RECOVERING WEIGHTED TANGENT GEOMETRY FROM A SINGLE-SCALE SCORE FIELD

Anonymous authors
Paper under double-blind review

ABSTRACT

Score fields learned by diffusion models offer a direct probe of local data geometry, but the usual tangent-space description is insufficient at branch points, where the first-order object is a weighted measure over branch directions. Recovering this geometry matters because junctions arise when trajectories merge, road segments intersect, or strata meet, and their directions and relative masses determine what a learned field represents. We ask whether a score field at one known noise level can recover the branch center, local homogeneity degree, directions, and relative masses. This inverse problem is difficult because the observed score is an uncentered vector field, a one-point Hessian cannot distinguish even regular junctions with different branch counts, and ordinary score accuracy need not preserve the angular information required for geometry recovery. We address these challenges by using homogeneity under Gaussian smoothing to obtain a linear, score-only calibration of the unknown center and degree. Once centered, the tangential score on one shell integrates to a scalar transform with strictly positive spherical-harmonic multipliers. This yields exact identification of the normalized angular measure in any ambient dimension $D \geq 2$ and constructive recovery of finite positive rays. In the plane, it further yields sharp moment requirements, finite-query certificates, and an $O(\sigma^{\beta})$ tangent-score approximation for finite $C^{1,\beta}$ branches. Experiments apply the full chain to population, kernel-density, and learned score fields, showing across four geometries that lower validation normalized-score error can coincide with higher angular-moment error, thereby distinguishing ordinary score fit from recoverable junction geometry.

1 INTRODUCTION

The score $s_\sigma(x) = \nabla_x \log(\mu * \varphi_\sigma)(x)$ is a local vector field used by score-based generative models (Song et al., 2021). Near a smooth manifold, its normal component and Jacobian reveal tangent and normal spaces (Stanczuk et al., 2024; Ventura et al., 2025). This description is incomplete wherever the support is not smooth, as it is at junctions and along strata boundaries: no single tangent space describes such a point. The first-order object is instead a normalized angular tangent measure, whose atoms record branch directions and their relative shares of local mass under the data measure. Recovering this object reveals what a score field represents at a candidate junction, enabling a local audit of trained models and a geometric reading of learned fields.

We call the local scaling exponent the homogeneity degree, the local measure dimension in our tangent model. At one known noise level we ask:

> Can local score queries recover a branch point, its homogeneity degree, and the weighted tangent geometry represented by the field?

Here one noise level means one available field rather than a controlled multiscale family, so the method applies whenever a single score slice is at hand; training and sampling may still use a full schedule.

The question is hard for three reasons that map to our contributions. The score is an uncentered vector field with an unknown reference point, so scalar-moment machinery has no centered density and the center and degree must be recovered first (Contribution 1). At a conic vertex the score Hessian

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.

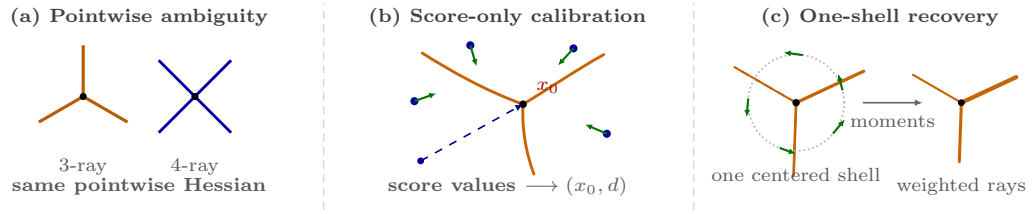

Figure 1: Single-scale recovery: (a) a pointwise score Hessian cannot distinguish junctions; (b) score values calibrate center and homogeneity; and (c) tangential score on one centered shell recovers directions and weights through angular moments.

depends only on the first two angular moments, so every uniform regular planar q-ray junction with $q \geq 3$ shares the same vertex Hessian, and the obvious local statistic is blind to the structure we want (Contribution 2). Real fields come from finite samples or networks, and a low denoising objective certifies no center, degree, count, direction, or mass (Contributions 3–4).

This differs from detecting a singularity in raw samples (von Rohrscheidt & Rieck, 2023; Lim et al., 2025), which can identify a neighborhood worth auditing and estimate sample geometry, but not what a trained score field represents there. We start from a local query window, not an exactly supplied center; global singularity search remains separate.

The blind Hessian has a partial exception. A separate trace identity encodes homogeneity, and its weak form lets a spatial score pattern calibrate the unknown center and degree without differentiating the score; once centered, one normalized shell carries the remaining angular information, encoding more than the first two angular moments so different counts stay distinguishable away from the vertex. The recovery chain then integrates tangential queries on the calibrated shell into a normalized scalar density whose harmonic coefficients give angular moments, reconstructs the branches with a finite-rank moment pencil, and reports a candidate, a certified stage, or an abstention. These steps convert vector log-gradient observations into the centered scalar-moment interface spectral estimators assume. Figure 1 summarizes this problem-to-solution path, while Figure 2 isolates the one-point obstruction and its shell conditioning.

For an exact homogeneous tangent model the center, homogeneity, and normalized angular measure are recoverable in any $D \geq 2$ under explicit identifiability conditions; quantitative discrete-query certificates are planar, though finite-branch reconstruction is not. Our contributions are:

1. **Score-only center and homogeneity calibration.** Homogeneity makes the smoothed conic density an Ornstein–Uhlenbeck eigenfunction; testing its score equation against localized functions gives a linear system for the unknown center and degree d, without differentiating the score. Full column rank suffices, its failure captures translational symmetries, and a direct least-squares perturbation bound quantifies approximate fields.

2. **Constructive recovery in arbitrary dimension.** In every $D \geq 2$, a centered exact shell identifies any finite positive angular measure. For at most K atoms, harmonic moments through degree $2K - 1$ form a finite-rank multivariate moment pencil whose rank, eigenvalues, and coefficients recover count, directions, and weights; any fixed scheme requires at least $KD - 1$ scalar tangential observations.

3. **Sharp planar recovery and conditional stability.** For finite $C^{1,\beta}$ planar branches with positive $C^{0,\beta}$ densities, the tangent-score bias is $O(\sigma^{\beta})$ on fixed normalized query sets. For positive planar rays moments through degree K suffice, whereas degree $K - 1$ does not; declared weight, angular separation, and conditioning bounds yield sufficient count, direction, and weight tests.

4. **Finite-data and learned-score diagnostics.** The error chain exposes tangent bias, local coverage, calibration error, harmonic amplification, network error, and query aliasing. Population, KDE, and learned fields pass through matched recovery stages, and long-horizon controls show lower validation normalized-score error can coexist with higher angular-moment error and incomplete geometry recovery.

Experiments validate the full chain across the tested dimensions. Every positive certificate report respects its declared bounds, and held-out KDE blind recovery reaches 99.74% at local coverage 1024. After 50k updates, normalized-score RMS falls by 5.4%–24.0% while moment error rises by 3.6%–20.9%, leaving full-geometry recovery at 2–4/10. Classical spectral methods start from scalar convolutions or moments; we derive that interface from an uncentered score field by calibrating center and homogeneity, recovering the shell transform, and propagating score errors to moments. Subsequent spectral reconstruction is classical (Kunis et al., 2016; 2019). Our scope is local identification from one score slice; Appendix J and Section 7 give boundary cases and limitations.

2 Self-calibration and one-shell tangent geometry

Let an unknown branch point $x_0 \in \operatorname{supp}\mu \subset \mathbb{R}^D$ lie in the local query window. Tangent measures formalize blow-up limits of measures (Preiss, 1987). We assume the rescaled local measures converge, in the Gaussian-weighted moments used below, to a nonzero d-homogeneous tangent measure

$$\mathrm{d}\nu_\Lambda(r,\theta) = r^{d-1}\mathrm{d}r\,\mathrm{d}\Lambda(\theta), \qquad \theta \in \mathbb{S}^{D-1}. \tag{1}$$

The noise scale σ is known from the score model. The center x_0 and homogeneity degree d will be recovered rather than supplied. The score cannot recover the total mass of the finite positive angular measure Λ. Its identifiable target is the normalized measure

$$\bar{\Lambda}(A) := \frac{\Lambda(A)}{\Lambda(\mathbb{S}^{D-1})} = \frac{\nu_\Lambda(\{r\theta : 0 < r \le 1,\ \theta \in A\})}{\nu_\Lambda(B(0,1))}, \qquad A \subseteq \mathbb{S}^{D-1}.$$

Thus a branch weight is a sector-wise fraction of tangent mass under the reference measure defining μ. Overall rescaling and reparameterizations that preserve μ leave it unchanged; replacing parameter measure by Hausdorff or arc-length measure can change it. For a data-distribution measure, the weights describe local probability shares, while for a geometric reference measure they describe relative branch density or multiplicity. We henceforth normalize Λ to unit mass when discussing weights. In the planar finite-ray case, $\Lambda = \sum_{j=1}^{s} w_j \delta_{\theta_j}$ with $w_j > 0$ and $\sum_j w_j = 1$ records branch count, directions, and relative mass.

The tangent-measure change of variables gives

$$\sigma s_\sigma(x_0 + \sigma z) \longrightarrow F_\Lambda(z) := \nabla_z \log q_\Lambda(z), \qquad q_\Lambda(z) = \int e^{-\|z-u\|^2/2}\mathrm{d}\nu_\Lambda(u). \tag{2}$$

For finite planar $C^{1,\beta}$ branches with $C^{0,\beta}$ positive densities, (2) holds uniformly at rate $O(\sigma^\beta)$ on compact normalized query sets. Appendix Theorem B.1 gives the conditions and proof; Appendix Figure 3 checks the predicted slopes but is not used in the proof.

2.1 Center and homogeneity from score values

Choose any reference c in the local query window and write $y = (x-c)/\sigma$, $b = (x_0 - c)/\sigma$, and $u(y) = \sigma s_\sigma(c + \sigma y)$. In the exact conic model, $u(y) = F_\Lambda(y - b)$. Homogeneity and the Gaussian heat equation imply

$$\nabla\cdot u(y) + \|u(y)\|^2 + (y-b)^\top u(y) + D - d = 0. \tag{3}$$

Homogeneity balances score energy, radial score, and local dimension, and a wrong center enters this balance linearly, so localized averages give a small linear system without differentiating the score. For a differentiable localized ψ, define

$$a_\psi = \left(\int \psi(y)u(y)\mathrm{d}y,\ \int \psi(y)\mathrm{d}y \right), \tag{4}$$

$$r_\psi = \int \left[-\nabla\psi(y)^\top u(y) + \psi(y)\{\|u(y)\|^2 + y^\top u(y) + D\} \right] \mathrm{d}y. \tag{5}$$

Theorem 1 (Single-scale center and homogeneity calibration)**.** *Stack the rows a_{ψ_j} into A and the values r_{ψ_j} into r. For every exact translated d-homogeneous cone,*

$$A\begin{pmatrix} b \\ d \end{pmatrix} = r. \tag{6}$$

If A has full column rank, score values at this one scale uniquely recover b and d. For perturbed quantities $\widehat{A} = A + E$ and $\widehat{r} = r + e$, least squares obeys

$$\left\| \begin{pmatrix} \widehat{b} \\ \widehat{d} \end{pmatrix} - \begin{pmatrix} b \\ d \end{pmatrix} \right\| \leq \frac{\|e\| + \|E\| \left\| \binom{b}{d} \right\|}{\sigma_{\min}(A) - \|E\|}, \qquad \|E\| < \sigma_{\min}(A). \tag{7}$$

Here E, e collect score, integration, and tangent-model errors; for a stochastic estimator, the bound holds on any event controlling their norms. Gaussian tests make the integrals expectations of score values. If the cone is translation invariant in a direction v, then $u(y)^\top v = 0$, the column for that center direction is unidentifiable, and shifting the nominal center along v changes no geometry. Thus the rank condition excludes a genuine symmetry rather than merely a numerical inconvenience. Appendix A proves the strong and weak identities, the perturbation bound, and the corresponding approximate tangent statement. We henceforth recenter at the recovered x_0 and use the recovered d.

2.2 One-shell angular recovery

Related boundary-layer analysis derives forward asymptotics for tangent cones and corners (Brosse & Dalalyan, 2026). Our inverse problem begins with the limiting field and asks whether it determines the angular measure.

Fix $R > 0$, write $z = R\omega$, and define

$$K_{d,R}(t) = \int_0^\infty r^{d-1} e^{-r^2/2 + Rrt} \mathrm{d}r, \qquad \mathsf{T}_{d,R}\Lambda(\omega) = \int K_{d,R}(\omega^\top \theta) \mathrm{d}\Lambda(\theta). \tag{8}$$

Polar integration and tangential differentiation yield the central identity

$$q_\Lambda(R\omega) = e^{-R^2/2} \mathsf{T}_{d,R}\Lambda(\omega), \qquad \boxed{g_R(\omega) := RP_{\omega^\perp} F_\Lambda(R\omega) = \nabla_{\mathbb{S}^{D-1}} \log \mathsf{T}_{d,R}\Lambda(\omega).} \tag{9}$$

We call g_R the *shell tangential score*. It reveals the positive scalar shell transform up to one multiplicative constant. The physical query radius is $R\sigma$.

The tangential score gives directional changes of a blurred angular density; integrating these on the connected shell recovers that density up to scale, normalization removes the scale, and Gaussian–cone smoothing preserves every harmonic mode.

Theorem 2 (One-shell injectivity)**.** *Fix $D \geq 2$, $d > 0$, and $R > 0$. If two nonzero finite positive angular measures have identical shell tangential scores g_R on $R\mathbb{S}^{D-1}$, then they are proportional. Unit-mass angular measures are therefore equal.*

The proof has two steps. Equal tangential log-gradients imply proportional scalar transforms because the sphere is connected. The Funk–Hecke formula, the spherical analogue of Fourier diagonalization for a rotationally symmetric kernel, diagonalizes the transform. Its degree-ℓ multiplier is a positive radial mixture of $(Rr)^{1-D/2} I_{\ell+D/2-1}(Rr)$; hence no harmonic mode vanishes. Full details are in Appendix C.

The shell has already converted score queries into polynomial moments; coordinate multiplication becomes a family of commuting small matrices whose joint eigenvalues are the branch directions, and a linear solve gives the weights.

Theorem 3 (Finite positive rays in arbitrary dimension)**.** *Let $D \geq 2$ and let $\Lambda = \sum_{j=1}^s w_j \delta_{\theta_j}$ contain $s \leq K$ distinct directions with positive weights. From one exact shell, spherical-harmonic moments through degree $2K-1$ constructively determine s, every $\theta_j \in \mathbb{S}^{D-1}$, and every normalized weight w_j.*

Once the shell moments are available, the remaining atom-recovery step is classical spherical multivariate Prony (Kunis et al., 2016; 2019): moment-matrix rank gives the count, commuting multiplication matrices give the directions, and a linear solve gives the weights. Appendix D verifies the interpolation rank and degree-$2K-1$ bound. The paper-specific step is the preceding conversion from an uncentered vector score to moments.

On the circle, spherical harmonics become Fourier modes, and positivity turns the moments into a positive semidefinite Toeplitz matrix whose rank gives the count and whose nullspace polynomials vanish at the branch directions.

Theorem 4 (Finite positive planar rays). *Let $D = 2, d = 1$, and suppose Λ contains $s \le K$ distinct positive rays. The reconstructed unit-mean shell density h has Fourier modes*

$$h_k = \rho_{|k|}(R) m_k, \qquad m_k = \sum_{j=1}^{s} w_j e^{-ik\theta_j}, \qquad \rho_k(R) = \frac{\kappa_k(1,R)}{\kappa_0(1,R)} > 0. \tag{10}$$

Modes $|k| \le K$ determine s, all directions, and all weights.

