

Score Geometry at Vanishing Density: Tilted-Gaussian Limits and Spectrally Calibrated Optimization

Johannes Herter

Independent Researcher

Abstract. For small Gaussian noise, denoiser values approximate nearest-point projection onto a data manifold, while denoiser Jacobians are often treated as tangent-space projectors. At a regular positive-density point, the Jacobian has this projector limit, whereas full support alone does not guarantee it: the density may vanish at an isolated point without removing that point from the support. We analyze zeros of finite order. If the first nonzero local density jet is a nonnegative homogeneous polynomial q_m of degree m , then the normalized posterior covariance, equivalently the denoiser Jacobian, converges to $UC_{q_m}U^\top$. The matrix C_{q_m} is the covariance of a q_m -tilted Gaussian. The limit is generally not a projector, but its range is the tangent space, it is positive definite there, and its positive eigenvalues sum to $k + m$. A smooth full-support circle makes the failure explicit. Density-floor and moving-query calculations show that the altered spectral scaling persists on the noise scale rather than only at the zero. Because the limiting tangent range remains separated from the normal space, top- k spectral projection recovers the tangent projector asymptotically, and it is stable to estimation error below half the tangent-normal eigengap. A residual form of the same bound replaces that error by its tangent-normal coupling block, so error confined to eigenvalue scale does not move the recovered tangent space. We use that observation to analyze a calibrated manifold optimizer, including population and learned oracle errors and a local invariant basin. Population quadrature, repeated empirical-posterior estimation, and learned-denoiser experiments on circles and a rotated $d = 16, k = 2$ model separate recovery of spectral range from recovery of eigenvalue scale. In an exactly linearized model problem, calibrating both the tangent update and the landing operation widens the local stability interval to one that no longer depends on the vanishing order, and in learned step-size sweeps it correspondingly enlarges the region of low final objective gap.

1 Introduction

Many high-dimensional distributions are modeled as being supported on, or concentrated near, a smooth low-dimensional manifold $\mathcal{M} \subset \mathbb{R}^d$. A Gaussian denoiser trained on such a distribution also contains information about this local geometry. Tweedie’s identity writes the posterior mean as a score step, while differentiation identifies the denoiser Jacobian with a normalized posterior covariance. These identities motivate the use of denoiser values as approximations to nearest-point projection and denoiser Jacobians as approximations to tangent projection [2, 3, 15, 18, 23].

At a point of positive manifold density, the posterior in tangent coordinates becomes asymptotically standard Gaussian, so its covariance tends to the identity. At a density zero, the same rescaling retains the first nonzero density jet. The limiting tangent law is a *polynomially tilted* Gaussian, whose covariance need not be the identity. Full topological support rules out an open set of zero mass but not an isolated density zero. This separates tangent range from eigenvalue scale and raises the questions of which geometry remains identifiable and whether it is sufficient for optimization.

At every finite-order zero, the denoiser Jacobian has a limit of the form $UC_qU^\top$, with tangent range and the jet-shape-independent trace law $\text{tr } C_q = k + m$. A two-order expansion gives an $O(\sigma^2)$ operator-norm remainder. A smooth full-support circle supplies a counterexample to an unqualified projector limit [15]. The density-floor crossover $\delta \asymp \sigma^m$ describes nearby positive floors. A uniform normal-offset boundary-layer

expansion treats queries near a fixed zero and gives the corresponding eigengap and landing bias.

The surviving range suggests a correction. Hard spectral calibration recovers the tangent projector, while the soft filter approximates it; both admit perturbation bounds under Jacobian error, and the uncalibrated trace continues to record the vanishing order asymptotically. For spectrally calibrated denoising Riemannian gradient descent (SC-DRGD), we prove global inexact-oracle bounds conditional on the visited set. A local invariant-basin result verifies that condition near an ambient stationary point whose tangent Hessian is positive definite. The evidence follows the same progression: population quadrature, 48-replicate finite-sample posterior estimation, and 12 independent training seeds per problem, for 36 noise-conditional MLP denoisers in total. The learned tests include a rotated $d = 16, k = 2$ anisotropic zero and end-to-end optimization ablations.

What the spectrum retains

At a finite-type density zero, the limiting eigenspace and eigenvalues answer different questions. The nonzero eigenspace of the limiting operator is $T_p\mathcal{M}$, but its eigenvalues depend on the leading density jet and sum to $k + m$. Spectral projection of that operator retains the tangent space while discarding this density-dependent scale.

2 Related work

Scores, denoisers, and posterior covariance. Tweedie’s formula provides the classical empirical-Bayes link between scores and posterior means [12]. Denoising score matching turns that identity into a learning principle [26], and Alain and Bengio [2] connected the local behavior of regularized autoencoders to the score. In diffusion models, posterior covariance is used to describe local uncertainty and geometry [13, 18]; its leading eigenvalues also serve as an out-of-distribution score [22]. At a finite-type zero, our calculation separates spectral range from scale: the nonzero eigenspace can equal the tangent space even when the leading eigenvalues are inflated.

Diffusion geometry and intrinsic dimension. That a score-based model recovers the support of the data distribution at all was established by Pidstrigach [20], and quantitative convergence under the manifold hypothesis by De Bortoli [9]. Recovering the support is a weaker property than recovering the tangent metric, and the two come apart at a density zero: the support is unchanged there, while the limiting eigenvalues are not. Small-noise score derivatives separate tangent and normal directions and support local intrinsic-dimension estimation [19, 23, 25]. The Fokker-Planck estimator of Kamkari et al. [14] reads local intrinsic dimension from the same small-noise limit, and its correctness argument treats affine submanifolds, so a finite-type zero is orthogonal to that analysis rather than in competition with it. Leung et al. [16] extend that analysis to disjoint unions of submanifolds of differing dimension, a form of heterogeneity across components rather than within a single chart. The density model matters for this interpretation. Yeats et al. [27] analyze score matching under a locally constant-density approximation, while Chen et al. [6] obtain $O(\sigma^2)$ density-derivative expansions and nonparametric estimators for tangent spaces and curvature under uniform manifold sampling. At a finite-type zero, the leading density jet instead survives the rescaling. Sampling-density correction has an analogous role in graph and kernel methods that recover intrinsic geometry [4, 7]. For denoiser Jacobians, the trace identity gives an exact counterpart: at a finite-type zero, the uncorrected spectral trace converges to $k + m$ rather than k in the small-noise limit. The local Hessian filtering of Osada [19] concerns high-dimensional normal noise rather than finite-order density zeros or the tilted-Gaussian limit.

Optimization over implicit manifolds. Classical Riemannian optimization assumes access to tangent projectors and retractions [1, 5]. Work on learned manifolds replaces these operations with denoisers and their Jacobians [15], learns a conditional carré-du-champ operator [3], or constructs pullback and iso-Riemannian geometries [10, 11]. Saito and Matsubara [21] instead build a training-free metric on noise space directly from the score Jacobian, relying on the same tangent-normal spectral separation that Proposition 5.1 exploits. Up to bandwidth normalization, the carré-du-champ target is the mean-centered conditional covariance studied here. Theorem 4.2 determines its limit at finite-type zeros. Their Corollary 4.2 states the tangent-projector limit for a smooth density without a positivity requirement, so the circle of Section 4.1 is a counterexample

there as well: since the conditional carré-du-champ operator equals $\frac{1}{2}H_\sigma$ at bandwidth $\varepsilon = \sigma^2$, its limit at that density zero is $\frac{3}{2}tt^\top$ rather than $tt^\top$. Other score-geometry analyses impose an explicit positive lower bound on the density [17, 23, 28]. The present analysis allows the density to vanish and identifies the geometric information that remains.

3 Setup and exact score identities

Let $\mathcal{M} \subset \mathbb{R}^d$ be an embedded k -dimensional manifold without boundary, and let the clean variable $X \sim \mu$ have density $\rho \geq 0$ with respect to Riemannian volume on $\mathcal{M}$. We assume $1 \leq k < d$ throughout. The full-dimensional case has no tangent-normal separation and is omitted. Appendix A collects the recurring notation. Throughout, $\|\cdot\|$ is the Euclidean norm on vectors and $\|\cdot\|_{\text{op}}$ the operator norm on matrices; applied to a map it is the Euclidean norm of the value at the stated point. For $G \sim \mathcal{N}(0, I_d)$ independent of X , observe

$$Y_\sigma = X + \sigma G, \quad p_\sigma(y) = (2\pi\sigma^2)^{-d/2} \int_{\mathcal{M}} e^{-\|y-x\|^2/(2\sigma^2)} \mu(dx). \quad (1)$$

Write $s_\sigma = \nabla \log p_\sigma$ and define the population denoiser and link Hessian

$$D_\sigma(y) = y + \sigma^2 s_\sigma(y), \quad H_\sigma(y) = \nabla D_\sigma(y) = I_d + \sigma^2 \nabla^2 \log p_\sigma(y). \quad (2)$$

Proposition 3.1 (Tweedie mean and covariance identities). *Whenever $p_\sigma(y) > 0$,*

$$D_\sigma(y) = \mathbb{E}[X \mid Y_\sigma = y], \quad H_\sigma(y) = \sigma^{-2} \text{Cov}(X \mid Y_\sigma = y) \succeq 0. \quad (3)$$

The second identity shows that the tangent eigenvalues of H_σ are variances of the rescaled local posterior rather than fixed geometric constants. Appendix B proves both identities by differentiating under the integral.

4 Finite-type density zeros

Fix $p \in \mathcal{M}$. Let $U_p : \mathbb{R}^k \rightarrow T_p \mathcal{M}$ have orthonormal columns, and use a normal graph chart

$$F_p(z) = p + U_p z + g_p(z), \quad g_p(z) \in N_p \mathcal{M}, \quad g_p(0) = Dg_p(0) = 0. \quad (4)$$

Let $J_p(z) = \sqrt{\det(DF_p(z)^\top DF_p(z))}$ and define the effective chart density $a_p(z) = \rho(F_p(z))J_p(z)$. Since $DF_p^\top DF_p = I + Dg_p^\top Dg_p$ and $Dg_p(0) = 0$, we have $J_p(z) = 1 + O(\|z\|^2)$, so a_p and $\rho \circ F_p$ have the same jets through order two; the leading jet in equation 5 may therefore be read off the density itself whenever $m \leq 2$.

Definition 4.1 (Finite-type point). The density has a *finite-type point of order m* at p if, for a nonzero homogeneous polynomial q_m of degree m ,

$$a_p(z) = q_m(z) + q_{m+1}(z) + O(\|z\|^{m+2}), \quad (5)$$

where q_{m+1} is homogeneous of degree $m+1$. For the limit without the stated rate, it is enough that $a_p(z) = q_m(z) + o(\|z\|^m)$. The case $m = 0$ is a regular positive-density point; a finite-type *zero* has $m \geq 2$. The terminology follows the classification of degenerate critical points by the order of vanishing in the asymptotic analysis of oscillatory and Laplace-type integrals [24].

Because $a_p \geq 0$, its first nonzero homogeneous jet is nonnegative. It cannot have odd degree: otherwise $q_m(-z) = -q_m(z)$ would force $q_m \equiv 0$. Thus every finite vanishing order is even. This parity conclusion is a statement about polynomial jets, not about density zeros in general. Polynomiality of q_m is used only here and in Lemma D.1, whose uniform constant ranges over the finite-dimensional space of degree- m forms. The limit equation 8 and the trace law need only that q_m be continuous, nonnegative, and homogeneous

of some degree $m > 0$, because the trace identity follows from Euler's relation. Effective densities with $a_p(z) \asymp \|z\|^3$ or $\|z\|^{1/2}$ are of that kind and satisfy $\text{tr } C_{q_m} = k + m$ with $m = 3$ and $m = 1/2$, yet they lie outside Definition 4.1; Section 8 records this restriction.

Because q_m is even, the q_m -tilted standard Gaussian has mean zero. For $Z \sim \mathcal{N}(0, I_k)$, define its covariance

$$C_{q_m} := \frac{\mathbb{E}[ZZ^\top q_m(Z)]}{\mathbb{E}[q_m(Z)]}. \quad (6)$$

When the degree is clear, we write $C_q = C_{q_m}$.

Theorem 4.2 (Score geometry at a finite-type point). *Suppose $\mathcal{M}$ is C^3 near p , its normal graph satisfies $g_p(z) = \frac{1}{2}\Pi_p[z, z] + R_g(z)$ with $\|R_g(z)\| \leq C\|z\|^3$, and the effective density satisfies equation 5. Then m is even, $q_m \geq 0$, and, as $\sigma \downarrow 0$,*

$$D_\sigma(p) = p + \sigma^2 b_{p,q} + O(\sigma^3), \quad (7)$$

$$H_\sigma(p) = U_p C_{q_m} U_p^\top + O(\sigma^2) \quad \text{in operator norm.} \quad (8)$$

Moreover,

$$C_{q_m} \succ 0, \quad \text{range}(U_p C_{q_m} U_p^\top) = T_p \mathcal{M}, \quad \text{tr}(C_{q_m}) = k + m. \quad (9)$$

If only $a_p(z) = q_m(z) + o(\|z\|^m)$ and $\mathcal{M}$ is C^1 near p with $g_p(z) = o(\|z\|)$, the convergence in equation 8 still holds without a rate. The C^1 requirement is what makes J_p , and hence a_p , defined.

Corollary 4.3 (Positive-density projector limit and coordinate invariance). *Under the regularity assumptions of Theorem 4.2, if $\rho(p) > 0$, then $m = 0$, $C_{q_0} = I_k$, and $H_\sigma(p) \rightarrow P_{T_p \mathcal{M}}$. More generally, the ambient limit in equation 8 is independent of the orthonormal tangent frame: replacing U_p by $U_p O$, $O \in O(k)$, replaces C_{q_m} by $O^\top C_{q_m} O$ and leaves $U_p C_{q_m} U_p^\top$ unchanged. The order m and the limit are likewise unchanged under any local parametrization ψ with $\psi(0) = p$ and $D\psi(0) = U_p$, since such a reparametrization is tangent to the identity and multiplies the effective density by a factor tending to one, so it preserves the leading jet. This is what licenses computing the jet in a convenient coordinate, such as arc length in Section 4.1, rather than in the normal graph chart of equation 4.*

Appendix D identifies $b_{p,q}$ as the sum of a contribution from the next odd density jet and a q_m -weighted contraction of the second fundamental form. For the covariance, the proof localizes and rescales $z = \sigma u$; parity removes the nominal order- σ correction, leaving a limit determined by the leading density jet in the chosen tangent frame. Figure 1 shows this mechanism. Positive definiteness follows because a nonzero polynomial cannot concentrate on a hyperplane. Gaussian integration by parts and Euler's homogeneous-function identity establish equation 9.

At the zero

The projector limit fails in scale, not in range: the limiting link Hessian has rank k and range $T_p \mathcal{M}$, while $\text{tr } H_\sigma(p) - k \rightarrow m$. The excess trace is exactly the local vanishing order.

Closed forms. If $q_m(z) = \|z\|^m$ (with even m), rotational symmetry and equation 9 give

$$C_{q_m} = \left(1 + \frac{m}{k}\right) I_k. \quad (10)$$

For a quadratic zero $q_2(z) = z^\top A z$, $A \succeq 0$ and $A \neq 0$, Isserlis' identity gives

$$C_{q_2} = I_k + \frac{2A}{\text{tr } A}. \quad (11)$$

Thus even the limiting tangent metric can be anisotropic.

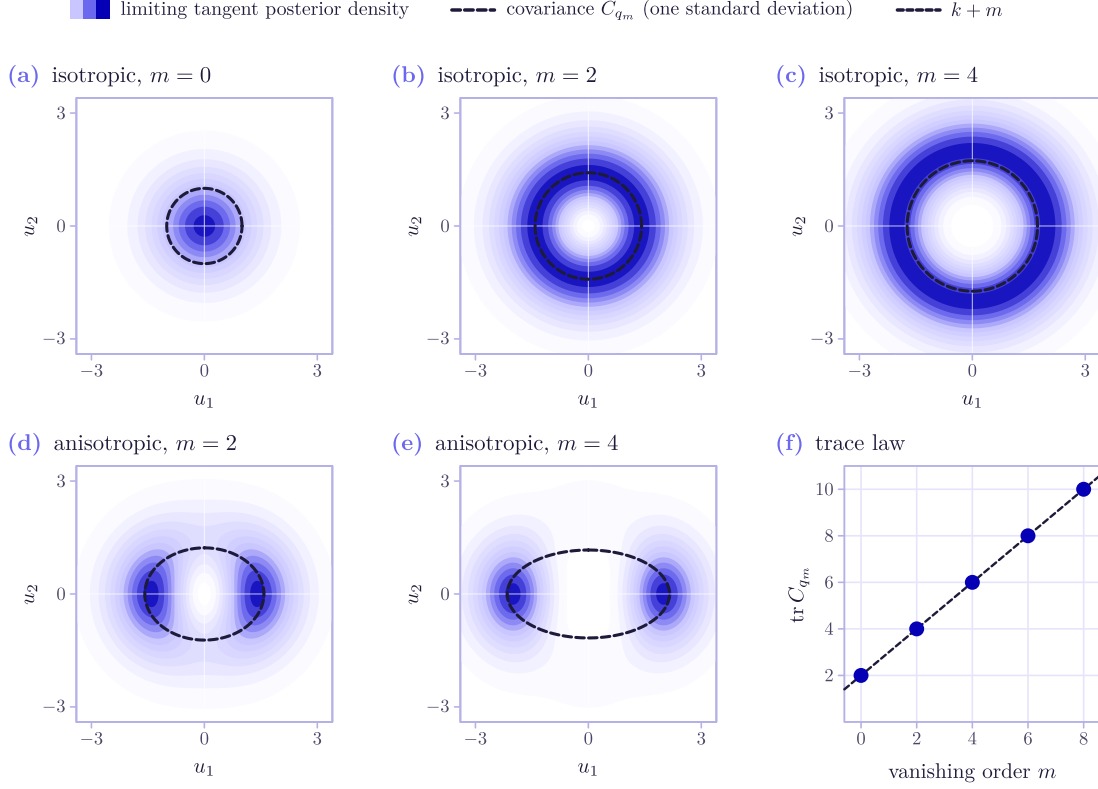


Figure 1 Why the vanishing order survives the small-noise rescaling, drawn in rescaled tangent coordinates with $k = 2$. Shading is the limiting tangent posterior density and the dashed curve is its covariance C_{q_m} . The top row uses isotropic jets $q_m(u) = \|u\|^m$ of increasing order: at a regular point (a) the posterior is standard Gaussian and $C_{q_0} = I$, so $H_\sigma(p)$ tends to the tangent projector, while at the zeros (b) and (c) the jet hollows out the centre and the covariance inflates to $2I$ and $3I$ without changing shape. The bottom row uses anisotropic jets at the same two orders, $q_2(u) = 3u_1^2 + u_2^2$ in (d) and the quartic $q_4(u) = u_1^4 + 0.1u_2^4$ of Section 7 in (e), giving $\text{diag}(2.5, 1.5)$ and $\text{diag}(4.636, 1.364)$. At matched order the trace is the same, 4 in both (b) and (d) and 6 in both (c) and (e), while the shape is not. The order therefore fixes the trace and the jet fixes the shape, and in every panel the ellipse stays centred and nondegenerate, so the tangent range is retained throughout. Panels (a) to (e) use equal scales on both axes, so the ellipses carry their true shape. (f) The trace against the vanishing order for the isotropic family, with the dashed line $k + m$ of equation 9; the plotted values agree with it exactly.

4.1 A smooth full-support counterexample

Let $\mathcal{M} = S^1$, $x(\theta) = (\cos \theta, \sin \theta)$, $p = (1, 0)$, and

$$\rho(\theta) = \frac{1 - \cos \theta}{2\pi}. \quad (12)$$

This density is C^∞ , nonnegative, and zero only at p . Consequently, $\text{supp}(\mu) = S^1$. Since $1 - \cos \theta = \theta^2/2 + O(\theta^4)$, Theorem 4.2 yields, for the tangent vector $t = (0, 1)$,

$$H_\sigma(p) \longrightarrow 3tt^\top \neq tt^\top = P_{T_p S^1}. \quad (13)$$

Indeed, after $\theta = \sigma u$, the tangent posterior is proportional to $u^2 e^{-u^2/2}$ and has variance $\mathbb{E}Z^4/\mathbb{E}Z^2 = 3$. Appendix E collects this circle family and the radial and quadratic jets.

At positive-density points, Corollary 4.3 recovers the usual projector limit. The stated assumptions of Theorem 1 in arXiv v2 (28 February 2026) of Kharitenko et al. [15] also permit isolated zeros, because full support is weaker than pointwise positivity.¹ The circle example shows that the conclusion requires a local

¹That theorem assumes “ $\mathcal{M} \subseteq \mathbb{R}^d$ a compact, embedded C^3 -submanifold and $\mu \in \mathcal{P}(\mathbb{R}^d)$ a Borel probability measure with

positivity condition. The local estimate in their Appendix E.2 carries $\mu(p)$ and $\mu(p)^2$ in denominators, and their Appendix E.4 supplies the missing condition by asserting that “since $\mu(\cdot) \in C^3(\mathcal{M})$ and $\text{supp } \mu = \mathcal{M}$, there exist positive lower and upper bounds on $\mu(\cdot)$ and its derivatives on $\mathcal{M}$ ”. With this condition, that result covers the positive-density regime, while Theorem 4.2 characterizes the complementary finite-type regime.

4.2 A density-floor phase transition

Although a zero is defined pointwise, its effect persists across a noise-dependent low-density regime. Fix a sufficiently small coordinate neighborhood $U = F_p(B(0, r_0))$. Let an unnormalized family satisfy, for $z \in B(0, r_0)$ and $0 \leq \delta \leq \delta_0$,

$$a_\delta(z) = q_m(z) + \delta h(z) + r_\delta(z), \quad h \text{ continuous}, \quad h(0) = a_0 > 0, \quad (14)$$

where $|r_\delta(z)| \leq C(\|z\|^m + \delta)$ on this fixed neighborhood and $r_\delta(z)/(\|z\|^m + \delta) \rightarrow 0$ uniformly in the joint limit $(z, \delta) \rightarrow (0, 0)$. After fixing the common local normalization, assume that the total unnormalized mass on $\mathcal{M} \setminus U$ is uniformly bounded as $\delta \downarrow 0$. The weaker subexponential condition $\exp(o(\sigma^{-2}))$ along $\delta = \delta_\sigma$ also suffices. Normalization is immaterial to the posterior. Let $\delta = \delta_\sigma \downarrow 0$ with $\delta_\sigma/\sigma^m \rightarrow \lambda \in [0, \infty]$.

