MATCHED SCORE QUERIES FOR CURVATURE AND DENSITY AT BRANCHING JUNCTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

At a junction a score field reveals weighted tangent rays, but these fix only first-order geometry. They say nothing about how a branch bends or how the density changes away from the center: the second-order structure needed to continue a branch past the junction. Small-noise expansions place this curvature and density in the first correction term, yet do not establish whether finite score observations can isolate the per-branch contributions when the center is also estimated. We answer this as an inverse problem, by matched queries at scales σ and $\lambda\sigma$. For a finite union of $C^{2,\alpha}$ half-branches in $\mathbb{R}^D$ the normalized score obeys $F_\sigma = F_0 + \sigma G + O(\sigma^{1+\alpha})$; matched subtraction cancels the leading field F_0 rather than estimating it, exposing G, which is linear in the branchwise curvature and log-density slope. Given the tangent directions and weights on distinct rays, G identifies all sD branch parameters. Point-coordinate queries attain this sharp scalar-information count, and any fixed continuous scheme that identifies every jet requires at least sD scalar outputs. This counts information rather than score-network evaluations. A geometry-dependent condition number sets stability as rays approach one another. Under coarse localization and full-rank calibration an $O(\sigma^2)$ center error contributes D translation modes, so the augmented problem needs $(s+1)D$ scalar observations except along a translation-invariant full line; we derive a perturbation bound and a conditional kernel-density rate. Across 180 population fits the fitted median exponent 0.998 matches the predicted $O(\sigma)$ order; matched-query systems remain full rank to $D = 20$ with 16 branches, and a known-count frontend fitted to finite-noise scores composes with the inverse in $D = 3$–5 (135 population systems, median jet error 0.132); under strong population first-order error, matched responses reduce the median parameter error by a factor of 49.4 relative to naive tangent subtraction. The contribution is an information count: matched subtraction turns a weak cross-scale signal into a well-posed inverse whose second-order branch geometry is identifiable at an sD (resp. $(s+1)D$ under center error) scalar-information cost.

1 INTRODUCTION

Score fields are learned by score matching (Hyvärinen, 2005), drive noise-conditional generative models (Song & Ermon, 2019; Song et al., 2021), and serve as local measurements of data geometry. Near a smooth manifold, they reveal tangent spaces, intrinsic dimension, and other first-order structure (Stanczuk et al., 2024; Ventura et al., 2025; Li et al., 2026). A junction, however, has no single tangent space. Its first-order description is a set of outgoing directions and relative masses. For local geometric recovery, that description is incomplete: it cannot tell how each branch continues away from the junction or how mass changes along it.

At leading order, a junction is therefore represented only by weighted tangent rays. The two neighborhoods in Figure 1 have the same such representation, yet bend differently and carry different outward density trends. Rescaling toward the junction removes these differences and leaves the same leading score field F_0. The missing quantities survive only in the next-order term. Recovering them requires isolating a weak cross-scale signal, assigning a superposition of responses to individual branches, and controlling center error at the same asymptotic order.

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.

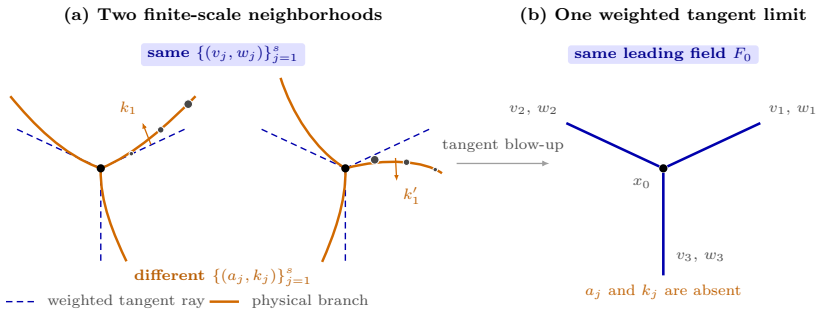


Figure 1: The same tangent geometry can hide different local structure. Tangent blow-up keeps branch directions and weights but removes curvature and outward density change.

Forward boundary-layer theory already gives two-term corner expansions whose first correction records density variation, the adapted Jacobian, and quadratic embedding geometry, together with score, log-Hessian, and scale-derivative expansions (Brosse & Dalalyan, 2026). We ask the converse finite-dimensional question: on supplied rays, which branch coefficients are identifiable from finite scalar observations, with what nullspace and information count, and under center error? Theorem 3.1 is the junction-specific forward bridge; Theorems 4.1, 4.2 and 5.3 solve the inverse.

The leading tangent field can be canceled rather than estimated. Consider a finite junction of one-dimensional branches in $\mathbb{R}^D$. Branch j has direction v_j, weight w_j, curvature vector $k_j \perp v_j$, and logarithmic density slope a_j; we call (a_j, k_j) its *branch jet*. The normalized score obeys

$$F_\sigma(z) = F_0(z) + \sigma G(z) + O(\sigma^{1+\alpha}), \tag{1}$$

where F_0 depends only on $\{v_j, w_j\}$ and G is linear in the branch jets. Curvature displaces mass normally to a tangent ray, whereas density slope changes mass along it. Gaussian smoothing preserves these distinct signatures inside G.

Querying the same normalized location z at scales σ and $\lambda\sigma$ exposes G directly:

$$\mathcal{D}_{\sigma,\lambda}(z) = \frac{\lambda\sigma s_{\lambda\sigma}(x_0 + \lambda\sigma z) - \sigma s_\sigma(x_0 + \sigma z)}{(\lambda - 1)\sigma} = G(z) + O(\sigma^\alpha). \tag{2}$$

The common field F_0 cancels. With known directions and weights, scalar component evaluations become rows of a linear inverse problem. The superposed field determines every branch jet without an angular-separation assumption, and sD scalar observations are necessary and sufficient. This is an information count, not a forward-pass count: a score network usually returns all D coordinates at one location. Quantitative stability is governed by the condition number, which captures ray collisions and vanishing weights.

The estimated center creates a second difficulty. Given an $O(\sigma)$ coarse localization and a full-rank weak calibration system, first-order score calibration gives $c_\sigma = x_0 + \sigma^2 b + O(\sigma^{2+\alpha})$. This physical error is small, but normalization promotes it to the same order as G. Reusing the provisional center at both scales gives

$$\mathcal{D}_{\sigma,\lambda}^{c_\sigma}(z) = G(z) - \lambda^{-1}\nabla F_0(z) b + O(\sigma^\alpha). \tag{3}$$

Thus $b \in \mathbb{R}^D$ contributes D known translation-shaped basis fields. Except when the tangent measure is invariant along a complete line, the branch jets and center bias remain jointly identifiable from $(s+1)D$ scalar observations, and estimating b refines the center to $O(\sigma^{2+\alpha})$.

Contributions.

1. Given the forward first correction on supplied distinct rays, the converse finite-dimensional problem exactly identifies every branch jet; quantitative stability is separated from exact injectivity.
2. Fixed continuous schemes have scalar-information complexity sD. The augmented inverse exposes the exact translation nullspace and jointly identifies the $O(\sigma^2)$ center bias and branch jets with complexity $(s+1)D$, except for the intrinsic full-line ambiguity.

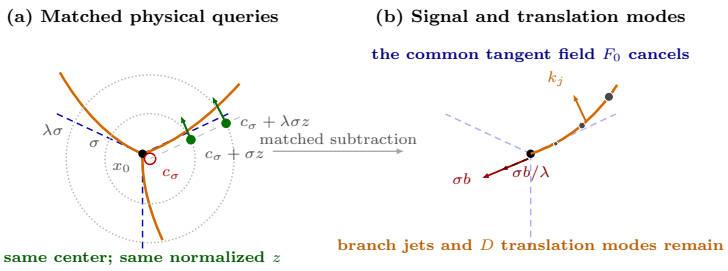


Figure 2: Matched queries expose the inverse signal and its translation modes. (a) Both scales use the same provisional center and normalized location. (b) Subtraction removes the tangent field F_0; branch jets remain, while the shared physical center produces the scale-dependent shifts σb and $\sigma b/\lambda$.

3. A locally uniform remainder and matched scale difference connect finite-noise score queries to the inverse by canceling the tangent field rather than estimating and subtracting it.
4. A perturbation bound separates finite-scale, score, center, and first-stage errors and yields a conditional KDE rate.

Evidence. Experiments test the complete path from score fields to branch jets. Population fits recover the predicted $O(\sigma)$ remainder; every known-count end-to-end system in $D = 3$–5 is full rank, with median relative jet error 0.132. Controlled perturbations extend to $D = 20$ and 16 supplied branches, while stability, response, KDE, and learned-score studies characterize separation, first-stage, sampling, and cross-scale effects.

2 Model and Second-Order Target

First-order geometry consists of tangent directions and relative masses; the second-order target is curvature and density change. Let $x_0 \in \mathbb{R}^D$ be the junction center. In a neighborhood of x_0, assume

$$\mu_{\mathrm{loc}} = \sum_{j=1}^{s} (\gamma_j)_{\#}\big(\rho_j(r)\,\mathrm{d}r\big), \qquad \gamma_j : [0, r_0] \to \mathbb{R}^D, \tag{4}$$

plus a finite remainder μ_{far} supported away from x_0. The curves are parameterized by arc length and satisfy

$$\gamma_j(0) = x_0, \qquad \gamma_j'(0) = v_j \in \mathbb{S}^{D-1}, \qquad \gamma_j''(0) = k_j \in v_j^{\perp}, \tag{5}$$

$$\rho_j(0) = w_j > 0, \qquad a_j = \partial_r \log \rho_j(0). \tag{6}$$

Distinct branches have distinct v_j. Arc-length parameterization implies $k_j \perp v_j$.

Assumption 2.1 (Uniform jet remainder). *For some $0 < \alpha \le 1$ and finite constants L_γ, L_ρ,*

$$\big\|\gamma_j(r) - x_0 - r v_j - \tfrac{1}{2} r^2 k_j\big\| \le L_\gamma r^{2+\alpha}, \tag{7}$$

$$|\rho_j(r) - w_j(1 + a_j r)| \le L_\rho r^{1+\alpha} \tag{8}$$

for $0 \le r \le r_0$, and $\mathrm{dist}(x_0, \mathrm{supp}\mu_{\mathrm{far}}) \ge \Delta > 0$.