Proposition 1 (Why degree K is necessary in the plane). *For every $K \ge 1$, normalized positive measures supported on at most K planar rays are not identified by the consecutive moments $m_0, \dots, m_{K-1}$. Hence the largest order K in Theorem 4 is necessary for this moment description.*

After deconvolution, the Toeplitz/Vandermonde step is standard in finite-rate-of-innovation, subspace, and super-resolution methods (Vetterli et al., 2002; Schmidt, 1986; Roy & Kailath, 1989; Candès & Fernandez-Granda, 2014; Moitra, 2015; Yang et al., 2016): Toeplitz rank gives the count, common nullspace roots give directions, and a linear solve gives weights. Our contribution is to supply these moments from a local score shell and carry score error to stability margins. Proposition 1 follows from rotated regular K-ray measures, whose first $K-1$ nonconstant moments all vanish; the proof is in Appendix E. This is a lower bound for the consecutive-moment interface, because one score query is a nonlinear combination of all angular modes. The next result gives a separate lower bound directly at the score-query interface.

Proposition 2 (Fixed scalar-query lower bound). *Fix K pairwise disjoint open spherical patches, one for each ray, and allow all positive normalized weights. If M scalar tangential-score components are observed through a fixed continuous scheme, exact identification of every measure in this class requires*

$$M \ge KD - 1. \tag{11}$$

This holds even when K is known and the observations are noiseless.

The directions contribute $K(D-1)$ local coordinates and the normalized weights contribute $K-1$. Their map to the observations is continuous, so invariance of domain gives the bound (Munkres, 2000, Section 62). A full tangential vector at one fixed shell location contains at most $D-1$ scalar components. The proposition does not cover adaptive observations. For $D = 2$ it reduces to $M \ge 2K-1$; the smallest uniform planar grid covered by our conditional analysis uses $M = 2K+1$. Appendix D gives the general proof.

3 Shell information beyond a one-point Hessian

For unit-mass Λ, let $m_\Lambda = \mathbb{E}_\Lambda \theta$, $Q_\Lambda = \mathbb{E}_\Lambda \theta\theta^\top$, and $a_d = \sqrt{2}\,\Gamma((d+1)/2)/\Gamma(d/2)$. Direct differentiation gives

$$\nabla F_\Lambda(0) = dQ_\Lambda - a_d^2 m_\Lambda m_\Lambda^\top - I_D. \tag{12}$$

Taking the trace and using $F_\Lambda(0) = a_d m_\Lambda$ recovers the scalar homogeneity through $d = D + \operatorname{tr} \nabla F_\Lambda(0) + \|F_\Lambda(0)\|^2$, the centered pointwise case of (3). The weak calibration theorem obtains the same nuisance parameter without knowing the center or differentiating the score. By contrast, the full vertex Hessian still keeps only two angular moments. Every uniform regular planar q-ray junction with $q \ge 3$ has $m_\Lambda = 0$, $Q_\Lambda = I_2/2$, and, for $d = 1$, $\nabla F(0) = -I_2/2$. Yet its first count-specific shell harmonic is nonzero. Figure 2 compares the common vertex Hessian, the distinct shell fields, and the inverse harmonic multipliers at the tested radii. Appendix F proves (12) and the full regular-q statement.

4 From finite queries to conditional guarantees

Exact identifiability assumes a continuous field. Here finitely many noisy queries produce a candidate, and predeclared margins determine what can be certified.

4.1 Finite-query reconstruction

For the planar estimator, query $M > 2K$ equally spaced angles $\phi_j = 2\pi j/M$. Let $g(\phi)$ be the planar parameterization of the shell tangential score g_R, so $g = (\log h)'$, and observe $y_j = g(\phi_j) + e_j$

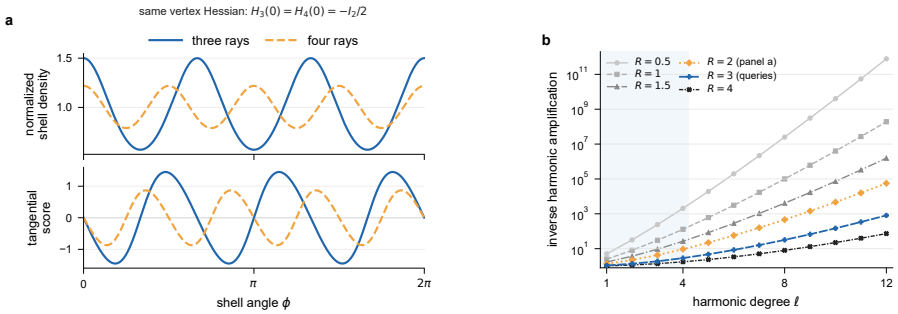

Figure 2: One-point collision and shell conditioning: (a) three and four rays share the vertex Hessian but differ on the $R = 2$ shell; (b) inverse harmonic amplification grows with degree (shaded: $1{:}4$).

with $M^{-1}\sum_j |e_j|^2 \le \varepsilon_2^2$; thus ε_2 is a root-mean-square (RMS) observation bound. The estimator spectrally integrates g, exponentiates and normalizes h, and removes the known shell multipliers from modes $1{:}K$. Toeplitz rank gives branch count, annihilating-polynomial roots give directions, and a final nonnegative solve gives weights.

Algorithm 1 (Finite-query one-shell estimator)**.** Inputs are R, K, and the M shell tangential score values.

1. Divide every nonzero discrete Fourier transform (DFT) mode by its frequency, then exponentiate and normalize to reconstruct the shell density.
2. Divide density modes $1{:}K$ by the known shell multipliers to obtain angular moments.
3. Infer Toeplitz rank, extract unit-circle roots, and fit normalized nonnegative weights.

A numerical candidate uses a relative eigengap. A certified output applies the thresholds in Theorem 5 and stops whenever a sufficient test fails.

Let $A_M(g)$ denote the explicit spectral-integration alias term, $B_{M,k}(R)$ the scalar-density alias term, and

$$C_M = \left(\sum_{k \in \mathcal{K}_M \setminus \{0\}} k^{-2} \right)^{1/2} < \pi/\sqrt{3}, \qquad \tau_M = A_M(g) + C_M \varepsilon_2. \tag{13}$$

Appendix G proves the finite-query moment bound

$$\boxed{|\widehat{m}_k - m_k| \le \rho_k(R)^{-1}(e^{2\tau_M} - 1) + \frac{B_{M,k}(R)/\rho_k(R) + B_{M,0}(R)}{1 - B_{M,0}(R)}}, \quad 1 \le k \le K, \tag{14}$$

where $\rho_k = \kappa_k/\kappa_0$ and $B_{M,0} < 1$. This separates score noise, log integration, density sampling alias, and harmonic amplification.

4.2 Conditional deterministic certification

The score-error bound first becomes a uniform moment-error bound; eigenvalues then certify branch count, separated polynomial roots certify directions, and a conditioned linear solve certifies weights, each stage using bounds fixed before inspecting the field and abstaining before any unverified output.

Theorem 5 (Finite-query certification)**.** *Assume $M > 2K$, $A_M(g) < \infty$, $B_{M,0}(R) < 1$, and the stated RMS observation error. Let $\bar{\eta}$ be the maximum of the right-hand side of* (14) *over $1 \le k \le K$, or any other valid bound on $\max_{k \le K} |\widehat{m}_k - m_k|$. Then*

$$\|\widehat{T}_K - T_K\|_{\mathrm{op}} \le (K+1)\bar{\eta} =: E_K. \tag{15}$$

For a declared geometry class with minimum positive Toeplitz eigenvalue Γ_K, branch count is guaranteed if $\Gamma_K > 2E_K$: it is the number of eigenvalues of $\widehat{T}_K$ above E_K. This certified rule is distinct from the relative-eigengap heuristic used only for an uncertified candidate. An error estimate

alone cannot exclude an extra branch of arbitrarily small weight, so this class-level resolution gap is needed for a uniform guarantee. Declared lower bounds on angular separation and the smallest Vandermonde singular value give the sufficient direction and weight tests in Appendix H. A failed sufficient test causes abstention; it does not invalidate the numerical candidate. Here "declared" means fixed before inspecting the tested field, using domain knowledge about the target class. Substituting the unknown true instance into these bounds is an oracle diagnostic, not a deployable certificate.

Appendix Table 2 exposes the three sufficient tests. The exact Rouché and Vandermonde inequalities, including the $s = 1$ case, are stated and proved in Appendix H.

The executable version derives every required lower bound from a minimum weight and angular separation fixed before inspecting the field. Appendix H gives these conservative substitutions; failure makes the procedure abstain rather than use the unknown true geometry.

4.3 Finite-data and learned-score errors

For N independent data samples, the normalized empirical KDE score at $x = x_0 + \sigma z$ is exactly

$$\sigma \widehat{s}_{N,\sigma}(x) = \frac{N^{-1} \sum_i (X_i - x)\sigma^{-1} e^{-\|x - X_i\|^2/(2\sigma^2)}}{N^{-1} \sum_i e^{-\|x - X_i\|^2/(2\sigma^2)}}. \tag{16}$$

For a general local measure, assume a Gaussian-weighted tangent approximation at rate σ^α on the translated query neighborhoods required by the center-error bound. Theorem B.1 verifies this condition with $\alpha = \beta$ for its planar branch class. Also assume $a_j \geq a_- \sigma^d$, validated calibration bounds $\|\widehat{b} - b\| \leq \varepsilon_{\mathrm{ctr}}$ and $|\widehat{d} - d| \leq \varepsilon_d$, and a validated pointwise network error $\varepsilon_\theta(\sigma)$. A simultaneous Bernstein ratio bound over the M queries then gives, with probability at least $1 - \delta$,

$$\boxed{\varepsilon_{\mathrm{shell}} \leq R \left\{ C_{\mathrm{geom}} \sigma^\alpha + C_{\mathrm{samp}} \left[\sqrt{\frac{\log(M/\delta)}{N a_- \sigma^d}} + \frac{\log(M/\delta)}{N a_- \sigma^d} \right] + \sigma \varepsilon_\theta(\sigma) + C_{\mathrm{ctr}} \varepsilon_{\mathrm{ctr}} \right\}.} \tag{17}$$

Here C_{geom} records the tangent approximation and density lower bound, while C_{ctr} controls local score variation. The exact sampling term is (101) with $t = \log(4M/\delta)$, $\lambda = N a_- \sigma^d$, and $S = \max_j \|b_j/a_j\|$; the simplified display requires $\sqrt{2t/\lambda} + t/(3\lambda) \leq 1/2$. Thus C_{samp} hides only universal constants and the declared bound on S, not an additional power of N or σ; none of these geometry-dependent quantities is distribution-free. On any compact declared interval for d, the finitely many positive shell multipliers used by the inverse have bounded logarithmic derivatives. Consequently, using $\widehat{d}$ adds at most $C_d \varepsilon_d$ to the recovered moment error in (14). Combining this term with $\varepsilon_2 \leq \varepsilon_{\mathrm{shell}}$ yields branch count, angle, and weight guarantees, or abstention. Balancing only tangent bias and sampling error gives $\sigma_{\mathrm{opt}} \asymp (N a_-)^{-1/(2\alpha + d)}$; this is estimator calibration, not a multiscale-necessity claim.

4.4 Stagewise stability

Proposition 3 (Stagewise perturbation margins)**.** *Let η bound the recovered moments through order K, and write $\gamma_K(\Lambda) = \lambda^+_{\min}(T_K)$. Correct branch count is guaranteed whenever*

$$\eta < \eta_{\mathrm{count}} := \frac{\gamma_K(\Lambda)}{2(K+1)}. \tag{18}$$

Conditional on that count, an angular tolerance a_0 is guaranteed if the Rouché inequality in (84) *with $r \leq \sin(a_0/2)$ holds, and a weight tolerance follows from the Vandermonde inequality in* (87)*. Full geometric recovery therefore has sufficient rank, root, and weight margins, whereas count uses only the rank margin. Replacing the geometry-dependent quantities by declared class-level lower bounds turns these perturbation statements into uniform certificates.*

The proposition follows by composing Weyl rank stability, Davis–Kahan and Rouché root stability, and the perturbed nonnegative weight solve. It gives a concrete mechanism for a count-correct but geometry-inaccurate regime: root separation or Vandermonde conditioning can become limiting after the Toeplitz rank has stabilized. Appendix H gives the complete proof and constants.

Table 1: Single-scale calibration and higher-dimensional reconstruction. Center error is normalized by σ. Exact-calibration rows report the maximum numerical quadrature error; perturbed and learned rows report the median. "Full geometry" counts trials meeting the generated count, direction, and weight criteria. For the three-dimensional score-shell rows, these criteria are correct count, maximum angular error 0.1°, and maximum weight error 10^{-3}. Every entry is generated from seed-level rows.

Task and field	D	Runs	Center error	$\lvert\widehat{d} - d\rvert$	Full geometry
Exact weak calibration	2	20	4.4×10^{-4}	8.7×10^{-5}	–
Exact weak calibration	3	20	8.5×10^{-4}	2.2×10^{-4}	–
Exact weak calibration	5	20	4.3×10^{-4}	9.7×10^{-5}	–
Score perturbation ($\mathrm{RMS} = 0.01$)	2	20	0.018	0.004	–
Score perturbation ($\mathrm{RMS} = 0.01$)	3	20	0.017	0.006	–
Score perturbation ($\mathrm{RMS} = 0.01$)	5	20	0.016	0.005	–
Exact score-shell reconstruction	3	20	–	–	20/20
Learned-score weak calibration	3	5	0.061	0.032	–

5 Experiments

We evaluate each link of the recovery chain in reproducible synthetic experiments and recompute every reported statistic from released seed-level rows. Before evaluation, we define planar full recovery as correct count, maximum circular angle error at most 5°, and maximum weight error at most 0.05. This descriptive criterion is separate from the certified tests, and Appendix Tables 8 and 3 report tolerance sensitivity and distinguish blind candidates, oracle diagnostics, and certified outputs.

5.1 Calibration and recovery beyond the plane

Table 1 evaluates score-only calibration and arbitrary-dimensional finite-ray recovery. Weak Gaussian tests recover center and homogeneity for exact ray cones in $D = 2, 3, 5$; a separate sweep adds smooth score perturbations with declared root-mean-square magnitude. The shell experiment uses non-coplanar two- and three-ray measures on $\mathbb{S}^2$, integrates their tangential score along meridians, deconvolves spherical harmonics through degree five, and applies the moment pencil of Theorem 3. Finally, the learned rows train a three-dimensional residual multilayer perceptron (MLP) by denoising score matching and apply the same derivative-free calibration. Models are evaluated in memory: no checkpoint or prediction dump is part of the experiment.

The exact and perturbed rows distinguish algebraic identifiability from conditioning of the chosen test family. The learned result shows that the calibration observable survives ordinary score training in this controlled three-dimensional setting. A neural-field guarantee additionally requires a validated external error bound. The non-coplanar shell rows exercise the full score-to-harmonic-to-pencil chain from score queries.

5.2 Finite-query and empirical KDE validation

Appendix Table 4 and Figure 5 evaluate finite-query reconstruction, certification, and sampling. The finite-query matrix contains 104,352 trials. It varies normalized radius, query count, score noise, seed, angular separation, minimum weight, ray clustering, rank selection, and center offset. At $R = 3$ and $M = 32$, all exact evaluation cases recover. Under noise, a four-ray cluster is less stable than an isolated close pair with the same minimum separation. This is consistent with the full Vandermonde matrix, rather than only the closest pair, controlling recovery. Artificial center offsets through 0.1σ retain 100% success in the baseline matrix; at 0.2σ, success falls to 68.75% as the declared angle tolerance is crossed.

The complete sufficient certificate is also executed under geometry bounds fixed before perturbation. Across 336 bounded-error trials, every positive report respects its count, direction, or weight bound; Appendix Table 4 shows how later stages abstain before earlier ones as error grows.