Theorem 4.4 (Density-floor crossover). *Under the local assumptions of Theorem 4.2, the tangent block of $H_\sigma(p)$ converges to*

$$C_{q,\lambda} = \frac{a_0 \lambda I_k + \mathbb{E}[ZZ^\top q_m(Z)]}{a_0 \lambda + \mathbb{E}[q_m(Z)]}, \quad \text{tr } C_{q,\lambda} = k + \frac{m \mathbb{E}[q_m(Z)]}{a_0 \lambda + \mathbb{E}[q_m(Z)]}. \quad (15)$$

For $\lambda = \infty$, the limit is I_k .

Appendix F proves the law by rescaling the floor against the leading jet. Thus $\delta \ll \sigma^m$ is zero-dominated, $\delta \gg \sigma^m$ is regular-density, and $\delta \asymp \sigma^m$ interpolates nontrivially. On the circle with $a_\delta(\theta) = \delta + 1 - \cos \theta$,

$$\lambda = \delta/\sigma^2, \quad C_{q,\lambda} = \frac{\lambda + 3/2}{\lambda + 1/2}. \quad (16)$$

There is a distinct spatial boundary layer for a *fixed* distribution. Query the denoiser at the nearby manifold point $y_\sigma = F_p(\sigma a)$, where $a \in \mathbb{R}^k$ is fixed, and let $\nu_{q,a}$ be the probability law with density proportional to

$$q_m(u) \exp\left(-\frac{1}{2}\|u - a\|^2\right). \quad (17)$$

Applying the rescaling from Theorem 4.2 gives the following result.

Proposition 4.5 (Spatial boundary layer). *Under the minimal convergence assumptions of Theorem 4.2,*

$$H_\sigma(F_p(\sigma a)) \longrightarrow U_p C_q^{\text{sp}}(a) U_p^\top, \quad C_q^{\text{sp}}(a) = \text{Cov}_{\nu_{q,a}}(u). \quad (18)$$

Equivalently, if $M_q(a) = \mathbb{E}[q_m(a + Z)]$, then

$$C_q^{\text{sp}}(a) = I_k + \nabla_a^2 \log M_q(a). \quad (19)$$

For $k = 1$ and $q_2(u) = u^2$,

$$C_q^{\text{sp}}(a) = \frac{a^4 + 3}{(a^2 + 1)^2}. \quad (20)$$

It equals 3 at the zero, crosses 1 at $|a| = 1$, reaches 3/4 at $|a| = \sqrt{3}$, and tends back to 1 from below.

$\text{supp } \mu = \mathcal{M}$ and $\mu \ll \text{Vol}_{\mathcal{M}}$ such that $\frac{d\mu}{d\text{Vol}_{\mathcal{M}}} \in C^3(\mathcal{M})$, and its Corollary 2 concludes $\lim_{\sigma \rightarrow 0} I + \sigma^2 \nabla^2 \log p_\sigma(x) = P_{T_x \mathcal{M}}$ for $x \in \mathcal{M}$. The density equation 12 satisfies each of these hypotheses.

Appendix F proves this by the same moving-query rescaling. Under the two-order density expansion and third-order graph bound in Theorem 4.2, the same moving-query calculation gives, uniformly for a in compact sets,

$$D_\sigma(F_p(\sigma a)) - F_p(\sigma a) = \sigma U_p(\mathbb{E}_{\nu_{q,a}} u - a) + O(\sigma^2), \quad (21)$$

$$H_\sigma(F_p(\sigma a)) = U_p C_q^{\text{sp}}(a) U_p^\top + O(\sigma). \quad (22)$$

The normal component of the first line is $O(\sigma^2)$, and Davis–Kahan applied to the second line gives $\|\mathcal{P}_k(H_\sigma(F_p(\sigma a))) - P_{\mathbb{T}_{F_p(\sigma a)}\mathcal{M}}\|_{\text{op}} = O(\sigma)$.

The on-manifold calculation extends to normal offsets of the same scale. Write $P = U_p U_p^\top$, $P^\perp = I - P$, and set

$$\mathcal{B}(u) = \frac{1}{2} \Pi_p[u, u], \quad \mu_a = \mathbb{E}_{\nu_{q,a}} u, \quad C_a = \text{Cov}_{\nu_{q,a}}(u). \quad (23)$$

Also define the normal-valued covariance and its calibrated version by

$$\Gamma_a = \mathbb{E}_{\nu_{q,a}}[(\mathcal{B}(u) - \bar{\mathcal{B}}_a)(u - \mu_a)^\top], \quad \Lambda_a = \Gamma_a C_a^{-1}, \quad \bar{\mathcal{B}}_a = \mathbb{E}_{\nu_{q,a}} \mathcal{B}(u). \quad (24)$$

For $b \in \mathbb{N}_p \mathcal{M}$, let the shape operator S_b and the normal map $\mathcal{B}_a$ be given by

$$\langle S_b v, w \rangle = \langle b, \Pi_p[v, w] \rangle, \quad \mathcal{B}_a v = \Pi_p[a, v]. \quad (25)$$

Theorem 4.6 (Uniform normal-offset boundary layer). *Assume the two-order density expansion and third-order graph expansion of Theorem 4.2, including $\|DR_g(z)\|_{\text{op}} \leq C\|z\|^2$. Fix compact sets $K_T \subset \mathbb{R}^k$ and $K_N \subset \mathbb{N}_p \mathcal{M}$. Assume also that, for some fixed chart radius r , constants $c, C > 0$, and all sufficiently small σ ,*

$$\sup_{a \in K_T, b \in K_N} \int_{\mathcal{M} \setminus F_p(B(0, r))} (1 + \|x - p\|^2) e^{-\|x - F_p(\sigma a) - \sigma b\|^2 / (2\sigma^2)} \mu(dx) \leq C e^{-c/\sigma^2}. \quad (26)$$

This localization condition is automatic when a fixed Euclidean neighborhood of p meets $\mathcal{M}$ only in the chart; in particular it holds for compact $\mathcal{M}$. For the two statements below that involve nearest-point projection we assume in addition that $\mathcal{M}$ has positive reach and that σ is small enough for $\mathcal{V}_\sigma(K_T, K_N)$ to lie within a tubular neighborhood of radius smaller than that reach, so π is single valued on it; the expansions of D_σ , H_σ , and Q_σ do not use π . Let

$$\mathcal{V}_\sigma(K_T, K_N) = \{F_p(\sigma a) + \sigma b : a \in K_T, b \in K_N\}. \quad (27)$$

Then, uniformly for $a \in K_T$ and $b \in K_N$,

$$D_\sigma(y) - y = \sigma \{U_p(\mu_a - a) - b\} + O(\sigma^2), \quad (28)$$

$$U_p^\top H_\sigma(y) U_p = C_a + O(\sigma), \quad (29)$$

$$P^\perp H_\sigma(y) U_p = \sigma \Gamma_a + O(\sigma^2), \quad P^\perp H_\sigma(y) P^\perp = \sigma^2 \text{Cov}_{\nu_{q,a}}(\mathcal{B}(u)) + O(\sigma^3).$$

In particular, if $c_0 = \inf_{a \in K_T} \lambda_{\min}(C_a) > 0$ and $Q_\sigma(y) = \mathcal{P}_k(H_\sigma(y))$, then, for all sufficiently small σ ,

$$\lambda_k(H_\sigma(y)) \geq c_0 - O(\sigma), \quad \lambda_{k+1}(H_\sigma(y)) = O(\sigma^2), \quad (30)$$

$$Q_\sigma(y) = P + \sigma \{\Lambda_a U_p^\top + U_p \Lambda_a^\top\} + O(\sigma^2). \quad (31)$$

Moreover, nearest projection and population normal-only landing obey

$$\pi(y) = F_p(\sigma a + \sigma^2 S_b a + O(\sigma^3)), \quad (32)$$

$$\bar{D}_{\sigma, Q_\sigma}(y) - \pi(y) = \sigma^2 \left[U_p \{\Lambda_a^\top b - S_b a\} + \bar{\mathcal{B}}_a - \mathcal{B}(a) - \Lambda_a(\mu_a - a) \right] + O(\sigma^3), \quad (33)$$

where $\bar{D}_{\sigma, Q}(y) = y + (I - Q(y))(D_\sigma(y) - y)$. All remainders are uniform on $K_T \times K_N$.

Finally, let $W_{F_p(z)} = DF_p(z)(DF_p(z)^\top DF_p(z))^{-1/2}$ and set $A_{0,\sigma}(y) = W_{\pi(y)}C_dW_{\pi(y)}^\top$. Then $\text{range}(A_{0,\sigma}(y)) = T_{\pi(y)}\mathcal{M}$ and, on $\mathcal{V}_\sigma(K_T, K_N)$,

$$\beta_H(\sigma) = O(\sigma), \quad \beta_L(\sigma) = O(\sigma^2), \quad g_\sigma \geq c_0/2, \quad \sup_y \|D_\sigma(y) - y\| = O(\sigma), \quad (34)$$

in the notation of Corollary 6.4, where β_H is the supremum over the query set of $\|H_\sigma - A_{0,\sigma}\|_{\text{op}}$, that is the population bias of the link Hessian against the reference field just defined, β_L is the corresponding supremum of the normal-only landing error $\|\bar{D}_{\sigma, Q_\sigma} - \pi\|$, and g_σ is the infimum of the tangent-normal eigengap $\lambda_k(H_\sigma) - \lambda_{k+1}(H_\sigma)$.

Appendix F.3 proves it. The theorem concerns a shrinking $O(\sigma)$ tube, the scale needed for the uniform population bounds used in local optimization. On a fixed-width tube, Lemma 6.1 shows why the same $O(\sigma^2)$ landing rate generally fails.

Because $C_{q,\lambda}$ is a scalar convex combination of C_q and I_k , each eigenvalue moves monotonically toward one as the floor grows. The matrix path need not be monotone in Loewner order. The spatial law need not be eigenvalue-monotone either, because shifting the Gaussian window changes both the posterior mean and variance. The experiments therefore treat the two mechanisms separately.

Two nearby regimes

A positive density floor does not restore regular-density geometry until it is resolved at the noise scale: floors $\delta = o(\sigma^m)$ retain the finite-zero limiting spectrum. With the density fixed, queries within $O(\sigma)$ of the zero instead follow the spatial boundary-layer law. The two transitions arise from different perturbations.

5 Spectral calibration

Let $A_0 = UC_qU^\top$ and $c_q = \lambda_{\min}(C_q) > 0$. Given a symmetric estimate $\hat{A}$, define the hard calibration

$$\mathcal{P}_k(\hat{A}) := \text{orthogonal projector onto the top-}k \text{ eigenspace of } \hat{A}. \quad (35)$$

If k is unknown, it can be selected by thresholding inside the tangent-normal spectral gap under the condition stated below. For a differentiable score network s_θ , the input is the symmetrized link Jacobian

$$\hat{A}_\theta(y) = \frac{1}{2}(A_\theta(y) + A_\theta(y)^\top), \quad A_\theta(y) = I_d + \sigma^2 \nabla s_\theta(y). \quad (36)$$

Proposition 5.1 (Stable hard calibration). *If $\|\hat{A} - A_0\|_{\text{op}} \leq \eta < c_q/2$, then*

$$\|\mathcal{P}_k(\hat{A}) - UU^\top\|_{\text{op}} \leq \frac{2\eta}{c_q}. \quad (37)$$

Write $P = UU^\top$ and $P^\perp = I_d - P$, and let $U_\perp$ have orthonormal columns spanning $\text{range}(U)^\perp$, so that its columns form a basis of the normal space. Set

$$\delta(\hat{A}) := \lambda_{\min}(U^\top \hat{A} U) - \lambda_{\max}(U_\perp^\top \hat{A} U_\perp). \quad (38)$$

Whenever $\delta(\hat{A}) > 0$, without any hypothesis on η ,

$$\|\mathcal{P}_k(\hat{A}) - UU^\top\|_{\text{op}} \leq \frac{\|P^\perp \hat{A} P\|_{\text{op}}}{\delta(\hat{A})}, \quad (39)$$

in which only the tangent-normal coupling block of the error appears: since $P^\perp A_0 P = 0$, we have $P^\perp \hat{A} P = P^\perp(\hat{A} - A_0)P$. In particular an error confined to the tangent block leaves $\mathcal{P}_k(\hat{A}) = UU^\top$

exactly, however large its eigenvalue bias. Under the standing hypothesis $\eta < c_q/2$, Weyl's inequality gives $\delta(\hat{A}) \geq c_q - 2\eta > 0$, so equation 39 is then available with a denominator depending only on c_q and η . Both blocks in equation 38 are computable from $\hat{A}$ and a tangent frame, so equation 39 remains informative in regimes where $\eta < c_q/2$ fails. For every threshold $\vartheta_{\text{sp}} \in (\eta, c_q - \eta)$, $\mathbf{1}_{(\vartheta_{\text{sp}}, \infty)}(\hat{A})$ has rank k and equals $\mathcal{P}_k(\hat{A})$; thus the same condition gives unknown- k rank recovery. In addition,

$$\left| \sum_{i=1}^k \lambda_i(\hat{A}) - (k + m) \right| \leq k\eta; \quad (40)$$

hence $m = \text{round}(\sum_{i=1}^k \lambda_i(\hat{A}) - k)$ whenever $k\eta < 1/2$.

A soft alternative is obtained after projecting $\hat{A}$ onto the positive-semidefinite cone:

$$\Phi_\tau(\hat{A}_+) = \hat{A}_+(\hat{A}_+ + \tau I)^{-1}. \quad (41)$$

Proposition 5.2 (Stable soft calibration). *If $\hat{A} \succeq 0$ and $\|\hat{A} - A_0\|_{\text{op}} \leq \eta$, then for every $\tau > 0$,*

$$\|\Phi_\tau(\hat{A}) - UU^\top\|_{\text{op}} \leq \frac{\tau}{c_q} + \frac{\eta}{\tau}. \quad (42)$$

Balancing the two terms gives an $O(\sqrt{\eta/c_q})$ error without forming a hard top- k eigenspace.

Appendix G proves the hard bound with Davis–Kahan and the soft bound with the resolvent identity. If the estimator is PSD by construction, the rational filter avoids a hard eigenspace cutoff. A general symmetric score Jacobian may still require an eigendecomposition to form $\hat{A}_+$. The estimate $\|\hat{A} - A_0\|_{\text{op}} \leq \eta$ yields $\|\hat{A}_+ - A_0\|_{\text{op}} \leq 2\eta$, so for a symmetric estimate that is not positive semidefinite the composed form of Proposition 5.2 reads $\|\Phi_\tau(\hat{A}_+) - UU^\top\|_{\text{op}} \leq \tau/c_q + 2\eta/\tau$, which balances at $2\sqrt{2\eta/c_q}$. The raw operator A_0 remains a positive-definite preconditioner on the tangent space and therefore preserves first-order stationary points. Calibration removes density-dependent conditioning and gives the step size the same interpretation as in Riemannian gradient descent. The optimization results below use the hard rank- k orthogonal projector. We evaluate the soft contraction in equation 41 only as a geometric estimator. The experiments assume known k , apart from a separately reported threshold-at-1/2 rank diagnostic. The present guarantees do not cover adaptive threshold selection. That diagnostic is licensed only when $c_q > 1/2 + \eta$, and c_q is not bounded below by $1/2$ uniformly in the vanishing order. For $k = 2$, minimizing $\lambda_{\min}(C_q)$ numerically over the nonnegative binary forms $q(z) = \prod_j (z_1 - t_j z_2)^2$ gives 0.879, 0.670, 0.534, 0.442, 0.380, and 0.333 at orders $m = 4, 8, 12, 16, 20, 24$. The diagnostic is therefore sound at the orders used in Section 7 but fails for deep zeros: at $m = 16$ a threshold of $1/2$ discards a genuine tangent direction even when the operator is exact. Lemma D.1 guarantees a positive constant depending on k and m , not a constant exceeding $1/2$.

6 Density-calibrated optimization

Calibrating only the gradient multiplier is insufficient if a raw denoiser is subsequently used as the landing map. We avoid the word retraction for $\overline{D}_Q$, which does not map into $\mathcal{M}$ and does not fix manifold points at positive σ , so it is not a retraction in the sense of Absil et al. [1]; the analysis never uses those axioms, only the landing error in equation 48. At a symmetric finite-type center, $D_\sigma(p) = p + O(\sigma^2)$ and $H_\sigma(p) \rightarrow UC_q U^\top$. For a locally isotropic quadratic objective centered at p , the tangent linearization of

$$x \mapsto D_\sigma(x - \gamma H_\sigma(x)) \nabla f(x) \quad (43)$$

converges to $C_q(I - \gamma C_q)$. Replacing the inner H_σ by $P_{T_p \mathcal{M}}$ still leaves the limiting factor $C_q(I - \gamma I)$: the outer denoiser reintroduces density-dependent expansion.

Normal-only landing avoids this effect. Given $Q(y)$, define

$$\overline{D}_Q(y) := y + (I - Q(y))(D(y) - y). \quad (44)$$

For the population denoiser of a distribution supported on an affine manifold, exact $Q = P_{T_p\mathcal{M}}$ makes this map the exact orthogonal projection for every σ and leaves tangent coordinates unchanged. On a curved manifold, it is an inexact landing map. The next result treats that error abstractly. Throughout this section, $\mathcal{M}$ is compact, C^3 , and has positive reach, and π denotes nearest-point projection in a reach tube $\mathcal{T}_r = \{x : \text{dist}(x, \mathcal{M}) < r\}$.

The spatial law quantifies the difference. For a one-dimensional quadratic zero and $y_\sigma = F_p(\sigma a)$,

$$D_\sigma(y_\sigma) - y_\sigma = \sigma U_p \frac{2a}{a^2 + 1} + O(\sigma^2). \quad (45)$$

Thus raw denoising has an $O(\sigma)$ tangent drift throughout the boundary layer, pointing away from the zero. Exact normal-only landing removes that term and leaves an $O(\sigma^2)$ curvature residual.

Lemma 6.1 (Normal-only landing decomposition). *Let $p = \pi(y)$, $v = y - p \in N_p\mathcal{M}$, $r_D = D(y) - p$, and let $Q(y)$ be an orthogonal projector satisfying $\|Q(y) - P_{T_p\mathcal{M}}\|_{\text{op}} \leq \epsilon_P$. Then*

$$\overline{D}_Q(y) - p = Q(y)v + (I - Q(y))r_D, \quad (46)$$

and therefore

$$\|\overline{D}_Q(y) - \pi(y)\| \leq \epsilon_P \text{dist}(y, \mathcal{M}) + \epsilon_P \|P_{T_p\mathcal{M}}r_D\| + \|(I - P_{T_p\mathcal{M}})r_D\|. \quad (47)$$

The first term cannot be omitted. Under the stronger moving-query expansion equation 21–equation 22, an on-manifold spatial boundary layer has $\epsilon_P = O(\sigma)$, an $O(\sigma)$ tangent component of r_D , and an $O(\sigma^2)$ normal component. Equation 47 then gives an $O(\sigma^2)$ error. On a fixed-width tube, the same bound permits only an $O(\sigma)$ guarantee because projector tilt can act on an $O(1)$ normal displacement. Appendix H proves the lemma and details this distinction.

Assumption 6.2 (Inexact normal landing and tangent oracles). Let $\mathcal{V} \subseteq \mathcal{T}_r$ be a set containing every iterate x_t and every intermediate point z_t of the recursion below. On $\mathcal{V}$, a landing map L and rank- k orthogonal projector Q satisfy

$$\|L(z) - \pi(z)\| \leq \epsilon_L, \quad \|Q(x) - P_{T_{\pi(x)}\mathcal{M}}\|_{\text{op}} \leq \epsilon_P. \quad (48)$$

The intended construction is $L = \overline{D}_Q$ from equation 44; the theorem also applies to an independently learned landing map satisfying the displayed error. We state the hypothesis on $\mathcal{V}$ rather than on all of $\mathcal{T}_r$ because the proof invokes these inequalities only at the visited points, and because the rates available from Theorem 4.6 are established on an $O(\sigma)$ layer rather than on a tube of fixed width. Taking $\mathcal{V} = \mathcal{T}_r$ is admissible whenever uniform bounds of this form are available there.

For an ambient differentiable objective f , define spectrally calibrated denoising Riemannian gradient descent (SC-DRGD):

$$z_t = x_t - \gamma Q(x_t) \nabla f(x_t), \quad x_{t+1} = L(z_t). \quad (49)$$

In practice $D = D_\theta$, $Q = \mathcal{P}_k(\hat{A}_\theta)$, and $L = \overline{D}_{\theta, Q}$. Evaluation-point and estimation differences are absorbed into equation 48.

Theorem 6.3 (Inexact SC-DRGD). *Suppose f has L_f -Lipschitz gradient and $\|\nabla f\| \leq B$ on $\mathcal{T}_r$.*

Let $F = f \circ \pi$ have L_F -Lipschitz gradient there, let π be L_π -Lipschitz, and assume

$$\|\mathrm{d}\pi_x - P_{T_{\pi(x)}\mathcal{M}}\|_{\mathrm{op}} \leq K_\pi \mathrm{dist}(x, \mathcal{M}). \quad (50)$$

Let L and Q satisfy Assumption 6.2 on a set $\mathcal{V}$ containing every x_t and z_t . Initialize $\mathrm{dist}(x_0, \mathcal{M}) \leq \epsilon_L$, which one landing step from any $x_{-1} \in \mathcal{T}_{r-\gamma B}$ achieves. If

$$\gamma B + \epsilon_L < r, \quad L_\pi \epsilon_L < r, \quad 0 < \gamma \leq (4L_F)^{-1}, \quad (51)$$

then every x_t is ϵ_L -feasible and, with

$$e = (K_\pi B + L_f) \epsilon_L + B \epsilon_P, \quad (52)$$

we have for $p_t = \pi(x_t)$

$$\frac{1}{N} \sum_{t=0}^{N-1} \|\mathrm{grad}_{\mathcal{M}} f(p_t)\|^2 \leq \frac{4(F(x_0) - \inf_{\mathcal{M}} f)}{\gamma N} + 9e^2 + \frac{4BL_\pi \epsilon_L}{\gamma}. \quad (53)$$

The term $4BL_\pi \epsilon_L / \gamma$ in equation 53 is the accumulated objective drift caused by landing error at successive iterations. The tangent-projector error appears through the quadratic term e^2 . Appendix H gives the perturbed descent argument for $F = f \circ \pi$.