The weighted tangent measure is

$$\nu_0 = \sum_{j=1}^{s} w_j \int_0^{\infty} \delta_{u v_j}\,\mathrm{d}u. \tag{9}$$

A score cannot see its total scale, so the weights may be normalized to sum to one. The known first-order geometry is $\{v_j, w_j\}_{j=1}^{s}$; the target is

$$\mathcal{J}_2(\mu, x_0) = \{(a_j, k_j) : j = 1, \dots, s\}. \tag{10}$$

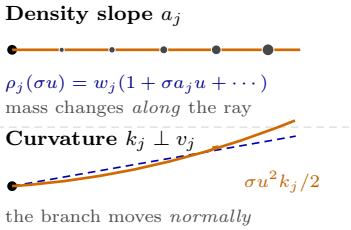
Here k_j measures bending and a_j measures outward log-density change.

Let $p_\sigma = \mu * \varphi_\sigma$, where φ_σ is the centered Gaussian density with covariance $\sigma^2 I_D$, and define

$$s_\sigma(x) = \nabla_x \log p_\sigma(x), \qquad F_\sigma(z) = \sigma s_\sigma(x_0 + \sigma z). \tag{11}$$

Write the repeated one-dimensional Gaussian integrals as

$$J_{j,m}(z) = \int_0^\infty u^m \exp\left(-\frac{\|z - uv_j\|^2}{2}\right) \mathrm{d}u. \tag{12}$$



The leading scalar transform and its first correction are

$$Q_0(z) = \sum_{j=1}^{s} w_j J_{j,0}(z), \tag{13}$$

$$Q_1(z) = \sum_{j=1}^{s} w_j \left[a_j J_{j,1}(z) + \frac{1}{2}\langle k_j, z\rangle J_{j,2}(z)\right]. \tag{14}$$

3 Why Second-Order Geometry Appears Across Scale

The mechanism is visible beside Equations (13) and (14). At physical radius $r = \sigma u$, both the density change $a_j r$ and the normal displacement $r^2 k_j/2$, after division by the observation scale σ, contribute at order σ. Gaussian smoothing therefore places both effects in the same correction G.

Theorem 3.1 (Second-order junction expansion)**.** *Under Assumption 2.1, for every compact $\mathcal{Q} \subset \mathbb{R}^D$ and every fixed integer $m \geq 0$, there are $C_{m,\mathcal{Q}} < \infty$ and $\sigma_0 > 0$ such that*

$$\left\|\widetilde{Q}_\sigma - Q_0 - \sigma Q_1\right\|_{C^m(\mathcal{Q})} \leq C_{m,\mathcal{Q}}\sigma^{1+\alpha}, \qquad 0 < \sigma \leq \sigma_0, \tag{15}$$

where $\widetilde{Q}_\sigma$ is the locally rescaled Gaussian convolution after removal of its z-independent normalization. Consequently, for $m \geq 1$,

$$\|F_\sigma - F_0 - \sigma G\|_{C^{m-1}(\mathcal{Q})} \leq C'_{m,\mathcal{Q}}\sigma^{1+\alpha}, \tag{16}$$

with

$$F_0 = \nabla \log Q_0, \qquad G = \nabla\left(\frac{Q_1}{Q_0}\right). \tag{17}$$

F_0 records the tangent rays, whereas G superposes all branch jets. This forward expansion is the technical bridge from finite-noise scores to the converse problem; Section 4 proves that the superposition is uniquely invertible.

The proof substitutes $r = \sigma u$ into the curve and density expansions, expands the Gaussian kernel once, and integrates under a uniform Gaussian envelope. The far-away measure is exponentially small, and positivity of Q_0 permits the logarithm. Section A gives the complete derivative and remainder argument.

For a fixed scale ratio $\lambda > 1$, define

$$\mathcal{D}_{\sigma,\lambda}(z) = \frac{F_{\lambda\sigma}(z) - F_\sigma(z)}{(\lambda - 1)\sigma}. \tag{18}$$

The two scores are evaluated at $x_0 + \lambda\sigma z$ and $x_0 + \sigma z$, respectively, so they share the same normalized location z.

Corollary 3.2 (Renormalized scale derivative)**.** *For fixed $1 < \lambda \leq \lambda_{\max}$ and compact $\mathcal{Q}$,*

$$\sup_{z\in\mathcal{Q}} \|\mathcal{D}_{\sigma,\lambda}(z) - G(z)\| \leq C_{\mathcal{Q},\lambda}\sigma^\alpha. \tag{19}$$

Equivalently, $G(z) = \partial_\sigma[\sigma s_\sigma(x_0 + \sigma z)]|_{\sigma=0^+}$.

Thus matched subtraction removes the common tangent field F_0 and exposes G without differentiating a noisy score.

If an exact tangent score were available, $(F_\sigma - F_0)/\sigma$ would have the same limit. The matched difference is preferable in practice because it cancels the leading field present in the observations themselves.

Algorithm 1 Two-scale branch-jet reconstruction

Require: First-order estimate $(c_\sigma, \{\hat{v}_j, \hat{w}_j\}_{j=1}^{\hat{s}})$, scales σ and $\lambda\sigma$, candidate normalized grid $\mathcal{Z}$.
1: Construct the branch basis fields in Equations (20) and (21); for center-robust recovery, append Equation (28).
2: Select scalar point-coordinate rows by rank-revealing QR, or retain an overdetermined design.
3: At each selected z, query the same center c_σ at $c_\sigma + \sigma z$ and $c_\sigma + \lambda\sigma z$.
4: Form Equation (26) and solve least squares.
5: Return $(\hat{a}_j, \hat{k}_j)$ and, when applicable, $\hat{x}_0 = c_\sigma - \sigma^2\hat{b}$.

4 Recovering Branch Jets

With the two-scale response in hand, two questions remain: does it uniquely determine the branch jets, and how many scalar observations are needed? Both become finite-dimensional once the directions and weights are known.

Choose an orthonormal basis $N_j = [n_{j,1}, \ldots, n_{j,D-1}]$ for $v_j^\perp$ and write $k_j = N_j\kappa_j$, with $\kappa_j \in \mathbb{R}^{D-1}$. Define

$$A_j(z) = \nabla\left(\frac{w_j J_{j,1}(z)}{Q_0(z)}\right), \tag{20}$$

$$K_{j,r}(z) = \nabla\left(\frac{w_j \langle n_{j,r}, z\rangle J_{j,2}(z)}{2Q_0(z)}\right), \quad r = 1, \ldots, D-1. \tag{21}$$

Then

$$G(z) = \sum_{j=1}^{s} a_j A_j(z) + \sum_{j=1}^{s}\sum_{r=1}^{D-1} \kappa_{j,r} K_{j,r}(z). \tag{22}$$

Thus each density slope and curvature coordinate multiplies a known response pattern.

Theorem 4.1 (Jet identifiability)**.** *Fix distinct directions $v_1, \ldots, v_s \in \mathbb{S}^{D-1}$ and positive weights $w_1, \ldots, w_s$. The map*

$$\{(a_j, k_j)\}_{j=1}^{s} \longmapsto G \tag{23}$$

is injective over $a_j \in \mathbb{R}$ and $k_j \in v_j^\perp$.

The geometric reason is that density changes act along a ray, while curvature acts normally to it; a localized observation can therefore separate both from every other branch.

The proof represents the correction as a distribution on the tangent rays. Because Gaussian convolution is injective, a zero field implies a zero distribution. A test supported in a narrowing tube isolates one ray, first its normal curvature and then its tangential density change. Repeating this test proves uniqueness for distinct rays, although stability may deteriorate as two rays approach. Section B gives the full argument.

Theorem 4.2 (Sharp scalar-information complexity)**.** *Under the conditions of Theorem 4.1, there exist sD scalar component evaluations of G whose values uniquely determine all branch jets. Conversely, any fixed continuous observation scheme that identifies every jet vector in a nonempty open subset of $\mathbb{R}^{sD}$ requires at least sD scalar outputs.*

Each branch contributes one density slope and $D-1$ curvature coordinates. One scalar observation is one coordinate at one location; because a network evaluation usually returns all D coordinates, sD is a scalar-information count, not a forward-pass count.

Point-coordinate evaluations span the dual of the sD-dimensional space generated by Equations (20) and (21), so sD of them suffice. For necessity, invariance of domain rules out a continuous injection from an open subset of $\mathbb{R}^{sD}$ into fewer coordinates. In computation, we select well-conditioned rows from a larger candidate matrix; Section C gives the formal proof.

5 Recovering Jets with an Imperfect Center

The previous section assumed the exact center x_0. A score-only first stage instead returns a nearby point. The key question is whether its error hides the order-σ branch signal or can be estimated with it.

Proposition 5.1 (Center accuracy after local calibration)**.** *Fix smooth localized score tests around a bounded reference offset. If the resulting exact tangent-model least-squares matrix has full column rank, then using the finite-noise score from Theorem 3.1 returns a physical center of the form*

$$c_\sigma = x_0 + \sigma^2 b + O(\sigma^{2+\alpha}) \tag{24}$$

for some bounded $b \in \mathbb{R}^D$.

In normalized coordinates, the calibration error is order σ; multiplying by the noise scale makes the physical error order σ^2.

The proposition is local: a coarse procedure must first place the calibration window within $O(\sigma)$ of x_0. Section D gives the full weak-system and least-squares expansion. More generally, we allow any first stage satisfying

$$c_\sigma = x_0 + \sigma^2 b + r_\sigma, \qquad \|r_\sigma\| \le C\sigma^{2+\alpha}. \tag{25}$$

Use this same physical center at both noise scales and define

$$\mathcal{D}^{c_\sigma}_{\sigma,\lambda}(z) = \frac{\lambda\sigma s_{\lambda\sigma}(c_\sigma + \lambda\sigma z) - \sigma s_\sigma(c_\sigma + \sigma z)}{(\lambda - 1)\sigma}. \tag{26}$$

The two red translation modes in Figure 2 show the resulting mismatch between normalized views.

Theorem 5.2 (Center-bias expansion)**.** *Under Assumption 2.1 and Equation* (25)*, uniformly on compact normalized query sets,*

$$\mathcal{D}^{c_\sigma}_{\sigma,\lambda}(z) = G(z) - \frac{1}{\lambda}\nabla F_0(z)b + O(\sigma^\alpha), \tag{27}$$

where ∇F_0 is the $D \times D$ Jacobian of the tangent score.

The shared physical offset becomes σb at the small scale and $\sigma b/\lambda$ at the large scale, so their mismatch has a known translation shape.

The proof expands F_σ at the two normalized offsets and subtracts them. All quadratic shift terms are smaller than the retained order. The complete Taylor calculation is in Section D.