The population/KDE evaluation contains 6,528 trials over eight named and twelve continuously generated held-out geometries. Here local coverage is the expected effective number of nearby samples receiving appreciable Gaussian weight at a query. Population inversion succeeds on all twenty geometries. For local coverages at least 32, fitted log–log slopes are -0.490 for normalized-score RMS and -0.499 for moment error, matching inverse-square-root coverage. On held-out geometries, blind recovery is 98.96%, 98.96%, and 99.74% at coverages 256, 512, and 1024. A fixed eight-template library and a one-point Hessian summary based on principal-component analysis (PCA) do not meet the same full-geometry criterion. These information-limited baselines are sanity checks for label templates and one-point information; the reconstruction procedure itself uses a classical spectral estimator. Under curvature, 0.1σ center error, ambient noise, and their combination, blind recovery is 83%–93%. An oracle rank gate reports 413 candidates in this matrix, all of which meet the full-geometry tolerance; this diagnostic uses the true synthetic signal gap and is not a deployable certificate. Appendix Figure 5 summarizes the coverage law and held-out recovery.

Controlled learned-score diagnostics. Appendix Table 7 and Figure 8 test whether longer training improves downstream inversion. Across ten seeds per cell, the 110 trajectories comprise 80 parameter-matched plain/residual models on four geometries and 30 larger-residual controls on three. All meet the plateau rule: less than 1% improvement in the best validation normalized-score RMS across a ten-evaluation window, recorded only after 20k updates. Validation selects the plain MLP at 50k updates for every geometry. Relative to 5k updates, normalized-score RMS improves by 5.4%– 24.0% while moment error increases by 3.6%–20.9%. Full-geometry recovery is 2–4/10, and weak-Y count recovery is 6/10.

6 Related work

Score geometry. Smooth-support work recovers normal bundles, dimensions, pullback geometry, or projections from scores and Jacobians (Stanczuk et al., 2024; Ventura et al., 2025; Diepeveen et al., 2025; Kharitenko et al., 2026). Other analyses treat tubes, singularities, curvature, structured supports, and forward junctions (Sakamoto et al., 2024; Liu et al., 2025; Li et al., 2026; Zhang et al., 2026; Yang et al., 2026; Mu et al., 2026; Brosse & Dalalyan, 2026; Chen et al., 2026), but none inverts weighted junction geometry; we recover its center, homogeneity, and tangent measure at one scale without a score Jacobian.

Spectral and spherical inversion. Finite-rate-of-innovation, MUSIC, ESPRIT, and super-resolution recover atoms from moments (Vetterli et al., 2002; Schmidt, 1986; Roy & Kailath, 1989; Candès & Fernandez-Granda, 2014; Moitra, 2015; Yang et al., 2016); spherical deconvolution, kernels, and multivariate Prony invert transforms or moments (Kerkyacharian et al., 2011; Simon-Gabriel & Schölkopf, 2018; Kunis et al., 2016; 2019); directional score matching assumes a known manifold (Mardia et al., 2016). Ours is score calibration, shell integration, all-frequency injectivity, and the finite-query score-to-moment error chain.

Neighboring targets. Global schedule mixture weights (Dennehy et al., 2026) differ from our local sector masses, and point-cloud methods locate but do not audit score fields (von Rohrscheidt & Rieck, 2023; Lim et al., 2025); ours is KDE sampling, not the memorization mechanism in Merger & Goldt (2026).

7 Limitations and conclusion

Scope is local: one known noise level, one approximately homogeneous junction window, a branch bound K, and convergence in the Gaussian-weighted score ratio. Theorem B.1 covers zero-thickness planar $C^{1,\beta}$ branches, not thickness, general strata, or arbitrary-dimensional rates, and the target may vary with scale without a stable tangent measure. Calibration requires full rank and can be ill-conditioned near symmetry; certification needs external error, weight, and separation bounds, else it abstains, and global search is outside scope. From one score slice we recover local weighted tangent geometry with exact arbitrary-dimensional identifiability and planar certificates. Although broader rates and pretrained scores remain future work, the audits confirm that score fit and geometry recovery are distinct, so the method isolates where a learned field's geometric claim breaks.

REPRODUCIBILITY STATEMENT

Appendices A–K state all assumptions, proofs, protocols, architectures, and seeds. Every reported number, table, and figure is generated from executable code and checked against saved raw results; the supporting-material README gives the commands. No external or private dataset is used.

ETHICS STATEMENT

The work uses only synthetic mathematical data and does not involve human subjects, personal data, or deployment decisions.

AI USE STATEMENT

Generative AI tools were used only to improve the grammar and readability of author-written text. The authors reviewed all edits and take full responsibility for the final manuscript.

REFERENCES

Nicolas Brosse and Arnak S. Dalalyan. Boundary-layer asymptotics for gaussian-smoothed singular measures. *arXiv preprint arXiv:2607.04514*, 2026. URL https://arxiv.org/abs/2607.04514.

Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014. doi: 10.1002/cpa.21455.

Zirui Chen, Shiliang Guo, and Shiyu Zhao. Guiding vector field generation via score-based diffusion model. *arXiv preprint arXiv:2604.24487*, 2026. doi: 10.48550/arXiv.2604.24487. URL https://arxiv.org/abs/2604.24487.

Andrew Dennehy, Ramchandran Muthukumar, Rebecca Willett, and Nisha Chandramoorthy. Diffusion models recover accurate mixture weights despite score function insensitivity. *arXiv preprint arXiv:2607.15485*, 2026. doi: 10.48550/arXiv.2607.15485. URL https://arxiv.org/abs/2607.15485.

Willem Diepeveen, Georgios Batzolis, Zakhar Shumaylov, and Carola-Bibiane Schönlieb. Score-based pullback Riemannian geometry: Extracting the data manifold geometry using anisotropic flows. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 13746–13773. PMLR, 2025. URL https://proceedings.mlr.press/v267/diepeveen25a.html.

Gérard Kerkyacharian, Thanh Mai Pham Ngoc, and Dominique Picard. Localized spherical deconvolution. *The Annals of Statistics*, 39(2):1042–1068, 2011. doi: 10.1214/10-AOS858.

Andrey Kharitenko, Zebang Shen, Riccardo De Santi, Niao He, and Florian Dorfler. Landing with the score: Riemannian optimization through denoising. In *International Conference on Learning Representations*, volume 2026, pp. 129064–129104, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/d1a6fe144bdff7b95138bca8097d4e69-Paper-Conference.pdf.

Stefan Kunis, Thomas Peter, Tim R"omer, and Ulrich von der Ohe. A multivariate generalization of Prony's method. *Linear Algebra and its Applications*, 490:31–47, 2016. doi: 10.1016/j.laa.2015.10.023. URL https://doi.org/10.1016/j.laa.2015.10.023.

Stefan Kunis, H. Michael Möller, and Ulrich von der Ohe. Prony's method on the sphere. *SMAI Journal of Computational Mathematics*, S5:87–97, 2019. doi: 10.5802/smai-jcm.53.

Xiang Li, Zebang Shen, Ya-Ping Hsieh, and Niao He. When scores learn geometry: Rate separations under the manifold hypothesis. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/4406cafe40d3ca4a7511a8996b6bd1b6-Abstract-Conference.html.

Uzu Lim, Harald Oberhauser, and Vidit Nanda. HADES: Fast singularity detection with local measure comparison. *SIAM Journal on Mathematics of Data Science*, 7(4):1882–1903, 2025. doi: 10.1137/24M1665763.

Zichen Liu, Wei Zhang, and Tiejun Li. Improving the euclidean diffusion generation of manifold data by mitigating score function singularity. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-3704.

Kanti V. Mardia, John T. Kent, and Arnab K. Laha. Score matching estimators for directional distributions, 2016. URL https://arxiv.org/abs/1604.08470.

Claudia Merger and Sebastian Goldt. Local coverage governs memorization in diffusion models. *arXiv preprint arXiv:2606.14390*, 2026. URL https://arxiv.org/abs/2606.14390.

Ankur Moitra. Super-resolution, extremal functions and the condition number of vandermonde matrices. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, pp. 821–830, 2015. doi: 10.1145/2746539.2746561.

Xinhe Mu, Zaijiu Shang, Zhaoqi Zhou, Chuan Zhou, Qi Meng, Guiying Yan, and Zhiming Ma. Score approximation for diffusion models on arbitrary low-dimensional structures, 2026. URL https://arxiv.org/abs/2606.19894.

James R. Munkres. *Topology*. Prentice Hall, 2 edition, 2000.

Iosif Pinelis. Optimum bounds for the distributions of martingales in banach spaces. *The Annals of Probability*, 22(4):1679–1706, 1994. doi: 10.1214/aop/1176988477.

David Preiss. Geometry of measures in $\mathbb{R}^n$: Distribution, rectifiability, and densities. *Annals of Mathematics*, 125(3):537–643, 1987. doi: 10.2307/1971410. URL https://annals.math.princeton.edu/1987/125-3/p05.

Richard Roy and Thomas Kailath. ESPRIT—estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(7):984–995, 1989. doi: 10.1109/29.32276.

Kotaro Sakamoto, Ryosuke Sakamoto, Masato Tanabe, Masatomo Akagawa, Yusuke Hayashi, Manato Yaguchi, Masahiro Suzuki, and Yutaka Matsuo. The geometry of diffusion models: Tubular neighbourhoods and singularities. In *Proceedings of the Geometry-grounded Representation Learning and Generative Modeling Workshop*, volume 251 of *Proceedings of Machine Learning Research*, pp. 332–363. PMLR, 2024. URL https://proceedings.mlr.press/v251/sakamoto24a.html.

Ralph O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986. doi: 10.1109/TAP.1986.1143830.

Carl-Johann Simon-Gabriel and Bernhard Schölkopf. Kernel distribution embeddings: Universal kernels, characteristic kernels and kernel metrics on distributions. *Journal of Machine Learning Research*, 19(44):1–29, 2018. URL https://jmlr.org/papers/v19/16-291.html.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Jan Pawel Stanczuk, Georgios Batzolis, Teo Deveney, and Carola-Bibiane Schönlieb. Diffusion models encode the intrinsic dimension of data manifolds. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 46412–46440. PMLR, 2024. URL https://proceedings.mlr.press/v235/stanczuk24a.html.

Enrico Ventura, Beatrice Achilli, Gianluigi Silvestri, Carlo Lucibello, and Luca Ambrogioni. Manifolds, random matrices and spectral gaps: The geometric phases of generative diffusion. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/dbdb2729f874e209b56ffecd23aa3b36-Abstract-Conference.html.

Martin Vetterli, Pina Marziliano, and Thierry Blu. Sampling signals with finite rate of innovation. *IEEE Transactions on Signal Processing*, 50(6):1417–1428, 2002. doi: 10.1109/TSP.2002.1003065.

Julius von Rohrscheidt and Bastian Rieck. Topological singularity detection at multiple scales. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 35175–35197. PMLR, 2023. URL https://proceedings.mlr.press/v202/von-rohrscheidt23a.html.

Ruofeng Yang, Yongcan Li, Bo Jiang, Cheng Chen, and Shuai Li. Multi-subspace multi-modal modeling for diffusion models: Estimation, convergence and mixture of experts. In *International Conference on Learning Representations*, volume 2026, pp. 35147–35183, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/3b493a9f0aa0ba8a42ae6cc93da5d81b-Paper-Conference.pdf.

Zai Yang, Lihua Xie, and Petre Stoica. Vandermonde decomposition of multilevel toeplitz matrices with application to multidimensional super-resolution. *IEEE Transactions on Information Theory*, 62(6):3685–3701, 2016. doi: 10.1109/TIT.2016.2553041.

Zixuan Zhang, Kaixuan Huang, Tuo Zhao, Mengdi Wang, and Minshuo Chen. Diffusion model for manifold data: Score decomposition, curvature, and statistical complexity, 2026. URL https://arxiv.org/abs/2603.20645.

A SINGLE-SCALE CENTER AND HOMOGENEITY CALIBRATION

This section proves Theorem 1. Let ν be a nonzero d-homogeneous measure on $\mathbb{R}^D$ and define its normalized Gaussian smoothing at variance t by

$$q_t(z) = (2\pi t)^{-D/2} \int \exp\left(-\frac{\|z-u\|^2}{2t}\right) \mathrm{d}\nu(u). \tag{19}$$

Changing variables $u = \sqrt{t}\, v$ and using $\nu(\sqrt{t}A) = t^{d/2}\nu(A)$ gives

$$q_t(z) = t^{(d-D)/2} q_1(z/\sqrt{t}). \tag{20}$$

The Gaussian convolution solves $\partial_t q_t = \Delta q_t/2$. Differentiating (20) at $t = 1$ and equating the two expressions for $\partial_t q_t$ yields

$$\Delta q_1(z) + z^\top \nabla q_1(z) + (D-d) q_1(z) = 0. \tag{21}$$

For a translated center b, put $q_b(y) = q_1(y-b)$ and $u = \nabla \log q_b$. Dividing (21) by q_b proves the strong score identity

$$\nabla\cdot u + \|u\|^2 + (y-b)^\top u + D - d = 0, \tag{22}$$

which is (3). In particular, if the center were already known, evaluation at $y = b$ would give $d = D + \nabla\cdot u(b) + \|u(b)\|^2$.

The usable calibration does not differentiate the score. Multiply the strong identity by a continuously differentiable test function ψ that is compactly supported, or decays sufficiently fast for the boundary term to vanish. Integration by parts gives

$$\int \left[-\nabla\psi(y)^\top u(y) + \psi(y)\{\|u(y)\|^2 + y^\top u(y) + D\}\right] \mathrm{d}y \tag{23}$$

$$= b^\top \int \psi(y) u(y) \mathrm{d}y + d \int \psi(y) \mathrm{d}y. \tag{24}$$

Stacking these identities proves (6). Full column rank of A makes its solution unique.

For completeness, let $\beta = (b, d)$ and suppose $\widehat{A} = A + E$ remains full column rank. Since $r = A\beta$,

$$\widehat{\beta} - \beta = \widehat{A}^\dagger(\widehat{r} - \widehat{A}\beta) = \widehat{A}^\dagger(e - E\beta). \tag{25}$$

Weyl's inequality gives $\sigma_{\min}(\widehat{A}) \geq \sigma_{\min}(A) - \|E\|$; taking norms proves (7). If $\widehat{u}$ has uniform score error at most ε on the test supports, the error in each response is bounded by

$$\|\nabla\psi\|_1 \varepsilon + \|\psi\|_1 \left[(2U + Y)\varepsilon + \varepsilon^2\right], \tag{26}$$

where U bounds $\|u\|$ and Y bounds $\|y\|$ there. The corresponding row error is at most $\|\psi\|_1\varepsilon$. Numerical quadrature and tangent-approximation errors enter the same E, e terms, so the theorem applies without changing the linear-algebra step.

The implementation uses normalized Gaussian tests $\psi_{z,h} = \mathcal{N}(z, h^2 I)$. Dividing (24) by their unit integral turns each row and response into

$$a_{z,h} = (\mathbb{E}u(Y), 1), \tag{27}$$

$$r_{z,h} = \mathbb{E}\left[h^{-2}(Y-z)^\top u(Y) + \|u(Y)\|^2 + Y^\top u(Y) + D\right], \qquad Y \sim \mathcal{N}(z, h^2 I), \tag{28}$$

which requires only score values. Gauss–Hermite quadrature evaluates these expectations in the reported low-dimensional experiments.

Finally, suppose the cone measure and hence q_1 are invariant under translation along a nonzero vector v. Then $u(y)^\top v = 0$ everywhere, so every weak row annihilates $(v, 0)$ and A cannot have full column rank. The same invariance means that b and $b + tv$ describe the same smoothed density. The lost center coordinate is therefore not statistically identifiable. A full line is the simplest example; a genuine junction without lineality is full rank for generic sufficiently rich test families.

B TANGENT-SCORE LIMIT AND SHELL IDENTITY

The remaining appendix gives the full proof chain behind the shell inverse, finite-query guarantees, counterexamples, and experiments. We begin with the forward tangent limit only to state the exact

assumptions needed by the inverse result; later sections then treat arbitrary-dimensional finite rays, the sharp planar specialization, Hessian collisions, stability, finite data, scope, and experimental protocols.