Corollary 6.4 (Learned-score reduction to population biases). *Let $\mathcal{V} \subset \mathcal{T}_r$ contain every tangent-step query x_t and landing query z_t of the run under consideration. For each $y \in \mathcal{V}$, suppose a PSD rank- k field $A_{0,\sigma}(y)$ satisfies*

$$\mathrm{range}(A_{0,\sigma}(y)) = T_{\pi(y)}\mathcal{M}, \quad \lambda_k(A_{0,\sigma}(y)) \geq c > 0. \quad (54)$$

Define $Q_\sigma = \mathcal{P}_k(H_\sigma)$ and the population quantities

$$\beta_H(\sigma) = \sup_{y \in \mathcal{V}} \|H_\sigma(y) - A_{0,\sigma}(y)\|_{\mathrm{op}}, \quad (55)$$

$$\beta_L(\sigma) = \sup_{y \in \mathcal{V}} \|\overline{D}_{\sigma, Q_\sigma}(y) - \pi(y)\|, \quad (56)$$

$$g_\sigma = \inf_{y \in \mathcal{V}} \{\lambda_k(H_\sigma(y)) - \lambda_{k+1}(H_\sigma(y))\}. \quad (57)$$

Assume $g_\sigma > 0$, $\sup_{\mathcal{V}} \|D_\sigma(y) - y\| \leq R_D$, and a learned score satisfies uniformly on $\mathcal{V}$

$$\|s_\theta - s_\sigma\| \leq \epsilon_s, \quad \|\nabla s_\theta - \nabla s_\sigma\|_{\mathrm{op}} \leq \epsilon_J. \quad (58)$$

Set $D_\theta(y) = y + \sigma^2 s_\theta(y)$, $\hat{A}_\theta = \mathrm{sym}(I + \sigma^2 \nabla s_\theta)$, and $Q_\theta = \mathcal{P}_k(\hat{A}_\theta)$. If

$$\beta_H(\sigma) + \sigma^2 \epsilon_J < c/2, \quad \sigma^2 \epsilon_J < g_\sigma/2, \quad (59)$$

then the two oracle inequalities in equation 48 hold on $\mathcal{V}$ for the learned normal-only landing and projector, with

$$\epsilon_P \leq \frac{2(\beta_H(\sigma) + \sigma^2 \epsilon_J)}{c}, \quad (60)$$

$$\epsilon_L \leq \beta_L(\sigma) + \sigma^2 \epsilon_s + \frac{2R_D \sigma^2 \epsilon_J}{g_\sigma}. \quad (61)$$

Appendix H derives both oracle errors from the uniform score bounds. Provided $\mathcal{V}$ contains every x_t

and z_t used in the descent proof, Corollary 6.4 supplies the two oracle bounds required by Theorem 6.3; that theorem's remaining hypotheses are unchanged. Evaluated at the finite-type center alone, with $A_{0,\sigma} = U_p C_{q_m} U_p^\top$, Theorem 4.2 gives $\beta_H = O(\sigma^2)$; this is a pointwise consistency check rather than a visited set, since a nonstationary run does not remain at p . Theorem 4.6 supplies $\beta_H = O(\sigma)$, $\beta_L = O(\sigma^2)$, and $g_\sigma = \Omega(1)$ uniformly on an $O(\sigma)$ boundary layer. If $\epsilon_s, \epsilon_J = O(1)$ and γ is fixed independently of σ , then $\epsilon_P = O(\sigma)$ and $\epsilon_L = O(\sigma^2)$, and after the $O((\gamma N)^{-1})$ transient in equation 53 the residual bound on the averaged squared stationarity measure is $O(\sigma^2)$.

This last instantiation carries a consistency requirement that we state explicitly. One step of Theorem 6.3 moves x_t by $\gamma \|h_t\|$, which is $\Theta(\gamma)$ wherever $\|\nabla f\|$ is bounded away from zero, whereas the layer of Theorem 4.6 has ambient radius $O(\sigma)$. For a σ -independent step size the visited set therefore stays inside that layer only when the run is already ambient-near-stationary, $\|\nabla f(x_t)\| = O(\sigma/\gamma)$ throughout; Corollary 6.5 establishes precisely that configuration and gives a stronger conclusion there. Away from it, Theorem 6.3 remains valid with whatever β_H , β_L , and g_σ hold uniformly on the actual visited set. Obtaining these on a macroscopic region requires a positive density lower bound away from the zero, in the manner of the positive-density analyses cited in Section 2; we do not prove such a uniform statement here. On a fixed-width tube, Lemma 6.1 permits only $\epsilon_L = O(\sigma)$ in general, and hence only an $O(\sigma)$ residual for fixed γ . Statistical guarantees converting an L^2 denoising loss into the local uniform errors above remain open.

Corollary 6.5 (Local invariant SC-DRGD basin). *Assume Theorem 4.6 on*

$$\mathcal{V}_\sigma^+ = \mathcal{V}_\sigma(\bar{B}(0, R_+), \{b \in \mathbb{N}_p \mathcal{M} : \|b\| \leq B_+\}) \quad (62)$$

for fixed $R_+, B_+ > 0$. Let f be C^3 near p and suppose

$$\nabla f(p) = 0, \quad G_f = U_p^\top \nabla^2 f(p) U_p \succ 0. \quad (63)$$

Fix a σ -independent step size γ such that

$$\rho = \|I_k - \gamma G_f\|_{\text{op}} < 1; \quad \text{equivalently,} \quad 0 < \gamma < 2/\lambda_{\max}(G_f). \quad (64)$$

Let $\hat{D}_\sigma$ and symmetric $\hat{H}_\sigma$ be defined on $\mathcal{V}_\sigma^+$ and satisfy

$$\begin{aligned} \delta_D(\sigma) &= \sup_{\mathcal{V}_\sigma^+} \|\hat{D}_\sigma - D_\sigma\|, \\ \delta_H(\sigma) &= \sup_{\mathcal{V}_\sigma^+} \|\hat{H}_\sigma - H_\sigma\|_{\text{op}}, \quad \omega_\sigma = \sigma + \delta_H(\sigma) + \frac{\delta_D(\sigma)}{\sigma} \rightarrow 0. \end{aligned} \quad (65)$$

Define

$$\hat{Q}_\sigma = \mathcal{P}_k(\hat{H}_\sigma), \quad \hat{L}_\sigma(y) = y + (I - \hat{Q}_\sigma(y))(\hat{D}_\sigma(y) - y). \quad (66)$$

Choose $0 < R_0 < R < R_+$ and initialize

$$x_0 = F_p(\sigma a_0) + \sigma b_0, \quad \|a_0\| \leq R_0, \quad \|b_0\| \leq M_0 \omega_\sigma \quad (67)$$

for fixed M_0 . Then, for all sufficiently small σ , the iterates

$$z_t = x_t - \gamma \hat{Q}_\sigma(x_t) \nabla f(x_t), \quad x_{t+1} = \hat{L}_\sigma(z_t) \quad (68)$$

remain in $\mathcal{V}_\sigma^+$ for every t . They have unique fixed-chart coordinates $x_t = F_p(\sigma a_t) + \sigma b_t$, $b_t \in \mathbb{N}_p \mathcal{M}$, satisfying, for constants C, M independent of t and σ ,

$$a_{t+1} = (I_k - \gamma G_f) a_t + e_t, \quad \|e_t\| \leq C \omega_\sigma, \quad (69)$$

$$\|a_t\| \leq \rho^t \|a_0\| + \frac{C}{1 - \rho} \omega_\sigma, \quad \|b_t\| \leq M \omega_\sigma. \quad (70)$$

In particular, with $p_t = \pi(x_t)$,

$$\text{dist}(x_t, \mathcal{M}) \leq C \sigma \omega_\sigma, \quad (71)$$

$$\|x_t - p\| + \|\text{grad}_{\mathcal{M}} f(p_t)\| \leq C \sigma \left(\rho^t \|a_0\| + \frac{\omega_\sigma}{1 - \rho} \right). \quad (72)$$

Appendix H.3 proves the basin is invariant. For population oracles, $\omega_\sigma = \sigma$, so the asymptotic feasibility and stationarity floors are $O(\sigma^2)$. If $\hat{D}_\sigma = y + \sigma^2 s_\theta$ and $\hat{H}_\sigma = \text{sym}(I + \sigma^2 \nabla s_\theta)$, then $\delta_D \leq \sigma^2 \epsilon_s$ and $\delta_H \leq \sigma^2 \epsilon_J$. Bounded ϵ_s, ϵ_J preserve the same $O(\sigma^2)$ floors. The conclusion is local. Appendix H.4 explains why that hypothesis is needed. It establishes the visited-set condition near an ambient stationary point with positive-definite tangent Hessian, but does not show that an arbitrary trajectory enters or remains in a finite-type boundary layer.

Both steps matter

Calibrating the tangent update alone does not remove density-dependent scaling from the full update. Normal-only landing removes the raw denoiser’s tangent displacement; without it, the outer derivative restores the scaling removed from the inner step.

7 Population, finite-sample, and learned-score evidence

The experiments separate population, finite-sample, and learned-model effects. Deterministic quadrature evaluates population posterior moments, empirical posterior estimation adds finite-sample error, and the learned-model studies add model and Jacobian error, first in geometric estimation and then inside an optimizer. Appendix K records the protocol, seeds and numerical self-checks.

7.1 Spectral law on the circle

We begin with deterministic quadrature, which removes sampling and learning error. For $r \in \{0, 1, 2, 4\}$, let $\rho_r(\theta) \propto (1 - \cos \theta)^r$ and evaluate $H_\sigma(p)$ at $p = (1, 0)$. Here the vanishing order is $m = 2r$, so Theorem 4.2 predicts tangent eigenvalue $1 + 2r$ and normal eigenvalue zero. Figure 2 approaches 1, 3, 5, and 9 without fitted constants.

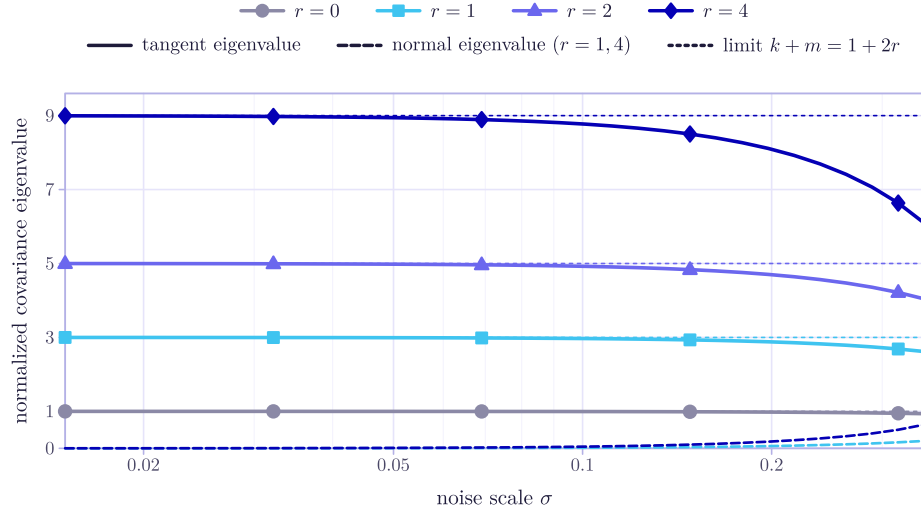


Figure 2 Population posterior-covariance spectra on S^1 , computed by deterministic quadrature. Solid curves are tangent eigenvalues for density zeros of order $m = 2r$; dashed curves show representative normal eigenvalues. The small-noise limits are the theorem’s $k + m = 1 + 2r$. The horizontal axis is the noise scale on a logarithmic scale, and the vertical axis is the normalized covariance eigenvalue.

7.2 Finite-sample range and order recovery

We draw samples from the same quadratic-zero circle and form the exact empirical Gaussian-KDE posterior

$$\hat{D}_N = \frac{\sum_i w_i X_i}{\sum_i w_i}, \quad \hat{H}_N = \frac{\sum_i w_i (X_i - \hat{D}_N)(X_i - \hat{D}_N)^\top}{\sigma^2 \sum_i w_i}, \quad w_i = e^{-\|X_i - p\|^2 / (2\sigma^2)}. \quad (73)$$

Population asymptotics do not indicate how estimation quality changes with realized local sample coverage. Figure 3 addresses this question with 48 independently seeded replications, $N \in \{20,000, 80,000, 320,000\}$, and five fixed noise levels. In panel (a), both $\hat{H}_N$ and its top-one projector are compared with finite-noise population references from deterministic quadrature, so deterministic finite- σ bias is not counted as sampling error. The horizontal coordinate is the realized effective sample size $(\sum_i w_i)^2 / \sum_i w_i^2$. Panel (b) asks a harder question: how often does $\text{round}(\lambda_1(\hat{H}_N) - 1)$ recover the true order $m = 2$? In this symmetric one-dimensional setting, the projector is already accurate at low effective sample size, whereas rounded-order recovery improves only with greater local coverage. We report this as a descriptive sweep, not as a concentration result.

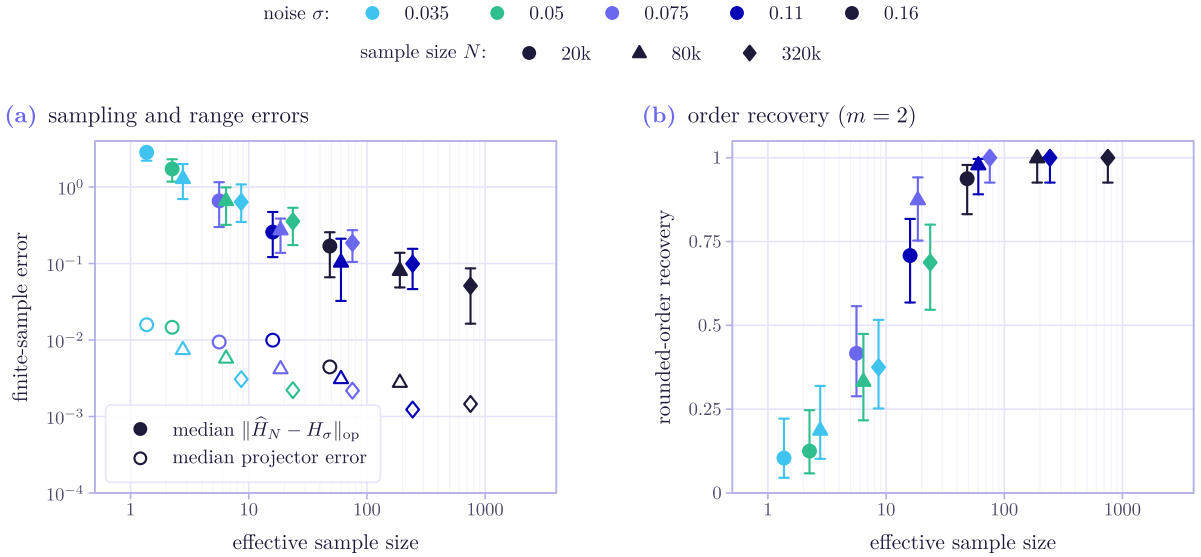


Figure 3 Finite-sample empirical-posterior experiment. Each point summarizes 48 replications. Color denotes σ , while circles, triangles, and diamonds denote $N = 20\text{k}, 80\text{k}, 320\text{k}$. (a) Filled points and bars are medians and interquartile ranges of $\|\hat{H}_N - H_\sigma\|_{\text{op}}$; open points are median top-one projector errors, both relative to finite-noise population references computed by deterministic quadrature. (b) Fraction recovering $m = 2$ after rounding $\lambda_1(\hat{H}_N) - 1$, with Wilson 95% binomial intervals. Both panels use a logarithmic horizontal axis giving the effective sample size; the left vertical axis is the finite-sample operator-norm error on a logarithmic scale, and the right vertical axis is the fraction of replications recovering the rounded order, on a linear scale.

7.3 Near-zero crossover

With $\rho_\delta(\theta) \propto \delta + 1 - \cos \theta$, Figure 4 plots the tangent eigenvalue against δ/σ^2 . Results from four noise levels collapse onto the parameter-free prediction equation 16. The experiment shows when a strictly positive density floor changes the score geometry at the current noise scale: the curves are organized by whether $\rho(p)$ is resolved at that scale, not by the binary distinction $\rho(p) = 0$ versus $\rho(p) > 0$.

7.4 Spatial boundary layer

To separate query displacement from the density-floor effect, we now hold the distribution fixed at $\rho(\theta) \propto 1 - \cos \theta$ and query at angle $\theta_\sigma = \sigma a$ from its zero. The tangent-normal cross block need not vanish at

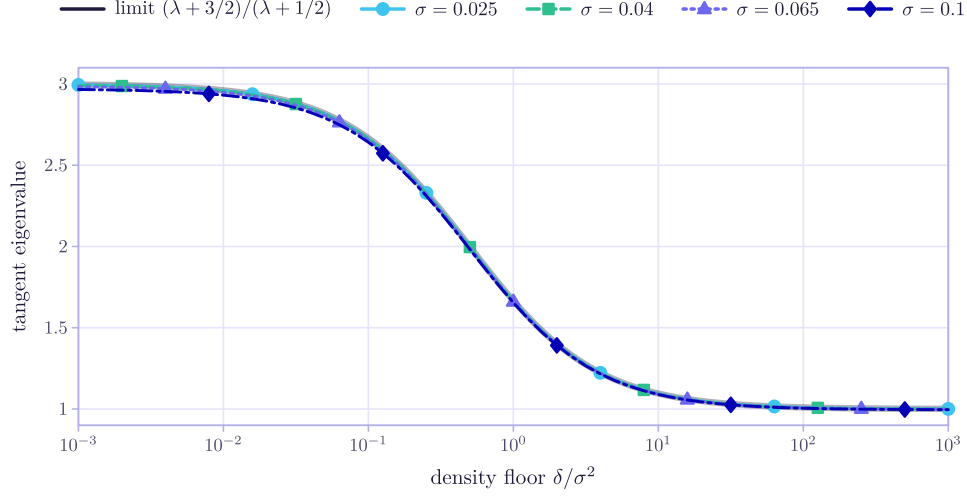


Figure 4 *Density-floor crossover. Deterministic quadrature at four noise levels collapses onto $(\lambda + 3/2)/(\lambda + 1/2)$, interpolating between the finite-zero value 3 and regular-density value 1. The horizontal axis is the density floor divided by σ^2 on a logarithmic scale, and the vertical axis is the tangent eigenvalue.*

finite σ , so Figure 5 reports the tangent block rather than labeling it an exact eigenvalue. It converges to the limit of Proposition 4.5, recorded in equation 20, but not monotonically: the quotient can fall below the regular value 1 before returning to it. This behavior comes from moving the Gaussian window through one distribution, rather than changing the distribution through a density floor. Both this experiment and the landing measurement below query on the manifold. Theorem 4.6 also covers queries displaced along the normal, and Table 4 checks that case separately: the tangent block is $C_a + O(\sigma)$ with a limit independent of the normal coordinate, while the landing bias is $O(\sigma^2)$ with a limit that depends on it.

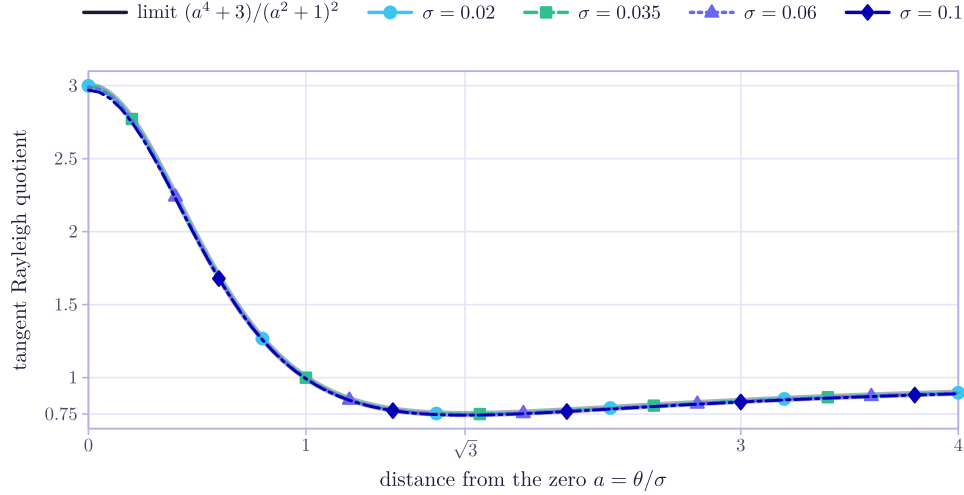


Figure 5 *Spatial boundary layer around a fixed quadratic zero. Population circle tangent Rayleigh quotients computed by deterministic quadrature at four noise levels converge to $(a^4 + 3)/(a^2 + 1)^2$. The nonmonotone undershoot is distinct from the monotone density-floor crossover in Figure 4. The horizontal axis is the query distance from the zero divided by the noise, and the vertical axis is the tangent Rayleigh quotient; both are linear.*

7.5 Raw versus normal-only landing

In the same fixed-distribution boundary layer, Figure 6 measures the denoiser’s value error. After division by σ , its raw tangent component converges to $2a/(a^2 + 1)$, which rules out a uniform $O(\sigma^2)$ retraction approximation there. Projecting the denoising displacement into the normal space removes that entire tangent term; the remaining error collapses after division by σ^2 .

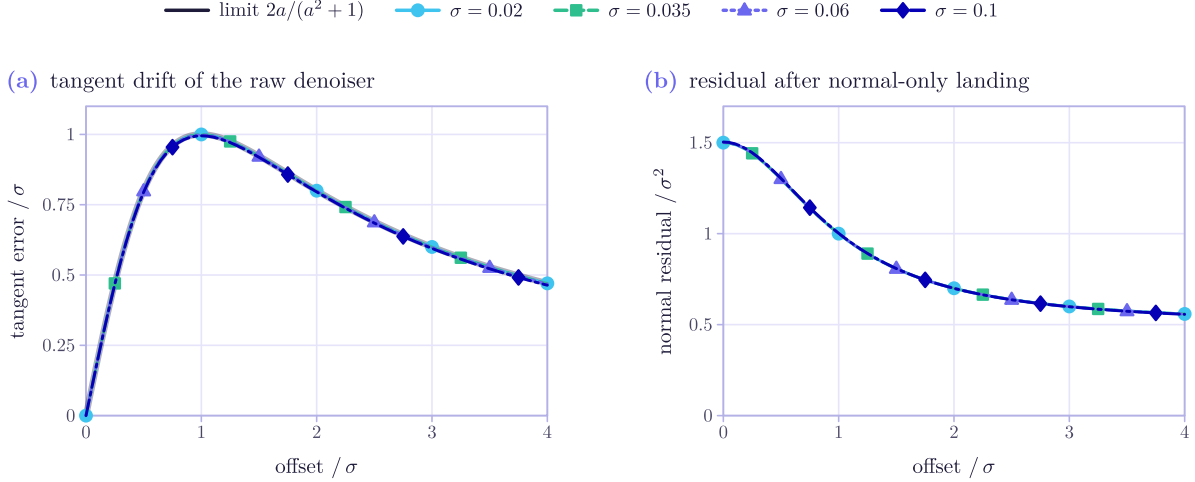


Figure 6 Effect of normal-only landing. Left: the raw denoiser has an $O(\sigma)$ tangent drift in the spatial boundary layer. Right: exact normal-only landing removes it and leaves an $O(\sigma^2)$ normal curvature residual. Both panels use a linear horizontal axis giving the offset divided by the noise; the vertical axes are the tangent error divided by σ on the left and the normal residual divided by σ^2 on the right, which are the normalizations that make the curves collapse.