Append the D known translation fields

$$C_r(z) = -\lambda^{-1}\nabla F_0(z)e_r, \qquad r = 1, \dots, D. \tag{28}$$

The inverse now has $(s+1)D$ coordinates. The only intrinsic exception is a tangent measure that is unchanged by translation along a nonzero direction; for a finite positive ray junction, this requires a complete line formed by opposite rays with matching density.

Theorem 5.3 (Center-robust identifiability and scalar information)**.** *If ν_0 is not invariant under any nonzero translation, the map*

$$(\{a_j, k_j\}_{j=1}^s, b) \longmapsto G - \lambda^{-1}\nabla F_0 b \tag{29}$$

is injective. Consequently, $(s+1)D$ fixed scalar component evaluations are sufficient, and fewer than $(s+1)D$ fixed continuous scalar observations cannot identify every parameter vector in an open set.

Unless the junction itself is translation-invariant along a full line, a center shift cannot imitate branchwise curvature and density changes.

A narrowing-tube test first removes curvature and center components normal to each ray. The endpoint term can cancel only in the full-line case. Otherwise $b = 0$, and Theorem 4.1 removes the density terms. Section D gives the proof and formal invariance criterion.

Estimating b immediately refines the center:

$$\hat{x}_0 = c_\sigma - \sigma^2\hat{b}. \tag{30}$$

If $\hat{b} - b = O(\sigma^\alpha)$, then $\hat{x}_0 - x_0 = O(\sigma^{2+\alpha})$.

6 Error Propagation and Finite Samples

Finite-scale approximation, score noise, and first-stage geometry error determine accuracy. Let $p = sD$ for a known center and $p = (s+1)D$ for joint center recovery. Select $M \geq p$ scalar evaluations, stack the exact basis rows into $B \in \mathbb{R}^{M\times p}$, and let y be the corresponding two-scale response for parameter vector θ.

Theorem 6.1 (Deterministic perturbation bound)**.** *Assume $\sigma_{\min}(B) = \gamma > 0$. Let the implemented design be $\widehat{B} = B + E$ and the observed response be $\widehat{y} = y + e$. If $\|E\|_{\mathrm{op}} < \gamma$, the least-squares estimator obeys*

$$\left\|\widehat{\theta} - \theta\right\| \leq \frac{C_{\mathrm{bias}}\sigma^{\alpha} + \|e\| + \|E\|_{\mathrm{op}}\|\theta\|}{\gamma - \|E\|_{\mathrm{op}}}. \tag{31}$$

If the two normalized score fields have stacked errors ε_σ and $\varepsilon_{\lambda\sigma}$, and the center remainder in Equation (25) *is r_σ, one may take*

$$\|e\| \leq \frac{\varepsilon_\sigma + \varepsilon_{\lambda\sigma}}{(\lambda-1)\sigma} + C_{\mathrm{ctr}}\frac{\|r_\sigma\|}{\sigma^2}. \tag{32}$$

Errors in recovered directions and weights enter through E by smooth dependence of the basis on the first-order geometry over any declared separated, positive-weight class.

The bound makes conditioning explicit: finite-scale bias and response noise are amplified by the inverse smallest singular value of the selected design.

The bound follows from the pseudoinverse identity and Weyl's inequality; Section E gives the proof. Its main message is the decomposition: scale differencing divides normalized-score noise by σ, while an unmodeled physical center remainder is divided by σ^2.

Corollary 6.2 (Kernel-density score rate)**.** *Assume a fixed finite query design, a positive lower bound on local tangent mass, bounded branch jets, and the conditions of Theorem 6.1. For N independent samples and fixed $\lambda > 1$, the score of the empirical Gaussian kernel density estimate (KDE) gives, with probability at least $1-\delta$,*

$$\left\|\widehat{\theta} - \theta\right\| \leq C\left[\sigma^{\alpha} + \sqrt{\frac{\log(M/\delta)}{N\sigma^3}} + \frac{\log(M/\delta)}{N\sigma^2} + \textit{first-stage error}\right]. \tag{33}$$

Ignoring the lower-order Bernstein term and assuming a comparably accurate first stage,

$$\sigma_{\mathrm{opt}} \asymp \left(\frac{\log(M/\delta)}{N}\right)^{1/(2\alpha+3)}, \qquad \left\|\widehat{\theta} - \theta\right\| = O_{\mathbb{P}}\left(\frac{\log(M/\delta)}{N}\right)^{\alpha/(2\alpha+3)}. \tag{34}$$

The variance term reflects two losses: only $N\sigma$ samples fall locally, and the scale difference contributes another factor $1/\sigma$.

The rate balances the deterministic remainder σ^α against the scale-differenced sampling fluctuation. A learned score enters the same response bound through its errors at both scales. Thus cross-scale consistency, rather than accuracy at either scale alone, is the quantity that controls second-order recovery.

7 Experiments

Each experiment tests one proof interface: finite-scale expansion, query conditioning, sampling, first-stage composition, or learned-score consistency. All results use the retained **paper** configuration; Section F gives full grids, seeds, hardware, and tables.

Finite-scale expansion. Across 180 population fits in $D \in \{2, 3, 5\}$, the median remainder exponent is 0.998 (5th–95th percentile 0.957–1.03), with median log–log R^2 1. Exact- and $O(\sigma^2)$-biased-center runs follow the predicted order in Theorems 3.1 and 5.2, providing an independent numerical check of the analytic results (Figure 4).

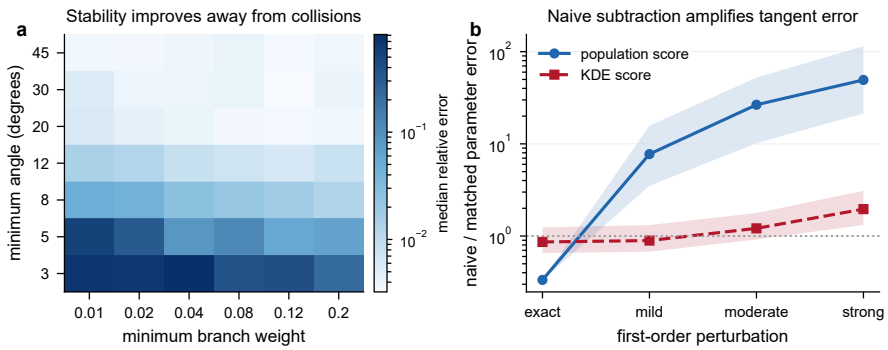


Figure 3: Stability and response construction. (a) Median error over 20 seeds per angle–weight cell grows as rays collide or weights vanish. (b) Paired naive/matched error ratio; values above one favor the matched response. Bands are interquartile ranges.

Scalar information versus stable designs. At response-noise RMS 10^{-3} and twice the parameter count, determinant-maximizing (D-optimal) row selection attains median relative error 0.00552, 2.89 times lower than random selection. Overdetermination and conditioning-aware rows increase the smallest singular value: Theorem 4.2 gives the information limit, whereas design controls noise amplification.

How geometry controls stability and scale. The 42-cell phase map in Figure 3a has 20 seeds per cell. From $(3°, 0.01)$ to $(45°, 0.2)$, median error falls from 0.698 to 0.00379 (184-fold) and condition number from 5.61×10^4 to 24.6. All 200 scale-up systems remain full rank through $D = 20$, 16 branches, 340 unknowns, and 680 queries. These runs use perturbed supplied geometry; the frontend study below covers $D = 3$–5. In the largest case, strong first-order perturbation raises error from 0.0825 to 85.6, so first-stage accuracy controls error even when rank is preserved.

Why match two scales? Naive subtraction divides first-order error by the base scale. Its population error is 7.75, 26.6, and 49.4 times the matched error under mild, moderate, and strong perturbations; the smallest-scale strong ratio is 143. KDE noise reduces it to 1.96. The exact-F_0 curve is an idealized reference for perfect nuisance removal; with estimated geometry, matched responses avoid the $1/\sigma$ amplification (Figure 3b).

Finite-sample kernel density estimates. Across four planar junctions and twenty seeds, all seed-bootstrap 95% intervals contain the predicted $N^{-1/5}$ slope, and error falls by 52.1–66%. Seed halves select bandwidth multipliers within 0.4 in 68.8% of comparisons; the dense $[1.8, 3.2]$ sweep identifies a broad stable operating region across geometries (Figure 5).

Estimated first-order frontends. On planar junctions, a Fourier-based frontend estimates branch count, directions, and weights. Blind count accuracy is 100% for population and 98.8% for KDE scores. A separate known-count frontend fits the same first-order quantities directly from finite-noise scores in $D = 3$–5 with up to 6 branches. All 270 second-stage systems are full rank. Population runs have median maximum tangent error 0.0197 radians and median relative jet error 0.132. With KDE scores, these medians are 0.0222 and 0.957; at $\sigma = 0.08$, the dimensionwise median jet errors range from 0.688 to 0.810 (Table 1). Changing the query rows has negligible effect (median error ratio 1). In population runs, the estimated basis absorbs part of the finite-scale remainder, giving an estimated/exact-basis error ratio of 0.611 (Figure 7).

Learned scores. Across 20 base jobs, the training trajectories separate single-scale score fit from recovered-jet accuracy. In the enlarged hard-case study (width 256, 6 blocks, 50000 updates), median jet error reaches 0.511 at 40000; score RMS continues to improve to 0.0143 at the final checkpoint, where jet error is 0.676. This controlled synthetic experiment isolates cross-scale consistency as the mechanism used by the inverse (Figures 7 and 8).

D	Pop. angle	Pop. jet error	KDE angle	KDE jet error
3	0.0212	0.142	0.0235	0.975
4	0.0167	0.124	0.0206	0.967
5	0.0209	0.125	0.0219	0.927

Table 1: Known-count score-to-jet composition in $D = 3$–5. Entries are medians over 45 cases per score source and dimension; angle errors are in radians, jet errors are relative, and every displayed design is full rank.

The support material contains the experiment code and its full configuration; each entry point writes its numerical observations as CSV files (Section F).

8 Related Work

Classical work reconstructs manifolds and estimates tangent spaces or curvature from samples (Federer, 1959; Preiss, 1987; Niyogi et al., 2008; Ozertem & Erdogmus, 2011; Genovese et al., 2012; Aamari & Levrard, 2019); score methods recover first-order and smooth-support geometry, including weighted singular tangents (Stanczuk et al., 2024; Ventura et al., 2025; Diepeveen et al., 2025; Kharitenko et al., 2026; Li et al., 2026; Zhao & Ni, 2026). Our target is instead the branchwise second-order continuation of a singular junction on supplied rays.