Let $T_{x_0,\sigma}(x) = (x - x_0)/\sigma$ and define the rescaled measure

$$\nu_\sigma = \sigma^{-d}(T_{x_0,\sigma})_{\#}\mu. \tag{29}$$

For a fixed normalized query z, put

$$W_z(u) = e^{-\|z-u\|^2/2}, \quad A_\sigma(z) = \int W_z(u)\mathrm{d}\nu_\sigma(u), \quad B_\sigma(z) = \int (u-z)W_z(u)\mathrm{d}\nu_\sigma(u). \tag{30}$$

Changing variables in the Gaussian convolution cancels all ambient Gaussian normalization constants from the score ratio and gives the exact identity

$$\sigma s_\sigma(x_0 + \sigma z) = \frac{B_\sigma(z)}{A_\sigma(z)}. \tag{31}$$

Assume ν_σ converges to ν_Λ in the two Gaussian-weighted moments in (30), locally uniformly over the declared query set. Since $A_0(z) > 0$, ratio continuity yields

$$\frac{B_\sigma(z)}{A_\sigma(z)} \longrightarrow \frac{B_0(z)}{A_0(z)} = \nabla_z \log q_\Lambda(z), \qquad q_\Lambda(z) = \int W_z(u)\mathrm{d}\nu_\Lambda(u). \tag{32}$$

This proves (2). Notice that weak convergence by itself is not the stated assumption; the Gaussian-weighted numerator and denominator are the quantities actually needed.

B.1 Finite-noise rate for planar branch junctions

Idea. After rescaling by σ, each branch differs from its tangent ray by $O(\sigma^\beta)$, and its density changes at the same order. Gaussian weights make these local errors integrable and suppress the separated remainder exponentially; positivity keeps the final score ratio stable.

Theorem B.1 (Finite-noise planar branch junctions)**.** *Let $D = 2$, $d = 1$, and $0 < \beta \le 1$. Suppose that, near x_0, the measure is a finite sum of half-branches $\gamma_j : [0, r_0] \to \mathbb{R}^2$ with densities ρ_j relative to arc length, plus a finite remainder supported a positive distance from x_0. Assume*

$$\gamma_j(0) = x_0, \quad \gamma_j'(0) = v_j, \quad \|v_j\| = 1, \quad \|\gamma_j'(r) - v_j\| \le L_\gamma r^\beta, \quad |\rho_j(r) - \rho_j(0)| \le L_\rho r^\beta, \tag{33}$$

where the v_j are distinct and $\rho_j(0) > 0$. Then, for every compact $Q \subset \mathbb{R}^2$ in the normalized coordinate $z = (x - x_0)/\sigma$, there are $C_Q < \infty$ and $\sigma_0 > 0$ such that

$$\sup_{z\in Q} \|\sigma s_\sigma(x_0 + \sigma z) - F_\Lambda(z)\| \le C_Q \sigma^\beta, \qquad 0 < \sigma \le \sigma_0, \tag{34}$$

with $\Lambda = \sum_j \rho_j(0)\delta_{v_j}$. In particular, C^2 branches with C^1 positive densities give an $O(\sigma)$ bias.

After shortening the local branch parameterizations if necessary, write the measure as

$$\int f\mathrm{d}\mu = \sum_{j=1}^m \int_0^{r_0} f(\gamma_j(r))\rho_j(r)\|\gamma_j'(r)\|\mathrm{d}r + \int f\mathrm{d}\mu_{\mathrm{far}}, \tag{35}$$

where $\mathrm{dist}(x_0, \operatorname{supp}\mu_{\mathrm{far}}) \ge \Delta > 0$. Let $w_j = \rho_j(0)$ and

$$\nu_0 = \sum_{j=1}^m w_j \int_0^\infty \delta_{uv_j}\mathrm{d}u, \qquad \Lambda = \sum_{j=1}^m w_j\delta_{v_j}. \tag{36}$$

Lemma 1 (Uniform Gaussian envelope)**.** *Fix $R < \infty$. Let v be a unit vector, $u \ge 0$, $\|z\| \le R$, and $q_t = uv + t\eta$ for $0 \le t \le 1$, where $\|\eta\| \le u/2$. There are $c > 0$ and $C_R < \infty$ such that*

$$W_z(q_t) + \|\nabla W_z(q_t)\| \le C_R(1 + u)e^{-cu^2}, \tag{37}$$

$$\|(q_t - z)W_z(q_t)\| + \|\nabla_q[(q - z)W_z(q)]_{q=q_t}\|_{\mathrm{op}} \le C_R(1 + u^2)e^{-cu^2}. \tag{38}$$

Proof. The bounds $u/2 \le \|q_t\| \le 3u/2$ imply $\|q_t - z\| \ge u/2 - R$. For $u \ge 4R$, this is at least $u/4$; for $0 \le u < 4R$, every polynomial factor is bounded by a constant depending on R. The identities

$$\nabla_q W_z(q) = -(q-z)W_z(q), \qquad \nabla_q[(q-z)W_z(q)] = [I - (q-z)(q-z)^\top]W_z(q)$$

then prove both displays, for example with any fixed $c \le 1/64$ after enlarging C_R. □

Proof of Theorem B.1. Define the density including the parameterization Jacobian by

$$a_j(r) = \rho_j(r)\|\gamma_j'(r)\|. \tag{39}$$

We may decrease r_0 until $L_\gamma r_0^\beta \le 1/2$; the omitted branch segments can be absorbed into μ_{far} while preserving a positive distance from x_0. Since $|\|\gamma_j'(r)\| - 1| \le L_\gamma r^\beta$, the density assumptions give constants $L_a, a_+ < \infty$ such that

$$|a_j(r) - w_j| \le L_a r^\beta, \qquad 0 \le a_j(r) \le a_+. \tag{40}$$

For example, if $\bar\rho = \max_{j,r} \rho_j(r)$ and $\rho_+ = \max_j w_j$, one may take $L_a = \frac{3}{2}L_\rho + \rho_+ L_\gamma$ and $a_+ = \frac{3}{2}\bar\rho$.

For $0 \le u \le r_0/\sigma$, define

$$q_{j,\sigma}(u) = \frac{\gamma_j(\sigma u) - x_0}{\sigma}. \tag{41}$$

Integrating the derivative estimate in (33) yields

$$\|q_{j,\sigma}(u) - uv_j\| \le L_0 \sigma^\beta u^{1+\beta}, \qquad L_0 = \frac{L_\gamma}{1+\beta}. \tag{42}$$

The right-hand side is at most $u/2$. Every segment from uv_j to $q_{j,\sigma}(u)$ therefore satisfies Lemma 1.

Changing variables $r = \sigma u$ in (35) gives the local contributions

$$A_\sigma^{\mathrm{loc}}(z) = \sum_{j=1}^m \int_0^{r_0/\sigma} a_j(\sigma u) W_z(q_{j,\sigma}(u))\mathrm{d}u, \tag{43}$$

$$B_\sigma^{\mathrm{loc}}(z) = \sum_{j=1}^m \int_0^{r_0/\sigma} a_j(\sigma u)(q_{j,\sigma}(u) - z) W_z(q_{j,\sigma}(u))\mathrm{d}u. \tag{44}$$

The tangent quantities replace $a_j(\sigma u)$ by w_j, $q_{j,\sigma}(u)$ by uv_j, and the upper limit by infinity.

Let Q be compact and $R_Q = \sup_{z\in Q}\|z\|$. For the denominator integrand, add and subtract $a_j(\sigma u)W_z(uv_j)$. The mean-value theorem, (40), (42), and Lemma 1 give, uniformly over $z \in Q$,

$$\begin{aligned} &|a_j(\sigma u)W_z(q_{j,\sigma}(u)) - w_j W_z(uv_j)| \\ &\quad \le C_{R_Q}\sigma^\beta \left[L_a u^\beta + a_+ L_0 u^{1+\beta}(1+u)\right] e^{-cu^2}. \end{aligned} \tag{45}$$

If $H_z(q) = (q-z)W_z(q)$, the same argument gives

$$\begin{aligned} &\|a_j(\sigma u)H_z(q_{j,\sigma}(u)) - w_j H_z(uv_j)\| \\ &\quad \le C_{R_Q}\sigma^\beta \left[L_a u^\beta (1+u) + a_+ L_0 u^{1+\beta}(1+u^2)\right] e^{-cu^2}. \end{aligned} \tag{46}$$

Both right-hand sides are integrable on $[0,\infty)$, so their integrals are $O(\sigma^\beta)$.

The tangent tails beyond r_0/σ have the same bound. Indeed, $W_z(uv_j) \le e^{-(u-R_Q)^2/2}$ and $\|H_z(uv_j)\| \le (u+R_Q)e^{-(u-R_Q)^2/2}$. A Gaussian tail times any fixed polynomial is $O(\sigma^p)$ for every $p > 0$ when its lower limit is r_0/σ.

It remains to control μ_{far}. Choose $\sigma_0 \le \Delta/(2\max\{R_Q, 1\})$. For $z \in Q$ and $x \in \operatorname{supp}\mu_{\mathrm{far}}$,

$$\left\|\frac{x - x_0}{\sigma} - z\right\| \ge \frac{\Delta}{2\sigma}.$$

The far contributions to A_σ and B_σ are consequently bounded by constant multiples of

$$\sigma^{-1}\mu_{\mathrm{far}}(\mathbb{R}^2)e^{-\Delta^2/(8\sigma^2)} \quad \text{and} \quad \sigma^{-2}\mu_{\mathrm{far}}(\mathbb{R}^2)e^{-\Delta^2/(8\sigma^2)},$$

respectively. Both are $O(\sigma^\beta)$ because, for $a, p > 0$,

$$\sup_{t>0} t^{-p} e^{-a/t^2} = a^{-p/2} \left(\frac{p}{2e}\right)^{p/2} < \infty. \tag{47}$$

Combining the local, tangent-tail, and far-field estimates proves

$$\sup_{z \in Q} \{|A_\sigma(z) - A_0(z)| + \|B_\sigma(z) - B_0(z)\|\} \le C_{\tan,\mathrm{Q}} \sigma^\beta. \tag{48}$$

Let $\rho_- := \min_j w_j > 0$. Restricting each tangent integral to $0 \le u \le 1$ gives the explicit lower bound

$$a_Q := \inf_{z \in Q} A_0(z) \ge m \rho_- e^{-(R_Q+1)^2/2} > 0. \tag{49}$$

If $C_{\tan,\mathrm{Q}} \sigma^\beta \le a_Q/2$, then $A_\sigma \ge a_Q/2$. With $S_Q = \sup_{z \in Q} \|B_0(z)/A_0(z)\|$, direct ratio subtraction yields

$$\sup_{z \in Q} \left\| \frac{B_\sigma}{A_\sigma} - \frac{B_0}{A_0} \right\| \le \frac{2(1 + S_Q) C_{\tan,\mathrm{Q}}}{a_Q} \sigma^\beta. \tag{50}$$

The exact identity (31) proves (34). Projecting on $z = R\omega$ multiplies the bound by at most R, so the same rate holds for the shell tangential score. Finally, C^2 branches and C^1 densities satisfy (33) with $\beta = 1$. □

The constant $C_{\tan,\mathrm{Q}}$ depends on the compact query radius, branch count, local chart radius, far-field distance and mass, the curve and density Hölder bounds, and density upper bounds. The ratio constant additionally depends on positive tangent mass through a_Q. A minimum angular separation does not enter the forward approximation; it enters only the conditioning of the downstream finite-atom inverse.

Numerical rate check. Figure 3 evaluates the two Gaussian-weighted moments and their ratio for the $\beta \in \{1/2, 3/4, 1\}$ branch/density cases in Theorem B.1. The fitted small-σ slopes track the corresponding exponents; this sweep checks the implementation and illustrates the asymptotic regime but is not used in the proof.

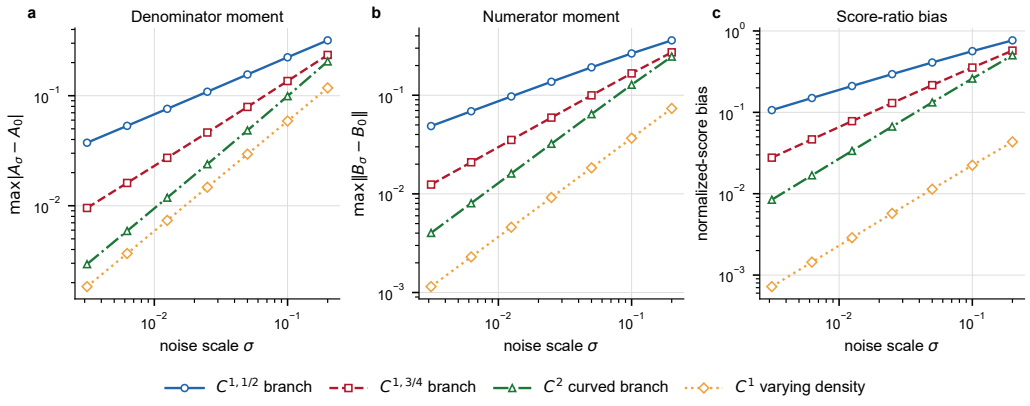

Figure 3: Finite-noise tangent-bias sweep. The panels report the maximum denominator error, numerator error, and normalized-score ratio error over a fixed normalized query set. Each curve is computed by Gaussian quadrature; the $\beta = 1/2$, $\beta = 3/4$, and $\beta = 1$ cases exhibit their predicted log–log exponents.

B.2 SHELL IDENTITY

For $z = R\omega$, substitute the polar form (1):

$$q_\Lambda(R\omega) = \int_{\mathbb{S}^{D-1}} \int_0^\infty e^{-(R^2+r^2-2Rr\omega^\top\theta)/2} r^{d-1} \mathrm{d}r \, \mathrm{d}\Lambda(\theta) \tag{51}$$

$$= e^{-R^2/2} \mathsf{T}_{d,R}\Lambda(\omega). \tag{52}$$

For any smooth scalar f in a neighborhood of the sphere, $\nabla_{\mathbb{S}^{D-1}} f(R\omega) = RP_{\omega^\perp} \nabla f(R\omega)$. Applying this to $\log q_\Lambda$ proves (9). The same identity holds weakly and then smoothly after convolution when Λ is a finite measure rather than a density.

C PROOF OF ONE-SHELL INJECTIVITY

Proof of Theorem 2. Let $H = \mathsf{T}_{d,R}\Lambda$ and $H' = \mathsf{T}_{d,R}\Lambda'$. The kernel is strictly positive and the measures are nonzero, so both transforms are positive. Equality of the observed fields gives

$$\nabla_{\mathbb{S}^{D-1}}(\log H - \log H') = 0. \tag{53}$$

The sphere is connected for $D \geq 2$, hence $H = cH'$ for some $c > 0$.

It remains to prove that $\mathsf{T}_{d,R}$ has no harmonic nullspace. Let $Y_{\ell,m}$ be a degree-ℓ spherical harmonic. The Funk–Hecke formula for the exponential kernel gives

$$\int_{\mathbb{S}^{D-1}} e^{Rr\omega^\top\theta} Y_{\ell,m}(\theta) \mathrm{d}\theta = c_{D,\ell}(Rr)^{1-D/2} I_{\ell+D/2-1}(Rr) Y_{\ell,m}(\omega), \tag{54}$$

where $c_{D,\ell} > 0$ under the usual harmonic normalization and the modified Bessel function $I_\nu(t)$ is strictly positive for $t > 0$. Integrating (54) against $r^{d-1}e^{-r^2/2}\mathrm{d}r$ shows that every degree-ℓ multiplier of $\mathsf{T}_{d,R}$ is strictly positive. Therefore $\mathsf{T}_{d,R}(\Lambda - c\Lambda') = 0$ implies that every spherical-harmonic moment of the finite signed measure $\Lambda - c\Lambda'$ vanishes. Finite linear combinations of spherical harmonics are uniformly dense in continuous functions on the sphere, so the signed measure is zero. Thus $\Lambda = c\Lambda'$. Unit mass forces $c = 1$. □

The theorem is injective but not uniformly well-conditioned over arbitrary angular measures. The multipliers in (54) can become small at large ℓ; finite-noise recovery must therefore restrict angular complexity or explicitly pay the inverse multiplier.

