WHAT DOES TWONN MEASURE? SAMPLE-RESOLUTION LAWS IN HIGH DIMENSIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

TwoNN estimates intrinsic dimension from ratios of first- and second-nearest-neighbor distances. In high dimensions, even nearest neighbors may lie outside the small neighborhoods underlying manifold-dimension interpretations. We study the pooled inverse-mean-log-ratio estimator as sample size and dimension grow together. For anisotropic Gaussian data under stated spectral assumptions, we prove concentration around an explicit function of sample size and covariance eigenvalues, accounting for dependence between ratios computed from the same sample. For uniform points on a noiseless sphere, polynomially many samples make TwoNN report a vanishing fraction of the true dimension. We also bound spherical mean-spacing approximation error and shallow Gaussian target-estimation error. Gaussian and spherical experiments test the phase laws. In a separate controlled Gaussian study, the trace-moment estimator closely recovers its oracle shallow target, while observed TwoNN retains finite-resolution and sampling error. In five CIFAR-10/ResNet-18 models, covariance-aligned rescaling fitted on disjoint images increases TwoNN by a median factor of 2.20 while preserving every prediction; random orientations with identical singular values give a factor of 0.995. Direct simulations from fitted global-Gaussian and class-Gaussian-mixture nulls compare neural TwoNN with those nulls: the mixture moves closer to real TwoNN at one tested layer and farther away at the other. TwoNN can therefore consistently track a sample-resolution-dependent target without recovering population manifold dimension. In the tested neural representations, its value should be interpreted together with coordinate controls and explicit distributional nulls.

1 INTRODUCTION

TwoNN uses the ratio of each point's second- to first-neighbor distance to estimate intrinsic dimension. We analyze the pooled inverse-mean-log-ratio estimator with numerator $n-1$, stated by Denti et al. (2022, Eq. (3)), which we call pooled TwoNN throughout. This MacKay-style aggregate differs from the empirical-CDF regression in the original implementation of Facco et al. (2017); both use Facco's Pareto ratio law. Our theorems concern the pooled estimator. Related distance-ratio estimators include Levina–Bickel (Levina & Bickel, 2004) and GRIDE (Denti et al., 2022). In neural networks, these estimates often rise in early layers and fall later, a pattern interpreted as expansion and compression of a latent manifold (Ansuini et al., 2019). Yet estimates can track covariance structure without recovering population dimension (Schulte & Rügamer, 2026; Hadfield, 2026).

The usual interpretation assumes that, sufficiently close to a point, the probability mass inside radius r is approximately proportional to r^d for dimension d. A nearest neighbor among n points lies at a radius containing probability mass of order $1/n$. That radius need not be small enough for the local approximation. TwoNN already uses sample size to study this change of scale (Facco et al., 2017), and Erba et al. (2019) study undersampling and spherical calibration. We ask: *what does pooled TwoNN converge to when the sample has not reached the local regime?*

Earlier theory describes this geometry around one fixed point, called the anchor. If $G_x(r)$ is the probability mass within distance r of anchor x, then $rG'_x(r)/G_x(r)$ equals d for the power law $G_x(r) \propto r^d$, but can vary at other radii. This is the established finite-radius local continuous intrinsic dimensionality (Houle, 2013; Amsaleg et al., 2018).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Random-codebook theory determines the nearest distance when one point is compared with many independent candidates (Lapidoth, 1997; Dembo & Kontoyiannis, 2002; Kontoyiannis & Zamir, 2006; Zhou et al., 2019). Pooled TwoNN aggregates ratios around every observed point. Two anchors can share candidate neighbors, so their ratios are dependent. A calculation for one anchor does not establish concentration of the pooled estimate.

Our main theorem treats this dependence for iid $\mathcal{N}(0, \Sigma_D)$ data. Assume bounded covariance eigenvalues and positive limiting first two spectral moments, with $\log n \to \infty$ and $\log n = O(D)$ as $D \to \infty$. For covariance eigenvalues λ_j, define

$$d_{\text{eff}}(n, \Sigma_D) = \sum_{j=1}^{D} \left[1 - (1 + 2s\lambda_j)^{-2}\right], \tag{1}$$

where $s > 0$ is the unique scale at which the Gaussian lower-tail exponent matches $\log(n-1)/D$; Section 3 gives the scalar equation. We prove $\widehat{d}_{2\text{NN}}/d_{\text{eff}} \to 1$ in probability for the ordinary pooled estimator. Each covariance direction contributes between zero and one, according to how well the sample resolves it. Thus d_{eff} need not equal the number of nonzero covariance directions.

With ridge weights $h_j = 2s\lambda_j/(1+2s\lambda_j)$, d_{eff} is the classical linear-smoother quantity $\sum_j (2h_j - h_j^2)$ (Buja et al., 1989; Tikhonov, 1963; Hoerl & Kennard, 1970), now evaluated at a sample-selected scale.

The proof first determines the distribution of one anchor's log ratio. We then show that two typical anchors share a point among their first two neighbors with probability tending to zero. We bound this probability using an established Gaussian-affinity identity (Abou-Moustafa & Ferrie, 2012). Moment bounds prevent rare large log ratios from dominating the average. Together, these estimates make the variance of the normalized average tend to zero. The estimates remain valid when $\log n/D \to 0$.

For isotropic Gaussian data, TwoNN requires exponentially many samples to report a fixed fraction of the ambient dimension. A separate theorem proves the corresponding law for uniform points on a noiseless sphere with Euclidean chord distance. Even though the sphere has a known manifold dimension, polynomially many samples make TwoNN report a vanishing fraction of it as dimension grows.

As a secondary contribution, we study variation due to hidden coordinates. At a fixed sample size, an invertible change can be undone before the next operation, preserving outputs while changing representation distances (Kornblith et al., 2019; Cain, 2026). A Gaussian construction shows that TwoNN can rise and fall under such maps while population dimension stays fixed.

Gaussian and spherical experiments measure the finite-sample accuracy of the phase laws, with a neighbor-window refinement improving later-order predictions. A separate 448-cloud Gaussian study finds that the trace-moment estimator closely recovers the oracle shallow target even when observed TwoNN remains below that target. In five CIFAR-10/ResNet-18 models, disjoint-fit PCA equalization increases TwoNN 2.20-fold while preserving predictions; matched random orientations give 0.995-fold. Fitted global-Gaussian and class-mixture simulations leave neural discrepancies that class conditioning alone does not explain.

Contributions.

1. We prove Gaussian and spherical convergence laws for TwoNN on one shared sample, with sample-size thresholds and extensions to fixed neighbor orders.

2. We quantify errors in approximating expected log ratios on the sphere and estimating the Gaussian target at subexponential sample sizes.

3. Controlled experiments measure phase accuracy and target-estimation error; five-model neural studies test disjoint-fit coordinate effects and compare observed TwoNN with fitted Gaussian and class-mixture nulls.

2 NEIGHBOR-RATIO ESTIMATORS AND SAMPLE RESOLUTION

Let $R_{i,k}$ be the kth nearest-neighbor radius of anchor X_i after excluding itself, and put $L_i = \log(R_{i,2}/R_{i,1})$. The pooled estimator is

$$\widehat{d}_{\mathrm{2NN}} = \frac{n-1}{\sum_{i=1}^{n} L_i}. \tag{2}$$

The numerator $n-1$ corrects the inverse-mean bias for independent exponential log ratios (Denti et al., 2022). In our setting, the L_i are exchangeable but dependent because all anchors share the cloud.

For an arbitrary anchor x, let F_x be the conditional CDF of squared radius $S = d(x,Y)^2$, and let $Q_x = F_x^{-1}$. When the radial support is an interval, F_x is continuously differentiable, and its density is positive in the interior, define

$$\beta_x(u) = \frac{\mathrm{d}\log\sqrt{Q_x(u)}}{\mathrm{d}\log u} = \frac{u}{2Q_x(u)f_x\{Q_x(u)\}}. \tag{3}$$

Writing $G_x(r) = F_x(r^2)$, the finite-radius local continuous intrinsic dimensionality and this slope obey the reciprocal identity

$$d_x(r) = \frac{rG_x'(r)}{G_x(r)} = \frac{1}{\beta_x\{F_x(r^2)\}}. \tag{4}$$

Its $r \downarrow 0$ limit, when it exists, is the limiting LID; the finite-radius quantity itself is classical. Conditioning on a sampled anchor leaves $m = n-1$ independent candidates. For their uniform order statistics $U_{(k)} = F_x(S_{(k)})$,

$$\log\frac{R_\ell}{R_k} = \int_{U_{(k)}}^{U_{(\ell)}} \beta_x(u)\ \mathrm{d}\log u, \qquad \frac{U_{(k)}}{U_{(\ell)}} \mid U_{(\ell)} \sim \mathrm{Beta}(k, \ell-k). \tag{5}$$

The probability spacing in TwoNN is therefore exactly exponential. Converting probability into distance uses reciprocal finite-radius LID near probability $1/n$. Only when this window reaches a local power-law regime does the statistic equal the limiting population quantity.

3 SHARED-SAMPLE CONVERGENCE LAWS

3.1 GAUSSIAN SPECTRAL LAW

Let $X_i \overset{\mathrm{iid}}{\sim} \mathcal{N}(0, \Sigma_D)$, with eigenvalues $\lambda_{D,j} \geq 0$. Assume (A1) $\sup_{D,j}\lambda_{D,j} < \infty$ and (A2) $D^{-1}\sum_j \lambda_{D,j} \to m_1 > 0$ and $D^{-1}\sum_j \lambda_{D,j}^2 \to m_2 > 0$. Exact zeros are allowed, but (A1)–(A2) imply rank $\Theta(D)$; fixed or sublinear intrinsic rank is outside this asymptotic regime. For $s > 0$, define

$$\mathcal{C}_D(s) = \frac{1}{2D}\sum_{j=1}^{D}\left[\log(1+2s\lambda_{D,j}) - \frac{2s\lambda_{D,j}}{(1+2s\lambda_{D,j})^2}\right], \tag{6}$$

$$\delta_D(s) = \frac{1}{D}\sum_{j=1}^{D}\left[1-(1+2s\lambda_{D,j})^{-2}\right]. \tag{7}$$

Both increase strictly for large D, with $\mathcal{C}_D(0) = 0$ and $\mathcal{C}_D(s) \to \infty$. Define $s_{D,n}$ by $\mathcal{C}_D(s_{D,n}) = \log(n-1)/D$ and $d_{\mathrm{eff}} = D\delta_D(s_{D,n})$. Larger samples increase $s_{D,n}$ and resolve more spectral directions. Equivalently, for $H_s = 2s\Sigma_D(I+2s\Sigma_D)^{-1}$, $d_{\mathrm{eff}} = \mathrm{tr}(2H_{s_{D,n}} - H_{s_{D,n}}^2)$. The functional is classical; the endogenous choice of $s = s_{D,n}$ and its connection to TwoNN are the point of the theorem.

Theorem 3.1 (Shared-cloud Gaussian phase law)**.** *Let $D \to \infty$, with $n = n_D$, $\log n \to \infty$, and $\log n = O(D)$. Under (A1)–(A2),*

$$\frac{\widehat{d}_{\mathrm{2NN}}}{d_{\mathrm{eff}}} \xrightarrow{\mathbb{P}} 1.$$

For each fixed r and fixed distinct indices $i_1, \dots, i_r$, $(d_{\text{eff}} L_{i_1}, \dots, d_{\text{eff}} L_{i_r})$ converges jointly to independent unit exponentials, with marginal first/second moments tending to $1, 2$ and pairwise cross moments to 1.

Despite candidate reuse, the pooled estimator converges to the explicit sample-selected finite-radius target d_{eff}, which can be much smaller than population dimension in the subexponential-sample regime.

Dependence between anchors. For $S_x = \|Y - x\|^2$, $M_x(s) = \mathbb{E}e^{-sS_x}$, and $C_s = \Sigma_D(I + 2s\Sigma_D)^{-1}$, Gaussian affinity gives

$$\mathbb{E}e^{-s(S_x+S_z)/2} = \sqrt{M_x(s)M_z(s)}e^{-s^2(x-z)^\top C_s(x-z)/2}.$$

At the sample-selected tilt, the last exponent is a positive fraction of $\log n$ for typical independent anchors. The marginal quantile estimate then makes the probability that one candidate enters both anchors' extreme neighbor windows $o(1/n)$ (Lemma C.8). The unconditional moment bound and exchangeable variance identity yield concentration (Lemma C.10).

3.2 Sample complexity and general neighbor orders

If the empirical spectral laws converge to H and $\log n/D \to c > 0$, then $\widehat{d}_{\text{2NN}}/D \to \Delta_H(c)$, obtained by replacing the sums in (6)–(7) by integrals. In the shallow regime $\log n \to \infty$, $\log n = o(D)$,

$$\widehat{d}_{\text{2NN}} = \frac{4}{\sqrt{3}}\sqrt{r_2(\Sigma_D)\log n}\,(1 + o_{\mathbb{P}}(1)), \qquad r_2(\Sigma) = \frac{(\operatorname{tr}\Sigma)^2}{\operatorname{tr}(\Sigma^2)}. \tag{8}$$

Participation ratio controls only the leading behavior when the sample is coarse relative to dimension, not the TwoNN target at every resolution.

For $\Sigma_D = I_D$, write $a = 1 + 2s$. Then $c(a) = \frac{1}{2}(\log a - a^{-1} + a^{-2})$ and the reported fraction is $1 - a^{-2}$. Hence the sharp threshold for a target fraction α is

$$c_*(\alpha) = \frac{1}{2}\left[\log\frac{1}{\sqrt{1-\alpha}} - \sqrt{1-\alpha} + (1-\alpha)\right]. \tag{9}$$

Below, at, and above $\log n/D = c_*(\alpha)$, pooled TwoNN reports less than, converges to, and reports more than αD, respectively. For the isotropic Gaussian model, recovering any fixed fraction of D requires exponentially many samples, with exponent $c_*(\alpha)$.

Estimating the Gaussian target. A same-cloud empirical spectrum need not reproduce the population target. If $n/D \to \gamma \in (0, \infty)$, then

$$\frac{d_{\text{eff}}(n, \widehat{\Sigma})}{d_{\text{eff}}(n, \Sigma_D)} \xrightarrow{\mathbb{P}} \left(1 + \frac{m_1^2}{\gamma m_2}\right)^{-1/2}. \tag{10}$$

For isotropic data and $n/D \to 1$, the limit is $1/\sqrt{2}$. This is classical sample-covariance trace distortion propagated through TwoNN's shallow phase (Proposition D.9). For the later neural comparison, it lowers the Gaussian baseline; at fixed observed TwoNN, it moves the ratio of observed TwoNN to that baseline upward rather than explaining a negative discrepancy.

A consistent shallow predictor instead uses established trace-moment estimators (Chen et al., 2010; Chun et al., 2026). Put $N = \lfloor n/2 \rfloor$, $Z_i = (X_{2i} - X_{2i-1})/\sqrt{2}$, $A = N^{-1}\sum_i \|Z_i\|^2$, and $B = \{N(N-1)\}^{-1}\sum_{i\neq j}(Z_i^\top Z_j)^2$. Then

$$d_U = \frac{4}{\sqrt{3}} A \sqrt{\frac{\log(n-1)}{B}}, \qquad \frac{\widehat{d}_{\text{2NN}}}{d_U} \xrightarrow{\mathbb{P}} 1 \tag{11}$$

under (A1)–(A2), $\log n \to \infty$, and $\log n = o(D)$, without a lower bound on n/D. At positive limiting resolution, the full saddle equation is needed. Proposition D.8 states exact variances and a finite-sample bound for the target-estimation error d_U/d_{eff}.

Later neighbors. For fixed $k < \ell$, the GRIDE-based log-moment estimator $\widehat{d}_{k,\ell} = \{\psi(\ell) - \psi(k)\}/\{n^{-1}\sum_i \log(R_{i,\ell}/R_{i,k})\}$ also converges to d_{eff}. The original Gaussian neighbor-window approximation integrates the reciprocal phase at probability $a/(n-1)$ against the Poisson coverage weight $\pi_{k,\ell}(a)/a$, where $\pi_{k,\ell}(a) = e^{-a}\sum_{j=k}^{\ell-1} a^j/j!$. Proposition D.6 states the formula and its first-order consistency law.

3.3 Spherical sample-resolution law

Let $d \to \infty$ and $n = n_d$. For iid uniform points on $S^d \subset \mathbb{R}^{d+1}$ with chordal distance, put $r_d = \log(n-1) \to \infty$ and assume $r_d \le Kd$. Let $t_{d,n} \in (0,1)$ solve $-\frac{1}{2}\log(1 - t_{d,n}^2) = r_d/d$.

Theorem 3.2 (Shared-cloud sphere law)**.**

$$\frac{\widehat{d}_{2\text{NN}}}{d\, 2t_{d,n}/(1+t_{d,n})} \xrightarrow{\mathbb{P}} 1.$$

If $\log n = o(d)$*, then* $\widehat{d}_{2\text{NN}} = 2\sqrt{2d\log n}(1 + o_{\mathbb{P}}(1))$.

The data lie on a smooth, noiseless manifold of known dimension d, yet polynomially many samples resolve a vanishing fraction of d. The leading constant depends on the chordal metric (Appendix E).

Spherical mean-calibration error. Put $m = n-1$, $r = \log m$, and let $\mu_{k,\ell}$ be the exact expected log spacing, given by Equation (133) of Appendix E. With $\tau(a) = \sqrt{1-(a/m)^{2/d}}$, define

$$M_{k,\ell}^{\text{sph}} = \int_0^{m/2} \frac{\pi_{k,\ell}(a)}{a}\frac{1+\tau(a)}{2d\,\tau(a)}\,\mathrm{d}a.$$

This accounts for neighbor order but drops the cap prefactor. For $0 < a < \sqrt{m}$, instead solve

$$\frac{c_d(1-v^2)^{d/2}}{d\,v} = \frac{a}{m}, \qquad c_d = \frac{\Gamma((d+1)/2)}{\sqrt{\pi}\,\Gamma(d/2)},$$

and integrate $\frac{1+v}{2d\,v}\{1-(1-v^2)/((d+2)v^2)\}$ against $\pi_{k,\ell}(a)\,\mathrm{d}a/a$ to obtain $M_{k,\ell}^{\text{pc}}$.

Proposition 3.3 (Spherical mean calibration)**.** *For fixed* $k < \ell$*,* $r \to \infty$*, and* $r \le Kd$*,*

$$\frac{\mu_{k,\ell}}{M_{k,\ell}^{\text{pc}}} = 1 + O(r^{-2}).$$

If also $r = o(d)$*, then* $\mu_{k,\ell}/M_{k,\ell}^{\text{sph}} - 1 = \frac{\log r}{4r}\{1 + o(1)\}$.

The proof retains two cap integration-by-parts terms and bounds quantile inversion and order-statistic tails (Proposition E.4). Proposition 3.3 shows that retaining the cap prefactor improves deterministic mean calibration relative to the phase-only neighbor-window approximation. The held-out experiment evaluates the phase-only neighbor-window formulas (Propositions D.6 and E.2); Appendix E evaluates this prefactor correction by deterministic quadrature.

4 Coordinate dependence of neural representations

How much can TwoNN change when hidden coordinates change while the network function stays fixed? This secondary question motivates the interventions in Section 5.

At a linear readout, an invertible hidden map B can be absorbed into its weights using B^{-1}, preserving all downstream outputs (Kornblith et al., 2019; Cain, 2026). It can nevertheless change Euclidean neighbor ratios. Our Gaussian construction has a rise-then-fall TwoNN curve under invertible contractions at both shallow and positive limiting resolution, with population dimension and predictions fixed (Theorem F.4 and Corollary F.6). The contractions change which covariance directions the available sample resolves, without discarding information. The appendix also gives a three-point example and the scale-invariance obstruction to universal samplewise monotonicity (Proposition F.1).

Table 1: Paired finite-sample accuracy on the same held-out central clouds: 40 records per neighbor pair, 200 pooled. Entries are median absolute relative errors (lower is better); bold marks the lower error in each row.

Neighbor pair	Single resolution	Neighbor window
$(1, 2)$	**2.17%**	3.97%
$(2, 4)$	7.00%	**4.71%**
$(4, 8)$	11.83%	**5.50%**
$(8, 16)$	16.25%	**6.25%**
$(16, 32)$	20.09%	**6.58%**
All pairs	10.72%	**5.55%**

5 EXPERIMENTS

Experimental roadmap. We ask four questions in sequence. (1) Do the asymptotic phase laws predict finite Gaussian and spherical samples? (2) Can the Gaussian target be estimated when its population spectrum is unknown? (3) Can TwoNN change under coordinate transformations that preserve a network's function? (4) When the Gaussian prediction misses neural measurements, can finite-sample effects or class structure explain the residual? The first two studies separate phase accuracy from target-estimation error; the neural interventions test coordinate sensitivity, and the fitted-null simulations then examine explanations for the remaining neural discrepancy.

Protocol status. The phase-law study uses its declared accuracy criterion, and the neural intervention follows prespecified rules (Appendix K.2). Additional analyses were specified before measurement; those reusing neural models and images are exploratory. Appendix N gives their protocols.

5.1 FINITE-SAMPLE ACCURACY ON GAUSSIAN AND SPHERICAL DATA

Do the phase laws remain accurate at finite sample sizes? Known population spectra and spherical dimensions let us compare predictions and measurements without covariance-estimation error. Agreement would show that the asymptotic laws are numerically accurate on these finite-sample Gaussian and spherical families; systematic error at later neighbors would motivate the neighbor-window correction.

Design. We sample isotropic, two-block, geometric-spectrum, and covariance-rank-$D/2$ Gaussians and uniform spheres over $D \in \{8, 16, 32, 64, 128\}$ and $n \in \{1024, 4096, 16384\}$, with five seeds: 375 clouds in total. Neighbor pairs $(1, 2), (2, 4), (4, 8), (8, 16), (16, 32)$ test whether a prediction at probability $1/n$ remains accurate for later neighbors. We compare each estimate with its model prediction using absolute relative error. The central subset, $D \geq 32$ and $n \geq 4096$, emphasizes the larger dimensions and samples.

We compare both formulas on the same 200 held-out central records. The evaluation split, decision criteria, and computational reproduction are documented in Appendix Sections G and K.5.

Observed accuracy. Pooled TwoNN has 2.51% median error on the full central subset. Later neighbors probe larger probability masses and have increasingly negative single-resolution signed errors. On the same held-out observations, the neighbor-window approximation lowers pooled median error from 10.72% to 5.55%. The gain comes from later orders; for $(1, 2)$, error increases from 2.17% to 3.97% (Table 1). The error reduction is consistent with the neighbor-window formula correcting the resolution shift induced by later neighbors. For spheres, the omitted cap prefactor causes an upward error in the inverse mean prediction (Proposition 3.3); it does not by itself explain the pooled five-family $(1, 2)$ difference.

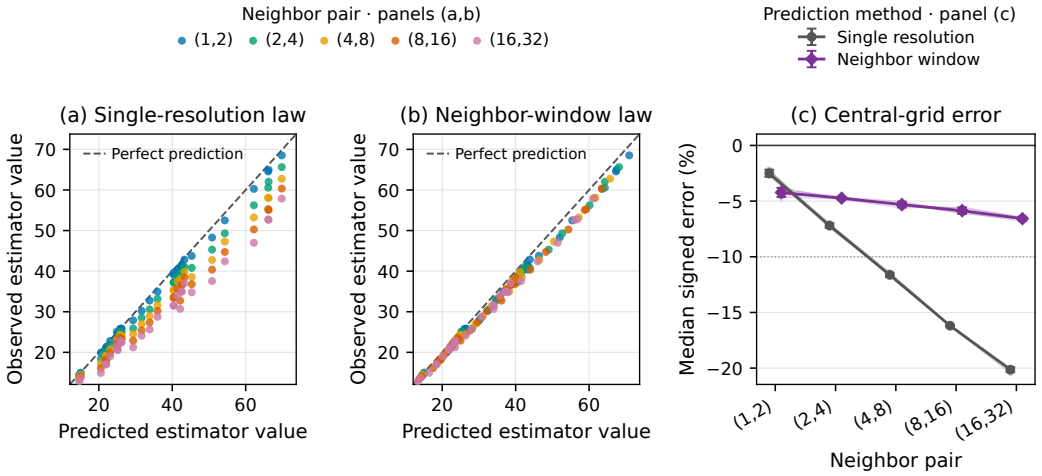


Figure 1: Finite-sample phase accuracy. In (a,b), color identifies the neighbor pair, points are five-seed condition means, and dashed diagonals mark perfect prediction. Panel (c) uses the full central grid: color identifies the prediction method; thin curves are seed-level median signed errors and bold curves are their five-seed means with 95% t-intervals. The neighbor-window formula reduces the systematic overprediction at later orders without fitting a coefficient.

5.2 Estimating the Gaussian target

Using the Gaussian law on data requires estimating its spectral target. We compare covariance plug-in and d_U estimates against known oracle targets; accurate recovery would separate target estimation from the remaining phase-approximation and TwoNN sampling errors.

To test Equation (11), we generated 448 independent Gaussian clouds with isotropic and two-block spectra. Dimensions 128 and 512 use $n/D \in \{1/2, 1, 2\}$ and 32 clouds per cell; a growing-dimension path adds $D = 1024, 2048$ at $n = D$ with 16 clouds per cell. We compare the oracle full phase, oracle shallow target, their empirical-covariance counterparts, d_U, and TwoNN (Figure 2c). At $n = D = 512$, empirical shallow targets are 29.4% and 15.9% low for the isotropic and two-block models, respectively, whereas d_U has median absolute target errors 0.245% and 0.385%. The attenuation agrees with Equation (10). The oracle shallow target is 8.3% and 7.6% above the full phase, while observed TwoNN remains about 12–13% below d_U. Accurate target estimation does not remove finite-resolution approximation error or TwoNN fluctuation (Tables 16–17).

5.3 Function-preserving coordinate interventions

Can TwoNN change substantially while every network prediction stays fixed? By compensating the readout, the coordinate maps hold classifier outputs fixed while changing hidden coordinates. Repeating the comparison with maps fitted on separate images tests whether the effect survives beyond the cloud that selected the map.

We trained five CIFAR-10 ResNet-18 models for 50 epochs and extracted six representations at epochs 0, 25, and 50 on nested test subsets of 1,000–8,000 images. At the final layer-4 linear readout, we apply $z' = \mu + (z - \mu)A$ and compensate the classifier by the inverse affine map. All predictions are preserved. The random-basis family has a median TwoNN range of 1.0097-fold; full-cloud PCA equalization gives 2.124-fold. Its Gaussian-predicted factor is 3.338, so the analytic null overstates the effect's magnitude. Control and estimator comparisons appear in Appendix Tables 13 and 5.

The disjoint-fit test stratifies each model's 8,000 images into two equal halves and reverses their roles. The fitting half determines the mean, PCA basis, eigenvalue floor, reference variance, and clipping scales; identity and transformed conditions share the evaluation half. Same-size in-sample fits separate fitting dependence from a change in n. Five random orientations per fit keep every

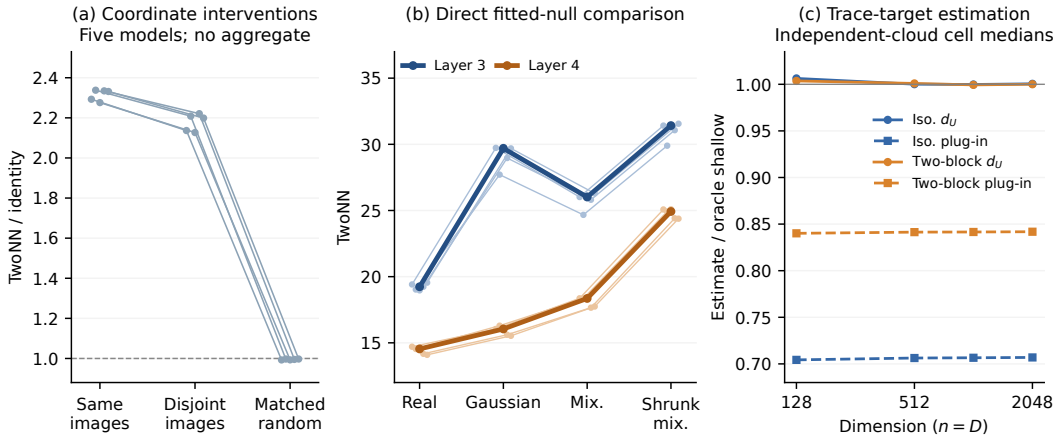


Figure 2: Exploratory analyses. (a) The PCA effect survives disjoint fitting; matched random orientations stay near identity. Lines show five trained models, each summarized over two evaluation folds. (b) Fitted Gaussian and class-mixture nulls remain above neural TwoNN; thin lines show individual models and thick lines their median. (c) At $n = D$, d_U tracks the oracle shallow target while the empirical-covariance plug-in remains attenuated. Curves are medians over 32 independent clouds at $D = 128, 512$ and 16 at $D = 1024, 2048$.

Table 2: Disjoint-fit calibration. Entries first average 16 surrogate draws, then combine folds geometrically within a model and take model medians. Real and surrogate clouds contain 4,000 points. Shrinkage is fixed at 10%.

Layer	Real	Analytic	Gaussian	Mixture	Shrunk mix.
3	19.24	31.84	29.69	26.02	31.42
4	14.54	17.14	16.05	18.35	24.92

singular value identical to PCA. We aggregate the two fold factors geometrically within each trained model.

Disjoint-fit PCA gives a median factor of **2.199** (model range 2.127–2.221), versus 2.331 for same-size in-sample fitting and 0.9952 for the matched random control (Table 6). All 120 maps preserve every prediction, with maximum logit discrepancy 7.49×10^{-6}. The large effect therefore survives separate fitting and evaluation images while remaining sensitive to basis alignment. The five trained models are the experimental units; repeated folds and orientations do not add independent training replicates. Because these images were previously analyzed, the disjoint-fit result is exploratory rather than an untouched-data confirmation.

5.4 Calibrating the neural Gaussian comparison

Coordinate sensitivity alone does not explain the neural residual. We now test two explanations in sequence: finite-sample formula error under a fitted global Gaussian, then class structure under a fitted Gaussian mixture. Agreement with a simulated null would support its corresponding explanation.

Covariance-matched neural Gaussian comparisons precede this study (Ansuini et al., 2019, Section 3.3). Our same-cloud analytic comparison has negative residuals in every model and all six representations at epoch 50 and $n = 8000$, ranging from -9.84% to -41.9% (Appendix Table 7). We now separate finite-sample formula error from departures from specific fitted distributions.