Denoising identities expose local scores (Vincent, 2011; Alain & Bengio, 2014), and forward asymptotics derive their density and curvature corrections (Brosse & Dalalyan, 2026; Zhang et al., 2026; Rawal, 2026). In particular, the boundary-layer expansion of Brosse & Dalalyan (2026) records density variation, the adapted Jacobian, and quadratic embedding geometry in its first correction. It also derives score, log-Hessian, and scale-derivative expansions. These results map known geometry to a correction. We solve the converse coefficient problem: uniquely decomposing the superposed correction with finite scalar information while accounting for center translation. Unlike Prony and super-resolution, the unknowns are coefficients on supplied rays, not atom locations (Kunis et al., 2016; 2019; Candès & Fernandez-Granda, 2014).

9 Limitations and Conclusion

Limitations. The theory covers zero-thickness $C^{2,\alpha}$ half-branches, positive $C^{1,\alpha}$ densities, Gaussian smoothing, two known scales, branch correspondence, and supplied first-order geometry. The condition number quantifies the effect of ray collisions and vanishing weights, and the KDE rate is conditioned on first-stage accuracy. Experiments separate full-pipeline recovery with known counts in $D = 3$–5 and blind planar counts from second-stage scaling through $D = 20$ and 16 branches with supplied perturbed geometry. The full-line case exhibits the intrinsic translation symmetry, while learned-score runs isolate cross-scale consistency on controlled synthetic junctions.

Conclusion. The first score correction has an inverse role: on supplied distinct rays, its superposed field uniquely determines how every branch bends and how its density changes, and matched subtraction makes it observable with sharp scalar-information complexity. Translation modes accommodate imperfect localization, and known-count experiments carry score-estimated first-order geometry through the inverse in $D = 3$–5, while the supplied-geometry interface scales to larger systems. The stability map, matched-response comparison, and learned-score trajectories identify the geometric and statistical conditions that govern accuracy, so branch jets test whether a score field retains junction continuation and provide a target for thicker supports, higher-dimensional strata, and pretrained models.

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.

Reproducibility statement

The model assumptions and theorem statements appear in the main text, with complete proofs in Sections A to E. Section F summarizes the experiment-to-claim map, data generation, numerical procedures, seeds, optimization, evaluation steps, hardware, and runtime. The accompanying anonymous support material contains the full configuration in `configs/paper.yaml`, the experiment entry points, and their shared numerical routines.

Ethics statement

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

References

Eddie Aamari and Clément Levrard. Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, 47(1):177–204, 2019. doi: 10.1214/18-AOS1685.

Guillaume Alain and Yoshua Bengio. What regularized auto-encoders learn from the data-generating distribution. *Journal of Machine Learning Research*, 15(110):3743–3773, 2014.

Nicolas Brosse and Arnak S. Dalalyan. Boundary-layer asymptotics for gaussian-smoothed singular measures. arXiv:2607.04514, 2026.

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.

Willem Diepeveen, Georgios Batzolis, Zakhar Shumaylov, and Carola-Bibiane Schönlieb. Score-based pullback riemannian geometry: Extracting the data manifold geometry using anisotropic flows. In *International Conference on Machine Learning*, 2025.

Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3): 418–491, 1959. doi: 10.1090/S0002-9947-1959-0110078-1.

Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Minimax manifold estimation. *Journal of Machine Learning Research*, 13(43):1263–1291, 2012.

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005.

Andrey Kharitenko, Zebang Shen, Riccardo De Santi, Niao He, and Florian Dörfler. Landing with the score: Riemannian optimization through denoising. In *International Conference on Learning Representations*, 2026.

Stefan Kunis, Thomas Peter, Tim Römer, and Ulrich von der Ohe. A multivariate generalization of prony's method. *Linear Algebra and its Applications*, 490:31–47, 2016.

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.

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.

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

Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1–3):419–441, 2008. doi: 10.1007/s00454-008-9053-2.

Umut Ozertem and Deniz Erdogmus. Locally defined principal curves and surfaces. *Journal of Machine Learning Research*, 12(34):1249–1286, 2011.

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

David Preiss. Geometry of measures in $\mathbb{R}^n$: Distribution, rectifiability, and densities. *Annals of Mathematics*, 125(3):537–643, 1987.

Divit Rawal. Rao-blackwellized score matching on manifolds. arXiv:2605.25567, 2026.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

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.

Jan Pawel Stanczuk, Georgios Batzolis, Teo Deveney, and Carola-Bibiane Schönlieb. Diffusion models encode the intrinsic dimension of data manifolds. In *International Conference on Machine Learning*, 2024.

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.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011. doi: 10.1162/NECO_a_00142.

Zixuan Zhang, Kaixuan Huang, Tuo Zhao, Mengdi Wang, and Minshuo Chen. Diffusion model for manifold data: Score decomposition, curvature, and statistical complexity. arXiv:2603.20645, 2026.

Ziqi Zhao and Qingjian Ni. Recovering weighted tangent geometry from a single-scale score field. arXiv:2608.22334, 2026.

A PROOF OF THE TWO-TERM EXPANSION

The local argument below tracks every power of σ. Constants may depend on the compact query set, the branch count, the jet bounds, r_0, Δ, and the far mass, but not on σ.

A.1 A GAUSSIAN ENVELOPE

Lemma A.1 (Polynomial Gaussian envelope)**.** *Fix a compact $\mathcal{Q} \subset \mathbb{R}^D$. There are $c > 0$ and constants $C_{m,\mathcal{Q}}$ such that the following holds. Let $v \in \mathbb{S}^{D-1}$, $u \geq 0$, and let $q_t = uv + t\xi$ for $0 \leq t \leq 1$, with $\|\xi\| \leq u/2$. For every fixed mixed-derivative order m of $K_z(q) = e^{-\|z-q\|^2/2}$ in (z, q),*

$$|\partial^m K_z(q_t)| \leq C_{m,\mathcal{Q}}(1+u)^m e^{-cu^2}, \qquad z \in \mathcal{Q}. \tag{35}$$

The same form holds for vector and matrix derivatives.

Proof. Let $R = \sup_{z\in\mathcal{Q}} \|z\|$. Since $u/2 \leq \|q_t\| \leq 3u/2$, for $u \geq 4R$ we have $\|z - q_t\| \geq u/4$. For bounded u, every polynomial factor is absorbed into the constant. Derivatives of a Gaussian equal a polynomial in $z - q_t$ times the same Gaussian, giving the claim. □

A.2 EXPANSION OF THE RESCALED CONVOLUTION

After changing variables $r = \sigma u$, remove the common factor $(2\pi)^{-D/2}\sigma^{1-D}$ and write

$$\widetilde{Q}_\sigma(z) = \sum_{j=1}^{s} \int_0^{r_0/\sigma} \rho_j(\sigma u) K_z(q_{j,\sigma}(u))\,\mathrm{d}u + R_{\mathrm{far},\sigma}(z), \tag{36}$$

where $q_{j,\sigma}(u) = (\gamma_j(\sigma u) - x_0)/\sigma$.

By Equation (7),

$$q_{j,\sigma}(u) = uv_j + \delta_{j,\sigma}(u), \qquad \delta_{j,\sigma}(u) = \frac{\sigma}{2}u^2 k_j + \epsilon_{j,\sigma}(u), \quad \|\epsilon_{j,\sigma}(u)\| \leq L_\gamma \sigma^{1+\alpha} u^{2+\alpha}. \tag{37}$$

Decrease r_0 if needed so that $\|\delta_{j,\sigma}(u)\| \leq u/2$ for $0 \leq u \leq r_0/\sigma$. Taylor's theorem and Lemma A.1 give, uniformly in $z \in \mathcal{Q}$,

$$K_z(q_{j,\sigma}(u)) = K_z(uv_j) + \nabla_q K_z(uv_j)^\mathsf{T} \delta_{j,\sigma}(u) + R_{j,\sigma}(z,u), \tag{38}$$

$$|R_{j,\sigma}(z,u)| + \|\nabla_z R_{j,\sigma}(z,u)\| \leq C\left[\sigma^{1+\alpha}P_1(u) + \sigma^2 P_2(u)\right] e^{-cu^2} \leq C\sigma^{1+\alpha}P(u)e^{-cu^2}. \tag{39}$$

The last inequality uses $0 < \alpha \leq 1$ and $\sigma \leq 1$. Since $\nabla_q K_z(uv_j) = (z - uv_j)K_z(uv_j)$ and $k_j \perp v_j$,

$$\nabla_q K_z(uv_j)^\mathsf{T} \frac{\sigma}{2}u^2 k_j = \frac{\sigma}{2}u^2 \langle k_j, z\rangle K_z(uv_j). \tag{40}$$

Likewise, Equation (8) gives

$$\rho_j(\sigma u) = w_j(1 + \sigma a_j u) + r_{j,\sigma}(u), \qquad |r_{j,\sigma}(u)| \leq L_\rho \sigma^{1+\alpha} u^{1+\alpha}. \tag{41}$$

Multiplying Equations (38) and (41), collecting the order-zero and order-σ terms, and applying the envelope to cross terms yields

$$\rho_j(\sigma u)K_z(q_{j,\sigma}(u)) = w_j K_z(uv_j) + \sigma w_j \left[a_j u + \frac{1}{2}u^2\langle k_j, z\rangle\right] K_z(uv_j) + \mathcal{R}_{j,\sigma}(z,u), \tag{42}$$

with

$$|\mathcal{R}_{j,\sigma}(z,u)| + \|\nabla_z \mathcal{R}_{j,\sigma}(z,u)\| \leq C\sigma^{1+\alpha}P(u)e^{-cu^2}. \tag{43}$$

The right-hand side is integrable on $[0,\infty)$. Extending the tangent integrals from r_0/σ to infinity contributes a Gaussian tail smaller than every power of σ.

For the far measure, $\|x - x_0\| \geq \Delta$. On a compact normalized query set and for small σ, $\|(x - x_0)/\sigma - z\| \geq \Delta/(2\sigma)$. Hence the rescaled far denominator and its z-gradient are bounded by a polynomial in σ^{-1} times $e^{-\Delta^2/(8\sigma^2)}$, again smaller than $\sigma^{1+\alpha}$. Integrating Equation (42) proves Equation (15).

Finally, Q_0 is strictly positive and has a positive minimum on $\mathcal{Q}$. Write $\widetilde{Q}_\sigma = Q_0[1 + \sigma Q_1/Q_0 + R_\sigma/Q_0]$. For every fixed m, the preceding envelope applies to all z-derivatives through order m. The map $x \mapsto \log x$ is smooth on a compact interval bounded away from zero, so composition in C^m gives

$$\log \widetilde{Q}_\sigma = \log Q_0 + \sigma Q_1/Q_0 + O_{C^m}(\sigma^{1+\alpha}) \tag{44}$$

which proves Equation (16).