D FINITE POSITIVE RAYS IN ARBITRARY DIMENSION

We prove Theorem 3 and the general fixed-query lower bound. Let $\mathcal{P}_n(\mathbb{S}^{D-1})$ be the restrictions to the sphere of real polynomials of total degree at most n. Their harmonic decomposition is

$$\mathcal{P}_n(\mathbb{S}^{D-1}) = \bigoplus_{\ell=0}^{n} \mathcal{H}_\ell(\mathbb{S}^{D-1}). \tag{55}$$

Score-to-moment interface. Integrating the shell tangential score reconstructs the positive scalar transform up to a constant; unit-mass normalization fixes it. Dividing each harmonic coefficient by the positive multiplier in (54) recovers

$$L(p) = \int_{\mathbb{S}^{D-1}} p(\theta) \mathrm{d}\Lambda(\theta) \tag{56}$$

for every $p \in \mathcal{P}_n$ when harmonics through degree n are available.

Classical atom-recovery step. Write $\Lambda = \sum_{j=1}^s w_j \delta_{\theta_j}$ with $s \leq K$. Degree- $(s-1)$ Lagrange polynomials on the distinct support points show that evaluation of $\mathcal{P}_{K-1}$ is surjective. Thus, for a spanning vector Φ and $V_{aj} = \phi_a(\theta_j)$, one has $\operatorname{rank} V = s$ and

$$H := L(\Phi\Phi^\top) = V \operatorname{diag}(w) V^\top, \qquad \operatorname{rank} H = s, \tag{57}$$

$$H_{B,r} := L(x_r \Phi_B \Phi_B^\top) = V_B \operatorname{diag}(w) \operatorname{diag}(\theta_{1r}, \ldots, \theta_{sr}) V_B^\top, \tag{58}$$

where Φ_B selects any nonsingular $s \times s$ row submatrix V_B, and $H_B := L(\Phi_B \Phi_B^\top) = V_B \operatorname{diag}(w) V_B^\top$ is positive definite. Therefore

$$H_B^{-1} H_{B,r} = V_B^{-\top} \operatorname{diag}(\theta_{1r}, \dots, \theta_{sr}) V_B^\top. \tag{59}$$

These matrices commute, and their joint eigenvalue tuples are the directions θ_j; then $L(\Phi_B) = V_B w$ recovers the weights. Entries of H and $H_{B,r}$ have degree at most $2K-2$ and $2K-1$, respectively. This standard multivariate/spherical Prony step (Kunis et al., 2016; 2019) proves the stated degree bound; the preceding score-to-moment conversion is the paper-specific interface.

Proof of Proposition 2. Choose K pairwise disjoint coordinate patches $U_1, \dots, U_K \subset \mathbb{S}^{D-1}$. One direction in each patch contributes $D-1$ Euclidean chart coordinates. Parameterize the last normalized weight as $w_K = 1 - \sum_{j<K} w_j$ and restrict the others to a nonempty open simplex. The resulting labeled family is an open subset of $\mathbb{R}^{K(D-1)+K-1} = \mathbb{R}^{KD-1}$.

Any fixed collection of M scalar tangential observations is a continuous map from this open parameter set to $\mathbb{R}^M$. If $M < KD-1$ and the observations identified every measure, composing with the standard embedding $\mathbb{R}^M \hookrightarrow \mathbb{R}^{KD-1}$ would give a continuous injection from an open subset of $\mathbb{R}^{KD-1}$ into $\mathbb{R}^{KD-1}$ whose image lies in a lower-dimensional coordinate subspace. Invariance of domain says the image must be open, a contradiction. Therefore $M \geq KD-1$. □

E Exact recovery of finite positive planar rays

Identify $\mathbb{S}^1$ with angles $\phi \in [0, 2\pi)$. Let $H(\phi) = \mathsf{T}_{1,R}\Lambda(\omega_\phi) > 0$ and $\bar{H} = (2\pi)^{-1} \int_0^{2\pi} H(\phi) \mathrm{d}\phi$. Equation (9) gives $(\log H)'$ exactly. Integrating around the circle determines H up to a constant; the unit-mean density $h = H/\bar{H}$ removes precisely the irrelevant total mass of Λ.

The circular Fourier expansion of the zonal kernel is

$$K_{1,R}(\cos\phi) = \sum_{k \in \mathbb{Z}} \kappa_k(1,R) e^{ik\phi}, \qquad \kappa_k(1,R) = c \int_0^\infty e^{-r^2/2} I_{|k|}(Rr) \mathrm{d}r > 0, \tag{60}$$

with a common positive convention constant c. Consequently the normalized Fourier coefficients satisfy (10).

Proof of Theorem 4. For $a = 0, \dots, K$ and $j = 1, \dots, s$, let $V_{aj} = e^{-ia\theta_j}$. Positivity gives

$$T_K = V \operatorname{diag}(w) V^*, \qquad c^* T_K c = \sum_{j=1}^{s} w_j \left| \sum_{a=0}^{K} c_a e^{ia\theta_j} \right|^2. \tag{61}$$

The distinct-node Vandermonde matrix has full column rank, so $\operatorname{rank} T_K = s$. Because the weights are positive, a kernel vector c is equivalent to a degree-at-most-K polynomial vanishing at every $z_j = e^{-i\theta_j}$. Conversely, $\prod_j (z - z_j)$ has degree $s \leq K$, so the common unit-circle zeros of all kernel polynomials are exactly the nodes. Finally, $m_0, \dots, m_{s-1}$ form a nonsingular Vandermonde system for the weights. □

Proof of Proposition 1. For $K \geq 1$ and an angle α, consider the normalized regular measure

$$\Lambda_\alpha = \frac{1}{K} \sum_{j=0}^{K-1} \delta_{\alpha + 2\pi j/K}. \tag{62}$$

Its moments satisfy

$$m_k(\Lambda_\alpha) = \frac{e^{-ik\alpha}}{K} \sum_{j=0}^{K-1} e^{-2\pi i k j/K} = 0, \qquad 1 \leq k < K, \tag{63}$$

while $m_0 = 1$. If $\alpha - \beta$ is not an integer multiple of $2\pi/K$, then $\Lambda_\alpha \neq \Lambda_\beta$, although their moments through order $K-1$ coincide. Their next moments are $m_K(\Lambda_\alpha) = e^{-iK\alpha}$ and $m_K(\Lambda_\beta) = e^{-iK\beta}$, which are distinct. This proves both the failure below order K and the stated sharpness. □

This proof also explains why positivity matters. Without it, the quadratic form need not be a sum of nonnegative terms, Toeplitz rank need not equal the number of atoms, and cancellation can hide directions.

F VERTEX HESSIANS AND AN INFINITE COLLISION FAMILY

Define the tilted probability law

$$\Pi_z(\mathrm{d}u) = q_\Lambda(z)^{-1} e^{-\|z-u\|^2/2} \nu_\Lambda(\mathrm{d}u). \tag{64}$$

Differentiation under the integral gives the standard Gaussian-mixture identities

$$F_\Lambda(z) = \mathbb{E}_{\Pi_z}[U] - z, \qquad \nabla F_\Lambda(z) = \mathrm{Cov}_{\Pi_z}(U) - I_D. \tag{65}$$

At $z = 0$, radius and angle are independent. The radial density is proportional to $r^{d-1}e^{-r^2/2}$, hence

$$\mathbb{E}R = a_d = \sqrt{2}\frac{\Gamma((d+1)/2)}{\Gamma(d/2)}, \qquad \mathbb{E}R^2 = d. \tag{66}$$

Therefore $\mathbb{E}U = a_d m_\Lambda$ and $\mathbb{E}UU^\top = dQ_\Lambda$. Substitution into (65) proves (12).

Proposition 4 (Regular-junction Hessian collision)**.** *For every uniform regular planar q-ray angular measure with $q \geq 3$, $m_\Lambda = 0$ and $Q_\Lambda = I_2/2$. When $d = 1$, every such junction has $\nabla F(0) = -I_2/2$.*

Proof. Write the directions as $e^{2\pi ij/q}$. The first moment is the sum of the qth roots of unity and vanishes. Using $\cos^2\theta = (1+\cos 2\theta)/2$, $\sin^2\theta = (1-\cos 2\theta)/2$, and $\sin\theta\cos\theta = \sin 2\theta/2$, the second moment is $I_2/2$ because the degree-two root sum vanishes for $q \geq 3$. Equation (12) completes the proof. □

Thus a one-point Hessian cannot identify even the branch count within this simple infinite family. By contrast, the regular q-ray angular measure has a nonzero degree-q moment, which survives the nonzero shell multiplier.

G FINITE-QUERY MOMENT STABILITY

We now prove (14). With the unit-mean h defined above, let

$$\ell(\phi) = \log h(\phi), \qquad g(\phi) = \ell'(\phi). \tag{67}$$

By construction the continuous mean of h is one. Let $\mathcal{K}_M = \{-\lfloor M/2 \rfloor, \ldots, \lceil M/2 \rceil - 1\}$ and let $a_M(n) \in \mathcal{K}_M$ be the representative congruent to n modulo M. Set

$$\beta_M(n) = \begin{cases} 1/a_M(n), & a_M(n) \neq 0, \\ 0, & a_M(n) = 0, \end{cases} \tag{68}$$

where the second case records that spectral integration discards the aliased zero mode. If g_n denotes the continuous Fourier coefficient, define

$$A_M(g) = \sum_{n \notin \mathcal{K}_M} |g_n| \left| \beta_M(n) - \frac{1}{n} \right|. \tag{69}$$

Let $\widetilde{y}_k = M^{-1}\sum_j y_j e^{-ik\phi_j}$ be the discrete Fourier transform (DFT). For nonzero $k \in \mathcal{K}_M$, spectral integration sets $\widetilde{\ell}_k = \widetilde{y}_k/(ik)$ and sets the zero coefficient to zero. The sampling identity

$$\widetilde{g}_k = \sum_{r \in \mathbb{Z}} g_{k+rM} \tag{70}$$

shows that the noiseless nodewise integration error is bounded by (69). For the noise, Parseval and Cauchy–Schwarz give

$$\max_j \left| \sum_{k \in \mathcal{K}_M \setminus \{0\}} \frac{\widetilde{e}_k}{ik} e^{ik\phi_j} \right| \le \left(\sum_{k \neq 0} k^{-2} \right)^{1/2} \left(\sum_k |\widetilde{e}_k|^2 \right)^{1/2} \tag{71}$$

$$\le C_M \varepsilon_2. \tag{72}$$

Thus, up to one additive log-density constant,

$$\|\widetilde{\ell} - \ell\|_{\infty,M} \le \tau_M = A_M(g) + C_M \varepsilon_2. \tag{73}$$

The bound does not grow with M. If g has $r > 1$ bounded derivatives, integration by parts also gives the explicit regularity estimate

$$A_M(g) \le \|g^{(r)}\|_\infty \sum_{n \notin \mathcal{K}_M} |n|^{-r} \left| \beta_M(n) - \frac{1}{n} \right|. \tag{74}$$

For the positive analytic cone kernel all such derivatives exist, and they are uniform over a compact declared geometry class.

Exponentiate the reconstructed log density at the nodes and normalize its discrete mean. Write $\bar{h}_M = M^{-1} \sum_j h(\phi_j)$. The correct comparison is with $h(\phi_j)/\bar{h}_M$, not with the continuously normalized $h(\phi_j)$. The additive log constant cancels, and the remaining discrete normalization changes by at most a factor e^{τ_M} in either direction. Thus (73) implies

$$\max_j \left| \frac{\widetilde{h}_j}{h_j/\bar{h}_M} - 1 \right| \le e^{2\tau_M} - 1. \tag{75}$$

The true normalized density has Fourier coefficients $h_k = \rho_{|k|}(R) m_k$. Sampling aliases them:

$$\frac{1}{M} \sum_j h(\phi_j) e^{-ik\phi_j} = \rho_k m_k + \sum_{r \neq 0} \rho_{|k+rM|} m_{k+rM}. \tag{76}$$

Because Λ is a unit-mass positive measure, $|m_n| \le 1$. Define

$$B_{M,k}(R) = \sum_{r \in \mathbb{Z} \setminus \{0\}} \rho_{|k+rM|}(R). \tag{77}$$

Then $B_{M,k}$ bounds the kth density alias uniformly over the angular class. Since $\bar{h}_M$ is the aliased discrete zero mode, it lies in $[1 - B_{M,0}, 1 + B_{M,0}]$. Combining this denominator perturbation with (75), and then dividing the kth mode by ρ_k, yields

$$|\widehat{m}_k - m_k| \le \rho_k^{-1}(e^{2\tau_M} - 1) + \frac{B_{M,k}/\rho_k + B_{M,0}}{1 - B_{M,0}}, \tag{78}$$

provided $B_{M,0} < 1$. This proves (14).

H RANK, ROOT, WEIGHT, AND CENTER STABILITY

Table 2: The conditional certificate proceeds from count to directions to weights. Every row uses class information fixed before the field is inspected. Failure stops the procedure before the next output.

Stage	Declared class information	Sufficient test	Output on success / failure
Count	Minimum positive Toeplitz eigenvalue	Signal gap exceeds twice E_K	Branch count / abstain
Directions	Count, angular separation, and root margin	Appendix Rouché inequality	Angles within the declared tolerance / no directions
Weights	Direction bound and minimum Vandermonde singular value	Appendix perturbed-solve inequality	Normalized weights within tolerance / no weights

Let $\eta = \max_{1\le k\le K} |\widehat{m}_k - m_k|$ and set $E_K = (K+1)\eta$. Every entry of the Hermitian Toeplitz perturbation changes by at most η, hence its Frobenius and operator norms obey

$$\|\widehat{T}_K - T_K\|_{\mathrm{op}} \le \|\widehat{T}_K - T_K\|_F \le E_K. \tag{79}$$

If $\gamma_K(\Lambda) = \lambda^+_{\min}(T_K) > 2E_K$, Weyl's inequality places all s positive eigenvalues above E_K and all $K+1-s$ zero eigenvalues below E_K after perturbation. Thresholding at E_K therefore recovers the true rank. For an honest uniform certificate, the declared class must supply $\gamma_K \ge \Gamma_K > 2E_K$.

The signal gap itself exposes two sources of difficulty. Since $T_K = V\,\mathrm{diag}(w)V^*$,

$$\gamma_K(\Lambda) \ge w_{\min}\sigma_{\min}(V)^2. \tag{80}$$

If the minimum circular angle separation is δ_0, the off-diagonal Vandermonde Gram entries are Dirichlet sums bounded by $\csc(\delta_0/2)$. For modes $0{:}q$ and at most u rays, Gershgorin gives

$$\underline{v}^2_{q,u} := q + 1 - (u-1)\csc(\delta_0/2) \le \sigma_{\min}(V_{0:q})^2. \tag{81}$$

Whenever the right-hand side is positive and every branch weight is at least w_0, (80) gives $\gamma_q \ge w_0\underline{v}^2_{q,u}$. The executable certificate uses $(q,u) = (K,K)$ before rank is known, (s,s) for root recovery, and (K,s) for the weight solve. It also uses $d_{\min} \ge 2\sin(\delta_0/2)$ and the universal leading-coefficient bound below. If any conservative lower bound is nonpositive, it abstains. Values computed from the true synthetic geometry are retained only as oracle conditioning diagnostics for the separate empirical sweeps.

