LATENT MANIFOLD DIMENSION FROM NOISY OBSERVATIONS: IDENTIFIABILITY AND RECOVERY VIA DIFFUSION SCORES

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models trained by denoising score matching (DSM) offer a promising approach to intrinsic dimension (ID) estimation. However, when observations are not exactly supported on a manifold, two questions remain: **(Q1)** Does the observed law determine a unique latent dimension, and **(Q2)** which diffusion scales recover it? We address these questions under bounded normal tubular noise, allowing a broad range of noise distributions. For (Q1), we define *latent manifold dimension* (LMD) relative to a prescribed model class and establish a uniform reach-separation condition for identifiability that cannot generally be relaxed at its boundary. To study recovery, we first introduce *local Hessian spectral dimension* (LHSD) as a score-based dimension statistic and examine how accurately it can be estimated from data at a given diffusion scale. Under suitable regularity and thin-tube conditions, we establish its minimax estimation rate, attained by an unrestricted empirical-DSM reference estimator. We then address (Q2): LHSD can fail at fine scales and tends to zero as the diffusion scale grows, while a sharp approximation bound characterizes recovery at intermediate scales. Experiments with diffusion transformers qualitatively support this scale dependence, including accurate estimation of the latent dimension at intermediate scales.

1 INTRODUCTION

Intrinsic dimension (ID) characterizes the geometric complexity of high-dimensional datasets [23, 44] and learned representations [7]. It guides dimensionality reduction [35, 40] and supports outlier [61] and out-of-distribution detection [62]. Applications include detecting AI-generated text [57] and mitigating structural hallucinations in diffusion-generated images [48]. In manifold learning, numerous ID estimators have been developed using neighborhood statistics [35, 17], packing numbers [30], entropic graphs [13], or Wasserstein distances [8]. Recently, diffusion-based methods for estimating ID from densities or scores have attracted growing interest [53, 29, 42], with theoretical results linking scores to support dimension under manifold assumptions [51, 34].

However, measurement error, ambient perturbations, and unmodeled variation can move observations away from a regular manifold [21, 45]. Even small perturbations can increase the support dimension beyond that of the latent structure. This poses a fundamental challenge for ID estimation: distinguishing latent geometric structure from variation introduced by noise. Noise sensitivity and scale dependence have been studied in neighborhood-based and multiscale ID estimation [17, 37, 10]. More recently, Martinez et al. [39] develop a diffusion-based approach to stratified learning and analyze dimension recovery under isotropic Gaussian observation noise. These developments motivate studying how latent-dimension identifiability interacts with scale-dependent recovery when observation noise may be non-Gaussian and depend on the latent point. We study this interaction through a framework for bounded normal tubular noise.

Our analysis addresses two questions:

- **(Q1)** Does the observed distribution determine a unique latent manifold dimension?
- **(Q2)** At which diffusion scales can this latent dimension be recovered from diffusion scores?

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

We first define **latent manifold dimension (LMD)** relative to prescribed bounds on tube width, reach, and regularity. We then adopt **local Hessian spectral dimension (LHSD)** [42] and analyze its estimation accuracy before determining which diffusion scales allow it to approximate the LMD.

Our main contributions are as follows.

- **Identifiability of LMD.** If $2\xi_0 < \tau_0$, where ξ_0 bounds the normal perturbation and τ_0 lower-bounds reach, all admissible representations of an observed law have the same dimension, so LMD is well-defined (Theorem 2.1). At $2\xi_0 = \tau_0$, counterexamples for the dimension pairs and common bounds in Theorem 2.2 show that this uniform condition cannot generally be relaxed.
- **LHSD and its estimation.** We introduce calibrated LHSD (Definition 3.1) and establish matching minimax bounds for its finite-scale field under the joint conditions of Theorems 3.1 and 3.2. The unrestricted empirical-DSM reference estimator attains this rate.
- **Scale-dependent recovery of LMD.** At diffusion scales much smaller than the tube width, we construct models with the same LMD and matching conditional noise means and covariances, yet different fine-scale LHSD limits (Theorem 4.1). We then characterize intermediate scales at which LHSD approximates the LMD and derive a sharp noise–geometry bias tradeoff (Theorems 4.2 and 4.3). At the opposite extreme, LHSD converges to zero as the diffusion scale tends to infinity (Section 4.3).
- **Empirical evaluation.** We compare population (or local-tube) LHSD references, LHSD estimated using diffusion transformers (DiTs) trained by DSM, and the known latent dimension on synthetic tubular models. LHSD outperforms the evaluated diffusion-based ID estimators at suitable intermediate scales.

Related work and geometric preliminaries are collected in Appendices A and B.1–B.2, respectively.

2 Tubular noise model and latent manifold dimension

We address **(Q1)** by specifying the bounded tubular noise model (Section 2.1), defining LMD relative to its admissible representations (Section 2.2), and establishing the identifiability condition and boundary counterexamples (Section 2.3).

2.1 Tubular noise model

Throughout the paper, suppose $D \geq 2$ and $d \in \{1, \dots, D-1\}$. For a d-dimensional manifold $\mathcal{M} \subset \mathbb{R}^D$ and $x \in \mathcal{M}$, let $T_x\mathcal{M}$ and $N_x\mathcal{M} = (T_x\mathcal{M})^\perp$ be the tangent and normal spaces, with orthogonal projectors $P_T(x)$ and $P_N(x)$. We separate latent variation from normal displacement:

$$Y = X + \xi, \qquad X \in \mathcal{M} \subset \mathbb{R}^D, \qquad \xi \in N_X\mathcal{M}, \qquad \|\xi\| \leq \xi_0 \quad \text{a.s.}$$

Let $Q = \mathcal{L}(Y)$ denote the observed law. Subject to the normal-direction constraint and the bound $\|\xi\| \leq \xi_0$, the conditional noise law is otherwise unrestricted. It may depend on X and be discrete or continuous, centered or noncentered, and symmetric or asymmetric.

Assumption 2.1 (Latent manifold)**.** *For a fixed integer $\beta \geq 2$, $\mathcal{M} \in \mathcal{C}^{\beta+1}_{d, L_\mathcal{T}, \tau_0}(\mathbb{R}^D)$.*

This class consists of compact connected embedded manifolds without boundary, with reach at least τ_0 and uniformly regular local coordinate maps. The reach is the supremum of radii r such that every point at distance less than r from $\mathcal{M}$ has a unique nearest point on $\mathcal{M}$; see Appendix B.2.

Assumption 2.2 (Latent density)**.** *Let $\mathcal{M}$ satisfy Assumption 2.1. $\mathcal{L}(X)$ has a density $f \in C^\beta(\mathcal{M})$ satisfying*

$$d\mathcal{L}(X) = f \, d\,\mathrm{vol}_\mathcal{M}, \quad \int_\mathcal{M} f \, d\,\mathrm{vol}_\mathcal{M} = 1, \quad 0 < f_- \leq f \leq f_+ < \infty, \quad \|f\|_{C^\beta(\mathcal{M})} \leq L_\mathcal{F}.$$

In particular, $\mathrm{supp}(\mathcal{L}(X)) = \mathcal{M}$.

Here $\mathrm{vol}_\mathcal{M}$ is the Riemannian volume induced by the ambient Euclidean metric. The condition $\beta \geq 2$ supplies second-order control of $\log f$ for the LHSD approximation bounds.

Assumption 2.3 (Arbitrary bounded normal perturbation). *Let $\mathcal{M}$ and X satisfy Assumptions 2.1–2.2. Assume that there exists a Borel probability kernel $x \mapsto K_x$ from $\mathcal{M}$ to $\mathbb{R}^D$ such that, for every Borel set $B \subseteq \mathbb{R}^D$,*

$$K_X(B) = \mathbb{P}(\xi \in B \mid X), \qquad K_X\big(\{v \in N_X\mathcal{M} : \|v\| \le \xi_0\}\big) = 1 \qquad \text{a.s.}$$

The kernel is a measurable family of conditional noise distributions at latent points x (Appendix B.1). As noted by Divol [16], a tubular noise model can admit multiple latent representations (X, ξ) of the same observed law $Q = \mathcal{L}(X+\xi)$. For representation-dependent geometric quantities, we fix an arbitrary admissible representation; the bounds below hold uniformly over this choice.

Definition 2.1 (Tubular noise class). *A Borel probability measure Q belongs to $\mathcal{P}_d(\xi_0, \tau_0, L_\mathcal{T}, L_\mathcal{F})$ if there exist a d-dimensional manifold $\mathcal{M}$ and random variables X, ξ satisfying Assumptions 2.1–2.3 such that $Q = \mathcal{L}(X+\xi)$.*

We suppress β, f_-, f_+ in the class notation. Unless stated otherwise, these parameters and $d, D, \tau_0, L_\mathcal{T}, L_\mathcal{F}$ are fixed independently of the sample size n, while $\xi_{0,n}$ may vary with n.

2.2 Latent manifold dimension

In general, the observed support alone does not specify which variation belongs to the latent manifold and which comes from normal perturbations. We therefore define the target through all admissible representations of the observed law.

Definition 2.2 (Latent manifold dimension). *For fixed $(\xi_0, \tau_0, L_\mathcal{T}, L_\mathcal{F})$ and $D \ge 2$, let*

$$\mathfrak{D}_{\xi_0,\tau_0,L_\mathcal{T},L_\mathcal{F}}(Q) := \big\{r \in [D-1] : Q \in \mathcal{P}_r(\xi_0, \tau_0, L_\mathcal{T}, L_\mathcal{F})\big\}.$$

If this set is a singleton, define the latent manifold dimension (LMD) $d_{\mathrm{LMD}}(Q; \xi_0, \tau_0, L_\mathcal{T}, L_\mathcal{F})$ as its unique element. Otherwise, the LMD is undefined. Once the class parameters are fixed and uniqueness is established, we write $d_{\mathrm{LMD}}(Q)$.

LMD is relative to prescribed tube-width, reach, and regularity bounds; relaxing them can admit additional latent representations. The observed support can include extra normal directions and, for suitable noise laws, even contain an open subset of $\mathbb{R}^D$ despite a d-dimensional latent manifold.

2.3 Identifiability under reach separation

The bounded noise condition and positivity of the latent density imply that every admissible manifold lies within Hausdorff distance ξ_0 of $\mathrm{supp}(Q)$ (Appendix B.2). Two such manifolds therefore have Hausdorff distance at most $2\xi_0$. When this distance is less than the common reach lower bound τ_0, the nearest-point projections compose to maps continuously deformable to the respective identity maps. This homotopy equivalence forces compact connected boundaryless manifolds to have equal dimensions [24].

Theorem 2.1 (LMD identifiability under reach separation). *If $2\xi_0 < \tau_0$, then every $Q \in \bigcup_{r\in[D-1]} \mathcal{P}_r(\xi_0, \tau_0, L_\mathcal{T}, L_\mathcal{F})$ has exactly one admissible dimension.*

The proof is given in Appendix C.2. Identifying the dimension is distinct from reconstructing the latent geometry, as in manifold estimation under Hausdorff loss [21, 31] and estimation of points and tangent spaces from noisy observations [6, 2].

The next theorem constructs a single observed law with two admissible latent dimensions for suitable common bounds. Counterexamples at equality show that the strict reach-separation condition cannot generally be weakened.

Theorem 2.2 (Nonidentifiable construction at the reach boundary). *Assume $D \ge 4$. Let $d_1, d_2 \in [D-2]$ satisfy $d_1 < d_2$, and suppose $2\xi_0 \ge \tau_0$. There exist common density bounds $0 < f_{-,0} < f_{+,0} < \infty$, finite common envelopes $L_{\mathcal{T},0}, L_{\mathcal{F},0}$, and a Borel probability measure*

$$Q \in \mathcal{P}_{d_1}(\xi_0, \tau_0, L_{\mathcal{T},0}, L_{\mathcal{F},0}) \cap \mathcal{P}_{d_2}(\xi_0, \tau_0, L_{\mathcal{T},0}, L_{\mathcal{F},0}),$$

where both classes use $f_- = f_{-,0}$ and $f_+ = f_{+,0}$. Thus, the LMD of Q is undefined for these common bounds.

The construction treats a spherical component as latent geometry in one representation and as normal perturbation in another, producing the same observed law with different latent dimensions; see Appendix C.3.

3 LHSD: CONSTRUCTION AND STATISTICAL ESTIMATION

Before addressing **(Q2)**, we introduce LHSD as the dimension statistic used in our recovery analysis and quantify its estimation error. Denoising score matching with Gaussian corruption targets the score of the Gaussian-smoothed observed law [60], as used in diffusion models [49, 25, 50]. The smoothing scale σ controls this Gaussian corruption and is distinct from the bound ξ_0 on the original, possibly non-Gaussian, normal noise.

Sections 3.1–3.2 construct LHSD from a local covariance and its spectrum; Section 3.3 establishes its minimax estimation rate. Section 4 then relates the field to the LMD.

3.1 GAUSSIAN LOCALIZATION AND THE NORMALIZED HESSIAN

For $\sigma > 0$, let $Y \sim Q$ and $G \sim \mathcal{N}(0, I_D)$ be independent, and set $W_\sigma = Y + \sigma G$. Its density is

$$q_{Q,\sigma}(y) := \int_{\mathbb{R}^D} \varphi_\sigma(y - z)\, dQ(z), \qquad \varphi_\sigma(v) := (2\pi\sigma^2)^{-D/2} \exp\!\big(-\|v\|^2/2\sigma^2\big)\,.$$

Since $q_{Q,\sigma}(y) > 0$,

$$\Pi^y_{Q,\sigma}(dz) := q_{Q,\sigma}(y)^{-1}\varphi_\sigma(y - z)Q(dz)$$

is the Gaussian-weighted distribution at query y and scale σ, equivalently, a version of the conditional law $\mathcal{L}(Y \mid W_\sigma = y)$. Let $\mu_{Q,\sigma}(y)$ and $\Sigma_{Q,\sigma}(y)$ denote its mean and covariance.

The score is $s_{Q,\sigma} := \nabla \log q_{Q,\sigma}$, whose Jacobian is $\nabla s_{Q,\sigma} = \nabla^2 \log q_{Q,\sigma}$. Define the normalized negative Hessian by

$$A_{Q,\sigma}(y) := -\sigma^2 \nabla_y^2 \log q_{Q,\sigma}(y) = I_D - \Sigma_{Q,\sigma}(y)/\sigma^2. \tag{1}$$

Here, differentiating the Gaussian convolution twice gives the identity. The normalization makes the projector geometry explicit: on a flat d-dimensional patch with nearly constant density and negligible normal variation, $\Sigma_{Q,\sigma}(y) \approx \sigma^2 P_T$, so $A_{Q,\sigma}(y) \approx I_D - P_T = P_N$, with projectors evaluated at $x = \pi_{\mathcal{M}}(y)$ within the reach tube. Thus d eigenvalues lie near zero and $D - d$ near one. Under tubular noise, this separation requires smoothing that averages normal variation while preserving local manifold geometry.

3.2 LOCAL HESSIAN SPECTRAL DIMENSION

For symmetric $A = U \operatorname{diag}(\lambda_j) U^\top$ with orthogonal U, filters act on eigenvalues: $h(A) := U \operatorname{diag}(h(\lambda_j)) U^\top$, so $\operatorname{tr} h(A) = \sum_j h(\lambda_j)$. Fix a spectral threshold $\varpi \in (0, 1)$ throughout. The fourth-order filter of Osada [42] is $\phi_\varpi(\lambda) := \varpi^4/(\varpi^4 + \lambda^4)$.

When applied to an ideal normal projector, the filter yields a trace of $d + (D - d)\phi_\varpi(1)$ rather than d. Therefore, we remove this deterministic offset by the affine rescaling

$$r_\varpi := \phi_\varpi(1), \qquad \bar\phi_\varpi(\lambda) := \frac{\phi_\varpi(\lambda) - r_\varpi}{1 - r_\varpi}, \qquad \operatorname{tr} \bar\phi_\varpi(A) = \frac{\operatorname{tr} \phi_\varpi(A) - D r_\varpi}{1 - r_\varpi}.$$

Then $\bar\phi_\varpi(0) = 1$, $\bar\phi_\varpi(1) = 0$, and every projector P_N of rank $D - d$ satisfies $\operatorname{tr} \bar\phi_\varpi(P_N) = d$. The calibrated filter is fourth-order flat at zero, $1 - \bar\phi_\varpi(\lambda) = O_\varpi(\lambda^4)$, whereas $\bar\phi'_\varpi(1) = -4\varpi^4/(1 + \varpi^4) \neq 0$. Small tangent-eigenvalue perturbations therefore enter at fourth order, and normal-eigenvalue perturbations enter at first order. The threshold ϖ acts on eigenvalues, while σ controls spatial localization.

Definition 3.1 (LHSD with affine calibration)**.** *For a Borel probability measure Q on $\mathbb{R}^D$, define*

$$T_{\sigma,\varpi}(Q)(y) := \operatorname{tr}\big[\bar\phi_\varpi(A_{Q,\sigma}(y))\big]\,.$$

Figure 1 illustrates how affine-calibrated LHSD approximates the LMD at intermediate scales under tubular noise. Empirically, Osada [42] reports lower average error than LIDL and FLIPD on high-dimensional and nonlinear benchmarks. Our experiments provide complementary evidence at suitable intermediate scales under tubular noise (Section 5).

