac{B}{K-1} = \frac{K}{\sqrt{3}} B\sigma_w. \quad (5.5)$$

To turn this local calculation into the stated supremum over positive-length paths, put $a = \sqrt{3}\sigma_w$ and $h = 1/(2B)$. For $|t| \leq \delta$, the integrand in (3.7) vanishes outside $|y| \leq h + a\delta$, while

$$R_t(y) \leq 2aB^2 \mathbf{1}_{\{|y| \leq h+a\delta\}}.$$

For almost every y , the defining indicators for the two uniform laws converge as $t \rightarrow 0$, so $F_t(y) \rightarrow F_\beta(y)$ and $R_t(y) \rightarrow \rho_\beta(y)^2 \mathbb{E}|X - X'|$. The displayed compactly supported bound supplies domination, first showing that the exact local intensity converges to (5.5) as $t \rightarrow 0$. For $\gamma_\varepsilon(s) = (-\varepsilon/2 + s)e_1$, $0 \leq s \leq \varepsilon$, averaging that intensity over the segment shows that its expected switch density has the same limit as $\varepsilon \downarrow 0$. The unused iid coordinates do not affect this probe. This proves (2.10). At $K = 2$, the limiting ratio to $(K-1)B\sigma_w$ is $2/\sqrt{3}$, proving (2.11). $\square$

The preceding example uses densities throughout. The next elementary counterexample explains why arbitrary atomic biases cannot be admitted.

Proposition 5.1 (Atomic biases can destroy local density bounds). *Let $K = 2, d = 1$, let the slopes W_1, W_2 be iid from any atomless law, and, independently of the slopes, let the iid biases satisfy $\mathbb{P}[\beta = 0] = p > 0$. Then, for every $\varepsilon > 0$,*

$$\mathbb{E}\lambda_{M_2}^{[-\varepsilon, \varepsilon]} \geq \frac{p^2}{2\varepsilon}. \quad (5.6)$$

In particular, no uniform length-proportional bound can hold over all bias laws that allow arbitrary atoms.

Proof. On the event $\{\beta_1 = \beta_2 = 0\}$, which has probability p^2 , the two affine candidates meet at the interior point zero. Their slopes are distinct almost surely, so they exchange upper-envelope order there and create one active switch. The path has length 2ε , proving (5.6). This shows that absolute continuity cannot simply be deleted; it does not assert that every particular law containing an atom must fail. $\square$

Even bounded densities for both marginals do not suffice when weights and biases may be dependent. The conditional formulation in (2.4) is therefore substantive rather than a technical convenience.

Proposition 5.2 (Marginal density bounds are insufficient). *Fix $a > 0$ and $x_0 \neq 0$, and let*

$$W \sim \text{Unif}[-a, a], \quad \beta = -x_0 W. \quad (5.7)$$

Both W and β have bounded densities. Nevertheless, for two iid candidate pairs and every $\varepsilon > 0$,

$$\mathbb{E} \lambda_{M_2}^{[x_0 - \varepsilon, x_0 + \varepsilon]} = \frac{1}{2\varepsilon}. \quad (5.8)$$

Proof. The candidate affine functions are $W_j(x - x_0)$. Since $W_1 \neq W_2$ almost surely, the larger slope is active to the right of x_0 and the smaller slope is active to the left. Thus there is exactly one switch at x_0 . The path length is 2ε , proving (5.8). Here $\beta \mid W$ is a point mass, which is precisely what the conditional-density hypothesis rules out. $\square$

6 Gaussian local intensity and spherical geometry

This section derives the exact Gaussian switch law from spherical geometry. We first state the multiplicity-preserving Crofton interface for normalized feature curves, then apply the covariance lift and specialize to straight segments.

Lemma 6.1 (Parameterized spherical Crofton formula). *Let $m \geq 2$, let $D \sim N(0, I_m)$, and let $\psi : [0, L] \rightarrow \mathbb{S}^{m-1}$ be absolutely continuous. Define the set of transverse interior parameter preimages*

$$\mathcal{Z}_\psi(D) := \{s \in (0, L) : D^\top \psi(s) = 0, \dot{\psi}(s) \text{ exists, } D^\top \dot{\psi}(s) \neq 0\}. \quad (6.1)$$

Then

$$\mathbb{E}_D \# \mathcal{Z}_\psi(D) = \frac{1}{\pi} \int_0^L \left\| \dot{\psi}(s) \right\|_2 ds. \quad (6.2)$$

The count retains parameter multiplicity and is finite almost surely. Endpoints, the nondifferentiable and zero-speed images, and tangencies have zero Crofton measure.