For directions, form the $(s+1)\times(s+1)$ Toeplitz matrix T_s and let c be its unit kernel vector, interpreted as the coefficients of the monic-up- to-scale annihilating polynomial. Put

$$E_s = (s+1)\eta, \qquad \zeta = \frac{\sqrt{2}E_s}{\gamma_s - E_s}, \qquad \gamma_s = \lambda^+_{\min}(T_s). \tag{82}$$

When $E_s < \gamma_s$, the Davis–Kahan theorem permits a phase alignment such that $\|\widehat{c} - c\|_2 \le \zeta$. Let $d_{\min} = 2\sin(\delta/2)$ and let c_s be the leading coefficient. For $s = 1$, set $d_{\min} = 2$ and interpret the empty product below as one; no separation assumption is needed. On the circle of radius $r < \min\{1, d_{\min}/2\}$ around a true root z_j,

$$|P_c(z)| \ge |c_s| r (d_{\min} - r)^{s-1}, \qquad |P_{\widehat{c}}(z) - P_c(z)| \le \sqrt{s+1}(1+r)^s\zeta. \tag{83}$$

Therefore the Rouché condition

$$\boxed{\sqrt{s+1}(1+r)^s\zeta < |c_s| r (d_{\min} - r)^{s-1}} \tag{84}$$

places one estimated root in each disjoint disk. Radial projection to the unit circle gives the safe angular bound $2\arcsin r$.

Let this angular bound be a and put

$$L_{K,s} = \sqrt{s\sum_{k=0}^{K} k^2}. \tag{85}$$

The leading-coefficient bound used by the executable certificate does not need the unknown roots. If $P_c(z) = c_s\prod_{j=1}^{s}(z - z_j)$ and $\|c\|_2 = 1$, then the monic coefficient of degree a has magnitude at most $\binom{s}{a}$, because every $|z_j| = 1$. Vandermonde's identity therefore gives

$$|c_s|^{-2} \le \sum_{a=0}^{s}\binom{s}{a}^2 = \binom{2s}{s}, \qquad |c_s| \ge \binom{2s}{s}^{-1/2}. \tag{86}$$

The perturbed Vandermonde matrix satisfies $\|\widehat{V} - V\|_{\mathrm{op}} \le L_{K,s}a$. Comparing the nonnegative least-squares residual at its minimizer with the residual at the true weights gives, before unit-mass renormalization,

$$\|\widetilde{w} - w\|_2 \le \frac{2\{\sqrt{K+1}\eta + L_{K,s}a\}}{\sigma_{\min}(V) - L_{K,s}a}, \tag{87}$$

provided the denominator is positive. Unit-mass renormalization is stable when the right-hand side, denoted q_w, is below $1/\sqrt{s}$. Indeed, $|\mathbf{1}^\top \widetilde{w} - 1| \le \sqrt{s} q_w$ and therefore

$$\left\| \frac{\widetilde{w}}{\mathbf{1}^\top \widetilde{w}} - w \right\|_2 \le \frac{(1+\sqrt{s})q_w}{1-\sqrt{s}q_w}. \tag{88}$$

Equations (80), (84), and (87) are sufficient certification tests when their required class-level bounds are declared in advance. Quantities evaluated using the true geometry are oracle diagnostics, not data-derived certificates.

H.1 Proof of the stagewise stability proposition

We prove Proposition 3. Weyl's inequality and $\|\widehat{T}_K - T_K\|_{\mathrm{op}} \le (K+1)\eta$ show that $\eta < \gamma_K/\big(2(K+1)\big)$ separates the s positive eigenvalues from the perturbed null eigenvalues. This establishes the count stage without solving for a single root or weight.

After count is known, Davis–Kahan controls the annihilating coefficient vector by

$$\|\widehat{c} - c\|_2 \le \frac{\sqrt{2}(s+1)\eta}{\gamma_s - (s+1)\eta}. \tag{89}$$

Substitution into (84), with $r \le \sin(a_0/2)$, gives an additional sufficient condition for the declared angular tolerance a_0. It depends on the root separation $d_{\min}$ and the annihilating polynomial, quantities absent from the count test. Given the angle bound, (87) adds the separate requirement $L_{K,s}a < \sigma_{\min}(V)$ and a tolerance on the renormalized weight error. Hence this sufficient full-geometry certificate is the intersection of the rank, root, and weight conditions, while the count certificate uses only the first. Whenever either later sufficient margin is smaller than η_{count}, the bounds exhibit an interval certified for count but not for full geometry. Failure of a later inequality is an abstention at that stage, not evidence that the earlier recovered rank is wrong.

Finally, condition on a validated calibration error $\|\widehat{x}_0 - x_0\| = \sigma\|\widehat{b} - b\|$. Querying around the estimated center translates the normalized shell by $\widehat{b} - b$. The mean-value theorem gives

$$|g_{\widehat{b}-b}(\phi) - g_0(\phi)| \le R\|\widehat{b} - b\| \sup_{\|z - R\omega_\phi\| \le \|\widehat{b}-b\|} \|\nabla F_\Lambda(z)\|_{\mathrm{op}}. \tag{90}$$

We denote the displayed supremum by C_{ctr}. This term is added before any rank, root, or weight threshold is applied.

I Finite-data and learned-score errors

We give the constants behind (17). Let $\mathcal{N}_{\mathrm{ctr}} = \bigcup_j \{z : \|z - z_j\| \le \varepsilon_{\mathrm{ctr}}\}$ contain every translated query allowed by the calibrated center-error bound. Assume the weighted tangent approximation

$$\sup_{z \in \mathcal{N}_{\mathrm{ctr}}} \{|A_\sigma(z) - A_0(z)| + \|B_\sigma(z) - B_0(z)\|\} \le C_{\tan}\sigma^\alpha \tag{91}$$

and define

$$a_* := \inf_{z \in \mathcal{N}_{\mathrm{ctr}}} A_0(z) > 0, \qquad S_0 := \sup_{z \in \mathcal{N}_{\mathrm{ctr}}} \|B_0(z)/A_0(z)\|. \tag{92}$$

If $C_{\tan}\sigma^\alpha \le a_*/2$, direct ratio subtraction yields

$$\max_j \|\sigma s_\sigma(x_j) - F_\Lambda(z_j)\| \le \frac{2(1+S_0)C_{\tan}}{a_*}\sigma^\alpha =: \varepsilon_{\mathrm{geom}}. \tag{93}$$

For the empirical KDE, let

$$W_{ij} = e^{-\|x_j - X_i\|^2/(2\sigma^2)}, \qquad V_{ij} = \frac{X_i - x_j}{\sigma} W_{ij}, \tag{94}$$

with expectations $a_j = \mathbb{E}W_{ij}$, $b_j = \mathbb{E}V_{ij}$ and $S = \max_j \|b_j/a_j\|$. The envelopes

$$0 \le W_{ij} \le 1, \qquad \|V_{ij}\| \le e^{-1/2}, \qquad \mathbb{E}\|V_{ij}\|^2 \le (2/e)a_j \tag{95}$$

follow from maximizing $re^{-r^2/2}$ and using $r^2 e^{-r^2} \le (2/e)e^{-r^2/2}$.

Centering gives $\|V_{ij} - b_j\| \le 2e^{-1/2}$ and $\mathbb{E}\|V_{ij} - b_j\|^2 \le (2/e)a_j$. We use the Hilbert-space Bernstein consequence of Pinelis (1994, Theorem 3.4):

$$\left\| \frac{1}{N}\sum_{i=1}^{N} Y_i \right\| \le \sqrt{\frac{2vt}{N}} + \frac{2Lt}{3N} \tag{96}$$

with probability at least $1 - 2e^{-t}$ for independent mean-zero vectors with $\|Y_i\| \le L$ and $\mathbb{E}\|Y_i\|^2 \le v$. Set $t = \log(4M/\delta)$ and

$$D_j = \sqrt{\frac{2a_j t}{N}} + \frac{t}{3N}, \qquad Q_j = \sqrt{\frac{4a_j t}{eN}} + \frac{4e^{-1/2}t}{3N}. \tag{97}$$

Scalar Bernstein for W_{ij}, vector Bernstein for V_{ij}, and a union bound imply that, with probability at least $1 - \delta$, simultaneously

$$|\overline{W}_j - a_j| \le D_j, \qquad \|\overline{V}_j - b_j\| \le Q_j. \tag{98}$$

Whenever $D_j < a_j$, another ratio subtraction gives

$$\left\| \frac{\overline{V}_j}{\overline{W}_j} - \frac{b_j}{a_j} \right\| \le \frac{Q_j + SD_j}{a_j - D_j}. \tag{99}$$

Suppose $a_j \ge a_- \sigma^d$ and put $\lambda = N a_- \sigma^d$. If

$$\sqrt{\frac{2t}{\lambda}} + \frac{t}{3\lambda} \le \frac{1}{2}, \tag{100}$$

then (99) simplifies to

$$\varepsilon_{\mathrm{KDE}} = 2\left[\sqrt{\frac{4t}{e\lambda}} + \frac{4e^{-1/2}t}{3\lambda} + S\left(\sqrt{\frac{2t}{\lambda}} + \frac{t}{3\lambda} \right) \right]. \tag{101}$$

This makes the denominator condition and the effective local coverage explicit, rather than hiding them inside big-O notation.

If the learned score has validated pointwise error

$$\max_j \|s_\theta(x_j, \sigma) - \widehat{s}_{N,\sigma}(x_j)\| \le \varepsilon_\theta(\sigma), \tag{102}$$

its normalized vector-field contribution is $\sigma\varepsilon_\theta$. Because $g(\phi) = R\langle t_\phi, F_\Lambda(R\omega_\phi)\rangle$, converting a normalized vector-field error into the scalar shell observation multiplies it by R. Adding (93), (101), the network term, and (90) therefore proves (17), with $\varepsilon_2 \le \varepsilon_{\mathrm{shell}}$. If d is replaced by an estimate $\widehat{d}$, fix a compact declared interval $[d_-, d_+] \subset (0, \infty)$ containing both values. Differentiation under the radial integral shows that each multiplier $\kappa_k(d, R)$ used by the finite inverse is continuously differentiable in d. Positivity and compactness therefore give

$$\max_{1 \le k \le K} \left| \log \rho_k(\widehat{d}, R) - \log \rho_k(d, R) \right| \le C_d |\widehat{d} - d|. \tag{103}$$

Since $|m_k| \le 1$, this changes the recovered moments by at most a constant multiple of ε_d. Inserting the combined result into the finite-query theorem and then the three perturbation tests proves the end-to-end guarantee with no additional probability loss.

J Counterexamples delimiting the observation model

The following examples separate limitations of particular detectors from the information available in the full score field. They are not assumptions or ingredients of the one-shell theorem.

J.1 Fine-scale Hessian collapse with retained score displacement

For an empirical Gaussian KDE, let $\pi_i(x)$ be the Gaussian posterior weight of sample X_i. Direct differentiation gives

$$\widehat{s}_{N,\sigma}(x) = \frac{\sum_i \pi_i(x) X_i - x}{\sigma^2}, \qquad \nabla^2 \log \widehat{p}_{N,\sigma}(x) = -\frac{I}{\sigma^2} + \frac{\mathrm{Cov}_{\pi(x)}(X)}{\sigma^4}. \tag{104}$$

If X_{i_*} is the unique nearest sample to x, every other posterior ratio decays as $\exp(-c_i/\sigma^2)$ for some $c_i > 0$. Hence

$$\sigma^2 \nabla^2 \log \widehat{p}_{N,\sigma}(x) \longrightarrow -I, \qquad \sigma^2 \widehat{s}_{N,\sigma}(x) \longrightarrow X_{i_*} - x. \tag{105}$$

At very fine noise, a Hessian-sign detector loses posterior-covariance information, whereas the full score retains the nearest-sample displacement.

J.2 The crossing/gap critical scale is detector- and measure-specific

For $t \in [-L, L]$ and half-gap $\Delta > 0$, consider

$$q_{\pm}^{0}(t) = (t, \pm|t|), \qquad q_{\pm}^{\Delta}(t) = \left(t, \pm\sqrt{t^2 + \Delta^2}\right). \tag{106}$$

We compare parameter measure $\mathrm{d}t$ with arc-length measure $w_\Delta(t)\mathrm{d}t$, where

$$w_\Delta(t) = \sqrt{\frac{2t^2 + \Delta^2}{t^2 + \Delta^2}}.$$

At the symmetric origin, the infinite parameter-measure control gives

$$H_{\mathrm{cross}}(0) = -\frac{1}{2\sigma^2} I_2, \qquad H_{\mathrm{gap,param}}(0) = \begin{pmatrix} -\frac{1}{2\sigma^2} & 0 \\ 0 & -\frac{1}{2\sigma^2} + \frac{\Delta^2}{\sigma^4} \end{pmatrix}. \tag{107}$$

Its separation-direction eigenvalue changes sign at $\sigma_c^{\mathrm{param}} = \sqrt{2}\Delta$.

Arc length changes the constant. Put $r = \Delta/\sigma$ and

$$m_2(r) = \frac{\int_{\mathbb{R}} u^2 \sqrt{(2u^2 + r^2)/(u^2 + r^2)} e^{-u^2} \mathrm{d}u}{\int_{\mathbb{R}} \sqrt{(2u^2 + r^2)/(u^2 + r^2)} e^{-u^2} \mathrm{d}u}. \tag{108}$$

The dimensionless separation-direction eigenvalue is $m_2(r) + r^2 - 1$.

Proposition 5 (Unique arc-length critical scale)**.** *Writing $s = r^2$ and $F(s) = m_2(\sqrt{s}) + s - 1$, F has one positive, transverse root in $(\sqrt{2} - 1, 1/2)$. Deterministic quadrature gives*

$$\frac{\sigma_c^{\mathrm{arc}}}{\Delta} = 1.5016756510. \tag{109}$$

Proof. Let π_s have density proportional to $\sqrt{(2t^2 + s)/(t^2 + s)} e^{-t^2}$ and write $m(s) = \mathbb{E}_s[t^2]$. Integration by parts yields

$$2m(s) = 1 + \mathbb{E}_s \frac{st^2}{(2t^2 + s)(t^2 + s)}, \tag{110}$$

so $1/2 < m(s) < 2 - \sqrt{2}$. These bounds give opposite signs at $s = \sqrt{2} - 1$ and $s = 1/2$. On that interval, differentiating under the integral and applying Popoviciu and Cauchy–Schwarz gives

$$F'(s) \geq 1 - \frac{\sqrt{2} - 1}{4} \sqrt{\frac{3\sqrt{2}}{4}} > 0.89335. \tag{111}$$

Thus the root exists, is unique, and is transverse. The displayed decimal is generated by deterministic quadrature. □

J.3 The coarse exponent depends on the reference measure

Let $L = \ell\sigma$, $r = \Delta/\sigma$, and define the physical score discrepancy on $\|x\| \leq c\sigma$ by

$$D_{\sigma,L}^{\mu}(c) := \sup_{\|x\| \leq c\sigma} \|s_{\sigma,L,\Delta}^{\mu}(x) - s_{\sigma,L,0}^{\mu}(x)\|, \tag{112}$$

where μ is either parameter measure or arc length.

Proposition 6 (Measure-dependent coarse-scale laws)**.** *For fixed $c, \ell > 0$ and $r \downarrow 0$, finite positive constants satisfy*

$$D^{\mathrm{arc}}_{\sigma,\ell\sigma}(c) = C_{\mathrm{arc}}(c,\ell)\frac{\Delta}{\sigma^2} + o\left(\frac{\Delta}{\sigma^2}\right),$$
$$D^{\mathrm{param}}_{\sigma,\ell\sigma}(c) = C_{\mathrm{param}}(c,\ell)\frac{\Delta^2}{\sigma^3} + o\left(\frac{\Delta^2}{\sigma^3}\right). \tag{113}$$

Proof. After reflection symmetry is used, the rescaled geometric kernel is analytic in r^2. Parameter measure therefore changes its density by $r^2 J + o_{C^1}(r^2)$. For arc length,

$$w_r(u) - \sqrt{2} = h(u/r), \qquad h(v) = \sqrt{\frac{2v^2+1}{v^2+1}} - \sqrt{2} \in L^1(\mathbb{R}), \tag{114}$$

and $\int h \neq 0$. This width-r boundary layer changes the density by $rH + o_{C^1}(r)$. A physical score is $\sigma^{-1}\nabla_z \log P$, producing the two orders in (113). The leading ratios H/P_0 and J/P_0 are nonconstant on coordinate lines, so both displayed constants are positive. □

The linear arc-length term comes from the mass element, not from branch displacement. Thus neither (Δ/σ) nor $(\Delta/\sigma)^2$ is a measure-independent score law; physical score units add another $1/\sigma$.