B Distributional proof of jet identifiability

Let $K(z) = e^{-\|z\|^2/2}$. The tangent transform is $Q_0 = K * \nu_0$. Associate to the jet parameters the tempered distribution

$$\langle \eta, \phi \rangle = \sum_{j=1}^{s} w_j \int_0^\infty \left[a_j u \phi(uv_j) + \frac{1}{2} u^2 \langle k_j, \nabla \phi(uv_j) \rangle \right] \mathrm{d}u. \tag{45}$$

Direct evaluation on the translated kernel $y \mapsto K(z-y)$ gives

$$\begin{aligned}(K * \eta)(z) &= \sum_j w_j \int_0^\infty \left[a_j u K(z - uv_j) + \frac{1}{2} u^2 \left\langle k_j, \nabla_y K(z-y)|_{y=uv_j} \right\rangle \right] \mathrm{d}u \\ &= \sum_j w_j \int_0^\infty \left[a_j u + \frac{1}{2} u^2 \langle k_j, z - uv_j \rangle \right] K(z - uv_j)\, \mathrm{d}u = Q_1(z), \end{aligned} \tag{46}$$

where $\langle k_j, v_j \rangle = 0$.

Suppose $G = 0$. Since $Q_0 > 0$ and $\mathbb{R}^D$ is connected, $\nabla(Q_1/Q_0) = 0$ implies $Q_1 = CQ_0$. Thus $K * (\eta - C\nu_0) = 0$. Taking Fourier transforms in the space of tempered distributions gives

$$\widehat{K}(\xi)\big(\widehat{\eta} - C\widehat{\nu_0}\big) = 0. \tag{47}$$

The Gaussian multiplier $\widehat{K}(\xi)$ is everywhere positive, so $\eta = C\nu_0$.

Fix branch j and an interval $I = (u_-, u_+) \Subset (0, \infty)$ such that the compact ray segment $\{uv_j : u \in \overline{I}\}$ has a tubular neighborhood disjoint from the other rays and the origin. In local coordinates $y = uv_j + n$, $n \in v_j^\perp$, choose $\phi(y) = \chi(u)\psi(n)$ with $\psi(0) = 0$ and arbitrary $\nabla\psi(0)$. Then $\nu_0(\phi) = 0$, and all branches except j vanish by support. Equality $\eta = C\nu_0$ gives

$$\frac{w_j}{2} \int_I u^2 \chi(u) \langle k_j, \nabla\psi(0) \rangle \, \mathrm{d}u = 0. \tag{48}$$

Arbitrary χ and $\nabla\psi(0)$ imply $k_j = 0$.

Now take $\psi(0) = 1$ and $\nabla\psi(0) = 0$. Then

$$w_j \int_I (a_j u - C)\chi(u)\, \mathrm{d}u = 0 \tag{49}$$

for every $\chi \in C_c^\infty(I)$. Hence $a_j u = C$ on I. Since I contains more than one point, $a_j = 0$ and $C = 0$. Repeating for every branch proves the nullspace is trivial. Applying the result to the difference of two parameter collections proves Theorem 4.1.

C Finite evaluations and the query lower bound

Lemma C.1 (Evaluation basis)**.** *Let $f_1, \dots, f_p : X \to \mathbb{R}^D$ be linearly independent as functions. Then there exist pairs $(x_m, r_m) \in X \times \{1, \dots, D\}$, $m = 1, \dots, p$, such that the matrix $[e_{r_m}^\mathsf{T} f_\ell(x_m)]_{m,\ell=1}^p$ is nonsingular.*

Proof. Let $V = \mathrm{span}\{f_1, \dots, f_p\}$. Point-coordinate evaluations $L_{x,r}(f) = e_r^\mathsf{T} f(x)$ separate points of V: if every such functional vanishes on $f \in V$, then f is the zero function. Therefore their restrictions span V^*. Select p of them forming a basis of V^*; their matrix in the basis $f_1, \dots, f_p$ is nonsingular. □

By Theorem 4.1, the $p = sD$ fields in Equations (20) and (21) are linearly independent, so Lemma C.1 proves the upper bound in Theorem 4.2. For the lower bound, restrict the parameter vector to a nonempty open box in $\mathbb{R}^{sD}$. Any fixed continuous scheme with M scalar outputs induces a continuous injective map from that box to $\mathbb{R}^M$. If $M < sD$, compose with the standard coordinate embedding $\mathbb{R}^M \hookrightarrow \mathbb{R}^{sD}$. Invariance of domain (Munkres, 2000) would make the image open in $\mathbb{R}^{sD}$, although it lies in a lower-dimensional coordinate subspace, a contradiction.

D Center Calibration and Center-Robust Identifiability

D.1 Why weak calibration has a physical $O(\sigma^2)$ error

Let $c_\sigma^{\mathrm{ref}} = x_0 - \sigma b_0$ with bounded b_0, and use normalized coordinates $y = (x - c_\sigma^{\mathrm{ref}})/\sigma$. The normalized finite-noise score can be written

$$u_\sigma(y) = F_0(y - b_0) + \sigma G(y - b_0) + O_{C^1}(\sigma^{1+\alpha}) \tag{50}$$

on every compact set containing the test supports. For a smooth localized test ψ, the corresponding weak-calibration row and response are

$$a_\psi(u) = \left(\int \psi(y) u(y)\, \mathrm{d}y,\ \int \psi(y)\, \mathrm{d}y \right), \tag{51}$$

$$r_\psi(u) = \int \left[-\nabla\psi(y)^\top u(y) + \psi(y)\{\|u(y)\|^2 + y^\top u(y) + D\} \right] \mathrm{d}y. \tag{52}$$

Both maps are smooth polynomial functionals of u on the compact test supports. Substitution of Equation (50) therefore gives

$$A_\sigma = A_0 + \sigma A_1 + O(\sigma^{1+\alpha}), \qquad r_\sigma = r_0 + \sigma r_1 + O(\sigma^{1+\alpha}). \tag{53}$$

The exact tangent identity gives $r_0 = A_0\beta_0$ with $\beta_0 = (b_0, 1)$ because each branch is one-dimensional. If A_0 has full column rank, the least-squares solution is

$$\widehat{\beta}_\sigma = (A_\sigma^\top A_\sigma)^{-1} A_\sigma^\top r_\sigma. \tag{54}$$

The full-rank least-squares map is smooth in a neighborhood of (A_0, r_0). Differentiating its normal equations, or expanding the pseudoinverse, yields

$$\widehat{\beta}_\sigma = \beta_0 + \sigma (A_0^\top A_0)^{-1} A_0^\top (r_1 - A_1\beta_0) + O(\sigma^{1+\alpha}), \tag{55}$$

where the apparent derivatives of $A_\sigma^\top A_\sigma$ cancel because $r_0 = A_0\beta_0$. If b_1 denotes the first D coordinates of the order-σ coefficient, the physical center returned by the calibration is

$$c_\sigma^{\mathrm{ref}} + \sigma \widehat{b}_\sigma = x_0 + \sigma^2 b_1 + O(\sigma^{2+\alpha}). \tag{56}$$

This proves Proposition 5.1. It also makes the role of the local-window assumption explicit: a separate coarse procedure must place the reference within $O(\sigma)$ of x_0.

D.2 Two-scale center expansion

To prove Theorem 5.2, put $h_\sigma = (c_\sigma - x_0)/\sigma = \sigma b + O(\sigma^{1+\alpha})$ and $h_{\lambda\sigma} = (c_\sigma - x_0)/(\lambda\sigma) = (\sigma/\lambda) b + O(\sigma^{1+\alpha})$. Uniform C^1 regularity of Equation (16) and a Taylor expansion give

$$\sigma s_\sigma(c_\sigma + \sigma z) = F_\sigma(z + h_\sigma) = F_0(z) + \sigma \nabla F_0(z) b + \sigma G(z) + O(\sigma^{1+\alpha}), \tag{57}$$

$$\lambda\sigma s_{\lambda\sigma}(c_\sigma + \lambda\sigma z) = F_{\lambda\sigma}(z + h_{\lambda\sigma}) = F_0(z) + \frac{\sigma}{\lambda} \nabla F_0(z) b + \lambda\sigma G(z) + O(\sigma^{1+\alpha}). \tag{58}$$

Terms quadratic in h and products of h with the order-σ correction are $O(\sigma^2)$, which is contained in $O(\sigma^{1+\alpha})$ for $\alpha \le 1$. Subtraction proves Equation (27).

Definition D.1 (Translation lineality)**.** The tangent measure ν_0 has translation lineality if there exists $b \neq 0$ such that $(T_{tb})_\# \nu_0 = \nu_0$ for every $t \in \mathbb{R}$. For a finite positive ray junction, this can occur only along a complete line represented by opposite rays with matching density.

For injectivity, suppose the augmented field in Equation (29) is zero. Since $\nabla F_0 b = \nabla(b^\mathsf{T}\nabla \log Q_0)$,

$$\nabla\left(\frac{Q_1 - \lambda^{-1} b^\mathsf{T}\nabla Q_0}{Q_0}\right) = 0. \tag{59}$$

Hence $Q_1 - \lambda^{-1} b^\mathsf{T}\nabla Q_0 = CQ_0$. Distributional differentiation commutes with convolution, so Gaussian injectivity yields

$$\eta - \lambda^{-1} b \cdot \nabla \nu_0 = C\nu_0. \tag{60}$$

Use the same tube around an interior segment of ray j. For a test function whose value is zero on the ray and whose normal gradient $g(u) \in v_j^\perp$ is arbitrary, Equation (60) becomes

$$w_j \int_I \left\langle \tfrac{1}{2} u^2 k_j + \lambda^{-1} P_{v_j^\perp} b, g(u) \right\rangle \mathrm{d}u = 0. \tag{61}$$

Therefore $\frac{1}{2}u^2 k_j + \lambda^{-1} P_{v_j^\perp} b = 0$ for all $u \in I$, which implies

$$k_j = 0, \qquad P_{v_j^\perp} b = 0 \tag{62}$$

for every j. Thus b is parallel to every ray direction.