Proof. This is the spherical Crofton formula with intersection multiplicity. In ambient dimension $m = 2$ it is Santaló [21, Eq. (37), p. 716] with spherical circle radius $\pi/2$; the higher-dimensional intersection formula is Santaló [22, Eq. (18.36), pp. 323–324], after normalizing the invariant hyperplane measure. Applied to the rectifiable curve measure induced by the parametrization, it counts each parameter preimage, so retraced arcs contribute repeatedly. Absolute continuity has the Lusin N property, so the image of the parameter-null set where the derivative does not exist has zero one-dimensional measure. The one-dimensional area formula gives the same conclusion for the zero-speed image. Crofton's formula then makes the average intersection count of these images zero; endpoint intersections also have probability zero. Its coarea integrand vanishes at tangencies, so they contribute zero as well. Finally, the right-hand side of (6.2) is finite, so the count is finite almost surely. $\square$

Lemma 6.2 (Gaussian upper-envelope Crofton law). *Let $m \geq 1$ be an integer, let $Z_1, \dots, Z_K$ be iid $N(0, I_m)$, and let $\psi : [0, L] \rightarrow \mathbb{S}^{m-1}$ be an absolutely continuous curve. If $N_K(\psi)$ counts active-index switches of $\max_{j \leq K} Z_j^\top \psi(s)$, then*

$$\mathbb{E} N_K(\psi) = c_K \int_0^L \left\| \dot{\psi}(s) \right\|_2 ds. \quad (6.3)$$

Proof. We first compute the intensity at regular points and then justify its integration along an arbitrary absolutely continuous curve by the parameterized Crofton formula. If $m = 1$, every absolutely continuous curve in $\mathbb{S}^0$ is constant, and both sides of (6.3) vanish. Assume henceforth that $m \geq 2$.

At almost every s , write $V = Z^\top \psi(s)$ and $A = Z^\top \dot{\psi}(s)$. Since $\psi(s)^\top \dot{\psi}(s) = 0$, these variables are independent Gaussians with variances 1 and $\|\dot{\psi}(s)\|_2^2$. Thus the expected absolute derivative difference for a fixed pair at common value y is $2 \|\dot{\psi}(s)\|_2 / \sqrt{\pi}$, while the other candidates lie below y with probability $\Phi(y)^{K-2}$. The visible-pair intensity is

$$\binom{K}{2} \frac{2 \|\dot{\psi}(s)\|_2}{\sqrt{\pi}} \int_{\mathbb{R}} \phi(y)^2 \Phi(y)^{K-2} dy. \quad (6.4)$$

Integration by parts, using $y\phi(y) = -\phi'(y)$, gives

$$\mathbb{E} \max_{j \leq K} G_j = K(K-1) \int_{\mathbb{R}} \phi(y)^2 \Phi(y)^{K-2} dy. \quad (6.5)$$

The boundary term vanishes at both ends, so (6.4) is $c_K \|\dot{\psi}(s)\|_2$.

For the global count, fix a pair and write

$$C = \frac{Z_i + Z_j}{\sqrt{2}}, \quad D = \frac{Z_i - Z_j}{\sqrt{2}}.$$

The vectors C, D are independent standard Gaussians, and, up to the null exceptions in Lemma 6.1, the pair ties precisely at $\mathcal{Z}_\psi(D)$. At any such intersection the common pair value is $C^\top \psi(s) / \sqrt{2} \sim N(0, 1/2)$, independently of D and of the remaining candidates. Its averaged visibility mark is

$$p_{K-2} = 2\sqrt{\pi} \int_{\mathbb{R}} \phi(y)^2 \Phi(y)^{K-2} dy. \quad (6.6)$$

Since $\mathcal{Z}_\psi(D)$ is finite almost surely, marked Tonelli gives

$$\mathbb{E}_{C, (Z_k)_{k \neq i, j}} \left[\sum_{s \in \mathcal{Z}_\psi(D)} \mathbf{1}_{\{Z_k^\top \psi(s) < C^\top \psi(s) / \sqrt{2} \ \forall k \neq i, j\}} \middle| D \right] = p_{K-2} \# \mathcal{Z}_\psi(D). \quad (6.7)$$

Only the marginal mark at each intersection is used; marks at different intersections need not be independent. The same conditioning, followed by a finite union, excludes triple ties. Hence every marked transverse pair intersection is one active switch, and every active switch has a unique pair. By Lemma 6.1, a fixed pair contributes $p_{K-2} \text{Len}(\psi) / \pi$ in expectation. Summing over pairs and using (6.5) gives

$$\binom{K}{2} \frac{p_{K-2}}{\pi} \text{Len}(\psi) = c_K \text{Len}(\psi),$$

which proves (6.3). $\square$

Proof of Theorem 2.8. Write $\Theta = \mu + \Omega^{1/2} Z$ with $Z \sim N(0, I_{d+1})$. The common deterministic affine function $\mu^\top q_x$ does not affect the active index. Along the path, put

$$r(s) := \left\| \Omega^{1/2} q_{\gamma(s)} \right\|_2, \quad \psi(s) := \Psi_\Omega(\gamma(s)).$$

Condition (2.17) and compactness give $\inf_s r(s) > 0$, so ψ is absolutely continuous. The centered candidate values are $r(s)Z_j^\top \psi(s)$. Multiplication by the common positive factor $r(s)$ leaves their ordering unchanged. Applying Lemma 6.2 proves (2.19). $\square$

Proof of Corollary 2.9. The block-diagonal covariance gives the lift (2.20). Put $s_x^2 = \sigma_\beta^2 + x^\top \Sigma x$. Direct differentiation gives, for every $u \in \mathbb{R}^d$,

$$\|D\Psi_\Sigma(x)[u]\|_2^2 = \frac{u^\top \Sigma u}{s_x^2} - \frac{(u^\top \Sigma x)^2}{s_x^4}. \quad (6.8)$$