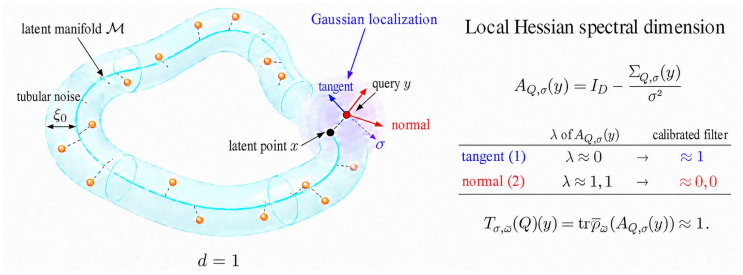


Figure 1: From tubular geometry to LHSD. Gaussian localization at scale σ produces a local covariance around the observed query. At an intermediate scale, the normalized covariance is close to the tangent projector and the normalized negative Hessian is close to the normal projector. The trace of the affine-calibrated filter applied to this matrix therefore approximates the LMD.

3.3 Minimax rates for estimating the LHSD field

Before relating LHSD to the LMD, we quantify the statistical difficulty of estimating its finite-scale field using an unrestricted empirical-DSM reference estimator.

Let $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} Q$ be the observed sample. We define the *LHSD generalization error* as the expected squared discrepancy between an estimated field and $T_{\sigma,\varpi}(Q)$ at a query independent of the training sample. For $\xi_0 \geq 0$, its minimax value over the tubular model class is

$$\mathfrak{R}^{\mathrm{LHSD}}_{n,\sigma}(\xi_0) := \inf_{\widehat{T}} \sup_{Q \in \mathcal{P}_d(\xi_0, \tau_0, L_{\mathcal{T}}, L_{\mathcal{F}})} \mathbb{E}_{Q^n} \|\widehat{T} - T_{\sigma,\varpi}(Q)\|^2_{L^2(Q)}, \tag{2}$$

where $\widehat{T}$ ranges over measurable estimators based on n observations and the $L^2(Q)$ norm integrates an independent query. The next two theorems give matching bounds in a common thin-tube regime. For the lower-bound construction, fix an auxiliary integer $\alpha \geq \beta \vee 3$. We consider scales $n^{-1/(2\alpha+d)} \leq \sigma_n \to 0$, with all model parameters other than $\xi_{0,n}$ fixed.

Theorem 3.1 (Minimax lower bound)**.** *If the fixed geometric and density bounds are sufficiently permissive, then, for every $\xi_{0,n} \geq 0$ and all sufficiently large n,*

$$\mathfrak{R}^{\mathrm{LHSD}}_{n,\sigma_n}(\xi_{0,n}) \gtrsim_{\varpi} n^{-1}\sigma_n^{4-d}.$$

Theorem 3.2 (Minimax upper bound)**.** *If $\xi_{0,n} \leq C_\xi \sigma_n^2/\tau_0$ for a fixed $C_\xi > 0$, then, for all sufficiently large n,*

$$\mathfrak{R}^{\mathrm{LHSD}}_{n,\sigma_n}(\xi_{0,n}) \lesssim_{\varpi} n^{-1}\sigma_n^{4-d}.$$

This upper bound is achieved by the unrestricted empirical-DSM reference estimator $\widetilde{T}_n = T_{\sigma_n,\varpi}(Q_n)$, where Q_n is the empirical measure $n^{-1}\sum_{i=1}^n \delta_{Y_i}$.

Here, δ_{Y_i} is the Dirac measure centered at Y_i. Under the joint conditions, the minimax rate is $n^{-1}\sigma_n^{4-d}$, with $n\sigma_n^d$ the order of the local sample count; small normal-covariance fluctuations and fourth-order suppression of tangent-eigenvalue errors yield the σ_n^4 factor. The proofs are given in Appendices D.1 and D.2. The lower bound uses Assouad's method [56]. Appendix D also compares the rate with the FLIPD lower bound of Seo et al. [47], for a different finite-scale target.

The experiments in Section 5 instead compute LHSD from DSM-trained DiTs. Extending the reference guarantee to these parameterized models requires control of the learned score Jacobian and is beyond the scope of this work.

4 Scale-dependent recovery of latent manifold dimension

We now address **(Q2)** by determining when LHSD approximates the LMD. Section 4.1 shows that LHSD can count additional noise directions when the diffusion scale is much smaller than the tube width. Section 4.2 establishes intermediate-scale recovery and a sharp noise–geometry bias tradeoff, then combines approximation and estimation bounds. Section 4.3 explains the loss of dimension information at large scales.

4.1 AT SMALL DIFFUSION SCALES

At scales much smaller than the tube width, LHSD can count directions introduced by noise as well as directions along the latent manifold. The next theorem constructs models that share the same latent manifold, density, and LMD and have matching conditional noise means and covariances, but whose LHSD limits attain every integer from d to D.

Theorem 4.1 (Different LHSD limits under matched noise moments)**.** *Fix $d \in [D-1]$. There exist fixed parameters $(\tau_0, L_{\mathcal{T},0}, L_{\mathcal{F},0})$ and a common latent variable X with a smooth positive density on a compact connected boundaryless d-dimensional manifold $\mathcal{M} \subset \mathbb{R}^D$ with the following property. For any positive sequences $(\xi_{0,n})$ and (σ_n) with $0 < \xi_{0,n} \le \tau_0/3$ and $\sigma_n/\xi_{0,n} \to 0$, there are models $Q_n^{(k)} := \mathcal{L}(X + \xi_n^{(k)})$, $k \in [D-d]_0$, whose Borel conditional normal-noise kernels satisfy, for every n, k, and $x \in \mathcal{M}$,*

$$\mathbb{E}(\xi_n^{(k)} | X = x) = 0, \quad \|\xi_n^{(k)}\| \le \xi_{0,n} \text{ a.s.}, \quad \mathrm{Cov}(\xi_n^{(k)} | X = x) = (D-d+2)^{-1} \xi_{0,n}^2 P_{N_x\mathcal{M}}. \tag{3}$$

For every n and k,

$$Q_n^{(k)} \in \mathcal{P}_d(\xi_{0,n}, \tau_0, L_{\mathcal{T},0}, L_{\mathcal{F},0}), \qquad d_{\mathrm{LMD}}(Q_n^{(k)}) = d. \tag{4}$$

Their LHSD fields nevertheless satisfy

$$\|T_{\sigma_n,\varpi}(Q_n^{(k)}) - (d+k)\|_{L^2(Q_n^{(k)})} \longrightarrow 0, \qquad k \in [D-d]_0. \tag{5}$$

Here $k \in [D-d]_0 = \{0, \dots, D-d\}$ counts extra noise directions. At a fixed latent point x, the normal perturbation is supported in $N_x\mathcal{M}$, and observations lie in $x + N_x\mathcal{M}$. The local support geometry of this conditional noise distribution matters beyond its first two moments. For example, matched-moment radial perturbations of a circle can produce two nearby circles or an annulus, with LHSD limits one or two despite a common LMD of one. Appendix E gives the constructions and proof.

4.2 AT INTERMEDIATE DIFFUSION SCALES

For $2\xi_0 < \tau_0$, define the worst-case error in approximating the LMD by LHSD:

$$\mathfrak{B}_{\xi_0}(\sigma) := \sup_{Q \in \mathcal{P}_d(\xi_0, \tau_0, L_{\mathcal{T}}, L_{\mathcal{F}})} \|T_{\sigma,\varpi}(Q) - d_{\mathrm{LMD}}(Q)\|_{L^\infty(Q)}.$$

We first bound this error at intermediate scales, then establish a matching lower bound to identify the scale minimizing the worst-case approximation bias.

Let Λ_f be a fixed common upper bound, over admissible representations, on the Riemannian derivatives (Appendix B.2):

$$\|\nabla^2_{\mathcal{M}} \log f\|_\infty + \|\nabla_{\mathcal{M}} \log f\|^2_\infty \le \Lambda_f.$$

Such a bound follows from $f_- > 0$ and the common smoothness envelopes. Define the dimensionless scale-error parameter by

$$\varepsilon(\xi_0, \sigma) := (\xi_0/\sigma)^2 + (\sigma/\tau_0)^2 + \sigma^2 \Lambda_f.$$

The terms capture normal perturbations, curvature, and density variation. Since $2\xi_0/\tau_0 \le (\xi_0/\sigma)^2 + (\sigma/\tau_0)^2$, sufficiently small $\varepsilon(\xi_0, \sigma)$ also ensures identifiability.

For the key mechanism, fix an admissible representation and $y \in \mathrm{supp}(Q)$, and write $x = \pi_{\mathcal{M}}(y)$ and $\varepsilon = \varepsilon(\xi_0, \sigma)$. When ε is sufficiently small, the covariance relative to $T_x\mathcal{M} \oplus N_x\mathcal{M}$ satisfies

$$\frac{\Sigma_{Q,\sigma}(y)}{\sigma^2} = \begin{pmatrix} I_d + E_y & G_y \\ G_y^\top & F_y \end{pmatrix}, \qquad \|E_y\|_{\mathrm{op}} \vee \|F_y\|_{\mathrm{op}} \lesssim \varepsilon, \qquad \|G_y\|_{\mathrm{op}} \lesssim \sqrt{\varepsilon}.$$

Eliminating the normal block in the eigenvalue equation (a Schur complement) makes the mixed block contribute quadratically, at order ε (Lemmas F.3 and F.4). The filter's fourth-order tangent and first-order normal responses then give the next bound uniformly over the allowed noise kernels.

Theorem 4.2 (Recovery at intermediate scales)**.** *There exists $\varepsilon_0 > 0$, depending only on the fixed model parameters, such that if $\varepsilon(\xi_0, \sigma) \le \varepsilon_0$, then $2\xi_0 < \tau_0$ and the LMD is well-defined. For every $Q \in \mathcal{P}_d(\xi_0, \tau_0, L_\mathcal{T}, L_\mathcal{F})$ and Q-almost every y, the eigenvalues of $A_{Q,\sigma}(y)$ can be indexed so that*

$$\max_{1\le j\le d} |\lambda_j(y)| \ \vee \ \max_{d<j\le D} |\lambda_j(y) - 1| \lesssim \varepsilon(\xi_0, \sigma). \tag{6}$$

Consequently,

$$\mathfrak{B}_{\xi_0}(\sigma) \lesssim_\varpi d\varepsilon(\xi_0, \sigma)^4 + (D-d)\varepsilon(\xi_0, \sigma). \tag{7}$$

The proof is given in Appendix F. We next ask whether this noise–geometry tradeoff can be improved. Under $\tau_0^2 \Lambda_f \le C_{\mathrm{den}}$ for fixed $0 < C_{\mathrm{den}} < \infty$, Theorem 4.2 gives $\mathfrak{B}_{\xi_0}(\sigma) = O_\varpi(\xi_0/\tau_0)$ at $\sigma^2 \asymp \xi_0\tau_0$ as $\xi_0 \downarrow 0$ with τ_0 fixed. The next theorem gives a matching lower bound under a class-richness condition.

Theorem 4.3 (Sharp scale bias tradeoff)**.** *Assume $\tau_0^2 \Lambda_f \le C_{\mathrm{den}}$ for a fixed dimensionless constant $0 < C_{\mathrm{den}} < \infty$, and that the model class contains the uniform law on $\mathcal{M}_\circ = \mathbb{S}^d(\tau_0) \times \{0\}^{D-d-1}$ and the observed laws obtained by adding the spherical normal perturbations constructed in Appendix F.4. There exists $c_{\mathrm{loc}} \in (0, 1/\sqrt{2})$ such that, whenever $\xi_0/\sigma \le c_{\mathrm{loc}}$ and $\sigma/\tau_0 \le c_{\mathrm{loc}}$,*

$$\mathfrak{B}_{\xi_0}(\sigma) \asymp_\varpi (\xi_0/\sigma)^2 + (\sigma/\tau_0)^2. \tag{8}$$

Consequently, as $\xi_0 \downarrow 0$ with τ_0 fixed,

$$\inf_{\substack{\sigma>0:\ \xi_0/\sigma\le c_{\mathrm{loc}} \\ \sigma/\tau_0 \le c_{\mathrm{loc}}}} \mathfrak{B}_{\xi_0}(\sigma) \asymp_\varpi \xi_0/\tau_0. \tag{9}$$

A scale satisfying $\sigma^2 \asymp \xi_0\tau_0$ attains this infimum up to constant factors.

The proof is given in Appendix F.4. This oracle bias-minimizing scale depends on tube width and reach. Selecting it from data when these are unknown remains open.

For the unrestricted empirical-DSM reference estimator, the approximation bound in Theorem 4.2 also yields a finite-sample guarantee for LMD recovery. Let $\varepsilon_n := \varepsilon(\xi_{0,n}, \sigma_n)$ and $N_n := n\sigma_n^d$. For $0 < \sigma_n < \tau_0$, sufficiently small ε_n, and $N_n \ge C_{\mathrm{est}}$, Theorem D.2 and the squared triangle inequality give

$$\begin{aligned}\sup_{Q\in\mathcal{P}_d(\xi_{0,n},\tau_0,L_\mathcal{T},L_\mathcal{F})} & \mathbb{E}_{Q^n}\|\widetilde{T}_n - d_{\mathrm{LMD}}(Q)\|^2_{L^2(Q)} \\ &\lesssim_\varpi \underbrace{\left[d\varepsilon_n^4 + (D-d)\varepsilon_n\right]^2}_{\text{squared approximation bias}} + \underbrace{\frac{\varepsilon_n^2}{N_n} + \frac{1}{N_n^2}}_{\text{finite-sample estimation error}}.\end{aligned} \tag{10}$$

The thresholds depend only on the fixed model parameters; the two terms separate approximation bias from finite-sample field estimation.

Under the joint conditions of Theorems 3.2 and 4.2, suppose $\mathfrak{B}_{\xi_{0,n}}(\sigma_n) \le 1/4$. For a query $Y' \sim Q$ independent of the training sample, the same reference estimator satisfies, for all sufficiently large n,

$$\sup_Q \mathbb{P}_{Q^{n+1}}\left\{\mathrm{round}\big(\widetilde{T}_n(Y')\big) \ne d_{\mathrm{LMD}}(Q)\right\} \lesssim_\varpi n^{-1}\sigma_n^{4-d}.$$

Here round rounds to the nearest integer with fixed tie-breaking, and the supremum is over the same tubular model class. The rounding argument is given in Appendix D.4.

Equation 10 also includes estimation error: increasing σ_n decreases the noise term in ε_n and increases N_n, while increasing the curvature and density-variation terms. Thus the bias-optimal scale is a population benchmark for finite-sample scale selection.

4.3 At large diffusion scales

Excessive smoothing also removes dimension information. For any fixed admissible law Q, compactness of the latent manifold and bounded noise imply bounded support. Every Gaussian-weighted distribution is supported on the same set, so $\Sigma_{Q,\sigma}(y)$ remains uniformly bounded in y and σ.

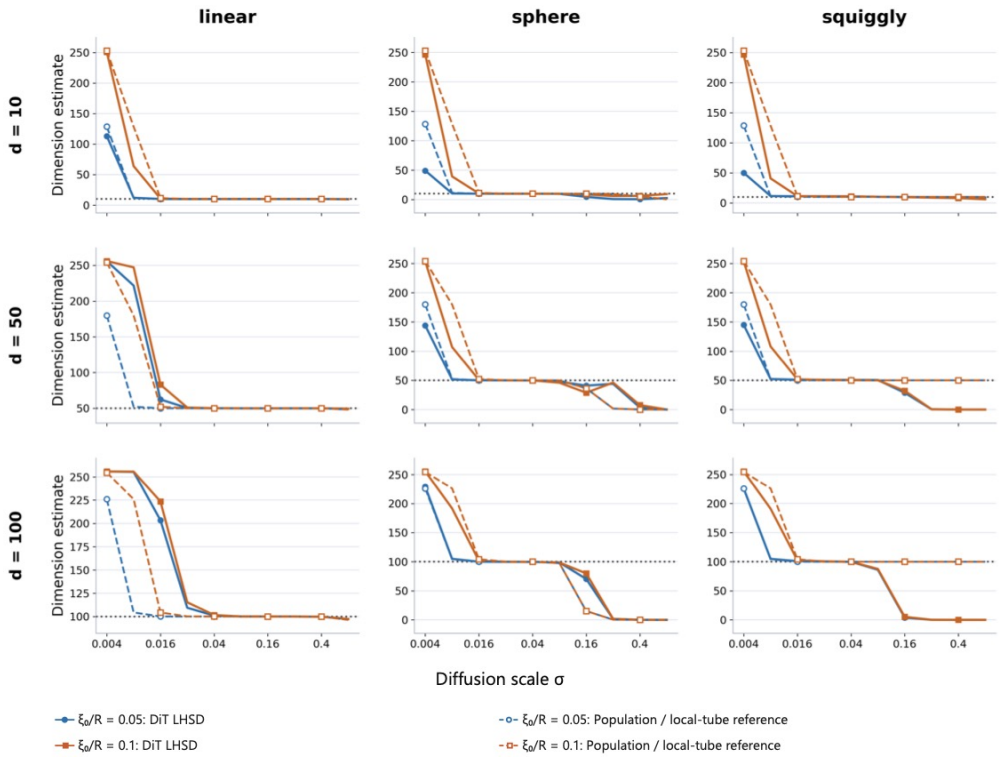


Figure 2: **Affine-calibrated LHSD across diffusion scales.** Columns show geometry; rows show latent dimension. Blue and orange denote $\xi_0/R = 0.05$ and 0.1, respectively. Solid curves with filled circles show query-averaged DiT estimates; dashed curves with open squares show numerical population (linear) or local-tube (sphere and squiggly) references. Estimates and references share affine calibration and queries. Dotted lines indicate the latent dimension d.