With the normal part removed, $b \cdot \nabla \nu_0$ is supported at the common endpoint. Indeed, for a smooth compactly supported ϕ,

$$\langle b \cdot \nabla \nu_0, \phi \rangle = -\sum_j w_j \int_0^\infty (b^\mathsf{T} v_j) \frac{\mathrm{d}}{\mathrm{d}u} \phi(u v_j)\, \mathrm{d}u = \left(\sum_j w_j b^\mathsf{T} v_j\right) \phi(0). \tag{63}$$

Neither $C\nu_0$ nor the remaining density-slope distribution has an atom at zero, so the coefficient must vanish. Under Equation (62), this cancellation is equivalent to b lying in a translation-lineality direction of the finite ray measure. The no-lineality assumption gives $b = 0$. The argument in Section B then gives $a_j = C = 0$. Linear independence and the query count follow exactly as in Lemma C.1 and the invariance-of-domain lower bound.

E Perturbation Composition and Finite-Sample Score Error

E.1 Deterministic least-squares composition

Let $\theta \in \mathbb{R}^p$ be the branch-jet vector, with the center coefficient appended in the augmented model. After selecting M scalar point-coordinate evaluations, write

$$y = B\theta + r_{\mathrm{fs}}, \qquad \|r_{\mathrm{fs}}\| \le C_{\mathrm{bias}} \sigma^\alpha, \tag{64}$$

where B is the limiting design and r_{fs} is the finite-scale remainder from Corollary 3.2 and Theorem 5.2. The implemented design and response are $\widehat{B} = B + E$ and $\widehat{y} = y + e$. Whenever $\widehat{B}$ has full column rank,

$$\widehat{\theta} - \theta = \widehat{B}^\dagger(\widehat{y} - \widehat{B}\theta) = \widehat{B}^\dagger(r_{\mathrm{fs}} + e - E\theta), \tag{65}$$

$$\left\|\widehat{\theta} - \theta\right\| \le \left\|\widehat{B}^\dagger\right\|_{\mathrm{op}} \left(C_{\mathrm{bias}}\sigma^\alpha + \|e\| + \|E\|_{\mathrm{op}} \|\theta\|\right). \tag{66}$$

Weyl's inequality gives $\sigma_{\min}(\widehat{B}) \ge \sigma_{\min}(B) - \|E\|_{\mathrm{op}}$, which proves Equation (31).

Suppose the normalized score vectors at the selected queries have stacked errors ε_σ and $\varepsilon_{\lambda\sigma}$. Scale differencing contributes at most

$$\frac{\varepsilon_\sigma + \varepsilon_{\lambda\sigma}}{(\lambda - 1)\sigma}. \tag{67}$$

If the provisional center has an additional physical remainder r_σ, its normalized displacements are r_σ/σ and $r_\sigma/(\lambda\sigma)$. A mean-value bound for the two normalized score fields, followed by division by $(\lambda - 1)\sigma$, gives

$$C_{\mathrm{ctr}} \frac{\|r_\sigma\|}{\sigma^2}. \tag{68}$$

This proves Equation (32).

Finally, for a fixed finite candidate set the map from distinct directions and positive normalized weights to every design entry is smooth: the ray-Gaussian integrals and their derivatives are dominated by polynomial Gaussian envelopes, and Q_0 is uniformly positive. Hence on a compact first-order class with angular separation $\delta_0 > 0$ and minimum weight $w_0 > 0$,

$$\|E\|_{\mathrm{op}} \le L_B \left(\max_j \|\widehat{v}_j - v_j\| + \|\widehat{w} - w\| \right) \tag{69}$$

after the branch correspondence is fixed. This is the explicit "first-stage design error" term used in the end-to-end experiment.

E.2 A SIMULTANEOUS EMPIRICAL KDE SCORE BOUND

At a physical query x, define

$$W_i = e^{-\|x - X_i\|^2/(2\sigma^2)}, \qquad V_i = \frac{X_i - x}{\sigma} W_i, \tag{70}$$

with means $a = \mathbb{E}W_i$ and $b = \mathbb{E}V_i$. The normalized population and empirical KDE scores are $F = b/a$ and $\widehat{F} = \overline{V}/\overline{W}$. The elementary envelopes

$$0 \le W_i \le 1, \qquad \|V_i\| \le e^{-1/2}, \qquad \mathbb{E}\,\|V_i\|^2 \le \frac{2}{e} a \tag{71}$$

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

Let $t = \log(4M/\delta)$. Scalar Bernstein for W_i and a Hilbert-space Bernstein inequality for V_i (Pinelis, 1994), followed by a union bound over M queries, imply with probability at least $1 - \delta$ that simultaneously

$$\left|\overline{W} - a\right| \le D(a) := \sqrt{\frac{2at}{N}} + \frac{t}{3N}, \tag{72}$$

$$\left\|\overline{V} - b\right\| \le Q(a) := \sqrt{\frac{4at}{eN}} + \frac{4e^{-1/2}t}{3N}. \tag{73}$$

Whenever $D(a) < a$, direct ratio subtraction gives

$$\left\|\widehat{F} - F\right\| \le \frac{Q(a) + \|F\|\, D(a)}{a - D(a)}. \tag{74}$$

On a declared compact normalized query set for a one-dimensional branch junction, positivity and the local density lower bound give $a \ge a_- \sigma$ for all sufficiently small σ, while $\|F\|$ is bounded. If $\sqrt{2t/(Na_-\sigma)} + t/(3Na_-\sigma) \le 1/2$, Equation (74) simplifies to

$$\max_{m \le M} \left\|\widehat{F}_\sigma(z_m) - F_\sigma(z_m)\right\| \le C \left[\sqrt{\frac{t}{N\sigma}} + \frac{t}{N\sigma} \right]. \tag{75}$$

The same bound holds at $\lambda\sigma$ with constants depending on fixed λ.

Substituting Equation (75) into Equation (67), absorbing the fixed finite design and λ into the constant, and adding Equations (64) and (69) yields

$$\left\|\widehat{\theta} - \theta\right\| \le C \left[\sigma^\alpha + \sqrt{\frac{\log(M/\delta)}{N\sigma^3}} + \frac{\log(M/\delta)}{N\sigma^2} + \varepsilon_{\mathrm{first}} \right]. \tag{76}$$

Here $\varepsilon_{\mathrm{first}}$ collects the center remainder and the tangent direction/weight error after multiplication by their deterministic Lipschitz constants. Under sample splitting, or under an independently validated first stage whose contribution is no larger than the displayed stochastic terms, balancing σ^α and $(N\sigma^3)^{-1/2}$ gives Equation (34). The term $N^{-1}\sigma^{-2}$ is lower order at that bandwidth for every $\alpha > 0$.

E.3 Learned-score error units

If a learned physical score has pointwise error $\|s_\theta(x,\tau) - s_\tau(x)\| \le \epsilon_{\mathrm{net}}(\tau)$, then its normalized score error is $\tau\epsilon_{\mathrm{net}}(\tau)$. Consequently the two-scale response error is bounded by

$$\frac{\epsilon_{\mathrm{net}}(\sigma) + \lambda\epsilon_{\mathrm{net}}(\lambda\sigma)}{\lambda - 1}. \tag{77}$$

The bound shows why a small average score loss does not certify the cross-scale derivative: the errors at the two scales must also be aligned, not simply small in isolation.

F Experimental Design and Reproduction Details

F.1 Experiment-to-claim map

Question	Script and primary output	Interpretation
Does the finite-scale remainder have order σ^α?	`run_population.py`; parameter error and refined-center error over decreasing σ	Numerical check of Theorems 3.1 and 5.2 alongside the analytic proof.
Can the sharp query theorem be implemented stably?	`run_query_design.py`; $\sigma_{\min}$, condition number, error under response noise	Compares D-optimal, pivoted-QR, random, and uniform scalar rows.
Does sampling create the predicted bias–variance tradeoff?	`run_kde.py`; score error, two-scale response error, jet error, effective sample size	Tests the conditional trend in Corollary 6.2 on the retained twenty-seed grid.
Does the first-order frontend compose with the jet inverse?	`run_end_to_end.py`; branch count, tangent error, jet/center proxy error	Separates known-count, blind-count, and exact-first-order interfaces.
Does score-estimated geometry compose beyond the plane?	`run_general_end_to_end.py`; center, tangent, weight, and jet errors in $D = 3$–5	Tests the known-count score-to-jet pipeline without supplied directions or weights.
Does ordinary score training preserve the two-scale derivative?	`run_learned_score.py`; score RMS, scale-difference RMS, jet error	Isolates cross-scale consistency on controlled synthetic junctions.
How do angle and minimum weight control stability?	`run_stability_phase.py`; singular values, condition number, relative error	Separates exact identifiability from uniform stability over 42 cells.
Does the inverse remain tractable in larger systems?	`run_scalability.py`; rank, error, timing through $D = 20$ and 16 branches	Tests computation and sensitivity to first-order perturbations.
Why match two scales instead of subtracting a tangent estimate?	`run_response_baselines.py`; paired response and parameter errors	Compares the proposed response with naive subtraction and an exact-F_0 diagnostic.
Which E4 interface produces the observed error?	`run_end_to_end_diagnostics.py`; crossed basis and query-row policies	Separates first-order basis error, row choice, finite-scale remainder, and KDE noise.
Does a larger learned score solve the inverse mismatch?	`run_learned_score.py`; six hard-case jobs through 50k updates	Width-256, six-block follow-up; checkpoint conclusions remain post hoc.

Table 2: Every experiment targets a distinct proof interface.

F.2 Geometries and exact population evaluation

Each synthetic branch is a constant-curvature arc

$$\gamma_j(r) - x_0 = \frac{\sin(\kappa_j r)}{\kappa_j} v_j + \frac{1 - \cos(\kappa_j r)}{\kappa_j} n_j, \tag{78}$$

with the continuous straight-ray limit at $\kappa_j = 0$. This curve is exactly arc-length parameterized, satisfies $\gamma_j''(0) = \kappa_j n_j$, and has nonzero higher-order terms. The density is

$$\rho_j(r) = w_j \exp(a_j r + d_j r^2), \tag{79}$$

where the quadratic coefficient keeps the finite-radius sampling distribution well behaved without changing the target slope a_j. The normalized population score is evaluated as a Gaussian-weighted first-moment/zeroth-moment ratio using fixed Gauss–Legendre quadrature. It is never obtained by numerical differentiation.

The analytic design uses the posterior moment identities

$$F_0(z) = \mathbb{E}_{\Pi_z}[U] - z, \qquad \nabla F_0(z) = \mathrm{Cov}_{\Pi_z}(U) - I_D, \tag{80}$$

and corresponding differentiated ray moments for the density and curvature columns. The implemented design uses this column order and sign convention, the center coefficient $-1/\lambda$, exact planar tangent reconstruction, and generic-rank constructions in $D \in \{2, 3, 5\}$.