Equivalently, if

$$v_{x,u}^2 := \text{Var}(W^\top u \mid W^\top x + \beta) = u^\top \Sigma u - \frac{(u^\top \Sigma x)^2}{s_x^2}, \quad (6.9)$$

then the exact local switch intensity is

$$I_K(x, u) = c_K \frac{v_{x,u}}{s_x}. \quad (6.10)$$

Here $I_K(x, u)$ denotes the switch rate per unit parameter at position x and tangent u ; this equation defines it through the continuous local-speed density. The identity remains valid for singular Σ . In particular,

$$\|D\Psi_\Sigma(x)[u]\|_2 \leq \frac{\sqrt{u^\top \Sigma u}}{s_x} \leq \frac{\sqrt{\lambda_{\max}(\Sigma)}}{\sigma_\beta} \quad (\|u\|_2 = 1). \quad (6.11)$$

Integrating and applying Theorem 2.8 proves (2.21). At $x = 0$, equality holds along a top eigenvector. Averaging the continuous local speed over shrinking segments in that direction proves (2.22). The rank statements (2.23) and (2.24) are proved in Section 7. $\square$

For $K = 2$, $c_2 = 1/\pi$, and Lemma 6.2 is exactly the spherical-length form of the Edelman–Kostlan zero-count formula [6]. The factor c_K is what upper-envelope visibility contributes for general rank.

The exact identity also gives a closed form for straight segments.

Corollary 6.3 (Jointly Gaussian straight-segment formula). *Under the assumptions of Theorem 2.8, suppose (2.17) holds on the segment $[a, b]$. Then*

$$\mathbb{E} \text{kt}_{M_K}^{[a,b]} = c_K \arccos \left(\frac{q_a^\top \Omega q_b}{\sqrt{q_a^\top \Omega q_a} \sqrt{q_b^\top \Omega q_b}} \right). \quad (6.12)$$

Here $\arccos$ is the principal branch in $[0, \pi]$. For independent weights and biases, the numerator is $\sigma_\beta^2 + a^\top \Sigma b$. For a CPWL path, sum (6.12) over its consecutive nonconstant segments.

Proof. The unnormalized lift of $(1-t)a + tb$ is the affine segment joining $\Omega^{1/2}q_a$ and $\Omega^{1/2}q_b$. It avoids the origin by assumption. Its normalization traces the minor great-circle arc (possibly degenerate) of length less than π between the normalized endpoints. Its spherical length is their arccosine distance; a degenerate arc has zero length and zero switch count. Thus Theorem 2.8 applies. $\square$

7 Gaussian rank constants and random polytopes

This section studies the Gaussian rank coefficient c_K . We establish finite-rank bounds and asymptotics, relate c_K to planar Gaussian polytopes, and then deduce the whole-line switch count. Write $m_K := \mathbb{E}M_K^G = \sqrt{\pi}c_K$.

Proposition 7.1 (Finite-rank bounds and asymptotics). *For $K \geq 2$, let $b_K = \Phi^{-1}(1 - 1/K)$. Then*

$$b_K - \frac{\Phi(b_K)^{K+1}}{K\phi(b_K)} \leq m_K \leq K\phi(b_K). \quad (7.1)$$

The sequence $(m_K)_{K \geq 1}$ is strictly increasing and discretely concave. As $K \rightarrow \infty$,

$$c_K = \sqrt{\frac{2 \log K}{\pi}} - \frac{\log \log K + \log(4\pi) - 2\gamma_E}{2\sqrt{2\pi \log K}} + o((\log K)^{-1/2}), \quad (7.2)$$

where γ_E is Euler's constant.

Proof. We prove the finite-rank bounds, discrete shape properties, and asymptotic expansion in turn. For any real b , the tail-integral identity gives

$$m_K = b + \int_b^\infty (1 - \Phi(t)^K) dt - \int_{-\infty}^b \Phi(t)^K dt. \quad (7.3)$$

At $b = b_K$, the union bound and $\int_b^\infty (1 - \Phi(t)) dt = \phi(b) - b(1 - \Phi(b))$ yield $m_K \leq K\phi(b_K)$. The function $\log \Phi$ is concave, so for $t \leq b_K$,

$$\Phi(t)^K \leq \Phi(b_K)^K \exp\left\{\frac{K\phi(b_K)}{\Phi(b_K)}(t - b_K)\right\}.$$

Integrating this bound and dropping the positive integral in (7.3) proves the lower bound.

For $x \in \mathbb{R}$, write $x_+ = \max\{x, 0\}$. Couple the nested maxima and let G be one auxiliary standard normal independent of all of them. Then

$$m_{K+1} - m_K = \mathbb{E}(G - M_K^G)_+.$$

This quantity is positive. Since $M_{K+1}^G \geq M_K^G$ almost surely, and the inequality is strict jointly with $G > M_{K+1}^G$ on an event of positive probability, the increments strictly decrease with K . This proves monotonicity and discrete concavity. Finally, Mills' ratio gives

$$b_K = \sqrt{2 \log K} - \frac{\log \log K + \log(4\pi)}{2\sqrt{2 \log K}} + o((\log K)^{-1/2}). \quad (7.4)$$