On layers 3 and 4, each 4,000-image fitting half supplies a global Gaussian and ten class Gaussians. We measure TwoNN directly on 16 independent 4,000-point surrogate clouds per fitted family. Mixtures retain evaluation class counts and fitted class means; a fixed sensitivity check replaces each class covariance by $0.9\widehat{\Sigma}_c + 0.1\,\mathrm{tr}(\widehat{\Sigma}_c)I/p$. All 960 surrogate clouds are retained.

Direct Gaussian measurements are about 6–7% below the fitted analytic phase. Even after this calibration, median paired real/Gaussian ratios are 0.658 and 0.907 for layers 3 and 4. Mixtures improve layer 3 but worsen layer 4; real/mixture ratios are 0.745 and 0.794. The discrepancy persists within classes: class-restricted real/mixture ratios are 0.737 and 0.783. At layer 4, both neighbors share the anchor's class for 84.6% of real evaluation points versus 96.0% under the mixture.

Monte Carlo SEs of null means are 0.22–0.65%, but fitting uncertainty remains: class fits contain only 386–415 images, and layer-4 covariances are rank deficient. Shrinkage substantially changes predictions and worsens agreement. Splitting removes same-cloud fitting dependence, not covariance estimation error. These comparisons establish departures from the tested fitted nulls, without uniquely identifying a geometric cause.

6 Related work

TwoNN explicitly varies sample size to study scale-dependent dimension (Facco et al., 2017). GRIDE develops general-order laws and scale analysis (Denti et al., 2022); mFSA studies finite-sample bias (Benko et al., 2022). Houle/Amsaleg's finite-radius LID is $rG'(r)/G(r)$ (Houle, 2013; Amsaleg et al., 2018); Erba et al. study local undersampling and full spherical correlation (Erba et al., 2019). Our target evaluates this established profile at a moving sample-selected quantile.

Random-codebook distortion exponents and refined Gaussian tails precede our one-anchor rate equation (Lapidoth, 1997; Dembo & Kontoyiannis, 2002; Kontoyiannis & Zamir, 2006; Zhou et al., 2019). The ridge filter, effective df, and moving-saddle strong-deviation frameworks are classical (Tikhonov, 1963; Hoerl & Kennard, 1970; Buja et al., 1989; Chaganty & Sethuraman, 1993; Bovier & Mayer, 2015). Our Gaussian theorem supplies the uniform moving-saddle estimates needed for noncentral quadratic forms, then proves the ratio moments and concentration of the shared-sample average. The dependence step uses standard Gaussian affinity (Abou-Moustafa & Ferrie, 2012). Fixed-dimensional spacing limits (Baryshnikov et al., 2009) and moment-matched spectral proxies (Hadfield, 2026) describe fixed-dimensional fluctuations and spectral approximations, respectively; our theorem establishes the joint-limit law and concentration of pooled TwoNN.

Zhang's per-query spherical maximum already has the phase point $\sqrt{1 - m^{-2/d}}$ and cap-tail refinements (Zhang, 2017, Theorems II.3–II.4). Differentiating the cap and converting inner product to chord radius yields our leading ratio center; we add the shared-cloud ratio/moment justification and quantitative neighbor-spacing calibration. Cai–Fan–Jiang study all-pair extrema (Cai et al., 2013). Other manifold estimators have guarantees under additional geometric assumptions (Kim et al., 2019; Block et al., 2022).

Ansuini et al. already report covariance-matched Gaussian surrogates and anisotropy-related TwoNN changes (Ansuini et al., 2019). Kornblith et al. describe invertible compensation and prove an invariance degeneracy theorem for representation similarity (Kornblith et al., 2019); Cain tests compensated metric and neighbor changes (Cain, 2026). Our neural increment is the TwoNN-specific comparison of disjoint-fit, matched-spectrum interventions and directly calibrated Gaussian predictions. The finite-cloud scale argument is supporting context. Neural spectral interpretations and participation-ratio corrections also precede this study (Schulte & Rügamer, 2026; Chun et al., 2026).

7 Scope and limitations

These laws are model- and estimator-specific, not minimax bounds. Neighbor-window formulas are first-order; auxiliary bounds address Gaussian target estimation and spherical mean spacing, not finite-sample TwoNN bias, variance, or mean squared error. Dependent-ratio bias correction, concentration rates, and deep-resolution plug-in bounds remain open.

Neural evidence is limited to one dataset, architecture, intervention site, and five models. Disjoint-fit PCA shows coordinate sensitivity on these representations; reused images preclude independent confirmation. Fitted Gaussian and class-mixture residuals depend on covariance estimates and model choice, and do not measure missing manifold dimension. The controlled d_U result concerns only its Gaussian shallow target. The finite-cloud obstruction assumes exact scale invariance and universal samplewise monotonicity. Other architectures and non-Gaussian nulls remain untested.

AI USE STATEMENT

The authors developed the theoretical model, formulated the mathematical claims, devised the proof ideas and proof strategies, designed the research methodology and experiments, and interpreted the results. Generative AI tools were not used to develop the theoretical model or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, design the research methodology or experiments, or interpret the results.

Generative AI tools were used in assisting roles for presentation, exposition, feedback, and straightforward execution of plans and ideas developed by the authors. In particular, they assisted with the writing and presentation of proofs after the proof ideas, strategies, and mathematical reasoning had been developed by the authors; this assistance included exposition, language choice, organization, and LaTeX typesetting. They also provided feedback on research methodology or experimental designs developed by the authors, assisted with implementing methods specified by the authors, and assisted with cleaning and reformatting datasets. Generative AI tools were not used to generate synthetic datasets. Translation and qualitative or thematic data analysis were not applicable to this work.

The authors reviewed and verified all AI-assisted work, including checking the mathematical exposition against the authors' proofs, testing and inspecting AI-assisted code and implementations, and reviewing the resulting text, figures, data processing, and references for correctness. The authors take responsibility for the final content of this work, including all text, claims, results, code, figures, and other artifacts produced with the assistance of generative AI.

ETHICS STATEMENT

The empirical study uses CIFAR-10 and does not involve a deployment or new data collection from human subjects. The paper concerns the interpretation of representation diagnostics; it does not use intrinsic-dimension estimates to make decisions about individuals. Function-preserving reparameterizations are used as scientific stress tests, not as mechanisms for evading model audits.

REPRODUCIBILITY STATEMENT

The main paper states the assumptions and theorems. Complete proofs and detailed experimental protocols appear in the appendix included with the paper. The supplementary material contains the code needed to reproduce the experiments.

REFERENCES

Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Estimating the reach of a manifold. *Electronic Journal of Statistics*, 13(1):1359–1399, 2019. ISSN 1935-7524. doi: 10.1214/19-EJS1551. Publisher Copyright: © 2019, Institute of Mathematical Statistics. All rights reserved.

Karim T. Abou-Moustafa and Frank P. Ferrie. A note on metric properties for some divergence measures: The gaussian case. In Steven C. H. Hoi and Wray L. Buntine (eds.), *Proceedings of the 4th Asian Conference on Machine Learning, ACML 2012, Singapore, Singapore, November 4-6, 2012*, volume 25 of *JMLR Proceedings*, pp. 1–15. JMLR.org, 2012. URL http://proceedings.mlr.press/v25/aboumoustafa12.html.

Laurent Amsaleg, Oussama Chelly, Teddy Furon, Stéphane Girard, Michael E. Houle, Ken-ichi Kawarabayashi, and Michael Nett. Extreme-value-theoretic estimation of local intrinsic dimensionality. *Data Min. Knowl. Discov.*, 32(6):1768–1805, 2018. doi: 10.1007/S10618-018-0578-6. URL https://doi.org/10.1007/s10618-018-0578-6.

Fabrizio Angiulli. On the behavior of intrinsically high-dimensional spaces: Distances, direct and reverse nearest neighbors, and hubness. *Journal of Machine Learning Research*, 18(170):1–60, 2018. URL http://jmlr.org/papers/v18/17-151.html.

Alessio Ansuini, Alessandro Laio, Jakob H. Macke, and Davide Zoccolan. Intrinsic dimension of data representations in deep neural networks. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (eds.), *Advances in*

Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, pp. 6109–6119, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/cfcce0621b49c983991ead4c3d4d3b6b-Abstract.html.

Yu Baryshnikov, Mathew D. Penrose, and J. E. Yukich. Gaussian limits for generalized spacings. *Annals of Applied Probability*, 19(1):158–185, February 2009. ISSN 1050-5164. doi: 10.1214/08-AAP537.

Zsigmond Benko, Marcell Stippinger, Roberta Rehus, Attila Bencze, Dániel Fabó, Boglárka Hajnal, Loránd G. Eröss, András Telcs, and Zoltán Somogyvári. Manifold-adaptive dimension estimation revisited. *PeerJ Comput. Sci.*, 8:e790, 2022. doi: 10.7717/PEERJ-CS.790. URL https://doi.org/10.7717/peerj-cs.790.

Adam Block, Zeyu Jia, Yury Polyanskiy, and Alexander Rakhlin. Intrinsic dimension estimation using wasserstein distance. *J. Mach. Learn. Res.*, 23:313:1–313:37, 2022. URL https://jmlr.org/papers/v23/21-1483.html.

Anton Bovier and Hannah Mayer. A conditional strong large deviation result and a functional central limit theorem for the rate function. *ALEA, Latin American Journal of Probability and Mathematical Statistics*, 12(1):533–550, 2015.

Andreas Buja, Trevor Hastie, and Robert Tibshirani. Linear Smoothers and Additive Models. *The Annals of Statistics*, 17(2):453 – 510, 1989. doi: 10.1214/aos/1176347115. URL https://doi.org/10.1214/aos/1176347115.

T. Tony Cai, Jianqing Fan, and Tiefeng Jiang. Distributions of angles in random packing on spheres. *J. Mach. Learn. Res.*, 14(1):1837–1864, 2013. doi: 10.5555/2567709.2567722. URL https://dl.acm.org/doi/10.5555/2567709.2567722.

Jericho Cain. Gauge freedom and metric dependence in neural representation spaces. *CoRR*, abs/2603.06774, 2026. doi: 10.48550/ARXIV.2603.06774. URL https://doi.org/10.48550/arXiv.2603.06774.

Narasinga Rao Chaganty and Jayaram Sethuraman. Strong Large Deviation and Local Limit Theorems. *The Annals of Probability*, 21(3):1671 – 1690, 1993. doi: 10.1214/aop/1176989136. URL https://doi.org/10.1214/aop/1176989136.

Song Xi Chen, Li-Xin Zhang, and Ping-Shou Zhong. Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819, 2010. ISSN 01621459. URL http://www.jstor.org/stable/29747085.

Chanwoo Chun, Abdulkadir Canatar, SueYeon Chung, and Daniel Lee. Estimating dimensionality of neural representations from finite samples. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=iM4o9a83F7.

Amir Dembo and Ioannis Kontoyiannis. Source coding, large deviations, and approximate pattern matching. *IEEE Trans. Inf. Theory*, 48(6):1590–1615, 2002. doi: 10.1109/TIT.2002.1003841. URL https://doi.org/10.1109/TIT.2002.1003841.

Francesco Denti, Diego Doimo, Alessandro Laio, and Antonietta Mira. The generalized ratios intrinsic dimension estimator. *Scientific Reports*, 12:20005, 2022. doi: 10.1038/s41598-022-20991-1.

Vittorio Erba, Marco Gherardi, and Pietro Rotondo. Intrinsic dimension estimation for locally undersampled data. *Scientific Reports*, 9(1):17133, 2019. doi: 10.1038/s41598-019-53549-9. URL https://doi.org/10.1038/s41598-019-53549-9.

Elena Facco, Maria d'Errico, Alex Rodriguez, and Alessandro Laio. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Scientific Reports*, 7(1):12140, 2017. doi: 10.1038/s41598-017-11873-y. URL https://doi.org/10.1038/s41598-017-11873-y.

Amir Massoud Farahmand, Csaba Szepesvári, and Jean-Yves Audibert. Manifold-adaptive dimension estimation. In Zoubin Ghahramani (ed.), *Machine Learning, Proceedings of the Twenty-Fourth International Conference (ICML 2007), Corvallis, Oregon, USA, June 20-24, 2007*, volume 227 of *ACM International Conference Proceeding Series*, pp. 265–272. ACM, 2007. doi: 10.1145/1273496.1273530. URL https://doi.org/10.1145/1273496.1273530.

Chris R. Giannella. Instability results for euclidean distance, nearest neighbor search on high dimensional gaussian data. *Inf. Process. Lett.*, 169:106115, 2021. doi: 10.1016/J.IPL.2021.106115. URL https://doi.org/10.1016/j.ipl.2021.106115.

Tom Hadfield. Patnaik-pearson intrinsic dimension for internal representations of neural networks. *CoRR*, abs/2606.19268, 2026. doi: 10.48550/ARXIV.2606.19268. URL https://doi.org/10.48550/arXiv.2606.19268.

Rob Heylen and Paul Scheunders. Hyperspectral intrinsic dimensionality estimation with nearest-neighbor distance ratios. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.*, 6(2-2):570–579, 2013. doi: 10.1109/JSTARS.2013.2256338. URL https://doi.org/10.1109/JSTARS.2013.2256338.

Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970. doi: 10.1080/00401706.1970.10488634. URL https://doi.org/10.1080/00401706.1970.10488634.

Michael E. Houle. Dimensionality, discriminability, density and distance distributions. In Wei Ding, Takashi Washio, Hui Xiong, George Karypis, Bhavani Thuraisingham, Diane J. Cook, and Xindong Wu (eds.), *13th IEEE International Conference on Data Mining Workshops, ICDM Workshops, TX, USA, December 7-10, 2013*, pp. 468–473. IEEE Computer Society, 2013. doi: 10.1109/ICDMW.2013.139. URL https://doi.org/10.1109/ICDMW.2013.139.

Jisu Kim, Alessandro Rinaldo, and Larry A. Wasserman. Minimax rates for estimating the dimension of a manifold. *J. Comput. Geom.*, 10(1):42–95, 2019. doi: 10.20382/JOCG.V10I1A3. URL https://doi.org/10.20382/jocg.v10i1a3.

Ioannis Kontoyiannis and Ram Zamir. Mismatched codebooks and the role of entropy coding in lossy data compression. *IEEE Trans. Inf. Theory*, 52(5):1922–1938, 2006. doi: 10.1109/TIT.2006.872845. URL https://doi.org/10.1109/TIT.2006.872845.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey E. Hinton. Similarity of neural network representations revisited. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3519–3529. PMLR, 2019. URL http://proceedings.mlr.press/v97/kornblith19a.html.

D. Kuonen. Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika*, 86(4):929–935, 1999. ISSN 00063444, 14643510. URL http://www.jstor.org/stable/2673596.

Amos Lapidoth. On the role of mismatch in rate distortion theory. *IEEE Trans. Inf. Theory*, 43(1):38–47, 1997. doi: 10.1109/18.567644. URL https://doi.org/10.1109/18.567644.

Elizaveta Levina and Peter Bickel. Maximum likelihood estimation of intrinsic dimension. In L. Saul, Y. Weiss, and L. Bottou (eds.), *Advances in Neural Information Processing Systems*, volume 17. MIT Press, 2004. URL https://proceedings.neurips.cc/paper_files/paper/2004/file/74934548253bcab8490ebd74afed7031-Paper.pdf.

Eng-Jon Ong, Omer Bobrowski, Gesine Reinert, and Primoz Skraba. A universal nearest-neighbor estimator for intrinsic dimensionality. *CoRR*, abs/2603.10493, 2026. doi: 10.48550/ARXIV.2603.10493. URL https://doi.org/10.48550/arXiv.2603.10493.

Alexandre Quemy. A hub of short rows inflates intrinsic dimension estimation of token embeddings, 2026. URL https://arxiv.org/abs/2608.29702.

Rickmer Schulte and David Rügamer. Rethinking intrinsic dimension estimation in neural representations. In Emtiyaz Khan, Yingzhen Li, Arno Solin, and Aaditya Ramdas (eds.), *Proceedings of The 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 10–18. PMLR, 02–05 May 2026. URL https://proceedings.mlr.press/v300/schulte26a.html.

Piotr Tempczyk, Dominik Filipiak, Łukasz Garncarek, Ksawery Smoczynski, and Adam Kurpisz. Why we need new benchmarks for local intrinsic dimension estimation. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 23495–23528, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/286888e6329c0af52777830ce2af00a5-Paper-Conference.pdf.

A. N. Tikhonov. On the solution of ill-posed problems and the method of regularization. *Dokl. Akad. Nauk SSSR*, 151(3):501–504, 1963. URL https://www.mathnet.ru/eng/dan28329.

Peter J. Verveer and Robert P. W. Duin. An evaluation of intrinsic dimensionality estimators. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(1):81–86, 1995. doi: 10.1109/34.368147. URL https://doi.org/10.1109/34.368147.

Nick Whiteley, Annie Gray, and Patrick Rubin-Delanchy. Statistical exploration of the manifold hypothesis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 88(2): 353–385, 04 2026. ISSN 1369-7412. doi: 10.1093/jrsssb/qkag055. URL https://doi.org/10.1093/jrsssb/qkag055.

Kai Zhang. Spherical cap packing asymptotics and rank-extreme detection. *IEEE Trans. Inf. Theory*, 63(7):4572–4584, 2017. doi: 10.1109/TIT.2017.2700202. URL https://doi.org/10.1109/TIT.2017.2700202.

Lin Zhou, Vincent Y. F. Tan, and Mehul Motani. Refined asymptotics for rate-distortion using gaussian codebooks for arbitrary sources. *IEEE Trans. Inf. Theory*, 65(5):3145–3159, 2019. doi: 10.1109/TIT.2018.2886959. URL https://doi.org/10.1109/TIT.2018.2886959.

A Related work

Neighbor-ratio dimension estimators. The Levina–Bickel likelihood estimator uses local k-nearest-neighbor spacings (Levina & Bickel, 2004). TwoNN specializes to the first two neighbors and derives a Pareto law from local constant density (Facco et al., 2017); GRIDE extends the ratio law to general neighbor orders and multiscale analysis (Denti et al., 2022). Revisiting the FSA estimator of Farahmand et al. (2007), Benko et al. (2022) derive normalized-neighbor-distance beta laws under local uniformity, introduce median aggregation, and calibrate corrections for finite-sample and boundary-induced underestimation. These are precedents for local ratio distributions and bias correction, rather than jointly growing-dimensional shared-cloud concentration. Earlier distance-ratio methods also include hyperspectral estimators (Heylen & Scheunders, 2013). Local continuous intrinsic dimensionality evaluates the radial distribution through $rG'_x(r)/G_x(r)$ at finite distance; its limiting LID is obtained as $r \downarrow 0$, when that limit exists (Houle, 2013; Amsaleg et al., 2018). Our inverse-CDF slope is its reciprocal at the corresponding finite radius; we state the identity in Lemma B.1. Finite-radius LID and its small-radius limit are established in this literature. TwoNN uses sample decimation to study scale-dependent dimension (Facco et al., 2017); locally undersampled geometry and full spherical correlation are studied by Erba et al. (2019). Our contribution is to evaluate the sample-selected profile in a growing-dimensional model and prove concentration for the actual estimator that reuses one cloud across anchors. Ong et al.'s L2N2 replaces the log ratio by a log–log statistic with a distribution-independent fixed-dimensional manifold limit and a learned finite-sample Gaussian calibration (Ong et al., 2026). We include both reported versions as comparators. Its fixed-d consistency and our jointly growing-n, D resolution phase address different limits.

Random codebooks and one-anchor extremes. Conditioned on one anchor, nearest-neighbor search compares one source point with an iid random codebook under squared-error distortion. The one-expected-codeword equation $I(q) = \log(n-1)/D$, mismatched codebooks, and refined Gaussian-codebook asymptotics therefore have direct precedents in source coding and approximate pattern matching (Lapidoth, 1997; Dembo & Kontoyiannis, 2002; Kontoyiannis & Zamir, 2006; Zhou et al., 2019). For $\Sigma_D = I_D$, putting $a = 1 + 2s$ gives

$$\mathcal{C}_D(s) = \tfrac{1}{2}\{\log a - a^{-1} + a^{-2}\},$$

the same-source iid-Gaussian codebook distortion-ball exponent in this parameterization. Our contribution is the reused-cloud ratio law: we control the probability that one candidate is extreme for two anchors and prove concentration of the all-anchor aggregate.

Spectral shrinkage and saddlepoint precedent. The ingredients of our Gaussian target have classical counterparts. Ridge and Tikhonov regularization use spectral weights $h_j = \lambda_j/(\lambda_j + \alpha)$ (Tikhonov, 1963; Hoerl & Kennard, 1970). For a symmetric linear smoother H, $\mathrm{tr}(2H - H^2) = \sum_j (2h_j - h_j^2)$ is one classical effective-degrees-of-freedom functional (Buja et al., 1989). With $\alpha = (2s)^{-1}$, this is the algebraic form of our d_{eff}. Gaussian quadratic-form saddlepoint methods also classically produce the log-determinant and rational terms appearing in $\mathcal{C}_D$ (Kuonen, 1999). We evaluate these classical spectral and saddlepoint quantities at the nearest-neighbor scale selected by $\log(n-1)/D$. Strong-deviation theorems allow moving tilts with $s\sqrt{D} \to \infty$ (Chaganty & Sethuraman, 1993); conditional random-environment refinements are developed by Bovier & Mayer (2015). We verify the needed uniform estimates for Gaussian quadratic forms, including random quantiles and unconditional moments. Candidate separation follows from classical Gaussian affinity (Abou-Moustafa & Ferrie, 2012), and the resulting shared-average convergence identifies the TwoNN target.

Finite samples and high-dimensional neighbors. High-dimensional distance concentration, direct and reverse neighbors, and hubness are well documented (Angiulli, 2018); empirical studies have long observed negative bias of dimension estimators on high-dimensional spheres (Verveer & Duin, 1995). For Gaussian data, covariance-dependent nearest-neighbor instability and its subexponential/exponential sample-size transition were established by Giannella (2021); that work does not identify the TwoNN ratio center. Fixed-dimensional shared-cloud spacing functionals also admit stabilization and Gaussian limits (Baryshnikov et al., 2009), but those local stabilization theorems do not directly cover $D \to \infty$ with neighborhood probabilities of order $1/n$. Hadfield proposes a Patnaik–Pearson moment-matched spectral approximation for TwoNN and applies it to neural

representations (Hadfield, 2026). Theorem C.2 derives the asymptotic spectral center for the random first-to-second ratio through $\log n = O(D)$, proves concentration of the shared-cloud average, and derives matching thresholds for this estimator.

Spherical extremes. Zhang's per-query maximum has phase point $t \sim \sqrt{1 - m^{-2/d}}$, with fluctuations and incomplete-beta tail estimates (Zhang, 2017, Theorems II.3–II.4 and Lemma A.1). The use of absolute rather than one-sided inner products changes the relevant tail mass by a factor two, which does not change this leading phase. Writing G for the one-sided cap tail and h for its density, $h(t)/G(t) \sim dt/(1 - t^2)$; chord radius $R = \sqrt{2(1 - t)}$ then gives $1/\beta \sim 2dt/(1 + t)$. Thus our leading spherical center follows from classical cap asymptotics and a radial change of variables. We derive the ratio law, the moment and two-cap estimates for its shared average, and the quantitative calibration comparison. Cai–Fan–Jiang study extrema over all approximately $n^2/2$ pairs (Cai et al., 2013).

Population manifold dimension. Under reach, density, smoothness, and noise assumptions, manifold-dimension estimators can have finite-sample or minimax guarantees (Farahmand et al., 2007; Kim et al., 2019; Block et al., 2022). Reach itself must also be estimated (Aamari et al., 2019). These guarantees estimate a declared population quantity under a stated model. We instead study what a widely used ratio statistic measures before it reaches the infinitesimal regime. The sphere theorem isolates under-resolution on a model with known dimension. Statistical formulations of the manifold hypothesis address a separate question (Whiteley et al., 2026).

Neural representations, spectra, and gauge freedom. Ansuini et al. compare neural TwoNN with covariance-matched Gaussian surrogates (Section 3.3, Figure 5B), finding larger, sample-dependent Gaussian values; their Section 3.4 also studies anisotropy-related TwoNN changes (Ansuini et al., 2019). We repeat that comparison with an analytic Gaussian spectral prediction across five training seeds. Recent work shows that common neural estimates need not track Hausdorff or pointwise dimension and are closely related to spectral quantities (Schulte & Rügamer, 2026); participation-ratio estimation itself has finite-sample corrections (Chun et al., 2026). Controlled benchmarks also show that estimator performance on simple known manifolds need not transfer across domains (Tempczyk et al., 2026). A contemporaneous token-embedding study shows that a short-row hub can inflate TwoNN and that normalization or hub removal changes the reported curve (Quemy, 2026); this is an empirical heterogeneous-cloud mechanism, not the homogeneous Gaussian phase analyzed here. Separately, hidden coordinates have a general-linear gauge freedom. Kornblith et al. explicitly state the compensation $X \mapsto XA$, $W \mapsto A^{-1}W$, and prove degeneracy of fully invertible-invariant representation similarities (Kornblith et al., 2019). Their statistic and axioms differ from our elementary finite-metric scale argument. Cain studies compensated invertible maps and reports changes in neural metric geometry, including nearest neighbors and whitening (Cain, 2026). Our increment is the TwoNN-specific magnitude test against the phase prediction. Function-preserving compensation is established prior work, and a metric statistic can remain useful for describing a trained representation even when it is not a function invariant.

Table 3: Theorem-level boundary between the closest precedents and the present theorem package. "Independent codebook" means one source is compared with fresh iid candidates; "shared cloud" means every observation is both source and candidate.

Precedent	Established component	Increment established here
TwoNN; mFSA; GRIDE (Facco et al., 2017; Benko et al., 2022; Denti et al., 2022)	Local-uniform neighbor-ratio distributions, aggregation, and finite-sample calibration	Moving Gaussian resolution and concentration for the estimator that reuses one cloud
Houle; Amsaleg et al.	Finite-radius local continuous ID/LID profile and its small-radius limit	Explicit growing-D Gaussian and spherical evaluations at the sample-selected quantile; reused-cloud estimator convergence
Random-codebook theory	One source, an independent codebook, distortion-ball rates, and refined Gaussian asymptotics	Uniform ratio hazard at a moving quantile; two-anchor overlap bound; all-anchor concentration
Giannella	Covariance-dependent Gaussian nearest-neighbor instability regimes	Asymptotically exact TwoNN center and estimator-specific thresholds
Baryshnikov–Penrose–Yukich	Shared-cloud spacing limits under fixed-dimensional stabilization	Growing dimension and exponentially rare neighborhoods with a moving saddle
Hadfield	Moment-matched Patnaik–Pearson spectral proxy	Asymptotically exact spectral center for the first-to-second ratio and concentration of pooled TwoNN
Zhang; Cai–Fan–Jiang	Per-query and global spherical extreme phases and cap-tail refinements	Shared-cloud ratio/moment law and quantitative neighbor-spacing calibration
Schulte–Rügamer	Neural ID estimates can track spectral structure rather than population ID	Canonical sample-resolution law and reused-cloud concentration
Ansuini; Kornblith; Cain	Gaussian-surrogate diagnostic, invertible compensation, invariance degeneracy, and neural metric sensitivity	TwoNN-specific magnitudes and comparison with the analytic target
Ong et al.	A different nearest-neighbor statistic with fixed-d consistency	Pooled TwoNN in a jointly growing finite-resolution regime

B NEIGHBOR-RATIO ESTIMATORS AND SAMPLE RESOLUTION

Let $X_1, \ldots, X_n$ be points in a metric space. For anchor X_i, write $R_{i,k}$ for its kth nearest-neighbor radius after excluding the anchor itself, and put

$$L_i = \log \frac{R_{i,2}}{R_{i,1}}.$$

We use the pooled inverse-mean-log-ratio estimator of Denti et al. (2022, Eq. (3)), with numerator $n-1$:

$$\widehat{d}_{\mathrm{2NN}} = \frac{n-1}{\sum_{i=1}^{n} L_i}. \tag{12}$$

We call this the MacKay aggregate to distinguish pooling log ratios from averaging their reciprocals. It also differs from Facco's empirical-CDF regression, which fits a line to transformed, ordered ratios. Throughout, TwoNN denotes the pooled estimator in Equation (12). The numerator $n-1$ is unbiased

in the idealized model of n independent exponential log ratios, since their sum is gamma with shape n; it is not an exact finite-sample bias correction for the dependent cloud. All anchors reuse that cloud, so the L_i are exchangeable but dependent.

We first isolate an identity that does not use a manifold or Gaussian model. Fix an anchor x, let $S_j = d(x, Y_j)^2$ for m independent reference points, and suppose its conditional support is an interval with positive interior, its conditional CDF F_x is continuously differentiable, and its density f_x is strictly positive in the interior of the support. Write $Q_x = F_x^{-1}$. These conditions hold for the Gaussian and spherical conditional radial laws used below. They also exclude gaps in the radial support, across which the ordinary derivative of the quantile would miss a jump. If $S_{(k)}$ is the kth order statistic, then

$$U_{(k)} = F_x(S_{(k)})$$

is the kth order statistic of m independent uniforms.

Lemma B.1 (Exact inverse-CDF representation)**.** *Define the radial inverse-CDF log slope*

$$\beta_x(u) = \frac{\mathrm{d}\log\sqrt{Q_x(u)}}{\mathrm{d}\log u} = \frac{u}{2Q_x(u)f_x(Q_x(u))}. \tag{13}$$

Let $G_x(r) = F_x(r^2)$ be the CDF of radius and define its finite-radius local continuous intrinsic dimensionality (LID profile) by

$$d_x(r) = \frac{rG_x'(r)}{G_x(r)}. \tag{14}$$

Then, for $u = F_x(r^2)$,

$$d_x(r) = \frac{1}{\beta_x(u)}. \tag{15}$$

When $\lim_{r\downarrow 0} d_x(r)$ exists, that limit is the limiting LID at x. For fixed $1 \le k < \ell$,

$$\log\frac{R_\ell}{R_k} = \int_{U_{(k)}}^{U_{(\ell)}} \beta_x(u)\ \mathrm{d}\log u. \tag{16}$$

Moreover,

$$\frac{U_{(k)}}{U_{(\ell)}}\ \Big|\ U_{(\ell)} \sim \mathrm{Beta}(k, \ell - k). \tag{17}$$

In particular, $\log(U_{(2)}/U_{(1)}) \sim \mathrm{Exp}(1)$ exactly.