The covariance identity then gives $A_{Q,\sigma}(y) - I_D = O_Q(\sigma^{-2})$ in operator norm. Since $\bar{\phi}_\varpi(1) = 0$ and the filter is smooth near one,

$$\|T_{\sigma,\varpi}(Q)\|_{L^\infty(Q)} = O_{Q,\varpi}(\sigma^{-2}) \longrightarrow 0, \qquad \sigma \to \infty.$$

As the diffusion scale grows, all spectral contributions tend to zero. This limit concerns scales large relative to the overall support size; reach alone does not determine its onset.

5 Experiments

We estimate LHSD from the Jacobians of DSM-trained DiT noise predictors and compare it across scales with reference curves, the known latent dimension, and estimates from LIDL and FLIPD.

Setup. We consider Gaussian linear-subspace, sphere, and radially perturbed sphere (squiggly) models with $D = 256$, $R = 4$, $n = 200{,}000$, and latent dimension $d \in \{10, 50, 100\}$. Here R denotes the curved models' reference radius or the linear model's tangential distribution scale. The noncompact linear model is a standard comparison baseline outside our regular-manifold assumptions. Uniform normal-ball noise has radius $\tau = \xi_0$, with $\tau/R \in \{0.05, 0.1\}$. Each condition uses a separately trained DiT and a common protocol. We show LHSD results with $\varpi = 0.05$ at ten scales in $[0.004, 0.6]$. RMSE at $\tau/R = 0.05$ is computed using the same DiT and 128 shared queries $Y \sim Q$, independent of training; it measures query-level error rather than seed variability. Appendix G gives implementation details and additional experiments over diffusion scales σ, noise levels ξ_0, and filter widths ϖ.

Discrepancies in Figure 2 may reflect learned-score and numerical-reference errors, plus local-tube approximation error for curved models.

Table 1: Query-level RMSE at every other scale in the ten-point grid ($n = 200{,}000$, $\tau/R = 0.05$, 128 queries). LHSD is affine-calibrated; bold marks the lowest RMSE for each row and scale.

(d, geometry)	$\sigma = 0.004$			$\sigma = 0.016$			$\sigma = 0.04$			$\sigma = 0.16$			$\sigma = 0.4$		
	LIDL	FLIPD	LHSD	LIDL	FLIPD	LHSD	LIDL	FLIPD	LHSD	LIDL	FLIPD	LHSD	LIDL	FLIPD	LHSD
10, linear	241.990	240.955	**107.370**	162.760	166.283	**0.045**	47.192	55.244	**0.012**	6.016	6.547	**0.007**	3.140	2.735	**0.001**
10, sphere	243.164	242.165	**39.103**	159.243	157.544	**0.014**	40.223	47.546	**0.015**	2.702	**2.052**	5.494	2.100	**0.948**	9.455
10, squiggly	243.045	242.153	**40.030**	165.484	158.020	**1.013**	53.732	48.828	**0.961**	4.752	4.362	**0.164**	2.210	**0.956**	1.298
50, linear	**204.930**	204.945	205.809	189.284	189.338	**15.002**	132.178	132.264	**0.055**	21.566	21.198	**0.058**	5.019	3.555	**0.038**
50, sphere	202.422	202.972	**93.986**	128.199	148.117	**0.023**	35.185	52.408	**0.266**	16.604	17.102	**9.126**	**13.988**	17.060	46.408
50, squiggly	203.672	202.952	**95.139**	154.485	148.253	**0.969**	58.390	51.574	**0.704**	6.530	**3.985**	21.356	**2.458**	5.056	49.959
100, linear	**155.752**	155.754	155.998	151.960	151.914	**104.312**	133.198	132.912	**1.012**	41.051	40.378	**0.043**	8.456	7.367	**0.239**
100, sphere	154.319	154.387	**128.966**	122.852	125.776	**0.065**	49.613	53.710	**0.318**	**9.632**	11.028	29.794	**6.987**	9.099	99.902
100, squiggly	154.690	154.364	**125.773**	129.786	125.054	**0.976**	61.043	52.385	**0.351**	6.078	**1.752**	96.179	**9.935**	25.361	99.986

Interpretation of simulation results. Fine-scale overestimation qualitatively agrees with Theorem 4.1: for $\sigma \ll \xi_0$, even exact-score LHSD can count normal-noise directions. Intermediate-scale agreement with d is consistent with noise averaging that preserves local geometry (Theorem 4.2); Theorem 4.3 formalizes this tradeoff in its local regime. At larger evaluated scales, higher-dimensional sphere and squiggly estimates approach zero, while the linear baseline retains a wider recovery range. This qualitatively agrees with the coarse-scale loss of dimension information in Section 4.3.

Comparison with LIDL and FLIPD. Table 1 reports five scales; all ten are in Appendix G. At $\sigma = 0.004$, all methods incur large errors, although LHSD improves on both baselines for the curved models. On the linear model, larger d shifts LHSD's low-error regime toward larger evaluated scales: at $\sigma = 0.016$, its RMSE rises from 0.045 for $d = 10$ to 104.312 for $d = 100$. At $\sigma = 0.04$, LHSD has the lowest RMSE in all nine conditions (at most 1.012). At $\sigma = 0.4$, FLIPD performs best on both curved models for $d = 10$, whereas LIDL performs best for $d \in \{50, 100\}$. LHSD remains best on the linear model at this scale, but its curved-model errors grow substantially with d. The highlighted scale is descriptive; selection without knowing the LMD remains open.

6 Discussion

Our results distinguish structural identifiability from scale-dependent recovery. Under $2\xi_0 < \tau_0$, the observed law determines a unique model-relative LMD, but LHSD need not reveal it at an arbitrary diffusion scale. The analysis reveals three scale regimes. At fine scales, LHSD can count normal-noise directions, with behavior not determined by the first two conditional noise moments. At intermediate scales, smoothing averages normal variation while preserving local geometry, yielding a sharp worst-case approximation bias of order $(\xi_0/\sigma)^2 + (\sigma/\tau_0)^2$ under the conditions of Theorem 4.3. For each fixed admissible law, excessive smoothing drives LHSD to zero as $\sigma \to \infty$. Our DiT experiments qualitatively support this scale dependence, including accurate recovery at suitable intermediate scales. Ultimately, interpreting a diffusion-based dimension estimate as LMD requires both an identifiable latent target and a scale that preserves its geometric signal.

6.1 Limitations and future work

Several limitations motivate further work. The oracle scale $\sigma^2 \asymp \xi_0\tau_0$ depends on unknown tube width and reach, and data-driven selection must also account for the finite-sample error in equation 10. The minimax guarantees concern the unrestricted empirical-DSM reference estimator, not trained DiTs. Extending these guarantees requires control of score-Jacobian error, motivating connections between diffusion adaptivity [52], statistical guarantees for DiTs [27], and derivative estimation [66]. Our theory also assumes bounded normal noise around a single compact manifold without boundary. Extending the analysis to noncompact manifolds, manifolds with boundary [4], perturbations beyond the bounded normal setting, and mixed-dimensional or stratified supports [1, 63, 36, 39] remains an important direction.

AI USE STATEMENT

In this work, we used generative AI tools for refining theoretical definitions and mathematical statements, assisting in the development and writing of mathematical arguments and proofs, providing feedback on experimental design, writing and revising experimental code, including code for generating synthetic datasets, and interpreting experimental results in relation to the theory. Additionally, we used generative AI tools for manuscript drafting and editing, LaTeX preparation, preliminary literature suggestions and feedback on citation accuracy, and generating and refining the illustration of the statistical setting and estimator.

We have reviewed all AI-assisted work. The authors independently conducted literature searches and made the final selection of references, treating AI-generated suggestions as preliminary leads. They verified cited claims against the original sources and directly checked the assumptions and proof steps of the mathematical results. They also reviewed the geometric depiction and mathematical expressions in the illustration for consistency with the statistical setting and estimator. The authors reviewed the AI-assisted experimental code and rechecked its correctness and the reproducibility of the reported experiments.

We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

Throughout the work, we focus on theoretical analysis and experiments on synthetic data. We do not identify any specific ethical concerns arising from this study.

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, the anonymous supplementary materials on OpenReview include a ZIP archive containing the complete experimental source code, software environment specifications, configuration files containing all hyperparameters and random seeds, and scripts for data generation, model training, estimator evaluation, and generation of the reported tables and figures. All experiments use synthetically generated data and require no external datasets. Section 5 and Appendix G provide the experimental setup and implementation details needed to use these materials.

REFERENCES

[1] Eddie Aamari and Clément Berenfeld. A theory of stratification learning. *arXiv preprint arXiv:2405.20066*, 2024.

[2] Eddie Aamari and Clément Levrard. Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, 47(1):177–204, 2019.

[3] Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Estimating the reach of a manifold. *Electronic Journal of Statistics*, 13(1):1359–1399, 2019.

[4] Eddie Aamari, Catherine Aaron, and Clément Levrard. Minimax boundary estimation and estimation with boundary. *Bernoulli*, 29(4):3334–3368, 2023.

[5] Armen Aghajanyan, Sonal Gupta, and Luke Zettlemoyer. Intrinsic dimensionality explains the effectiveness of language model fine-tuning. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: long papers)*, pp. 7319–7328, 2021.

[6] Yariv Aizenbud and Barak Sober. Non-parametric estimation of manifolds from noisy data. *arXiv preprint arXiv:2105.04754*, 2021.

[7] Alessio Ansuini, Alessandro Laio, Jakob H Macke, and Davide Zoccolan. Intrinsic dimension of data representations in deep neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.

[8] Adam Block, Zeyu Jia, Yury Polyanskiy, and Alexander Rakhlin. Intrinsic dimension estimation using wasserstein distance. *Journal of Machine Learning Research*, 23(313):1–37, 2022.

[9] Bradley CA Brown, Anthony L. Caterini, Brendan Leigh Ross, Jesse C Cresswell, and Gabriel Loaiza-Ganem. Verifying the union of manifolds hypothesis for image data. In *The Eleventh International Conference on Learning Representations*, 2023.

[10] Francesco Camastra and Antonino Staiano. Intrinsic dimension estimation: Advances and open problems. *Information Sciences*, 328:26–41, 2016.

[11] Saptarshi Chakraborty, Quentin Berthet, and Peter L Bartlett. Generalization properties of score-matching diffusion models for intrinsically low-dimensional data. *arXiv preprint arXiv:2603.03700*, 2026.

[12] Junhao Chen, Ruowei Li, and Zhigang Yao. Estimation of riemannian quantities from noisy data via density derivatives. *arXiv preprint arXiv:2603.27171*, 2026.

[13] Jose A Costa and Alfred O Hero. Geodesic entropic graphs for dimension and entropy estimation in manifold learning. *IEEE Transactions on Signal Processing*, 52(8):2210–2221, 2004.

[14] Ju L Daletskii and Selim Grigorievich Krein. Integration and differentiation of functions of hermitian operators and applications to the theory of perturbations. *AMS Translations (2)*, 47 (1-30):10–1090, 1965.

[15] Antonio Di Noia, Iuri Macocco, Aldo Glielmo, Alessandro Laio, and Antonietta Mira. Scale adaptive and robust intrinsic dimension estimation via optimal neighbourhood identification. *Scientific Reports*, 2026.

[16] Vincent Divol. Measure estimation on manifolds: an optimal transport approach. *Probability Theory and Related Fields*, 183(1):581–647, 2022.

[17] Elena Facco, Maria d'Errico, Alex Rodriguez, and Alessandro Laio. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Scientific reports*, 7(1):12140, 2017.

[18] Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93 (3):418–491, 1959.

[19] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.

[20] Keinosuke Fukunaga and David R Olsen. An algorithm for finding intrinsic dimensionality of data. *IEEE Transactions on computers*, C-20(2):176–183, 1971.

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

[22] Anand Jerry George and Nicolas Macris. Asymptotic learning curves for diffusion models with random features score and manifold data. *arXiv preprint arXiv:2603.22962*, 2026.

[23] Daniele Granata and Vincenzo Carnevale. Accurate estimation of the intrinsic dimension using graph distances: Unraveling the geometric complexity of datasets. *Scientific reports*, 6(1): 31377, 2016.

[24] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, 2002.

[25] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

[26] Christian Horvat and Jean-Pascal Pfister. On gauge freedom, conservativity and intrinsic dimensionality estimation in diffusion models. In *International Conference on Learning Representations*, volume 2024, pp. 48710–48726, 2024.

[27] Jerry Yao-Chieh Hu, Weimin Wu, Yi-Chen Lee, Yu-Chao Huang, Minshuo Chen, and Han Liu. On statistical rates of conditional diffusion transformers: Approximation, estimation and minimax optimality. In *International Conference on Learning Representations*, volume 2025, pp. 12591–12675, 2025.

[28] Michael F Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, 18(3):1059–1076, 1989.

[29] 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*, volume 37, pp. 38307–38354, 2024.

[30] Balázs Kégl. Intrinsic dimension estimation using packing numbers. *Advances in neural information processing systems*, 15, 2002.

[31] Arlene K. H. Kim and Harrison H. Zhou. Tight minimax rates for manifold estimation under hausdorff loss. *Electronic Journal of Statistics*, 9:1562–1582, 2015.

[32] Jisu Kim, Alessandro Rinaldo, and Larry Wasserman. Minimax rates for estimating the dimension of a manifold. *Journal of Computational Geometry*, 10(1):42–95, 2019.

[33] Jisu Kim, Jaehyeok Shin, Alessandro Rinaldo, and Larry Wasserman. Uniform convergence rate of the kernel density estimator adaptive to intrinsic volume dimension. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3398–3407. PMLR, 2019.

[34] Kin Kwan Leung, Rasa Hosseinzadeh, and Gabriel Loaiza-Ganem. On convolutions, intrinsic dimension, and diffusion models. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856.

[35] Elizaveta Levina and Peter Bickel. Maximum likelihood estimation of intrinsic dimension. *Advances in neural information processing systems*, 17, 2004.

[36] Xin Li and Anand Sarwate. Unraveling the localized latents: Learning stratified manifold structures in llm embedding space with sparse mixture-of-experts. *arXiv preprint arXiv:2502.13577*, 2025.

[37] Anna V Little, Yoon-Mo Jung, and Mauro Maggioni. Multiscale estimation of intrinsic dimensionality of data sets. In *AAAI fall symposium: manifold learning and its applications*, volume 9, pp. 04, 2009.

[38] Xingjun Ma, Bo Li, Yisen Wang, Sarah M. Erfani, Sudanthi Wijewickrema, Grant Schoenebeck, Michael E. Houle, Dawn Song, and James Bailey. Characterizing adversarial subspaces using local intrinsic dimensionality. In *International Conference on Learning Representations*, 2018.

[39] Randy Martinez, Rong Tang, and Lizhen Lin. A deep generative approach to stratified learning. *arXiv preprint arXiv:2604.10650*, 2026.

[40] Dengyao Mo and Samuel H Huang. Fractal-based intrinsic dimension estimation and its application in dimensionality reduction. *IEEE Transactions on Knowledge and Data Engineering*, 24(1):59–71, 2010.

[41] Hariharan Narayanan and Sanjoy Mitter. Sample complexity of testing the manifold hypothesis. *Advances in neural information processing systems*, 23, 2010.

[42] Genki Osada. Local hessian spectral filtering for robust intrinsic dimension estimation. In *Forty-third International Conference on Machine Learning*, 2026.

[43] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182. IEEE, 2023.

[44] Phil Pope, Chen Zhu, Ahmed Abdelkader, Micah Goldblum, and Tom Goldstein. The intrinsic dimension of images and its impact on learning. In *International Conference on Learning Representations*, 2021.

[45] Nikita Puchkin and Vladimir Spokoiny. Structure-adaptive manifold estimation. *Journal of Machine Learning Research*, 23(40):1–62, 2022.

[46] Benjamin Matthias Ruppik, Julius von Rohrscheidt, Carel van Niekerk, Michael Heck, Renato Vukovic, Shutong Feng, Hsien chin Lin, Nurul Lubis, Bastian Rieck, Marcus Zibrowius, and Milica Gasic. Less is more: Local intrinsic dimensions of contextual language models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[47] Jaehee Seo, Wontae Jeong, and Jisu Kim. Minimax lower bound for estimating diffusion-based local intrinsic dimension. *arXiv preprint arXiv:2609.04822*, 2026.

[48] Bartlomiej Sobieski, Matthew Tivnan, Dawid Płudowski, Michał Jan Włodarczyk, Pengfei Jin, Przemyslaw Biecek, and Quanzheng Li. Local intrinsic dimension unveils hallucinations in diffusion models. *arXiv preprint arXiv:2605.05026*, 2026.

[49] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

[50] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

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

[52] Rong Tang and Yun Yang. Adaptivity of diffusion models to manifold structures. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 1648–1656. PMLR, 2024.

[53] Piotr Tempczyk, Rafał Michaluk, Lukasz Garncarek, Przemysław Spurek, Jacek Tabor, and Adam Golinski. Lidl: Local intrinsic dimension estimation using approximate likelihood. In *International Conference on Machine Learning*, pp. 21205–21231. PMLR, 2022.

[54] Piotr Tempczyk, Łukasz Garncarek, Dominik Filipiak, and Adam Kurpisz. A wiener process perspective on local intrinsic dimension estimation methods. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20859–20866, 2025.

[55] Piotr Tempczyk, Dominik Filipiak, Łukasz Garncarek, Ksawery Smoczynski, and Adam Kurpisz. Why we need new benchmarks for local intrinsic dimension estimation. In *International Conference on Learning Representations*, 2026.

[56] Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, 2009.

[57] Eduard Tulchinskii, Kristian Kuznetsov, Laida Kushnareva, Daniil Cherniavskii, Sergey Nikolenko, Evgeny Burnaev, Serguei Barannikov, and Irina Piontkovskaya. Intrinsic dimension estimation for robust detection of ai-generated texts. *Advances in Neural Information Processing Systems*, 36:39257–39276, 2023.