Moreover, Gaussian extreme-value theory [16, Theorem 1.5.3] gives convergence of $X_K := b_K(M_K^G - b_K)$ to the standard Gumbel law. We record the first-moment justification. For $K \geq 3$, Gaussian hazard bounds give

$$\mathbb{P}(X_K > x) \leq e^{-x} \quad (x \geq 0), \quad \mathbb{P}(X_K < -x) \leq \exp(-e^{x/2}) \quad (0 \leq x \leq b_K^2).$$

For $x = b_K^2 + b_K r$, $r \geq 0$, Mills' bound yields

$$\mathbb{P}(X_K < -x) \leq 2^{-K} e^{-Kr^2/2}, \quad \int_{b_K^2}^\infty \mathbb{P}(X_K < -x) dx \leq b_K 2^{-K} \sqrt{\frac{\pi}{2K}}.$$

The two tail estimates, with the finitely many smaller K handled separately, prove uniform integrability of (X_K) . Hence $m_K = b_K + \gamma_E/b_K + o(1/b_K)$. Combining with (7.4) and dividing by $\sqrt{\pi}$ proves (7.2), and in particular (2.23). $\square$

The same coefficient has two exact planar interpretations. Let $P_K = \text{conv}\{Z_1, \dots, Z_K\}$ for iid $Z_j \sim N(0, I_2)$, and let $f_0(P_K)$ and $\text{Per}(P_K)$ denote its number of vertices and perimeter (with the usual doubled-segment convention when $K = 2$).

Proposition 7.2 (Planar Gaussian-polytope identities). *For every $K \geq 2$,*

$$\mathbb{E}f_0(P_K) = 2\pi c_K = 2\sqrt{\pi} m_K, \quad \mathbb{E}\text{Per}(P_K) = 2\pi m_K. \quad (7.5)$$

Consequently,

$$\mathbb{E}f_0(P_K) = \frac{1}{\sqrt{\pi}} \mathbb{E}\text{Per}(P_K). \quad (7.6)$$

Proof. Apply Lemma 6.2 to $\psi(\theta) = (\cos \theta, \sin \theta)$, $0 \leq \theta \leq 2\pi$, with the endpoints identified. The seam has a tie with probability zero. The active candidate is the exposed vertex of P_K ; as θ crosses a boundary between adjacent normal cones, this vertex changes. For $K = 2$, the two support candidates exchange twice and both sample points are vertices. For $K \geq 3$, the hull is almost surely a nondegenerate polygon and the number of transitions around the circle equals its number of vertices. Thus $\mathbb{E}f_0(P_K) = 2\pi c_K$. The Cauchy support-function formula and rotational invariance give

$$\mathbb{E}\text{Per}(P_K) = \int_0^{2\pi} \mathbb{E} \max_{j \leq K} Z_j^\top \psi(\theta) d\theta = 2\pi m_K.$$

Eliminating m_K between the preceding identities proves (7.6). $\square$

These formulas place the rank coefficient within classical Gaussian-polytope theory [1, 5, 13, 14]. Since $f_0(P_K) \leq K$,

$$c_K \leq \frac{K}{2\pi} \leq \frac{K-1}{\pi}. \quad (7.7)$$

This proves (2.24). For $K = 2$,

$$m_2 = \frac{1}{2} \mathbb{E}|G_1 - G_2| = \frac{1}{\sqrt{\pi}}, \quad c_2 = \frac{1}{\pi}.$$

Together with shrinking-path equality in (6.11), this shows that $1/\pi$ is the optimal constant uniform in K in the rank-linear Gaussian normalization.

One further consequence makes the global one-dimensional scale explicit. If (A_j, B_j) , $1 \leq j \leq K$, are iid Gaussian pairs with positive-definite covariance and $K \geq 2$, the normalized feature curve of $t \mapsto A_j t + B_j$ over $t \in \mathbb{R}$ traces an open semicircle of length π . Define the whole-line count by the monotone limit $N_{\mathbb{R}} := \lim_{R \rightarrow \infty} \text{kt}_{M_K}^{[-R, R]}$. Monotone convergence and Lemma 6.2 therefore give

$$\mathbb{E}N_{\mathbb{R}} = \pi c_K = \sqrt{\pi} m_K. \quad (7.8)$$

Every finite interval has strictly smaller expected count, and the supremum over finite intervals equals the right-hand side.

8 Propagation through deep Maxout networks

We now propagate the single-unit estimates through a feedforward Maxout network. We first give a distribution-free switch-and-length recursion and then sharpen both factors for Gaussian layers.

Fix a depth $D \geq 1$, widths $q_0 = d, q_1, \dots, q_D \geq 1$, and unit ranks $K_{\ell r} \geq 2$. Set $z_0(x) = x$, and for $1 \leq \ell \leq D$ define

$$F_{\ell r}(z) := \max_{1 \leq j \leq K_{\ell r}} (W_{\ell r j}^\top z + \beta_{\ell r j}), \quad z \in \mathbb{R}^{q_{\ell-1}}, \quad 1 \leq r \leq q_\ell, \quad (8.1)$$

$$F_\ell := (F_{\ell 1}, \dots, F_{\ell q_\ell}), \quad z_\ell := F_\ell \circ z_{\ell-1}. \quad (8.2)$$

Thus $z_\ell = (z_{\ell 1}, \dots, z_{\ell q_\ell})$, where $z_{\ell r} = F_{\ell r} \circ z_{\ell-1}$; $W_{\ell r j} \in \mathbb{R}^{q_{\ell-1}}$, and $W_{\ell r}$ denotes a generic candidate weight in unit (ℓ, r) . Put $\Sigma_{\ell r} := \text{Cov}(W_{\ell r})$. Parameter collections in different layers are independent. Within each unit the candidates are iid, with conditional bias density bounded by $B_{\ell r} < \infty$ and square-integrable weight $W_{\ell r}$. Units in the same layer may have arbitrary dependence subject to these marginal assumptions.