Proof. The probability integral transform gives the uniform order statistics. The fundamental theorem of calculus applied to $u \mapsto \log\sqrt{Q_x(u)}$ gives (16). Conditional on $U_{(\ell)} = u$, the lower $\ell - 1$ points are distributed as the order statistics of $\ell - 1$ independent uniforms on $(0, u)$. Their rescaled kth order statistic is $\mathrm{Beta}(k, \ell - k)$, which proves (17). For $(k, \ell) = (1, 2)$ this beta variable is uniform, and minus its logarithm is unit exponential. Finally, $G_x'(r) = 2rf_x(r^2)$, so

$$\frac{rG_x'(r)}{G_x(r)} = \frac{2r^2 f_x(r^2)}{F_x(r^2)} = \frac{1}{\beta_x\{F_x(r^2)\}},$$

which proves (15). □

Lemma B.1 isolates the role of sample resolution. The probability spacing has a known beta law, but converting it to a distance ratio uses reciprocal finite-radius continuous ID over a random probability window. The statistic estimates the local limit only after this window has entered a power-law regime in which $d_x(r)$ is approximately constant.

C The Gaussian sample-resolution law

C.1 Model and the sample-selected finite-radius target

Let

$$X_1, \ldots, X_n \overset{\mathrm{iid}}{\sim} \mathcal{N}(0, \Sigma_D),$$

where Σ_D has eigenvalues $\lambda_{D,1}, \ldots, \lambda_{D,D} \ge 0$. We impose the following spectral regularity conditions.

(A1) Bounded operator norm.
There is an $L < \infty$ such that $\max_j \lambda_{D,j} \le L$ for every D.

(A2) Nondegenerate spectral mass.
For constants $m_1, m_2 > 0$,

$$\frac{1}{D}\sum_j \lambda_{D,j} \to m_1, \qquad \frac{1}{D}\sum_j \lambda_{D,j}^2 \to m_2.$$

Exact zero eigenvalues are allowed.

These conditions require macroscopic spectral mass: since $\mathrm{rank}(\Sigma_D) \ge \mathrm{tr}(\Sigma_D)/L$, the rank is $\Theta(D)$. A fixed or sublinear rank embedded in growing ambient dimension is outside this regime, although a fixed positive fraction of zero eigenvalues is permitted.

For $s > 0$, define

$$\mathcal{C}_D(s) = \frac{1}{2D}\sum_{j=1}^{D}\left[\log(1+2s\lambda_{D,j}) - \frac{2s\lambda_{D,j}}{(1+2s\lambda_{D,j})^2}\right], \tag{18}$$

$$\delta_D(s) = \frac{1}{D}\sum_{j=1}^{D}\left[1-(1+2s\lambda_{D,j})^{-2}\right], \tag{19}$$

$$y_D(s) = \frac{1}{D}\sum_{j=1}^{D}\left[\frac{\lambda_{D,j}}{1+2s\lambda_{D,j}} + \frac{\lambda_{D,j}}{(1+2s\lambda_{D,j})^2}\right]. \tag{20}$$

Direct algebra gives

$$\delta_D(s) = 2s y_D(s). \tag{21}$$

Lemma C.1 (Shape of the finite spectral curve)**.** *For all sufficiently large D, $\mathcal{C}_D$ and δ_D are strictly increasing, $\mathcal{C}_D(0) = 0$, and $\mathcal{C}_D(s) \to \infty$ as $s \to \infty$. Moreover,*

$$\mathcal{C}_D'(s) = \frac{1}{D}\sum_j \frac{2s\lambda_{D,j}^2(3+2s\lambda_{D,j})}{(1+2s\lambda_{D,j})^3} > 0, \tag{22}$$

$$\delta_D'(s) = \frac{4}{D}\sum_j \frac{\lambda_{D,j}}{(1+2s\lambda_{D,j})^3} > 0. \tag{23}$$

As $s \downarrow 0$,

$$\mathcal{C}_D(s) = 3m_{2,D}s^2 + O(s^3), \tag{24}$$

$$\delta_D(s) = 4m_{1,D}s - 12m_{2,D}s^2 + O(s^3), \tag{25}$$

where $m_{r,D} = D^{-1}\sum_j \lambda_{D,j}^r$ and the remainders are uniform in D.

Proof. Differentiate (18) and (19). Positivity follows from (A2). For all large D, $m_{2,D} \ge m_2/2$. Set $\lambda_0 = \sqrt{m_2}/2$ and $\kappa = m_2/(4L^2)$, and let J_D count eigenvalues exceeding λ_0. If $J_D < \kappa D$, then

$$m_{2,D} \le \frac{J_D}{D}L^2 + \left(1-\frac{J_D}{D}\right)\frac{m_2}{4} \le \frac{J_D}{D}L^2 + \frac{m_2}{4} < \frac{m_2}{2},$$

a contradiction. Thus at least κD eigenvalues exceed λ_0. Their log terms grow as $\log s$, whereas the rational terms are bounded, so $\mathcal{C}_D(s) \to \infty$. The expansions follow from Taylor's theorem and (A1). □

C.2 Main shared-cloud theorem

Put

$$c_{D,n} = \frac{\log(n-1)}{D}.$$

This ratio controls sample resolution. A nearest neighbor probes an event of probability about $1/n$; in high dimension, reaching substantially smaller radii therefore requires n to grow exponentially with D. By Lemma C.1, for all sufficiently large D there is a unique $s_{D,n} > 0$ satisfying

$$\mathcal{C}_D(s_{D,n}) = c_{D,n}. \tag{26}$$

Define the sample-selected effective dimension

$$d_{\text{eff}}(n, \Sigma_D) = D\delta_D(s_{D,n}) = \sum_{j=1}^{D} \left[1 - (1 + 2s_{D,n}\lambda_{D,j})^{-2}\right]. \tag{27}$$

The parameter $s_{D,n}$ records how deeply the sample reaches into the lower tail of pairwise distances. Each covariance direction contributes a number between zero and one to d_{eff}: well-resolved directions count almost fully, while weak or unresolved directions count only partially. To display the classical algebra explicitly, define

$$H_s = 2s\Sigma_D(I + 2s\Sigma_D)^{-1}.$$

Its eigenvalues are $h_j = 2s\lambda_{D,j}/(1 + 2s\lambda_{D,j}) = \lambda_{D,j}/(\lambda_{D,j} + (2s)^{-1})$, and hence

$$d_{\text{eff}} = \operatorname{tr}(2H_{s_{D,n}} - H_{s_{D,n}}^2) = \sum_j (2h_j - h_j^2). \tag{28}$$

Equation (28) is a classical effective-df spectral response evaluated at a scale chosen by the extreme-neighbor probability equation. Theorem C.2 additionally proves, uniformly in growing dimension, that the ordinary shared-cloud statistic concentrates at this finite-radius target.

Theorem C.2 (TwoNN on one shared Gaussian cloud)**.** *Assume* (A1)–(A2)*. Let $D \to \infty$ along any sequence $n = n_D$ satisfying*

$$\log n \to \infty, \qquad \log n = O(D). \tag{29}$$

For the pooled shared-cloud estimator (12)*,*

$$\frac{\widehat{d}_{\text{2NN}}}{d_{\text{eff}}(n, \Sigma_D)} \xrightarrow{\mathbb{P}} 1. \tag{30}$$

More strongly, for each fixed r and fixed distinct indices $i_1, \dots, i_r$,

$$\left(d_{\text{eff}} L_{i_1}, \dots, d_{\text{eff}} L_{i_r}\right) \xrightarrow{\mathcal{L}} (E_1, \dots, E_r), \tag{31}$$

where $E_1, \dots, E_r$ are independent $\text{Exp}(1)$ *variables. The convergence also holds in first and second joint moments: for fixed $i \neq j$,*

$$\mathbb{E}[d_{\text{eff}} L_i] \to 1, \qquad \mathbb{E}[(d_{\text{eff}} L_i)^2] \to 2, \qquad \mathbb{E}[(d_{\text{eff}} L_i)(d_{\text{eff}} L_j)] \to 1.$$

Theorem C.2 states that the ordinary TwoNN estimate, despite reusing one dependent data cloud, converges to d_{eff}. That quantity need not equal the covariance rank or the ambient dimension. When n is subexponential in D, it can be much smaller than D even for a full-rank isotropic Gaussian. Recovering a fixed fraction of D requires exponentially many samples. The proof dependencies are: the uniform tilted local limit (Lemma C.3) and unconditional ratio tails (Lemma C.4) give the moving-saddle law and moment control; the two-anchor gap (Lemma C.8) gives independent extreme processes (Lemma C.9); moment control then yields cross-moment factorization and exchangeable concentration (Lemma C.10).

C.3 Conditional saddlepoint and one-anchor law

Work in the eigenbasis of Σ_D. Conditional on an anchor $X = x$, let

$$S_x = \|Y - x\|^2, \qquad Y \sim \mathcal{N}(0, \Sigma_D).$$

Its normalized conditional cumulant generating function is

$$\Lambda_{D,x}(t) = -\frac{1}{2D}\sum_{j=1}^{D} \log(1 - 2t\lambda_{D,j}) + \frac{1}{D}\sum_{j=1}^{D} \frac{tx_j^2}{1 - 2t\lambda_{D,j}}. \tag{32}$$

For a typical Gaussian anchor, weighted laws of large numbers make this and its first three derivatives uniformly close on compact negative t-sets to

$$\Lambda_D(t) = -\frac{1}{2D}\sum_j \log(1-2t\lambda_{D,j}) + \frac{1}{D}\sum_j \frac{t\lambda_{D,j}}{1-2t\lambda_{D,j}}. \tag{33}$$

At $t=-s$,

$$\Lambda_D'(-s) = y_D(s). \tag{34}$$

The Legendre value at this saddlepoint is

$$-sy_D(s) - \Lambda_D(-s) = \mathcal{C}_D(s). \tag{35}$$

The required sharp lower-tail statement is a triangular-array version of the standard exponential-tilting or strong large-deviation estimate (Chaganty & Sethuraman, 1993). We first prove the uniform local theorem needed for that estimate; this is the analytic step on which both the moving quantile and the moment bounds depend.

Lemma C.3 (Uniform tilted local limit)**.** *Assume* (A1)–(A2)*. Conditional on $x \sim \mathcal{N}(0,\Sigma_D)$, let $S_x = \|Y-x\|^2$ and, under exponential tilting by $-s$, let $g_{D,x,s}$ be its density. Put $a_j(s) = 1+2s\lambda_{D,j}$ and define*

$$y_{D,x}(s) = \frac{1}{D}\sum_j \left\{ \frac{\lambda_{D,j}}{a_j(s)} + \frac{x_j^2}{a_j(s)^2} \right\},$$
$$v_{D,x}(s) = \frac{1}{D}\sum_j \left\{ \frac{2\lambda_{D,j}^2}{a_j(s)^2} + \frac{4\lambda_{D,j}x_j^2}{a_j(s)^3} \right\},$$
$$\mathcal{C}_{D,x}(s) = -sy_{D,x}(s) - \Lambda_{D,x}(-s).$$

There are deterministic sets $\mathcal{G}_D$ with $\mathbb{P}(X\in\mathcal{G}_D)\to 1$ and $\epsilon_D\to 0$ such that, uniformly for $x\in\mathcal{G}_D$, $s\ge 0$, and $z\in\mathbb{R}$,

$$\left| \sqrt{Dv_{D,x}(s)}\, g_{D,x,s}\{Dy_{D,x}(s)+z\} - \phi\left(\frac{z}{\sqrt{Dv_{D,x}(s)}}\right)\right| \le \epsilon_D. \tag{36}$$

Moreover,

$$\frac{c\sqrt{D}}{1+s} \le \sqrt{Dv_{D,x}(s)} \le \frac{C\sqrt{D}}{1+s}, \tag{37}$$

$$\frac{c}{\sqrt{Dv_{D,x}(s)}} \le g_{D,x,s}\{Dy_{D,x}(s)\} \le \sup_q g_{D,x,s}(q) \le \frac{C}{\sqrt{Dv_{D,x}(s)}}. \tag{38}$$

If $0<s_D\le S$ and $s_D\sqrt{D}\to\infty$, then uniformly on $\mathcal{G}_D$,

$$\frac{f_{D,x}\{Dy_{D,x}(s_D)\}}{F_{D,x}\{Dy_{D,x}(s_D)\}} = s_D(1+o(1)). \tag{39}$$

Finally, with $H_D(r) = \min\{D, \sqrt{D(1+r)}\}$,

$$\beta_x(u) \le \frac{C}{H_D(\log(1/u))}, \qquad 0<u\le\tfrac{1}{2}, \qquad x\in\mathcal{G}_D. \tag{40}$$

Proof. Work on the positive-eigenvalue support and write $x_j = \sqrt{\lambda_{D,j}}g_j$, where $g_j = x_j/\sqrt{\lambda_{D,j}}$ on the positive support and $g_j=0$ when $\lambda_{D,j}=0$. The active-block bound in Lemma C.1 gives fixed $\lambda_0 = \sqrt{m_2}/2$ and $\kappa = m_2/(4L^2) > 0$ such that at least κD eigenvalues exceed λ_0 for all large D. Define

$$\mathcal{G}_D = \left\{ x : D^{-1}\sum_j g_j^2 \le M, \quad D^{-1}\sum_j g_j^6 \le D^{1/12} \right\}. \tag{41}$$

For sufficiently large fixed M, Gaussian concentration and Markov's inequality give, for every fixed q,

$$\mathbb{P}(X \notin \mathcal{G}_D) \le e^{-cD} + C_q D^{-q/12}. \tag{42}$$

Put $a_j(s) = 1 + 2s\lambda_{D,j}$ and $b_s = 1 + s$. Under the $-s$ tilt,

$$Y_j - x_j = \mu_{j,s} + \sqrt{v_{j,s}} Z_j, \qquad \mu_{j,s} = -\frac{x_j}{a_j(s)}, \quad v_{j,s} = \frac{\lambda_{D,j}}{a_j(s)}.$$

Center the rescaled squared coordinates,

$$\xi_{j,s} = b_s \left[(\mu_{j,s} + \sqrt{v_{j,s}} Z_j)^2 - (\mu_{j,s}^2 + v_{j,s}) \right],$$

and set

$$B_{D,x,s}^2 = \sum_j \mathrm{Var}(\xi_{j,s}) = b_s^2 D v_{D,x}(s).$$

Uniformly in j, s,

$$\frac{b_s \lambda_{D,j}}{a_j(s)} \le C, \qquad \frac{b_s x_j^2}{a_j(s)^2} \le C g_j^2,$$

and the first quantity is bounded below on the active κD coordinates. Since a squared Gaussian has variance $2v^2 + 4v\mu^2$, this proves

$$cD \le B_{D,x,s}^2 \le CD \tag{43}$$

and hence (37). Also,

$$\mathbb{E}|(\mu + \sqrt{v}Z)^2 - (\mu^2 + v)|^3 \le C(v^3 + |\mu|^3 v^{3/2}).$$

Therefore, on $\mathcal{G}_D$,

$$\sum_j \mathbb{E}|\xi_{j,s}|^3 \le C D^{1+1/24}, \qquad B_{D,x,s}^{-3} \sum_j \mathbb{E}|\xi_{j,s}|^3 \le C D^{-11/24}. \tag{44}$$

We now prove the Fourier bounds. Let

$$\Psi_{D,x,s}(u) = \mathbb{E} \exp\left(\frac{iu}{B_{D,x,s}} \sum_j \xi_{j,s} \right), \qquad U_D = D^{1/24}.$$

The good-anchor bound implies

$$\max_j \frac{\mathrm{Var}(\xi_{j,s})}{B_{D,x,s}^2} \le C D^{-23/36}.$$

Thus, for $|u| \le U_D$, the logarithmic characteristic-function expansion and (44) give

$$\log \Psi_{D,x,s}(u) = -u^2/2 + \rho_{D,x,s}(u), \qquad |\rho_{D,x,s}(u)| \le C D^{-11/24} |u|^3, \tag{45}$$

$$|\Psi_{D,x,s}(u)| \le e^{-cu^2}. \tag{46}$$

For a squared Gaussian,

$$\left| \mathbb{E} e^{it(\mu + \sqrt{v}Z)^2} \right| = (1 + 4t^2 v^2)^{-1/4} \exp\left\{ -\frac{2t^2 v \mu^2}{1 + 4t^2 v^2} \right\} \le (1 + 4t^2 v^2)^{-1/4}. \tag{47}$$

Using only the active block in (47),

$$|\Psi_{D,x,s}(u)| \le \left(1 + c \frac{u^2}{B_{D,x,s}^2} \right)^{-\kappa D/4} \le e^{-cu^2}, \quad U_D < |u| \le B_{D,x,s}, \tag{48}$$

$$\int_{|u| > B_{D,x,s}} |\Psi_{D,x,s}(u)| \, \mathrm{d}u \le C\sqrt{D} e^{-cD}. \tag{49}$$

Fourier inversion, (45)–(49), and the same Gaussian bounds for $e^{-u^2/2}$ yield

$$\sup_y |p_{D,x,s}(y) - \phi(y)| \le C D^{-11/24},$$

where $p_{D,x,s}$ is the density of $B_{D,x,s}^{-1}\sum_j \xi_{j,s}$. Since

$$\sum_j \xi_{j,s} = b_s\{S_x - Dy_{D,x}(s)\}, \qquad B_{D,x,s} = b_s\sqrt{Dv_{D,x}(s)},$$

this proves (36) with $\epsilon_D = CD^{-11/24}$ and also (38). The identical frequency split in the smoothing inequality gives uniform convergence of the corresponding CDFs to Φ.

For the hazard, set $q_s = Dy_{D,x}(s)$, $\sigma_s = \sqrt{Dv_{D,x}(s)}$, and $h = s\sigma_s$. Inverse tilting gives

$$f_{D,x}(q_s) = e^{-D\mathcal{C}_{D,x}(s)} g_{D,x,s}(q_s), \tag{50}$$

$$F_{D,x}(q_s) = e^{-D\mathcal{C}_{D,x}(s)} \int_0^\infty e^{-sw} g_{D,x,s}(q_s - w)\, \mathrm{d}w. \tag{51}$$

Consequently,

$$s\frac{F_{D,x}(q_s)}{f_{D,x}(q_s)} = \int_0^\infty e^{-t} \frac{g_{D,x,s}(q_s - t/s)}{g_{D,x,s}(q_s)}\, \mathrm{d}t. \tag{52}$$

If $s \le S$ and $h \to \infty$, split the integral at $h^{1/4}$. On the first part, (36) gives the density ratio $1 + O(\epsilon_D + h^{-3/2})$; on the second, (38) bounds the ratio by C and the exponential weight contributes $O(e^{-h^{1/4}})$. This proves (39), including the required $z = O(1/s)$ window.

It remains to prove the global slope claim. Equations (51) and (38) give

$$\frac{F_{D,x}(q_s)}{f_{D,x}(q_s)} \le C\min\{\sigma_s, s^{-1}\} \le \frac{C}{s + \sigma_s^{-1}}. \tag{53}$$

The active block gives $q_s \ge cD/(1+s)$, hence

$$2q_s \frac{f_{D,x}(q_s)}{F_{D,x}(q_s)} \ge c\left\{\sqrt{D} + \frac{Ds}{1+s}\right\}. \tag{54}$$

A lower bound on the tilted density over $0 \le w \le c\min\{\sigma_s, s^{-1}\}$ gives

$$-\log F_{D,x}(q_s) \le D\mathcal{C}_{D,x}(s) + C + \log(1 + s\sigma_s). \tag{55}$$

By (37),

$$\mathcal{C}_{D,x}(s) = \int_0^s t v_{D,x}(t)\, \mathrm{d}t \le \begin{cases} Cs^2, & s \le 1, \\ C\log(1+s), & s \ge 1. \end{cases}$$

Combining this with (54) proves (40) for the probabilities parameterized by negative saddles. At $s = 0$, uniform CDF convergence gives

$$F_{D,x}\{\mathbb{E}(S_x \mid x)\} = \tfrac{1}{2} + o(1).$$

If this value is below $1/2$, the uncovered probability interval has vanishing width. Its quantiles are $o(\sqrt{D})$ from the mean, so (36) at $s = 0$ gives density at least $cD^{-1/2}$, while the quantile is at least cD. Thus the elasticity is at least $c\sqrt{D}$, the order of $H_D(\log(1/u))$ there. This closes the central interval and proves the lemma. □

Lemma C.4 (Unconditional Gaussian ratio tails)**.** *Assume* (A1)–(A2)*, let* $m = n - 1$ *satisfy* $\log m \le KD$*, and fix* $1 \le k < \ell$*. For the* k*th and* ℓ*th neighbor radii of one anchor in the shared Gaussian cloud, there are constants* $z_0, c, C > 0$ *such that, for all sufficiently large* D *and all* $z \ge z_0$*,*

$$\mathbb{P}\left\{\log\frac{R_{i,\ell}}{R_{i,k}} > z\right\} \le C\{e^{-cDz} + e^{-cDe^z}\}. \tag{56}$$

Consequently, for every fixed $q < \infty$*,*

$$\sup_D \mathbb{E}\left(\log\frac{R_{i,\ell}}{R_{i,k}}\right)^q < \infty, \qquad \mathbb{E}\left\{d_* \log\frac{R_{i,\ell}}{R_{i,k}}\right\}^q \le C_{q,k,\ell} D^q \tag{57}$$

for every deterministic $0 < d_* \le D$*.*

Proof. Assumptions (A1)–(A2) give constants $\kappa, \lambda_0 > 0$ such that at least κD eigenvalues of Σ_D are at least λ_0. For independent $X, Y \sim \mathcal{N}(0, \Sigma_D)$,

$$X - Y \overset{\mathcal{L}}{=} \sqrt{2}\,\Sigma_D^{1/2} G, \qquad G \sim \mathcal{N}(0, I_D).$$

Hence standard chi-square Chernoff bounds, using the active block for the lower tail and $\|\Sigma_D\|_{\mathrm{op}} \le L$ for the upper tail, give, for any fixed $B > 0$ and all sufficiently large fixed z,

$$p_-(z) := \mathbb{P}\{\|X - Y\| < B\sqrt{D}e^{-z/2}\} \le e^{C_0 D - c_0 D z}, \tag{58}$$

$$p_+(z) := \mathbb{P}\{\|X - Y\| > B\sqrt{D}e^{z/2}\} \le e^{C_0 D - c_0 D e^z}. \tag{59}$$

Indeed, the first event implies $\chi^2_{\lfloor \kappa D \rfloor} \le CDe^{-z}$, whereas the second implies $\chi^2_D \ge cDe^z$.

Let $N_-(z)$ and $N_+(z)$ count candidates below and above the two radii in (58)–(59). No independence after averaging over the anchor is needed: linearity of expectation and the marginal difference law give

$$\mathbb{E}N_-(z) = mp_-(z), \qquad \mathbb{E}N_+(z) = mp_+(z).$$

Therefore Markov's inequality yields

$$\mathbb{P}\{R_{i,k} < B\sqrt{D}e^{-z/2}\} \le \frac{mp_-(z)}{k},$$

$$\mathbb{P}\{R_{i,\ell} > B\sqrt{D}e^{z/2}\} \le \frac{mp_+(z)}{m - \ell + 1} \le 2p_+(z)$$

for all large m. If neither event occurs, then $R_{i,\ell}/R_{i,k} \le e^z$. Since $m \le e^{KD}$, increasing z_0 absorbs the factor m in the first bound and proves (56). Integrating $qz^{q-1}\mathbb{P}\{\log(R_{i,\ell}/R_{i,k}) > z\}$ proves the first assertion in (57); the finitely many smaller D have finite moments because Gaussian interpoint distances have smooth densities. The second assertion follows from $d_* \le D$. □

Corollary C.5 (Bounded positive tilts and upper-radius control)**.** *Under (A1)–(A2), fix $0 < t_0 < 1/(4L)$. On $\mathcal{G}_D$, the positive tilts $0 \le t \le t_0$ lie strictly inside the MGF domain $t < 1/(2\lambda_{\max})$. The tilted means and variances of the squared radius S_x are both of order D, uniformly in x, t for all sufficiently large D. There exist $C, \varepsilon > 0$ such that*

$$\beta_x(u) \le \frac{C}{\sqrt{D}(1-u)\sqrt{1+T}}, \qquad T = -\log(1-u) \le \varepsilon D, \quad u \ge \tfrac{1}{2}. \tag{60}$$

Proof. Put $b_j = 1 - 2t\lambda_{D,j} \ge 1/2$. Under the tilt, the independent coordinates of $Y - x$ have means $-x_j/b_j$ and variances $\lambda_{D,j}/b_j$. Thus, writing $q_t = D\Lambda'_{D,x}(t)$ and $\sigma_t^2 = D\Lambda''_{D,x}(t)$,

$$q_t = \sum_j \left(\frac{\lambda_{D,j}}{b_j} + \frac{x_j^2}{b_j^2} \right) \asymp D, \qquad \sigma_t^2 = \sum_j \left(\frac{2\lambda_{D,j}^2}{b_j^2} + \frac{4\lambda_{D,j} x_j^2}{b_j^3} \right) \asymp D.$$

The active spectral block gives the lower bounds, and $\mathcal{G}_D$ and $b_j \ge 1/2$ give the upper bounds. The Fourier estimates of Lemma C.3, now with bounded denominators and $b_s = 1$, give a uniform normal density approximation, center lower bound, and supremum upper bound for the tilted density $g_{x,t}$. Inverse tilting therefore gives

$$\frac{1 - F_x(q_t)}{f_x(q_t)} = \int_0^\infty e^{-tw} \frac{g_{x,t}(q_t + w)}{g_{x,t}(q_t)}\, \mathrm{d}w \le \frac{C}{t + D^{-1/2}}.$$

Integrating the center lower bound over $0 \le w \le c\min(\sqrt{D}, t^{-1})$ also gives, with $I_x(t) = t\Lambda'_{D,x}(t) - \Lambda_{D,x}(t)$,

$$-\log\{1 - F_x(q_t)\} \le DI_x(t) + C + \log(1 + t\sqrt{D}) \le C(1 + Dt^2),$$

because $I_x(t) = \int_0^t v\Lambda''_{D,x}(v)\, \mathrm{d}v \asymp t^2$. At t_0, Chernoff's inequality gives a tail exponent at least cD; choose $\varepsilon < c$ to cover $T \le \varepsilon D$. Combining the last two bounds with $q_t \ge cD$ proves (60). If $F_x(q_0) > 1/2$, the remaining central interval has width $o(\sqrt{D})$ in radius squared and is covered by the untilted local limit. Integrating this envelope over a fixed order-statistic window yields (84); the unconditional ratio-tail lemma handles the exceptional region $T > \varepsilon D$ in the moment proof. □

Lemma C.6 (Moving-saddlepoint tail hazard)**.** *Under the assumptions of Theorem C.2, let $F_{D,x}$ and $f_{D,x}$ be the conditional CDF and density of S_x. Uniformly over every fixed multiplicative probability window $u = a/(n-1)$ with a in a compact subset of $(0,\infty)$, let $q_{D,x}(u) = F_{D,x}^{-1}(u)$. For a random Gaussian anchor,*

$$\frac{q_{D,x}(u)}{D} = y_D(s_{D,n}) + o_{\mathbb{P}}(1), \tag{61}$$

$$\frac{f_{D,x}(q_{D,x}(u))}{F_{D,x}(q_{D,x}(u))} = s_{D,n}\big(1 + o_{\mathbb{P}}(1)\big). \tag{62}$$

Consequently,

$$\beta_x(u) = \frac{1}{D\delta_D(s_{D,n})}\big(1 + o_{\mathbb{P}}(1)\big). \tag{63}$$

Moreover, for every fixed $1 \le k < \ell$, if $L_i^{(k,\ell)} = \log(R_{i,\ell}/R_{i,k})$, then for every fixed $p < \infty$,

$$\sup_D \mathbb{E}\left[\{D\delta_D(s_{D,n})L_i^{(k,\ell)}\}^p\right] < \infty. \tag{64}$$

Proof. Write $m = n-1$, $r_D = \log m$, $s_D = s_{D,n}$, and $x_j = \sqrt{\lambda_{D,j}}g_j$. Put $a_j(s) = 1 + 2s\lambda_{D,j}$. The conditional tilted mean, variance, and rate are

$$y_{D,x}(s) = \frac{1}{D}\sum_j \left\{\frac{\lambda_{D,j}}{a_j(s)} + \frac{x_j^2}{a_j(s)^2}\right\}, \tag{65}$$

$$v_{D,x}(s) = \frac{1}{D}\sum_j \left\{\frac{2\lambda_{D,j}^2}{a_j(s)^2} + \frac{4\lambda_{D,j}x_j^2}{a_j(s)^3}\right\}, \tag{66}$$

$$\mathcal{C}_{D,x}(s) = -sy_{D,x}(s) - \Lambda_{D,x}(-s), \qquad \mathcal{C}'_{D,x}(s) = sv_{D,x}(s). \tag{67}$$

The anchor perturbation has the exact first-order cancellation

$$\mathcal{C}_{D,x}(s) - \mathcal{C}_D(s) = \frac{2s^2}{D}\sum_j \frac{\lambda_{D,j}^2(g_j^2-1)}{a_j(s)^2}. \tag{68}$$

Also,

$$y_{D,x}(s) - y_D(s) = \frac{1}{D}\sum_j \frac{\lambda_{D,j}(g_j^2-1)}{a_j(s)^2}, \tag{69}$$

$$v_{D,x}(s) - v_D(s) = \frac{4}{D}\sum_j \frac{\lambda_{D,j}^2(g_j^2-1)}{a_j(s)^3}. \tag{70}$$

For each weighted centered sum $W_D(s) = D^{-1}\sum_j w_j(s)(g_j^2-1)$ appearing here (divide the rate error by s^2), the coefficients and their first derivatives are bounded on $[0,S]$. Independence and Cauchy–Schwarz give

$$\mathbb{E}\sup_{0\le s\le S}|W_D(s)| \le \mathbb{E}|W_D(0)| + \int_0^S \mathbb{E}|W_D'(t)|\,\mathrm{d}t \le C_S D^{-1/2}.$$