[58] Shashanka Ubaru, Jie Chen, and Yousef Saad. Fast estimation of tr(f(a)) via stochastic lanczos quadrature. *SIAM Journal on Matrix Analysis and Applications*, 38(4):1075–1099, 2017.

[59] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[60] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

[61] Jonathan Von Brünken, Michael E Houle, and Arthur Zimek. Intrinsic dimensional outlier detection in high-dimensional data. *NII Tech. Rep.*, 2015(3):1–12, 2015.

[62] Qizhou Wang, Sarah M Erfani, Christopher Leckie, and Michael E Houle. A dimensionality-driven approach for unsupervised out-of-distribution detection. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, pp. 118–126. SIAM, 2021.

[63] Yi Wang and Zhiren Wang. CW complex hypothesis for image data. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 51596–51613. PMLR, 2024.

[64] Jingda Wu and Changxiao Cai. Diffusion models are statistically optimal for learning low-dimensional multi-modal distributions. In *Forty-third International Conference on Machine Learning*, 2026.

[65] Konstantin Yakovlev and Nikita Puchkin. Generalization error bound for denoising score matching under relaxed manifold assumption. In *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 5824–5891. PMLR, 2025.

[66] Konstantin Yakovlev, Anna Markovich, and Nikita Puchkin. Implicit score matching meets denoising score matching: improved rates of convergence and log-density hessian estimation. *arXiv preprint arXiv:2512.24378*, 2025.

[67] Eric Yeats, Aaron Jacobson, Darryl Hannan, Yiran Jia, Timothy Doster, Henry Kvinge, and Scott Mahan. A connection between score matching and local intrinsic dimension. In *NeurIPS 2025 Workshop on Structured Probabilistic Inference & Generative Modeling*, 2025.

APPENDIX

In the Appendix, we collect the skipped related works, notation and model classes, theoretical proofs of main results, and experimental details.

Section	Main content	Location
Related works	ID estimation, diffusion-based methods, and geometric inference	Appendix A
Notation and model classes	notation, reach, smoothness, and uniform regularity bounds	Appendix B
Structural identifiability	support proximity, projection and homology arguments, and the boundary counterexample	Appendix C
Minimax theory	the lower bound, empirical DSM estimator, and matching upper bound	Appendix D
Fine-scale noise sensitivity	matched-moment normal fibers with different local support dimensions	Appendix E
LHSD approximation and sharpness	covariance and spectral bounds, intermediate-scale recovery, and the sharp scale tradeoff	Appendix F
Experimental details	synthetic models, DiT training, reference construction, and additional results	Appendix G

A RELATED WORK

Direct geometric estimators. Numerous methods estimate dimension directly from geometric statistics of the observed sample. These include local covariance spectra [20], packing numbers [30], entropic graphs and graph-based distances [13, 23], nearest-neighbor scaling laws [35, 17], and fractal scaling [40]. Wasserstein-based estimators instead exploit how distances between empirical distributions vary with sample size [8]. These approaches use different geometric summaries and regularity assumptions. The survey of Camastra & Staiano [10] reviews statistical and computational tradeoffs in intrinsic-dimension estimation.

Observation noise and scale selection. Noise and neighborhood size affect which geometric features a dimension estimator detects. Little et al. [37] analyze local singular values across scales to distinguish noise, tangent variation, and nonlinear geometry. The TWO-NN estimator of Facco et al. [17] uses first- and second-neighbor distances; its block analysis probes dimension as the effective sampling scale changes. More recently, Di Noia et al. [15] jointly adapt dimension estimates and pointwise neighborhood sizes to identify regions of approximately constant density. These studies establish scale dependence as a central issue for geometric estimators. In a diffusion-based analysis, the smoothing scale must likewise be distinguished from the magnitude and geometry of observation noise.

Generative model-based estimators. Generative models offer an alternative route to dimension estimation through learned densities and scores. Score-based generative modeling and diffusion models provide distributions at multiple Gaussian smoothing levels [49, 25, 50], while DSM relates denoising to estimation of the corresponding scores [60]. LIDL estimates local dimension from the scale dependence of convolved likelihoods [53]; FLIPD uses a density-evolution identity to evaluate a local dimension field [29]. LHSD instead applies a smooth spectral filter to the normalized negative log-density Hessian [42]. Yeats et al. [67] connect LID to the denoising score-matching loss, motivate a loss-based estimator, and relate implicit score matching to FLIPD. The normal alignment of low-noise scores provides another route to recovering manifold dimension [51]. Leung et al. [34] establish small-noise convolution identities supporting FLIPD on smooth embedded manifolds and separated unions of manifolds. For computational efficiency, LHSD uses stochastic Lanczos quadrature to approximate its spectral trace [42, 58], building on randomized trace estimation [28].

Finite-scale effects and learned-score geometry. Tempczyk et al. [54] relate likelihood-scaling and diffusion-based dimension estimators and analyze how nonuniform density, boundaries, and nearby support components affect their finite-scale values. Horvat & Pfister [26] study gauge freedom and conservativity, showing why properties sufficient for diffusion sampling need not determine the local geometric information used for dimension estimation. Tempczyk et al. [55] demonstrate sensitivity to representations and architectures and propose controlled benchmarks with known geometric changes. Osada [42] reports improved LHSD robustness on several high-dimensional and nonlinear benchmarks, with its spectral interpretation relying on a useful tangent–normal separation. Under uniform latent sampling on smooth compact manifolds with positive reach and isotropic Gaussian observation noise, Chen et al. [12] study the recovery of tangent spaces, intrinsic dimension, and curvature from density derivatives in the small-noise regime. Together, these studies motivate distinguishing the finite-scale population functional from errors in its learned evaluation.

Statistical theory for dimension and geometry. Minimax bounds for estimating manifold dimension have been established for exactly manifold-supported distributions [32], while manifold-hypothesis tests assess compatibility with low-dimensional geometric structure [41, 19]. Related theory studies manifold estimation under Hausdorff loss [31, 21], reach and local geometric estimation [3, 2], and manifold recovery from noisy observations [6, 45]. Reach, introduced by Federer [18], controls the scale on which nearest-point projection is unique. In tubular models, Divol [16] studies measure estimation and notes that distinct admissible latent signal–noise decompositions can generate the same observed law. This representation ambiguity motivates distinguishing uniqueness of the latent manifold from identifiability of its dimension. Intrinsic geometry also controls statistical difficulty beyond dimension estimation: Kim et al. [33] derive uniform kernel-estimation rates adaptive to intrinsic volume dimension. For a diffusion-based target, Seo et al. [47] establish a minimax lower bound for estimating the finite-scale FLIPD field.

Statistical learning of scores and Hessians. Tang & Yang [52] establish adaptation of diffusion-based distribution estimation to low-dimensional manifold structure. Yakovlev & Puchkin [65] derive approximation and generalization bounds for DSM under nonparametric Gaussian-mixture observation models, allowing samples to deviate from a manifold. Wu & Cai [64] obtain near-optimal Wasserstein-1 distribution-learning rates for unions of low-dimensional subspaces, while Chakraborty et al. [11] derive Wasserstein-p generalization bounds governed by an intrinsic Wasserstein dimension under finite-moment conditions. In a random-feature score model, George & Macris [22] study asymptotic learning curves and the role of manifold structure in score estimation. Transformer architectures [59] have also been incorporated into diffusion models [43]; Hu et al. [27] analyze approximation and estimation rates of conditional diffusion transformers under smoothness and latent-subspace assumptions. Since LHSD depends on the score Jacobian, its learned evaluation requires derivative accuracy in addition to score accuracy. Yakovlev et al. [66] establish log-density Hessian estimation bounds for implicit and denoising score matching under Gaussian-convolved low-dimensional generator models and smoothness constraints on the score class. These results motivate the derivative control needed for learned LHSD.

Heterogeneous and stratified geometry. Empirical studies motivate geometric models with spatially varying dimension: Brown et al. [9] investigate unions of manifolds for image data, Wang & Wang [63] propose a CW complex hypothesis allowing dimension changes within connected components, and Li & Sarwate [36] examine stratified structure in language-model embeddings using sparse mixture-of-experts models. Aamari & Berenfeld [1] develop statistical guarantees for recovering stratified mixtures, including layer dimensions and localized tangent structure. Martinez et al. [39] develop likelihood- and diffusion-based methods for stratified learning under isotropic Gaussian observation noise. Their dimension-consistency analysis treats regular points and extends to vanishing Gaussian noise controlled by the diffusion scale, under score-approximation conditions. This provides an important noisy-data comparison: their Gaussian observation model and our bounded normal-noise class impose different restrictions on the perturbation distribution.

Applications of intrinsic dimension. ID has been used to characterize image datasets [44], neural representations [7], and the training dynamics of contextual language models [46]. Applications include dimensionality reduction [40], outlier and out-of-distribution detection [61, 62], characterization of adversarial examples [38], detection of AI-generated text [57], and analysis of structural

hallucinations in diffusion models [48]. A related parameter-space notion of intrinsic dimension explains aspects of language-model fine-tuning [5]; its target is the dimension of a useful parameter subspace rather than the dimension of the data support.

Relation to the present work. We study a model-relative latent dimension under bounded normal perturbations whose conditional laws may depend on the latent point and may be discrete, asymmetric, or noncentered. Our analysis first determines when this dimension is identifiable from the observed law, then separates estimation of a finite-scale LHSD field from its approximation to the latent dimension. Under the joint conditions of Theorems 3.1 and 3.2, the field has minimax squared-risk rate $\sigma_n^4/(n\sigma_n^d)$, attained by the nonparametric empirical DSM reference estimator. Its comparison with the FLIPD lower bound concerns different finite-scale targets; LMD recovery additionally depends on approximation bias. The recovery analysis characterizes an intermediate scale and shows that matching conditional noise means and covariances does not determine the fine-scale LHSD limit. Thus, identifiability, finite-sample precision, and scale-dependent geometric bias address distinct parts of dimension recovery from noisy observations.

B Notation and model classes

B.1 Notation

Probability. For every positive integer m, write $[m] := \{1, \dots, m\}$ and $[m]_0 := \{0, \dots, m\}$; also set $[0]_0 := \{0\}$. Throughout, Q denotes a Borel probability measure on $\mathbb{R}^D$, and $\mathcal{L}(U) := \mathbb{P} \circ U^{-1}$ denotes the Borel probability measure induced by a random variable U. The topological support of Q is

$$\operatorname{supp}(Q) := \{y \in \mathbb{R}^D : Q(\mathbb{B}_D(y, r)) > 0 \text{ for every } r > 0\}.$$

Here $\mathbb{B}_D(y, r)$ is an open Euclidean ball. $\mathbb{E}_Q$ and $\mathbb{P}_Q$ denote expectation and probability under Q. If $Y_1, \dots, Y_n$ are independent with common probability measure Q, then $Q^n := Q^{\otimes n}$ is their product probability measure; $\mathbb{E}_{Q^n}$ and $\mathbb{P}_{Q^n}$ refer to this measure. The Dirac measure δ_y is the unit mass at y, so $n^{-1}\sum_{i=1}^n \delta_{Y_i}$ is the empirical probability measure. "Q-almost everywhere" means outside a set of Q-measure zero, and "almost surely" (a.s.) means with probability one. A Borel probability kernel $x \mapsto K_x$ is a family of probability measures for which $x \mapsto K_x(B)$ is measurable for every Borel set B; in the tubular model it represents the conditional law of ξ given $X = x$. A Rademacher variable takes each of -1 and 1 with probability $1/2$. For probability measures P, Q on the same space, relative entropy and total variation distance use the conventions

$$D_{\mathrm{KL}}(P, Q) := \begin{cases} \displaystyle\int \log\left(\frac{dP}{dQ}\right) dP, & P \ll Q, \\ +\infty, & \text{otherwise}, \end{cases} \qquad \|P - Q\|_{\mathrm{TV}} := \sup_B |P(B) - Q(B)|,$$

where $P \ll Q$ means that every Q-null set is P-null and the supremum is over measurable sets. For a finite signed measure μ, $|\mu|$ is its total variation measure; in particular, $|P-Q|(\mathbb{R}^D) = 2\|P-Q\|_{\mathrm{TV}}$ for laws on $\mathbb{R}^D$. The weighted norms $\int(1 + \|t\|^r)\, d|\mu|(t)$ used below are signed-measure norms, distinct from this probability-distance convention. The bounded tubular class $\mathcal{P}_d(\xi_0, \tau_0, L_{\mathcal{T}}, L_{\mathcal{F}})$ and the LMD $d_{\mathrm{LMD}}(Q; \xi_0, \tau_0, L_{\mathcal{T}}, L_{\mathcal{F}})$ are defined in Definitions 2.1–2.2. The spectral threshold $\varpi \in (0, 1)$ is fixed throughout all limits and risk bounds.

Linear algebra and function norms. The unsubscripted norm is Euclidean. For a linear subspace $V \subseteq \mathbb{R}^D$, P_V denotes the orthogonal projector onto V. For a matrix A, $\|A\|_{\mathrm{op}}$, $\|A\|_F$, $\operatorname{tr}(A)$, and I_k denote its operator norm, Frobenius norm, trace, and the $k \times k$ identity matrix, respectively. For symmetric matrices, $A \preceq B$ means that $B - A$ is positive semidefinite (the Loewner order). For a vector-valued map m, $\nabla \mathsf{m}$ denotes its input Jacobian. For a measurable matrix field H and $1 \le p < \infty$, set

$$\|H\|_{L^p(Q;F)} := \left(\int \|H(y)\|_F^p \, dQ(y)\right)^{1/p}, \qquad \|H\|_{L^\infty(Q;F)} := \operatorname*{ess\,sup}_{y \sim Q} \|H(y)\|_F. \tag{11}$$

The essential supremum is the least upper bound valid outside a Q-null set. For scalar fields we omit "; F". Pointwise tensor sup norms use the Euclidean-induced operator norm. Scalar functions of symmetric matrices are understood spectrally: if $A = U \operatorname{diag}(\lambda_j) U^\top$, then $h(A) =$

$U\,\mathrm{diag}(h(\lambda_j))U^\top$. A spectral projector is the orthogonal projector onto the sum of eigenspaces in a specified eigenvalue cluster; the spectral gap between two clusters is the minimum distance between their eigenvalues.

Asymptotic conventions. For nonnegative quantities, $a \lesssim b$ means $a \le Cb$ for a constant depending only on the fixed model envelopes; $a \gtrsim b$ reverses the inequality and $a \asymp b$ means both. A subscript, as in $\lesssim_\varpi$, permits additional dependence on that parameter. For positive sequences, $a_n \ll b_n$ means $a_n/b_n \to 0$. For real scalars, $a \wedge b$ and $a \vee b$ denote the smaller and larger of a, b, respectively.

B.2 Smoothness and geometric classes

Smoothness classes. Let $U \subset \mathbb{R}^m$ be open, let E be a finite-dimensional normed vector space, and let $r \in \mathbb{N}_0$. The class $C^r(U;E)$ consists of maps $h : U \to E$ with continuous Fréchet derivatives $D^j h$ for $0 \le j \le r$, where $D^0 h = h$. For such maps with bounded derivatives, define

$$\|h\|_{C^r(U;E)} := \max_{0\le j\le r} \|D^j h\|_\infty. \tag{12}$$

For $j \ge 1$, the derivative $D^j h$ is equipped with its induced multilinear operator norm. We omit E when the target space is clear. The notation $C_c^\infty(U)$ denotes infinitely differentiable scalar functions whose support is a compact subset of U.

For an integer $r \ge 1$, an embedded d-dimensional C^r manifold $\mathcal{M} \subset \mathbb{R}^D$ is a set that is locally parametrized by C^r maps of full rank: for every $x \in \mathcal{M}$, there are an open set $U_x \subset \mathbb{R}^d$ and a C^r map $g_x : U_x \to \mathbb{R}^D$ that is a homeomorphism onto a relatively open neighborhood of x in $\mathcal{M}$ and satisfies $\mathrm{rank}(Dg_x(u)) = d$ for every $u \in U_x$. Using open subsets of $\mathbb{R}^d$ at every point means the manifold has no boundary. An atlas is a collection of such parametrizations whose images cover the manifold. Equivalently, after shrinking the neighborhood and identifying tangent coordinates with $T_x\mathcal{M} \simeq \mathbb{R}^d$, the manifold is a C^r normal graph: for some $r_x > 0$ and a C^r map G_x,

$$\begin{aligned} &\Psi_x(v) = x + v + G_x(v), && v \in \mathbb{B}_{T_x\mathcal{M}}(0, r_x), \\ &G_x(v) \in N_x\mathcal{M}, && G_x(0) = 0, \qquad DG_x(0) = 0. \end{aligned} \tag{13}$$

For an atlas $\{g_j : \mathbb{B}_d(0, r_0) \to \mathcal{M}\}_{j\in[J_0]}$ whose charts are C^r and satisfy the uniform conditions in Definition B.1, the class $C^r(\mathcal{M})$ consists of scalar functions $f : \mathcal{M} \to \mathbb{R}$ such that $f \circ g_j \in C^r(\mathbb{B}_d(0, r_0))$ for every $j \in [J_0]$. The chartwise norm is

$$\|f\|_{C^r(\mathcal{M})} := \max_{j\in[J_0]} \|f \circ g_j\|_{C^r(\mathbb{B}_d(0,r_0))}. \tag{14}$$

The chartwise norm is evaluated using the witnessing atlas included in the admissible representation. It is an envelope used to make the estimates uniform, not a claim that equation 14 is a canonical intrinsic norm. The common metric-distortion and derivative bounds on the admissible atlases make the resulting proof constants comparable across the class. The function class associated with Assumption 2.2 is specified in Definition B.2 below.