7.6 Exact local stability, with method-specific step sizes

For even $m \geq 0$, choose the realizable jet $q_m(z) = z_1^m$ in $k = 2$, so that $C_q = \text{diag}(m + 1, 1)$, and take $f(z) = \|z\|^2/2$. Raw inner stepping followed by raw denoising has linearization $C_q(I - \gamma C_q)$. Calibrating only the inner step gives $C_q(I - \gamma I)$; raw additive preconditioning gives $I - \gamma C_q$; calibrating both the tangent step and normal-only landing gives $I - \gamma I$. Figure 7 places their spectral radii on their own step-size ranges. The last map has the density-independent stable interval $(0, 2)$. By comparison, the raw inner-and-outer interval has width $2/(m + 1)^2$, and the inner-only and raw-additive intervals each have width $2/(m + 1)$. All three baselines can be stabilized by tuning, but their admissible intervals contract as the vanishing order grows.

7.7 Learned circle geometry

Quadrature leaves open whether a learned Jacobian preserves the same separation between spectral range and scale. We train a two-hidden-layer noise-conditional residual MLP by denoising regression at three noise levels and compute $\hat{H}_\theta = \text{sym}(\nabla D_\theta)$ from its analytic input Jacobian. Both the uniform circle and the quadratic-zero circle use 12 independent seeds. Figure 8 shows biased eigenvalues, most visibly at the smallest noise where local coverage is hardest, alongside an accurate top eigenspace. The contrast is quantitative at the quadratic zero with $\sigma = 0.12$: replacing raw $\hat{H}_\theta$ by its hard top-one projector improves the median operator error by 1.483, with paired percentile-bootstrap interval $[1.443, 1.528]$, and improves all 12 seeds; a two-sided sign test over the seeds gives $p = 2^{-11} \approx 4.9 \times 10^{-4}$, which we regard as the primary evidence here, since a percentile bootstrap for a median with 12 replications is coarse and its interval largely reflects seed spread. On this benchmark, tangent-range recovery is more accurate than eigenvalue recovery.

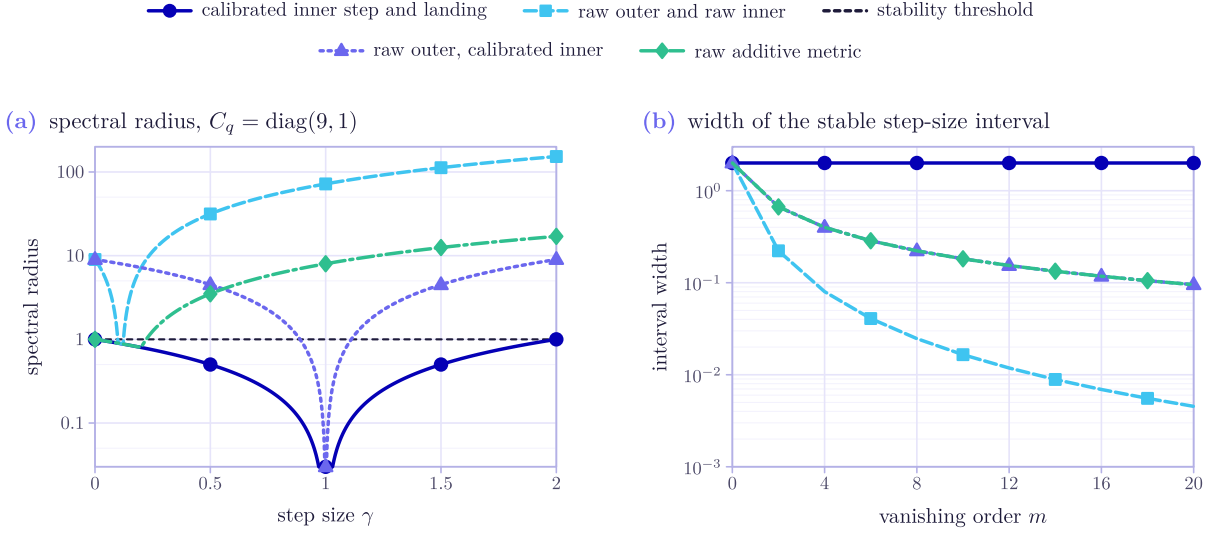


Figure 7 Exact local linear stability. (a) Spectral radius versus method-specific step size for $C_q = \text{diag}(9, 1)$; values below 0.03 are shown at the display floor, and the dashed horizontal line marks stability radius one. (b) Width of the exact stable step-size interval as the vanishing order increases. The inner-only and raw-additive widths coincide, although their intervals have different locations. In (a) the horizontal axis is the step size and the vertical axis is the spectral radius on a logarithmic scale; in (b) the horizontal axis is the vanishing order and the vertical axis is the interval width, also logarithmic.

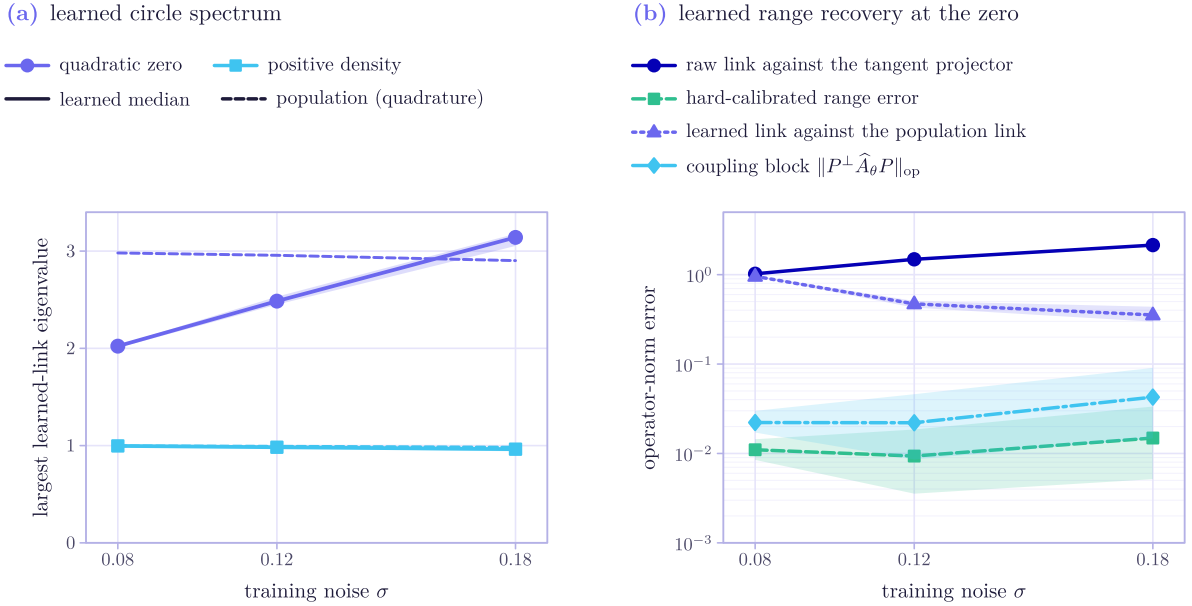


Figure 8 Twelve independently trained models per circle density (24 total); curves show medians and shaded bands show interquartile ranges. (a) Largest learned-link eigenvalue for positive density and for a quadratic zero; solid curves are the learned medians and dashed curves the population values computed by deterministic quadrature. (b) Four operator-norm errors at the zero: the raw link against the tangent projector, the hard-calibrated range error, the learned link against the finite-noise population link, and the tangent-normal coupling block $\|P^\perp \hat{A}_\theta P\|_{\text{op}}$ that replaces the full error in the residual bound equation 39. The learned spectrum is biased, while hard calibration gives substantially smaller operator error for the tangent range. In both panels the horizontal axis is the training noise level; the vertical axis is the tangent eigenvalue in (a) and the operator-norm error on a logarithmic scale in (b).

7.8 Rotated anisotropic finite-type plane

The circle has $k = 1, d = 2$, so it cannot show whether calibration survives multiple tangent directions, anisotropy, and a larger ambient space. We therefore use a fixed random orthonormal frame $U \in \mathbb{R}^{16 \times 2}$ and latent density

$$\rho(z) \propto (z_1^4 + 0.1z_2^4) \exp\left(-\frac{\|z\|^2}{2(0.35)^2}\right), \quad X = Uz. \quad (74)$$

The resulting order-four zero has unequal population tangent eigenvalues and a nonzero tangent-normal eigengap. Over the same 12 seeds, Figure 9 and Table 1 report both learned-link eigenvalues and projector error, keeping eigenvalue scale separate from tangent alignment. At $\sigma = 0.18$, hard top-two calibration lowers the median projector error from 1.314 to 0.016. The paired median improvement is 1.301, with bootstrap interval $[1.266, 1.603]$, and every seed improves, again $p = 2^{-11}$ by the sign test. The fixed soft filter with $\tau = 0.25$ reduces raw scaling error, but it is not an orthogonal projector.

Table 2 identifies which guarantee is actually in force. At this noise level the learned link is $\eta = 1.355$ from the population link in operator norm, against a threshold $c_q/2 = 0.539$, so the hypothesis of Proposition 5.1 fails by a factor of about 2.5 and the bound equation 37 is vacuous; the two circle settings satisfy it. The accuracy observed on the plane is therefore not an instance of that bound. It is explained by the residual form equation 39: the tangent-normal coupling block measures only $\|P^\perp \hat{A}_\theta P\|_{\text{op}} = 0.0374$, a factor of 36 below the full operator-norm error, and the computed separation $\delta(\hat{A}_\theta) = 0.840$ stays positive, so equation 39 certifies 0.0422 against an observed 0.0160. The learned error at a finite-type zero is thus concentrated in eigenvalue scale rather than in the eigenvectors, which is exactly the structure the operator-norm bound cannot exploit and the residual bound can. On the two circles the residual bound is attained to three decimal places.

The population spectrum at these noise levels is itself pre-asymptotic, which bounds what this experiment can test. Because the latent density in equation 74 is a degree-four tilt of an isotropic Gaussian envelope of scale $s = 0.35$, homogeneity of q_4 makes the tilt scale free and gives $H_\sigma(p) = \frac{s^2}{s^2 + \sigma^2} UC_q U^\top$ exactly, with $C_q = \text{diag}(4.636, 1.364)$ and $\text{tr } C_q = k + m = 6$. At $\sigma = 0.12, 0.18$, and 0.25 the scalar factor is 0.895, 0.791, and 0.662, so the population traces are 5.37, 4.75, and 3.97 rather than 6. Rounding the trace statistic would return the correct order only for $\sigma < 0.106$, below every noise level trained here. Comparisons on this problem are therefore made against the exact finite-noise population operator rather than against its limit, and the experiment tests anisotropic range recovery and calibration rather than the asymptotic trace law; the circle experiments test the trace law directly.

Table 1 *Learned geometry at the middle training noise. Entries are medians with interquartile ranges over 12 independent seeds. Errors are operator norms relative to the true tangent projector. The last column thresholds learned eigenvalues at 1/2; it is a diagnostic, not an adaptive-rank guarantee.*

Setting	Population spectrum	Largest learned-link spectrum [IQR]	Raw error [IQR]	Hard error [IQR]	Rank success
Positive circle, 0.12	0.993	0.983 [0.974, 0.996]	0.046 [0.036, 0.066]	0.008 [0.006, 0.021]	12/12
Zero circle, 0.12	2.957	2.485 [2.449, 2.532]	1.485 [1.449, 1.532]	0.009 [0.004, 0.019]	12/12
Rotated plane, 0.18	3.667	2.314 [2.282, 2.619]	1.314 [1.282, 1.619]	0.016 [0.014, 0.017]	12/12
	1.078	0.816 [0.808, 0.965]			

7.9 End-to-end learned optimization ablation

The final experiment places the learned geometric estimates inside SC-DRGD on the quadratic-zero circle and rotated plane. We compare four variants: oracle Riemannian gradient descent, a raw link followed by the raw denoiser, hard calibration only in the inner step, and hard calibration in both the step and normal-only learned landing. Figure 10 gives each method its own step-size sweep. On the circle, the reported gap is the projected objective $F(x_t) - F(p) = f(\pi(x_t)) - f(p)$ from Theorem 6.3, with feasibility reported separately. Both objectives have ambient gradient zero at the target, as required by Corollary 6.5.

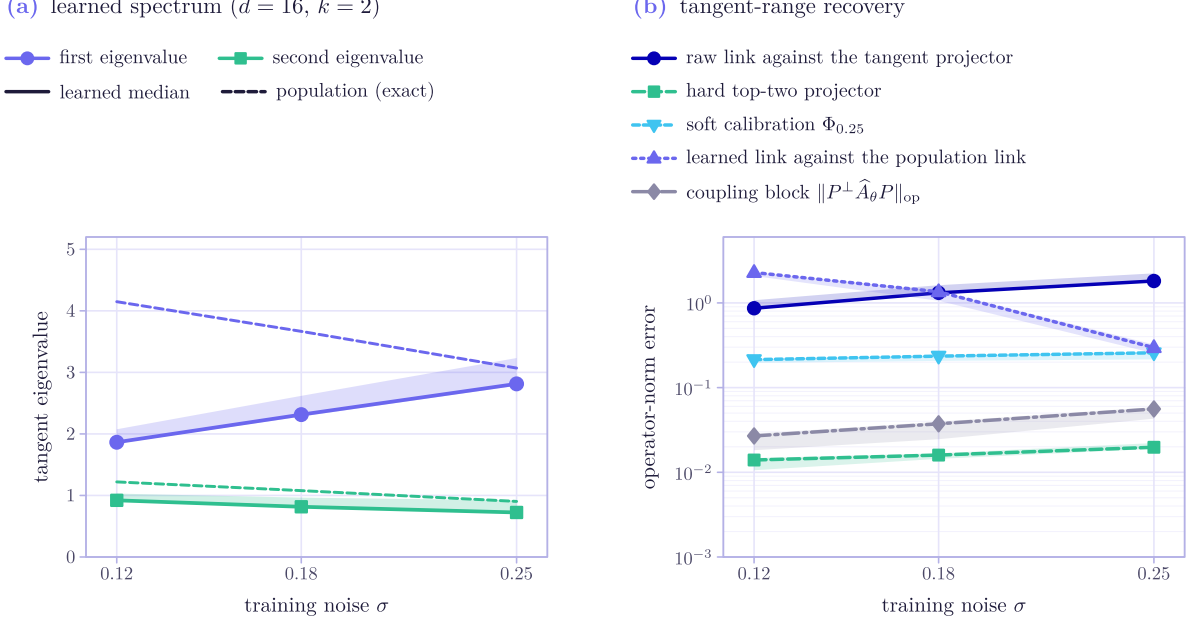


Figure 9 Twelve independently trained denoisers for the rotated $d = 16, k = 2$ order-four model. (a) The two largest learned-link eigenvalues, solid, against the exact finite-noise population tangent eigenvalues, dashed. (b) Five operator-norm errors: the raw link, the hard top-two projector, the soft calibration $\Phi_{0.25}$, the learned link against the population link, and the tangent-normal coupling block $\|P^\perp \hat{A}_\theta P\|_{\text{op}}$ of the residual bound equation 39. Bands are interquartile ranges. In both panels the horizontal axis is the training noise level; the vertical axis is the tangent eigenvalue in (a) and the operator-norm error on a logarithmic scale in (b).

Table 2 Which calibration guarantee is in force at the middle training noise. Medians over 12 seeds. Here $\eta = \|\hat{A}_\theta - H_\sigma\|_{\text{op}}$ is the quantity constrained by Proposition 5.1, whose operator-norm bound equation 37 requires $\eta < c_q/2$ and is reported as *vacuous* when it exceeds one. The residual bound equation 39 replaces η by the tangent-normal coupling block $\|P^\perp \hat{A}_\theta P\|_{\text{op}}$ and divides by the computed separation $\delta(\hat{A}_\theta)$ of equation 38. The plane fails the operator-norm hypothesis while the residual bound remains informative there.

Setting	η	$c_q/2$	Bound equation 37	$\ P^\perp \hat{A}_\theta P\ _{\text{op}}$	$\delta(\hat{A}_\theta)$	Bound equation 39	Observed hard error
Positive circle, 0.12	0.042	0.496	0.084	0.0077	0.948	0.0081	0.0081
Zero circle, 0.12	0.471	1.478	0.319	0.0221	2.461	0.0094	0.0094
Rotated plane, 0.18	1.355	0.539	vacuous	0.0374	0.840	0.0422	0.0160

Using the descriptive criterion of median final gap below 10^{-3} and median feasibility error below 0.1, SC-DRGD succeeds at 19/20 tested circle step sizes and 23/24 plane step sizes. The corresponding counts are 0/20 and 0/24 for the raw method, and 7/20 and 12/24 for the inner-only ablation. Two checks indicate that these counts are not artifacts of aggregating by the median or of the particular threshold. Per-seed success rates, averaged over the tested step sizes, are 95.0%, 35.8%, and 12.1% on the circle and 95.8%, 50.3%, and 3.8% on the plane, in the same order. Loosening the gap threshold to 10^{-2} gives circle counts 20/20, 9/20, and 7/20, and tightening it to 10^{-4} gives 19/20, 4/20, and 0/20; the ordering is unchanged throughout.

The counts track the exact linear stability intervals of Section 7.6 more closely than the criterion alone suggests. At $\sigma = 0.12$ the population tangent eigenvalue is $c = 2.957$, so those intervals are $((c - 1)/c^2, (c + 1)/c^2) = (0.224, 0.453)$ for the raw additive update, $(1 - 1/c, 1 + 1/c) = (0.662, 1.338)$ for inner-only calibration, and $(0, 2)$ after full calibration. The circle grid places 5, 7, and 20 of its 20 step sizes inside these intervals, against observed counts of 0, 7, and 19. Inner-only calibration therefore agrees with the linearization exactly,

and full calibration agrees at every step size but one.

The residual disagreements have different causes, which the criterion does not separate. The single calibrated failure is the smallest step size, $\gamma = 0.05$, in both sweeps; there the median gap is still halving over the final fifth of the run, so the horizon of 35 and 30 steps is binding rather than stability. The raw method instead remains stable inside its own interval while converging to a biased point: at its five in-interval step sizes the median final gap lies between 5.4×10^{-2} and 8.2×10^{-2} , above the criterion at every threshold tested. Its 0/20 therefore records density-dependent bias rather than divergence, which is the effect calibration is intended to remove. These grid counts describe this benchmark and are not universal stability guarantees.

Because each denoiser is noise conditional and trained at three levels, the same models can be run at every trained level without retraining, which tests the noise dependence that Corollary 6.5 predicts. Taking for each noise level the smallest median final value over the step-size grid, the feasibility floor of SC-DRGD is 6.7×10^{-3} , 1.7×10^{-2} , and 6.1×10^{-2} at $\sigma = 0.08, 0.12, 0.18$ on the circle, and 1.2×10^{-3} , 3.5×10^{-3} , and 1.2×10^{-2} at $\sigma = 0.12, 0.18, 0.25$ on the plane. Fitting a power law over each range gives exponents 2.7 and 3.2. The theory predicts an $O(\sigma^2)$ floor, which is an upper bound, so decay at least this fast is consistent with it; the ratios feasibility/ σ^2 stay within a factor of two over the circle range and a factor of 2.4 over the plane range. The objective gap is not a useful floor diagnostic on these problems, since the calibrated iteration contracts it to 10^{-10} or below within the horizon and the residual is set by the number of iterations rather than by σ .

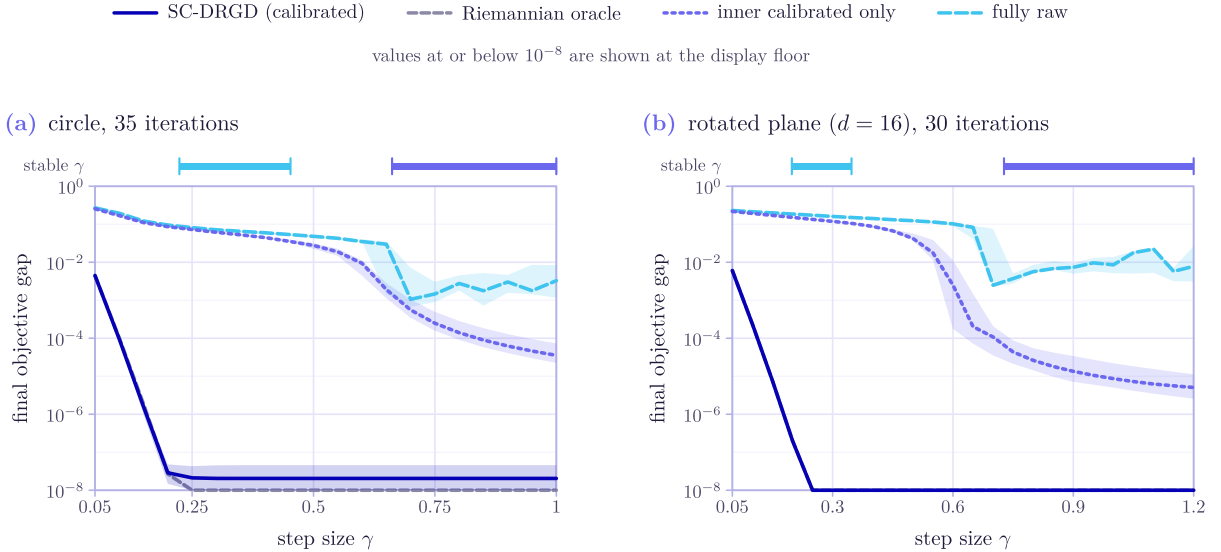


Figure 10 End-to-end learned optimization over method-specific step-size sweeps. Curves are median final objective gaps and bands are interquartile ranges over 12 independently trained models per problem; dashed gray is oracle Riemannian gradient descent. Left: quadratic-zero circle after 35 steps. Right: rotated order-four plane after 30 steps. Values at or below 10^{-8} are shown at the display floor. The horizontal axis is the step size and the vertical axis is the final objective gap on a logarithmic scale. Shading inside each panel is the interquartile range and nothing else. The two bars above each panel, marked stable γ , are annotations rather than data: they give the exact linear stability intervals of Section 7.6 at the plotted noise level, cyan for the fully raw update and violet for inner-only calibration. The fully calibrated interval is $(0, 2)$ and covers each panel entirely, so it is not drawn. The inner-only curve descends where its bar begins, whereas the raw update is stable throughout its own bar yet converges to a biased point there. Stability and accuracy therefore separate for the raw update, which is what calibration removes.