The rate-error weights extend continuously to zero. Markov's inequality therefore yields

$$\sup_s \frac{|\mathcal{C}_{D,x}(s) - \mathcal{C}_D(s)|}{s^2} = O_{\mathbb{P}}(D^{-1/2}), \tag{71}$$

$$\sup_s |y_{D,x}(s) - y_D(s)| + \sup_s |v_{D,x}(s) - v_D(s)| = O_{\mathbb{P}}(D^{-1/2}). \tag{72}$$

Assumptions (A1)–(A2) imply that some fixed fraction of the eigenvalues is bounded below by a positive constant. Hence $v_D(s)$ is bounded above and below on compact intervals and $\mathcal{C}_D(s) \asymp s^2$

there. Since r_D/D is bounded, s_D stays in one deterministic compact interval. If $r_D/D \to 0$, Lemma C.1 gives

$$s_D^2 = \frac{r_D}{3Dm_{2,D}}(1+o(1)), \qquad s_D\sqrt{D} \asymp \sqrt{r_D} \longrightarrow \infty. \tag{73}$$

We first locate the conditional quantile without applying a sharp estimate at its as-yet unknown saddle. Fix $\varepsilon \in (0,1/2)$ and put $s_D^{\pm} = (1\pm\varepsilon)s_D$. Lemma C.3 applies at these deterministic bounded saddles. In both the $c_{D,n} \to 0$ and positive-c regimes, $s_D^{\pm}\sqrt{Dv_{D,x}(s_D^{\pm})} \to \infty$. Equations (50), (36), and (39) therefore give, at $s = s_D^{\pm}$,

$$F_{D,x}\{Dy_{D,x}(s)\} = \frac{e^{-D\mathcal{C}_{D,x}(s)}}{s\sqrt{2\pi Dv_{D,x}(s)}}(1+o_{\mathbb{P}}(1)), \tag{74}$$

$$\frac{f_{D,x}\{Dy_{D,x}(s)\}}{F_{D,x}\{Dy_{D,x}(s)\}} = s(1+o_{\mathbb{P}}(1)). \tag{75}$$

The curve bounds used in (73) and the identity $\mathcal{C}_D'(s) = sv_D(s)$ imply

$$D\{\mathcal{C}_D(s_D) - \mathcal{C}_D(s_D^-)\} \geq c_\varepsilon r_D, \qquad D\{\mathcal{C}_D(s_D^+) - \mathcal{C}_D(s_D)\} \geq c_\varepsilon r_D \tag{76}$$

for some $c_\varepsilon > 0$. When $s_D \to 0$, both differences are fixed multiples of $Ds_D^2 \asymp r_D$; when s_D is bounded away from zero, they are fixed multiples of $D \asymp r_D$. Every sequence has subsequences in one of these two cases, so the same bounds cover oscillating regimes. Moreover, (71) gives

$$D|\mathcal{C}_{D,x}(s_D^{\pm}) - \mathcal{C}_D(s_D^{\pm})| = O_{\mathbb{P}}(\sqrt{D}\,(s_D^{\pm})^2) = o_{\mathbb{P}}(r_D),$$

and the logarithmic prefactor in (74) is $O(\log(1+r_D)) = o(r_D)$. Hence, uniformly for a in the stated compact set, with probability tending to one,

$$F_{D,x}\{Dy_{D,x}(s_D^+)\} < \frac{a}{m} < F_{D,x}\{Dy_{D,x}(s_D^-)\}.$$

Since $y_{D,x}'(s) = -v_{D,x}(s) < 0$, there is a unique $\widehat{s}_{D,x,a} \in (s_D^-, s_D^+)$ satisfying $q_{D,x}(a/m) = Dy_{D,x}(\widehat{s}_{D,x,a})$. Letting $\varepsilon \downarrow 0$ along a countable sequence proves

$$\widehat{s}_{D,x,a}/s_D \xrightarrow{\mathbb{P}} 1. \tag{77}$$

In particular, the random saddle stays in a deterministic compact interval and

$$\widehat{s}_{D,x,a}\sqrt{Dv_{D,x}(\widehat{s}_{D,x,a})} \longrightarrow \infty \quad \text{in probability.}$$

The uniformity in Lemma C.3 now permits (74)–(75) to be evaluated at this random saddle. Taking logarithms gives, uniformly for a in the stated compact set,

$$D\mathcal{C}_{D,x}(\widehat{s}_{D,x,a}) + \log\left[\widehat{s}_{D,x,a}\sqrt{2\pi Dv_{D,x}(\widehat{s}_{D,x,a})}\right] = r_D - \log a + o_{\mathbb{P}}(1). \tag{78}$$

Equations (71) and (73), together with $\mathcal{C}_{D,x}'(s) \asymp s$, imply

$$\frac{\widehat{s}_{D,x,a}}{s_D} = 1 + O_{\mathbb{P}}\left(D^{-1/2} + \frac{\log(1+r_D)}{r_D}\right). \tag{79}$$

Combining this with (72), (75), and (21) proves (61)–(63).

It remains to justify moments. This cannot be inferred from convergence on fixed probability windows, so we record the global bound. Define

$$H_D(r) = \min\{D, \sqrt{D(1+r)}\}, \qquad d_* = D\delta_D(s_D).$$

The spectral assumptions and $r_D \leq KD$ give

$$d_* \asymp H_D(r_D). \tag{80}$$

Use the deterministic anchor event $\mathcal{G}_D$ from (41). For every $q < \infty$,

$$\mathbb{P}(\mathcal{G}_D^c) \le e^{-cD} + C_q D^{-q/12}. \tag{81}$$

Lemma C.3 already proves the explicit lower-tail slope envelope

$$\beta_x(u) \le \frac{C}{H_D(\log(1/u))}, \qquad 0 < u \le \tfrac{1}{2}, \qquad x \in \mathcal{G}_D. \tag{82}$$

For $u \ge 1/2$, Corollary C.5 bounds the complementary inverse-CDF slope.

Let $U_{(k)}, U_{(\ell)}$ be fixed order statistics of m uniforms. Integrating (82), using (80), gives below $1/2$

$$d_* L_i^{(k,\ell)} \le C_{k,\ell} \left[\{-\log(mU_{(k)})\}_+ + \{\log(mU_{(\ell)})\}_+ \right]. \tag{83}$$

Indeed the integrand is at most one below $1/m$, while above $1/m$, with $v = \log(1/u)$,

$$\int_v^{r_D} \sqrt{\frac{1+r_D}{1+w}}\, \mathrm{d}w \le 2(r_D - v).$$

The part above $1/2$, on $T_\ell = -\log(1 - U_{(\ell)}) \le \varepsilon D$, is at most

$$C_{k,\ell} \sqrt{r_D T_\ell}\, \mathbf{1}\{U_{(\ell)} > 1/2\}. \tag{84}$$

The order-statistic tails satisfy

$$\begin{aligned} &\mathbb{P}\{-\log(mU_{(k)}) > t\} \le C_k e^{-kt}, \\ &\sup_m \mathbb{E}[\{\log(mU_{(\ell)})\}_+^p] < \infty, \\ &\mathbb{P}(T_\ell > t) \le C_\ell m^{\ell-1} e^{-(m-\ell+1)t}. \end{aligned}$$

Thus (83)–(84) have uniformly bounded fixed moments on $\mathcal{G}_D$ outside the final event $T_\ell > \varepsilon D$.

It remains to remove both exceptional events without using a conditional moment estimate. Fix $p < q < \infty$. Lemma C.4, which was proved without restricting the anchor, gives

$$\mathbb{E}(d_* L_i^{(k,\ell)})^q \le C_{q,k,\ell} D^q.$$

Consequently, for any event A, Hölder's inequality gives

$$\mathbb{E}\left[(d_* L_i^{(k,\ell)})^p \mathbf{1}_A \right] \le C D^p \mathbb{P}(A)^{1-p/q}. \tag{85}$$

For $A = \mathcal{G}_D^c$, choose the polynomial order in (81) sufficiently large; the right side of (85) then tends to zero. For $A = \{T_\ell > \varepsilon D\}$, the order-statistic tail above implies

$$\mathbb{P}(A) \le C_\ell m^{\ell-1} e^{-(m-\ell+1)\varepsilon D},$$

so the same conclusion holds. On the complement of these two events, (83)–(84) have uniformly bounded pth moments. This proves (64). □

Proposition C.7 (One-anchor ratio law)**.** *For a uniformly selected anchor in the shared Gaussian cloud,*

$$d_{\mathrm{eff}} L_i \xrightarrow{\mathcal{L}} \mathrm{Exp}(1)$$

with convergence of the first two moments.

Proof. For $m = n - 1$ uniforms, $mU_{(2)}$ is tight and $U_{(1)}/U_{(2)}$ is uniform independently of $U_{(2)}$. Lemma C.6 makes the inverse-CDF slope uniformly $d_{\mathrm{eff}}^{-1}(1 + o_{\mathbb{P}}(1))$ across every fixed random multiplicative window containing $[U_{(1)}, U_{(2)}]$. Lemma B.1 gives

$$d_{\mathrm{eff}} L_i = \log \frac{U_{(2)}}{U_{(1)}} \big(1 + o_{\mathbb{P}}(1)\big).$$

The right-hand logarithm is exactly unit exponential. Truncate its exponential tail, apply the uniform approximation on the truncated event, and then remove the truncation using the envelope in Lemma C.6. This also proves moment convergence. □

C.4 TWO-ANCHOR SEPARATION

The one-anchor law does not by itself control the reused cloud: one candidate can be close to several anchors. A classical Gaussian Bhattacharyya-affinity identity reduces the joint lower-tail bound to a quadratic form in the two anchors (Abou-Moustafa & Ferrie, 2012). The argument combines Gaussian integration with the moving-quantile estimates in Lemma C.6.

For an anchor x and an independent $Y \sim \mathcal{N}(0, \Sigma_D)$, put $S_x = \|Y - x\|^2$, $M_x(s) = \mathbb{E}e^{-sS_x}$, and $C_s = \Sigma_D(I + 2s\Sigma_D)^{-1}$. For $s > 0$, the tilted law with density proportional to e^{-sS_x} relative to the law of Y is Gaussian with covariance C_s and mean $2sC_sx$. The equal-covariance Gaussian affinity therefore gives the identity

$$\mathbb{E}e^{-s(S_x+S_z)/2} = \sqrt{M_x(s)M_z(s)}\,\exp\left\{-\frac{s^2}{2}(x-z)^\top C_s(x-z)\right\}. \tag{86}$$

If Σ_D is singular, the densities and Gaussian integration are taken on its positive-eigenvalue support; the displayed matrix identity is unchanged. Equivalently, (86) follows by completing the square separately in each nonzero covariance direction.

Lemma C.8 (Separation from Gaussian affinity)**.** *Let X, Z, Y be independent $\mathcal{N}(0, \Sigma_D)$ variables, and put $U = F_{D,X}(\|Y - X\|^2)$ and $V = F_{D,Z}(\|Y - Z\|^2)$. Under the assumptions of Theorem C.2, for every fixed $a, b > 0$,*

$$n\mathbb{P}\left(U \le \frac{a}{n},\ V \le \frac{b}{n} \mid X, Z\right) \xrightarrow{\mathbb{P}} 0. \tag{87}$$

More precisely, if $\log(n-1) \le KD$, there is a constant $\eta_K > 0$ such that the conditional probability is at most $n^{-1-\eta_K+o_{\mathbb{P}}(1)}$. When $\log n = o(D)$, the same argument gives the sufficient upper bound $n^{-4/3+o_{\mathbb{P}}(1)}$.

Proof. Write $m = n - 1$, $r = \log m$, and $s = s_{D,n}$. We first use thresholds $a/m, b/m$; the change to $a/n, b/n$ is covered by the compact-window uniformity in Lemma C.6. Let $q_x = F_{D,x}^{-1}(a/m)$, $q_z = F_{D,z}^{-1}(b/m)$, and define

$$J_x(t; q_x) = -tq_x - \log M_x(t), \qquad G_{x,z}(s) = \frac{s^2}{2}(x-z)^\top C_s(x-z).$$

Chernoff's inequality and (86) give

$$\mathbb{P}(S_x \le q_x, S_z \le q_z \mid x, z) \le \exp\left\{-\frac{J_x(s; q_x) + J_z(s; q_z)}{2} - G_{x,z}(s)\right\}. \tag{88}$$

We justify the use of the same deterministic saddle s in the two marginal objectives. The active spectral block in (A1)–(A2) and $r/D \le K$ imply $s \le S$ for a fixed $S < \infty$, and $Ds^2 \asymp r$. Let $\widehat{s}_x = \widehat{s}_{D,x,a}$ be the conditional saddle from (77). On events of probability tending to one, $\widehat{s}_x/s = 1 + o(1)$, and (78) gives

$$J_x(\widehat{s}_x; q_x) = D\mathcal{C}_{D,x}(\widehat{s}_x) = r + O_{\mathbb{P}}(\log(1 + r)) = r + o_{\mathbb{P}}(r).$$

Since $J_x'(\widehat{s}_x; q_x) = 0$ and $J_x''(t; q_x) = -Dv_{D,x}(t)$, the compact-interval variance bound in (72) and Taylor's theorem give

$$0 \le J_x(\widehat{s}_x; q_x) - J_x(s; q_x) \le O_{\mathbb{P}}(D)(\widehat{s}_x - s)^2 = o_{\mathbb{P}}(Ds^2) = o_{\mathbb{P}}(r).$$

Thus $J_x(s; q_x) = r + o_{\mathbb{P}}(r)$, and the same statement holds for z. This step depends only on the marginal lemma, without using any joint-tail or concentration conclusion.

For independent Gaussian anchors, write $X_j - Z_j = \sqrt{2\lambda_{D,j}}\,W_j$ with iid standard normal W_j on the positive support. Then

$$G_{X,Z}(s) = s^2 \sum_j \frac{\lambda_{D,j}^2}{1 + 2s\lambda_{D,j}} W_j^2.$$

The sum has mean $A_D(s) = \sum_j \lambda_{D,j}^2/(1 + 2s\lambda_{D,j}) \asymp D$ and variance at most $2DL^4$. Hence

$$G_{X,Z}(s) = s^2 A_D(s)\{1 + o_{\mathbb{P}}(1)\}. \tag{89}$$

Finally, $\mathcal{C}'_D(t) = tv_D(t)$ and $v_D(t) \le 6m_{2,D}$ give

$$r = D\mathcal{C}_D(s) \le 3Dm_{2,D}s^2, \qquad A_D(s) \ge \frac{Dm_{2,D}}{1+2SL}.$$

Therefore

$$\frac{G_{X,Z}(s)}{r} \ge \frac{1+o_{\mathbb{P}}(1)}{3(1+2SL)}.$$

Substituting in (88) proves the asserted bound, for example with $\eta_K = \{3(1+2SL)\}^{-1}$, and consequently (87). In the shallow regime, $s \to 0$ and (24) gives $r = 3Dm_{2,D}s^2\{1+o(1)\}$, whereas $A_D(s) = Dm_{2,D}\{1+o(1)\}$. Thus $G_{X,Z}(s)/r \to 1/3$ in probability, giving the stated $4/3$ upper-bound exponent. No matching lower bound or optimal joint exponent is asserted or needed for the shared-cloud law. □

Lemma C.9 (Independent limiting extreme processes)**.** *Let (U_j, V_j), $j = 1, \dots, m$, be independent pairs with uniform marginals. If, for every fixed $a, b > 0$,*

$$m\mathbb{P}(U_1 \le a/m, V_1 \le b/m) \to 0,$$

then the point processes $\sum_j \delta_{mU_j}$ and $\sum_j \delta_{mV_j}$ converge jointly to independent unit-rate Poisson point processes on $(0, \infty)$.

Proof. For bounded disjoint intervals on the two axes, the vector of cell counts is multinomial. Each marginal cell has probability equal to its length divided by m, up to $o(1/m)$, while the probability that one pair contributes to cells on both axes is $o(1/m)$. The multinomial Laplace transform converges to the product of the corresponding Poisson Laplace transforms. Tightness follows from the unit-rate marginal means. □

Lemma C.8 verifies Lemma C.9 conditional on two typical anchors. The mutual anchor-to-anchor distance itself lies above the lower-tail neighbor radius with probability tending to one, so deleting that one candidate does not change the limiting first two points. Hence the two normalized ratios converge to independent exponentials. For a fixed number r of anchors, intersect the high-probability typical-pair events over the $\binom{r}{2}$ pairs. Lemma C.8 then makes the probability that one candidate enters extreme cells for any two anchors $o(1/m)$, uniformly over this finite collection. The multinomial Laplace-transform argument in Lemma C.9, now with r axes, yields r independent unit-rate Poisson processes. Finally, each anchor has only $r-1$ mutual-anchor candidates, and each corresponding conditional PIT is uniform; a union bound shows that none enters a fixed $O(1/m)$ extreme window with probability tending to one. This proves the fixed-r assertion in Theorem C.2.

C.5 Concentration of the shared-sample average

Lemma C.10 (Exchangeable pair criterion)**.** *Let $(Z_{n,1}, \dots, Z_{n,n})$ be an exchangeable triangular array. If*

$$\mathbb{E}Z_{n,1} \to \mu, \qquad \sup_n \mathbb{E}Z_{n,1}^2 < \infty, \qquad \mathbb{E}[Z_{n,1}Z_{n,2}] \to \mu^2,$$

then $n^{-1}\sum_i Z_{n,i} \to \mu$ in L^2.

Proof. Exchangeability gives

$$\mathrm{Var}\left(\frac{1}{n}\sum_i Z_{n,i}\right) = \frac{\mathrm{Var}(Z_{n,1})}{n} + \frac{n-1}{n}\,\mathrm{Cov}(Z_{n,1}, Z_{n,2}),$$

and both terms tend to zero. □

Completion of the proof of Theorem C.2. Take $Z_{D,n,i} = d_{\mathrm{eff}}L_i$. Proposition C.7, Lemma C.8, and Lemma C.9 give pair convergence in distribution to independent unit exponentials. The $p = 4$ case of (64) and Cauchy–Schwarz imply that $\{Z_{D,n,1}Z_{D,n,2}\}_{D,n}$ is uniformly integrable. Hence pair convergence transfers to the cross moment, and

$$\mathbb{E}Z_{D,n,1} \to 1, \qquad \mathbb{E}Z_{D,n,1}^2 \to 2, \qquad \mathbb{E}[Z_{D,n,1}Z_{D,n,2}] \to 1.$$

Lemma C.10 yields

$$\frac{1}{n}\sum_i Z_{D,n,i} \xrightarrow{L^2} 1.$$

Finally,

$$\frac{\widehat{d}_{\mathrm{2NN}}}{d_{\mathrm{eff}}} = \frac{(n-1)/n}{n^{-1}\sum_i Z_{D,n,i}} \xrightarrow{\mathbb{P}} 1.$$

The fixed-anchor joint statement was established above. □

D Spectral consequences and neighbor-order extensions

D.1 Spectral contributions to the finite-radius target

Suppose the empirical spectral measures

$$H_D = \frac{1}{D}\sum_{j=1}^{D} \delta_{\lambda_{D,j}}$$

converge weakly to a compactly supported law H with positive first two moments. Define

$$\mathcal{C}_H(s) = \frac{1}{2}\int \left[\log(1+2s\lambda) - \frac{2s\lambda}{(1+2s\lambda)^2}\right] \mathrm{d}H(\lambda), \tag{90}$$

$$\delta_H(s) = \int \left[1-(1+2s\lambda)^{-2}\right] \mathrm{d}H(\lambda). \tag{91}$$

For $c > 0$, let s_c be the unique solution of $\mathcal{C}_H(s_c) = c$ and put

$$\Delta_H(c) = \delta_H(s_c).$$

Corollary D.1 (Exponential-resolution spectral phase diagram)**.** *If* $\log n/D \to c \in (0,\infty)$*, then*

$$\frac{\widehat{d}_{\mathrm{2NN}}}{D} \xrightarrow{\mathbb{P}} \Delta_H(c).$$

An eigenvalue λ *contributes*

$$w_{s_c}(\lambda) = 1-(1+2s_c\lambda)^{-2}$$

to the normalized effective dimension. Thus directions with $s_c\lambda \gg 1$ *count almost fully, while directions with* $s_c\lambda \ll 1$ *contribute* $4s_c\lambda + O((s_c\lambda)^2)$.

Thus the estimate changes continuously with both the covariance spectrum and the available sample resolution. It is a continuous spectral weighting of resolved directions, not a binary count of nonzero eigenvalues.

If H has an atom at zero, then

$$\lim_{c\to\infty} \Delta_H(c) = 1 - H(\{0\}).$$

This also exposes a noncommuting limit: a sequence of technically full-rank covariances may have eigenvalues tending to zero so quickly that no fixed $\log n/D$ ratio resolves them.

D.2 Participation ratio at coarse sample resolution

Define

$$r_2(\Sigma_D) = \frac{(\operatorname{tr}\Sigma_D)^2}{\operatorname{tr}(\Sigma_D^2)}.$$

After normalizing the eigenvalues to sum to one, this is the exponential of their Rényi-2 entropy.

Corollary D.2 (Participation-ratio law)**.** *If* $\log n \to \infty$ *and* $\log n = o(D)$*, then*

$$\widehat{d}_{\mathrm{2NN}} = \frac{4}{\sqrt{3}}\sqrt{r_2(\Sigma_D)\log n}\,\big(1+o_{\mathbb{P}}(1)\big). \tag{92}$$

Proof. Equations (24)–(25) give

$$s_{D,n} = \sqrt{\frac{\log n}{3Dm_{2,D}}}\,(1 + o(1))$$

and

$$D\delta_D(s_{D,n}) = \frac{4}{\sqrt{3}}\frac{Dm_{1,D}}{\sqrt{Dm_{2,D}}}\sqrt{\log n}\,(1 + o(1)).$$

The ratio before $\sqrt{\log n}$ is $\sqrt{r_2(\Sigma_D)}$. Apply Theorem C.2. □

Corollary D.2 does not identify the participation ratio with true intrinsic dimension. Within the Gaussian model, participation ratio is the leading spectral quantity controlling TwoNN before the sample resolves the full phase curve.

D.3 Isotropic sample-complexity threshold

For $\Sigma_D = I_D$, put $a = 1 + 2s > 1$. The phase curve becomes

$$c(a) = \frac{1}{2}\left(\log a - \frac{1}{a} + \frac{1}{a^2}\right), \tag{93}$$

$$\Delta_{\mathrm{iso}}(c(a)) = 1 - \frac{1}{a^2}. \tag{94}$$

Corollary D.3 (Matching estimator-specific resolution boundary)**.** *For a target fraction $\alpha \in (0, 1)$, define*

$$c_*(\alpha) = \frac{1}{2}\left[\log\frac{1}{\sqrt{1-\alpha}} - \sqrt{1-\alpha} + (1-\alpha)\right]. \tag{95}$$

Then:

(i) if $\log n/D \to c < c_*(\alpha)$, then $\widehat{d}_{2\mathrm{NN}}/D < \alpha + o_{\mathbb{P}}(1)$;

(ii) if $\log n/D \to c > c_*(\alpha)$, then $\widehat{d}_{2\mathrm{NN}}/D > \alpha - o_{\mathbb{P}}(1)$;

(iii) at $c = c_*(\alpha)$, $\widehat{d}_{2\mathrm{NN}}/D \to \alpha$ in probability.

Moreover,

$$c_*(\alpha) = \frac{3}{16}\alpha^2 + O(\alpha^3) \quad \text{as } \alpha \downarrow 0, \qquad c_*(\alpha) \to \infty \quad \text{as } \alpha \uparrow 1.$$

The lower and upper statements form a matching boundary for pooled TwoNN on this model. They do not constitute an information-theoretic lower bound for arbitrary dimension estimators. In particular, estimating a fixed positive fraction of D requires $n \asymp \exp\{c_*(\alpha)D\}$ samples. Polynomial and other subexponential sample sizes therefore cannot recover any fixed positive fraction of D in this regime.

D.4 Resolution at general neighbor orders

Let ψ denote the digamma function.

Theorem D.4 (Fixed neighbor-order ratio law)**.** *Under the assumptions of Theorem C.2, fix $1 \le k < \ell$. For every fixed collection of anchors,*

$$d_{\mathrm{eff}} \log\frac{R_{i,\ell}}{R_{i,k}} \xrightarrow{\mathcal{L}} -\log B_{k,\ell-k},$$

jointly to independent limits, where $B_{k,\ell-k} \sim \mathrm{Beta}(k, \ell - k)$. Consequently,

$$\widehat{d}_{k,\ell} = \frac{\psi(\ell) - \psi(k)}{n^{-1}\sum_i \log(R_{i,\ell}/R_{i,k})} \tag{96}$$

satisfies $\widehat{d}_{k,\ell}/d_{\mathrm{eff}} \to 1$ in probability.

Proof. Lemma B.1 gives the beta law, and Lemma C.6 makes the inverse-CDF slope asymptotically constant across the fixed random probability window. Since $\mathbb{E}[-\log B_{k,\ell-k}] = \psi(\ell) - \psi(k)$, the shared-cloud conclusion follows from the same two-anchor and exchangeability argument as Theorem C.2: the fixed-(k,ℓ) case of (64) supplies fourth-moment uniform integrability, so joint weak convergence transfers to the one- and two-anchor moments needed by Lemma C.10. □

The constant-slope statement is the asymptotic center for every fixed neighbor order, but it discards a visible pre-asymptotic effect: later neighbors probe probabilities near k/n rather than $1/n$. Retaining this phase variation produces a neighbor-window approximation with no fitted parameters. Let

$$\pi_{k,\ell}(a) = \mathbb{P}\{k \le N_a \le \ell - 1\} = e^{-a} \sum_{j=k}^{\ell-1} \frac{a^j}{j!}, \qquad N_a \sim \mathrm{Poisson}(a).$$

For $0 < a < m/2$, let $s_D(a)$ solve

$$\mathcal{C}_D\{s_D(a)\} = \frac{\log(m/a)}{D}, \qquad d_D(a) = D\delta_D\{s_D(a)\},$$

and define

$$M_{k,\ell}(D,m) = \int_0^{m/2} \frac{\pi_{k,\ell}(a)}{a\, d_D(a)}\, \mathrm{d}a, \qquad d^{\mathrm{rw}}_{k,\ell} = \frac{\psi(\ell) - \psi(k)}{M_{k,\ell}(D,m)}. \tag{97}$$

The cutoff $a < m/2$ keeps $\log(m/a) \ge \log 2 > 0$, so the positive saddle is defined and the lower-tail bounds apply uniformly. The exact-order-statistic contribution above $m/2$ is exponentially negligible after the upper-radius control in the proof below; the Poisson coverage weight also has an exponential tail.

Lemma D.5 (Integrable neighbor-window envelope)**.** *Fix integers $1 \le k < \ell$ and $m \ge \ell$, and let $w_m(a) = \mathbb{P}\{k \le \mathrm{Binomial}(m, a/m) \le \ell - 1\}$. For $q > 0$, put $H_q(r) = \min\{q, \sqrt{q(1+r)}\}$ and suppose the positive phase scales $d(a)$ and d_{base} are comparable to $H_q(\log(m/a))$ and $H_q(\log m)$, respectively, for $0 < a < m/2$, with comparability constants independent of m, q, a. Then both $w_m(a)d_{\mathrm{base}}/\{ad(a)\}$ and its Poisson analogue are bounded by the integrable function below. Its constant may also depend on the fixed comparability and slope-envelope constants:*

$$g_{k,\ell}(a) = C_{k,\ell}\left[a^{k-1}\mathbf{1}_{(0,1]}(a) + a^{\ell-2}e^{-a/2}\sqrt{1 + \max(0, \log a)}\,\mathbf{1}_{(1,\infty)}(a)\right].$$

The same bound holds for $w_m(a)d_{\mathrm{base}}\beta_x(a/m)/a$ whenever $\beta_x(a/m) \le C/H_q(\log(m/a))$.

Proof. Standard binomial bounds, and directly the Poisson series, give

$$w_m(a) \le \begin{cases} C_{k,\ell}a^k, & 0 < a \le 1, \\ C_{k,\ell}a^{\ell-1}e^{-a/2}, & 1 < a \le m/2. \end{cases} \tag{98}$$

For $a \le 1$, monotonicity of H_q bounds its ratio by one. For $1 < a \le \sqrt{m}$, that ratio is bounded by $\sqrt{2}$. For $\sqrt{m} < a < m/2$, it is at most $C\sqrt{1 + \log m}$, and $\log m < 2\log a$ absorbs this factor into the displayed envelope. Integrability follows from $k \ge 1$ at zero and the exponential tail at infinity. The statement concerns $a < m/2$ only; upper-radius contributions must be bounded separately. □

Proposition D.6 (Poisson neighbor-window approximation)**.** *Under the assumptions of Theorem C.2, for fixed $1 \le k < \ell$,*

$$\frac{n^{-1}\sum_i \log(R_{i,\ell}/R_{i,k})}{M_{k,\ell}(D, n-1)} \xrightarrow{\mathbb{P}} 1, \qquad \frac{\widehat{d}_{k,\ell}}{d^{\mathrm{rw}}_{k,\ell}} \xrightarrow{\mathbb{P}} 1.$$

Moreover $d^{\mathrm{rw}}_{k,\ell}/d_{\mathrm{eff}} \to 1$, so this changes no first-order claim; it retains the phase movement across the finite neighbor window. It is a parameter-free approximation. This proposition establishes no improvement in bias, mean squared error, or asymptotic error order.

Proof. For m uniforms and $0 < u < 1$, the exact probability that u lies between $U_{(k)}$ and $U_{(\ell)}$ is

$$\mathbb{P}\{k \le \mathrm{Binomial}(m, u) \le \ell - 1\}.$$

Writing

$$w_m(a) = \mathbb{P}\{k \le \mathrm{Binomial}(m, a/m) \le \ell - 1\},$$

Tonelli's theorem and Lemma B.1 give the conditional mean representation

$$\mathbb{E}\left[\log \frac{R_{i,\ell}}{R_{i,k}} \middle| X_i = x\right] = \int_0^m \frac{w_m(a)}{a}\, \beta_x(a/m)\, \mathrm{d}a. \tag{99}$$

On every compact interval $a \in [A^{-1}, A]$, the binomial-to-Poisson limit is uniform, $w_m(a) = \pi_{k,\ell}(a) + o(1)$, and the moving-saddle proof, with $\log(m/a)$ in place of $\log m$, gives

$$\sup_{a\in[A^{-1},A]} |d_D(a)\beta_X(a/m) - 1| \xrightarrow{\mathbb{P}} 0 \tag{100}$$

for a random anchor X. On $\mathcal{G}_D$, Equation (82) bounds the conditional slope by the deterministic envelope used below. Convergence to the typical-anchor center is in probability, rather than uniform over all anchors in $\mathcal{G}_D$.