F.3 Population and finite-query matrices

The population matrix crosses ambient dimension, branch count, random seed, scale ratio, true versus $O(\sigma^2)$-biased center, and decreasing base scale. Directions are generated with a minimum angular separation, weights are positive and normalized, and curvature/density-slope coordinates are independently sampled inside declared bounds. The fitted exponent uses only the smallest configured scales.

For query design, a candidate normalized grid is built independently of the unknown second-order coefficients. Every point contributes D possible scalar component rows. We compare:

1. *Greedy D-optimal*: sequentially maximize $b^\top A^{-1} b$ with a small ridge;
2. *Pivoted QR*: select a rank-revealing minimal basis, then add high-norm rows;
3. *Random*: uniform sampling without replacement;
4. *Uniform*: equally spaced row indices, retained as a deliberately geometry-agnostic baseline.

The response perturbation has a configured RMS and is regenerated from a recorded seed. The code records scalar-query count, number of unique physical locations, rank, $\sigma_{\min}$, condition number, and relative parameter error.

F.4 Stability phase, scale-up, and response baselines

The stability phase experiment uses three planar rays with configured minimum angles $\{3, 5, 8, 12, 20, 30, 45\}^\circ$ and minimum weights $\{0.01, 0.02, 0.04, 0.08, 0.12, 0.20\}$. Each of the 42 cells has 20 seeds, response-noise RMS 10^{-3}, and twice as many scalar queries as unknowns. We retain the entire grid, including the high-error near-collision cells. The full grid characterizes the transition across angle and weight without imposing a universal separation threshold.

The scale-up crosses $(D, s) \in \{(2, 4), (3, 8), (5, 8), (10, 16), (20, 16)\}$, ten seeds, and four first-order perturbation levels. Direction/weight RMS pairs are $(0, 0)$, $(0.01, 0.005)$, $(0.03, 0.015)$, and $(0.06, 0.03)$. Every case uses $\sigma = 0.04$, $\lambda = 2$, and twice the parameter count in scalar queries. Timing separates population-response evaluation, row design, and least-squares solution.

The response baseline uses four planar geometries, ten seeds, four base scales, population and KDE scores, and the same four perturbation levels. Three methods share fixed exact-geometry query rows: the matched two-scale response, naive subtraction of a tangent field, and an exact-F_0 diagnostic. The exact-F_0 diagnostic is an idealized reference that isolates the cost of estimating the nuisance field. Method ratios are paired before aggregation.

F.5 KDE and estimated frontends

A branch is selected according to its finite-radius mass and its radial coordinate is sampled by an inverse CDF. Empirical Gaussian KDE scores are evaluated exactly from the stored samples, with a per-query log-weight shift for numerical stability. The effective sample size

$$\mathrm{ESS}(x) = \frac{(\sum_i W_i(x))^2}{\sum_i W_i(x)^2} \tag{81}$$

is recorded at both scales. One sampled dataset is reused across the bandwidth sweep for a paired comparison.

The planar first-stage frontend queries a tangential score on an equally spaced shell, spectrally integrates the circular log density, divides by the positive Gaussian–ray Fourier multipliers, infers Toeplitz rank, recovers directions through a matrix pencil, and fits nonnegative weights. The end-to-end experiment reports known-count and blind-count variants separately. The second-order error is evaluated only when the recovered count matches, and a separate exact-first-order row isolates the second stage.

The nonplanar frontend treats the branch count as known and fits the leading tangent-ray score directly. Its variables are a normalized center offset, hyperspherical branch directions, and softmax-normalized weights. Multistart nonlinear least squares uses shell and interior score queries; the fitted center and first-order geometry then determine the second-stage query rows. Ground-truth directions and weights enter only the final permutation alignment and error calculation. The grid crosses $D \in \{3, 4, 5\}$, branch counts $\{3, 4, 6\}$, five seeds, population and KDE scores, and base scales $\{0.080, 0.055, 0.038\}$. KDE rows use 131,072 curve samples. Every case also includes an exact-first-order diagnostic evaluated at the same fitted center.

The crossed E4 diagnostic uses the same four geometries, five seeds, four scales, known/blind counts, population/KDE scores, and 32,768 samples. For every successful first stage, it independently crosses an estimated or exact first-order basis with rows selected from estimated or exact geometry. This separates coefficient-basis error from row-selection error. A failed blind-count case produces one explicit sentinel instead of four inverse rows; count accuracy includes that sentinel, whereas error summaries contain only successful inversions. Median relative error changes by a factor of 1 when the row policy is switched. In population scores, the estimated-basis/exact-basis ratio 0.611 reflects finite-scale bias cancellation: the estimated tangent basis can absorb part of the remainder, so the comparison diagnoses bias absorption rather than geometric accuracy.

F.6 Learned-score protocol and hardware scaling

The network predicts the normalized score σs_σ from a normalized noisy location and a bounded-frequency sinusoidal encoding of $\log \sigma$. Clean samples are restricted to radius 1.25 around the junction, and the noise scale is drawn log-uniformly across the configured interval. Training samples

$$\widetilde{X} = X/\sigma + Z, \qquad Z \sim \mathcal{N}(0, I_D), \tag{82}$$

then minimizes $\frac{1}{2}\mathbb{E}\|f_\theta(\widetilde{X}, \sigma) + Z\|^2$. The residual MLP has width 128, four blocks, eight sinusoidal frequencies up to 8, and 135042 trainable parameters. Each job starts from 32,768 physical samples, of which 7544–8110 fall within the radius cutoff. AdamW uses batch size 1,024, initial learning rate 1.5×10^{-3}, weight decay 10^{-6}, and gradient-norm clipping at 10. A cosine schedule decays to 5% of the initial rate over 20,000 updates. Evaluation uses an exponential moving average with decay 0.999 after a 1,000-update warmup at checkpoints $\{1000, 2500, 5000, 7500, 10000, 15000, 20000\}$.

All geometries, initializations, sample draws, minibatches, and noise draws use recorded integer seeds. The base configuration has one independent job per geometry–seed pair. The launcher assigns the 20 jobs round-robin across two NVIDIA GeForce RTX 3090 GPUs; each process sees one device as `cuda:0`. Median job time is 241 seconds, aggregate GPU time is 1.33 hours, and dual-GPU wall time is 2.47×10^3 seconds.

The scale-up retains the two harder four-ray geometries and three seeds, increases the model to width 256 with 6 residual blocks, uses 65,536 physical samples, and trains through 50000 updates. The seven checkpoints are $\{1, 5, 10, 20, 30, 40, 50\}$k. The 6 jobs were run concurrently on one RTX 3090; median per-job wall time was 736 seconds. Independent population and empirical score evaluations diagnose the same controlled synthetic distribution. Jet-error checkpoint selection is reported post hoc; the nonmonotone parameter trajectory is the mechanism under study.

F.7 Support Material

The anonymous package contains `configs/paper.yaml`, ten experiment entry points, and the shared numerical routines they import. The entry points cover the eleven studies above because `run_learned_score.py` accepts both the base and scale-up configurations. Each script writes

CSV output to an ignored `results/` directory. Exact commands and the mapping from scripts to experiments appear in `README.md`.

F.8 Generated tables and additional figures

The following tables report aggregates from the experiment CSVs and are included here for completeness.

D	center	median slope	IQR	median R^2	instances
2	biased	1.005	[0.993, 1.021]	1.0000	30
2	true	0.992	[0.978, 1.003]	1.0000	30
3	biased	1.003	[0.989, 1.009]	1.0000	30
3	true	0.999	[0.992, 1.010]	1.0000	30
5	biased	0.996	[0.988, 1.001]	1.0000	30
5	true	0.999	[0.992, 1.002]	1.0000	30

Table 3: Population finite-scale fits from the active result profile.

design	median $\sigma_{\min}$	median cond.	median relative error	95th pct.
D-optimal	0.468	23.2	0.00552	0.011
pivoted QR	0.321	42.7	0.00754	0.0151
random	0.138	33.6	0.016	0.0629
uniform	0.109	46.6	0.0195	0.0714

Table 4: Query-design summary at approximately twice the parameter dimension.

geometry	error slope	seed-bootstrap 95% CI	theory	bandwidth slope
clustered four-ray	-0.189	[-0.248, -0.134]	-0.200	-0.184
regular Y	-0.256	[-0.327, -0.187]	-0.200	-0.163
weak four-ray	-0.231	[-0.283, -0.196]	-0.200	-0.200
weak Y	-0.201	[-0.266, -0.126]	-0.200	-0.159

Table 5: Best bandwidth and error slopes on the retained twenty-seed grid.

field	tangent input	count rate	median jet error	95th pct.	median angle
kde	blind count	98.8%	1.77	5.78	0.0216
kde	known count	100.0%	1.79	5.76	0.0214
kde	exact first-order	100.0%	1.8	5.77	0
population	blind count	100.0%	0.0396	0.0892	0.0177
population	known count	100.0%	0.0396	0.0892	0.0177
population	exact first-order	100.0%	0.0672	0.15	0

Table 6: Composition of the planar tangent frontend and the second-order inverse.

geometry	seeds	score RMS	scale-diff. RMS	jet error	KDE jet error
clustered four-ray	5	0.0149	0.231	0.972	0.861
regular Y	5	0.0164	0.215	0.34	0.277
weak four-ray	5	0.0169	0.238	0.746	0.599
weak Y	5	0.0203	0.303	0.761	0.677

Table 7: Final-checkpoint learned-score diagnostics.

Regime	Angle	Min. weight	Median $\sigma_{\min}$	Median cond.	Median rel. error
near collision	3	0.01	0.0018	5.61e+04	0.698
well separated	45	0.2	0.553	24.6	0.00379

Table 8: Opposite corners of the angle–minimum-weight stability phase map.

Score source	Perturbation	Median naive/matched	IQR	n
population	exact	0.334	[0.332, 0.337]	160
population	mild	7.75	[3.5, 15.7]	160
population	moderate	26.6	[10.2, 52.4]	160
population	strong	49.4	[21.4, 115]	160
KDE	exact	0.863	[0.652, 1.24]	160
KDE	mild	0.893	[0.676, 1.32]	160
KDE	moderate	1.21	[0.916, 1.78]	160
KDE	strong	1.96	[1.32, 3.09]	160

Table 9: Paired naive-subtraction error divided by matched-response error.