For each layer define its switch factor and length factor by

$$\mathcal{D}_\ell := \frac{1}{2} \sup_{\|v\|_2=1} \sum_{r=1}^{q_\ell} K_{\ell r} B_{\ell r} \mathbb{E} \left| v^\top (W_{\ell r} - W'_{\ell r}) \right|, \quad (8.3)$$

$$\mathcal{A}_\ell := \sup_{\|v\|_2=1} \left(\sum_{r=1}^{q_\ell} \mathbb{E} \max_{j \leq K_{\ell r}} \left| v^\top W_{\ell r j} \right|^2 \right)^{1/2}, \quad (8.4)$$

where $W'_{\ell r}$ is an independent copy of one candidate weight. For each layer, write $\Gamma_{\ell-1} := z_{\ell-1} \circ \gamma$. Let $\mathcal{S}_{\ell r}(\gamma)$ be the set of parameter values $t \in (0, L)$ that satisfy the active-switch sign-change criterion of Definition 2.1 for the candidate scores

$$W_{\ell r j}^\top \Gamma_{\ell-1}(t) + \beta_{\ell r j}, \quad 1 \leq j \leq K_{\ell r},$$

with time measured in the original parametrization of γ . Set

$$\mathcal{S}_D(\gamma) := \bigcup_{\ell=1}^D \bigcup_{r=1}^{q_\ell} \mathcal{S}_{\ell r}(\gamma), \quad S_D(\gamma) := |\mathcal{S}_D(\gamma)|. \quad (8.5)$$

Thus $S_D(\gamma)$ counts distinct switch times; simultaneous events count once, so it is not the sum of the individual unit counts.

The factor $\mathcal{D}_\ell$ controls switches created per unit incoming length, whereas $\mathcal{A}_\ell$ controls expected length growth. The next theorem iterates these two one-layer estimates.

Theorem 8.1 (Layerwise switch propagation). *For every deterministic finite-length rectifiable input path,*

$$\mathbb{E} \text{Len}(z_\ell \circ \gamma) \leq \text{Len}(\gamma) \prod_{h=1}^{\ell} \mathcal{A}_h, \quad 1 \leq \ell \leq D, \quad (8.6)$$

$$\mathbb{E} S_D(\gamma) \leq \text{Len}(\gamma) \sum_{\ell=1}^D \mathcal{D}_\ell \prod_{h < \ell} \mathcal{A}_h. \quad (8.7)$$

For every $1 \leq \ell \leq D$, the factors admit the covariance-based estimates

$$\mathcal{D}_\ell \leq \frac{1}{\sqrt{3}} \sup_{\|v\|_2=1} \sum_{r=1}^{q_\ell} K_{\ell r} B_{\ell r} \sqrt{v^\top \Sigma_{\ell r} v}, \quad (8.8)$$

$$\mathcal{A}_\ell \leq \sqrt{\lambda_{\max} \left(\sum_{r=1}^{q_\ell} K_{\ell r} \mathbb{E}[W_{\ell r} W_{\ell r}^\top] \right)}. \quad (8.9)$$

Proof. We condition layer by layer, prove one-step switch and length bounds for the incoming curve, and then iterate while taking the union of switch times. Write $\Gamma_\ell = z_\ell \circ \gamma$. Conditional on the first $\ell - 1$ layers, $\Gamma_{\ell-1}$ is a fixed absolutely continuous rectifiable curve independent of layer ℓ . If its length is zero, it is constant and both one-step bounds below are zero. Otherwise, remove constant portions and use its arclength representation, preserving the order and multiplicity of retraced portions. Constant portions cannot create an active switch. Applying the first bound of Theorem 2.3 to this representation for every unit and taking a union bound gives

$$\mathbb{E}[\text{switches created in layer } \ell \mid \Gamma_{\ell-1}] \leq \mathcal{D}_\ell \text{Len}(\Gamma_{\ell-1}). \quad (8.10)$$

At almost every parameter where $\Gamma'_{\ell-1}$ exists and is nonzero and every unit has a unique maximizing candidate, put $v = \Gamma'_{\ell-1} / \|\Gamma'_{\ell-1}\|_2$. The derivative of one unit is the projection of its active weight, so the chain rule gives

$$\|\Gamma'_\ell\|_2 \leq \|\Gamma'_{\ell-1}\|_2 \left(\sum_{r=1}^{q_\ell} \max_{j \leq K_{\ell r}} |v^\top W_{\ell r j}|^2 \right)^{1/2}. \quad (8.11)$$