We next remove the compact truncation using Lemma D.5. The curve bounds give $d_D(a) \asymp H_D(\log(m/a))$ and $d_{\mathrm{eff}} \asymp H_D(\log m)$; the actual conditional slope on $\mathcal{G}_D$ obeys (82). The lemma therefore dominates both normalized integrands by $g_{k,\ell}$ for $a < m/2$. The contributions of $a < A^{-1}$ and $A < a < m/2$ vanish after normalization as $A \to \infty$, uniformly along the asymptotic sequence. The positive-tail envelope (60), followed by Lemma C.4 on the exceptional anchor and radius events, handles $a > m/2$. On each compact window, the normalized integrals restricted to $\mathcal{G}_D$ are uniformly bounded, so convergence in probability implies convergence of expectations. The contribution of $\mathcal{G}_D^c$ is $o(1)$ after normalization by (85). Equations (99)–(100) therefore give

$$\mathbb{E}\log(R_{i,\ell}/R_{i,k}) = M_{k,\ell}(D, m)\{1 + o(1)\}.$$

For the shared average, first retain only the contribution of $a \in [A^{-1}, A]$ in the inverse-CDF integrals for two anchors. Lemmas C.8–C.9, uniformly on this compact window, make the two truncated contributions asymptotically independent. The preceding tail estimates and the fourth-moment case of (64) make the discarded products uniformly negligible as $A \to \infty$. Hence the cross moment factorizes, and Lemma C.10 proves the first convergence in the proposition; the second follows algebraically from (96).

Finally, $s_D(a)/s_D \to 1$ and $d_D(a)/d_{\mathrm{eff}} \to 1$ on every fixed a-window. Since

$$\int_0^\infty \frac{\pi_{k,\ell}(a)}{a}\, \mathrm{d}a = \psi(\ell) - \psi(k),$$

the same uniform truncation proves $d_{\mathrm{eff}} M_{k,\ell}(D, m) \to \psi(\ell) - \psi(k)$ and hence the last assertion. □

Proposition D.7 (Growing-order resolution law)**.** *Suppose, in addition to the assumptions of Theorem C.2, that $H_D \Rightarrow H$ as at the start of this section. If $k_D \to \infty$, $\ell_D/k_D \to a > 1$, $\ell_D = o(n)$, and*

$$\frac{\log(n/k_D)}{D} \to c \in (0, \infty),$$

then, for every fixed collection of anchors,

$$D\Delta_H(c) \log \frac{R_{i,\ell_D}}{R_{i,k_D}} \xrightarrow{\mathbb{P}} \log a. \tag{101}$$

Proof. Conditional on an anchor, the probability integral transforms are iid uniform. The beta mean and variance, or the exponential-spacing representation, therefore give

$$\frac{mU_{(k_D)}}{k_D} \xrightarrow{\mathbb{P}} 1, \qquad \frac{mU_{(\ell_D)}}{\ell_D} \xrightarrow{\mathbb{P}} 1.$$

With probability tending to one, every u between these two order statistics satisfies

$$\frac{\log(1/u)}{D} = \frac{\log(n/k_D)}{D} + o(1) \longrightarrow c$$

uniformly across the interval. Repeat the saddle bracketing in Lemma C.6 with $\log(1/u)$ in place of $\log m$. Because the limiting saddle s_c is bounded away from zero, the uniform tilted local limit applies directly and gives

$$\sup_{u\in[U_{(k_D)},U_{(\ell_D)}]} |D\Delta_H(c)\beta_{X_i}(u) - 1| \xrightarrow{\mathbb{P}} 0.$$

Lemma B.1 now yields

$$D\Delta_H(c)\log\frac{R_{i,\ell_D}}{R_{i,k_D}} = \{1 + o_{\mathbb{P}}(1)\}\log\frac{U_{(\ell_D)}}{U_{(k_D)}} \xrightarrow{\mathbb{P}} \log a.$$

Intersecting the anchor-typical events proves the same statement for any fixed collection. □

Thus varying the neighbor order probes the same phase curve at resolution $\log(n/k)/D$. Under the Gaussian model, a neighbor-order sweep traces the phase curve at several sample resolutions. This is marginal convergence and joint convergence for any fixed number of anchors. It does not assert concentration of the all-anchor pooled estimator for growing orders: the fixed-extreme overlap bound does not control candidate reuse in a growing neighbor window.

D.5 Estimating the shallow Gaussian target from the same cloud

Estimating a spectral target and estimating the covariance matrix are different statistical tasks. Classical covariance tests estimate trace moments directly by averaging over distinct sample indices (Chen et al., 2010, Section 2); Chun et al. (2026, Section 4) apply the same principle to participation-ratio estimation. The following proposition applies these established moment estimators to the shallow Gaussian TwoNN target. It does not introduce a new covariance estimator. Disjoint pairing removes the unknown mean and yields simple independent-difference variance bounds, at a possible cost in statistical efficiency compared with using all distinct tuples. The predictor targets the shallow participation-ratio law; positive limiting $\log n/D$ requires the full nonlinear spectral equation.

Proposition D.8 (Same-cloud estimation of the shallow Gaussian target)**.** *Let $X_1,\dots,X_n$ be iid $\mathcal{N}(\mu,\Sigma)$, where $n \geq 4$, $\Sigma \neq 0$, and $T_j = \mathrm{tr}(\Sigma^j)$. Put $N = \lfloor n/2 \rfloor$, $r = \log(n-1)$, and*

$$Z_i = \frac{X_{2i} - X_{2i-1}}{\sqrt{2}}, \qquad A_N = \frac{1}{N}\sum_{i=1}^{N}\|Z_i\|^2, \qquad B_N = \frac{1}{N(N-1)}\sum_{i\neq j}(Z_i^\top Z_j)^2.$$

Define

$$d_U = \frac{4}{\sqrt{3}}A_N\sqrt{\frac{r}{B_N}}, \qquad d_{\mathrm{PR}} = \frac{4}{\sqrt{3}}T_1\sqrt{\frac{r}{T_2}}.$$

Then $B_N > 0$ almost surely, $\mathbb{E}A_N = T_1$, $\mathbb{E}B_N = T_2$, and

$$\mathrm{Var}(A_N) = \frac{2T_2}{N}, \tag{102}$$

$$\mathrm{Var}(B_N) = \frac{8T_4}{N} + \frac{4(T_2^2 + T_4)}{N(N-1)}. \tag{103}$$

For every $0 < \varepsilon < 1$, with probability at least $1 - V_N/\varepsilon^2$,

$$\frac{1-\varepsilon}{\sqrt{1+\varepsilon}} \leq \frac{d_U}{d_{\mathrm{PR}}} \leq \frac{1+\varepsilon}{\sqrt{1-\varepsilon}}, \qquad V_N = \frac{2T_2}{NT_1^2} + \frac{8T_4}{NT_2^2} + \frac{4}{N(N-1)}\left(1 + \frac{T_4}{T_2^2}\right). \tag{104}$$

In particular, $V_N \leq 10/N + 8/[N(N-1)]$. Under the paper's spectral assumptions (A1)–(A2), $V_N = O((ND)^{-1} + N^{-2})$. If also $\log n \to \infty$ and $\log n = o(D)$, then

$$\frac{d_U}{d_{\mathrm{eff}}(n,\Sigma_D)} \xrightarrow{\mathbb{P}} 1, \qquad \frac{\widehat{d}_{\mathrm{2NN}}}{d_U} \xrightarrow{\mathbb{P}} 1. \tag{105}$$

No lower bound on n/D and no independent covariance-estimation sample are required.

Proof. The disjoint differences are iid $\mathcal{N}(0,\Sigma)$, including when μ is unknown. The first two moments of $\|Z\|^2$ give (102). For $h(z,w)=(z^\top w)^2$,

$$\mathbb{E}h(Z_1,Z_2)=T_2,\quad \mathbb{E}[h(Z_1,Z_2)\mid Z_1]=Z_1^\top\Sigma Z_1,\quad \mathrm{Var}(Z_1^\top\Sigma Z_1)=2T_4.$$

Conditional Gaussian fourth moments give

$$\mathbb{E}h(Z_1,Z_2)^2=3T_2^2+6T_4.$$

In the order-two Hoeffding decomposition, the first projection therefore has variance $2T_4$ and the degenerate remainder has variance $2T_2^2+2T_4$. Their orthogonality yields (103). A nonzero Gaussian vector has a nonzero inner product with an independent Gaussian vector almost surely on the positive support of Σ, proving $B_N>0$ almost surely.

Chebyshev's inequality and a union bound show that $|A_N/T_1-1|\le\varepsilon$ and $|B_N/T_2-1|\le\varepsilon$ simultaneously with the stated probability. Taking the ratio proves (104). The elementary bounds $T_2\le T_1^2$ and $T_4\le T_2^2$ prove the dimension-free bound on V_N; (A1)–(A2) give $T_1,T_2\asymp D$ and $T_4=O(D)$. Consequently $d_U/d_{\mathrm{PR}}\to 1$ in probability. The shallow spectral expansion gives $d_{\mathrm{eff}}/d_{\mathrm{PR}}\to 1$; combining this with the Gaussian shared-cloud theorem proves (105). These are joint probability statements on the original cloud; no independence between d_U and $\widehat{d}_{2\mathrm{NN}}$ is used. □

A finite-sample bound for the spectral target. The approximation of d_{eff} by d_{PR} can also be bounded deterministically. If $\|\Sigma\|\le L$, set $s_0=\sqrt{r/(3T_2)}$ and $\theta=Ls_0$. For $\theta\le 1/8$,

$$(1+4\theta)^{-2}\le\frac{d_{\mathrm{eff}}}{d_{\mathrm{PR}}}\le(1+4\theta)^{3/2}. \tag{106}$$

Indeed, differentiating the spectral curve and integrating gives

$$\frac{3T_2s^2}{(1+2Ls)^3}\le D\mathcal{C}_D(s)\le 3T_2s^2.$$

Thus its solution $D\mathcal{C}_D(s)=r$ satisfies $s_0\le s\le 2s_0$ because $(1+4\theta)^3\le(3/2)^3<4$. More precisely, $s/s_0\le(1+4\theta)^{3/2}$. The bounds

$$\frac{4sT_1}{(1+2Ls)^2}\le D\delta_D(s)\le 4sT_1$$

then prove (106). Combining it with (104), on the same event of probability at least $1-V_N/\varepsilon^2$,

$$\frac{1-\varepsilon}{\sqrt{1+\varepsilon}(1+4\theta)^{3/2}}\le\frac{d_U}{d_{\mathrm{eff}}}\le\frac{(1+\varepsilon)(1+4\theta)^2}{\sqrt{1-\varepsilon}}. \tag{107}$$

This bounds estimation of the Gaussian spectral target in the shallow regime.

Proposition D.9 (Bias of the empirical-covariance substitution)**.** *Under (A1)–(A2), suppose $n/D\to\gamma\in(0,\infty)$ and let*

$$\widehat{\Sigma}=\frac{1}{n-1}\sum_{i=1}^n(X_i-\overline{X})(X_i-\overline{X})^\top.$$

Let $d_{\mathrm{plug}}=d_{\mathrm{eff}}(n,\widehat{\Sigma})$, obtained by substituting the empirical eigenvalues into the population spectral equations. Then

$$\frac{d_{\mathrm{plug}}}{d_{\mathrm{eff}}(n,\Sigma_D)}\xrightarrow{\mathbb{P}}\left(1+\frac{m_1^2}{\gamma m_2}\right)^{-1/2}<1. \tag{108}$$

The same limit holds for the ratio of empirical and population PR targets. For $\Sigma_D=I_D$, the factor is $\sqrt{\gamma/(1+\gamma)}$.

Proof. This is the classical sample-covariance trace-moment distortion, propagated through the shallow TwoNN law. Gaussian orthogonal invariance in the sample index gives the distributional representation $\widehat{\Sigma}\overset{d}{=}m^{-1}\sum_{i=1}^m Y_iY_i^\top$, where $m=n-1$ and the Y_i are iid $\mathcal{N}(0,\Sigma_D)$. Hence

$$\frac{\mathrm{tr}\,\widehat{\Sigma}}{D}\xrightarrow{\mathbb{P}}m_1.$$

Writing $Q_i = \|Y_i\|^2$ and B_m for the off-diagonal statistic above,

$$\operatorname{tr}(\widehat{\Sigma}^2) = \frac{m-1}{m} B_m + \frac{1}{m^2} \sum_{i=1}^{m} Q_i^2.$$

The variance formula for B_m gives $B_m/D \to m_2$ in probability. Moreover,

$$\mathbb{E}Q_i^2 = T_1^2 + 2T_2, \qquad \operatorname{Var}(Q_i^2) = 8T_1^2T_2 + 8T_2^2 + 32T_1T_3 + 48T_4 = O(D^3).$$

Thus $D^{-1}m^{-2}\sum_i Q_i^2 \to m_1^2/\gamma$ in probability, and

$$\frac{\operatorname{tr}(\widehat{\Sigma}^2)}{D} \xrightarrow{\mathbb{P}} m_2 + \frac{m_1^2}{\gamma}.$$

The ratio of empirical to population participation ratios therefore converges to $m_2/(m_2 + m_1^2/\gamma)$. Finally, $\|\widehat{\Sigma}\| = O_{\mathbb{P}}(1)$: writing the sample matrix as $G\Sigma_D^{1/2}$, a fixed-radius sphere net and the chi-square Chernoff bound give $\|G\|^2/m = O_{\mathbb{P}}(1)$ when $m/D \to \gamma > 0$. Consequently the deterministic bound (106) applies with probability tending to one to the empirical spectrum, since $\log n = o(D)$ and its second spectral moment stays positive. Its phase target is asymptotically equivalent to its PR target. Taking square roots of the PR ratio proves (108). □

The trace-moment predictor consistently estimates the population Gaussian target at shallow resolution. The nonlinear ratios d_U and A_N^2/B_N need not be unbiased. Under the proposition's Gaussian assumptions, empirical spectral dispersion lowers the baseline: for fixed observed TwoNN, $e = \widehat{d}_{\mathrm{2NN}}/d_{\mathrm{null}} - 1$ then increases. This mechanism acts against the negative neural residual, rather than explaining it. Its sign and magnitude are not guaranteed for finite neural clouds outside those assumptions; d_U there would also be a diagnostic requiring validation.

E SPHERICAL SAMPLE-RESOLUTION LAW

Noise, anisotropy, or unbounded support could account for the Gaussian phase. A sphere removes all three explanations: it is smooth, compact, noiseless, and has a known manifold dimension. High-dimensional pairwise-angle extremes already have classical phase laws (Cai et al., 2013); Theorem E.1 converts the cap hazard into the TwoNN-specific first/second-spacing center and proves it for the ordinary shared-cloud aggregate.

Let $X_1, \ldots, X_n$ be independent and uniform on the unit sphere $S^d \subset \mathbb{R}^{d+1}$. Its Hausdorff and pointwise dimension is d.

Theorem E.1 (TwoNN on one shared sphere sample)**.** *Let $d \to \infty$ along any sequence $n = n_d$. Put $m = n - 1$, $r_d = \log m$, and assume*

$$r_d \to \infty, \qquad r_d \le Kd$$

for some fixed $K < \infty$. Let

$$I(t) = -\frac{1}{2}\log(1 - t^2),$$

let $t_{d,n} \in (0,1)$ solve $I(t_{d,n}) = r_d/d$, and define

$$d_{\mathrm{sph}}(d,n) = d\,\frac{2t_{d,n}}{1 + t_{d,n}}. \tag{109}$$

Then

$$\frac{\widehat{d}_{\mathrm{2NN}}}{d_{\mathrm{sph}}(d,n)} \xrightarrow{\mathbb{P}} 1. \tag{110}$$

For fixed $1 \le k < \ell$, fixed q, and fixed distinct indices $i_1, \ldots, i_q$, the scaled log ratios at these indices obey

$$d_{\mathrm{sph}}(d,n)\log\frac{R_{i,\ell}}{R_{i,k}} \xrightarrow{\mathcal{L}} -\log B_{k,\ell-k}$$

jointly to independent limits, with convergence in every fixed moment.

Consequently, if $\log n/d \to c \in (0,\infty)$*, put*

$$t_c = \sqrt{1-e^{-2c}}, \qquad \Delta_{\mathrm{sph}}(c) = \frac{2t_c}{1+t_c}. \tag{111}$$

Then

$$\frac{\widehat{d}_{2\mathrm{NN}}}{d} \xrightarrow{\mathbb{P}} \Delta_{\mathrm{sph}}(c). \tag{112}$$

If instead $\log n = o(d)$*, then*

$$\widehat{d}_{2\mathrm{NN}} = 2\sqrt{2d\log n}\big(1+o_{\mathbb{P}}(1)\big). \tag{113}$$

Proof. Fix an anchor x. For an independent uniform point Y, let $T = \langle Y, x\rangle$. Its density and upper-tail function are

$$h_d(t) = c_d(1-t^2)^{(d-2)/2}, \qquad G_d(t) = \mathbb{P}(T \ge t),$$

where

$$c_d = \frac{\Gamma((d+1)/2)}{\sqrt{\pi}\,\Gamma(d/2)} = \sqrt{\frac{d}{2\pi}}\{1+O(d^{-1})\}.$$

For $t > 0$, integration by parts gives the identity

$$\frac{G_d(t)}{c_d} = \frac{(1-t^2)^{d/2}}{dt} - \frac{1}{d}\int_t^1 z^{-2}(1-z^2)^{d/2}\,\mathrm{d}z. \tag{114}$$

The second term is nonnegative and at most $G_d(t)/(dt^2c_d)$. Hence

$$\frac{c_d(1-t^2)^{d/2}}{dt\{1+(dt^2)^{-1}\}} \le G_d(t) \le \frac{c_d(1-t^2)^{d/2}}{dt}. \tag{115}$$

Uniformly when $0 < t \le t_K < 1$ and $dt^2 \to \infty$,

$$G_d(t) = \frac{c_d(1-t^2)^{d/2}}{dt}\{1+O((dt^2)^{-1})\}, \tag{116}$$

$$\frac{h_d(t)}{G_d(t)} = \frac{dt}{1-t^2}\{1+O((dt^2)^{-1})\}. \tag{117}$$

Let $t_d(a)$ be the cap height satisfying $G_d\{t_d(a)\} = a/m$. Equation (115) gives

$$-\log G_d(t) = dI(t) + \log(t\sqrt{d}) + O(1) + O((dt^2)^{-1}).$$

Since $dt_{d,n}^2 \asymp r_d \to \infty$ and

$$dt_{d,n}I'(t_{d,n}) = \frac{dt_{d,n}^2}{1-t_{d,n}^2} \asymp r_d,$$

while the prefactor is $O(\log(1+r_d)) = o(r_d)$, uniformly for a in compact subsets of $(0,\infty)$,

$$\frac{t_d(a)}{t_{d,n}} = 1+o(1). \tag{118}$$

The squared distance is $S = 2(1-T)$. At $q = 2(1-t)$, $F_S(q) = G_d(t)$ and $f_S(q) = h_d(t)/2$, so

$$\beta(u) = \frac{G_d(t)}{2(1-t)h_d(t)}. \tag{119}$$

Equations (117)–(119) therefore give

$$\beta(a/m) = \frac{1+t_{d,n}}{2dt_{d,n}}\{1+o(1)\} = \frac{1+o(1)}{d_{\mathrm{sph}}(d,n)}. \tag{120}$$

This remains uniform when $t_{d,n} \downarrow 0$ because $t_{d,n}\sqrt{d} \asymp \sqrt{r_d} \to \infty$.

We next establish the moments required by the shared-cloud argument. Put

$$H_d(r) = \min\{d, \sqrt{d(1+r)}\}.$$

Then

$$d_{\mathrm{sph}}(d,n) \asymp H_d(r_d). \tag{121}$$

For $0 < u \le 1/2$, let $t = G_d^{-1}(u) \ge 0$. If $t \ge d^{-1/2}$, (115) and (119) give $\beta(u) \le 1/(dt)$. If $t < d^{-1/2}$, then $h_d(t) \ge c\sqrt{d}$ and $\beta(u) \le C/\sqrt{d}$. For $t \le 1/2$, (115) also gives $\log(1/u) \le C(1+dt^2)$; for $t \ge 1/2$, $H_d(\log(1/u)) \le d \le 2dt$. Thus

$$\beta(u) \le \frac{C}{H_d(\log(1/u))}, \qquad 0 < u \le \tfrac{1}{2}. \tag{122}$$

Let $U_{(k)} < U_{(\ell)}$ be fixed uniform order statistics. The part of the inverse-CDF integral below $1/2$ is bounded, by (121)–(122), by

$$C\left[\{-\log(mU_{(k)})\}_+ + \{\log(mU_{(\ell)})\}_+\right]. \tag{123}$$

For the part above $1/2$, write $u = 1 - G_d(w)$, $w \in (0,1)$. The squared-radius quantile is $2(1+w)$ and $du = h_d(w)\ \mathrm{d}w$, so

$$\beta(u)\ \mathrm{d}\log u = \frac{\mathrm{d}w}{2(1+w)}. \tag{124}$$

If w_ℓ corresponds to $U_{(\ell)} > 1/2$, the upper integral is at most $w_\ell/2$. Put $T_\ell = -\log(1-U_{(\ell)})$. Spherical concentration and (115) give

$$w_\ell \le C \begin{cases} \sqrt{(1+T_\ell)/d}, & T_\ell \le d, \\ 1, & T_\ell > d. \end{cases}$$

Together with (121), the normalized upper part is at most

$$C\left[\sqrt{(1+r_d)(1+T_\ell)}\mathbf{1}\{U_{(\ell)} > 1/2,\ T_\ell \le d\} + d\,\mathbf{1}\{T_\ell > d\}\right].$$

For fixed k, ℓ,

$$\sup_m \mathbb{E}[\{-\log(mU_{(k)})\}_+^p] < \infty,$$
$$\sup_m \mathbb{E}[\{\log(mU_{(\ell)})\}_+^p] < \infty,$$
$$\mathbb{P}(T_\ell > t) \le C_\ell m^{\ell-1} e^{-(m-\ell+1)t}, \qquad t \ge \log 2.$$

Therefore, for every fixed $p < \infty$,

$$\sup_{d,n} \mathbb{E}\left[\left\{d_{\mathrm{sph}} \log \frac{R_{i,\ell}}{R_{i,k}}\right\}^p\right] < \infty. \tag{125}$$

The local convergence from (120) consequently extends to convergence in every fixed moment.

It remains to handle two anchors. For anchors x, z, put $\rho = \langle x, z\rangle$ and

$$\mathcal{T}_d = \{(x,z) : |\rho| \le d^{-1/4}\}.$$

Since $\mathbb{E}\rho^2 = 1/(d+1)$, $\mathbb{P}\{(X,Z) \in \mathcal{T}_d^c\} \le Cd^{-1/2}$. For a shared candidate Y, let $A = \langle Y, x\rangle$ and $B = \langle Y, z\rangle$. Its squared projection onto $\mathrm{span}\{x,z\}$ is

$$Q_\rho(A,B) = \frac{A^2 - 2\rho AB + B^2}{1-\rho^2}.$$

On $A, B \ge t > 0$, $Q_\rho(A,B) \ge 2t^2/(1+|\rho|)$. The squared length of a uniform sphere point's projection onto a fixed two-dimensional subspace is $\mathrm{Beta}(1,(d-1)/2)$, and hence

$$\mathbb{P}(R_2^2 \ge q) = (1-q)^{(d-1)/2}.$$

If t_a, t_b are the cap heights at probabilities $a/m, b/m$ and $t_{\min} = \min(t_a, t_b)$, then uniformly on $\mathcal{T}_d$,

$$\mathbb{P}\left(U_x \le \frac{a}{m},\, U_z \le \frac{b}{m} \mid x, z\right) \le \left[1 - \frac{2t_{\min}^2}{1+|\rho|}\right]_+^{(d-1)/2}. \tag{126}$$

If $r_d/d \to 0$, then $t_{\min}^2 = 2r_d/d\{1+o(1)\}$ and the right side is $m^{-2+o(1)}$. If $r_d/d \in [\varepsilon, K]$, the cap heights stay in a compact subset of $(0,1)$. Where the bracket is positive, the joint rate

$$J(t) = -\frac{1}{2}\log(1-2t^2)$$

is uniformly larger than the marginal rate $I(t) = -\frac{1}{2}\log(1-t^2)$; otherwise the bound is zero. Thus some $\eta_K > 0$ satisfies

$$\mathbb{P}\left(U_x \le \frac{a}{m},\, U_z \le \frac{b}{m} \mid x, z\right) \le m^{-1-\eta_K+o(1)} \tag{127}$$

throughout the stated regime. Lemma C.9 now implies independent limiting extreme processes. The mutual anchor candidate is negligible because its cap PIT is uniform conditional on either anchor.

For pooled TwoNN put $Z_i = d_{\mathrm{sph}} L_i$. The $p = 4$ case of (125) gives $\sup_{d,n} \mathbb{E}Z_i^4 < \infty$, and Cauchy–Schwarz gives $\sup_{d,n} \mathbb{E}(Z_1 Z_2)^2 < \infty$. Thus the products are uniformly integrable. Pair convergence to independent exponentials implies

$$\mathbb{E}Z_1 \to 1, \qquad \mathbb{E}Z_1^2 \to 2, \qquad \mathbb{E}(Z_1 Z_2) \to 1.$$

Lemma C.10 yields $n^{-1}\sum_i Z_i \to 1$ in L^2, proving (110). Finally, $t_{d,n} \to t_c$ when $\log n/d \to c$, while $t_{d,n} = \sqrt{2\log n/d}\{1+o(1)\}$ when $\log n = o(d)$, proving (112) and (113). □

The same neighbor-window correction used for the Gaussian models has a distinct spherical specialization. For $0 < a < m/2$, define

$$\tau_d(a;m) = \sqrt{1 - \exp\left\{-\frac{2\log(m/a)}{d}\right\}}, \qquad d_{\mathrm{sph}}(d,m;a) = d\,\frac{2\tau_d(a;m)}{1+\tau_d(a;m)}, \tag{128}$$

and

$$M_{k,\ell}^{\mathrm{sph}}(d,m) = \int_0^{m/2} \frac{\pi_{k,\ell}(a)}{a\, d_{\mathrm{sph}}(d,m;a)}\, \mathrm{d}a, \qquad d_{k,\ell}^{\mathrm{rw,sph}} = \frac{\psi(\ell)-\psi(k)}{M_{k,\ell}^{\mathrm{sph}}(d,m)}. \tag{129}$$

Here $d_{\mathrm{sph}}(d,m;1) = d_{\mathrm{sph}}(d,n)$ when $m = n-1$.

Proposition E.2 (Spherical Poisson neighbor-window approximation)**.** *Under the assumptions of Theorem E.1, for every fixed* $1 \le k < \ell$*,*

$$\frac{n^{-1}\sum_i \log(R_{i,\ell}/R_{i,k})}{M_{k,\ell}^{\mathrm{sph}}(d,n-1)} \xrightarrow{\mathbb{P}} 1, \qquad \frac{\widehat{d}_{k,\ell}}{d_{k,\ell}^{\mathrm{rw,sph}}} \xrightarrow{\mathbb{P}} 1.$$

Moreover, $d_{k,\ell}^{\mathrm{rw,sph}}/d_{\mathrm{sph}}(d,n) \to 1$*. Thus the formula in* (129) *is a parameter-free finite-order approximation of the spherical phase law and leaves its first-order limit unchanged.*

Proof. Rotational invariance makes the conditional radial CDF independent of the anchor. Therefore the same Tonelli argument as in (99) gives the identity

$$\mathbb{E}\log\frac{R_{i,\ell}}{R_{i,k}} = \int_0^m \frac{w_m(a)}{a}\,\beta(a/m)\,\mathrm{d}a. \tag{130}$$

On each compact a-interval in $(0,\infty)$, $w_m(a) = \pi_{k,\ell}(a) + o(1)$ uniformly. Repeating (118)–(120) with $\log(m/a)$ in place of $\log m$ yields

$$\beta(a/m) = d_{\mathrm{sph}}(d,m;a)^{-1}\{1+o(1)\} \tag{131}$$

uniformly on the same interval. Notice that $d\tau_d(a;m)^2 \asymp \log(m/a) \to \infty$ there, including when $\tau_d(a;m) \downarrow 0$, so the moving-cap error remains uniform.

We record the truncation needed to pass from compact intervals to the full integral. Equations (121) and (122), with $\log(m/a)$ replacing r_d, give

$$d_{\mathrm{sph}}(d, m; a) \asymp H_d\{\log(m/a)\}.$$

Lemma D.5, with $q = d$, therefore dominates both normalized model and actual integrands by $g_{k,\ell}$ for $a < m/2$. Their tails outside $[A^{-1}, A]$ vanish as $A \to \infty$ by integrability. Finally, the event represented by $a > m/2$ is exponentially unlikely; the upper-radius change of variables (124) and the tail calculation used for (125) make its normalized contribution vanish. Consequently, first $m, d \to \infty$ and then $A \to \infty$ in (130) gives

$$\mathbb{E} \log \frac{R_{i,\ell}}{R_{i,k}} = M^{\mathrm{sph}}_{k,\ell}(d, m)\{1 + o(1)\}.$$

For two anchors, truncate both inverse-CDF integrals to $a \in [A^{-1}, A]$. The two-cap gap (127) and Lemma C.9 make these truncated contributions asymptotically independent. The preceding tail bounds together with the fourth-moment envelope (125) make the discarded cross terms uniformly negligible as $A \to \infty$. The cross moment therefore factorizes, and Lemma C.10 proves convergence of the shared average; the estimator claim follows algebraically.

For every fixed a, the definitions imply $d_{\mathrm{sph}}(d, m; a)/d_{\mathrm{sph}}(d, n) \to 1$. Applying the same uniform truncation and $\int_0^\infty \pi_{k,\ell}(a)\ \mathrm{d}a/a = \psi(\ell) - \psi(k)$ shows

$$d_{\mathrm{sph}}(d, n) M^{\mathrm{sph}}_{k,\ell}(d, m) \longrightarrow \psi(\ell) - \psi(k),$$

which proves the last assertion. □

Corollary E.3 (Matching spherical resolution boundary)**.** *For $\alpha \in (0, 1)$, let*

$$t_\alpha = \frac{\alpha}{2 - \alpha}, \qquad c_{\mathrm{sph}}(\alpha) = -\frac{1}{2} \log(1 - t_\alpha^2). \tag{132}$$

Then $\log n/d = c_{\mathrm{sph}}(\alpha) + o(1)$ is the sharp asymptotic threshold at which pooled TwoNN reports the fraction α of the true sphere dimension. Below it the limiting fraction is smaller, and above it the limiting fraction is larger.