8 Limitations and scope

Limits of the assumptions. The finite-type assumption is substantive. For an infinitely flat one-dimensional density such as $\rho(u) \propto e^{-1/u^2}$, Laplace’s method gives typical $u^2 \propto \sigma$ and hence $\sigma^{-2} \text{Cov}(X \mid Y = p) =$

$\Theta(\sigma^{-1})$. For this example, the tangent metric has no finite limit. Definition 4.1 is also narrower than the argument requires in a second way: it asks for a polynomial leading jet, so effective densities vanishing at a finite but non-even order, such as $\|z\|^3$ or $\|z\|^{1/2}$, are excluded even though the limit and the trace law hold for them. Extending the uniform eigengap of Lemma D.1 beyond polynomial jets is open. Boundaries and singular supports pose a different problem, since their local geometry is described by tangent cones rather than tangent spaces. Appendix I works that case out in full.

Theorem 4.2 is also local. A full learning guarantee would need uniform control of the population biases β_H, β_L , the eigengap, and the optimizer’s visited set. It would further have to turn denoising training risk, typically an L^2 quantity, into the local score-Jacobian errors required by Corollary 6.4.

Coverage of the experiments. The learned studies are controlled synthetic tests. They provide no evidence yet for a large pretrained diffusion model or real data. Every experiment has ambient dimension $d \leq 16$ and intrinsic dimension $k \leq 2$, and the population and finite-sample studies are $d = 2, k = 1$ throughout. Two jet families are tested, the radial θ^{2r} on the circle and one anisotropic quartic on the plane, and that plane is affine, so every curvature-dependent term in Theorem 4.6 vanishes on it. Appendix J partly closes this by placing a finite-type zero on a spiral whose curvature varies by a factor of five along the arc, where the limit is still $k + m$ and the $O(\sigma^2)$ remainder is found to scale as the square of the curvature κ ; that check is nonetheless still $k = 1$, so curvature combined with several tangent directions remains untested. A single random frame U was drawn; because the construction is equivariant under rotation this is not a loss of generality, but it does mean frame-to-frame variability is not sampled. Hard calibration uses the known value of k , while the fixed threshold is reported only as a rank diagnostic. Because the largest learned-link eigenvalues can remain materially biased even when the eigenspace is accurate, the trace law is not a reliable empirical estimator of vanishing order without further sample and Jacobian control. Query-focused resampling depends only on the noisy input and therefore preserves the population conditional mean, but it remains part of the benchmark design. Larger architectures, adaptive eigengap selection, and real-data validation are open empirical questions.

The population claims can be checked directly against the q_m -tilted Gaussian spectrum, trace law, and density-floor crossover. For a learned Jacobian, Proposition 5.1 guarantees stable calibration when estimation error is sufficiently smaller than the tangent-normal eigengap. It makes no converse claim once the error becomes comparable to the gap. The per-seed tables include failures as well as successes, allowing the empirical behavior to be inspected without relying on plotted medians.

Practical scope. Low-probability regions can change the scale of score-derived geometry, including when an optimizer leaves high-density modes. Calibration addresses that geometric effect. It does not certify that a learned manifold represents valid or safe designs; domain constraints and downstream validation remain necessary.

9 Conclusion

At a regular positive-density point, the denoiser-Jacobian limit is the tangent projector. At a finite-type zero, its range and nonzero eigenvalue scale have different interpretations. The limit is $UC_{q_m}U^\top$: its range is the tangent space, while its positive eigenvalues depend on the leading density jet and sum to $k + m$. The density-floor calculation and the $O(\sigma)$ moving-query analysis give explicit transition laws and identify where this distinction persists.

When population bias and learned-Jacobian error are below the tangent-normal eigengap, spectral projection estimates the tangent projector with controlled error. The optimization analysis treats calibration in both the tangent update and the landing map, and the experiments compare this choice with partially calibrated and raw variants. Population quadrature, finite-sample posterior estimation, and learned-denoiser experiments exhibit the predicted separation between tangent range and eigenvalue scale. In the finite-type small-noise limit, the nonzero eigenspace retains tangent geometry, whereas the raw eigenvalue scale remains density dependent.

References

- [1] P.-A. Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.
- [2] 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.
- [3] Jacob Bamberger, Adam Gosztolai, Pierre Vandergheynst, Michael M. Bronstein, and Iolo Jones. Riemannian metric matching for scalable geometric modeling of distributions. In *International Conference on Machine Learning*, 2026. URL <https://arxiv.org/abs/2606.14334>. Oral.
- [4] Tyrus Berry and John Harlim. Variable bandwidth diffusion kernels. *Applied and Computational Harmonic Analysis*, 40(1):68–96, 2016.
- [5] Nicolas Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023.
- [6] Junhao Chen, Ruowei Li, and Zhigang Yao. Estimation of Riemannian quantities from noisy data via density derivatives. *arXiv preprint arXiv:2603.27171*, 2026. URL <https://arxiv.org/abs/2603.27171>.
- [7] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.
- [8] Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.
- [9] Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022. URL <https://arxiv.org/abs/2208.05314>.
- [10] Willem Diepeveen and Melanie Weber. Iso-Riemannian optimization on learned data manifolds. *arXiv preprint arXiv:2510.21033*, 2025. URL <https://arxiv.org/abs/2510.21033>.
- [11] 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*, volume 267 of *Proceedings of Machine Learning Research*, pages 13746–13773, 2025. URL <https://proceedings.mlr.press/v267/diepeveen25a.html>.
- [12] Bradley Efron. Tweedie’s formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.
- [13] Zahra Kadkhodaie, Florentin Guth, Eero P. Simoncelli, and Stéphane Mallat. Generalization in diffusion models arises from geometry-adaptive harmonic representations. In *International Conference on Learning Representations*, 2024. URL <https://arxiv.org/abs/2310.02557>.
- [14] Hamidreza Kamkari, Brendan Leigh Ross, Rasa Hosseinzadeh, Jesse C. Cresswell, and Gabriel Loaiza-Ganem. A geometric view of data complexity: Efficient local intrinsic dimension estimation with diffusion models. In *Advances in Neural Information Processing Systems*, 2024. URL <https://arxiv.org/abs/2406.03537>. Spotlight.
- [15] 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. URL <https://arxiv.org/abs/2509.23357>.
- [16] Kin Kwan Leung, Rasa Hosseinzadeh, and Gabriel Loaiza-Ganem. On convolutions, intrinsic dimension, and diffusion models. *Transactions on Machine Learning Research*, 2025. URL <https://openreview.net/forum?id=xSzBf1te4s>.

- [17] 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://arxiv.org/abs/2509.24912>.
- [18] Hila Manor and Tomer Michaeli. On the posterior distribution in denoising: Application to uncertainty quantification. In *International Conference on Learning Representations*, 2024. URL <https://arxiv.org/abs/2309.13598>.
- [19] Genki Osada. Local Hessian spectral filtering for robust intrinsic dimension estimation. In *International Conference on Machine Learning*, 2026. URL <https://arxiv.org/abs/2605.01221>.
- [20] Jakiw Pidstrigach. Score-based generative models detect manifolds. In *Advances in Neural Information Processing Systems*, 2022. URL <https://arxiv.org/abs/2206.01018>.
- [21] Shinnosuke Saito and Takashi Matsubara. Be tangential to manifold: Discovering Riemannian metric for diffusion models. *arXiv preprint arXiv:2510.05509*, 2025. URL <https://arxiv.org/abs/2510.05509>.
- [22] Shirin Shoushtari, Yi Wang, Xiao Shi, M. Salman Asif, and Ulugbek S. Kamilov. EigenScore: OOD detection using posterior covariance in diffusion models. In *International Conference on Learning Representations*, 2026. URL <https://arxiv.org/abs/2510.07206>.
- [23] 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*, pages 46412–46440, 2024. URL <https://proceedings.mlr.press/v235/stanczuk24a.html>.
- [24] A. N. Varchenko. Newton polyhedra and estimation of oscillating integrals. *Functional Analysis and Its Applications*, 10(3):175–196, 1976.
- [25] 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://arxiv.org/abs/2410.05898>.
- [26] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.
- [27] Eric Yeats, Aaron Jacobson, Darryl Hannan, Yiran Jia, Timothy Doster, Henry Kvinge, and Scott Mahan. A connection between score matching and local intrinsic dimension. *arXiv preprint arXiv:2510.12975*, 2025. URL <https://arxiv.org/abs/2510.12975>. 3rd SPIGM Workshop at NeurIPS 2025.
- [28] Zixuan Zhang, Kaixuan Huang, Tuo Zhao, Mengdi Wang, and Minshuo Chen. Diffusion model for manifold data: Score decomposition, curvature, and statistical complexity. *arXiv preprint arXiv:2603.20645*, 2026. URL <https://arxiv.org/abs/2603.20645>.

A Notation

Recurring symbols are collected here in order of first appearance. Each is also defined at its first use in the text. Throughout, $\|\cdot\|$ is the Euclidean norm on vectors and $\|\cdot\|_{\text{op}}$ the operator norm on matrices.

Manifold and ambient geometry

$\mathcal{M}, k, d$	Embedded C^3 manifold without boundary, its dimension, and the ambient dimension, with $1 \leq k < d$.
$T_p\mathcal{M}, N_p\mathcal{M}$	Tangent and normal spaces at $p \in \mathcal{M}$.

$U_p, U_\perp$	Matrices with orthonormal columns spanning $T_p\mathcal{M}$ and $N_p\mathcal{M}$; $P = U_p U_p^\top$ and $P^\perp = I_d - P$.
F_p, g_p, J_p	Normal graph chart at p , its normal component, and the area Jacobian $\sqrt{\det(DF_p^\top DF_p)}$.
$\Pi_p, S_b, \mathcal{B}_a$	Second fundamental form; shape operator in the normal direction b ; the map $v \mapsto \Pi_p[a, v]$.
κ	Curvature of a one-dimensional manifold, used only in Appendix J.
$\mathcal{T}_r, \pi$	Reach tube $\{x : \text{dist}(x, \mathcal{M}) < r\}$ and nearest-point projection onto $\mathcal{M}$.
<i>Density and its leading jet</i>	
μ, ρ	Law of the clean variable X and its density with respect to Riemannian volume.
a_p	Effective chart density $\rho(F_p(z))J_p(z)$, whose jets drive every limit.
q_m, m	First nonzero homogeneous jet of a_p and its degree, the vanishing order.
$\mathcal{Z}_q, C_{q_m}$	Normalizer $\int q_m e^{-\ u\ ^2/2} du$ and the covariance equation 6 of the q_m -tilted Gaussian.
<i>Noise, score and denoiser</i>	
$\sigma, Y_\sigma, p_\sigma$	Noise scale, the observation $X + \sigma G$, and its density.
$s_\sigma, D_\sigma, H_\sigma$	Score $\nabla \log p_\sigma$, denoiser $y + \sigma^2 s_\sigma$, and link Hessian $\nabla D_\sigma = \sigma^{-2} \text{Cov}(X \mid Y_\sigma = y)$.
$s_\theta, \hat{A}_\theta$	Learned score and its symmetrized link Jacobian.
<i>Spectral calibration</i>	
A_0, c_q	Limiting operator $U_p C_{q_m} U_p^\top$ and its smallest nonzero eigenvalue $\lambda_{\min}(C_{q_m})$.
$\mathcal{P}_k, Q_\sigma$	Orthogonal projector onto a top- k eigenspace, and $\mathcal{P}_k(H_\sigma)$.
$\eta, \delta(\hat{A})$	Operator-norm estimation error $\ \hat{A} - A_0\ _{\text{op}}$ and the computed separation equation 38.
$\vartheta_{\text{sp}}, \tau, \Phi_\tau$	Hard spectral cutoff, soft filter strength, and the filter $\hat{A}(\hat{A} + \tau I)^{-1}$.
<i>Boundary layer at a finite-type point</i>	
a, b	Rescaled tangential and normal coordinates of a query, so the query is $F_p(\sigma a) + \sigma b$.
$\nu_{q,a}, C_a, \mu_a$	Shifted tilted law with density $\propto q_m(a + u)e^{-\ u\ ^2/2}$, its covariance and its mean.
Γ_a, Λ_a	Normal-valued covariance equation 24 and its calibrated version $\Gamma_a C_a^{-1}$.
K_T, K_N, c_0	Compact sets of admissible a and b , and $\inf_{a \in K_T} \lambda_{\min}(C_a)$.
<i>Optimization</i>	
f, F, γ	Ambient objective, its pullback $f \circ \pi$, and the step size.
$L, Q, \bar{D}_Q$	Landing map, tangent projector oracle, and the normal-only landing equation 44.
ϵ_P, ϵ_L	Oracle errors of Assumption 6.2, for the projector and the landing map.
$\beta_H, \beta_L, g_\sigma$	Population Jacobian bias, population landing bias, and tangent-normal eigengap, all defined in Corollary 6.4.
$\epsilon_s, \epsilon_J, R_D$	Uniform score error, score-Jacobian error, and a bound on $\ D_\sigma(y) - y\ $.
$\mathcal{V}$	Query set containing every iterate x_t and intermediate point z_t .
$B, L_f, L_F, L_\pi, K_\pi$	Gradient bound and Lipschitz constants of $\nabla f, \nabla F, \pi$ and $d\pi$.
G_f	Tangent block $U_p^\top \nabla^2 f(p) U_p$ of the ambient Hessian.

B Exact posterior identities

For Proposition 3.1, let

$$\varphi_\sigma(y - x) = (2\pi\sigma^2)^{-d/2} \exp\left(-\frac{\|y - x\|^2}{2\sigma^2}\right).$$

The Gaussian factor and all of its derivatives are bounded as functions of x for fixed $\sigma > 0$. Since μ is finite, differentiation under the integral is justified by dominated convergence. We obtain

$$\nabla p_\sigma(y) = \int \frac{x - y}{\sigma^2} \varphi_\sigma(y - x) \mu(dx), \quad (75)$$

$$\nabla \log p_\sigma(y) = \frac{\mathbb{E}[X \mid Y_\sigma = y] - y}{\sigma^2}. \quad (76)$$

This is the posterior-mean identity.

To differentiate the conditional mean, write $m(y) = \mathbb{E}[X \mid Y_\sigma = y]$. In coordinates,

$$\partial_{y_j} m_i(y) = \frac{1}{\sigma^2} (\mathbb{E}[X_i X_j \mid Y_\sigma = y] - \mathbb{E}[X_i \mid Y_\sigma = y] \mathbb{E}[X_j \mid Y_\sigma = y]). \quad (77)$$

Therefore $\nabla m(y) = \sigma^{-2} \text{Cov}(X \mid Y_\sigma = y)$. Since $m = D_\sigma$, differentiating equation 2 proves the second identity and symmetry of H_σ .

C A localized Laplace expansion

In the normal graph equation 4, assume

$$g_p(z) = \frac{1}{2} \Pi_p[z, z] + R_g(z), \quad \|R_g(z)\| \leq C\|z\|^3, \quad (78)$$

$$a_p(z) = q_m(z) + q_{m+1}(z) + R_a(z), \quad |R_a(z)| \leq C\|z\|^{m+2}. \quad (79)$$

Here $\Pi_p : \mathbb{R}^k \times \mathbb{R}^k \rightarrow N_p \mathcal{M}$ is the second fundamental form in the chosen frame. The normal graph is orthogonal to $U_p z$, so

$$\|F_p(z) - p\|^2 = \|z\|^2 + \|g_p(z)\|^2 = \|z\|^2 + O(\|z\|^4). \quad (80)$$

Choose the chart so that $\mathcal{M} \cap B(p, \eta)$ is contained in its image. The contribution from $\mathcal{M} \setminus B(p, \eta)$ to a posterior moment of order $s \in \{0, 1, 2\}$ is at most

$$O\left(e^{-\eta^2/(4\sigma^2)}\right), \quad (81)$$

after absorbing a bounded polynomial factor into the exponential. This term is smaller than every power of σ .

Inside the chart set $z = \sigma u$. For $\|u\| < \eta/\sigma$,

$$\sigma^{-m} a_p(\sigma u) = q_m(u) + \sigma q_{m+1}(u) + O(\sigma^2 \|u\|^{m+2}), \quad (82)$$

$$\exp\left(-\frac{\|g_p(\sigma u)\|^2}{2\sigma^2}\right) = 1 + O(\sigma^2 \|u\|^4). \quad (83)$$

The expansion equation 83 is to be read on bounded sets of u ; on the whole chart, where $\|u\|$ may reach η/σ , the elementary bound $|e^{-t} - 1| \leq t$ for $t \geq 0$ gives instead $0 \leq 1 - \exp(-\|g_p(\sigma u)\|^2/2\sigma^2) \leq C\sigma^2 \|u\|^4$ with the same constant, and $\|g_p(\sigma u)\| \leq C\sigma^2 \|u\|^2$ makes the exponent nonnegative, so the factor never exceeds one. On a smaller fixed chart, these remainders are dominated by a polynomial in $\|u\|$ times $e^{-c\|u\|^2}$. The annulus between that smaller chart and the original one is exponentially negligible. Dominated convergence therefore permits termwise integration over $\mathbb{R}^k$.

Let

$$\mathcal{Z}_q = \int_{\mathbb{R}^k} q_m(u) e^{-\|u\|^2/2} du = (2\pi)^{k/2} \mathbb{E}[q_m(Z)], \quad Z \sim \mathcal{N}(0, I_k).$$

The unnormalized $\mathcal{Z}_q$ and the expectation $\mathbb{E}[q_m(Z)]$ used in equation 6 therefore differ by the Gaussian normalizing constant, which cancels in every ratio below. Because q_m is even and q_{m+1} is odd,

$$\int_{\mathcal{M}} e^{-\|x-p\|^2/(2\sigma^2)} \mu(dx) = \sigma^{k+m} (\mathcal{Z}_q + O(\sigma^2)). \quad (84)$$

The order- σ term vanishes by parity. Multiplication by the ambient Gaussian normalization gives

$$p_\sigma(p) = (2\pi)^{-d/2} \mathcal{Z}_q \sigma^{k+m-d} (1 + O(\sigma^2)). \quad (85)$$

The smoothed-density power law therefore contains the vanishing order.

D Proof of the finite-type theorem

D.1 Parity and positivity

For every fixed u ,

$$q_m(u) = \lim_{t \downarrow 0} \frac{a_p(tu)}{t^m} \geq 0.$$

If m were odd, homogeneity would give $q_m(-u) = -q_m(u)$. Nonnegativity in both directions would imply $q_m \equiv 0$, a contradiction. Hence m is even, q_m is even, and it is positive on some nonempty open set. In particular $\mathcal{Z}_q > 0$.

D.2 Posterior mean

Let N_1 denote the unnormalized first centered moment about p . Its tangent component is

$$\begin{aligned} P_{T_p \mathcal{M}} N_1 &= U_p \sigma^{k+m+1} \int_{\mathbb{R}^k} u (q_m(u) + \sigma q_{m+1}(u) + O(\sigma^2 \text{poly}(u))) e^{-\|u\|^2/2} du \\ &= U_p \sigma^{k+m+2} \int_{\mathbb{R}^k} u q_{m+1}(u) e^{-\|u\|^2/2} du + O(\sigma^{k+m+3}). \end{aligned} \quad (86)$$

The leading integral vanishes because $u q_m(u)$ is odd. From equation 78, the normal component is

$$P_{N_p \mathcal{M}} N_1 = \frac{\sigma^{k+m+2}}{2} \int_{\mathbb{R}^k} \Pi_p[u, u] q_m(u) e^{-\|u\|^2/2} du + O(\sigma^{k+m+3}). \quad (87)$$

Dividing equation 86 and equation 87 by equation 84 yields equation 7, with

$$b_{p,q} = U_p \frac{\int u q_{m+1}(u) e^{-\|u\|^2/2} du}{\mathcal{Z}_q} + \frac{1}{2} \frac{\int \Pi_p[u, u] q_m(u) e^{-\|u\|^2/2} du}{\mathcal{Z}_q}. \quad (88)$$

D.3 Posterior covariance

Write N_0 for the normalizer equation 84 and N_2 for the unnormalized second moment about p ,

$$N_0 = \int_{\mathcal{M}} e^{-\|x-p\|^2/(2\sigma^2)} \mu(dx), \quad N_2 = \int_{\mathcal{M}} (x-p)(x-p)^\top e^{-\|x-p\|^2/(2\sigma^2)} \mu(dx), \quad (89)$$

so that $\mathbb{E}[(X-p)(X-p)^\top | Y_\sigma = p] = N_2/N_0$. We compute the three blocks of N_2 and then divide.

In the chart, $\mu(dx) = a_p(z) dz$ and $z = \sigma u$ gives $dz = \sigma^k du$, while equation 80 splits the kernel as

$$e^{-\|F_p(\sigma u) - p\|^2/(2\sigma^2)} = e^{-\|u\|^2/2} \exp\left(-\frac{\|g_p(\sigma u)\|^2}{2\sigma^2}\right) = e^{-\|u\|^2/2} (1 + O(\sigma^2 \|u\|^4)), \quad (90)$$

by equation 83. Combining this with equation 82, the unnormalized measure element is

$$a_p(\sigma u) \sigma^k e^{-\|F_p(\sigma u) - p\|^2/(2\sigma^2)} du = \sigma^{k+m} (q_m(u) + \sigma q_{m+1}(u) + O(\sigma^2 \text{poly}(u))) e^{-\|u\|^2/2} du, \quad (91)$$

the kernel remainder of equation 90 having been absorbed into the $O(\sigma^2 \text{poly}(u))$ term. The displacement itself splits as

$$x - p = U_p(\sigma u) + g_p(\sigma u), \quad g_p(\sigma u) = \frac{\sigma^2}{2} \Pi_p[u, u] + O(\sigma^3 \|u\|^3), \quad (92)$$

using equation 78.