When $\Gamma'_{\ell-1} = 0$, the output derivative also vanishes. Conditional Jensen, which does not require independence among same-layer units, and (8.4) therefore give for almost every parameter

$$\mathbb{E}[\|\Gamma'_\ell\|_2 \mid \Gamma_{\ell-1}] \leq \mathcal{A}_\ell \|\Gamma'_{\ell-1}\|_2.$$

Integration gives

$$\mathbb{E}[\text{Len}(\Gamma_\ell) \mid \Gamma_{\ell-1}] \leq \mathcal{A}_\ell \text{Len}(\Gamma_{\ell-1}).$$

Iteration proves (8.6). Summing (8.10) over layers, observing that simultaneous switches can only reduce the number of distinct parameter values, proves (8.7). If a vertex created in a lower layer propagates through later layers, it remains at the same parameter value already included in $\mathcal{S}_D(\gamma)$; it does not require a new higher-layer tie and creates no additional distinct-time term.

For clarity, if $\text{Break}(\eta)$ denotes the parameter values where a CPWL curve η is not locally affine, then for a CPWL input probe γ , almost surely,

$$\text{Break}(z_D \circ \gamma) \subseteq \text{Break}(\gamma) \cup \mathcal{S}_D(\gamma). \quad (8.12)$$

Hence an affine probe has at most $\mathcal{S}_D(\gamma)$ final-output breakpoints; equality need not hold. Finally, Lemma 4.2 proves (8.8), while $\max_j |v^\top W_j|^2 \leq \sum_j |v^\top W_j|^2$ proves (8.9). $\square$

For Gaussian layers, both factors can be sharpened. Assume now that in each unit (ℓ, r) the candidate pairs are iid as above and, for every j ,

$$W_{\ell r j} \sim N(0, \Sigma_{\ell r}), \quad \beta_{\ell r j} \sim N(0, \sigma_{\beta, \ell r}^2) \quad (8.13)$$

with $W_{\ell r j}$ independent of $\beta_{\ell r j}$ and $\sigma_{\beta, \ell r} > 0$. Let

$$m_{2,K} := \mathbb{E}[(M_K^G)^2], \quad q_K := \max\{1, m_{2,K}\}. \quad (8.14)$$

The Gaussian length factor can be improved because the active weight is not a generic candidate weight. The next proposition identifies its exact second moment.

Proposition 8.2 (Exact active-weight second moment for a Gaussian unit). *Let $K \geq 2$ and $q \geq 1$, and let the candidate pairs be iid with $W_j \in \mathbb{R}^q$, $W_j \sim N(0, \Sigma)$, $\beta_j \sim N(0, \sigma_\beta^2)$, and W_j independent of β_j , where $\Sigma \succeq 0$ and $\sigma_\beta > 0$. For every deterministic $z \in \mathbb{R}^q$, let $J(z)$ denote the almost surely unique active candidate of the unit at z . Then*

$$\mathbb{E}[W_{J(z)} W_{J(z)}^\top] = Q_K(z) := \Sigma + (m_{2,K} - 1) \frac{\Sigma z z^\top \Sigma}{\sigma_\beta^2 + z^\top \Sigma z}. \quad (8.15)$$

Moreover, $Q_K(z) \preceq q_K \Sigma$ and

$$\pi c_K^2 \leq m_{2,K} \leq q_K \leq 2 \log(2K) + 2. \quad (8.16)$$

Consequently, $q_K = \Theta(\log K)$.

Proof. Fix a direction v . For candidate j , set $V_j = W_j^\top z + \beta_j$, $A_j = v^\top W_j$, $s_z^2 = \sigma_\beta^2 + z^\top \Sigma z$, and $\tau = v^\top \Sigma z$. Gaussian regression gives

$$A_j = \frac{\tau}{s_z^2} V_j + \varepsilon_j, \quad \mathbb{E} \varepsilon_j^2 = v^\top \Sigma v - \frac{\tau^2}{s_z^2}.$$

The residual vector $(\varepsilon_1, \dots, \varepsilon_K)$ is independent of the full score vector $(V_1, \dots, V_K)$. The almost surely unique winner $J(z)$ is score-measurable, so selecting $J(z)$ does not change the residual variance or create a cross term. Since $V_{J(z)}$ is distributed as $s_z M_K^G$,

$$\mathbb{E}(v^\top W_{J(z)})^2 = v^\top \Sigma v + (m_{2,K} - 1) \frac{(v^\top \Sigma z)^2}{s_z^2},$$

which is (8.15). Put

$$P(z) := \frac{\Sigma z z^\top \Sigma}{\sigma_\beta^2 + z^\top \Sigma z}.$$

Cauchy–Schwarz in the Σ -seminorm gives $0 \preceq P(z) \preceq \Sigma$. If $m_{2,K} \geq 1$, then $Q_K(z) \preceq m_{2,K} \Sigma$; if $m_{2,K} < 1$, then $Q_K(z) \preceq \Sigma$. Thus $Q_K(z) \preceq q_K \Sigma$. Jensen gives $m_{2,K} \geq m_K^2 = \pi c_K^2$. Finally, $(M_K^G)^2 \leq \max_{j \leq K} |G_j|^2$, and

$$\mathbb{P} \left[\max_j |G_j|^2 > t \right] \leq \min\{1, 2K e^{-t/2}\}.$$

Integrating at the threshold $2 \log(2K)$ proves (8.16). $\square$

Define

$$\mathcal{D}_\ell^G := \sup_{\|v\|_2=1} \sum_{r=1}^{q_\ell} c_{K_{\ell r}} \frac{\sqrt{v^\top \Sigma_{\ell r} v}}{\sigma_{\beta, \ell r}}, \quad (8.17)$$

$$\mathcal{A}_\ell^G := \sqrt{\lambda_{\max} \left(\sum_{r=1}^{q_\ell} q_{K_{\ell r}} \Sigma_{\ell r} \right)}. \quad (8.18)$$