In particular, with polynomially many samples TwoNN reports a vanishing fraction of the true dimension even on a smooth, compact, noiseless manifold.

Dependence on the distance convention. The theorem uses chord length $R = \sqrt{2(1 - t)}$, not geodesic distance $\vartheta = \arccos t$. Their log derivatives satisfy

$$\frac{\mathrm{d} \log \vartheta}{\mathrm{d} \log R} = \frac{2(1 - t)}{\arccos(t)\sqrt{1 - t^2}}.$$

At shallow resolution $t \to 0$, this factor tends to $4/\pi$; the corresponding leading inverse-log-spacing center is therefore $(\pi/2)\sqrt{2d \log n}$, or $\pi/4$ of the chordal center. This follows at a fixed anchor from the same quantile localization and the chain rule. The stated shared-cloud theorem and mean-calibration bounds use chordal distance. As $t \to 1$ the factor tends to one, recovering the local metric equivalence.

E.1 Quantitative spherical mean calibration

On the sphere, retaining the cap prefactor improves approximation of the mean log spacing beyond the first-order neighbor-window equivalence. Write

$$\mu_{k,\ell}(d, m) = \mathbb{E} \log \frac{R_{i,\ell}}{R_{i,k}}, \qquad c_{k,\ell} = \psi(\ell) - \psi(k).$$

Rotational invariance makes $\mu_{k,\ell}$ also the expectation of the ordinary shared-cloud average. An elementary exact benchmark is available: if Q_d is the quantile function of $\mathrm{Beta}(d/2, d/2)$ and $b_{j,m-j+1}$ is the corresponding uniform-order-statistic density, then

$$\mu_{k,\ell}(d, m) = \frac{1}{2} \int_0^1 \log Q_d(u)\{b_{\ell,m-\ell+1}(u) - b_{k,m-k+1}(u)\}\ \mathrm{d}u. \tag{133}$$

Indeed, $R^2/4$ has this beta distribution and each radial probability transform is uniform. This identity computes the mean log spacing, not the expectation of its reciprocal.

Proposition E.4 bounds the relative mean-calibration error by $O((\log m)^{-2})$ and identifies the larger term omitted by the phase-only neighbor window. Retain the exact spherical normalizer c_d from the proof of Theorem E.1. For $0 < a < \sqrt{m}$, let $v_d(a; m) \in (0, 1)$ be the unique solution of

$$A_d(v) := \frac{c_d(1 - v^2)^{d/2}}{d\, v} = \frac{a}{m}. \tag{134}$$

Define

$$b_d^{\mathrm{pc}}(a; m) = \frac{1 + v_d(a; m)}{2d\, v_d(a; m)} \left\{ 1 - \frac{1 - v_d(a; m)^2}{(d + 2) v_d(a; m)^2} \right\}, \tag{135}$$

$$M_{k,\ell}^{\mathrm{pc}}(d, m) = \int_0^{\sqrt{m}} \frac{\pi_{k,\ell}(a)}{a} b_d^{\mathrm{pc}}(a; m)\ \mathrm{d}a. \tag{136}$$

The superscript refers to retaining the cap prefactor and its first integration-by-parts correction. The integrand is positive for all sufficiently large d, m in the regime below.

Proposition E.4 (Quantitative spherical mean calibration)**.** *Fix $1 \le k < \ell$ and $K < \infty$. Let $d \to \infty$, $m \to \infty$, $r = \log m \to \infty$, and $r \le Kd$. Then*

$$\frac{\mu_{k,\ell}(d, m)}{M_{k,\ell}^{\mathrm{pc}}(d, m)} = 1 + O_{K,k,\ell}(r^{-2}). \tag{137}$$

If additionally $r = o(d)$, the phase-only neighbor-window mean in (129) *satisfies*

$$\frac{\mu_{k,\ell}(d, m)}{M_{k,\ell}^{\mathrm{sph}}(d, m)} - 1 = \frac{\log r}{4r}\{1 + o(1)\}. \tag{138}$$

Consequently, with $d_{k,\ell}^{\mathrm{mean}} = c_{k,\ell}/\mu_{k,\ell}$ and $d_{k,\ell}^{\mathrm{pc}} = c_{k,\ell}/M_{k,\ell}^{\mathrm{pc}}$, $d_{k,\ell}^{\mathrm{pc}}/d_{k,\ell}^{\mathrm{mean}} = 1 + O(r^{-2})$, whereas $d_{k,\ell}^{\mathrm{rw,sph}}/d_{k,\ell}^{\mathrm{mean}} - 1$ has the leading term in (138) *in the shallow regime. These are deterministic calibration errors for the mean log spacing. They give neither a finite-sample concentration rate for the reused-cloud average nor a bias or mean-squared-error bound for its reciprocal.*

Proof. We first retain a second term in the cap expansion. A further integration by parts in (114) gives, for every $0 < t < 1$,

$$\frac{G_d(t)}{A_d(t)} = 1 - \alpha_d(t) + \varepsilon_d(t), \quad \alpha_d(t) = \frac{1 - t^2}{(d + 2)t^2}, \quad 0 \le \varepsilon_d(t) \le \frac{3(1 - t^2)^2}{(d + 2)(d + 4)t^4}. \tag{139}$$

For completeness, the remainder after the first integration by parts is

$$\int_t^1 z^{-2}(1 - z^2)^{d/2}\ \mathrm{d}z = \frac{(1 - t^2)^{(d+2)/2}}{(d + 2)t^3} - \frac{3}{d + 2} \int_t^1 z^{-4}(1 - z^2)^{(d+2)/2}\ \mathrm{d}z.$$

The last integral is nonnegative and at most $(1 - t^2)^{(d+4)/2}/\{(d + 4)t^5\}$, which proves the displayed bound.

Put $t = t_d(a) = G_d^{-1}(a/m)$ and $v = v_d(a; m)$. Uniformly for $0 < a \le \sqrt{m}$,

$$dt^2 \ge c_K r. \tag{140}$$

To verify this, use the cap sandwich at the endpoint $a = \sqrt{m}$. It gives $dt_d(\sqrt{m})^2 \asymp_K r$ by the same inversion as (118), with $r/2$ in place of r; smaller a only increases the cap height. The endpoint height stays bounded away from one because $r/(2d) \le K/2$. Since $G_d \le A_d$ and A_d decreases strictly from infinity to zero on $v \in (0, 1)$, $v \ge t$. Moreover

$$-\frac{\mathrm{d}}{\mathrm{d}\log z} \log A_d(z) = 1 + \frac{dz^2}{1 - z^2} \ge dt^2, \qquad t \le z \le v.$$

Equations (139) and (140) therefore give

$$0 \le \log(v/t) \le \frac{\log\{A_d(t)/G_d(t)\}}{dt^2} \le \frac{C}{d^2 t^4}. \tag{141}$$

Let $b_0(z) = (1+z)/(2dz)$. Its logarithmic derivative with respect to $\log z$ is $-1/(1+z)$, while $|\,\mathrm{d}\alpha_d(z)/\,\mathrm{d}\log z| = 2/\{(d+2)z^2\}$. Using (119) and (139), we obtain

$$\begin{aligned}\beta(a/m) &= b_0(t)\{1-\alpha_d(t)+\varepsilon_d(t)\}\\ &= b_0(v)\{1-\alpha_d(v)\}\{1+O_K(r^{-2})\}\\ &= b_d^{\mathrm{pc}}(a;m)\{1+O_K(r^{-2})\}, \qquad 0<a\le\sqrt{m}.\end{aligned}$$

All errors here are uniform, including arbitrarily small a. Also $1-\alpha_d(v) \ge 1 - C_K/r > 0$ and both slopes are at most $C_K/\sqrt{dr}$ on this interval. On every fixed interval of positive a they are comparable to $1/\sqrt{dr}$, with constants depending on K.

We next check that replacing the coverage weight and truncating the integral cost less than r^{-2}. For $a \le \sqrt{m}$ and fixed k, ℓ, expansion of each binomial probability gives

$$|w_m(a) - \pi_{k,\ell}(a)| \le \frac{C_{k,\ell}}{m}(1+a^2)\pi_{k,\ell}(a). \tag{142}$$

Indeed, the ratio of its jth binomial term to $e^{-a}a^j/j!$ is

$$\frac{(m)_j}{m^j}\exp(a)(1-a/m)^{m-j},$$

whose logarithm has absolute value at most $C_j(1+a^2)/m \le C_j$; use $|e^x - 1| \le e^{|x|}|x|$ and sum over $k \le j \le \ell - 1$. Since $\int_0^\infty (1+a^2)\pi_{k,\ell}(a)\,\mathrm{d}a/a < \infty$, (142) changes the mean by $O_{K,k,\ell}\{m^{-1}(dr)^{-1/2}\}$.

For $\sqrt{m} < a < m/2$, the global slope bound (122) and the coverage bound (98) bound the discarded mean by

$$\frac{C}{\sqrt{d}}\int_{\sqrt{m}}^{\infty} a^{\ell-2}e^{-a/2}\,\mathrm{d}a.$$

For the part above $u = 1/2$, the change of variables (124) bounds the log-radius contribution by $w_\ell/2$, where $U_{(\ell)} = 1 - G_d(w_\ell)$ if $U_{(\ell)} > 1/2$. With $T_\ell = -\log(1-U_{(\ell)})$, the spherical concentration bound used above gives $w_\ell \le C\sqrt{(1+T_\ell)/d}$ for every such T_ℓ, since for $T_\ell > d$ its right side already exceeds one. The beta-order-statistic tail then yields a mean contribution at most

$$\frac{C}{\sqrt{d}}\, m^{\ell-1}2^{-(m-\ell+1)}.$$

Both discarded terms are $o\{r^{-2}(dr)^{-1/2}\}$. The corrected integral is comparable to $(dr)^{-1/2}$, by its uniform upper bound, positivity on a fixed interval, and the positivity of $\pi_{k,\ell}$ there. Combining these estimates with the Tonelli identity (130) proves (137).

Finally assume $r = o(d)$ and put $L = \log(m/a)$ and $\tau = \tau_d(a;m)$ from (128). On the expanding interval $e^{-\sqrt{r}} \le a \le e^{\sqrt{r}}$, $L/r \to 1$ uniformly. The definition of v and the gamma-ratio formula for c_d give

$$d\{I(\tau) - I(v)\} = \log(dv/c_d) = \tfrac{1}{2}\log r + O(1) + o(1), \qquad v^2 \sim \tau^2 \sim 2r/d.$$

Applying the mean-value theorem in the logarithmic height coordinate, $dzI'(z) = dz^2/(1-z^2) = 2r\{1+o(1)\}$ for $v \le z \le \tau$, uniformly on this interval. Hence

$$\log(\tau/v) = \frac{\log r}{4r}\{1+o(1)\}.$$

Since $\,\mathrm{d}\log b_0(z)/\,\mathrm{d}\log z = -1/(1+z) = -1 + o(1)$ here and $\alpha_d(v) = O(r^{-1})$, it follows that

$$\frac{b_d^{\mathrm{pc}}(a;m)}{b_0(\tau_d(a;m))} - 1 = \frac{\log r}{4r}\{1+o(1)\}$$

uniformly on that interval. Outside it, the lower Poisson tail is $O(e^{-k\sqrt{r}})$ after normalization and the upper Poisson tail is smaller still. To see uniformity, both slopes are at most $C/\sqrt{dr}$ for $a \le \sqrt{m}$, while the phase slope is at most $C/\sqrt{d}$ for $\sqrt{m} < a < m/2$; integrate the explicit Poisson series. Thus the same relative expansion holds for the ratio of the integrals. Equation (137), whose remainder is of smaller order, proves (138). □

Proposition E.4 controls the spherical mean log spacing. Table 4 evaluates this correction separately from the phase-only experiments.

Deterministic quadrature check. We evaluate the exact beta-quantile mean, the original phase-only mean, and (136) by one-dimensional quadrature. Table 4 reports $100(\mu/M - 1)$ for neighbor orders $(1, 2)$. These are deterministic quadrature checks, separate from the sampling experiments; no clouds are sampled and no estimator-MSE claim is made. The largest reported quadrature error over both checked neighbor pairs $(1, 2)$ and $(2, 4)$ is 1.14×10^{-13}. This error estimate excludes the truncation and special-function errors. The integration window is $e^{-32} \le a \le 100$, shortened to $\sqrt{m}$ for the corrected formula; the explicit coverage tails bound the omitted terms.

Table 4: Mean-calibration errors in percent from deterministic quadrature. Here m is the nearest integer to e^r; the displayed r is rounded.

d	r	Phase-only	Prefactor correction
4096	8	8.0803	2.0694
4096	12	5.9190	0.8329
4096	16	4.7147	0.4462
16384	24	3.6923	0.1920
16384	32	2.9285	0.1052

F Coordinate dependence at fixed population dimension

The convergence laws above hold for a fixed distribution and metric. This secondary contribution asks how the statistic changes when coordinates vary at a fixed sample size. The finite-cloud argument is independent of the convergence proofs; the Gaussian construction applies the phase law to three invertibly related distributions. Together they motivate the compensated neural interventions without assuming that every coordinate map has a large effect.

F.1 Scale invariance and samplewise monotonicity

Hausdorff dimension cannot increase under a Lipschitz map. Requiring the same behavior from every finite-sample estimate, however, is incompatible with exact scale invariance unless the statistic is constant.

Proposition F.1 (Elementary finite-cloud obstruction)**.** *Fix $n \ge 2$. Let T assign a real number to every n-point metric space and suppose:*

(i) $T(sX) = T(X)$ for every uniform scale $s > 0$;

(ii) $T(Y) \le T(X)$ whenever $f : X \to Y$ is a bijective non-expansive map between n-point metric spaces.

Then T is constant on the class of n-point metric spaces with distinct points.

Proof. Let $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_n\}$. Set

$$a = \min_{i \ne j} d_X(x_i, x_j) > 0, \qquad b = \max_{i \ne j} d_Y(y_i, y_j) < \infty.$$

For $0 < s \le a/b$, the labeled map $x_i \mapsto y_i$, with the target metric scaled by s, is non-expansive. Hence

$$T(Y) = T(sY) \le T(X).$$

Interchanging X and Y gives the reverse inequality. □

The same scale comparison bounds the oscillation of a statistic by its deviations from scale invariance and samplewise monotonicity. Define

$$\varepsilon_{\text{scale}} = \sup_{X,s} |T(sX) - T(X)|,$$

$$\varepsilon_{\mathrm{mono}} = \sup_{X,Y,f:X\to Y \text{ bijective non-expansive}} [T(Y) - T(X)]_+,$$

where X, Y range over n-point metric spaces with distinct points, $s > 0$, and $\mathrm{osc}(T) = \sup_X T(X) - \inf_X T(X)$.

Corollary F.2 (Quantitative no-go)**.** *Every finite-cloud statistic satisfies*

$$\mathrm{osc}(T) \le \varepsilon_{\mathrm{scale}} + \varepsilon_{\mathrm{mono}}.$$

Proof. For arbitrary X, Y, scale Y so that the labeled map $X \to sY$ is non-expansive. Then

$$T(Y) \le T(sY) + \varepsilon_{\mathrm{scale}} \le T(X) + \varepsilon_{\mathrm{mono}} + \varepsilon_{\mathrm{scale}}.$$

Take the supremum over Y and infimum over X. □

Finite ReLU interpolation realizes these cloud maps after uniform shrinkage.

Proposition F.3 (Realization by two-layer ReLU networks)**.** *Let $x_1, \dots, x_n \in \mathbb{R}^p$ be distinct and $y_1, \dots, y_n \in \mathbb{R}^q$ be arbitrary. There is a finite-width two-layer ReLU network $g : \mathbb{R}^p \to \mathbb{R}^q$ and an $s > 0$ such that $g(x_i) = sy_i$ for all i and $\mathrm{Lip}(g) \le 1$.*

Proof. Choose $v \in \mathbb{R}^p$ outside the finite union of hyperplanes $\{v : \langle v, x_i - x_j\rangle = 0\}$. Then $t_i = \langle v, x_i\rangle$ are distinct. For each output coordinate, interpolate the target values at $t_1, \dots, t_n$ by a continuous piecewise linear function, extended affinely outside their range. Every such function has a finite hinge representation

$$a + bt + \sum_j c_j (t - t_j)_+,$$

and hence a two-layer ReLU representation after writing the affine term using two ReLUs if necessary. Combining coordinates gives a globally Lipschitz map h with $h(x_i) = y_i$. Taking $g = sh$ for $0 < s \le 1/\max\{1, \mathrm{Lip}(h)\}$ completes the proof, including constant h. □

After enough uniform shrinkage, any labeled finite cloud can be reached from any other by a non-expansive map. Scale invariance then forces the comparison in both directions. Therefore a nonconstant diagnostic must relax exact scale invariance or universal samplewise monotonicity, or use information beyond the induced metric cloud, such as cross-layer or derivative information.

F.2 Nonmonotone TwoNN under invertible contractions

Three points suffice for a rise-then-fall TwoNN curve under invertible contractions. For $X(a) = \{(0,0), (a,0), (0,1)\}$, $a > 0$, $a \ne 1$, the three distances are $a, 1, \sqrt{1+a^2}$, so direct summation gives

$$\widehat{d}_{2\mathrm{NN}}(X(a)) = \frac{2}{\log\{1 + \max(a^2, a^{-2})\}}.$$

Starting at $a = 0.1$, apply $\mathrm{diag}(1, 1/9)$ and then $\mathrm{diag}(1, 0.09)$. Up to uniform scale the three shapes have $a = 0.1, 0.9, 10$, and the estimates are $0.4334, 2.4874, 0.4334$. Both maps are invertible contractions and no anchor has a distance tie. The following construction derives a corresponding Gaussian population example from the phase law.

Layerwise TwoNN curves are often read as evidence that a network first expands and then compresses the dimension of its representation. The next construction shows why that conclusion need not follow: the estimate rises and falls even though every layer is invertible and every distribution has the same population dimension. Fix $D > b_D$, $0 < \varepsilon < 1$, and set

$$\Sigma_0 = \mathrm{diag}(I_{b_D}, \varepsilon I_{D-b_D}).$$

Let

$$A_1 = \mathrm{diag}(\sqrt{\varepsilon} I_{b_D}, I_{D-b_D}), \qquad A_2 = \mathrm{diag}(I_{b_D}, \varepsilon I_{D-b_D}).$$

Then

$$\Sigma_1 = A_1 \Sigma_0 A_1^\top = \varepsilon I_D,$$

and

$$\Sigma_2 = A_2 \Sigma_1 A_2^\top = \mathrm{diag}(\varepsilon I_{b_D}, \varepsilon^3 I_{D-b_D}).$$

Both maps are invertible contractions.

Theorem F.4 (Rise then fall under invertible linear layers)**.** *Suppose* $b_D/D \to \theta \in (0,1)$ *and* $\log n/D \to c > 0$*. Choose*

$$\theta < \Delta_{\text{iso}}(c). \tag{143}$$

There exists $\varepsilon_0(\theta, c) > 0$ *such that, for every fixed* $0 < \varepsilon < \varepsilon_0$*, the limiting TwoNN values at the three layers satisfy*

$$\lim \frac{\widehat{d}_1}{D} > \max\left\{\lim \frac{\widehat{d}_0}{D}, \lim \frac{\widehat{d}_2}{D}\right\}.$$

Every distribution has pointwise and Hausdorff dimension D*, and both layer maps are information-preserving.*

Proof. Uniform scale does not affect neighbor ratios. Up to scale, the limiting spectra at the outer layers are

$$H_0 = \theta\delta_1 + (1-\theta)\delta_\varepsilon, \qquad H_2 = \theta\delta_1 + (1-\theta)\delta_{\varepsilon^2},$$

whereas the middle spectrum is $H_1 = \delta_1$. As $\varepsilon \downarrow 0$, both outer laws converge to

$$H_* = \theta\delta_1 + (1-\theta)\delta_0.$$

For this law,

$$\Delta_{H_*}(c) = \theta\Delta_{\text{iso}}(c/\theta) < \theta.$$

By (143), this is strictly smaller than $\Delta_{\text{iso}}(c)$. To make the continuity step explicit, all four laws are supported on $[0,1]$, so weak convergence gives uniform convergence of $\mathcal{C}_H(s)$ and $\delta_H(s)$ on every compact s-interval. At the unique solution s_* of $\mathcal{C}_{H_*}(s_*) = c$, equation (22) has a strictly positive limit derivative because H_* assigns mass θ to one. Monotonicity then brackets the solutions for H_0, H_2 between $s_* \pm \eta$ for every fixed $\eta > 0$, and hence their saddles converge to s_*. Uniform convergence of δ_H now gives $\Delta_{H_0}(c), \Delta_{H_2}(c) \to \Delta_{H_*}(c)$ as $\varepsilon \downarrow 0$. Thus the strict inequality persists for all sufficiently small positive ε. The estimator limits follow from Theorem C.2. Since all singular values of A_1, A_2 lie in $(0,1]$, the maps are invertible contractions, and invertible linear maps preserve the full Gaussian support dimension. □

The construction changes anisotropy while neighbor ratios remain invariant to uniform scale. The rise and fall comes from changing which directions are resolved at the available sample size.

Remark F.5 (The same construction at shallow resolution). Let $b_D/D \to \theta \in (0,1)$, $\log n \to \infty$, and $\log n = o(D)$. For any fixed $0 < \varepsilon < 1$, set

$$q_\theta(x) = \frac{\{\theta + (1-\theta)x\}^2}{\theta + (1-\theta)x^2}.$$

The normalized participation ratios at the three layers converge to $q_\theta(\varepsilon)$, 1, and $q_\theta(\varepsilon^2)$. For $0 < x < 1$,

$$1 - q_\theta(x) = \frac{\theta(1-\theta)(1-x)^2}{\theta + (1-\theta)x^2} > 0.$$

Corollary D.2 therefore gives

$$\frac{(\widehat{d}_0, \widehat{d}_1, \widehat{d}_2)}{(4/\sqrt{3})\sqrt{D\log n}} \xrightarrow{\mathbb{P}} (\sqrt{q_\theta(\varepsilon)}, 1, \sqrt{q_\theta(\varepsilon^2)}).$$

A union bound applies the three marginal convergence statements to the same transformed sample. Hence the middle estimate exceeds both outer estimates with probability tending to one. This needs no positive-resolution condition (143); all maps remain invertible and contractive.

The same construction can be made pointwise function preserving, rather than merely information preserving.

Corollary F.6 (Function-equivalent networks can have different TwoNN)**.** *Under the hypotheses of Theorem F.4 or Remark F.5, put* $B_0 = I$*,* $B_1 = A_1$*, and* $B_2 = A_2A_1$*. For* $r \in \{0,1,2\}$*, consider the two-linear-layer factorization*

$$f_r(z) = B_r^{-1}(B_r z).$$

Then $f_0 = f_1 = f_2$ *pointwise (and an arbitrary fixed downstream predictor composed with them has identical logits and predictions), while the hidden clouds* B_rX *have the strict rise-then-fall TwoNN pattern of Theorem F.4 or Remark F.5.*

Proof. Invertibility implies $B_r^{-1}B_r = I$. The hidden covariances are $\Sigma_0, \Sigma_1, \Sigma_2$, respectively, so Theorem F.4 or Remark F.5 establishes the estimator comparison. □

G EXPERIMENTAL DESIGN

Experimental roadmap. The studies form a sequence of four questions. First, do the asymptotic phase laws accurately predict finite Gaussian and spherical samples when the population spectrum or dimension is known? Second, can the Gaussian target itself be recovered when its spectrum must be estimated from finite data? Third, can TwoNN change under a coordinate transformation that preserves a trained network's function, including when the map is fitted on separate images? Fourth, can finite-sample formula error or class structure explain the Gaussian prediction's residual on neural representations? The first two questions separate phase approximation from target estimation. The third isolates coordinate sensitivity without changing predictions. For the fourth, direct global-Gaussian simulations calibrate formula error, then class-Gaussian mixtures test whether the measured class structure closes the remaining discrepancy. Agreement at either step would support that specific fitted-null explanation.

Additional analyses separate covariance-target estimation, dependence on fitting images, and direct Gaussian/class-mixture calibration. Their protocols were fixed before those measurements. Because the neural analyses reuse previously analyzed models and images, they are exploratory rather than independent confirmatory training runs; the prespecified analyses and their outcomes are reported separately.

G.1 GAUSSIAN AND SPHERICAL PREDICTION EXPERIMENT

This experiment tests the numerical content of the phase laws before introducing covariance estimation or learned representations. Comparing predictions with known-distribution samples distinguishes accurate finite-sample prediction from a systematic neighbor-order shift that the window correction should reduce.

We generated five model families: isotropic Gaussian, a two-block Gaussian spectrum, a geometrically decaying Gaussian spectrum, a Gaussian with covariance rank $D/2$, and the uniform sphere. The grid used

$$D \in \{8, 16, 32, 64, 128\}, \qquad n \in \{1024, 4096, 16384\},$$

five independent seeds, and neighbor pairs $(1,2), (2,4), (4,8), (8,16), (16,32)$. The grid contains 375 independent clouds and 1,875 cloud–neighbor-pair records. The two-block spectrum places unit variance on the first $\max(1, D/4)$ coordinates and variance 0.1 on the remainder. The geometric spectrum is $\lambda_j = \exp\{(j-1)\log(0.05)/(D-1)\}$, $j = 1, \ldots, D$. The rank-$D/2$ spectrum has $\max(1, D/2)$ unit eigenvalues and zeros elsewhere. Sphere points are normalized standard Gaussian vectors in $\mathbb{R}^{D+1}$ and use chordal distance.

All pairwise distances were computed blockwise in float64. For Gaussian models, the base prediction is d_{eff} from Equations (26)–(27). For spheres, it is Equation (109). The naive fixed-order prediction assigns the same $1/n$ phase point to every neighbor pair. The neighbor-window prediction is the parameter-free Poisson integral in Proposition D.6 for the Gaussian families and Proposition E.2 for the sphere; neither contains a fitted coefficient.

The primary accuracy criterion used all model families and neighbor orders with $D \geq 32$ and $n \geq 4096$, emphasizing the larger dimensions and samples. It required median absolute relative error below 10%, a study-chosen tolerance for numerical accuracy rather than a theorem-derived bound. After a discovery sweep identified systematic movement across neighbor windows, we specified the refinement and held-out evaluation before examining the confirmation set. Its untouched confirmation set contains seeds 310004 and 310005 in cells with either $n = 16384$ or $D \in \{8, 128\}$; the central subset contains 200 records and uses the same 10% threshold. The phase-only prediction and its error remain in the comparison. The refinement was developed on the discovery portion and evaluated on held-out cells. Its protocol was recorded internally rather than externally preregistered.

G.2 CONTROLLED GAUSSIAN TARGET ESTIMATION

The preceding phase test assumes a known spectrum; practical use requires estimating a target from the sample. Comparing d_U and covariance plug-in estimates with oracle quantities tests target recovery separately from the shallow approximation and the fluctuation of observed TwoNN.

We distinguish the population full phase d_{eff} from the shallow participation-ratio target d_{PR} in Proposition D.8. The controlled study uses two known spectra: all eigenvalues one, or $D/4$ eigenvalues one and the remaining $3D/4$ equal to 0.1. For each family, $D \in \{128, 512\}$ and $n/D \in \{1/2, 1, 2\}$ have 32 independent clouds per cell. An extension at $n = D$ uses $D \in \{1024, 2048\}$ with 16 clouds per cell, for 448 clouds in total. Every cloud records TwoNN, the oracle full and shallow targets, the empirical-spectrum full and shallow targets, and d_U.

The disjoint differences defining d_U follow the generated row order, fixed before outcomes. We report d_U/d_{PR} and its median absolute relative error separately from $d_{\text{PR}}/d_{\text{eff}}$ and $\widehat{d}_{\text{2NN}}/d_{\text{eff}}$. The empirical shallow target is compared with the attenuation reference $(1 + m_1^2/(\gamma m_2))^{-1/2}$ at $\gamma = n/D$; this is an asymptotic reference. A fixed translation of the first cloud in each cell checks the unknown-mean construction without adding an independent replicate. These simulations test the intended Gaussian target, not a universal neural dimension correction.

G.3 Neural model, data, and representation extraction

The neural studies use five independently trained models to connect the controlled laws to learned representations. Shared extraction settings make coordinate interventions and natural-layer comparisons paired within each model; replication is across training seeds.

We trained five ResNet-18 models from scratch on the 50,000-image CIFAR-10 training split, using seeds 270101–270105. The ImageNet stem was replaced by a 3×3, stride-one convolution and no max pooling. Training used 50 epochs, batches of 512, SGD with learning rate 0.1, momentum 0.9, Nesterov acceleration, weight decay 5×10^{-4}, and cosine decay with horizon 50. Random crops with four-pixel padding and random horizontal flips were followed by the standard CIFAR-10 channel normalization. Training used CUDA bfloat16 autocast; feature extraction and geometric analyses used the precisions stated below.

For training seed s, a CPU PyTorch generator seeded with $s + 17$ draws one permutation without replacement of all 10,000 test images using `torch.randperm`; its first 8,000 indices are retained. Nested prefixes of size 1,000, 2,000, 4,000, and 8,000 eliminate sample-composition confounding in resolution comparisons. At epochs 0, 25, and 50 we extracted global-average-pooled representations after the stem and residual stages 1–4, plus the ten-dimensional logits. Thus each seed has 72 natural cells (3 checkpoints, 6 layers, and 4 sample sizes). The function-preserving interventions use the epoch-50 layer-4 representation.

G.4 Function-preserving coordinate interventions

These interventions ask whether the same classifier can have different TwoNN values solely because of its hidden coordinates. Compensating the readout holds predictions fixed, while orthogonal and uniform-scale controls distinguish anisotropic effects from numerical changes.

Let $z \in \mathbb{R}^{1\times p}$ be a row-vector layer-4 activation, μ its empirical mean, and A an invertible matrix. We replace the activation by

$$z' = \mu + (z - \mu)A \tag{144}$$

and insert the inverse before the frozen linear classifier. If the original row-vector logits are $zW^\top + b$, the compensated classifier reconstructs $z = \mu + (z' - \mu)A^{-1}$ before applying $zW^\top + b$. This is the finite-network counterpart of Corollary F.6. We verify logits and predictions numerically for every seed and transformation rather than infer equivalence from algebra alone.