Tangent-tangent block. The tangent part of equation 92 contributes $\sigma^2 U_p u u^\top U_p^\top$, so by equation 91

$$\begin{aligned} P_{T_p \mathcal{M}} N_2 P_{T_p \mathcal{M}} &= \sigma^{k+m+2} U_p \int_{\mathbb{R}^k} u u^\top \left(q_m(u) + \sigma q_{m+1}(u) + O(\sigma^2 \text{poly}(u)) \right) e^{-\|u\|^2/2} du U_p^\top \\ &= \sigma^{k+m+2} U_p \left(\int_{\mathbb{R}^k} u u^\top q_m(u) e^{-\|u\|^2/2} du + O(\sigma^2) \right) U_p^\top. \end{aligned} \quad (93)$$

The order- σ term drops because $u u^\top q_{m+1}(u)$ is odd, $u u^\top$ being even and q_{m+1} odd. The surviving integral is exactly $\mathcal{Z}_q C_{q_m}$: writing $Z \sim \mathcal{N}(0, I_k)$,

$$\int_{\mathbb{R}^k} u u^\top q_m(u) e^{-\|u\|^2/2} du = (2\pi)^{k/2} \mathbb{E}[Z Z^\top q_m(Z)] = (2\pi)^{k/2} \mathbb{E}[q_m(Z)] C_{q_m} = \mathcal{Z}_q C_{q_m}, \quad (94)$$

where the middle step is the definition equation 6 of C_{q_m} and the last is the definition of $\mathcal{Z}_q$. Hence

$$P_{T_p \mathcal{M}} N_2 P_{T_p \mathcal{M}} = \sigma^{k+m+2} \mathcal{Z}_q (U_p C_{q_m} U_p^\top + O(\sigma^2)). \quad (95)$$

Tangent-normal block. Pairing the tangent part of equation 92 with the normal part gives a prefactor $\sigma \cdot \sigma^2/2$, so

$$\begin{aligned} P_{T_p \mathcal{M}} N_2 P_{N_p \mathcal{M}} &= \frac{\sigma^{k+m+3}}{2} U_p \int_{\mathbb{R}^k} u \Pi_p[u, u]^\top q_m(u) e^{-\|u\|^2/2} du + O(\sigma^{k+m+4}) \\ &= O(\sigma^{k+m+4}). \end{aligned} \quad (96)$$

The displayed integral vanishes by parity: u is odd while $\Pi_p[u, u]$ and $q_m(u)$ are both even, so the integrand is odd. The nominal order- σ^{k+m+3} term is therefore absent, and what remains comes from the $O(\sigma^3 \|u\|^3)$ remainder in equation 92 and from the σq_{m+1} term in equation 91.

Normal-normal block. Both factors are normal, so by equation 92 the integrand carries $\|g_p(\sigma u)\|^2 = O(\sigma^4 \|u\|^4)$ and

$$P_{N_p \mathcal{M}} N_2 P_{N_p \mathcal{M}} = O(\sigma^{k+m+4}). \quad (97)$$

Division and centering. Assembling equation 95, equation 96 and equation 97, and dividing by $N_0 = \sigma^{k+m}(\mathcal{Z}_q + O(\sigma^2))$ from equation 84,

$$\mathbb{E}[(X - p)(X - p)^\top \mid Y_\sigma = p] = \frac{N_2}{N_0} = \frac{\sigma^{k+m+2} \mathcal{Z}_q (U_p C_{q_m} U_p^\top + O(\sigma^2))}{\sigma^{k+m} (\mathcal{Z}_q + O(\sigma^2))} = \sigma^2 U_p C_{q_m} U_p^\top + O(\sigma^4), \quad (98)$$

where the last step uses $\mathcal{Z}_q > 0$ from Appendix D and expands $(\mathcal{Z}_q + O(\sigma^2))^{-1} = \mathcal{Z}_q^{-1}(1 + O(\sigma^2))$. Finally equation 7 gives $\mathbb{E}[X \mid Y_\sigma = p] - p = \sigma^2 b_{p,q} + O(\sigma^3)$, so subtracting the outer product of the mean removes only an $O(\sigma^4)$ matrix:

$$\text{Cov}(X \mid Y_\sigma = p) = \sigma^2 U_p C_{q_m} U_p^\top + O(\sigma^4). \quad (99)$$

Dividing by σ^2 and applying Proposition 3.1 proves equation 8.

D.4 Rank and trace

For $v \neq 0$,

$$v^\top C_{q_m} v = \frac{\int (v^\top u)^2 q_m(u) e^{-\|u\|^2/2} du}{\mathcal{Z}_q} > 0. \quad (100)$$

The strict inequality holds because q_m is positive on an open set, which cannot be contained in the hyperplane $v^\top u = 0$. Hence $C_{q_m} \succ 0$, and the ambient limit has rank k and range $T_p \mathcal{M}$.

Euler's identity gives $u^\top \nabla q_m(u) = m q_m(u)$. Integrating the divergence of $u q_m(u) e^{-\|u\|^2/2}$ gives

$$0 = \int_{\mathbb{R}^k} (k q_m(u) + u^\top \nabla q_m(u) - \|u\|^2 q_m(u)) e^{-\|u\|^2/2} du \quad (101)$$

$$= \int_{\mathbb{R}^k} (k + m - \|u\|^2) q_m(u) e^{-\|u\|^2/2} du. \quad (102)$$

Dividing by Z_q proves $\text{tr } C_{q_m} = k + m$.

D.5 Positive density and change of tangent frame

If $\rho(p) > 0$, then the first jet is the positive constant $q_0 = \rho(p)$ (the chart volume factor equals one at the origin). Its Gaussian tilt is unchanged, so $C_{q_0} = \mathbb{E}[ZZ^\top] = I_k$ and $U_p C_{q_0} U_p^\top = P_{T_p \mathcal{M}}$.

For coordinate invariance, let $\tilde{U}_p = U_p O$ for $O \in O(k)$. The same density jet in the new coordinates is $\tilde{q}_m(\tilde{z}) = q_m(O\tilde{z})$. Rotational invariance of the standard Gaussian gives

$$C_{\tilde{q}_m} = O^\top C_{q_m} O.$$

It follows that $\tilde{U}_p C_{\tilde{q}_m} \tilde{U}_p^\top = U_p C_{q_m} U_p^\top$, proving Corollary 4.3.

D.6 Minimal assumptions and rate assumptions

The operator limit requires less regularity than the stated rate. If $g_p(z) = o(\|z\|)$ and $a_p(z) = q_m(z) + o(\|z\|^m)$, dominated convergence in the rescaled posterior gives

$$\sigma^{-2} \text{Cov}(X \mid Y_\sigma = p) \rightarrow U_p C_{q_m} U_p^\top.$$

For $D_\sigma(p) - p = O(\sigma^2)$ it suffices that $g_p(z) = O(\|z\|^2)$ and $a_p(z) = q_m(z) + O(\|z\|^{m+1})$. The sharper normalized-covariance error $O(\sigma^2)$ requires the two-order expansion equation 79, or a comparable assumption such as $a_p \in C^{m+1,1}$.

The additional rate assumptions are necessary. For even m and $0 < \alpha < 1$, the local one-dimensional density

$$a(z) = z^m (1 + \text{sgn}(z)|z|^\alpha)$$

has posterior mean $\Theta(\sigma^{1+\alpha})$, rather than $O(\sigma^2)$. Likewise $a(z) = z^m (1 + |z|^{1+\alpha})$ creates a $\Theta(\sigma^{1+\alpha})$ correction to the normalized covariance. Finite order alone gives convergence, not a universal polynomial remainder.

D.7 A uniform finite-degree eigengap

Lemma D.1. *For fixed k and even m , there is a constant $c_{k,m} > 0$ such that*

$$\lambda_{\min}(C_q) \geq c_{k,m}$$

for every nonzero nonnegative homogeneous polynomial q of degree m .

Proof. Normalize by $\mathbb{E}q(Z) = 1$. Homogeneous degree- m polynomials form a finite-dimensional vector space, and the Gaussian L^1 norm is equivalent to any coefficient norm. The normalized nonnegative polynomials therefore form a compact set. The map $q \mapsto \lambda_{\min}(\mathbb{E}[ZZ^\top q(Z)])$ is continuous and, by the positivity argument above, strictly positive at every point of that compact set. It therefore has a positive minimum. $\square$

The constant may deteriorate with m . For a bounded collection $m \leq m_{\max}$, taking the minimum over the finitely many even degrees yields a uniform gap. This gap applies on a controlled finite-type stratum; uniform spectral recovery also requires bounds on the Taylor remainders.

E Explicit examples and the full-support gap

E.1 Radial and quadratic jets

If $q_m(u) = \|u\|^m$, rotational invariance makes $C_q = cI_k$. The trace identity gives $c = 1 + m/k$. If $q_2(u) = u^\top A u$ with $A \succeq 0$ and $\text{tr } A > 0$, Isserlis' formula gives

$$\mathbb{E}[ZZ^\top(Z^\top A Z)] = (\text{tr } A)I_k + 2A, \quad C_{q_2} = I_k + \frac{2A}{\text{tr } A}.$$

E.2 Circle family

For $x(\theta) = (\cos \theta, \sin \theta)$ and $\rho_r(\theta) \propto (1 - \cos \theta)^r$, the zero at $\theta = 0$ has $m = 2r$ and $q_m(u) \propto u^{2r}$. Hence

$$C_q = \frac{\mathbb{E}Z^{2r+2}}{\mathbb{E}Z^{2r}} = 2r + 1.$$

For $r = 1$, the posterior density in θ at $p = (1, 0)$ is proportional to

$$(1 - \cos \theta) \exp\left(-\frac{1 - \cos \theta}{\sigma^2}\right).$$

The rescaling $\theta = \sigma u$ gives

$$\mathbb{E}[X \mid Y_\sigma = p] = p - \frac{3}{2}\sigma^2(1, 0) + O(\sigma^4), \quad (103)$$

$$\text{Cov}(X \mid Y_\sigma = p) = 3\sigma^2(0, 1)(0, 1)^\top + O(\sigma^4). \quad (104)$$

The density is smooth, nonnegative, and positive except at one point, so every open arc has positive mass and $\text{supp } \mu = S^1$. Equation equation 13 follows, and the example shows that full support is weaker than a positive density lower bound.

F Two low-density crossover laws

F.1 Changing the density floor

The stated bound on mass outside the fixed neighborhood U makes that contribution exponentially negligible after multiplication by the Gaussian kernel: an $\exp(o(\sigma^{-2}))$ mass factor is dominated by the fixed separation penalty $\exp(-c/\sigma^2)$. Inside U , the uniform envelope on r_δ provides a Gaussian-integrable dominator after rescaling. The local rescaling therefore determines the posterior.

In the setting of Theorem 4.4, substitute $z = \sigma u$ and divide the effective density by σ^m :

$$\sigma^{-m} a_{\delta_\sigma}(\sigma u) \longrightarrow q_m(u) + a_0 \lambda.$$

The rescaled tangent posterior therefore has density proportional to $(q_m(u) + a_0 \lambda)e^{-\|u\|^2/2}$. Its mean is zero by parity, and its covariance is

$$\frac{a_0 \lambda \mathbb{E}[ZZ^\top] + \mathbb{E}[ZZ^\top q_m(Z)]}{a_0 \lambda + \mathbb{E}[q_m(Z)]},$$

which is equation 15. Taking the trace and using $\text{tr } C_q = k + m$ proves the second formula. If $\delta_\sigma/\sigma^m \rightarrow \infty$, dividing instead by δ_σ leaves the standard Gaussian law and covariance I_k .

For $a_\delta(\theta) = \delta + 1 - \cos \theta$, the local jet is $q_2(u) = u^2/2$. Here $\mathbb{E}q_2(Z) = 1/2$ and $\mathbb{E}[Z^2 q_2(Z)] = 3/2$, giving

$$C_{q,\lambda} = \frac{\lambda + 3/2}{\lambda + 1/2}.$$

F.2 Moving through a fixed distribution

Let the query be $y_\sigma = F_p(\sigma a)$. With the clean coordinate $z = \sigma u$, the leading kernel becomes

$$\exp\left(-\frac{1}{2}\|u - a\|^2\right),$$

while the density contributes $\sigma^m q_m(u)$. Dominated convergence shows that the rescaled tangent posterior converges to $\nu_{q,a}$, proving equation 18.

The rates used in Section 6 follow under the stronger assumptions of Theorem 4.2, uniformly for a in a fixed compact set:

$$\sigma^{-m} a_p(\sigma u) = q_m(u) + \sigma q_{m+1}(u) + O(\sigma^2 \text{poly}(u)), \quad (105)$$

$$g_p(\sigma u) - g_p(\sigma a) = \frac{\sigma^2}{2} (\Pi_p[u, u] - \Pi_p[a, a]) + O(\sigma^3 \text{poly}(u, a)). \quad (106)$$

The tangent and normal graph components are orthogonal, so

$$\frac{\|F_p(\sigma u) - F_p(\sigma a)\|^2}{2\sigma^2} = \frac{1}{2}\|u - a\|^2 + O(\sigma^2 \text{poly}(u, a)). \quad (107)$$

The same Gaussian domination as in Appendix C therefore shows that posterior tangent moments of order at most two differ from their $\nu_{q,a}$ limits by $O(\sigma)$. Unlike at the centered query, the order- σ term need not vanish because the shifted Gaussian breaks parity. It follows that

$$D_\sigma(F_p(\sigma a)) - F_p(\sigma a) = \sigma U_p(\mathbb{E}_{\nu_{q,a}} u - a) + O(\sigma^2), \quad (108)$$

and its normal component is $O(\sigma^2)$. The tangent-tangent, tangent-normal, and normal-normal blocks of H_σ are respectively

$$C_q^{\text{sp}}(a) + O(\sigma), \quad O(\sigma), \quad O(\sigma^2). \quad (109)$$

These are the stated moving-query mean and link-Hessian rates. The argument in Appendix D shows that the limiting tangent covariance is positive definite; on a compact set of a its minimum eigenvalue is uniformly positive. Since $P_{T_{F_p(\sigma a)}\mathcal{M}} = U_p U_p^\top + O(\sigma)$, Davis–Kahan gives the stated $O(\sigma)$ projector tilt.

For the alternative representation, change variables $u = a + z$:

$$M_q(a) = \mathbb{E}[q_m(a + Z)].$$

Differentiation under this Gaussian expectation gives

$$\nabla_a \log M_q(a) = \mathbb{E}_{\nu_{q,a}}[u - a], \quad \nabla_a^2 \log M_q(a) = \text{Cov}_{\nu_{q,a}}(u) - I_k.$$

For $q_2(u) = u^2$,

$$\mathbb{E}_{\nu_{q,a}} u = \frac{a^3 + 3a}{a^2 + 1}, \quad \mathbb{E}_{\nu_{q,a}} u^2 = \frac{a^4 + 6a^2 + 3}{a^2 + 1}.$$

Subtracting the square of the mean yields $(a^4 + 3)/(a^2 + 1)^2$. Direct differentiation shows its minimum is $3/4$ at $|a| = \sqrt{3}$.

F.3 Uniform normal offsets

We prove Theorem 4.6. All estimates in this subsection are uniform for $(a, b) \in K_T \times K_N$. The localization assumption makes each posterior-moment contribution outside the fixed chart exponentially smaller than the local normalizer $\asymp \sigma^{k+m}$. If the chart contains $\mathcal{M} \cap B(p, \eta)$ for some $\eta > 0$, this assumption follows from $\|F_p(\sigma a) + \sigma b - p\| = O(\sigma)$: outside the chart the Gaussian exponent is at most $-c/\sigma^2$, and the factor $1 + \|x - p\|^2$ is absorbed by a slightly smaller Gaussian exponent. We therefore work in the chart and set $z = \sigma u$.

For compactly varying a and b , the graph expansion gives

$$F_p(\sigma u) - F_p(\sigma a) - \sigma b = \sigma \{U_p(u - a) - b\} + \sigma^2 \{\mathcal{B}(u) - \mathcal{B}(a)\} + O(\sigma^3 \text{poly}(u)), \quad (110)$$

$$\frac{\|F_p(\sigma u) - F_p(\sigma a) - \sigma b\|^2}{2\sigma^2} = \frac{1}{2}\|u - a\|^2 + \frac{1}{2}\|b\|^2 - \sigma \langle b, \mathcal{B}(u) - \mathcal{B}(a) \rangle + O(\sigma^2 \text{poly}(u)). \quad (111)$$

The order- σ cross term between $U_p(u - a)$ and the graph difference vanishes because the latter is normal to $T_p\mathcal{M}$. On a smaller fixed chart the remainders in equation 110–equation 111 are bounded by a fixed polynomial times a Gaussian $e^{-c\|u\|^2}$; the remaining annulus is exponentially negligible. The following expansions may therefore be integrated term by term.

Let

$$Z_a = \int_{\mathbb{R}^k} q_m(u) e^{-\|u-a\|^2/2} du, \quad \langle \phi \rangle_a = Z_a^{-1} \int_{\mathbb{R}^k} \phi(u) q_m(u) e^{-\|u-a\|^2/2} du, \quad (112)$$

so $\langle \phi \rangle_a = \mathbb{E}_{\nu_{q,a}} \phi$. Define

$$r_{a,b}(u) = q_{m+1}(u) + q_m(u) \langle b, \mathcal{B}(u) - \mathcal{B}(a) \rangle, \quad (113)$$

$$\Delta_{a,b}[\phi] = Z_a^{-1} \int_{\mathbb{R}^k} \{\phi(u) - \langle \phi \rangle_a\} r_{a,b}(u) e^{-\|u-a\|^2/2} du. \quad (114)$$

Equations equation 82 and equation 111 show that, after cancelling the constant factor $e^{-\|b\|^2/2}$, the unnormalized rescaled posterior weight is

$$e^{-\|u-a\|^2/2} [q_m(u) + \sigma r_{a,b}(u) + O(\sigma^2 \text{poly}(u))]. \quad (115)$$

The continuous positive function $a \mapsto Z_a$ has a positive minimum on K_T . Expanding the ratio of numerator and denominator in equation 115 therefore yields, for every vector- or matrix-valued polynomial ϕ needed below,

$$\mathbb{E}[\phi(X_{\text{chart}}/\sigma) \mid Y_\sigma = F_p(\sigma a) + \sigma b] = \langle \phi \rangle_a + \sigma \Delta_{a,b}[\phi] + O(\sigma^2). \quad (116)$$

Here X_{chart} denotes the tangent coordinate of X . The error is uniform because all Gaussian-weighted polynomial moments are continuous in a and bounded on K_T .

Set

$$\eta_{a,b} = \Delta_{a,b}[u], \quad d_a = \mu_a - a. \quad (117)$$

Using $F_p(\sigma u) = p + \sigma U_p u + \sigma^2 \mathcal{B}(u) + O(\sigma^3 \text{poly}(u))$ in equation 116 gives

$$D_\sigma(y) = p + \sigma U_p \mu_a + \sigma^2 \{U_p \eta_{a,b} + \bar{\mathcal{B}}_a\} + O(\sigma^3), \quad (118)$$

$$D_\sigma(y) - y = \sigma \{U_p d_a - b\} + \sigma^2 \{U_p \eta_{a,b} + \bar{\mathcal{B}}_a - \mathcal{B}(a)\} + O(\sigma^3). \quad (119)$$

Equation equation 28 follows. The second-order term will be used below.

Let

$$C_{1,a,b} = \Delta_{a,b}[(u - \mu_a)(u - \mu_a)^\top], \quad \Omega_a = \text{Cov}_{\nu_{q,a}}(\mathcal{B}(u)). \quad (120)$$

Centering the tangent and normal graph coordinates in equation 116 gives

$$U_p^\top H_\sigma(y) U_p = C_a + \sigma C_{1,a,b} + O(\sigma^2), \quad (121)$$

$$P^\perp H_\sigma(y) U_p = \sigma \Gamma_a + O(\sigma^2), \quad (122)$$

$$P^\perp H_\sigma(y) P^\perp = \sigma^2 \Omega_a + O(\sigma^3). \quad (123)$$

The cross covariance before division by σ^2 is $\sigma^3 \text{Cov}_{\nu_{q,a}}(\mathcal{B}(u), u) + O(\sigma^4)$. Equations equation 121–equation 123 prove equation 29.

The same hyperplane argument used in Appendix D shows $C_a \succ 0$: the shifted Gaussian is positive everywhere and q_m is positive on a nonempty open set. The entries of C_a depend continuously on a , hence

$$c_0 = \inf_{a \in K_T} \lambda_{\min}(C_a) > 0. \quad (124)$$

The min-max principle applied to the tangent space and equation 121 gives $\lambda_k(H_\sigma) \geq c_0 - O(\sigma)$. Applied with the fixed normal space and equation 123, it gives $0 \leq \lambda_{k+1}(H_\sigma) \leq O(\sigma^2)$. These inequalities prove equation 30 and give a uniform gap of at least $c_0/2$.

To expand the top- k projector, use the fixed decomposition $T_p\mathcal{M} \oplus N_p\mathcal{M}$:

$$H_\sigma = \begin{pmatrix} C_a + O(\sigma) & \sigma\Gamma_a^\top + O(\sigma^2) \\ \sigma\Gamma_a + O(\sigma^2) & O(\sigma^2) \end{pmatrix}. \quad (125)$$

Because of the uniform gap, its top space is the graph of a map $L_\sigma : \mathbb{R}^k \rightarrow N_p\mathcal{M}$ with $\|L_\sigma\|_{\text{op}} = O(\sigma)$. Invariance of that graph is equivalent to

$$H_{NT} + H_{NN}L_\sigma = L_\sigma(H_{TT} + H_{TN}L_\sigma). \quad (126)$$

Substitution of equation 125, followed by right multiplication by C_a^{-1} , gives

$$L_\sigma = \sigma\Gamma_a C_a^{-1} + O(\sigma^2) = \sigma\Lambda_a + O(\sigma^2). \quad (127)$$

The orthogonal projector onto the graph of L_σ has off-diagonal blocks L_σ and $L_\sigma^\top$ up to $O(\|L_\sigma\|_{\text{op}}^2)$ and diagonal blocks I_k and 0 up to the same order. Hence equation 31.