Geometry and reach. Write $\mathbb{B}_k(x, r)$ for the open Euclidean ball in $\mathbb{R}^k$ and abbreviate $\mathbb{B}(x, r) := \mathbb{B}_D(x, r)$. For a linear subspace $E \subset \mathbb{R}^D$, let $\mathbb{B}_E(0, r) := \{v \in E : \|v\| < r\}$. Furthermore, $\mathbb{S}^k(r) := \{u \in \mathbb{R}^{k+1} : \|u\| = r\}$ is the radius-r sphere and $\mathbb{S}^k := \mathbb{S}^k(1)$. We use k-dimensional Hausdorff measure $\mathcal{H}^k$ normalized to agree with Lebesgue measure on each k-dimensional Euclidean plane; $\mathcal{H}^0$ is counting measure. For a measure μ, restriction to a measurable set A means $(\mu \restriction A)(B) := \mu(A \cap B)$. The ambient Euclidean inner product induces the Riemannian metric on $\mathcal{M}$, whose volume measure is $\mathrm{vol}_\mathcal{M} = \mathcal{H}^d \restriction \mathcal{M}$. For a chart parametrization $g : U \subset \mathbb{R}^d \to \mathcal{M}$, its volume Jacobian is $J_g(u) := \sqrt{\det(Dg(u)^\top Dg(u))}$: for every nonnegative measurable h,

$$\int_{g(U)} h(x)\, d\,\mathrm{vol}_\mathcal{M}(x) = \int_U h(g(u)) J_g(u)\, du.$$

Thus $d\mathcal{L}(X) = f\, d\,\mathrm{vol}_\mathcal{M}$ means that, for every Borel set $B \subseteq \mathbb{R}^D$,

$$\mathbb{P}(X \in B) = \int_{B\cap\mathcal{M}} f(x)\, d\,\mathrm{vol}_\mathcal{M}(x).$$

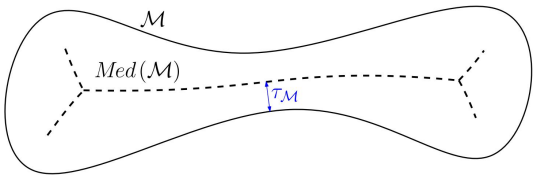


Figure 3: The medial axis $\mathrm{Med}(\mathcal{M})$ consists of points having multiple nearest points on $\mathcal{M}$. The shortest distance from $\mathcal{M}$ to this axis is $\mathrm{reach}(\mathcal{M})$. A narrow bottleneck brings the medial axis closer to $\mathcal{M}$ and therefore reduces the reach.

For a nonempty closed set $\mathcal{M} \subset \mathbb{R}^D$, define

$$\mathrm{dist}(y, \mathcal{M}) := \inf_{x \in \mathcal{M}} \|y - x\|.$$

Its *medial axis* is the set of points having more than one nearest point on $\mathcal{M}$:

$$Med(\mathcal{M}) := \left\{ y \in \mathbb{R}^D : \exists x_1 \neq x_2 \in \mathcal{M} \text{ such that } \|y - x_1\| = \|y - x_2\| = \mathrm{dist}(y, \mathcal{M}) \right\}. \quad (15)$$

The *reach* of $\mathcal{M}$ is its distance to the medial axis [18]:

$$\mathrm{reach}(\mathcal{M}) := \inf_{y \in Med(\mathcal{M})} \mathrm{dist}(y, \mathcal{M}), \quad (16)$$

where the infimum is $+\infty$ if $Med(\mathcal{M})$ is empty. Equivalently,

$$\mathrm{reach}(\mathcal{M}) = \sup \left\{ r > 0 : \forall y \in \mathbb{R}^D,\ \mathrm{dist}(y, \mathcal{M}) < r \Longrightarrow \exists! x \in \mathcal{M} :\ \|y - x\| = \mathrm{dist}(y, \mathcal{M}) \right\}. \quad (17)$$

Figure 3 illustrates the medial axis and the reach of the closed set.

At $x \in \mathcal{M}$, $T_x\mathcal{M}$ and $N_x\mathcal{M} = (T_x\mathcal{M})^\perp$ are the tangent and normal spaces. Let $P_T(x)$ and $P_N(x)$ denote their ambient orthogonal projectors, so $P_N(x) = I_D - P_T(x)$. Whenever $\mathrm{dist}(y, \mathcal{M}) < \mathrm{reach}(\mathcal{M})$, denote the unique nearest point by $\pi_{\mathcal{M}}(y)$. On this tubular neighborhood, set $P_T(y) := P_T(\pi_{\mathcal{M}}(y))$ and $P_N(y) := I_D - P_T(y)$. The normal bundle is $N\mathcal{M} := \{(x, u) : x \in \mathcal{M},\ u \in N_x\mathcal{M}\}$. Its fiber at x is $\{x\} \times N_x\mathcal{M}$, identified with the vector space $N_x\mathcal{M}$; the corresponding affine normal space is $x + N_x\mathcal{M}$. The offset map $(x, u) \mapsto x + u$ produces observations from bundle points. For $\|u\| < \mathrm{reach}(\mathcal{M})$, one has $\pi_{\mathcal{M}}(x + u) = x$.

For a differentiable scalar function h on $\mathcal{M}$, its Riemannian gradient is the tangent vector characterized by $\langle \nabla_{\mathcal{M}} h(x), v \rangle = dh_x(v)$ for $v \in T_x\mathcal{M}$. Writing $\nabla^{\mathcal{M}}$ for the Levi–Civita covariant derivative of the induced metric, the Hessian of $h \in C^2(\mathcal{M})$ is $\nabla^2_{\mathcal{M}} h(x)[v, w] := \langle \nabla^{\mathcal{M}}_v(\nabla_{\mathcal{M}} h), w \rangle$. For an embedded manifold, $\nabla^{\mathcal{M}}_v V$ is the tangent projection of the ambient directional derivative of a tangent vector field V. A geodesic has zero intrinsic acceleration, $\nabla^{\mathcal{M}}_{\gamma'} \gamma' = 0$; it has unit speed when $\|\gamma'(t)\| = 1$. Compact boundaryless Riemannian manifolds are geodesically complete, meaning every geodesic extends to all real times. For subspaces V, W of the same positive dimension, their principal angles are the arccosines of the singular values of $B_V^\top B_W$, where B_V, B_W have orthonormal columns spanning the respective subspaces. Writing $\theta_{\max}$ for the largest angle, we use

$$\|P_V - P_W\|_{\mathrm{op}} = \|(I_D - P_V)P_W\|_{\mathrm{op}} = \sin\theta_{\max}.$$

For a C^2 manifold, the second fundamental form is the symmetric normal-valued bilinear map $\mathrm{II}_x(v, w) := D^2 G_x(0)[v, w]$ in the tangent–normal graph equation 13. At C^3 regularity, as in our model, this convention gives

$$G_x(v) = \tfrac{1}{2}\,\mathrm{II}_x(v, v) + O(\|v\|^3).$$

For each unit normal $\nu \in N_x\mathcal{M}$, the shape operator $S_{x,\nu}$ is defined by $\langle S_{x,\nu} v, w \rangle = \langle \mathrm{II}_x(v, w), \nu \rangle$. Its eigenvalues are the principal curvatures in the normal direction ν; this convention applies in arbitrary codimension.

For a C^2 embedded manifold with positive reach, Federer's secant characterization gives

$$\mathrm{reach}(\mathcal{M}) = \inf_{\substack{p, q \in \mathcal{M} \\ p \neq q}} \frac{\|p - q\|^2}{2\,\mathrm{dist}(q - p, T_p\mathcal{M})}. \quad (18)$$

Consequently, if $\mathrm{reach}(\mathcal{M}) \geq \tau_0$, then

$$\|P_N(p)(q-p)\| \leq \frac{\|p-q\|^2}{2\tau_0}, \qquad p, q \in \mathcal{M}. \tag{19}$$

We also use

$$\mathcal{M}^r := \{y : \mathrm{dist}(y, \mathcal{M}) \leq r\}, \qquad d_{\mathrm{H}}(A, B) := \left(\sup_{a\in A} \mathrm{dist}(a, B)\right) \vee \left(\sup_{b\in B} \mathrm{dist}(b, A)\right). \tag{20}$$

The lower bound $\mathrm{reach}(\mathcal{M}) \geq \tau_0$ in Assumption 2.1 provides both local and global geometric control. Locally, it bounds the principal curvatures in magnitude by τ_0^{-1} and supplies tangent–normal graph neighborhoods on a radius proportional to τ_0. Globally, it prevents distinct parts of $\mathcal{M}$ from approaching closely enough to make nearest-point projection ambiguous below that scale. Consequently, if $\xi_0 < \tau_0$, the tubular observation $Y = X + \xi$ satisfies $\pi_{\mathcal{M}}(Y) = X$. The proofs use this projection to identify tangent and normal blocks, and use the same reach scale to exclude geometrically unrelated sheets from the Gaussian localization region. The $C^{\beta+1}$ chart bounds control higher-order Taylor remainders, whereas the reach lower bound determines the uniform spatial scale on which these local representations remain valid.

Definition B.1 (Regular embedded manifold class)**.** *Fix $D \geq 2$, $d \in [D-1]$, an integer $\beta \geq 2$, and $\tau_0 > 0$. Let*

$$L_{\mathcal{T}} := (R_0, V_0, J_0, r_0, c_g, C_g, L_g)$$

collect fixed positive constants, with $J_0 \in \mathbb{N}$. Here R_0 and V_0 control global size, J_0 and r_0 provide a finite common chart scale, c_g and C_g prevent metric degeneration, and L_g bounds chart derivatives. We write $\mathcal{M} \in \mathcal{C}^{\beta+1}_{d,L_{\mathcal{T}},\tau_0}(\mathbb{R}^D)$ if $\mathcal{M} \subset \mathbb{R}^D$ is a compact, connected, boundaryless, embedded d-dimensional $C^{\beta+1}$ manifold satisfying

$$\mathrm{reach}(\mathcal{M}) \geq \tau_0, \qquad \mathcal{M} \subset \mathbb{B}_D(0, R_0), \qquad \mathrm{vol}_{\mathcal{M}}(\mathcal{M}) \leq V_0,$$

and admitting an atlas of at most J_0 charts $g_j : \mathbb{B}_d(0, r_0) \to \mathcal{M}$ whose restricted images $g_j(\mathbb{B}_d(0, r_0/2))$ cover $\mathcal{M}$ and satisfy

$$c_g I_d \preceq Dg_j(u)^\top Dg_j(u) \preceq C_g I_d, \qquad \|g_j\|_{C^{\beta+1}} \leq L_g. \tag{21}$$

The envelope $L_{\mathcal{T}}$ is common to the entire class and does not depend on an individual manifold $\mathcal{M}$. Each admissible representation includes one witnessing atlas satisfying the displayed conditions, and all chartwise norms for that representation use this atlas.

Definition B.2 (C^β function class)**.** *Let $\beta \geq 2$ be the fixed integer from Assumption 2.1, and let $\mathcal{M} \in \mathcal{C}^{\beta+1}_{d,L_{\mathcal{T}},\tau_0}(\mathbb{R}^D)$ with a witnessing atlas as in Definition B.1. For $L_{\mathcal{F}} > 0$, define*

$$\mathcal{F}^{\beta}_{L_{\mathcal{F}}}(\mathcal{M}) := \left\{f \in C^\beta(\mathcal{M}) : \|f\|_{C^\beta(\mathcal{M})} \leq L_{\mathcal{F}}\right\},$$

where the norm is given by equation 14. The admissible latent densities in Assumption 2.2 are the elements of this class satisfying, for fixed f_-, f_+,

$$\int_{\mathcal{M}} f(x)\, d\,\mathrm{vol}_{\mathcal{M}}(x) = 1, \qquad 0 < f_- \leq f(x) \leq f_+ < \infty, \qquad x \in \mathcal{M}. \tag{22}$$

The radius $L_{\mathcal{F}}$ is common to the density class and does not depend on f. All components of $L_{\mathcal{T}}$, as well as f_-, f_+, and $L_{\mathcal{F}}$, are fixed independently of n. An envelope contains a representation when its upper-bound components dominate the corresponding quantities and its positive lower-bound components do not exceed them. The common density bounds f_-, f_+ and the integer β are suppressed in the notation for $\mathcal{P}_d$ and the LMD. Fixed-class statements use these prescribed bounds. In an existence construction that chooses common model envelopes, the density bounds are specified explicitly along with the geometric envelope and the smoothness radius.

For statements comparing density variation with the reach scale, fix a dimensionless constant $0 < C_{\mathrm{den}} < \infty$ and impose

$$\tau_0^2 \Lambda_f \leq C_{\mathrm{den}}. \tag{23}$$

This is the uniform interpretation of $\Lambda_f \lesssim \tau_0^{-2}$ in those statements; their constants may depend on C_{den}.

C STRUCTURAL IDENTIFIABILITY

We first relate the observed support to every admissible latent manifold, then prove uniqueness of its dimension and construct a counterexample at the reach boundary.

C.1 TWO-SIDED SUPPORT PROXIMITY

Bounded perturbations place the observed support within distance ξ_0 of the latent manifold. Positivity of the latent density gives the reverse proximity, as formalized by the following lemma.

Lemma C.1 (Two-sided support proximity)**.** *Let $Q \in \mathcal{P}_d(\xi_0, \tau_0, L_\mathcal{T}, L_\mathcal{F})$. Fix any $(\mathcal{M}, X, \xi)$ satisfying Assumptions 2.1–2.3 such that $Q = \mathcal{L}(X + \xi)$. Then*

$$\operatorname{supp}(Q) \subset \mathcal{M}^{\xi_0}, \qquad \mathcal{M} \subset \big(\operatorname{supp}(Q)\big)^{\xi_0}, \qquad d_{\mathrm{H}}(\mathcal{M}, \operatorname{supp}(Q)) \le \xi_0, \tag{24}$$

where $A^r = \{y : \operatorname{dist}(y, A) \le r\}$.

Proof of Lemma C.1. The noise constraint gives $Y \in \mathcal{M}^{\xi_0}$ almost surely. Since $\mathcal{M}^{\xi_0}$ is closed, $\operatorname{supp}(Q) \subset \mathcal{M}^{\xi_0}$.

Conversely, fix $x \in \mathcal{M}$ and $r > 0$. The density lower bound and local Riemannian volume give

$$\mathbb{P}(X \in \mathcal{M} \cap \mathbb{B}_D(x, r)) > 0.$$

On this event, $\|Y - x\| \le \|X - x\| + \|\xi\| < r + \xi_0$. Consequently, $Q(\mathbb{B}(x, r + \xi_0)) > 0$, so this Euclidean ball intersects $\operatorname{supp}(Q)$. Letting $r \downarrow 0$ yields $\operatorname{dist}(x, \operatorname{supp}(Q)) \le \xi_0$. Taking suprema in the two directions proves equation 24. □

C.2 REACH-SEPARATED LMD IDENTIFIABILITY

Theorem 2.1 (restated) (LMD identifiability under reach separation)**.** *If $2\xi_0 < \tau_0$, then every probability measure Q in*

$$\bigcup_{r \in [D-1]} \mathcal{P}_r(\xi_0, \tau_0, L_\mathcal{T}, L_\mathcal{F})$$

has exactly one admissible dimension. Thus, $d_{\mathrm{LMD}}(Q)$ is well-defined.

Proof of Theorem 2.1. Suppose the same Borel probability measure Q has admissible representations with compact connected boundaryless manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$ of dimensions d_1 and d_2. By Lemma C.1 and the triangle inequality for Hausdorff distance,

$$\delta := d_{\mathrm{H}}(\mathcal{M}_1, \mathcal{M}_2) \le 2\xi_0 < \tau_0. \tag{25}$$

Let p_j be the nearest-point projection onto $\mathcal{M}_j$. Positive reach and equation 25 make the continuous maps

$$f := p_2|_{\mathcal{M}_1} : \mathcal{M}_1 \to \mathcal{M}_2, \qquad g := p_1|_{\mathcal{M}_2} : \mathcal{M}_2 \to \mathcal{M}_1$$

well-defined. For $x \in \mathcal{M}_1$, the homotopy

$$H_1(x, t) := p_1\big((1 - t)x + tf(x)\big), \qquad 0 \le t \le 1, \tag{26}$$

is well-defined since the point inside p_1 is at distance at most $t\|f(x) - x\| \le \delta < \tau_0$ from $\mathcal{M}_1$. Its endpoints are $H_1(x, 0) = x$ and $H_1(x, 1) = g(f(x))$. Hence, $g \circ f \simeq \mathrm{id}_{\mathcal{M}_1}$. Interchanging the indices and using

$$H_2(y, t) := p_2\big((1 - t)y + tg(y)\big)$$

shows that $f \circ g \simeq \mathrm{id}_{\mathcal{M}_2}$. Thus, $\mathcal{M}_1$ and $\mathcal{M}_2$ are homotopy equivalent.