The inverse can be absorbed into the linear classifier weights and bias. An arbitrary linear map cannot in general be moved through an intervening nonlinearity. For the same-cloud intervention analysis, maps are fitted once to the full epoch-50, layer-4 cloud of $N = 8000$ rows, then evaluated on that cloud and its nested prefixes. Section G.6 applies the same fitting rule to separate fitting and evaluation images. With centered rows collected in Z_c, the frozen implementation uses

$$\widehat{\Sigma} = \frac{Z_c^\top Z_c}{N-1} = V \operatorname{diag}(\lambda_1, \dots, \lambda_p) V^\top, \qquad \lambda_1 \le \cdots \le \lambda_p,$$

computed by float64 `torch.linalg.eigh`. Define

$$f = \max\{10^{-4}\lambda_p, \epsilon_{64}\}, \qquad J = \{j : \lambda_j > f\},$$

$$\lambda_{\mathrm{ref}} = \text{lower median}\{\lambda_j : j \in J\}, \qquad a_j = \min\{5, \max\{0.2, \sqrt{\lambda_{\mathrm{ref}} / \max(\lambda_j, f)}\}\},$$

where ϵ_{64} is float64 machine epsilon and $\lambda_{\mathrm{ref}} = 1$ if J is empty. For even $|J|$, the lower of the two middle order statistics is used, as in `torch.median`. Thus the order is: compute the floor, select eigenvalues strictly above it to obtain the reference, floor the denominators, take square roots, and clip. This also handles zero or negative numerical eigenvalues. PCA equalization applies $A = V \operatorname{diag}(a_j) V^\top$. Tail contraction uses the same original PCA basis and replaces a_j by $0.1a_j$ for $j \le q = \max(1, \lfloor 3p/4 \rfloor)$, after clipping. Eigenvector signs cancel in both matrices. Equalization assigns equal scales to repeated eigenvalues and is independent of rotations within an exactly degenerate eigenspace. A tie crossing the tail cutoff follows the eigensolver's returned order.

For the random basis, a CPU PyTorch generator with seed $271828 + 1009p$ draws a float64 standard-normal $p \times p$ matrix. Its QR factorization is sign-normalized by multiplying each column of Q by $+1$ when the corresponding diagonal entry of R is positive and by -1 otherwise. The same Q is reused across training seeds and condition numbers. For $\kappa \in \{2, 4, 8, 16\}$, the applied matrix is

$$A_\kappa = Q \operatorname{diag}(b_j) Q^\top, \qquad b_j = \exp\left[-\tfrac{1}{2} \log \kappa + \frac{j-1}{p-1} \log \kappa \right].$$

The negative controls apply I, Q, and $2I$. Condition numbers throughout refer to the applied row-vector matrix A.

The random-basis family through condition number eight is the primary effect-size test. Condition 16 is a descriptive extension. PCA interventions were specified as secondary data-adapted transformations. The primary random-basis and PCA families use different bases and full scaling spectra.

G.5 Estimators and the spectral-null diagnostic

We report MacKay-aggregated TwoNN from Equation (12); the GRIDE-based log-moment estimator for all five neighbor pairs; local and pooled Levina–Bickel estimates at $k = 10, 20$; and asymptotic and published finite-sample-calibrated L2N2. Appendix L lists the formulas. Nearest-neighbor distances in the archived primary analysis were computed in float32 CUDA blocks of 256 with neighbors through order 32; the computational reproduction and additional analyses use float64. For each ratio window separately, nonfinite or nonpositive computed log ratios are excluded; the estimate is missing if fewer than $\max(32, \lfloor n/2 \rfloor)$ anchors remain. TwoNN keeps the declared numerator $n - 1$. This numerical convention is not part of the theorem; Appendix K.3 reports its sensitivity and the float64 control check. The function-equivalence check used float64 matrices and inverses before casting the transformed activations used by the geometric estimators.

For a neural cloud we compute empirical covariance eigenvalues from that same cloud and insert them into Equations (26)–(27). This plug-in prediction uses the estimated spectrum without a fitted correction; its residual also contains finite-sample prediction error. The downward trace-distortion direction proved for Gaussian shallow resolution is discussed in Proposition D.9. We define

$$e = \frac{\widehat{d}_{\mathrm{2NN}} - d_{\mathrm{null}}}{d_{\mathrm{null}}}. \tag{145}$$

We report d_{null} and e together: the former is the fitted Gaussian prediction and the latter is its relative error on the observed cloud.

G.6 Disjoint-fit coordinate interventions

A map fitted on the measured cloud can exploit that cloud's covariance. We therefore fit and evaluate on separate images, alongside same-size in-sample fits and matched random orientations. A persistent PCA effect would support coordinate sensitivity beyond the fitting images; the random comparison isolates basis alignment at fixed singular values.

For each trained model, stratify the saved 8,000 epoch-50 images by class into two halves of exactly 4,000 and reverse their fitting/evaluation roles. The fitting half determines every component of the layer-4 map: mean, covariance basis, eigenvalue floor, reference variance, and clipped scales. The

evaluation half is identical for identity and transformed conditions. A second fit on the evaluation half uses the same evaluation size, isolating the fitting effect from changes in sample resolution.

For each fit, five deterministic random orientations have exactly the same singular values as PCA. Both fitting modes and both folds therefore produce $5 \times 2 \times 2 \times 6 = 120$ transformed-cloud measurements. Classifier compensation uses the fitting mean and inverse map, with a 10^{-4} maximum logit-error tolerance and exact prediction agreement. Within a model, take the geometric mean of the two PCA factors; for random maps first take the median over five orientations within each fold. Model summaries then use the median across the five training seeds. Separate fitting and evaluation images do not make previously analyzed images an untouched confirmation set.

G.7 Class-aware and direct Gaussian calibration

A negative neural residual can reflect finite-sample formula error or a mismatch between a single Gaussian and class-structured data. We first replace the analytic prediction by direct simulation from the fitted global Gaussian, then replace that null by fitted class Gaussians. Agreement with real TwoNN would support the corresponding explanation, conditional on the estimated null.

We first measure unrestricted TwoNN, first/second-neighbor class purity, class-restricted TwoNN, and global and class spectra on all six natural 8,000-image clouds per model. For class c with n_c points, let $L_{i,c}$ be the log ratio when candidates are restricted to that class, and let $d_c = d_{\text{eff}}(n_c, \widehat{\Sigma}_c)$. We pool log spacings as

$$\widehat{d}_{\text{within}} = \frac{n-1}{\sum_c \sum_{i \in c} L_{i,c}}, \qquad d_{\text{within,null}} = \frac{n-1}{\sum_c n_c/d_c}. \tag{146}$$

The analytic comparator uses the class candidate count $n_c - 1$ inside each phase prediction. It is the leading log-spacing prediction for this class-restricted statistic, rather than an arithmetic mean of dimensions or a replacement null for unrestricted TwoNN.

For layers 3 and 4, the same disjoint halves fit a global Gaussian and ten class Gaussians using unbiased empirical covariances and empirical class means. Each fitted family generates 16 independent 4,000-point surrogate clouds. Mixture draws condition on the real evaluation class counts; within each class, points are independent under its fitted Gaussian. A fixed sensitivity family uses the same means and replaces each class covariance by

$$\widehat{\Sigma}_{c,0.1} = 0.9\widehat{\Sigma}_c + 0.1\frac{\operatorname{tr}\widehat{\Sigma}_c}{p} I_p. \tag{147}$$

The two mixture versions share standard-normal draws within each replicate. There are $5 \times 2 \times 2 \times 3 \times 16 = 960$ surrogate clouds. We compare unrestricted and class-restricted measurements with real data, and the direct global-Gaussian mean with its fitted analytic phase. When fit and evaluation class counts differ, the fitted class covariance is paired with the evaluation count in the phase equation.

Surrogate means and Monte Carlo SEs summarize the 16 draws conditional on the fitted null; they exclude covariance-fitting uncertainty. Fold ratios are combined geometrically within each trained model before model medians are reported. Purity uses arithmetic fold means. These repeated draws and folds are not additional trained models. Splitting avoids same-cloud fitting/evaluation coupling, but does not eliminate covariance-estimation bias. All shrinkage outcomes are retained without selecting a value for better agreement.

G.8 Effect sizes and statistical comparisons

The five training seeds are the independent experimental units. Percentile intervals from 10,000 seed-bootstrap resamples summarize variation across these networks; the resamples add no independent replication. For the primary random-basis family, the per-seed fold range is the largest divided by the smallest TwoNN value over identity and conditions 2, 4, and 8. The prespecified 1.5-fold threshold requires a 50% spread, with an effect above the orthogonal and scale controls. The 1.5-fold threshold is a study-chosen material-effect criterion. PCA equalization is a separately planned secondary transformation.

The natural-layer comparison asks whether residual signs and layer ordering replicate across seeds, and whether the residual at 8,000 images exceeds its change from 4,000 images. These compar-

isons distinguish a persistent model discrepancy from sensitivity to the last sample-size doubling. Appendix K.2 records the full thresholds and stop rule.

H RESULTS

We follow the same questions: finite-sample prediction, estimation of the Gaussian target, coordinate sensitivity with predictions held fixed, and successive explanations for the neural residual. The intervention controls establish function preservation before their TwoNN effects are interpreted.

H.1 FINITE-SAMPLE PREDICTION ACCURACY

On the complete central grid, the $(1, 2)$ pooled-TwoNN phase prediction has 2.51% median absolute relative error. Reusing the $1/n$ single-resolution prediction across all five neighbor pairs gives 10.64% error, exceeding the 10% accuracy criterion. All 30 central model–dimension–sample groups have negative signed error, and the error grows from 2.51% for $(1, 2)$ to 20.10% for $(16, 32)$. This pattern is consistent with a finite-order resolution shift. The spherical mean-calibration bound additionally identifies an omitted cap-prefactor term. Its leading shallow contribution $\log r/(4r)$ is about 6.37% at $m = 4095$ and 5.85% at $m = 16383$. These are asymptotic mean-calibration terms, not predictions of the finite-cloud median absolute error. They give a possible source of spherical overprediction; neither an exact explanation of the sphere's 3.68% error nor of the pooled five-family $(1, 2)$ comparison follows without separating family and sampling effects.

On the same 200 held-out central records, the neighbor-window approximation reduces median absolute relative error from 10.72% to 5.55%. The gain comes from later orders; for $(1, 2)$, error increases from 2.17% to 3.97% (Table 1). Single-resolution signed errors decrease from -2.17% to -20.09% as the pair advances from $(1, 2)$ to $(16, 32)$. Later neighbors probe larger probability masses, and integrating over that window reduces this systematic overprediction. Its held-out errors remain below 7.5% for every model and 6.6% for every neighbor pair. Across the complete grid, error falls from 8.30% to 4.28%; on the complete central grid it falls from 10.64% to 5.34%. We report these full-grid summaries in addition to the prespecified held-out comparison; they do not replace the held-out criterion.

H.2 TRACE-TARGET ESTIMATION ON CONTROLLED GAUSSIANS

The ordinary empirical-covariance plug-in shows the predicted attenuation, whereas d_U closely estimates its oracle shallow target (Figure 2c and Table 16). At $n = D = 512$, the isotropic and two-block empirical shallow targets have median ratios 0.7063 and 0.8414 to their oracle values; the corresponding asymptotic references are 0.7071 and 0.8421. Median absolute relative errors of d_U are only 0.245% and 0.385%. The smallest cells, $D = 128, n = 64$, have errors 2.40% and 2.67%; at $n = D = 2048$ these fall to 0.07% and 0.04%. The unknown-mean translation check agrees within 2.74×10^{-14} relative error for both TwoNN and d_U.

These comparisons separate target estimation from phase approximation and estimator fluctuation (Table 17). At $n = D = 512$, the oracle shallow/full ratios are 1.0833 and 1.0763, while median TwoNN/full ratios are 0.9504 and 0.9357. Consequently TwoNN remains about 12–13% below d_U. The controlled result supports Proposition D.8 for its stated target, without supplying a finite-sample correction for TwoNN.

H.3 PREDICTION PRESERVATION AND NUMERICAL CONTROLS

All five training runs reached epoch 50 and all planned cells were recorded. Held-out accuracy ranges from 92.225% to 93.213%. Every compensated map preserves all predictions, and the maximum logit discrepancy over every seed and map is 7.60×10^{-6}, below 10^{-4}. Across every reported ratio estimator, the largest relative change under an orthogonal map or uniform scaling is 0.155%, below the 2% threshold. These controls distinguish a geometric intervention from a classifier change and bound the numerical floor against which anisotropic effects are judged.

Table 5: PCA equalization changes every reported neighbor-ratio estimator, with median factors from 2.124 to 3.099 at $n = 8000$, relative to identity. The classifier remains function-equivalent in each case.

Estimator	Factor	Estimator	Factor
TwoNN / GRIDE $(1, 2)$	2.124	GRIDE $(2, 4)$	2.371
GRIDE $(4, 8)$	2.477	GRIDE $(8, 16)$	2.558
GRIDE $(16, 32)$	2.616	L2N2 asymptotic	2.411
L2N2 calibrated	3.099	Levina–Bickel local $k = 10$	2.578
Levina–Bickel local $k = 20$	2.622	Levina–Bickel pooled $k = 10$	2.435
Levina–Bickel pooled $k = 20$	2.517		

A post-hoc float64 recomputation attributes the small control effect to numerical precision. The complete sensitivity measurements, including filtered anchors and alternative numerator calculations, appear in Appendix K.3.

H.4 Random-basis intervention effects

Across identity and condition numbers 2, 4, and 8, the random-basis family has a median per-seed TwoNN fold range of 1.0097, below the predeclared 1.5-fold material-effect threshold. Its paired log effect exceeds the control floor in these five models: median 0.00967, with descriptive 95% seed-bootstrap interval $[0.00777, 0.02066]$. At condition eight, the null predicts a ratio of 0.991 to identity and the observed median is 0.990. A generic, randomly oriented, moderately conditioned map therefore need not produce a practically large change in this trained representation.

H.5 Covariance-aligned intervention effects

Compensated metric changes follow the intervention principle of Kornblith et al. (2019) and Cain (2026). Here PCA equalization has median condition number 8.142 and changes TwoNN by a median factor 2.124 (seed range 2.064–2.174) while satisfying the same function-equivalence checks. Every comparison estimator also changes, with median factors from 2.124 to 3.099 (Table 5). Across the five seeds, function-preserving coordinate changes therefore range from negligible to large. The matched-spectrum control in Appendix K.5 fixes each PCA scaling vector and varies only its basis. Across all five models, PCA approximately doubles TwoNN while the tested random bases remain near identity.

The spectral null distinguishes the weak random-basis and strong aligned effects. Across 30 seed-by-map measurements (random conditions $2, 4, 8, 16$ and both PCA maps within each of five models), predicted and observed log effects have Pearson correlation 0.994 and Spearman correlation 0.867. The null captures the relative ordering of interventions, but not their absolute magnitudes: for PCA equalization it predicts 3.338-fold versus 2.124-fold observed. Appendix L specifies the seed-by-map clustering used to interpret these descriptive correlations.

H.6 Coordinate effects on disjoint evaluation images

The approximately twofold effect survives fitting on different images. Across the five models, disjoint-fit PCA has median factor 2.1989 (range 2.127–2.221), compared with 2.3314 for same-size in-sample fits and 0.9952 for disjoint-fit matched random orientations (Figure 2a and Table 6). The disjoint/in-sample factor is 0.932–0.951 within each model: fitting on separate images moderates the response but does not remove it. Both comparisons use 4,000 evaluation points, so this difference is not attributed to changing sample size from the original 8,000-point experiment.

All 120 transformed clouds preserve every prediction; the maximum logit error is 7.49×10^{-6} and the singular-value matching error is below 10^{-10}. The result establishes a basis-sensitive response on disjoint evaluation images for these fitted maps and models. It does not establish a change in population dimension or generalization to other architectures.

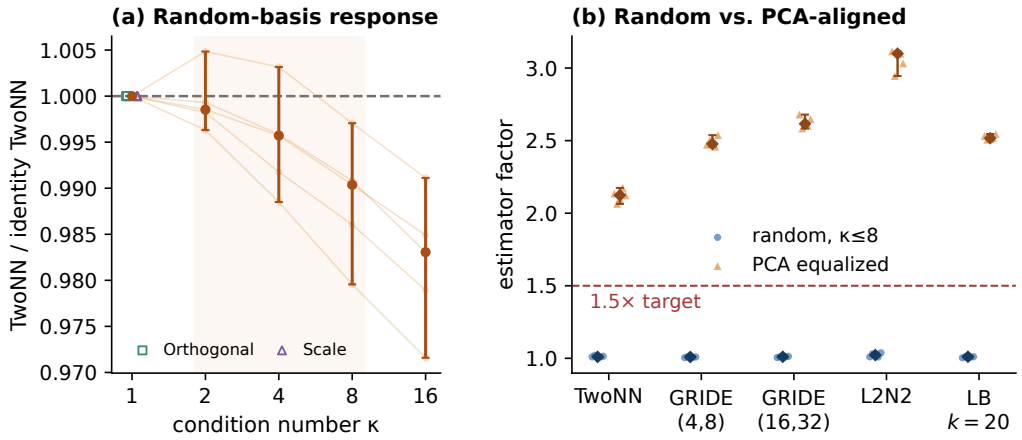


Figure 3: Function-preserving coordinate interventions. All predictions are preserved; the maximum compensated-logit error is $7.60 \times 10^{-6} < 10^{-4}$. (a) The orange series shows the small changes from random-basis maps; the dashed line marks identity. (b) PCA equalization approximately doubles TwoNN and increases the other estimators. Points and faint lines show five independently trained models; prominent markers and error bars show medians and descriptive 95% seed-bootstrap intervals. Panel (b) compares the random-family fold range with the PCA-to-identity factor.

Table 6: Disjoint-fit PCA on five trained models. Each factor divides transformed TwoNN by identity on the same 4,000 evaluation images. Geometric means combine the two complementary folds; random factors first take the median of five orientations per fold. Folds and orientations are repeated measurements within a model.

Seed	In-sample PCA	Disjoint PCA	Disjoint random	Disjoint/in-sample
270101	2.2928	2.1368	0.9927	0.9320
270102	2.3378	2.2085	0.9971	0.9447
270103	2.2760	2.1268	0.9931	0.9344
270104	2.3346	2.2209	0.9952	0.9513
270105	2.3314	2.1989	0.9973	0.9432

H.7 Gaussian prediction errors in unmodified representations

At epoch 50 and $n = 8000$, observed TwoNN is below the plug-in Gaussian null in every layer and seed. Median residuals range from -9.84% at the logits to -41.9% at layer 3 (Table 7). Every layer has the same negative sign in all five seeds, every seed has leave-one-out layer-order correlation 1.0, and the bootstrap interval excludes zero in every layer. The residual magnitude also exceeds the paired 4,000-to-8,000 change everywhere. The Gaussian spectral comparator leaves a negative, layer-dependent residual that repeats across all five seeds; we therefore use it as a structured comparator rather than a quantitatively calibrated predictor. Under the Gaussian shallow assumptions, classical covariance trace distortion lowers this baseline, so it cannot explain a negative residual by itself (Proposition D.9). That asymptotic direction does not calibrate the finite neural cloud.

Because global covariance includes within-class covariance and between-class mean variation, we measure those components and neighbor-label purity directly. The population decomposition is

$$\Sigma_{\text{global}} = \sum_c p_c \Sigma_c + \sum_c p_c (\mu_c - \mu)(\mu_c - \mu)^\top.$$

At layer 4, the first neighbor shares the anchor's class for 89.3% of the full-cloud points and both neighbors do so for 84.6% (Table 20). Between-class covariance contributes 72.4% of total variance, yet global participation ratio is 10.04 versus 16.11 for pooled within-class covariance (Table 21). Adding the positive-semidefinite between-class term therefore does not imply an increase

Table 7: Relative residual e from Equation (145) at epoch 50 and $n = 8000$. Intervals bootstrap five independent training seeds.

Representation	Median e	95% interval	Absolute 4k–8k change
Stem	-29.1%	$[-32.2, -28.3]\%$	1.76%
Layer 1	-36.5%	$[-37.3, -34.5]\%$	1.94%
Layer 2	-38.8%	$[-41.7, -37.3]\%$	1.71%
Layer 3	-41.9%	$[-42.8, -38.1]\%$	1.93%
Layer 4	-17.2%	$[-17.5, -16.1]\%$	0.712%
Logits	-9.84%	$[-11.3, -8.42]\%$	1.83%

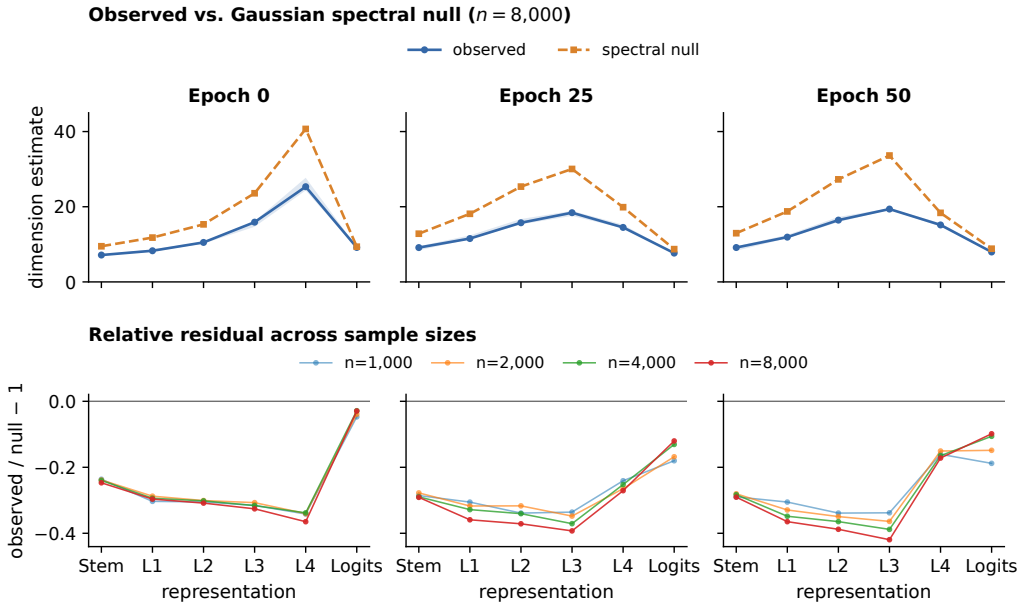


Figure 4: Observed TwoNN and its Gaussian spectral comparator across epochs and layers (top, $n = 8000$), and relative residuals across nested sample sizes (bottom). Each row uses a common vertical scale across epochs. The spectral prediction leaves a systematic negative residual at epoch 50 and $n = 8000$. Shading shows descriptive 95% seed-bootstrap intervals for median TwoNN across five trained models; residual curves are seed medians.

in participation ratio. Class purity and covariance decomposition motivate the direct calibration below, rather than determining its outcome.

H.8 Direct Gaussian and class-mixture calibration

Direct Gaussian measurements fall approximately 6–7% below the analytic phase computed from the same fitted population covariance. This is finite-sample formula error under a specified null. Comparing neural measurements with the simulated null still leaves median paired real/Gaussian ratios 0.658 in layer 3 and 0.907 in layer 4 (Table 2). The analytic discrepancy is therefore not wholly explained by that finite-sample calibration.

Class mixtures improve agreement at layer 3 but worsen it at layer 4; this direction holds for every trained model. Median paired real/mixture ratios are 0.745 and 0.794 (Table 18). Within-class measurements also remain lower than their mixture counterparts, with ratios 0.737 and 0.783 (Table 19). The mixture further overpredicts both-neighbor purity: 77.8% versus 65.9% in layer 3 and 96.0% versus 84.6% in layer 4 on the disjoint evaluation halves. Thus the measured class structure alone does not calibrate these local neighbor statistics.

The fixed 10% shrinkage sensitivity raises surrogate TwoNN to 31.42 and 24.92, worsening agreement in both layers. Fitting uncertainty remains: there are only 386–415 fitting points per class; layer-3 class covariances have rank 256, but layer-4 covariances have rank 385–414 in 512 coordinates. Monte Carlo SEs of the fitted-null means are 0.22–0.65% of those means, much smaller than the observed discrepancies, but do not include covariance uncertainty. The 16 surrogate draws support a conditional Monte Carlo precision estimate; we do not make a formal significance claim from them.

On the full neural clouds, d_U is close to the empirical shallow target: model-median ratios across the six representations range from 0.994 to 1.001. Their empirical participation ratios are only about 4–14. The large attenuation in the controlled high-effective-rank Gaussian study therefore does not automatically explain these neural residuals. The direct comparisons show departures from the particular fitted Gaussian and class-Gaussian nulls; non-Gaussian class shapes, local heterogeneity, and other geometry are not separated by this experiment.

I Discussion and limitations

I.1 Interpretation of the finite-radius target

The shared mathematical mechanism is a moving probability scale. A nearest neighbor radius corresponds roughly to probability $1/n$, not to an abstract infinitesimal neighborhood. The inverse-CDF slope at that probability determines the log-radius spacing and is the reciprocal of finite-radius continuous ID. In the Gaussian model this dimension profile has a classical ridge-like effective-df form whose scale is assigned endogenously by $1/n$; on the sphere it becomes a cap-hazard phase law. The two targets are different because the population geometries are different, but both depend on sample resolution. Pooled TwoNN is accurate for those targets in the canonical sweep, but its finite-sample output should not be called the population manifold dimension without qualification.

I.2 Coordinate transformations and intervention magnitude

The finite-cloud proposition and rise-then-fall construction establish that coordinate changes can alter TwoNN while leaving population dimension and function fixed. The matched-spectrum control shows that a large basis effect remains when singular values are held fixed in these five representations (Appendix K.5). Disjoint fitting preserves the large PCA effect (Section H.6), addressing fitting/evaluation coupling for these models. Generalization beyond this architecture and these data-adapted scaling vectors remains open.

For geometric intuition, let Q be Haar orthogonal, $A = Q\,\mathrm{diag}(a_j)Q^\top$, and $v \neq 0$ be fixed independently of Q. With $w_j = a_j^2$ and $S_v = \|vA\|^2/\|v\|^2$, spherical moments give

$$\mathbb{E}_Q S_v = \overline{w}, \qquad \frac{\mathrm{Var}_Q(S_v)}{(\mathbb{E}_Q S_v)^2} = \frac{2}{p+2}\left\{\frac{p\sum_j w_j^2}{(\sum_j w_j)^2} - 1\right\}.$$

When $p\sum_j w_j^2/(\sum_j w_j)^2$ stays bounded, each fixed vector's squared length concentrates around the same multiplier. Scale invariance then helps explain the small random-basis effects. This calculation does not provide uniform control of all pairwise lengths or nearest-neighbor rankings, and it fails as an explanation when the scaling spectrum is too concentrated. A PCA basis depends on the data and need not satisfy this fixed-vector random-orientation argument.

I.3 Manifold dimension and estimator resolution

The sphere theorem describes estimation on a known manifold. Even when data lie on a smooth compact manifold with known dimension, polynomially many samples can make pooled TwoNN report a vanishing fraction of that dimension as $d \to \infty$. A different estimator supplied with reach, density, noise, and sampling assumptions may consistently estimate manifold dimension (Farahmand et al., 2007; Kim et al., 2019; Block et al., 2022). Hence the manifold hypothesis alone does not guarantee that a chosen finite-sample neighbor statistic measures the manifold dimension.

I.4 Dimension labels for neural activations

For a canonical sphere or a rank-deficient Gaussian, the generating law names a dimension. For CIFAR-10 activations, there is no analogous observable label. With continuous noise the support can be full-dimensional; with a latent-noise model the scientifically relevant dimension depends on a scale and on the model class. A finite point cloud is compatible with many smooth latent descriptions. We therefore evaluate measurable invariance, stability, spectral prediction, and residuals on the neural data.

I.5 Assumptions and theorem limits

The convergence laws concern the pooled estimator in Equation (12), rather than Facco's empirical-CDF regression. The phase centers are asymptotically exact in the stated models. The inverse-CDF/Beta representation and finite-cloud no-go are finite-sample identities or statements; the phase laws and parameter-free neighbor-window approximations are first-order asymptotic statements. Proposition E.4 separately controls spherical mean calibration. Proposition D.8 bounds estimation of the shallow Gaussian target using classical trace moments. These bounds control the Gaussian target and spherical mean log spacing, not TwoNN's finite-sample bias, variance, or mean squared error. Bias correction for dependent ratios, including the role of the $n-1$ numerator, remains open.

The Gaussian theorem concerns a triangular array of iid centered Gaussian vectors with Euclidean distance, bounded covariance eigenvalues, nondegenerate first two spectral moments, and $\log n \to \infty$ with $\log n = O(D)$. Exact zero eigenvalues are allowed, with rank $\Theta(D)$ under these assumptions. The notation d_{eff} denotes the sample-selected finite-radius target. "Parameter-free" means that no coefficient is fitted once n, D, and the population spectrum are fixed; the neural plug-in uses an estimated spectrum.

The sphere theorem concerns iid uniform points, chordal Euclidean distance, and a named joint growth regime. It is not a minimax impossibility theorem for all manifold estimators. The finite-cloud no-go theorem concerns exact scale invariance and universal samplewise monotonicity for scalars of the induced metric cloud. Extensions involving guarantees in expectation, restricted map classes, task-aware quantities, derivatives, cross-layer coupling, and architecture-specific definitions require separate analysis. The ReLU realization allows width to depend on the finite cloud; learning such a realization by SGD is a separate question.

I.6 Empirical limits and next steps

The neural experiment uses one dataset, one architecture family, five seeds, and one primary intervention site. The same-cloud covariance plug-in is a structured null whose consistency is unproved in the jointly growing neural setting. Its residual should not be interpreted as missing manifold dimension. The intervals summarize variation across the five training seeds.

The additional exploratory analyses separate fitting from evaluation and test class-conditioned nulls. They do not change the prespecified random-basis stop rule. They do not add independent trained models or an untouched neural dataset. Class covariances are estimated from small samples, and the shrinkage sensitivity changes the modeled distribution substantially. The neural residual is robust to the tested Gaussian calibration, but its cause remains unidentified; class structure alone does not resolve it in both layers.