We next compare the landing with nearest projection. Let $\zeta_{\sigma,a,b}$ be the chart coordinate of $\pi(y)$ and put $h = \zeta_{\sigma,a,b} - \sigma a$. The nearest-point equation is

$$DF_p(\zeta_{\sigma,a,b})^\top \{y - F_p(\zeta_{\sigma,a,b})\} = 0. \quad (128)$$

The implicit function theorem first gives $h = O(\sigma^2)$. Since

$$Dg_p(\sigma a) = \sigma\mathcal{B}_a + O(\sigma^2), \quad DF_p(\sigma a)^\top DF_p(\sigma a) = I_k + O(\sigma^2), \quad (129)$$

expanding equation 128 gives

$$0 = \sigma^2\mathcal{B}_a^\top b - h + O(\sigma^3) = \sigma^2 S_b a - h + O(\sigma^3). \quad (130)$$

Therefore

$$\zeta_{\sigma,a,b} = \sigma a + \sigma^2 S_b a + O(\sigma^3), \quad (131)$$

which is equation 32. The tangent space at this projected point is the graph of $\sigma\mathcal{B}_a + O(\sigma^2)$ over $T_p\mathcal{M}$, so

$$P_{T_{\pi(y)}\mathcal{M}} = P + \sigma\{\mathcal{B}_a U_p^\top + U_p \mathcal{B}_a^\top\} + O(\sigma^2). \quad (132)$$

Combining this with equation 31 gives the projector bias

$$Q_\sigma(y) - P_{T_{\pi(y)}\mathcal{M}} = \sigma\{(\Lambda_a - \mathcal{B}_a)U_p^\top + U_p(\Lambda_a - \mathcal{B}_a)^\top\} + O(\sigma^2). \quad (133)$$

Substitute equation 119 and equation 31 into the normal-only displacement:

$$(I - Q_\sigma(y))(D_\sigma(y) - y) = -\sigma b \quad (134)$$

$$+ \sigma^2\{U_p \Lambda_a^\top b + \bar{\mathcal{B}}_a - \mathcal{B}(a) - \Lambda_a d_a\} + O(\sigma^3). \quad (135)$$

Adding $y = p + \sigma U_p a + \sigma b + \sigma^2 \mathcal{B}(a) + O(\sigma^3)$ and subtracting the expansion of $F_p(\zeta_{\sigma,a,b})$ from equation 131 gives exactly equation 33.

For the uniform rates, note that $W_{\pi(y)} = U_p + O(\sigma)$ uniformly. Hence

$$\|H_\sigma(y) - W_{\pi(y)} C_a W_{\pi(y)}^\top\|_{\text{op}} = O(\sigma). \quad (136)$$

This comparison field has exact range $T_{\pi(y)}\mathcal{M}$ and smallest nonzero eigenvalue at least c_0 . Equation equation 33 gives the $O(\sigma^2)$ landing bias, equation 30 gives the uniform gap, and equation 28 gives $\sup_y \|D_\sigma(y) - y\| = O(\sigma)$. These are the claims in equation 34.

G Spectral calibration proofs

G.1 Hard calibration

The eigenvalues of $A_0 = UC_qU^\top$ consist of the k positive eigenvalues of C_q , all at least c_q , and $d - k$ zeros. Weyl's inequality places the top k eigenvalues of $\hat{A}$ above $c_q - \eta$ and the remainder below η . Under $\eta < c_q/2$, this separates the two groups and determines the rank.

The Davis–Kahan sin- Θ theorem [8] gives

$$\|\mathcal{P}_k(\hat{A}) - UU^\top\|_{\text{op}} \leq \frac{\eta}{c_q - \eta} \leq \frac{2\eta}{c_q}.$$

Here the separation required by the sin- Θ theorem is between the spectrum of A_0 on $\text{range}(U)$, contained in $[c_q, \lambda_{\max}(C_q)]$, and the spectrum of $\hat{A}$ on the orthogonal complement of its top- k eigenspace, contained in $[-\eta, \eta]$; and $\|\mathcal{P}_k(\hat{A}) - UU^\top\|_{\text{op}}$ equals the sine of the largest principal angle between the two ranges because both projectors have rank k .

We now prove equation 39, in which only the coupling block enters. Write $P = UU^\top$ and $P^\perp = I_d - P$, and let $U_\perp$ have orthonormal columns spanning $\text{range}(U)^\perp$. Since A_0 is block diagonal for $\mathbb{R}^d = \text{range}(U) \oplus \text{range}(U)^\perp$, we have $P^\perp A_0 P = 0$ and therefore $P^\perp \hat{A} P = P^\perp (\hat{A} - A_0) P$: the coupling block of $\hat{A}$ is exactly the coupling block of the error.

Apply the residual form of the sin- Θ theorem [8] to the symmetric matrix $\hat{A}$ and the trial subspace $\text{range}(U)$, which need not be invariant for $\hat{A}$. Its residual is $P^\perp \hat{A} P$, the compression to the trial subspace is $U^\top \hat{A} U$, and the compression to the complement is $U_\perp^\top \hat{A} U_\perp$; these two compressions are separated by $\delta(\hat{A})$ of equation 38. Whenever $\delta(\hat{A}) > 0$, the theorem bounds the sine of the largest principal angle between $\text{range}(U)$ and the top- k eigenspace of $\hat{A}$ by $\|P^\perp \hat{A} P\|_{\text{op}} / \delta(\hat{A})$, and that sine equals $\|\mathcal{P}_k(\hat{A}) - UU^\top\|_{\text{op}}$ because both projectors have rank k . This is equation 39, and it uses no bound on η .

Under the standing hypothesis $\eta < c_q/2$, Weyl's inequality bounds $\lambda_{\min}(U^\top \hat{A} U)$ below by $c_q - \eta$ and $\lambda_{\max}(U_\perp^\top \hat{A} U_\perp)$ above by η , so $\delta(\hat{A}) \geq c_q - 2\eta > 0$ and equation 39 holds with that denominator. Because $\|P^\perp \hat{A} P\|_{\text{op}} \leq \|\hat{A} - A_0\|_{\text{op}} \leq \eta$, this recovers equation 37 up to the constant, but it is strictly smaller whenever the estimation error is concentrated in the tangent block. Section 7 reports exactly that regime: the learned link eigenvalues are biased while the top eigenspace is accurate.

For the trace statistic, Weyl's inequality gives

$$|\lambda_i(\hat{A}) - \lambda_i(A_0)| \leq \eta \quad (1 \leq i \leq k).$$

Summing and using $\sum_{i=1}^k \lambda_i(A_0) = k + m$ proves Proposition 5.1.

G.2 Soft calibration

For A_0 , functional calculus gives

$$\|\Phi_\tau(A_0) - UU^\top\|_{\text{op}} = \max_i \frac{\tau}{\lambda_i(C_q) + \tau} \leq \frac{\tau}{c_q}.$$

For positive-semidefinite $\hat{A}$, the resolvent identity gives

$$\Phi_\tau(\hat{A}) - \Phi_\tau(A_0) = \tau((A_0 + \tau I)^{-1} - (\hat{A} + \tau I)^{-1}) \quad (137)$$

$$= \tau(\hat{A} + \tau I)^{-1}(\hat{A} - A_0)(A_0 + \tau I)^{-1}. \quad (138)$$

Both inverse norms are at most τ^{-1} , so the last display has norm at most η/τ . The triangle inequality proves Proposition 5.2.

If a symmetric estimate is not PSD, let $\hat{A}_+$ be its orthogonal spectral projection onto the PSD cone. Since $A_0 \succeq 0$, Weyl's inequality gives $\|\hat{A}_+ - \hat{A}\|_{\text{op}} \leq \|\hat{A} - A_0\|_{\text{op}}$. Hence

$$\|\hat{A}_+ - A_0\|_{\text{op}} \leq 2\|\hat{A} - A_0\|_{\text{op}}, \quad (139)$$

which is the conservative raw-to-PSD error conversion stated in the main text.

Limits of spectral calibration. Recovery requires estimation error below the smallest tangent eigenvalue. For example, on a positive-density Gaussian plane with tangent variances κ_i^2 , the population link eigenvalues are $\kappa_i^2/(\kappa_i^2 + \sigma^2)$. If the smallest is below the score-Jacobian estimation error, tangent and normal eigenvectors cannot be guaranteed distinguishable under this operator-error bound. For a fixed κ_i this eigenvalue tends to one as σ decreases, so the difficulty arises only when some tangent scale is comparable to σ or smaller. Spectral calibration changes eigenvalue scaling but cannot create an eigengap. Lemma D.1 gives a lower bound for finite-type zeros under its assumptions.

H Proof of the optimization theorem

Let $p = \pi(x)$, $P = P_{T_p\mathcal{M}}$, and

$$g = P\nabla f(p) = \text{grad}_{\mathcal{M}} f(p), \quad h = Q(x)\nabla f(x).$$

Since $F = f \circ \pi$,

$$\nabla F(x) = (d\pi_x)^\top \nabla f(p).$$

For any iterate with $\text{dist}(x, \mathcal{M}) \leq \epsilon_L$, the theorem's assumptions imply

$$\|\nabla F(x) - g\| \leq K_\pi B \epsilon_L, \quad (140)$$

$$\|h - g\| \leq B \epsilon_P + L_f \epsilon_L. \quad (141)$$

Write the two residuals as r_a and r_h , and let $e = \|r_a\| + \|r_h\|$ be bounded by equation 52. Then

$$\langle \nabla F(x), h \rangle = \langle g + r_a, g + r_h \rangle \geq \frac{1}{2} \|g\|^2 - 2e^2, \quad (142)$$

$$\|h\|^2 \leq 2\|g\|^2 + 2e^2. \quad (143)$$

These bounds follow from $\|g\|e \leq \|g\|^2/4 + e^2$ and $\|r_a\|\|r_h\| \leq e^2/4$; the displayed constants are not optimized.

Because Q is an orthogonal projector, $\|h\| \leq B$. Every point on the segment from x_t to $z_t = x_t - \gamma h_t$ has distance at most $\epsilon_L + \gamma B < r$ from $\mathcal{M}$. The descent lemma for F and equation 142–equation 143 therefore yield

$$F(z_t) \leq F(x_t) - \gamma \langle \nabla F(x_t), h_t \rangle + \frac{L_F \gamma^2}{2} \|h_t\|^2 \quad (144)$$

$$\leq F(x_t) + \left(-\frac{\gamma}{2} + L_F \gamma^2\right) \|g_t\|^2 + (2\gamma + L_F \gamma^2) e^2 \quad (145)$$

$$\leq F(x_t) - \frac{\gamma}{4} \|g_t\|^2 + \frac{9}{4} \gamma e^2. \quad (146)$$

The last line uses $\gamma L_F \leq 1/4$.

The landing map satisfies $\|L(z_t) - \pi(z_t)\| \leq \epsilon_L$, so $\text{dist}(x_{t+1}, \mathcal{M}) \leq \epsilon_L$. It also satisfies

$$\|\pi(L(z_t)) - \pi(z_t)\| \leq L_\pi \epsilon_L.$$

Under $L_\pi \epsilon_L < r$, the line segment between these two manifold points remains in $\mathcal{T}_r$, and the gradient bound gives

$$F(x_{t+1}) - F(z_t) = f(\pi(L(z_t))) - f(\pi(z_t)) \leq B L_\pi \epsilon_L. \quad (147)$$

Combining equation 146 and equation 147, summing from $t = 0$ to $N - 1$, and using $F(x_N) \geq \inf_{\mathcal{M}} f$ gives

$$\frac{\gamma}{4} \sum_{t=0}^{N-1} \|g_t\|^2 \leq F(x_0) - \inf_{\mathcal{M}} f + \frac{9}{4} \gamma N e^2 + N B L_\pi \epsilon_L.$$

Dividing by $\gamma N/4$ proves equation 53. The preceding two bounds also give feasibility and the query-tube condition by induction.

H.1 Normal-only landing error

Let $p = \pi(y)$, $v = y - p$, and $r_D = D(y) - p$. Substituting $y = p + v$ and $D(y) = p + r_D$ into equation 44 gives

$$\overline{D}_Q(y) - p = v + (I - Q)(r_D - v) \quad (148)$$

$$= Qv + (I - Q)r_D. \quad (149)$$

Because $v \in \mathbb{N}_p \mathcal{M}$,

$$\|Qv\| = \|(Q - P_{\mathbb{T}_p \mathcal{M}})v\| \leq \epsilon_P \text{dist}(y, \mathcal{M}).$$

For the second term, split r_D into tangent and normal components:

$$\|(I - Q)r_D\| \leq \|(I - Q)P_{\mathbb{T}_p \mathcal{M}}r_D\| + \|(I - Q)(I - P_{\mathbb{T}_p \mathcal{M}})r_D\| \quad (150)$$

$$\leq \epsilon_P \|P_{\mathbb{T}_p \mathcal{M}}r_D\| + \|(I - P_{\mathbb{T}_p \mathcal{M}})r_D\|. \quad (151)$$

This is Lemma 6.1.

The identity separates two regimes. At an on-manifold query in a generic spatial boundary layer, $v = 0$, the tangent part of r_D is $O(\sigma)$, its normal part is $O(\sigma^2)$, and the score-derived projector can have $\epsilon_P = O(\sigma)$. The normal-only landing error is therefore $O(\sigma^2)$. At a point a fixed normal distance from the manifold, the term $\epsilon_P \text{dist}(y, \mathcal{M})$ is instead $O(\sigma)$. A uniform $O(\sigma^2)$ landing bound on a fixed-width tube is therefore an additional hypothesis, not a consequence of the fixed-center expansion.

H.2 From score error to the two oracle errors

Let

$$D_\theta(y) = y + \sigma^2 s_\theta(y), \quad \hat{A}_\theta(y) = \text{sym}(I + \sigma^2 \nabla s_\theta(y)).$$

Then

$$\|D_\theta - D_\sigma\| \leq \sigma^2 \epsilon_s, \quad \|\hat{A}_\theta - H_\sigma\|_{\text{op}} \leq \sigma^2 \epsilon_J,$$

where symmetry of H_σ and contractivity of symmetrization give the second inequality. For each y in the visited set we apply the projector bound of Proposition 5.1 to $A_{0,\sigma}(y)$. Only that clause is used here, and its proof needs only that $A_{0,\sigma}(y)$ is positive semidefinite of rank k with $\lambda_k \geq c$, which the eigengap hypothesis supplies; the trace clause of the proposition is specific to the centered limit A_0 , since $\text{tr } C_a = k + m$ holds at $a = 0$ but not for general a . This gives

$$\|Q_\theta(y) - P_{\mathbb{T}_{\pi(y)} \mathcal{M}}\|_{\text{op}} \leq \frac{2(\beta_H(\sigma) + \sigma^2 \epsilon_J)}{c},$$

which proves equation 60. The population eigengap assumption and Davis–Kahan separately give

$$\|Q_\theta(y) - Q_\sigma(y)\|_{\text{op}} \leq \frac{2\sigma^2 \epsilon_J}{g_\sigma}. \quad (152)$$

Let $Q_\sigma = P_k(H_\sigma)$ and assume $\sup_{\mathcal{V}} \|D_\sigma(y) - y\| \leq R_D$. The learned and population normal-only landings satisfy

$$\|\overline{D}_{\theta, Q_\theta} - \overline{D}_{\sigma, Q_\sigma}\| \leq \|D_\theta - D_\sigma\| + \|Q_\theta - Q_\sigma\|_{\text{op}} \|D_\sigma - y\| \quad (153)$$

$$\leq \sigma^2 \epsilon_s + \frac{2R_D \sigma^2 \epsilon_J}{g_\sigma}, \quad (154)$$

where the last line uses equation 152. Adding the population landing bias yields

$$\epsilon_L \leq \beta_L(\sigma) + \sigma^2 \epsilon_s + \frac{2R_D \sigma^2 \epsilon_J}{g_\sigma}. \quad (155)$$

Equations equation 60 and equation 155 prove Corollary 6.4. The two learned errors must be tracked separately: a value bound on $D_\theta - D_\sigma$ does not control its Jacobian, as a small-amplitude, high-frequency perturbation shows. The gap assumption is separate as well. Even if both H_σ and $\hat{A}_\theta$ are close to $A_{0,\sigma}$, their top- k spaces need not be stably comparable unless their mutual perturbation is smaller than the population gap g_σ .

H.3 A local invariant basin

We prove Corollary 6.5. Applied on $\mathcal{V}_\sigma^+$, Theorem 4.6 gives constants $c, C > 0$ such that, for all sufficiently small σ ,

$$\|Q_\sigma(y) - P\|_{\text{op}} \leq C\sigma, \quad (156)$$

$$\|\overline{D}_{\sigma, Q_\sigma}(y) - \pi(y)\| \leq C\sigma^2, \quad (157)$$

$$\|D_\sigma(y) - y\| \leq C\sigma, \quad (158)$$

$$\lambda_k(H_\sigma(y)) - \lambda_{k+1}(H_\sigma(y)) \geq c. \quad (159)$$

Since $\omega_\sigma \rightarrow 0$ implies $\delta_H(\sigma) \rightarrow 0$, Davis–Kahan and equation 159 yield

$$\|\widehat{Q}_\sigma(y) - Q_\sigma(y)\|_{\text{op}} \leq C\delta_H(\sigma). \quad (160)$$

It follows that

$$\|\widehat{Q}_\sigma(y) - P\|_{\text{op}} \leq C\{\sigma + \delta_H(\sigma)\} \leq C\omega_\sigma. \quad (161)$$

The learned and population normal-only landing maps satisfy the exact identity

$$\widehat{L}_\sigma(y) - \overline{D}_{\sigma, Q_\sigma}(y) = (I - \widehat{Q}_\sigma(y))\{\widehat{D}_\sigma(y) - D_\sigma(y)\} + (Q_\sigma(y) - \widehat{Q}_\sigma(y))\{D_\sigma(y) - y\}. \quad (162)$$

Combining the preceding bounds gives

$$\|\widehat{L}_\sigma(y) - \pi(y)\| \leq C\sigma^2 + \delta_D(\sigma) + C\sigma\delta_H(\sigma) \quad (163)$$

$$\leq C\sigma\omega_\sigma. \quad (164)$$

Every point sufficiently close to p has a unique fixed-chart representation

$$x = F_p(\sigma a) + \sigma b, \quad b \in \mathbb{N}_p \mathcal{M}, \quad (165)$$

with coordinates

$$a = \sigma^{-1}U_p^\top(x - p), \quad b = \sigma^{-1}P^\perp\{x - p - g_p(\sigma a)\}. \quad (166)$$

For the induction, suppose that $x_t = F_p(\sigma a_t) + \sigma b_t$ with $\|a_t\| \leq R$ and $\|b_t\| \leq M\omega_\sigma$. Since $\nabla f(p) = 0$, Taylor's theorem and $\omega_\sigma \geq \sigma$ give

$$\nabla f(x_t) = \nabla^2 f(p)(x_t - p) + O(\|x_t - p\|^2) \quad (167)$$

$$= \sigma \nabla^2 f(p)U_p a_t + O(\sigma\omega_\sigma). \quad (168)$$

This uses $g_p(\sigma a_t) = O(\sigma^2)$ and $b_t = O(\omega_\sigma)$. Equation 161 then implies

$$U_p^\top \widehat{Q}_\sigma(x_t) \nabla f(x_t) = \sigma G_f a_t + O(\sigma\omega_\sigma), \quad (169)$$

$$P^\perp \widehat{Q}_\sigma(x_t) \nabla f(x_t) = O(\sigma\omega_\sigma). \quad (170)$$

Write the exact fixed-chart coordinates of the prelanding point as

$$z_t = F_p(\sigma \tilde{a}_t) + \sigma \tilde{b}_t, \quad \tilde{b}_t \in \mathbb{N}_p \mathcal{M}. \quad (171)$$

The tangent coordinate of a normal graph is exactly $U_p^\top(F_p(z) - p) = z$, so equation 169 gives

$$\tilde{a}_t = (I_k - \gamma G_f)a_t + O(\omega_\sigma). \quad (172)$$

For the normal coordinate,

$$\sigma \tilde{b}_t = \sigma b_t + g_p(\sigma a_t) - g_p(\sigma \tilde{a}_t) - \gamma P^\perp \widehat{Q}_\sigma(x_t) \nabla f(x_t). \quad (173)$$

Both a_t and $\tilde{a}_t$ remain in a fixed compact set. Since $\|Dg_p(z)\|_{\text{op}} = O(\|z\|)$, the graph difference in equation 173 is $O(\sigma^2)$; using equation 170 therefore gives

$$\|\tilde{b}_t\| \leq \|b_t\| + C\omega_\sigma. \quad (174)$$

For sufficiently small σ , the fixed margins $R < R_+$ and $B_+ > 0$ ensure that $z_t \in \mathcal{V}_\sigma^+$.

Let $\pi(z_t) = F_p(\sigma a_t^\pi)$. The projection expansion equation 32, applied with $(a, b) = (\tilde{a}_t, \tilde{b}_t)$, and equation 174 give

$$a_t^\pi - \tilde{a}_t = O(\sigma\omega_\sigma + \sigma^2) = O(\sigma\omega_\sigma). \quad (175)$$

The last equality uses $\omega_\sigma \geq \sigma$. By equation 164,

$$\|x_{t+1} - F_p(\sigma a_t^\pi)\| \leq C\sigma\omega_\sigma. \quad (176)$$

Taking the fixed tangent coordinate in equation 176 and using equation 172–equation 175 proves

$$a_{t+1} = (I_k - \gamma G_f)a_t + e_t, \quad \|e_t\| \leq C\omega_\sigma. \quad (177)$$

The same landing estimate controls the new normal coordinate. Write $x_{t+1} = F_p(\sigma a_t^\pi) + r_t$ with $\|r_t\| \leq C\sigma\omega_\sigma$. The definition of the fixed-chart coordinates gives

$$\sigma b_{t+1} = P^\perp r_t + g_p(\sigma a_t^\pi) - g_p(\sigma a_{t+1}). \quad (178)$$

In addition, $\|a_{t+1} - a_t^\pi\| = O(\omega_\sigma)$, so the graph difference is $O(\sigma^2\omega_\sigma)$. Hence

$$\|b_{t+1}\| \leq C\omega_\sigma. \quad (179)$$

Choose M larger than both this uniform constant and M_0 .

Iterating equation 177 yields

$$\|a_t\| \leq \rho^t \|a_0\| + C\omega_\sigma \sum_{j=0}^{t-1} \rho^j \leq \rho^t \|a_0\| + \frac{C}{1-\rho} \omega_\sigma. \quad (180)$$

Choose σ small enough that $C\omega_\sigma/(1-\rho) < R - R_0$. A first-exit argument using equation 179 and equation 180 shows that no x_t can leave the narrower set $\{\|a\| < R, \|b\| \leq M\omega_\sigma\}$; the fixed margins then keep every z_t in $\mathcal{V}_\sigma^+$. This closes the induction and proves the recurrence, coordinate bounds, and visited-set claim.