Case	Perturbation	Median rel. error	Median cond.	Response time (s)
D2/s4	exact	0.205	40.2	0.011
D2/s4	mild	0.215	40.1	0.011
D2/s4	moderate	0.259	38.2	0.011
D2/s4	strong	0.609	38.5	0.011
D3/s8	exact	0.524	169	0.022
D3/s8	mild	0.52	143	0.022
D3/s8	moderate	0.831	174	0.022
D3/s8	strong	0.989	307	0.022
D5/s8	exact	0.162	44	0.030
D5/s8	mild	0.163	49.7	0.030
D5/s8	moderate	0.311	60.3	0.030
D5/s8	strong	0.556	117	0.030
D10/s16	exact	0.124	99.8	0.096
D10/s16	mild	0.161	104	0.096
D10/s16	moderate	0.485	175	0.096
D10/s16	strong	79.2	8.55e+03	0.096
D20/s16	exact	0.0825	128	0.121
D20/s16	mild	0.131	137	0.121
D20/s16	moderate	0.408	162	0.121
D20/s16	strong	85.6	2.43e+04	0.121

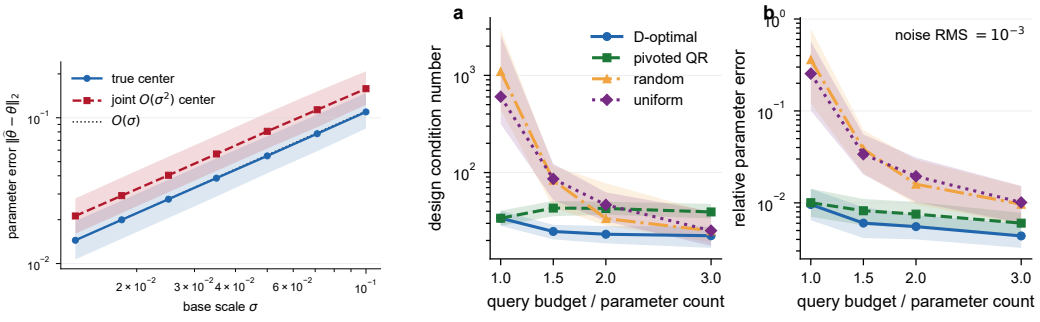
Table 10: Scale-up errors and timings. Every displayed cell has ten seeds and full column rank.

Source	Basis	Rows	Projected error	Relative proxy error	Remainder	Cond.
population	estimated	estimated	0.039	0.0781	0.0219	20.7
population	estimated	exact	0.0376	0.0741	0.0219	20.3
population	exact	estimated	0.0785	0.164	0.0387	20.8
population	exact	exact	0.0675	0.131	0.0375	20.4
kde	estimated	estimated	1.68	3.29	0.852	19.8
kde	estimated	exact	1.67	3.14	0.844	19.7
kde	exact	estimated	1.63	3.3	0.851	19.5
kde	exact	exact	1.66	3.44	0.833	19.5

Table 11: Crossed E4 diagnostic on successful inverse rows. “Basis” and “Rows” vary independently.

Step	Score RMS	Response RMS	Parameter error	Parameter-error IQR
1k	0.0872	0.675	2.11	[1.81, 2.9]
5k	0.0345	0.184	0.725	[0.705, 0.768]
10k	0.0215	0.201	0.768	[0.674, 1.03]
20k	0.0187	0.222	0.739	[0.686, 0.977]
30k	0.0148	0.217	0.65	[0.452, 0.931]
40k	0.0145	0.192	0.511	[0.473, 0.648]
50k	0.0143	0.21	0.676	[0.665, 0.708]

Table 12: Width-256, six-block hard-case learned-score follow-up. Values are medians over six jobs.



(a) Predicted population order.

(b) Conditioning controls recovery.

Figure 4: Two theorem-facing checks. Lines show medians and bands show interquartile ranges. Left: exact- and joint-center errors follow the $O(\sigma)$ reference. Right: additional, well-selected scalar queries stabilize noisy recovery.

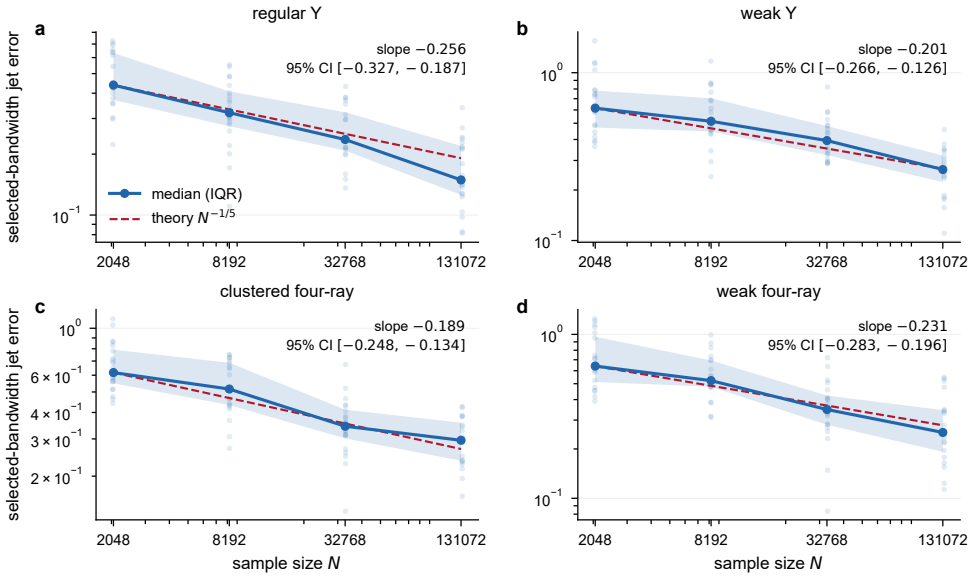


Figure 5: KDE finite-sample rate by geometry. Faint points are individual seeds at the median-selected bandwidth; solid curves and bands show the seed median and interquartile range. Red references have slope $-1/5$. Insets report 5,000-resample seed-bootstrap percentile intervals.

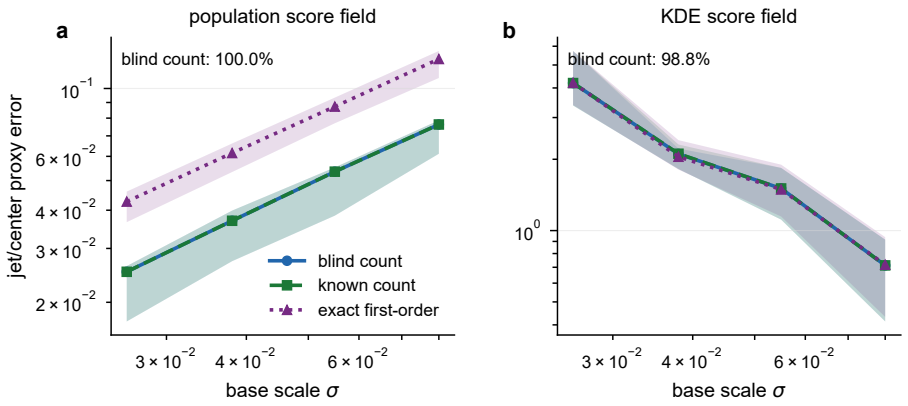


Figure 6: First-order frontend followed by second-order recovery. Bands are interquartile ranges over correct-count rows. Population error decreases with scale; at fixed sample size, KDE variance reverses the trend. Blind-count accuracy is computed on all rows.

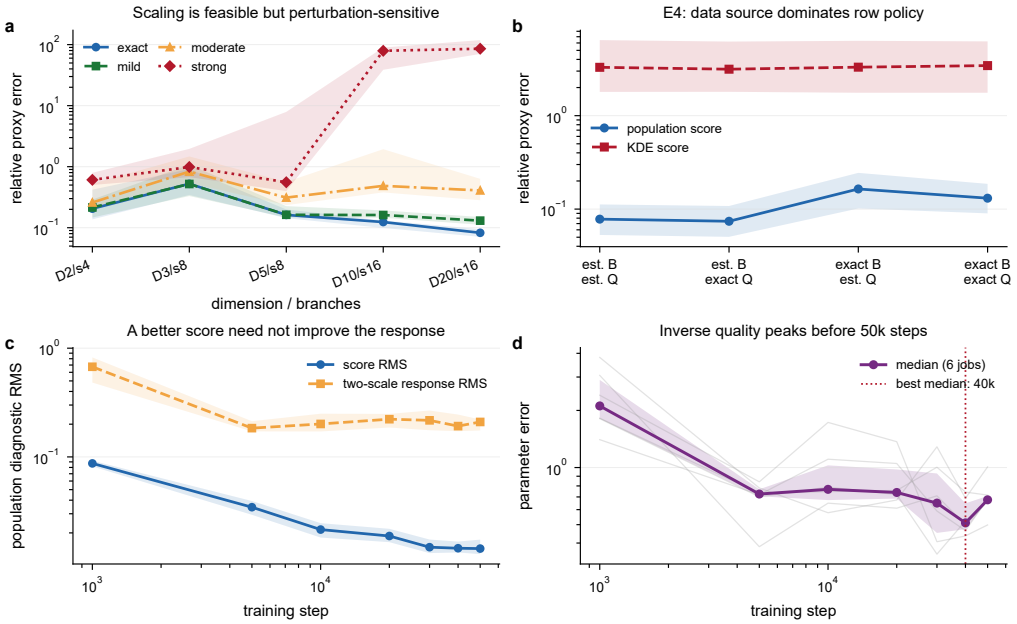


Figure 7: Supplemental diagnostics. (a) Dimension/branch scale-up under exact through strong first-order perturbations. (b) E4 with first-order basis (B) and query-row geometry (Q) crossed. (c–d) Width-256 learned-score follow-up through 50k updates; bands are interquartile ranges, and thin curves in (d) are all six jobs.

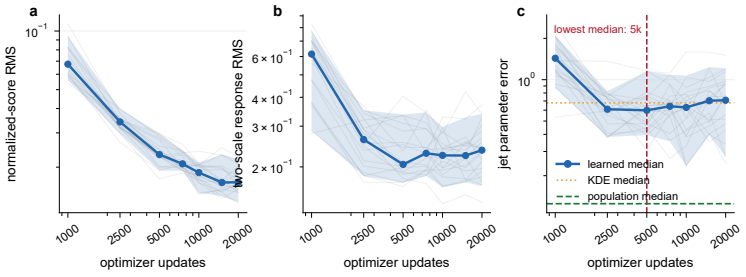


Figure 8: Learned-score diagnostics across seven checkpoints. Thin curves show all 20 jobs; thick curves and bands show the median and 5th–95th percentiles. Median jet error reaches its minimum at 5000 updates before score RMS reaches its lowest value; dotted and dashed lines are matched KDE and population baselines.