A compact connected boundaryless r-manifold has $H_r(\mathcal{M}; \mathbb{Z}_2) \cong \mathbb{Z}_2$ and $H_j(\mathcal{M}; \mathbb{Z}_2) = 0$ for $j > r$ [24]. If $d_1 < d_2$, homotopy invariance would give

$$0 = H_{d_2}(\mathcal{M}_1; \mathbb{Z}_2) \cong H_{d_2}(\mathcal{M}_2; \mathbb{Z}_2) \cong \mathbb{Z}_2,$$

which is a contradiction. Therefore, $d_1 = d_2$, proving uniqueness of the admissible dimension. □

C.3 A NONIDENTIFIABLE LAW AT THE REACH BOUNDARY

Theorem 2.2 (restated) (Nonidentifiable construction at the reach boundary)**.** *Assume $D \geq 4$. Let $d_1, d_2 \in [D-2]$ satisfy $d_1 < d_2$, and suppose $2\xi_0 \geq \tau_0$. There exist common density bounds $0 < f_{-,0} < f_{+,0} < \infty$, finite common envelopes $L_{\mathcal{T},0}, L_{\mathcal{F},0}$, and a Borel probability measure*

$$Q \in \mathcal{P}_{d_1}(\xi_0, \tau_0, L_{\mathcal{T},0}, L_{\mathcal{F},0}) \cap \mathcal{P}_{d_2}(\xi_0, \tau_0, L_{\mathcal{T},0}, L_{\mathcal{F},0}),$$

where both classes use $f_- = f_{-,0}$ and $f_+ = f_{+,0}$. Thus, the LMD of Q is undefined for these common bounds.

Proof of Theorem 2.2. Let $k = d_2 - d_1$, choose

$$R \geq \tau_0, \tag{27}$$

and decompose the first d_2+2 ambient coordinates as $\mathbb{R}^{d_1+1} \times \mathbb{R}^{k+1}$. Extra coordinates, when $D > d_2+2$, are set to zero. The condition $2\xi_0 \geq \tau_0$ gives $\tau_0/2 \leq \xi_0$. Let U and V be independent uniform random variables on the unit spheres $\mathbb{S}^{d_1}$ and $\mathbb{S}^k$, and define the common observed probability measure

$$Q = \mathcal{L}(RU, \tau_0 V/2, 0). \tag{28}$$

The same observed point is decomposed in two ways:

	latent manifold	latent point	normal perturbation
Representation 1	$\mathbb{S}^{d_1}(R) \times \{0\}^{k+1}$	$(RU, 0)$	$(0, \tau_0 V/2)$
Representation 2	$\mathbb{S}^{d_1}(R) \times \mathbb{S}^k(\tau_0)$	$(RU, \tau_0 V)$	$(0, -\tau_0 V/2)$

For the first representation take

$$\mathcal{M}_1 = \mathbb{S}^{d_1}(R) \times \{0\}^{k+1} \times \{0\}, \qquad X_1 = (RU, 0, 0), \qquad \xi_1 = (0, \tau_0 V/2, 0).$$

The second-coordinate block is contained in $N_{X_1}\mathcal{M}_1$, and $\|\xi_1\| = \tau_0/2 \leq \xi_0$. For the second representation take

$$\mathcal{M}_2 = \mathbb{S}^{d_1}(R) \times \mathbb{S}^k(\tau_0) \times \{0\}, \qquad X_2 = (RU, \tau_0 V, 0), \qquad \xi_2 = (0, -\tau_0 V/2, 0).$$

Since V is radial to $\mathbb{S}^k(\tau_0)$, the second perturbation belongs to $N_{X_2}\mathcal{M}_2$ and has norm $\tau_0/2 \leq \xi_0$. Both representations produce $(RU, \tau_0 V/2, 0)$. Hence, they induce exactly the same probability measure Q.

Write $\omega_j := \mathcal{H}^j(\mathbb{S}^j)$. The two latent densities, relative to their Riemannian volume measures, are

$$f_1 = \frac{1}{\omega_{d_1} R^{d_1}}, \qquad f_2 = \frac{1}{\omega_{d_1} R^{d_1} \omega_k \tau_0^k}.$$

Choose the common density bounds

$$f_{-,0} := \tfrac{1}{2}(f_1 \wedge f_2), \qquad f_{+,0} := 2(f_1 \vee f_2),$$

and use $f_- = f_{-,0}$, $f_+ = f_{+,0}$ in both model classes. Both manifolds are compact, connected, and boundaryless, with dimensions d_1 and d_2. Their reaches are

$$\operatorname{reach}(\mathcal{M}_1) = R, \qquad \operatorname{reach}(\mathcal{M}_2) = R \wedge \tau_0 = \tau_0.$$

Indeed, within the smaller of the two sphere radii the nearest point onto the orthogonal product is obtained uniquely by coordinatewise radial projection; the coordinate axes show equality. Choose smooth finite atlases for both manifolds, take a common geometric envelope $L_{\mathcal{T},0}$ containing their atlas, support, and volume bounds, and choose $L_{\mathcal{F},0} \geq f_1 \vee f_2$. The last choice bounds both chart-wise density norms because all positive-order derivatives of f_1, f_2 vanish. Thus, both representations satisfy the same density bounds and regularity envelopes, and both induce exactly the probability measure Q defined in equation 28. □

D FINITE-SAMPLE MINIMAX THEORY FOR LHSD

This section proves Theorems 3.1 and 3.2 through stronger finite-sample bounds. We first prove a minimax lower bound on a curved exact-support submodel, then construct an empirical DSM estimator and prove its matching upper bound. The risk averages over the training sample and an independent query with law Q. The geometric, covariance, and spectral estimates used below are proved in Appendix F.

For a prescribed sequence σ_n, write $T_n(Q) := T_{\sigma_n,\varpi}(Q)$.

We use the minimax integrated squared risk $\mathfrak{R}^{\mathrm{LHSD}}_{n,\sigma}(\xi_0)$ defined in equation 2.

D.1 MINIMAX LOWER BOUND

This subsection proves Proposition D.1. The Assouad construction has five ingredients:

1. place $M_\sigma \asymp \sigma^{-d}$ disjoint density bumps of width σ on one fixed curved manifold;
2. make adjacent alternatives have KL divergence $O(\theta^2\sigma^d)$;
3. use curvature to turn one bump into an LHSD response of order $\theta\sigma^2$;
4. choose $\theta \asymp (n\sigma^d)^{-1/2}$ so adjacent experiments remain statistically indistinguishable; and
5. sum the disjoint squared separations to obtain $\sigma^4/(n\sigma^d)$.

Since a zero normal perturbation is allowed in every bounded tubular class, the exact-support submodel below yields a lower bound for every tube radius $\xi_0 \geq 0$.

Fix $R_\circ > 0$ and embed $\mathcal{M}_\circ := \mathbb{S}^d(R_\circ)\times\{0\}^{D-d-1}$ in $\mathbb{R}^D$. Let $f_\circ := \mathrm{vol}_d(\mathcal{M}_\circ)^{-1}$. Fix a finite C^∞ atlas on this sphere satisfying the geometric-envelope bounds. For an integer $\alpha \geq \beta \geq 2$, evaluate all $C^\alpha(\mathcal{M}_\circ)$ norms using this fixed smooth atlas and the formula equation 14 in Appendix B.2. Its derivative bounds through order α are fixed submodel constants.

Definition D.1 (Curved density submodel)**.** *For a sufficiently small fixed $\eta_\circ > 0$, let*

$$\mathcal{F}^\circ_\alpha := \Bigg\{ f_\circ + g :\ g \in C^\alpha(\mathcal{M}_\circ),\quad \int_{\mathcal{M}_\circ} g\, d\,\mathrm{vol}_d = 0, \\ \|g\|_{C^\alpha(\mathcal{M}_\circ)} \leq \eta_\circ,\, f_\circ/2 \leq f_\circ + g \leq 3f_\circ/2 \Bigg\}. \tag{29}$$

For $f \in \mathcal{F}^\circ_\alpha$, let P_f have density f with respect to vol_d on $\mathcal{M}_\circ$, and set $\mathcal{Q}^\circ_{\alpha,d} := \{P_f : f \in \mathcal{F}^\circ_\alpha\}$. The envelope condition in Proposition D.1 means that $\mathcal{Q}^\circ_{\alpha,d}$ is contained in $\mathcal{P}_d(\xi_0, \tau_0, L_\mathcal{T}, L_\mathcal{F})$ through the representation $Y = X$ and $\xi = 0$, using the fixed smooth atlas above and the common density bounds f_-, f_+.

The sphere provides a concrete uniformly curved submodel on which the leading normal-covariance response can be computed explicitly.

The sufficiently permissive bounds in Theorem 3.1 can be specified using $R_\circ = \tau_0$ and a fixed smooth finite atlas on $\mathcal{M}_\circ$ with chart radius r_0 and a positive uniform metric lower bound, whose inner-half chart images cover $\mathcal{M}_\circ$. Choose $R_0 > \tau_0$, $V_0 \geq \mathrm{vol}_d(\mathcal{M}_\circ)$, and J_0 at least as large as the number of charts. Choose $c_g > 0$ no larger than the atlas's metric lower bound, and C_g, L_g no smaller than its metric upper bound and $C^{\beta+1}$ chart bound, respectively. For the density bounds, it suffices that

$$f_- < f_\circ < f_+, \qquad L_\mathcal{F} > f_\circ.$$

Indeed, choose a fixed $0 < \eta_\circ < f_\circ/2$ smaller than each of $f_\circ - f_-$, $f_+ - f_\circ$, and $L_\mathcal{F} - f_\circ$. Since $\alpha \geq \beta$, every density in $\mathcal{F}^\circ_\alpha$ then satisfies $f_- \leq f \leq f_+$ and $\|f\|_{C^\beta(\mathcal{M}_\circ)} \leq f_\circ + \eta_\circ \leq L_\mathcal{F}$, using the same atlas for both norms. The sphere has reach τ_0, and zero normal noise is admissible for every $\xi_0 \geq 0$. Hence these fixed bounds ensure the submodel inclusion required below.

Lemma D.1 (Localized alternatives and LHSD separation)**.** *There are constants $L_\circ, c_0, c_1, C_1, \sigma_0 > 0$, depending only on the fixed submodel and ϖ, with the following property. For every $0 < \sigma \leq \sigma_0$*

and $0 < \theta \le c_0\sigma^\alpha$, one can find $M_\sigma \asymp \sigma^{-d}$ smooth functions $b_{1,\sigma}, \dots, b_{M_\sigma,\sigma}$ with pairwise disjoint supports and densities

$$f_\omega := f_\circ + \theta \sum_{j=1}^{M_\sigma} \omega_j b_{j,\sigma}, \qquad \omega \in \{-1,1\}^{M_\sigma}, \tag{30}$$

in $\mathcal{F}_\alpha^\circ$. If $\omega^{(j)}$ denotes ω with its jth sign reversed, then

$$D_{\mathrm{KL}}(P_{f_\omega}, P_{f_{\omega^{(j)}}}) \le C_1\theta^2\sigma^d, \tag{31}$$

$$\int_{U_{j,\sigma}} \left| T_{\sigma,\varpi}(P_{f_\omega})(x) - T_{\sigma,\varpi}(P_{f_{\omega^{(j)}}})(x) \right|^2 d\,\mathrm{vol}_d(x) \ge c_1\theta^2\sigma^{d+4}, \tag{32}$$

where the query regions $U_{j,\sigma} \subset \mathcal{M}_\circ$ are pairwise disjoint.

Proof. **Step 1: A mean-zero profile with nonzero curvature response.** Fix $a \in \mathcal{M}_\circ$ and identify $T_a\mathcal{M}_\circ$ with $\mathbb{R}^d$. In a tangent–normal chart,

$$F_a(v) = a + v + G_a(v), \qquad G_a(v) = Q_a(v) + O(\|v\|^3), \qquad Q_a(v) := \tfrac{1}{2}\mathrm{II}_a(v,v). \tag{33}$$

For the sphere, Q_a is nonzero at every a. Let π_z denote the $N(z, I_d)$ probability measure and $W \sim \pi_z$. For each ψ, define

$$\mathcal{G}_z(\psi) := \operatorname{tr}\mathrm{Cov}_{\pi_z}\big(\{Q_a(W-z) - \mathbb{E}_{\pi_z}Q_a(W-z)\}\{Q_a(W-z) - \mathbb{E}_{\pi_z}Q_a(W-z)\}^\top, \psi(W)\big). \tag{34}$$

The polynomial $H(w) := \|Q_a(w) - \mathbb{E}_{\pi_0}Q_a(W)\|^2$ is nonconstant. Hence the linear functional

$$\psi \longmapsto \int_{\mathbb{R}^d} \{H(w) - \mathbb{E}_{\pi_0}H(W)\}\psi(w)\, d\pi_0(w) \tag{35}$$

does not vanish on the space of compactly supported smooth functions with zero Lebesgue integral. Otherwise, the density of this functional with respect to Lebesgue measure would be constant; its Gaussian decay would force that constant to be zero, contradicting that H is nonconstant. Choose one such function ψ for which $\mathcal{G}_0(\psi) \ne 0$. Continuity in z then gives fixed $r_q > 0$ and $c_q > 0$ such that

$$\int_{\mathbb{B}_d(0,r_q)} |\mathcal{G}_z(\psi)|^2\, dz \ge c_q. \tag{36}$$

Rotational invariance permits the same profile and constants at every center after identifying tangent and normal frames.

Step 2: Manifold bumps and packing. Let J_a be the volume Jacobian of F_a. Choose a fixed $\chi \in C_c^\infty(\mathbb{R}^d)$ with $\int \chi > 0$ whose support contains that of ψ, and put

$$c_{a,\sigma} := \frac{\int \psi(w) J_a(\sigma w)\, dw}{\int \chi(w) J_a(\sigma w)\, dw}, \qquad b_{a,\sigma}(F_a(\sigma w)) := \psi(w) - c_{a,\sigma}\chi(w), \tag{37}$$

extended by zero outside the chart. Since $J_a(0) = 1$, $J_a(\sigma w) = 1 + O(\sigma^2\|w\|^2)$ on the sphere, and $\int \psi = 0$,

$$\int_{\mathcal{M}_\circ} b_{a,\sigma}\, d\,\mathrm{vol}_d = 0, \quad |c_{a,\sigma}| \le C\sigma^2, \quad \|b_{a,\sigma}\|_{C^\alpha(\mathcal{M}_\circ)} \le C\sigma^{-\alpha}, \quad \int b_{a,\sigma}^2\, d\,\mathrm{vol}_d \asymp \sigma^d. \tag{38}$$

Take a maximal $L\sigma$-separated set of centers in a fixed positive-volume geodesic patch, where L is chosen larger than twice the support and query radii. Uniform volume growth on the sphere gives $M_\sigma \asymp \sigma^{-d}$. Set $b_{j,\sigma} := b_{a_j,\sigma}$ and $U_{j,\sigma} := F_{a_j}(\sigma\mathbb{B}_d(0,r_q))$. The bump supports and query regions are pairwise disjoint. By equation 38,

$$\left\| \theta \sum_j \omega_j b_{j,\sigma} \right\|_{C^\alpha} \le C\theta\sigma^{-\alpha}. \tag{39}$$

Decreasing c_0 makes every density in equation 30 positive and places it in $\mathcal{F}_\alpha^\circ$.

For adjacent vertices, positivity and the elementary inequality $D_{\mathrm{KL}}(p,q) \le \int (p-q)^2/q$ give

$$D_{\mathrm{KL}}(P_{f_\omega}, P_{f_{\omega^{(j)}}}) \le C\theta^2 \int b_{j,\sigma}^2 \, d\,\mathrm{vol}_d \le C_1 \theta^2 \sigma^d, \tag{40}$$

which proves equation 31.

Step 3: First-order response of affine-calibrated LHSD. First isolate one bump and write $f_t = f_\circ + t b_{a,\sigma}$, $|t| \le \theta$. The goal of this step is to show that changing the density by an amount t produces an LHSD response of order $t\sigma^2$. The factor σ^2 comes from curvature: a tangential displacement of size σ produces a normal displacement of size σ^2, hence a normal covariance of size σ^4, which is divided by σ^2 in the matrix defining LHSD.

For a query $x = F_a(\sigma z)$ with $z \in \mathbb{B}_d(0, r_q)$, expectations without an indicated measure refer to the Gaussian-localized law induced by $f_\circ$ at x. Thus, points X closer to x receive larger weights, proportional to $\exp(-\|X-x\|^2/(2\sigma^2))$. Define $W := \sigma^{-1} P_{T_a\mathcal{M}_\circ}(X-a)$ in the chosen tangent frame, and denote its law by $\nu_{a,z,\sigma}$. The variable W records location in the tangent plane in units of σ; the query has rescaled tangent coordinate z.

We first establish the quantitative comparison

$$\|\nu_{a,z,\sigma} - \pi_z\|_{\mathrm{TV},r} \le C_r \sigma^2, \qquad r \in \{0, \dots, 12\}, \tag{41}$$

uniformly in $a \in \mathcal{M}_\circ$ and $z \in \mathbb{B}_d(0, r_q)$, where $\pi_z = N(z, I_d)$ and

$$\|\nu - \pi\|_{\mathrm{TV},r} := \int_{\mathbb{R}^d} (1 + \|w\|^r)\, d|\nu - \pi|(w).$$

This weighted comparison allows us to replace expectations of polynomially growing functions under the actual tangent law by Gaussian expectations. We retain moments through order twelve because the cubic Taylor remainders below will be controlled in L^4.