J.4 Why these facts do not prove multiscale necessity

At the common center, $g(y) = y_2^2 - y_1^2$ is zero on the crossing and equals Δ^2 on the gap. By (105), a sufficiently fine raw-score query recovers a nearest sample and distinguishes these supports. For a class with $\Delta \geq \Delta_{\min} > 0$, one common sufficiently fine scale works. The Hessian detector's bandwidth window is therefore not a fixed-scale lower bound for the full collection of score queries.

Small far-field density is also irrelevant to log-score magnitude. For a point mass,

$$p_\sigma(x) \propto e^{-\|x-y\|^2/(2\sigma^2)}, \qquad \nabla \log p_\sigma(x) = \frac{y-x}{\sigma^2}. \tag{115}$$

A deterministic check reaches $\log_{10} p_\sigma = -435.40$ while $\|s_\sigma\| = 800.99$. Density vanishes exponentially while the score grows with distance.

Consequently, heterogeneous feature sizes alone do not imply multiscale necessity under an absolute-error score oracle. The main result instead asks what one shell identifies.

K Experimental protocols and additional diagnostics

Table 3: Information permissions for the three reported inference roles. Only a certified output carries a sufficient guarantee, and only when its external error budget and declared class bounds are valid.

Role	Information beyond the queried field	Interpretation
Blind candidate	R, K, query grid, and numerical rank rule	Uncertified count, directions, and weights
Oracle diagnostic	True synthetic signal gap or geometry	Controlled diagnosis only; not deployable
Certified output	Valid moment-error budget and class bounds fixed in advance	Stagewise guarantee, otherwise abstention

K.1 Single-scale calibration and three-dimensional recovery

The weak-calibration sweep draws twenty separated positive ray cones for each $D \in \{2, 3, 5\}$. It uses $3(D+1)$ Gaussian test functions with normalized bandwidth $h = 0.75$ and tensor Gauss–Hermite rules of orders 15, 13, and 8, respectively. Each exact score is also perturbed by a deterministic smooth random field at root-mean-square magnitudes 0.003, 0.01, and 0.03. Test centers, cone

centers, directions, weights, and perturbations are all generated from recorded integer seeds. The table reports exact numerical quadrature error separately from perturbation sensitivity.

The spherical experiment alternates two and three non-coplanar rays over twenty seeds. At normalized radius $R = 2$, it reconstructs log shell density by 24-point Gauss–Legendre integration along each north-pole meridian on a 24×56 polar–azimuth quadrature grid. A 96-point rule computes the Gaussian–cone multipliers. Spherical harmonics through degree five are converted to polynomial moments, after which a pivoted moment block and three coordinate multiplication matrices recover the joint eigenvalue tuples. Full geometry requires the correct count, maximum angular error at most 0.1°, and maximum weight error at most 10^{-3}.

The learned calibration uses a non-coplanar weighted three-ray distribution in $\mathbb{R}^3$ translated away from the coordinate origin. Each of five seeds samples 50,000 points on length-four rays and trains a four-block residual SiLU MLP of width 192 for 8,000 AdamW updates with batches of 1,024. Denoising score matching draws noise levels 0.12, 0.18, and 0.25; calibration is evaluated at $\sigma = 0.18$ with the same weak Gaussian tests. The implementation records one aggregate row per seed and discards the model immediately after evaluation. It writes no checkpoint, score-prediction array, run manifest, or hash ledger.

K.2 Shell feasibility and finite-query evaluation

The deterministic shell-feasibility experiment verifies the regular three-ray/four-ray Hessian collision, computes their shell-field separation, and scans the inverse harmonic multipliers over normalized radius and degree. The finite-query matrix then varies

$$\begin{aligned} R &\in \{2, 3, 4\}, \\ M &\in \{12, 16, 24, 32, 48, 64\}, \\ \text{score-noise standard deviation} &\in \{0, 0.001, 0.01, 0.05\}. \end{aligned} \tag{116}$$

over seeds 0:11, angular separations from 10° to 90°, and minimum weights from 0.02 to 0.25. It includes two-, three-, and four-ray cases, an isolated close pair, a four-ray cluster, blind and oracle rank, and normalized center offsets through 0.8. These crossed factors produce 104,352 trials. Exact evaluation cases at $R = 3, M = 32$ all recover. The noisy results are reported as an empirical conditioning map, not as a replacement for the worst-case perturbation theorem.

Every numerical candidate uses $K = 4$. The blind rank rule sorts the five Toeplitz eigenvalues in decreasing order and maximizes the adjacent relative gap

$$r_s = \frac{\max\{\lambda_s, \epsilon_{\text{eig}}\}}{\max\{|\lambda_{s+1}|, \epsilon_{\text{eig}}\}}, \qquad s = 1{:}4, \tag{117}$$

after flooring magnitudes at

$$\epsilon_{\text{eig}} = \max\{10(K+1)\epsilon_{\text{mach}}\lambda_{\max}, 10^{-15}\}. \tag{118}$$

This relative-gap rule produces only an uncertified candidate. Whenever a valid moment-error bound $\bar{\eta}$ and a declared class gap are supplied, the certified path instead sets $E_K = (K+1)\bar{\eta}$ and counts the eigenvalues strictly above E_K. It then verifies that the declared gap exceeds $2E_K$ and abstains if this test fails. It never substitutes the largest internal signal eigengap for the zero-versus-positive threshold. For the selected rank s, the smallest-eigenvalue eigenvector of T_s supplies the annihilating coefficients. Reversing these coefficients exposes the roots, which are projected radially to the unit circle. Nonnegative least squares on the real and imaginary parts of modes $0{:}K$ then gives the normalized weights. The reported total comprises 103,680 conditioning trials and 672 artificial center-offset trials.

The separate certificate execution fixes its geometry class before any perturbation: $K = 4$, $w_0 = 0.10$, $\delta_0 = 80^\circ$, angle tolerance 5°, and weight tolerance 0.05. It includes regular and unequal three- and four-ray measures, all satisfying those bounds. For each of seven moment-error budgets and twelve seeds, every nonzero moment receives a random complex perturbation whose magnitude is at most the stated budget. A seed keeps the same perturbation direction while its magnitude changes across error budgets, producing paired error paths rather than independent trials. The code uses the worst-case Gershgorin bounds in (81), the coefficient bound (86), and bisection for the smallest radius satisfying the Rouché test. It then applies the normalized weight-error bound. This produces 336 trials. Automated checks verify that every count report is correct and every reported direction or weight error

lies below its computed bound. The implementation assumes a valid input error budget and standard floating-point arithmetic; its numerical floor protects linear algebra but is not an interval-arithmetic proof.

Table 4: Execution of the pre-specified count–direction–weight certificate. Entries are positive reports out of 48 bounded perturbations at each error level; unreported cases abstain at that stage.

Moment-error bound	Count	Directions	Full geometry
$1e-09$	48/48	48/48	48/48
$1e-08$	48/48	48/48	48/48
$1e-07$	48/48	48/48	48/48
$1e-06$	48/48	48/48	24/48
$1e-04$	48/48	24/48	0/48
$1e-03$	48/48	0/48	0/48
$1e-02$	0/48	0/48	0/48

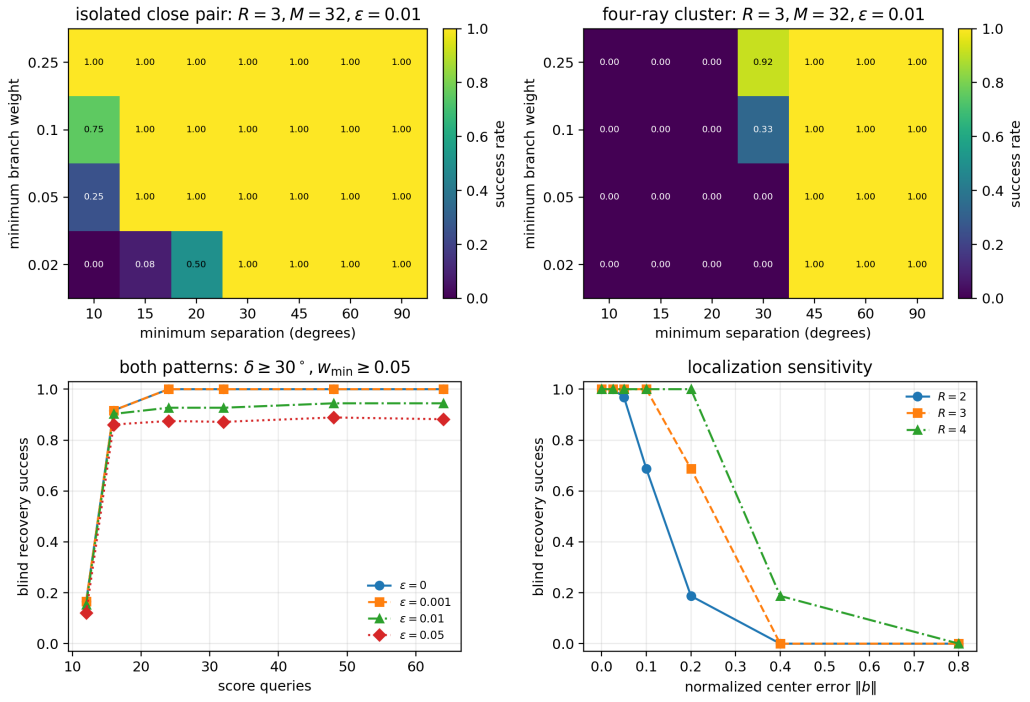

Figure 4: Finite-query recovery separates conditioning, query, noise, and center-offset failure modes. Rates aggregate 12 seeds per configuration; success requires correct count, maximum angular error 5°, and maximum weight error 0.05. Top: an isolated close pair and a four-ray cluster at matched minimum separation and weight. Bottom: query/noise sensitivity and center-offset sensitivity by shell radius. Cell numbers, markers, and line styles make the rates readable without relying on color.

K.3 Population, KDE, held-out geometry, and perturbations

The named geometry set contains an endpoint, a line, two corners, a regular Y, a crossing, an unequal four-ray junction, and a clustered four-ray junction. Twelve held-out instances are deterministically generated with two to four rays, continuous angles and weights, and minimum separation 30°; they are not selected from the named template library. Radial Poisson coverage ranges from 16 to 1024 over seeds $0{:}31$.

For each sample, the script computes the exact population score, the exact empirical Gaussian KDE score, blind recovery, the fixed-template baseline, and a one-point Hessian baseline based on principal-component analysis (PCA). The full target is count, angles, and weights rather than a geometry label. In normalized coordinates, a curved branch is parameterized as $r\theta + (0.02/2)r^2\theta^{\perp}$ for $0 \le r \le 9$; ambient perturbations add isotropic Gaussian noise of coordinate standard deviation 0.1, and center error translates every query by $(0.1, 0)$. Additional matrices apply each perturbation and all three together at coverages 256 and 1024. Twice the split-sample moment discrepancy plus the exactly computed population-to-tangent moment bias feed an oracle rank gate, which compares that proxy with half the true synthetic Toeplitz signal gap. This is neither a calibrated high-probability bound nor an implementation of the later root and weight tests. We therefore report its abstention and its full-recovery rate conditional on reporting a rank.

The 6,528 empirical trials comprise 4,480 coverage trials and 2,048 perturbation trials. In the coverage matrix, 810 fields are non-abstaining and 791 of those meet the full-geometry tolerance. In the perturbation matrix, the diagnostic reports a rank for 413 fields, and all 413 meet the same tolerance. These oracle-assisted conditional counts are kept separate from blind numerical recovery.

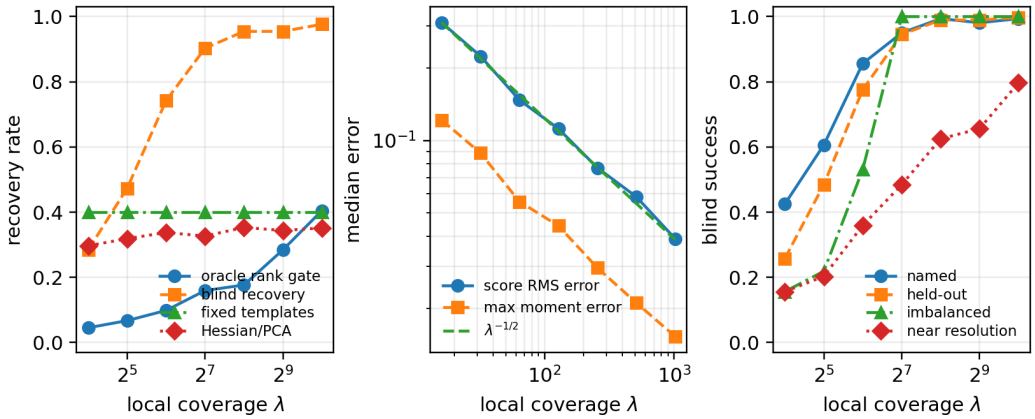

Figure 5: Empirical KDE recovery approaches the population field as local coverage grows. Across 6,528 trials, local coverage is the expected effective number of nearby samples receiving appreciable Gaussian weight at a query. Left: oracle-rank and blind recovery, with fixed-template and one-point Hessian/PCA checks; the rank gate uses the true synthetic signal gap. Middle: normalized-score RMS, maximum moment error, and an inverse-square-root reference slope. Right: blind recovery on named, held-out, imbalanced, and near-resolution subsets. Markers and line styles identify every curve without relying on color.

K.4 Single-geometry training-budget sensitivity

The sensitivity study fixes one weak four-ray geometry: angles $(0.05, 1.18, 3.02, 5.10)$ and weights $(0.05, 0.25, 0.30, 0.40)$. It crosses three dataset sizes (512, 2,048, and 8,192), three noise scales (0.1, 0.2, and 0.4), widths 32 and 128, 5 seeds, batch size 512, and 1,000 updates. We evaluate a time-conditioned network and single-noise specialists, extending four width-128 controls to 5,000 updates.

At the standard budget, 3 configurations had empirical KDE full success in every seed but learned undercount in 3/5 seeds. The 4 long controls then restored count in several settings; full weighted geometry remained outside tolerance in 80%–100% of seeds. These runs do not support irreversible branch loss as the cause of the earlier count errors. This single-geometry, single-MLP matrix measures training-budget sensitivity; the fixed evaluation below covers multiple geometries and parameterizations.

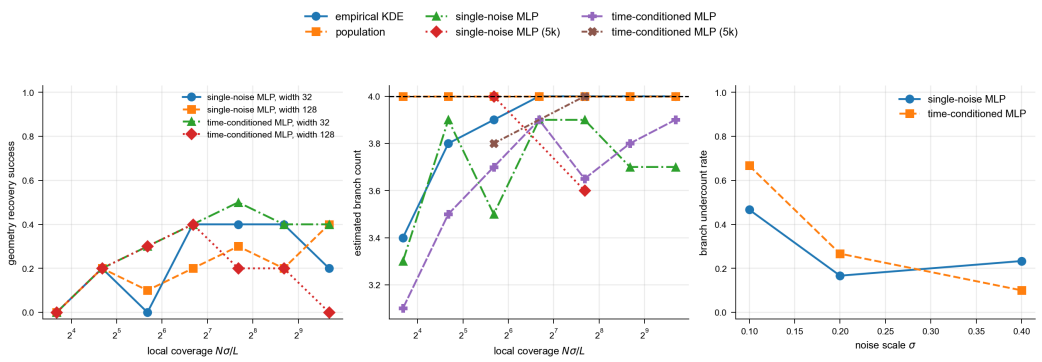

Figure 6: Longer training can restore count without full geometry. The study uses 5 seeds per configuration; its local-coverage proxy is $N\sigma/L$. Left: full-geometry success versus coverage. Middle: estimated count, with the true count marked by a dashed line. Right: undercount rate versus noise scale. Markers and line styles distinguish empirical, time-conditioned, and single-noise fields and their training budgets. The external legend identifies the six middle-panel fields; the other panels use local legends.