Because $F_p(\sigma a_t) \in \mathcal{M}$,

$$\text{dist}(x_t, \mathcal{M}) \leq \sigma \|b_t\| \leq C\sigma\omega_\sigma. \quad (181)$$

The graph expansion and equation 180 give

$$\|x_t - p\| \leq C\sigma \left(\rho^t \|a_0\| + \frac{\omega_\sigma}{1-\rho} \right). \quad (182)$$

$\nabla f(p) = 0$ and local Lipschitz continuity of ∇f imply the same upper bound for $\|\nabla f(x_t)\|$. Local Lipschitz continuity of nearest projection gives the same bound for $\|\pi(x_t) - p\|$; applying $\text{grad}_{\mathcal{M}} f(\pi(x_t)) = P_{T_{\pi(x_t)}\mathcal{M}} \nabla f(\pi(x_t))$ proves equation 72.

H.4 Why the local corollary assumes ambient stationarity

The condition $\nabla f(p) = 0$ in Corollary 6.5 cannot in general be weakened to Riemannian stationarity. Suppose only $P\nabla f(p) = 0$ and write $n = \nabla f(p) \in N_p\mathcal{M}$. For a population iterate with $x_t = F_p(\sigma a_t) + O(\sigma^2)$, equation 31 gives the scaled tangent recurrence

$$a_{t+1} = a_t - \gamma \{U_p^\top \nabla^2 f(p) U_p a_t + \Lambda_{a_t}^\top n\} + O(\sigma). \quad (183)$$

The corresponding exact-tangent recurrence would contain $\mathcal{B}_{a_t}^\top n = S_n a_t$ instead of $\Lambda_{a_t}^\top n$; together with the ambient Hessian this is the Riemannian Hessian of $f|_{\mathcal{M}}$ at p . At a positive-density point $\Lambda_a = \mathcal{B}_a$ by Gaussian integration by parts, but at a finite-type zero these maps generally differ. Riemannian strong convexity alone therefore does not make equation 183 a contraction.

For a learned oracle, a projector error of size δ_H acting on the nonzero normal vector n creates an $O(\delta_H)$ tangent step. In the scaled a -recurrence, this becomes $O(\delta_H/\sigma)$. At minimum one needs $\delta_H = O(\sigma)$ to remain in the boundary layer and $\delta_H = o(\sigma)$ to recover the population recurrence. Ambient stationarity removes both leakage terms. If $P\nabla f(p) \neq 0$, a fixed step has an $O(1)$ tangent displacement and immediately leaves the $O(\sigma)$ layer; choosing $\gamma = O(\sigma)$ only yields a stopped or finite-horizon statement unless a separate inward-invariance argument is supplied.

H.5 Effect of raw outer denoising

Take the limiting local model of an affine manifold and a symmetric finite-type zero at the origin, so $D(0) = 0$ and $DD(0) = UC_qU^\top$. In tangent coordinates let $f(z) = \|z\|^2/2$. The raw inner-and-outer update has derivative

$$C_q(I - \gamma C_q).$$

Calibrating only the inner multiplier changes this to $C_q(I - \gamma I)$, whereas the normal-only landing equation 44 has tangent derivative I and gives $I - \gamma I$. If C_q has an eigenvalue $c > 1$, the inner-only map expands that direction whenever $c|1 - \gamma| > 1$. Figure 7 plots this instability.

I An infinitely flat zero

Finite type is essential for a finite link-Hessian limit. Consider a locally flat one-dimensional manifold and the smooth density

$$\rho(u) = \begin{cases} e^{-1/u^2}, & u \neq 0, \\ 0, & u = 0. \end{cases}$$

At a query at the origin, the posterior exponent is

$$\frac{u^2}{2\sigma^2} + \frac{1}{u^2}.$$

Writing $u = \sqrt{\sigma}v$ turns this into $\sigma^{-1}(v^2/2 + v^{-2})$, whose two minima satisfy $v^4 = 2$. Laplace's method therefore gives

$$\mathbb{E}[u] = 0, \quad \mathbb{E}[u^2] \sim \sqrt{2}\sigma, \quad H_\sigma(0) \sim \frac{\sqrt{2}}{\sigma}.$$

The tangent eigenvalue diverges. Quadrature confirms this: for $\sigma = 5 \times 10^{-2}$ down to 10^{-3} , the product $\sigma H_\sigma(0)$ takes the values 1.464214, 1.434214, 1.424214, 1.419214, 1.416214, and 1.415214, so it converges to $\sqrt{2} = 1.414214$ with an excess equal to σ at every level, giving $H_\sigma(0) = \sqrt{2}/\sigma + 1 + o(1)$. This is the sense in which the finite-type hypothesis of Definition 4.1 cannot simply be dropped: no rescaling of H_σ by a fixed power of σ produces a finite nonzero limit here. A smooth compact example is obtained by replacing the local coordinate with a circle coordinate and smoothly completing the density away from the zero.

J A curved manifold with non-constant curvature

The circle has constant curvature and the rotated plane of Section 7 is affine, so neither separates the role of curvature from the role of the density jet. This appendix places a finite-type zero on a manifold whose curvature varies along the arc, and isolates where curvature can and cannot appear in Theorem 4.2.

We use the Archimedean spiral that serves as a standard two-dimensional benchmark for generative models, with two changes that the present setting requires: the law is supported exactly on the curve rather than blurred around it, since the ambient blur here is supplied by the forward noising $Y_\sigma = X + \sigma G$, and the density along the arc is made to vanish at one point rather than being roughly uniform. Take $\mathcal{M} = \{x(t) : t \in [0, 4\pi]\}$ with

$$x(t) = (a + bt)(\cos t, \sin t), \quad a = 0.6, \quad b = 0.35, \quad (184)$$

whose speed and signed curvature are

$$\|x'(t)\| = \sqrt{b^2 + (a + bt)^2}, \quad \kappa(t) = \frac{(a + bt)^2 + 2b^2}{((a + bt)^2 + b^2)^{3/2}}. \quad (185)$$

Curvature decreases monotonically along the arc, from $\kappa = 1.106$ at $t = 1$ to $\kappa = 0.209$ at $t = 12$, a factor of 5.3. Writing $s(t)$ for arc length and fixing a base point $p = x(t_*)$, give $\mathcal{M}$ the clean density

$$\rho(x(t)) \propto (s(t) - s(t_*))^2, \quad (186)$$

which is positive except at p , where it vanishes to order two. Because arc length agrees with the normal-graph coordinate of equation 4 to first order, the leading jet is $q_2(z) = z^2$, so $k = 1$, $m = 2$, and Theorem 4.2 predicts $H_\sigma(p) \rightarrow 3\tau\tau^\top$ with τ the unit tangent, whatever the curvature at p .

Deterministic quadrature along the arc confirms this and separates the two effects. Figure 11 reports five choices of $t_\star$ spanning the available curvature range. In every case the tangent Rayleigh quotient converges to 3, so the limit is curvature free, exactly as the theorem states: curvature enters the localization only through the second fundamental form, which contributes at second order and is therefore absent from the limit. What curvature does control is the remainder. At $\sigma = 5 \times 10^{-3}$ the measured deficits satisfy

$$3 - \lambda_1(H_\sigma(p)) = 2.5 \kappa^2 \sigma^2 + o(\sigma^2), \quad \lambda_2(H_\sigma(p)) = 1.5 \kappa^2 \sigma^2 + o(\sigma^2), \quad (187)$$

with both constants reproduced to four decimal places at every tested point and noise level; the fitted ratios are 2.5000 and 1.5000 throughout. The $O(\sigma^2)$ rate of equation 8 is therefore attained rather than merely bounded, and its constant is proportional to κ^2 for this density family. The normal eigenvalue obeys the same law, consistent with the $O(\sigma^2)$ normal block of Theorem 4.6.

Two limitations of this check are worth stating. It is still $k = 1$, so it does not test curvature together with several tangent directions, which remains the gap noted in Section 8. And the constants 2.5 and 1.5 are specific to the density equation 186; a different jet with the same order would change them while leaving the limit and the κ^2 scaling intact.

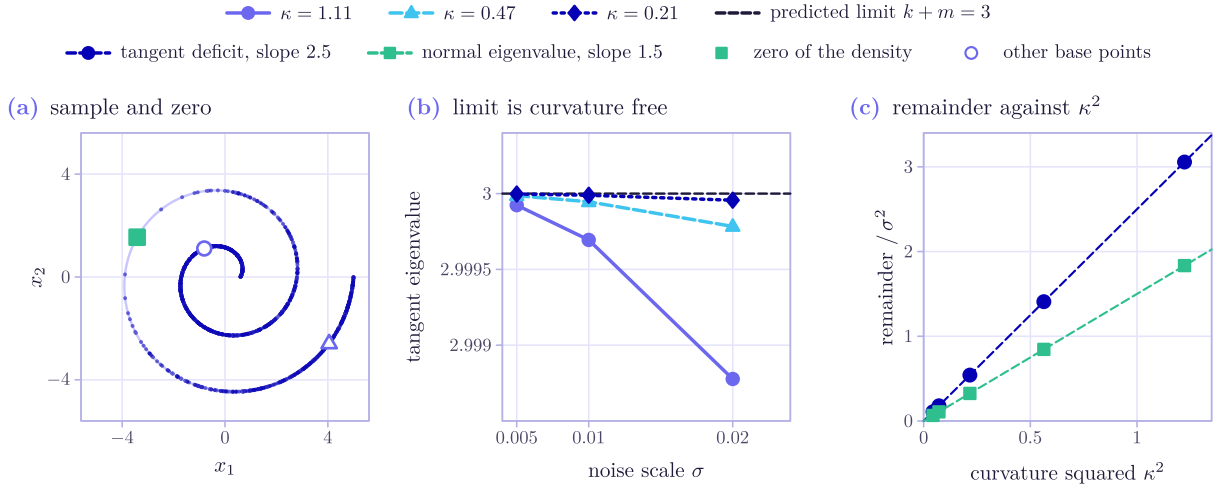


Figure 11 A finite-type zero on a manifold with non-constant curvature. (a) Fourteen hundred points drawn from the clean density equation 186 for the zero at the filled square, with the manifold itself drawn as a faint guide. The samples thin to nothing around that point, which is the feature the theorem is about. The hollow circle and triangle mark two of the other tested base points; each carries its own density vanishing there, so each would empty a different part of the arc. Both axes use the same scale, so the spiral keeps its true shape. (b) The tangent Rayleigh quotient against noise scale at the three curvatures given in the legend, all converging to the predicted $k + m = 3$ shown dashed, at rates that differ. (c) The remainder at $\sigma = 5 \times 10^{-3}$ against κ^2 , with the dashed reference lines of slope 2.5 for the tangent deficit and 1.5 for the normal eigenvalue. The limit does not depend on curvature; only the $O(\sigma^2)$ remainder does.

K Experimental details and numerical checks

This appendix records the protocol of the population, finite-sample, and exact-linearization experiments and of the trained-denoiser and end-to-end optimization experiments.

Circle quadrature. We use a periodic uniform grid of 200,001 points on $[-\pi, \pi)$ for the spectrum experiment and 120,001 points for each crossover curve. The right endpoint is excluded, so the same circle point is not

counted twice. At query angle α , unnormalized weights are

$$w_j = \rho(\theta_j) \exp\left(-\frac{1 - \cos(\theta_j - \alpha)}{\sigma^2}\right).$$

After subtracting the largest log-kernel value for numerical stability, the weights are normalized and the 2×2 posterior covariance is computed. Projecting it onto the tangent and normal vectors at α and dividing by σ^2 gives the plotted link blocks. At the symmetric center these blocks are eigenvalues. At a moving query the cross block is generally nonzero, so Figure 5 labels the tangent quantity as a Rayleigh quotient.

At the smallest plotted noise $\sigma = 0.015$, the tangent eigenvalues for $r = 0, 1, 2, 4$ are respectively

$$0.999887, \quad 2.999325, \quad 4.998312, \quad 8.994937,$$

against predicted limits 1, 3, 5, 9. The corresponding normal eigenvalues are at most 1.02×10^{-3} . The same sweep also exhibits the rate in equation 8 rather than only its limit. Regressing $\log |\lambda_1(H_\sigma) - (k + m)|$ on $\log \sigma$ over the small-noise end of the grid gives slope 2.000 for each of $r = 0, 1, 2, 4$, and the deviations divided by σ^2 settle at 0.5000, 3.0002, 7.5004, and 22.5013. On this family the observed constants agree with $(m + 1)(m + 2)/4$ to four digits, with $m = 2r$.

Crossover accuracy. Across all plotted floor ratios and noise levels, the maximum absolute deviation from $(\lambda + 3/2)/(\lambda + 1/2)$ is 3.01×10^{-2} and the median deviation is 3.19×10^{-3} . For the spatial boundary layer, the maximum absolute deviation from $(a^4 + 3)/(a^2 + 1)^2$ is 3.01×10^{-2} and the median is 2.59×10^{-3} . The largest errors occur at the largest noise, consistent with curvature corrections.

Normal offsets. The experiments plotted in Figures 5 and 6 query on the manifold, that is at $b = 0$. To check Theorem 4.6 at genuinely nonzero normal offset we evaluate the same circle quadrature at $y = (1 + \sigma b)(\cos \sigma a, \sin \sigma a)$, whose rescaled tangential and normal coordinates are a and b . Table 4 reports the tangent block against its predicted limit $C_a = (a^4 + 3)/(a^2 + 1)^2$, the smaller eigenvalue of H_σ divided by σ^2 , and the normal-only landing error divided by σ^2 . Three predictions are visible. The tangent-block deviation divided by σ is constant in σ , so that block is $C_a + O(\sigma)$ as stated. The limit does not depend on b : at $a = 1$ both $b = 0$ and $b = 1.3$ approach 1, which is the uniformity in the normal coordinate. Finally λ_2/σ^2 and the landing error over σ^2 both converge, confirming $\lambda_{k+1} = O(\sigma^2)$ and the $O(\sigma^2)$ landing bias, and the landing limit does depend on b , as equation 33 requires through the term $U_p\{\Lambda_a^\top b - S_b a\}$.

Table 4 Circle quadrature at nonzero normal offset, checking Theorem 4.6. The deviation column is $|U_p^\top H_\sigma U_p - C_a|/\sigma$, which is constant in σ when the tangent block is $C_a + O(\sigma)$.

a	b	σ	$U_p^\top H_\sigma U_p$	C_a	deviation/ σ	landing/ σ^2
0.0	0.0	0.020	2.99880	3.00000	0.060	1.500
0.0	0.0	0.005	2.99992	3.00000	0.015	1.500
1.0	0.0	0.020	0.99967	1.00000	0.017	0.500
1.0	0.0	0.005	0.99998	1.00000	0.004	0.500
1.0	1.3	0.020	0.96200	1.00000	1.900	0.808
1.0	1.3	0.005	0.99031	1.00000	1.937	0.817
-0.8	1.3	0.020	1.21816	1.26770	2.477	0.876
-0.8	1.3	0.005	1.25508	1.26770	2.523	0.898
2.0	-1.1	0.020	0.77537	0.76000	0.769	0.941
2.0	-1.1	0.005	0.76383	0.76000	0.766	0.930

Exact optimization linearization. For $C_q = \text{diag}(c, 1)$ with $c = m + 1$, the four tangent update matrices in Figure 7 are

$$C_q(I - \gamma C_q), \quad C_q(I - \gamma I), \quad I - \gamma C_q, \quad I - \gamma I.$$

The stable interval of the raw inner-and-outer update is $((c - 1)/c^2, (c + 1)/c^2)$ and has width $2/c^2$. The inner-only interval is $(1 - 1/c, 1 + 1/c)$, and the raw-additive interval is $(0, 2/c)$; both have width $2/c$. The doubly calibrated inner-step-and-normal-only-landing map has interval $(0, 2)$. The plotted curves are their exact spectral radii rather than simulated trajectories. Each method is evaluated over its own step-size range.

Finite-sample KDE. Samples from $\rho(\theta) = (1 - \cos \theta)/(2\pi)$ are generated by exact rejection sampling from a uniform proposal, with acceptance probability $(1 - \cos \theta)/2$. For every seed we reuse prefixes of one sample stream at $N \in \{20,000, 80,000, 320,000\}$ and evaluate equation 73 at $\sigma \in \{0.035, 0.05, 0.075, 0.11, 0.16\}$. We use 48 fixed consecutive seeds, 20260819 through 20260866. For each σ , deterministic quadrature first computes the finite-noise population reference H_σ . We then record

$$\|\hat{H}_N - H_\sigma\|_{\text{op}}, \quad \|\mathcal{P}_1(\hat{H}_N) - \mathcal{P}_1(H_\sigma)\|_{\text{op}}, \quad |\lambda_1(H_\sigma) - 3|,$$

as separate sampling, subspace, and finite- σ quantities. The exact population top eigenvalues across the plotted noise levels are respectively 2.99632, 2.99250, 2.98310, 2.96359, and 2.92269. Rounded-order success means $\text{round}(\lambda_1(\hat{H}_N) - 1) = 2$. Figure 3 shows medians and interquartile ranges for the filled operator-error markers, median-only open projector-error markers, and Wilson 95% intervals for this binomial success rate. Since prefixes from a replication are nested and effective sample size is itself random, the pooled display is descriptive; no independence across plotted cells or universal $\text{ESS}^{-1/2}$ law is claimed.

Learned denoisers. Each denoiser is a residual MLP with two tanh hidden layers of width 96, and all of its parameters are trained. Its inputs are the scaled noisy vector and a normalized log σ coordinate. If $r_\theta(y, \sigma)$ is the network output, then

$$D_\theta(y, \sigma) = y + \sigma r_\theta(y, \sigma), \quad \ell(\theta) = \mathbb{E} \left\| r_\theta(Y, \sigma) - \frac{X - Y}{\sigma} \right\|^2.$$

The three circle noise levels are 0.08, 0.12, 0.18, and the plane levels are 0.12, 0.18, 0.25; during training σ is drawn uniformly from these three values rather than from a continuous range, and the same three values are used for evaluation. Training pairs are generated afresh at every step from the exact sampler, so there is no fixed finite training set and no repetition, and the learned-model error therefore reflects approximation and optimization error rather than memorization of a sample. The scaled noisy input is y divided by 1.25 for the circles and by 0.85 for the plane. We optimize every parameter for 3500 Adam steps with batch size 768, cosine-decayed learning rate from 2.5×10^{-3} to 2.5×10^{-4} , weight decay 2×10^{-6} , and gradient-norm clipping at 8. Validation uses 20,000 fresh pairs.

To increase query coverage without changing the population regression function, each minibatch mixes 15% unweighted samples with 85% samples resampled from an eight-times-larger pool according to the strictly positive weight $w(y) = \exp\{-\|y - c\|^2/(2r^2)\} + 10^{-4}$, normalized over the pool, which depends only on Y ; here c is the evaluation point and r the query radius below. Multiplication of the joint density by a function of Y leaves $\mathbb{E}[X | Y = y]$ unchanged wherever the weight is positive. The query radii are 0.22 for the circle and 0.38 for the plane. This resampling changes finite-model optimization but not the population conditional target.

The residual MLP has an analytic input Jacobian obtained by the chain rule through both tanh layers. We form $\hat{H}_\theta = \text{sym}(\nabla_y D_\theta)$, then evaluate the raw matrix, its hard top- k projector, and $\Phi_{0.25}((\hat{H}_\theta)_+)$. For the regime diagnostic of Table 2 we also record, in the true tangent frame at the evaluation point, the coupling block $\|P^\perp \hat{H}_\theta P\|_{\text{op}}$ and the computed separation $\delta(\hat{H}_\theta)$ of equation 38, obtained from the two compressions $U^\top \hat{H}_\theta U$ and $U_\perp^\top \hat{H}_\theta U_\perp$, where the columns of $U_\perp$ are an orthonormal basis of the normal space. A centered finite-difference audit with step 10^{-6} and seed 2026081998 found an operator-norm Jacobian discrepancy of 1.69×10^{-10} . The plane population reference is evaluated in closed form rather than by quadrature: homogeneity of q_4 makes the tilt scale free, so $H_\sigma(p) = \frac{s^2}{s^2 + \sigma^2} U C_q U^\top$ with $s = 0.35$ and $C_q = \text{diag}\{(m+1+\epsilon)/(1+\epsilon), (1+\epsilon(m+1))/(1+\epsilon)\}$ for $m = 4$ and anisotropy $\epsilon = 0.1$, whose trace is $k + m = 6$. The fixed plane frame is obtained by seeded QR in $\mathbb{R}^{16 \times 2}$. Its latent distribution is sampled exactly as a mixture of polynomially tilted Gaussians; a 4,000,000-draw check with seed 2026081997 obtained second moments (0.568106, 0.167006) against analytic values (0.567955, 0.167045), with maximum relative error 2.67×10^{-4} .

There are 12 independent base replications indexed by seeds 2026081900 through 2026081911. Each replication trains one model for each of the three problems, for 36 denoisers total; the plane frame seed is 2026081901, and the paired bootstrap seed is 2026081999. Shaded figure bands show interquartile ranges, not confidence intervals. The two quoted paired intervals use 20,000 percentile-bootstrap resamples of the within-seed raw-minus-hard differences.

Learned optimization. For the circle we minimize $f(x) = \|x - (1, 0)\|^2/2$ from angle 0.55 for 35 steps; for the plane we minimize $f(x) = \|U^\top x\|^2/2$ from latent coordinate (0.40, -0.32) for 30 steps. Both targets are ambient critical points, as required by Corollary 6.5. For each method, we sweep 20 circle step sizes from 0.05 to 1 and 24 plane step sizes from 0.05 to 1.2. Each sweep is run at all three noise levels the corresponding denoiser was trained on, namely 0.08, 0.12, 0.18 for the circle and 0.12, 0.18, 0.25 for the plane, which requires no additional training because the networks are noise conditional. Figure 10 plots the middle level of each problem, $\sigma = 0.12$ and $\sigma = 0.18$, and the noise dependence of the achieved floor is reported in Section 7. The raw method uses $\hat{H}_\theta$ in the inner step and D_θ as the outer landing. The inner-only method replaces the first by $\mathcal{P}_k(\hat{H}_\theta)$ but retains raw landing. SC-DRGD uses the hard projector in both the tangent step and the normal-only landing. Oracle Riemannian gradient descent is included as a deterministic reference. We report the projected objective gap $F(x_t) - F(p) = f(\pi(x_t)) - f(p)$ and distance to the manifold for every iteration, seed, method, and step size; for the affine plane, $F = f$. Values are clipped only when drawing the log-scale figure.

Compute. The reported run used the full training configuration with Python 3.12.13 and NumPy 2.3.5 on an arm64 CPU without an accelerator; its recorded wall-clock time was 169.8 seconds.