The controlled trace study validates estimation of d_{PR}, not an exact finite-sample predictor of TwoNN. Further work should distinguish non-Gaussian radial and angular structure, assess other architectures, and obtain shared-cloud concentration rates and deep-resolution plug-in guarantees.

J Conclusion

We derived what pooled shared-cloud TwoNN and fixed-order neighbor ratios measure in two high-dimensional models. Under the stated Gaussian assumptions, TwoNN converges to an explicit sample- and spectrum-dependent finite-radius ID with a classical effective-df algebraic form. Its scale is selected by the nearest-neighbor tail probability rather than tuned externally. On the sphere TwoNN follows a separate cap-resolution law. Pooled TwoNN requires exponentially many samples

to resolve a fixed fraction of dimension on both models. The shared-cloud reuse does not remove this boundary: two-anchor extreme-process separation makes the ordinary aggregate concentrate around its sample-resolution target.

The canonical experiments support the TwoNN law and identify a later-neighbor resolution shift addressed by the neighbor-window approximation. Controlled Gaussians validate the trace-moment shallow target while exposing its remaining finite-resolution difference from observed TwoNN. In neural representations, the approximately twofold PCA effect survives disjoint fitting and preserves the classifier, while matched random orientations stay near identity. Direct Gaussian and class-mixture calibration leaves model-dependent residuals, including an unfavorable layer-4 mixture comparison. A layerwise number therefore requires interpretation together with its sample resolution, metric, fitted null, and sensitivity to function-preserving coordinate changes.

K Implementation and reproducibility details

K.1 Canonical computation

Pairwise Euclidean distances are computed in float64 blocks, with self-distances removed and the first 32 neighbors retained. Float64 is used because float32 can create artificial nearest-neighbor ties in these continuous high-dimensional models.

K.2 Prespecified decision rules and statistical analysis

The training seed is the independent unit. All reported confidence intervals are percentile intervals from 10,000 seed-level bootstrap samples. The retained neural protocol specified the following decision rules.

1. Every compensated map must have prediction agreement one and maximum absolute logit error at most 10^{-4}.
2. Every ratio estimator must change by at most 2% under orthogonal and uniform-scale controls.
3. For each seed, the TwoNN fold range across identity and random-basis conditions 2, 4, and 8 is computed. The median must be at least 1.5, and its paired log effect must exceed the control effect.
4. Natural residual signs must agree in at least four of five seeds for at least four layers, or leave-one-seed-out layer ordering must have Spearman correlation at least 0.6 for four seeds.
5. At headline layers, the magnitude of the 8,000-sample residual must exceed the paired change from 4,000 to 8,000 samples.

The protocol also declared a stop rule: dataset and architecture expansion would not proceed if the primary random-basis material-effect gate failed. The stop rule keeps the secondary PCA intervention separate from the primary random-basis hypothesis.

K.3 Numerical sensitivity measurements

A post hoc numerical sensitivity audit recomputes the relevant cells without changing any prespecified threshold or estimator definition. The audit counts every filtered first/second-neighbor ratio in all 540 neural cells and recomputes the complete epoch-50, layer-4, $n = 8000$ identity, orthogonal, and uniform-scale cells for all five seeds. It uses float64 distances first on the archived rounded clouds and then on transformations and distances kept in float64 throughout. It also replaces the declared MacKay numerator $n - 1$ by $N_{\mathrm{valid}} - 1$ to measure the effect of numerical filtering.

The post-hoc precision audit identifies the numerical origin of the 0.155% control effect. Across 2,025,000 anchor–cell instances, 44 (0.00217%) in 33 of 540 cells have tied first and second float32 distances; no nearest distance is zero. Replacing $n - 1$ by $N_{\mathrm{valid}} - 1$ changes any affected TwoNN value by at most 0.1001%. Across all reported estimators in the fifteen primary negative-control cells, the maximum control effect falls from 0.155% in the archived float32 pipeline to 0.000103% when only distance computation is promoted to float64, and to 8.36×10^{-11}% when the coordinate transform and distances are both float64. No first/second tie remains in those float64 cells, and the

Table 8: Prespecified neural decision rules and observed measurements. Prediction preservation and numerical controls meet their tolerances. The random-basis family falls short of the 1.5-fold effect-size threshold, while natural residual signs, ordering, and resolution comparisons replicate.

Decision rule	Threshold	Observed value	Verdict
Prediction agreement	agreement $= 1$	1.000	pass
Maximum compensated-logit error	$\leq 10^{-4}$	7.60×10^{-6}	pass
Negative-control estimator change	$\leq 2\%$	0.155%	pass
Random-basis TwoNN fold range	≥ 1.5	1.0097	fail
Stable natural residual sign/order	at least four layers/seeds	all six / all five	pass
Residual exceeds 4k–8k change	headline layers	all six	pass

maximum float32-to-float64 change in TwoNN itself is 0.000164%. These discrepancies are far below the prespecified control thresholds and headline intervention effects, so the float32 artifact does not change any reported decision-rule verdict or the stated intervention conclusions.

K.4 Neural computation

The five-seed confirmatory campaign ran on one NVIDIA RTX 5090 with PyTorch, and all seeds reached epoch 50. A separately labeled post hoc precision analysis compared the original float32 results with rounded-cloud and full-float64 recomputations for all five seeds in the primary $n = 8000$ cell. One launch exceeded the system's Unix-socket path limit before producing analyzable model output; it is outside the five completed epoch-50 runs. A post-unblinding reporting-only export changed no decision threshold, exclusion, or verdict.

K.5 Computational reproduction and matched-spectrum control

We reproduced the complete canonical grid from its original seeds and recomputed all 360 natural and 180 original neural intervention cells from the saved activations of the five trained networks. This is computational reproduction, not additional independent training replication. Every canonical estimate reproduced without a numerical difference. For computational reproduction, both formulas are evaluated on the same previously designated 200-record held-out central subset; this reuse does not constitute a new holdout.

For the neural rerun, covariance calculations, transformations, compensated classifiers, and nearest-neighbor distances use float64 arithmetic on the archived activations. The original estimator formulas and filtering conventions remain fixed. Across the 540 repeated cells, the relative change in pooled TwoNN from the original float32 pipeline is at most 0.00456%; among epoch-50, $n = 8000$ cells it is at most 0.000860%. Across all 12 estimator outputs the maximum change is 1.51%, so the TwoNN precision bound should not be attributed to every comparator. The random-family fold range remains 1.0097, and the six natural residual signs, layer ordering, and sample-size stability conclusions are unchanged.

We additionally evaluated an exploratory control that holds the complete scaling spectrum fixed while changing its basis. For each network, use the original PCA equalization scales a_j defined above and compare

$$A_{\text{aligned}} = V \operatorname{diag}(a_j) V^\top, \qquad A_{\text{random},b} = Q_b \operatorname{diag}(a_j) Q_b^\top, \quad b = 1, \ldots, 5.$$

The deterministic QR construction uses base seeds 918001–918005, with the same dimension offset $1009p$ as the original helper. These bases are reused across the five models. All maps are evaluated at epoch 50, layer 4, and $n = 8000$; all 25 random-orientation measurements are retained. The largest absolute singular-value mismatch within a matched comparison is below 2.45×10^{-15}. All classifier predictions are preserved; the largest logit discrepancy across the original and additional rerun maps is below 7.50×10^{-6}.

The median PCA factor is 2.1235, the median of within-network random medians is 0.9978, and the median paired PCA/random factor is 2.1282 (Table 9). This control isolates a basis effect in these

Table 9: Exploratory matched-spectrum control on five trained models. Factors divide TwoNN by its identity value. Each random median and range summarizes five orientations within one model; these orientations are not independent training replicates.

Seed	PCA factor	Random median	Random range	PCA/random median
270101	2.1350	0.9982	[0.9896, 1.0030]	2.1388
270102	2.0639	0.9976	[0.9937, 0.9996]	2.0689
270103	2.0981	0.9968	[0.9882, 0.9987]	2.1050
270104	2.1743	0.9984	[0.9973, 1.0015]	2.1778
270105	2.1235	0.9978	[0.9904, 1.0044]	2.1282

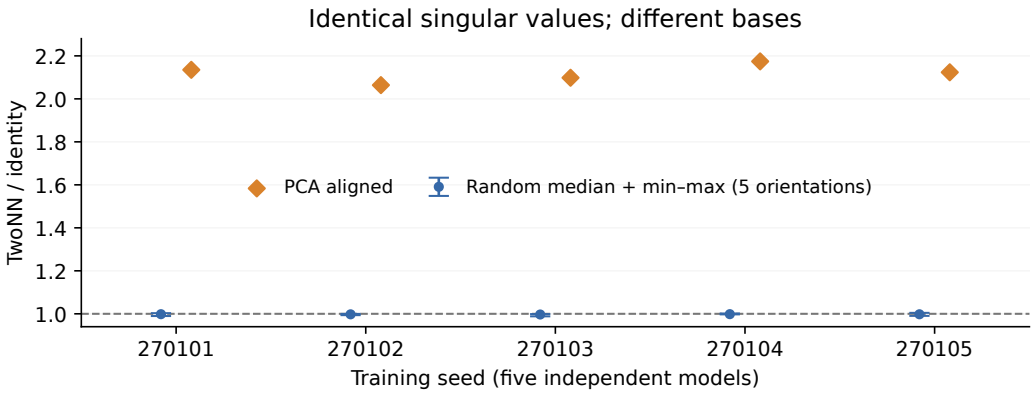


Figure 5: TwoNN under maps with identical singular values and different bases. Each training seed identifies one independent model. Circles show the median of five random orientations, with bars spanning their minimum and maximum (not confidence intervals); diamonds show the PCA-aligned value. The dashed line marks identity. All predictions remain unchanged. Random orientations remain near identity, while PCA alignment produces an approximately twofold increase in every model.

five representations. Generalization beyond these models and data-adapted scaling vectors remains open. The matched-spectrum control is exploratory and was not part of the prespecified thresholds or confirmatory outcomes.

K.6 Per-seed training outcomes

Table 10: Final training accuracy and loss, with held-out accuracy.

Seed	Train accuracy	Held-out accuracy	Final loss
270101	0.99924	0.924000	0.29231
270102	0.99842	0.922250	0.31061
270103	0.99932	0.923125	0.28982
270104	0.99920	0.924625	0.28690
270105	0.99970	0.932125	0.26204

L Estimator definitions and statistical summaries

L.1 Neighbor-ratio estimators

For anchor i, let $R_{i,j}$ be its jth nearest-neighbor radius after excluding itself. MacKay TwoNN is Equation (12). For fixed $k < \ell$, we use the GRIDE-based log-moment estimator in Equation (96), based on the ratio law of Denti et al. (2022); its probability spacing has the beta law in Equation (17). This aggregate matches the mean log ratio, whereas the general GRIDE maximum-likelihood estimator solves a nonlinear score equation. Throughout the empirical results, the label GRIDE denotes this GRIDE-based log-moment estimator. We report all five adjacent doubling windows through neighbor order 32.

For the Levina–Bickel estimator, define the bias-corrected local score

$$\widehat{d}_{i,k} = \frac{k-2}{\sum_{j=1}^{k-1} \log(R_{i,k}/R_{i,j})}. \tag{148}$$

The local version averages $\widehat{d}_{i,k}$ over anchors. The pooled version uses $N(k-1)$ divided by the sum of all retained anchor spacings, where N is the number of valid anchors. We use $k = 10, 20$.

For each GRIDE ratio window and TwoNN separately, nonpositive or nonfinite computed log ratios are discarded; fewer than $\max(32, \lfloor n/2 \rfloor)$ valid anchors makes that estimate missing. L2N2 uses the same $(1, 2)$ validity rule. Levina–Bickel requires all local spacings to be finite and their sum to be positive, and is missing if no valid anchor remains. Across 540 confirmatory clouds, 33 $(1, 2)$ cells discard a total of 44 anchors because float32 makes the first two positive distances equal. This is 0.00217% of 2,025,000 anchor–cell instances; the minimum retained fraction is 99.9%, no cloud contains a zero nearest distance, and no estimator cell is missing, infinite, or nonpositive. MacKay TwoNN retains its declared numerator $n - 1$ after this numerical filtering. The post-hoc alternative $(N_{\text{valid}} - 1)/\sum_{i \in \text{valid}} L_i$ differs by at most 0.1001% in an affected cell. Full-float64 recomputation of all five complete primary negative-control triples eliminates these ties and reduces the maximum control effect over every reported ratio estimator to 8.36×10^{-13} in relative units.

For L2N2, put

$$\overline{L} = \frac{1}{n} \sum_{i=1}^{n} \left[-\log \left\{ \log(R_{i,2}/R_{i,1}) \right\} \right]. \tag{149}$$

The asymptotic version is $\exp(\overline{L} - \gamma)$, where γ is Euler's constant. The finite-sample version is $\exp\{(\overline{L} - \beta_n)/\alpha_n\}$ using the published calibration table of Ong et al. (2026). Sizes 1,000 and 2,000 use exact table entries; 4,000 and 8,000 use log-n interpolation between the published 2,000, 5,000, and 10,000 entries. The (α_n, β_n) values are $(0.732002, 1.165663)$, $(0.752371, 1.131156)$, $(0.7649427882, 1.1147602519)$, and $(0.7779242754, 1.0948234416)$ at the four analyzed sizes.

L.2 Empirical spectral prediction

For activation rows centered into $Z_c \in \mathbb{R}^{n \times p}$, the same-cloud covariance is $\widehat{\Sigma} = Z_c^\top Z_c/(n-1)$. We use float64 `torch.linalg.eigvalsh`, with eigenvalues in ascending order and negative numerical values clamped to zero. If $p > n$, the implementation uses $Z_c Z_c^\top/(n-1)$ and appends $p - n$ zeros. We solve Equation (26) numerically and evaluate Equation (27). Zero eigenvalues are kept. For interventions, predicted log effects are

$$\log d_{\text{null}}(A) - \log d_{\text{null}}(I), \tag{150}$$

and observed log effects replace d_{null} by TwoNN. Correlations are descriptive secondary statistics over 30 seed-by-map pairs: six maps are repeated within each of five trained models: the four random conditions $2, 4, 8, 16$ and both PCA maps, excluding identity and the two controls. These are not 30 independent experimental units. They do not replace the frozen fold-range gate.

L.3 Bootstrap and stability calculations

All inferential resampling treats the five training seeds as exchangeable units. A bootstrap draw samples five seeds with replacement and recomputes the declared median or paired effect. For natural

layers, sign stability is counted before bootstrapping. Leave-one-seed-out ordering compares each held-out seed's six-layer residual vector with the median vector of the other four using Spearman correlation. Resolution stability compares $|e_{8000}|$ with the paired absolute change $|e_{8000} - e_{4000}|$.

M EXPANDED EMPIRICAL TABLES

M.1 CANONICAL HELD-OUT ERRORS

Table 11: Fresh central neighbor-window median absolute relative error by model and neighbor pair.

Model	Error	Neighbor pair	Error
Isotropic Gaussian	5.44%	$(1, 2)$	3.97%
Two-block Gaussian	7.43%	$(2, 4)$	4.71%
Geometric spectrum	6.00%	$(4, 8)$	5.50%
Rank-$D/2$ Gaussian	4.05%	$(8, 16)$	6.25%
Sphere	3.68%	$(16, 32)$	6.58%

M.2 PER-SEED PRIMARY INTERVENTION

Table 12: Per-seed random-basis TwoNN fold ranges and paired effects. Fold range uses identity and conditions 2, 4, and 8. The control column is the largest absolute log effect of the orthogonal and uniform-scale controls.

Seed	Fold range	Control log effect	Paired log effect
270101	1.009722	3.33×10^{-7}	0.009675
270102	1.020872	1.25×10^{-6}	0.020656
270103	1.007805	5.81×10^{-7}	0.007774
270104	1.009273	2.49×10^{-7}	0.009230
270105	1.014152	2.79×10^{-6}	0.014050

M.3 EXPANDED COORDINATE-INTERVENTION COMPARISON

Table 13: Expanded coordinate-intervention comparison at epoch 50, layer 4, $n = 8000$. Entries retain the six-decimal factors underlying the rounded original-study estimates. Factors are relative to identity; random-basis condition numbers are prescribed, and PCA condition numbers are seed medians. PCA equalization produces the largest observed response, which the Gaussian prediction overstates. Condition 16 is outside the primary comparison.

Map	$\kappa(A)$	Observed factor	Spectral factor
Orthogonal control	1.000	1.000000	1.000000
Uniform scale	1.000	0.999999	1.000000
Random basis, $\kappa = 2$	2.000	0.998538	1.000338
Random basis, $\kappa = 4$	4.000	0.995718	0.997811
Random basis, $\kappa = 8$	8.000	0.990372	0.990879
Random basis, $\kappa = 16$	16.000	0.983055	0.981486
PCA equalization	8.142	2.123535	3.338461
Equalization + tail contraction	9.965	1.138066	1.392996

Table 14: Epoch-50 layer-4 effects at $n = 8000$. Factors are relative to the identity map; the random range uses identity and $\kappa = 2, 4, 8$. Values are displayed to six decimals.

Seed	PCA factor	Random range	Maximum logit error
270101	2.135018	1.009722	7.589362×10^{-6}
270102	2.063914	1.020872	6.913943×10^{-6}
270103	2.098136	1.007805	6.937368×10^{-6}
270104	2.174280	1.009273	7.600155×10^{-6}
270105	2.123535	1.014152	7.037943×10^{-6}

Table 15: Per-seed natural relative residuals at epoch 50 and $n = 8000$, in dimensionless units. The full-precision export also reports each TwoNN value and its same-cloud null.

Seed	Stem	Layer 1	Layer 2	Layer 3	Layer 4	Logits
270101	-0.286298	-0.364758	-0.411171	-0.428426	-0.174997	-0.095692
270102	-0.282645	-0.362787	-0.372635	-0.380785	-0.168123	-0.084197
270103	-0.290515	-0.373021	-0.416510	-0.427833	-0.174892	-0.098437
270104	-0.314825	-0.372411	-0.387884	-0.408581	-0.160699	-0.112828
270105	-0.321709	-0.344930	-0.383038	-0.419441	-0.171962	-0.106621

M.4 Per-seed intervention factors and natural residuals

M.5 Natural residual stability

This extends the covariance-matched Gaussian-surrogate comparison of Ansuini et al. (2019, Section 3.3) using the analytic target. At epoch 50 and $n = 8000$, all six residual signs are negative for every seed. The seed-median absolute changes between 4,000 and 8,000 samples are 0.0176, 0.0194, 0.0171, 0.0193, 0.00712, and 0.0183 for stem through logits, respectively; each is smaller than the corresponding residual magnitude. Every seed's leave-one-out six-layer rank correlation is 1.0.

N Additional exploratory protocols and full results

For these additional analyses, the models, layers, splits, covariance sensitivity, seeds, and replication counts were fixed before measurement. The same five epoch-50 models and saved activations are used throughout. No network was retrained, no clipping constant or prespecified decision rule was changed, and no outcome was excluded. Prespecified and exploratory analyses are documented separately. Five models remain the units of neural replication.

N.1 Splits and coordinate maps

For each model, a pseudorandom generator with a distinct fixed seed shuffles image positions within each class. One half receives $\lfloor n_c/2 \rfloor$ from each class; randomly allocated odd-count remainders make its total exactly 4,000. Both halves are then independently shuffled using the same generator state. The retained positions and original image indices verify complete, disjoint coverage of the 8,000 images. Fitting on each half and evaluating on the other gives two complementary folds, whose results are dependent within one model.

At layer 4, both disjoint and evaluation-half fits use the original lower-median reference variance, $10^{-4}\lambda_{\max}$ floor, and $[0.2, 5]$ clipping rule. Random bases use base seeds 918001–918005 and the original dimension offset $1009p$; these bases are reused across models. Every map uses its own fitting mean in Equation (144). Its inverse may be absorbed explicitly into the readout as

$$(W')^\top = A^{-1}W^\top, \qquad b' = b + \mu W^\top - \mu A^{-1}W^\top.$$

All singular-value matching errors are below 10^{-10}, every prediction agrees, and the maximum logit discrepancy is 7.49×10^{-6}. Table 6 reports every model-level factor.

The follow-ups use CPU float64 distances, covariance estimates, maps, and trace statistics on the archived float32 activations. Self-neighbors are excluded. TwoNN keeps numerator $n - 1$ and the original positive-finite-log-spacing rule; no recorded main or real class cell discards an anchor, and no zero nearest distance occurs.

N.2 Controlled trace-target grid

The controlled grid uses isotropic and two-block spectra with $n/D = 1/2, 1, 2$. All 448 clouds use distinct random seeds. The pairing used by d_U follows row order. Table 16 reports every target-estimation cell; Table 17 separately reports empirical-full, shallow-full, and TwoNN comparisons. No accuracy gate or replication count was changed after seeing results.

Table 16: Controlled Gaussian target estimation. Entries are medians over R independent clouds; error is the median absolute relative error of d_U against the oracle shallow target. Plug-in denotes the empirical-covariance shallow target divided by its oracle value. The attenuation column is the asymptotic reference, not an exact finite-dimensional prediction.

Spectrum	D	n/D	R	d_U/shallow	Error	Plug-in	Attenuation
Isotropic	128	0.5	32	0.9956	2.40%	0.5732	0.5774
Isotropic	128	1	32	1.0063	1.19%	0.7043	0.7071
Isotropic	128	2	32	0.9988	0.49%	0.8155	0.8165
Isotropic	512	0.5	32	0.9982	0.52%	0.5760	0.5774
Isotropic	512	1	32	1.0001	0.24%	0.7063	0.7071
Isotropic	512	2	32	0.9997	0.13%	0.8160	0.8165
Isotropic	1024	1	16	0.9998	0.13%	0.7066	0.7071
Isotropic	2048	1	16	1.0005	0.07%	0.7069	0.7071
Two-block	128	0.5	32	1.0109	2.67%	0.7378	0.7412
Two-block	128	1	32	1.0036	0.92%	0.8401	0.8421
Two-block	128	2	32	1.0008	0.82%	0.9089	0.9109
Two-block	512	0.5	32	0.9983	0.49%	0.7394	0.7412
Two-block	512	1	32	1.0009	0.38%	0.8414	0.8421
Two-block	512	2	32	0.9997	0.13%	0.9103	0.9109
Two-block	1024	1	16	0.9994	0.20%	0.8415	0.8421
Two-block	2048	1	16	1.0001	0.04%	0.8418	0.8421

N.3 Fitted-null simulation and Monte Carlo precision

Each fitting covariance uses divisor $n_{\mathrm{fit}} - 1$ and negative numerical eigenvalues are clamped to zero. Global-Gaussian simulations may be generated in the covariance eigenbasis because translations and orthogonal rotations leave TwoNN unchanged. For mixtures, draws are transformed back into the common feature coordinates and shifted by their class means. Evaluation class counts are fixed; these are Gaussian-mixture simulations conditional on those counts, rather than random multinomial class assignments.

Each model, fold, layer, and replicate uses a distinct random stream. The global Gaussian and class-mixture draws use separate streams. The shrunk and unshrunk mixtures reuse the same standard normals, with fixed covariance shrinkage (147). Thus 960 clouds use 640 distinct random streams; 320 mixture pairs intentionally share normals.

For each fitted family the reported Monte Carlo SE is its sample standard deviation over 16 draws divided by four. Relative SE ranges are 0.219–0.627% for global Gaussians, 0.239–0.647% for class mixtures, and 0.245–0.614% for shrunk mixtures. These quantify simulation precision conditional on the fitted parameters. They omit fitting uncertainty and between-model variation. Every real value is below all 16 surrogate values in its corresponding cell, but we report no formal significance from

Table 17: Separating target estimation, the shallow approximation, and observed TwoNN on the same controlled clouds as Table 16. Full phase uses the exact population spectrum in the saddle equation. Each column reports a cell median.

Spectrum	D	n/D	Empirical full/ oracle full	Oracle shallow/ oracle full	TwoNN/ oracle full	TwoNN/ d_U
Isotropic	128	0.5	0.5612	1.1418	0.9139	0.7967
Isotropic	128	1	0.7182	1.1548	0.9446	0.8251
Isotropic	128	2	0.8397	1.1670	0.9382	0.8023
Isotropic	512	0.5	0.5690	1.0782	0.9347	0.8660
Isotropic	512	1	0.7132	1.0833	0.9504	0.8758
Isotropic	512	2	0.8288	1.0881	0.9460	0.8700
Isotropic	1024	1	0.7115	1.0610	0.9323	0.8822
Isotropic	2048	1	0.7104	1.0447	0.9491	0.9089
Two-block	128	0.5	0.7192	1.1154	0.9020	0.8055
Two-block	128	1	0.8390	1.1223	0.9076	0.8112
Two-block	128	2	0.9146	1.1281	0.9253	0.8161
Two-block	512	0.5	0.7319	1.0723	0.9333	0.8709
Two-block	512	1	0.8432	1.0763	0.9357	0.8711
Two-block	512	2	0.9145	1.0800	0.9457	0.8741
Two-block	1024	1	0.8435	1.0581	0.9374	0.8862
Two-block	2048	1	0.8433	1.0437	0.9570	0.9178

these comparisons. The class-covariance rank limitation and shrinkage sensitivity remain part of the interpretation. Table 18 includes every model rather than selecting a favorable layer or seed.

Table 18: Disjoint-fit null calibration by trained model. Surrogate means use 16 clouds of 4,000 points; geometric means combine the two folds. Ratios compare real TwoNN with direct Gaussian, class-mixture, and fixed 10% shrunk-mixture means. The final column compares direct Gaussian measurements with the fitted analytic phase.

Layer	Seed	Real/Gauss.	Real/mix.	Real/shrunk	Gauss./phase
L3	270101	0.6529	0.7461	0.6179	0.9338
L3	270102	0.6868	0.7714	0.6365	0.9349
L3	270103	0.6364	0.7257	0.6029	0.9332
L3	270104	0.6641	0.7453	0.6192	0.9301
L3	270105	0.6583	0.7435	0.6195	0.9320
L4	270101	0.9165	0.8008	0.5865	0.9360
L4	270102	0.8922	0.7921	0.5833	0.9279
L4	270103	0.9028	0.7923	0.5810	0.9350
L4	270104	0.9078	0.8027	0.5806	0.9404
L4	270105	0.9074	0.7942	0.5780	0.9359

N.4 Natural class structure and covariance decomposition

Equation (146) uses class-specific candidate counts and pools log spacings. Its within-class residual is not the unrestricted residual with a different denominator. Table 20 reports both neighbor purity and the class-restricted comparison across all six layers. The exact finite-sample covariance decomposition uses

$$\widehat{\Sigma}_{\text{within}} = \sum_c \frac{n_c - 1}{n - 1} \widehat{\Sigma}_c, \qquad \widehat{\Sigma}_{\text{between}} = \sum_c \frac{n_c}{n - 1} (\widehat{\mu}_c - \widehat{\mu})(\widehat{\mu}_c - \widehat{\mu})^\top,$$

Table 19: Class-restricted direct comparisons on disjoint evaluation halves. Surrogate values average 16 draws, then use geometric fold means and model medians. Purity instead uses arithmetic fold means. Ratios are aggregated within each model, so their medians need not equal ratios of marginal medians.

Layer	Real within	Mixture within	Real/mixture	Real both purity	Mix. both purity
L3	18.073	24.602	0.737	65.92%	77.76%
L4	13.961	18.118	0.783	84.61%	96.00%

with $\widehat{\Sigma} = \widehat{\Sigma}_{\text{within}} + \widehat{\Sigma}_{\text{between}}$. Table 21 retains both the variance decomposition and participation ratios. The full-cloud d_U/empirical-shallow ratios are 0.999, 0.994, 1.001, 0.999, 1.000, and 1.000 from stem through logits (model medians). They are descriptive neural quantities; the Gaussian guarantee is not assumed to transfer.

Table 20: Natural class diagnostics on the full 8,000-image clouds, summarized by medians across five models. Purity is the fraction of anchors whose first neighbor, or both first two neighbors, share its class. The last column uses class-restricted neighbors and class-count-specific phase predictions; it is a different statistic from unrestricted TwoNN.

Layer	TwoNN	Global phase	First purity	Both purity	Within residual
Stem	9.149	12.960	30.12%	11.79%	-30.12%
L1	11.913	18.754	39.24%	19.55%	-33.80%
L2	16.441	27.282	52.45%	33.91%	-35.16%
L3	19.382	33.624	78.97%	67.77%	-37.05%
L4	15.157	18.372	89.28%	84.64%	-32.46%
Logits	7.961	8.830	88.80%	84.41%	-5.93%

Table 21: Covariance decomposition on the full neural clouds. Entries are medians across five models. Participation ratio (PR) is $(\operatorname{tr}\Sigma)^2/\operatorname{tr}(\Sigma^2)$; it need not increase when the positive-semidefinite between-class term is added.

Layer	Between/total trace	Global PR	Within PR	Between PR
Stem	0.1721	4.1133	4.4805	1.7155
L1	0.2161	6.5471	8.1761	2.0803
L2	0.2088	10.8603	13.4842	3.3020
L3	0.3656	14.1490	24.4745	4.5452
L4	0.7237	10.0387	16.1126	8.1159
Logits	0.7924	7.2275	8.1534	6.7127

O THEOREM SCOPE AND DEPENDENCY GUIDE

The theoretical package contains three independent branches. The Gaussian branch starts with the exact inverse-CDF representation, establishes a uniform tilted local limit and moving-saddle hazard, proves a two-anchor rare-event gap, and combines Poisson separation with an exchangeable-pair criterion to obtain the shared-cloud law. Its corollaries state the spectral phase diagram, participation-ratio expansion, sharp isotropic boundary, fixed-order law, and neighbor-window approximation. Continuity of the spectral phase establishes the strict ordering in the rise-then-fall theorem and its function-equivalent network corollary.

The sphere branch replaces Gaussian saddlepoints by spherical-cap hazards and two-cap separation, then uses the same extreme-process and exchangeability logic. The finite-cloud branch is independent:

scale comparison proves Proposition F.1 and Corollary F.2, while finite ReLU interpolation shows that the relevant cloud maps can be realized after uniform shrinkage.

These dependencies delimit the scope of each claim. Fixed-anchor extreme processes become independent; the entire nearest-neighbor graph is not claimed to do so. The isotropic and spherical sample thresholds are estimator-specific, not minimax lower bounds for dimension estimation. The growing-order expression is a Gaussian-null interpretation and is not used as a headline empirical claim.