Choose a fixed sufficiently small chart radius $r_\circ > 0$, and take σ small enough that $\|\sigma z\| < r_\circ/2$. On $\|\sigma w\| < r_\circ$, orthogonality of the graph coordinates gives

$$\|F_a(\sigma w) - F_a(\sigma z)\|^2 = \sigma^2 \|w - z\|^2 + \|G_a(\sigma w) - G_a(\sigma z)\|^2.$$

The first term produces the Gaussian reference law π_z; the second term and the volume Jacobian account for the effect of curvature. After division by $f_\circ (2\pi)^{d/2}\sigma^d$, the unnormalized local pushforward has density

$$J_a(\sigma w) \exp\left(-\frac{\|G_a(\sigma w) - G_a(\sigma z)\|^2}{2\sigma^2} \right)$$

relative to π_z on this coordinate domain. The graph bounds imply $J_a(\sigma w) \le C$, $|J_a(\sigma w) - 1| \le C\sigma^2 \|w\|^2$, and

$$\|G_a(\sigma w) - G_a(\sigma z)\| \le C\sigma^2 (\|w\|^2 + \|z\|^2).$$

Using $0 \le 1 - e^{-v} \le v$ for $v \ge 0$, we obtain

$$\left| J_a(\sigma w) \exp\left(-\frac{\|G_a(\sigma w) - G_a(\sigma z)\|^2}{2\sigma^2} \right) - 1 \right| \le C\sigma^2 (1 + \|w\|^4). \tag{42}$$

Integrating against $(1 + \|w\|^r)\, d\pi_z(w)$ gives $C_r\sigma^2$, uniformly for bounded z. Points outside this fixed spherical cap are at a fixed positive distance from x for sufficiently small σ. Their weighted contribution, with the same normalization, is at most $C_r \sigma^{-d-r} \exp(-c/\sigma^2)$; the Gaussian mass outside the coordinate domain has an analogous exponential bound. These tails are $O(\sigma^2)$, and the total normalizer is $1 + O(\sigma^2)$. Dividing by it proves equation 41. This bound controls polynomial moments through order twelve, including their products with the bounded bump profile ψ.

Next, we express displacements in the tangent and normal spaces at the query x, rather than at the bump center a. Use the fixed orthogonal decomposition $T_a\mathcal{M}_\circ \oplus N_a\mathcal{M}_\circ$ and put $K := DG_a(\sigma z)$. The maps

$$U_{a,z,\sigma} := \begin{pmatrix} I_d \\ K \end{pmatrix} (I_d + K^\top K)^{-1/2}, \qquad V_{a,z,\sigma} := \begin{pmatrix} -K^\top \\ I_{D-d} \end{pmatrix} (I_{D-d} + KK^\top)^{-1/2}$$

are isometries onto $T_x\mathcal{M}_\circ$ and $N_x\mathcal{M}_\circ$, respectively. They account for the small rotation of the tangent and normal spaces between a and x. For bounded z, $K = \sigma DQ_a(z)+O(\sigma^2)$, so both inverse square roots equal the corresponding identity plus $O_{\mathrm{op}}(\sigma^2)$. Writing $\Delta_a(w,z) := F_a(\sigma w)-F_a(\sigma z)$, we obtain on the fixed chart

$$\begin{aligned} U_{a,z,\sigma}^\top \Delta_a(w,z) &= (I_d + K^\top K)^{-1/2}\big[\sigma(w-z) + K^\top\{G_a(\sigma w) - G_a(\sigma z)\}\big] \\ &= \sigma(w-z) + O\big(\sigma^3(1+\|w\|^3)\big), \\ V_{a,z,\sigma}^\top \Delta_a(w,z) &= (I_{D-d} + KK^\top)^{-1/2}\big[G_a(\sigma w) - G_a(\sigma z) - K\sigma(w-z)\big] \\ &= \sigma^2 Q_a(w-z) + O\big(\sigma^3(1+\|w\|^3)\big). \end{aligned}$$

The normal projection subtracts the tangent-plane contribution at x, leaving the quadratic curvature term. Algebraically, this is the identity $Q_a(w)-Q_a(z)-DQ_a(z)(w-z) = Q_a(w-z)$. The uniform third-order Taylor bounds for the fixed smooth sphere give the displayed remainders, and the twelfth-moment bound in equation 41 makes them $O_{L^4}(\sigma^3)$. Outside the fixed cap, the displacements and the scaled leading terms $\sigma(W-z)$ and $\sigma^2 Q_a(W-z)$ are uniformly bounded, while the localized probability is at most $C\sigma^{-d}e^{-c/\sigma^2}$. Thus this contribution has the same L^4 bound. Finally, centering a remainder increases its L^4 norm by at most a factor of two. Consequently, uniformly in a and bounded z,

$$P_T(x)\{X-\mathbb{E}X\} = \sigma U_{a,z,\sigma}\{W-\mathbb{E}W\} + O_{L^4}(\sigma^3), \tag{43}$$

$$P_N(x)\{X-\mathbb{E}X\} = \sigma^2 V_{a,z,\sigma}\{Q_a(W-z) - \mathbb{E}Q_a(W-z)\} + O_{L^4}(\sigma^3), \tag{44}$$

Moreover, the bump construction in Step 2 gives, on the chart,

$$\frac{b_{a,\sigma}(F_a(\sigma w))}{f_\circ} = \frac{\psi(w)}{f_\circ} + O(\sigma^2). \tag{45}$$

When this relation is used inside a localized expectation, the off-chart contribution is absorbed by the same exponential tail bound.

We now translate these displacement expansions into a covariance response. If a localized probability measure is tilted from a positive density g along $g+tb$, differentiation of normalized moments gives

$$\frac{d}{dt}\operatorname{Cov}_t(V)\Big|_{t=0} = \operatorname{Cov}_0\left((V-\mathbb{E}_0V)(V-\mathbb{E}_0V)^\top, \frac{b}{g}\right). \tag{46}$$

Thus the covariance changes according to how the density perturbation b/g correlates with the centered quadratic displacement. The covariance on the right also accounts for the change in the normalizing constant; no mean-zero condition on b/g under the localized law is needed.

Apply equation 46 to equation 43– equation 44. If a dot denotes differentiation at $t=0$, then the blocks of $\dot A = -\dot\Sigma/\sigma^2$ satisfy

$$\|P_T\dot A P_T\|_{\mathrm{op}} \le C, \qquad \|P_T\dot A P_N\|_{\mathrm{op}} \le C\sigma, \tag{47}$$

$$P_N\dot A P_N = -\frac{\sigma^2}{f_\circ} V_{a,z,\sigma}\operatorname{Cov}_z(\psi)V_{a,z,\sigma}^\top + O_{\mathrm{op}}(\sigma^3), \tag{48}$$

where $\operatorname{Cov}_z(\psi)$ is the matrix covariance inside equation 34. The orders of these blocks reflect the displacement scales: tangent–tangent, tangent–normal, and normal–normal products have sizes σ^2, σ^3, and σ^4, respectively, before division by σ^2.

For the normal block, the replacement of $\nu_{a,z,\sigma}$ by π_z also changes the centering mean of $Q_a(W-z)$ by $O(\sigma^2)$. Since Q_a is quadratic, the centered covariance can be expanded into moments of degree at most four, with or without the bounded factor ψ. The weighted comparison therefore controls both the moments and their centering errors. Together with equation 45, it changes the leading order-σ^4 covariance derivative by $O(\sigma^6)$. The cross products with the $O_{L^4}(\sigma^3)$ remainder in equation 44 contribute $O(\sigma^5)$. Division by σ^2 yields the stated $O_{\mathrm{op}}(\sigma^3)$ remainder.

Let $q_{a,z,\sigma}(t) := T_{\sigma,\varpi}(P_{f_t})(x)$. The standard derivative formula for a trace spectral function [14] gives

$$q'_{a,z,\sigma}(0) = \operatorname{tr}\Big[\bar\phi'_\varpi\{A_{P_{f_\circ},\sigma}(x)\}\dot A\Big]. \tag{49}$$

Rotational invariance of the uniform sphere makes $A_{P_{f_\circ},\sigma}(x)$ block diagonal in the tangent, radial, and remaining normal directions. Writing Σ_0 for the localized covariance under $f_\circ$, the moment comparison and centered displacement expansions above give

$$P_T\Sigma_0 P_T = \sigma^2 P_T + O_{\mathrm{op}}(\sigma^4), \qquad P_N\Sigma_0 P_N = O_{\mathrm{op}}(\sigma^4).$$

Hence the tangent eigenvalue of $A_{P_{f_\circ},\sigma}(x)$ is $O(\sigma^2)$, its radial eigenvalue is $1 + O(\sigma^2)$, and its remaining normal eigenvalues are exactly one, since the sphere has no variation in those directions.

The filter suppresses the tangent response because $\bar\phi'_\varpi(\lambda) = O_\varpi(\lambda^3)$ near zero, whereas it retains the normal response because $\bar\phi'_\varpi(1) = -4\varpi^4/(1+\varpi^4) \neq 0$. In particular, the tangent contribution is $O_\varpi(\sigma^6)$, and the mixed blocks do not contribute to the trace against the block-diagonal matrix $\bar\phi'_\varpi(A_{P_{f_\circ},\sigma}(x))$. Substituting equation 47– equation 48 into equation 49 therefore yields

$$q'_{a,z,\sigma}(0) = -\frac{\sigma^2\bar\phi'_\varpi(1)}{f_\circ}\mathcal{G}_z(\psi) + O_\varpi(\sigma^3). \tag{50}$$

This is the desired infinitesimal response: the nonzero Gaussian curvature response selected in Step 1 survives in the actual LHSD field, with a leading factor σ^2.

To turn this derivative into a separation between the two densities $f_\circ \pm \theta b_{a,\sigma}$, we also need to control the Taylor remainder. We establish a uniform third-derivative bound, since the second-order terms cancel in the symmetric difference $q_{a,z,\sigma}(\theta) - q_{a,z,\sigma}(-\theta)$. For a polynomially bounded scalar-, vector-, or matrix-valued H, write

$$\mathcal{M}_H(g;x) := \frac{\int e^{-\|u-x\|^2/(2\sigma^2)}H(u)g(u)d\,\mathrm{vol}_d(u)}{\int e^{-\|u-x\|^2/(2\sigma^2)}g(u)d\,\mathrm{vol}_d(u)}. \tag{51}$$

Consider a path $g_t = g_0 + tb$, $|t| \le \theta$, with $f_\circ/2 \le g_t \le 3f_\circ/2$ and $\|b\|_\infty \le C$, as supplied by Step 2. Apply this formula componentwise to $Z(u) := (u-x)/\sigma$ and $Z(u)Z(u)^\top$. The numerator and denominator have the form $N(t) = N_0 + tN_1$ and $D(t) = D_0 + tD_1$. The localized moment bounds, the off-chart tail estimate, and the density bounds imply

$$|N_0| + |N_1| + |D_0| + |D_1| \le C\sigma^d, \qquad \inf_{|t|\le\theta} D(t) \ge c\sigma^d.$$

For $1 \le \ell \le 3$, direct differentiation gives

$$\frac{d^\ell}{dt^\ell}\frac{N(t)}{D(t)} = (-1)^{\ell-1}\ell!\frac{D_1^{\ell-1}(N_1D_0 - N_0D_1)}{D(t)^{\ell+1}}. \tag{52}$$

The powers of σ^d cancel in this expression: although the localized mass shrinks with σ, normalization does not make these derivatives diverge. Thus the normalized moments and their first three derivatives are uniformly bounded. Consequently, for

$$A(t) := A_{P_{g_t},\sigma}(x) = I_D - \mathcal{M}_{ZZ^\top}(g_t;x) + \mathcal{M}_Z(g_t;x)\mathcal{M}_Z(g_t;x)^\top,$$

we have $\sup_{|t|\le\theta}\|A^{(\ell)}(t)\|_F \le C$ for $0 \le \ell \le 3$. The affine-calibrated filter satisfies

$$\operatorname{tr}\bar\phi_\varpi(A) = \varpi^4(1+\varpi^4)\operatorname{tr}(\varpi^4 I_D + A^4)^{-1} - D\varpi^4.$$

For symmetric A and fixed $\varpi > 0$, $\varpi^4 I_D + A^4 \succeq \varpi^4 I_D$. Repeated differentiation of this polynomial and matrix inverse therefore bounds the first three derivatives of the trace map uniformly on the bounded matrices above. The chain rule gives

$$\left|\frac{d^3}{dt^3}\operatorname{tr}\bar\phi_\varpi(A(t))\right| \le C_\varpi\big(\|A'''(t)\|_F + \|A'(t)\|_F\|A''(t)\|_F + \|A'(t)\|_F^3\big) \le C_\varpi.$$

In particular, the isolated path satisfies

$$\sup_{|t|\le\theta}|q^{(3)}_{a,z,\sigma}(t)| \le C_\varpi. \tag{53}$$

Finally, set $c_* := -\bar\phi'_\varpi(1)/f_\circ > 0$. Taylor's formula for the symmetric difference and equation 50 give

$$q_{a,z,\sigma}(\theta) - q_{a,z,\sigma}(-\theta) = 2c_*\theta\sigma^2\mathcal{G}_z(\psi) + O_\varpi(\theta\sigma^3 + \theta^3).$$

The response from Step 1 is nonzero in L^2 over the query ball; a pointwise lower bound at every query is not required. To see that the remainder cannot cancel this response, the reverse triangle inequality gives

$$\begin{aligned}&\|q_{a,\cdot,\sigma}(\theta) - q_{a,\cdot,\sigma}(-\theta)\|_{L^2(\mathbb{B}_d(0,r_q))}\\ &\quad\geq 2c_*\theta\sigma^2\|\mathcal{G}_\cdot(\psi)\|_{L^2(\mathbb{B}_d(0,r_q))} - C_\varpi(\theta\sigma^3+\theta^3).\end{aligned}$$

By equation 36, the first norm on the right is bounded below by a positive constant. Moreover,

$$\frac{\theta\sigma^3+\theta^3}{\theta\sigma^2} = \sigma + \frac{\theta^2}{\sigma^2} \leq \sigma + c_0^2\sigma^{2\alpha-2},$$

using $\theta \leq c_0\sigma^\alpha$ and $\alpha \geq \beta \geq 2$. After reducing c_0 and σ_0 if needed, the remainder is absorbed by the leading term. Squaring and changing variables $x = F_a(\sigma z)$, whose volume Jacobian is $\sigma^d J_a(\sigma z)$ with $J_a(\sigma z)$ uniformly bounded below, therefore yields

$$\int_{U_{a,\sigma}} |T_{\sigma,\varpi}(P_{f_\circ+\theta b_{a,\sigma}})(x) - T_{\sigma,\varpi}(P_{f_\circ-\theta b_{a,\sigma}})(x)|^2 d\,\mathrm{vol}_d(x) \geq c_\varpi\theta^2\sigma^{d+4}. \tag{54}$$

In terms of integrated squared orders, the leading response, geometric remainder, and third-order Taylor remainder have sizes $\theta^2\sigma^{d+4}$, $\theta^2\sigma^{d+6}$, and $\sigma^d\theta^6$, respectively. Thus curvature supplies the factor σ^4 in the squared separation, while the query region supplies the factor σ^d.

Step 4: Stability under the remaining bumps. It remains to verify that the other coordinates of the hypercube do not cancel the response in equation 54. We use the normalized moments in equation 51 and their derivatives in equation 52. If the centers are $L\sigma$-separated and $x \in U_{j,\sigma}$, Gaussian tails and shell counting give

$$\sum_{k\neq j}\int e^{-\|u-x\|^2/(2\sigma^2)}(1+\|u-x\|/\sigma)^m |b_{k,\sigma}(u)| d\,\mathrm{vol}_d(u) \leq C_m\sigma^d\delta_L, \tag{55}$$

where

$$\delta_L := C\sum_{r\geq 1}(r+1)^d e^{-cL^2r^2} \longrightarrow 0 \quad \text{as } L\to\infty. \tag{56}$$

To make the dependence on the bump amplitude explicit, fix the remaining signs and put

$$g_\omega := f_\circ + \theta\sum_{k\neq j}\omega_k b_{k,\sigma}, \qquad q_{j,g}(t;x) := T_{\sigma,\varpi}(P_{g+tb_{j,\sigma}})(x), \qquad |t|\leq\theta.$$

All densities along these paths are bounded below by $f_\circ/2$ after the choice of c_0 in Step 2. For a fixed query $x \in U_{j,\sigma}$, apply equation 55 to the components of $V(u) := (u-x)/\sigma$ and $V(u)V(u)^\top$. The changes in their unnormalized numerators, and in the denominator, are at most $C\theta\sigma^d\delta_L$. Their derivatives along $b_{j,\sigma}$ are the same for the backgrounds g_ω and $f_\circ$, and are bounded by $C\sigma^d$. Since the denominators are at least $c\sigma^d$, equation 52 gives, for each such component H,

$$\begin{aligned}|\mathcal{M}_H(g_\omega;x) - \mathcal{M}_H(f_\circ;x)| &\leq C\theta\delta_L,\\ \left|\frac{d}{dt}\mathcal{M}_H(g_\omega+tb_{j,\sigma};x)\Big|_{t=0} - \frac{d}{dt}\mathcal{M}_H(f_\circ+tb_{j,\sigma};x)\Big|_{t=0}\right| &\leq C\theta\delta_L.\end{aligned}$$

The covariance equals the second moment minus the outer product of the first moment. Thus, writing $A_g := A_{P_g,\sigma}(x)$ and $\dot{A}_g := \frac{d}{dt}A_{P_{g+tb_{j,\sigma}},\sigma}(x)\Big|_{t=0}$, we obtain

$$\|A_{g_\omega} - A_{f_\circ}\|_F + \|\dot{A}_{g_\omega} - \dot{A}_{f_\circ}\|_F \leq C\theta\delta_L, \qquad \|\dot{A}_{f_\circ}\|_F \leq C. \tag{57}$$

All constants are uniform in j, x, σ and the remaining signs.