For later use, let $Q_{\ell r}(z)$ denote $Q_K(z)$ in (8.15) with $(K, \Sigma, \sigma_\beta) = (K_{\ell r}, \Sigma_{\ell r}, \sigma_{\beta, \ell r})$. The preceding quantities replace the generic switch and length factors by their Gaussian counterparts.

Corollary 8.3 (Deep Gaussian rank law). *Under the preceding Gaussian assumptions,*

$$\mathbb{E}S_D(\gamma) \leq \text{Len}(\gamma) \sum_{\ell=1}^D \mathcal{D}_\ell^G \prod_{h<\ell} \mathcal{A}_h^G. \quad (8.19)$$

Moreover, fix $1 \leq \ell \leq D$ and $0 < T < \infty$, and let $\eta : [0, T] \rightarrow \mathbb{R}^{q_{\ell-1}}$ be absolutely continuous and independent of the parameters in layer ℓ . Conditional on this input curve,

$$\mathbb{E}[\text{Len}(F_\ell \circ \eta) \mid \eta] \leq \int_0^T \sqrt{\dot{\eta}(s)^\top \left(\sum_{r=1}^{q_\ell} Q_{\ell r}(\eta(s)) \right) \dot{\eta}(s)} ds, \quad (8.20)$$

where $\dot{\eta}(s)$ exists for almost every s , and the expectation is over the current layer.

Proof. For each unit, (6.10) and (6.11) bound its local switch intensity in direction v by

$$c_{K_{\ell r}} \frac{\sqrt{v^\top \Sigma_{\ell r} v}}{\sigma_{\beta, \ell r}},$$

which gives $\mathcal{D}_\ell^G$. For length, Jensen's inequality and Proposition 8.2 give the pointwise conditional bound

$$\mathbb{E}[\|(F_\ell \circ \eta)'(s)\|_2 \mid \eta] \leq \sqrt{\dot{\eta}(s)^\top \left(\sum_{r=1}^{q_\ell} Q_{\ell r}(\eta(s)) \right) \dot{\eta}(s)},$$

and integration gives (8.20). The bound $Q_{\ell r}(z) \preceq q_{K_{\ell r}} \Sigma_{\ell r}$ then gives $\mathcal{A}_\ell^G$. The proof of Theorem 8.1 completes the recursion. $\square$

The relations $c_K = \Theta(\sqrt{\log K})$ and $q_K = \Theta(\log K)$ determine the unit-level rank dependence. More precisely, let $R \rightarrow \infty$ along a family with fixed depth and uniformly bounded widths. Suppose constants $0 < a \leq b < \infty$, $0 < \underline{\sigma} \leq \bar{\sigma} < \infty$, and $C < \infty$, independent of R , satisfy

$$aR \leq K_{\ell r} \leq bR, \quad \underline{\sigma} \leq \sigma_{\beta, \ell r} \leq \bar{\sigma}, \quad \|\Sigma_{\ell r}\|_{\text{op}} \leq C$$

for every unit. Then $\mathcal{D}_\ell^G, \mathcal{A}_\ell^G = O(\sqrt{\log R})$ uniformly in ℓ . If, in addition, each layer contains a unit with $\lambda_{\max}(\Sigma_{\ell r})$ bounded below by a positive constant independent of R , both factors are $\Theta(\sqrt{\log R})$. Without this lower nondegeneracy assumption, only the stated upper order follows. This is a network-level consequence of the sharp single-unit law, not a new curve-length recursion.

9 Relation to prior work, limitations, and outlook

Our results concern expected pathwise active-candidate switches under random parameters, rather than global activation-region counts. This section separates those notions, locates the proof tools in prior work, and records the limits of the conclusions.

The all-pairs estimate of Goujon et al. [10] pays for hidden intersections; their discussion after equation (31) conjectures a tighter linear-in-rank bound. The exact $(K - 1)$ normalization, the sharp distribution-free constants, and the Gaussian law proved here strengthen that qualitative conjecture. The area formula accounts for crossings through the variation of projected candidate differences, including parameter multiplicity, while the two counterexamples in Section 5 show why some conditional bias anti-concentration is needed.

Global region counts for piecewise-linear and Maxout networks study a different complexity measure [17, 18, 23]. In particular, Tseran and Montúfar [24, Proposition 29] obtain a one-dimensional $O(\sqrt{\log K})$ bound for expected nonempty regions, not the local switch density studied here. Existing composition and curve-length arguments already yield deep recursions [10, 25]; our contribution is the sharp single-unit coefficient inserted into them.