K.5 Cross-geometry evaluation

The fixed evaluation matrix uses the following geometries:

Table 5: Synthetic geometries used in the fixed cross-geometry learned-score evaluation.

Geometry	Angles (radians)	Weights
regular Y	$0.20, 2.2944, 4.3888$	$1/3, 1/3, 1/3$
weak Y	$0.25, 2.15, 4.55$	$0.05, 0.40, 0.55$
clustered four	$0.10, 0.75, 1.40, 4.20$	$0.25, 0.25, 0.25, 0.25$
weak four	$0.05, 1.18, 3.02, 5.10$	$0.05, 0.25, 0.30, 0.40$

The plain network has three hidden SiLU layers of width 128 and 33,794 parameters. The residual network has four residual SiLU blocks of width 64 and 33,666 parameters. Both are time-conditioned and trained with the same denoising score-matching construction, optimizer, batch size 512, and 5 seeds indexed from zero. A clean point first selects a branch according to its weight and a radius uniformly from $[0, 4]$. For each training item,

$$\sigma \sim \mathrm{Unif}\{0.1, 0.2, 0.4\}, \qquad Z \sim \mathcal{N}(0, I_2), \qquad \widetilde{X} = X/\sigma + Z, \tag{119}$$

and the per-coordinate denoising loss is

$$\mathcal{L}(\theta) = \frac{1}{2}\mathbb{E}\left\| f_\theta(\widetilde{X}, \sigma) + Z \right\|_2^2. \tag{120}$$

Optimization uses AdamW with learning rate 2×10^{-3}, weight decay 10^{-6}, and no learning-rate schedule. Dataset generation, model initialization, and minibatch/noise draws use deterministic separately offset random streams. The saved final-loss diagnostic is an exponential moving average over the last 50 updates with coefficient 0.9; it is not used as evidence for geometry recovery.

The standard matrix has $N \in \{512, 8192\}$, $\sigma \in \{0.1, 0.2, 0.4\}$, and 1,000 updates. There are 80 distinct standard trained models, each evaluated at all three noise scales. The long matrix has $N = 8192$, evaluation $\sigma = 0.1$, 5,000 updates from the same initial seed as its standard counterpart, and 40 trained models. Together with matched population and exact empirical KDE fields, this yields 520 saved rows and 280,000 optimizer updates.

The two complete learned-score scripts train 260 models for 500,000 optimizer updates. On the reported RTX 5070 Ti, the training-budget sensitivity script takes 8.16 minutes and the cross-geometry script takes 15.51 minutes (23.67 minutes in total); the larger PyTorch peak CUDA reservation is

24.0 MiB. The CPU fallback is exercised by the released pilots on the reported Ryzen 5 5600: the sensitivity pilot takes 18.5 seconds for 12,000 updates, while the cross-geometry pilot takes 12.0 seconds for 3,200 updates. These are observed costs, not extrapolated full-CPU runtimes. Run-metadata CSV rows record the software versions, device, driver, workload, wall time, and both peak CUDA allocation and reservation.

The seed-level outcomes are reported directly, without a combined study-level decision rule. At 5,000 updates, both parameterizations recover count in at least 4/5 seeds but full geometry in at most 2/5 on regular Y, clustered four, and weak four. On weak Y, count is correct in 3/5 plain MLP seeds and 2/5 residual MLP seeds. From 1,000 to 5,000 updates, count improves for both parameterizations on clustered four and weak four; regular Y is already count-correct at the shorter budget.

An oracle-assisted learned-field rank-screening budget is reported as a deliberately conservative diagnostic. It is

$$\eta_{\mathrm{rank}} = \eta_{\mathrm{learned}\to\mathrm{empirical}} + \eta_{\mathrm{empirical}\to\mathrm{population}} + \eta_{\mathrm{population}\to\mathrm{truth}}, \tag{121}$$

and it is compared with half the true synthetic Toeplitz signal gap. Every long learned field fails this inequality, so the diagnostic always abstains. Because both the population-to-truth term and the gap use the known synthetic geometry, this is not a deployable confidence statement and it does not execute the root or weight tests. It records only that this worst-case triangle budget is too conservative to validate individual learned reconstructions; it does not replace the directly evaluated count, angle, and weight outcomes.

Those outcomes are consistent with the decomposition in Proposition 3. Clustered four reaches 5/5 and 5/5 count-correct seeds in the plain and residual parameterizations while its median angular or weight error crosses the declared full-geometry tolerance. Weak four reaches 5/5 count-correct seeds in the plain network and 4/5 in the residual network; its median angular error is within the declared tolerance, but its median weight error remains outside it. Weak-Y count recovery varies across seeds under this budget. Across the tested geometries, correct count recovery can coexist with angular or weight errors above tolerance.

Table 6: Fixed learned-score evaluation (successes out of 5 seeds). "Emp." is exact empirical KDE full recovery; count and full columns show learned fields at 1k→5k updates. Angle (degrees) and weight columns are medians over count-correct 5k seeds only.

Geometry	Network	Emp.	Count 1k→5k	Full 1k→5k	Angle 5k	Weight 5k
clustered four	plain	4/5	3/5→5/5	0/5→1/5	6.501	0.055
clustered four	residual	4/5	3/5→5/5	0/5→0/5	5.344	0.068
regular Y	plain	5/5	5/5→5/5	2/5→2/5	1.853	0.065
regular Y	residual	5/5	5/5→5/5	2/5→1/5	2.221	0.092
weak four	plain	5/5	3/5→5/5	1/5→0/5	2.505	0.099
weak four	residual	5/5	2/5→4/5	0/5→0/5	3.083	0.111
weak Y	plain	5/5	1/5→3/5	1/5→2/5	1.836	0.048
weak Y	residual	5/5	0/5→2/5	0/5→0/5	2.217	0.072

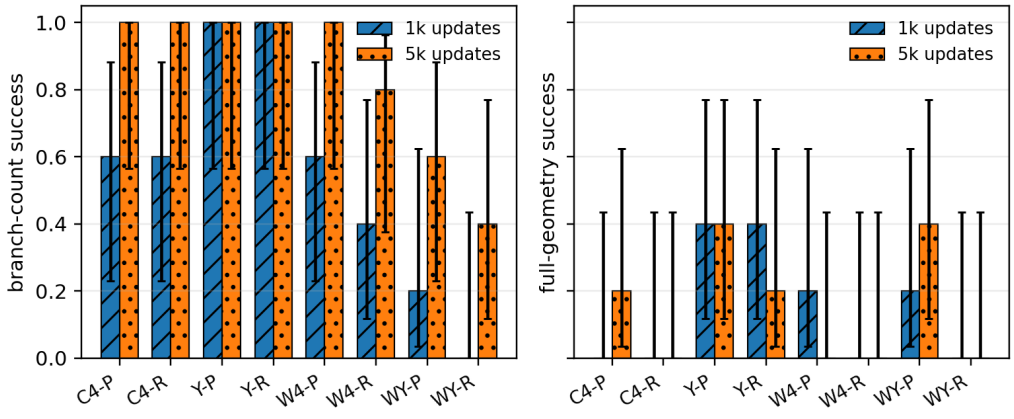

Figure 7: Fixed evaluation across four geometries and two parameter-matched multilayer-perceptron (MLP) parameterizations. Left: count recovery; right: full-geometry recovery. Hatching distinguishes 1k from 5k updates, and error bars are 95% Wilson intervals over 5 seeds. C4, Y, W4, and WY denote clustered four, regular Y, weak four, and weak Y; P and R denote plain and residual MLPs.

K.6 Train-to-convergence evaluation

The 1k-to-5k comparison above asks whether additional optimization stabilizes count recovery under the fixed inverse. The long-horizon study instead asks whether further improvements in normalized-score RMS also improve recovered moments and full geometry.

The convergence study retains the four geometries in Table 5, fixes $N = 8192$, and uses ten seeds per architecture–geometry cell. Plain and residual MLPs are parameter matched as above. A larger residual MLP with 132,866 parameters provides a capacity control on regular Y, weak Y, and clustered four. Every model trains through 50,000 optimizer updates, with evaluation every 500 updates and saved checkpoints at 1,000, 5,000, 10,000, 20,000, and 50,000 updates. Each trajectory uses one sampled training set. Validation and test evaluations use disjoint rotated shell grids against the analytic population score; empirical score diagnostics reuse the sampled training set.

The plateau diagnostic is the first evaluation at which the best validation normalized-score RMS improves by less than 1% across ten consecutive evaluations. It is recorded only after 20,000 updates, and all models continue to 50,000 updates. The reported checkpoint minimizes mean validation normalized-score RMS over the ten seeds and three trained noise scales among the saved checkpoints at or beyond 20,000 updates. Neither moment error nor a geometry-recovery outcome participates in selection.

All 110 trajectories satisfy the plateau rule. For the main plain/residual comparison, validation selects the plain MLP at 50,000 updates for every geometry. From 5,000 updates to that checkpoint, normalized-score RMS improves in every geometry, whereas maximum moment error increases in every geometry. The larger residual control also has higher validation score error than the selected plain model on each of its three geometries. Within this protocol, the downstream errors were not resolved by training through 50,000 updates or by the tested roughly fourfold parameter increase; other schedules and architectures remain untested.

Table 7: Validation-selected train-to-convergence results at $\sigma = 0.1$ (ten seeds per row). "Score RMS" is the root-mean-square error of the normalized score against its population value; moment error is the maximum error through order K. Full recovery requires correct count, maximum angle error at most $5°$, and maximum weight error at most 0.05.

Geometry	Arch.	Updates	Score RMS	Moment err.	Count	Full
clustered four	plain	50,000	0.0824	0.105	10/10	2/10
regular Y	plain	50,000	0.0993	0.133	10/10	2/10
weak four	plain	50,000	0.104	0.132	10/10	4/10
weak Y	plain	50,000	0.12	0.165	6/10	2/10

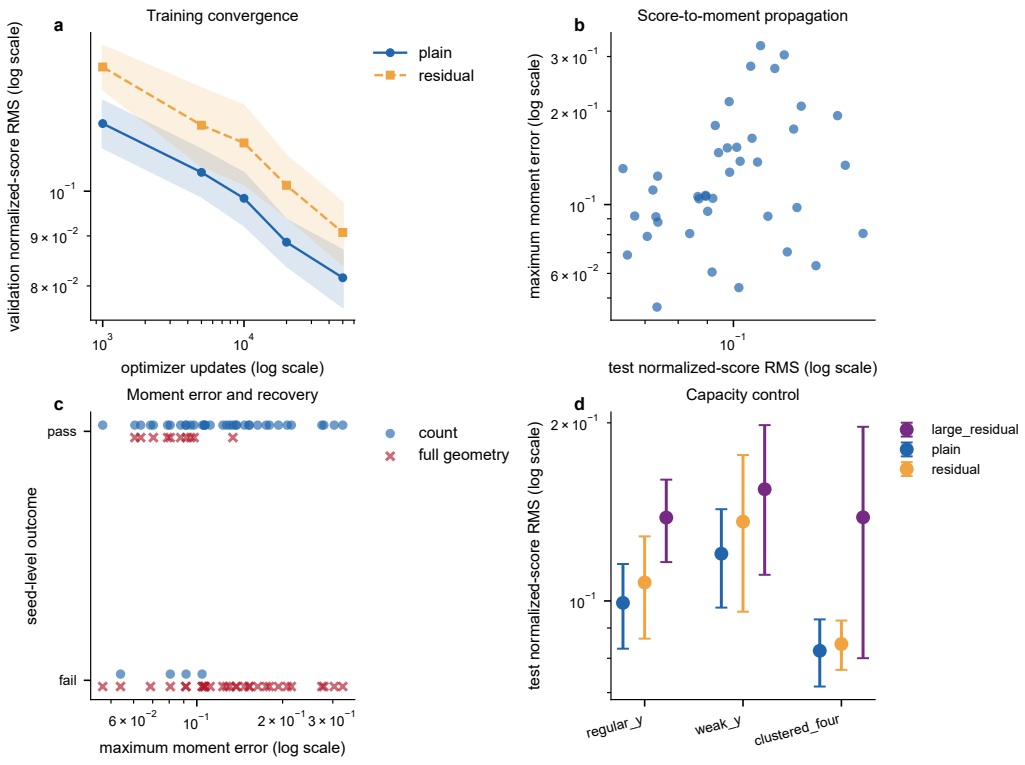

Figure 8: Train-to-convergence learned-score diagnostics. (a) Validation normalized-score RMS across optimizer updates. (b) Test normalized-score RMS versus maximum angular-moment error at saved checkpoints. (c) Seed-level count and full-geometry outcomes versus moment error. (d) Validation-selected comparison of parameter-matched plain/residual MLPs with a larger residual MLP on three geometries. Curves and intervals aggregate ten seeds; checkpoint selection uses validation score error only.

K.7 Tolerance sensitivity from the released rows

The main evaluation declares $5°$ maximum angular error and 0.05 maximum weight error. To expose the dependence on that choice without retraining or adding trials, we reaggregate the same 40 long-run learned fields over a grid of angle and weight tolerances.

Table 8: Pooled full-geometry successes over the released long-run learned fields. Rows give the maximum angle error and columns the maximum weight error; each entry is successes out of 40 fields. These fields span four geometries, two parameterizations, and 5 seeds, so the pooled entries are descriptive reaggregations rather than additional independent replications.

Angle tolerance	0.02	0.05	0.10
2°	0/40	2/40	7/40
5°	1/40	6/40	19/40
10°	1/40	8/40	23/40

The changes across columns show that the absolute full-recovery fraction is sensitive to the weight tolerance. Failures remain even at the most permissive displayed pair, while the stricter pairs expose different root and weight bottlenecks. We therefore retain the separate count, angle, and weight measurements as the primary evidence rather than treating any one threshold as a universal success definition.

L SCOPE OF THE RESULTS

Together, the proofs, crossing/gap derivations, and counterexamples establish the following scope:

- **Proved:** weak single-scale recovery of center and homogeneity under a full-rank test matrix; exact one-shell injectivity; constructive at-most-K finite-ray recovery in arbitrary dimension from moments through degree $2K-1$; the $KD-1$ fixed scalar-query lower bound; sharp degree-K planar recovery; finite-query moment, count, direction, and weight tests; and the finite-data local-coverage bound under the stated assumptions. For finite planar $C^{1,\beta}$ junctions with $C^{0,\beta}$ positive densities, the Gaussian-weighted tangent moments and normalized score converge at rate $O(\sigma^{\beta})$ on fixed compact normalized query sets.
- **Observed synthetically:** weak calibration in $D = 2, 3, 5$ under exact, perturbed, and GPU-trained three-dimensional score fields; non-coplanar three-dimensional score-shell recovery; held-out weighted-geometry recovery beyond fixed templates; inverse-square-root coverage scaling; valid execution of the complete bounded-error certificate; and the separation between count and full-geometry fidelity across four planar geometries and two MLP parameterizations. The train-to-convergence extension further observes that, across four tested geometries, the lower validation-selected normalized-score RMS at 50k updates can coexist with higher angular-moment error than at 5k updates. The oracle rank gate on learned fields is reported separately from these metrics.
- **Ruled out under the stated score-oracle formulation:** heterogeneous feature sizes alone implying multiscale necessity, universal fixed-scale failure, irreversible branch loss, and far-field density decay implying a weak log-score. We also do not claim universality across diffusion architectures or real data.

All reported numbers, tables, and figures are generated from executable code; the supporting-material README lists the environment, commands, and output files. Seeded neural optimization may vary across hardware, so the released raw rows can be reaggregated without retraining.