Visibility marks are classical in random convex-hull geometry [8, 20], and the Gaussian proof belongs to the Kac–Rice and integral-geometric tradition [2, 6]. Our additions are the sharp bounded-density visibility inequality and the resulting Maxout constants; the planar-polytope identity in Proposition 7.2 is presented as an interpretation, not as a new face-number formula.

The results do not imply a global region count, a statement for trained weights, or concentration under the present weak assumptions. The deep bound also requires independence between layers. Natural next questions are sharp visibility inequalities for dependent candidates, concentration under stronger distributional assumptions, and analogous laws for broader upper-envelope spline operators.

References

- [1] F. Affentranger and R. Schneider. Random projections of regular simplices. *Discrete & Computational Geometry*, 7(3):219–226, 1992. doi:10.1007/BF02187839.
- [2] J.-M. Azaïs and M. Wschebor. *Level Sets and Extrema of Random Processes and Fields*. Wiley, 2009. doi:10.1002/9780470434642.
- [3] R. Balestrierio and R. G. Baraniuk. A spline theory of deep learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 374–383, 2018. arXiv:1805.06576.
- [4] R. Balestrierio and R. G. Baraniuk. Mad Max: Affine spline insights into deep learning. *Proceedings of the IEEE*, 109(5):704–727, 2021. doi:10.1109/JPROC.2020.3042100.
- [5] Y. M. Baryshnikov and R. A. Vitale. Regular simplices and Gaussian samples. *Discrete & Computational Geometry*, 11(2):141–147, 1994. doi:10.1007/BF02574000.
- [6] A. Edelman and E. Kostlan. How many zeros of a random polynomial are real? *Bulletin of the American Mathematical Society*, 32(1):1–37, 1995. doi:10.1090/S0273-0979-1995-00571-9.
- [7] H. Federer. *Geometric Measure Theory*. Springer, 1969. doi:10.1007/978-3-642-62010-2.
- [8] J. Gates. The mean number of extreme lines in a convex hull of lines. *Discrete & Computational Geometry*, 27(4):485–499, 2002. doi:10.1007/s00454-001-0093-0.
- [9] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout networks. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28(3) of *Proceedings of Machine Learning Research*, pages 1319–1327, 2013. arXiv:1302.4389.
- [10] A. Goujon, A. Etemadi, and M. Unser. On the number of regions of piecewise linear neural networks. *Journal of Computational and Applied Mathematics*, 441:115667, 2024. doi:10.1016/j.cam.2023.115667. Extended version: arXiv:2206.08615.
- [11] B. Hanin and D. Rolnick. Complexity of linear regions in deep networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2596–2604, 2019a. arXiv:1901.09021.

- [12] B. Hanin and D. Rolnick. Deep ReLU networks have surprisingly few activation patterns. In *Advances in Neural Information Processing Systems 32*, 2019b. arXiv:1906.00904.
- [13] D. Hug, G. O. Munsonius, and M. Reitzner. Asymptotic mean values of Gaussian polytopes. *Beiträge zur Algebra und Geometrie*, 45(2):531–548, 2004.
- [14] Z. Kabluchko and D. Zaporozhets. Absorption probabilities for Gaussian polytopes and regular spherical simplices. *Advances in Applied Probability*, 52(2):588–616, 2020. doi:10.1017/apr.2020.7.
- [15] R. La Haye and P. Zizler. The Gini mean difference and variance. *METRON*, 77(1):43–52, 2019. doi:10.1007/s40300-019-00149-2.
- [16] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. Springer, 1983. doi:10.1007/978-1-4612-5449-2.
- [17] G. Montúfar, R. Pascanu, K. Cho, and Y. Bengio. On the number of linear regions of deep neural networks. In *Advances in Neural Information Processing Systems 27*, pages 2924–2932, 2014. arXiv:1402.1869.
- [18] G. Montúfar, Y. Ren, and L. Zhang. Sharp bounds for the number of regions of Maxout networks and vertices of Minkowski sums. *SIAM Journal on Applied Algebra and Geometry*, 6(4):618–649, 2022. doi:10.1137/21M1413699.
- [19] M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, and J. Sohl-Dickstein. On the expressive power of deep neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2847–2854, 2017. arXiv:1606.05336.
- [20] A. Rényi and R. Sulanke. Über die konvexe Hülle von n zufällig gewählten Punkten. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2(1):75–84, 1963. doi:10.1007/BF00535300.
- [21] L. A. Santaló. Integral formulas in Crofton’s style on the sphere and some inequalities referring to spherical curves. *Duke Mathematical Journal*, 9(4):707–722, 1942. doi:10.1215/S0012-7094-42-00949-9.
- [22] L. A. Santaló. *Integral Geometry and Geometric Probability*. Cambridge Mathematical Library. Cambridge University Press, second edition, 2004. doi:10.1017/CBO9780511617331.
- [23] T. Serra, C. Tjandraatmadja, and S. Ramalingam. Bounding and counting linear regions of deep neural networks. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 4558–4566, 2018. arXiv:1711.02114.
- [24] H. Tseran and G. Montúfar. On the expected complexity of Maxout networks. In *Advances in Neural Information Processing Systems 34*, pages 28995–29008, 2021. arXiv:2107.00379.
- [25] H. Tseran and G. Montúfar. Expected gradients of Maxout networks and consequences to parameter initialization. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 34491–34532, 2023. arXiv:2301.06956.