The scalar function $\bar{\phi}'_\varpi$ has bounded derivative. Its matrix divided-difference formula therefore gives

$$\|\bar{\phi}'_\varpi(B) - \bar{\phi}'_\varpi(C)\|_F \leq \sup_{\lambda\in\mathbb{R}}|\bar{\phi}''_\varpi(\lambda)|\,\|B-C\|_F$$

for symmetric B, C. Apply this inequality and the trace derivative formula to equation 57. Since $\bar{\phi}'_\varpi$ is bounded, this yields

$$|q'_{j,g_\omega}(0;x) - q'_{j,f_\circ}(0;x)| \leq C_\varpi\theta\delta_L \leq C_\varpi\sigma^2\delta_L, \tag{58}$$

where the last step uses $\theta \le c_0\sigma^\alpha$, $\alpha \ge \beta \ge 2$, and $\sigma \le \sigma_0 \le 1$. These paths have the same uniform density and bump bounds as those used in Step 3, so its quotient and trace-map argument also gives $\sup_{|t|\le\theta} |q^{(3)}_{j,g_\omega}(t;x)| \le C_\varpi$.

In the rescaled query coordinate $x = F_{a_j}(\sigma z)$, set

$$H_\varpi(z) := -\frac{\bar\phi'_\varpi(1)}{f_\circ}\mathcal{G}_z(\psi), \qquad c_{\mathrm{resp}} := \|H_\varpi\|_{L^2(\mathbb{B}_d(0,r_q))} > 0.$$

Equations equation 50 and equation 58 imply

$$\left\|\sigma^{-2}q'_{j,g_\omega}(0;F_{a_j}(\sigma\,\cdot)) - H_\varpi\right\|_{L^2(\mathbb{B}_d(0,r_q))} \le C_\varpi(\sigma+\delta_L).$$

Choose $L = L_\circ$ so that the second term is at most $c_{\mathrm{resp}}/4$, and then decrease σ_0 so that the first term is at most $c_{\mathrm{resp}}/4$. The derivative consequently has L^2 norm at least $c_{\mathrm{resp}}\sigma^2/2$ on the rescaled query ball. The symmetric Taylor remainder has norm at most $C_\varpi\theta^3$, which is absorbed by the leading $\theta\sigma^2$ term because $\theta^2/\sigma^2 \le c_0^2\sigma^{2\alpha-2}$ tends to zero uniformly as σ_0 decreases. The query change of variables contributes σ^d with a uniformly bounded positive volume Jacobian. This proves equation 32, uniformly over all remaining signs, and completes the construction. □

Under the stated submodel-inclusion condition, the exact-support curved submodel is admissible for every noise bound $\xi_0 \ge 0$, since zero normal perturbation is allowed. Curvature turns a tangent-density perturbation into an order-σ^2 response in LHSD through the normalized normal covariance; squaring this response produces the factor σ^4.

The next proposition gives a finite-sample form of Theorem 3.1.

Proposition D.1 (Minimax lower bound for finite-scale LHSD)**.** *Let $\alpha \ge \beta \ge 2$ be an integer, and suppose the fixed model envelopes contain the curved sphere-density submodel $\mathcal{Q}^\circ_{\alpha,d}$ in Definition D.1. Set $h_n := n^{-1/(2\alpha+d)}$. There exists $\sigma_{\mathrm{L}} > 0$, depending only on the fixed model parameters and ϖ, such that, for all sufficiently large n, every $\xi_0 \ge 0$, and every $h_n \le \sigma \le \sigma_{\mathrm{L}}$,*

$$\mathfrak{R}^{\mathrm{LHSD}}_{n,\sigma}(\xi_0) \gtrsim_\varpi \frac{\sigma^4}{n\sigma^d}. \tag{59}$$

Proof of Proposition D.1. Fix $h_n \le \sigma \le \sigma_{\mathrm{L}}$ and choose

$$\theta := \frac{\kappa}{\sqrt{n\sigma^d}}, \tag{60}$$

where $\kappa > 0$ is a sufficiently small fixed constant. The inequality $\sigma \ge h_n = n^{-1/(2\alpha+d)}$ is equivalent to $(n\sigma^d)^{-1/2} \le \sigma^\alpha$, so Lemma D.1 applies. For adjacent vertices, product additivity of relative entropy and equation 31 give

$$D_{\mathrm{KL}}(P^{\otimes n}_{f_\omega}, P^{\otimes n}_{f_{\omega^{(j)}}}) \le C_1\kappa^2. \tag{61}$$

Choose κ so that Pinsker's inequality makes the corresponding total variation distance at most $1/2$.

For each j, pair the hypercube vertices that differ only in coordinate j. The usual two-point testing inequality in the Hilbert space $L^2(U_{j,\sigma}, d\,\mathrm{vol}_d)$, followed by equation 32, gives a contribution at least $c\theta^2\sigma^{d+4}$ per pair. Since the query regions are disjoint and every $f_\omega \ge f_\circ/2$, averaging over all vertices yields, for every estimator $\widehat{T}$,

$$\begin{aligned}\sup_\omega \mathbb{E}_{P^{\otimes n}_{f_\omega}}\|\widehat{T} - T_{\sigma,\varpi}(P_{f_\omega})\|^2_{L^2(P_{f_\omega})} &\ge cM_\sigma\theta^2\sigma^{d+4} \\ &\ge c\theta^2\sigma^4 = c\frac{\sigma^4}{n\sigma^d}.\end{aligned} \tag{62}$$

This is Assouad's hypercube argument; see Tsybakov [56, Theorem 2.12]. The submodel belongs to the full tubular class for every $\xi_0 \ge 0$ by Definition D.1. Taking the infimum over all estimators in equation 62 proves equation 59. For the scales in Theorem 3.1, $\sigma_n \to 0$ ensures $\sigma_n \le \sigma_{\mathrm{L}}$ for all sufficiently large n, proving the main-text statement. □

D.2 Empirical DSM and minimax upper bound

We now work with the full tubular class $\mathcal{P}_d(\xi_{0,n}, \tau_0, L_{\mathcal{T}}, L_{\mathcal{F}})$. We construct the Gaussian-kernel estimator through fixed-scale empirical DSM, control its covariance blocks, and transfer these bounds through the LHSD filter. The resulting risk bound yields the matching minimax upper bound in the thin-tube regime.

D.2.1 Gaussian-kernel estimator and connection to fixed-scale DSM

Let $Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} Q$, let $G \sim \mathcal{N}(0, I_D)$, and fix $\sigma_n > 0$. The Gaussian-averaged empirical denoising objective is

$$\widehat{\mathcal{R}}_{n,\mathrm{DSM}}(\mathsf{m}) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_G \|\mathsf{m}(Y_i + \sigma_n G) - Y_i\|^2. \tag{63}$$

Minimization is over measurable maps $\mathsf{m} : \mathbb{R}^D \to \mathbb{R}^D$ with finite objective. Conditionally on the sample, its smooth minimizer has weights

$$w_{i,n}(w) := \frac{\exp\{-\|w - Y_i\|^2/(2\sigma_n^2)\}}{\sum_{j=1}^{n} \exp\{-\|w - Y_j\|^2/(2\sigma_n^2)\}}, \tag{64}$$

and Gaussian-weighted mean

$$\widetilde{\mathsf{m}}_{n,\sigma_n}(w) := \sum_{i=1}^{n} w_{i,n}(w) Y_i. \tag{65}$$

If

$$\widehat{q}_{n,\sigma_n}(w) := \frac{1}{n}\sum_{i=1}^{n} \varphi_{\sigma_n}(w - Y_i), \tag{66}$$

then

$$\nabla \log \widehat{q}_{n,\sigma_n}(w) = \frac{\widetilde{\mathsf{m}}_{n,\sigma_n}(w) - w}{\sigma_n^2}, \qquad -\sigma_n^2 \nabla^2 \log \widehat{q}_{n,\sigma_n}(w) = I_D - \nabla \widetilde{\mathsf{m}}_{n,\sigma_n}(w), \tag{67}$$

and

$$\nabla \widetilde{\mathsf{m}}_{n,\sigma_n}(w) = \frac{1}{\sigma_n^2}\sum_{i=1}^{n} w_{i,n}(w)\big(Y_i - \widetilde{\mathsf{m}}_{n,\sigma_n}(w)\big)\big(Y_i - \widetilde{\mathsf{m}}_{n,\sigma_n}(w)\big)^{\top}. \tag{68}$$

The fixed-scale empirical objective in equation 63 is a conditional least-squares problem. Its optimizer and the corresponding covariance identity are proved next.

Proposition D.2 (Empirical DSM minimizer and covariance identity)**.** *Among measurable maps with finite objective, the minimizer of equation 63 is unique up to null sets of the empirical Gaussian mixture. Equation 65 is an everywhere-defined C^∞ version, its Jacobian is given by equation 68, and it satisfies equation 67.*

Proof of Proposition D.2. Condition on the observed sample and let I be uniform on $[n]$, independently of G, with $W = Y_I + \sigma_n G$. The empirical objective equation 63 is

$$\mathbb{E}\big(\|\mathsf{m}(W) - Y_I\|^2 \mid Y_1, \dots, Y_n\big).$$

The conditional Pythagorean identity shows that its unique minimizer in $L^2(\widehat{q}_{n,\sigma_n}(w)\,dw; \mathbb{R}^D)$ is $\mathbb{E}(Y_I \mid W = w, Y_1, \dots, Y_n)$. Direct normalization of the empirical Gaussian mixture gives the weights in equation 64. Hence, this denoising regression function is $\widetilde{\mathsf{m}}_{n,\sigma_n}(w)$. Since the empirical Gaussian mixture has a strictly positive density on $\mathbb{R}^D$, this formula is C^∞ and is an everywhere-defined version of the otherwise almost-everywhere defined L^2 minimizer.

Writing $\widetilde{\mathsf{m}} = \widetilde{\mathsf{m}}_{n,\sigma_n}$ and $w_i = w_{i,n}$, direct differentiation gives

$$\nabla w_i(w) = \sigma_n^{-2} w_i(w)\big(Y_i - \widetilde{\mathsf{m}}(w)\big).$$

Therefore,

$$\nabla\widetilde{\mathsf{m}}(w) = \sum_i Y_i \nabla w_i(w)^\top = \sigma_n^{-2} \sum_i w_i(w)\big(Y_i - \widetilde{\mathsf{m}}(w)\big)\big(Y_i - \widetilde{\mathsf{m}}(w)\big)^\top,$$

which proves equation 68. Directly differentiating equation 66 also gives

$$\nabla \log \widehat{q}_{n,\sigma_n} = \sigma_n^{-2}(\widetilde{\mathsf{m}}_{n,\sigma_n} - \mathrm{id}), \qquad -\sigma_n^2 \nabla^2 \log \widehat{q}_{n,\sigma_n} = I_D - \nabla\widetilde{\mathsf{m}}_{n,\sigma_n},$$

which proves equation 67. □

Define the Gaussian-kernel normalized Hessian and LHSD field by

$$\widetilde{A}_{n,\sigma_n}(w) := -\sigma_n^2 \nabla^2 \log \widehat{q}_{n,\sigma_n}(w), \qquad \widetilde{T}_n(w) := \operatorname{tr} \bar{\phi}_\varpi\Big(\widetilde{A}_{n,\sigma_n}(w)\Big). \tag{69}$$

The covariance representation in equation 68 makes $\widetilde{A}_{n,\sigma_n}$ symmetric. Set

$$\varepsilon_n := \varepsilon(\xi_{0,n}, \sigma_n), \qquad N_n := n\sigma_n^d.$$

The quantity N_n is the effective local sample size in the thin-tube regime: localization averages across the bounded normal fiber, while local mass along the latent manifold is of order σ_n^d.

Constants for the finite-sample bounds. The covariance and spectral bounds in Appendix F supply fixed constants $\varepsilon_0 > 0$ and $C_0 < \infty$. For the finite-sample analysis, set

$$b_n := C_0 \varepsilon_n. \tag{70}$$

D.2.2 Kernel localization and covariance concentration

The concentration proof has four steps: the local kernel mass is of order σ_n^d; tangent displacements have scale σ_n; normal displacements have scale $\xi_{0,n} + \sigma_n^2/\tau_0$; and raw kernel moments are converted to covariances on a stable-denominator event. This explains both the effective sample size $N_n = n\sigma_n^d$ and the anisotropic block rates.

Fix $y \in \operatorname{supp}(Q)$, put $x = \pi_{\mathcal{M}}(y)$, and let P_T, P_N be the tangent and normal projectors at x. For $w \in \operatorname{supp}(Q)$ define

$$\mathsf{k}_{y,n}(w) := \exp\left(-\frac{\|w-y\|^2}{2\sigma_n^2}\right), \qquad D_T(w) := P_T(w-y), \quad D_N(w) := P_N(w-y). \tag{71}$$

Write $\delta_n = \xi_{0,n} + \sigma_n^2/\tau_0$ and set $r_T = \sigma_n$, $r_N = \delta_n$. The two terms in δ_n are, respectively, actual fiber variation and curvature-induced normal displacement over a tangent neighborhood of radius σ_n.

Lemma D.2 (Uniform local kernel moments)**.** *Under the geometric and density assumptions, there is a fixed $c_\sigma \in (0,1)$ such that, if $\sigma_n/\tau_0 \le c_\sigma$ and $\varepsilon_n \le \varepsilon_0$, then, uniformly over $Q \in \mathcal{P}_d(\xi_{0,n}, \tau_0, L_{\mathcal{T}}, L_{\mathcal{F}})$ and $y \in \operatorname{supp}(Q)$,*

$$c\sigma_n^d \le \mu_0(y) := \mathbb{E}_Q \mathsf{k}_{y,n}(Y) \le C\sigma_n^d. \tag{72}$$

Moreover, for $1 \le q \le 8$ and nonnegative integers k, ℓ with $k + \ell \le 16$,

$$\mathbb{E}_Q\big[\mathsf{k}_{y,n}(Y)^q \|D_T(Y)\|^k \|D_N(Y)\|^\ell\big] \le C\sigma_n^d r_T^k r_N^\ell. \tag{73}$$

Proof. Recall that $x = \pi_{\mathcal{M}}(y)$; uniqueness follows from the tube and reach conditions. Write $y = x + \eta_y$, where $\eta_y \in N_x\mathcal{M}$ and $\|\eta_y\| \le \xi_{0,n}$. In the reach chart around x, a candidate sample point has the form $Y = z(v) + \xi$, with $z(v) = x + v + h_x(v)$ and, for $K_{z(v)}$-almost every ξ, $\xi \in N_{z(v)}\mathcal{M}$ and $\|\xi\| \le \xi_{0,n}$. The graph and normal-rotation bounds equation 190–equation 192 imply

$$\begin{aligned} &\|D_T(Y)\| \le C\|v\|, \\ &\|D_N(Y)\| \le C\left(\xi_{0,n} + \frac{\|v\|^2}{\tau_0}\right), \\ &\|Y - y\|^2 \ge c\|v\|^2 - C\xi_{0,n}^2. \end{aligned} \tag{74}$$

Here the harmless factor $1+\xi_{0,n}/\tau_0$ in the first line is absorbed into C. Since $(\xi_{0,n}/\sigma_n)^2 \le \varepsilon_n \le \varepsilon_0$, the last line bounds $\mathsf{k}_{y,n}(Y)^q$ by $C\exp\bigl(-c_q\|v\|^2/\sigma_n^2\bigr)$.

Use the chart $\mathcal{U}_x$ from Lemma F.1; it contains $\mathcal{M}\cap\mathbb{B}_D(x, c_{\mathrm{sep}}\tau_0)$. Integrate first with respect to the arbitrary conditional noise kernel and then in the chart variable. The noise integral has mass one, the density and chart Jacobian are uniformly bounded, and

$$\int_{\mathbb{R}^d} \exp\bigl(-c\|v\|^2/\sigma_n^2\bigr)\,\|v\|^k \left(\xi_{0,n}+\frac{\|v\|^2}{\tau_0}\right)^{\ell} dv \le C\sigma_n^{d+k}\left(\xi_{0,n}+\frac{\sigma_n^2}{\tau_0}\right)^{\ell}.$$

This proves the local upper bounds. Every latent point outside this chart is at distance at least $c_{\mathrm{sep}}\tau_0$ from x. Hence, since $2\xi_{0,n}/\tau_0 \le (\xi_{0,n}/\sigma_n)^2 + (\sigma_n/\tau_0)^2 \le \varepsilon_n$, after decreasing ε_0 if necessary, every corresponding observed point is at distance at least $c_{\mathrm{sep}}\tau_0/2$ from y. The fixed support envelope bounds all powers of D_T and D_N, so the off-chart contribution is at most

$$C\exp\left(-c\frac{\tau_0^2}{\sigma_n^2}\right).$$

For the finitely many indices in the lemma, the elementary bound

$$\exp\left(-c\frac{\tau_0^2}{\sigma_n^2}\right) \le C_{k,\ell}\sigma_n^{d+k}\left(\frac{\sigma_n^2}{\tau_0}\right)^{\ell} \le C_{k,\ell}\sigma_n^{d+k}\delta_n^{\ell}$$

absorbs this term into equation 73.

For the lower bound in equation 72, decrease c_σ if needed and restrict to $\|v\| \le \sigma_n$, which lies inside the graph domain. The density and chart Jacobian are bounded below. Moreover, for $K_{z(v)}$-almost every u,

$$\|z(v)+u-y\| \le C\left(\sigma_n+\xi_{0,n}+\frac{\sigma_n^2}{\tau_0}\right) \le C'\sigma_n.$$

The last inequality follows from $\xi_{0,n}/\sigma_n + \sigma_n/\tau_0 \le \sqrt{2\varepsilon_n}$. Thus the conditional noise integral is bounded below by a fixed positive constant, and the restricted chart has volume comparable to σ_n^d. This completes the proof. □

Define the population and empirical kernel-weighted probability measures

$$\Pi_y(dw) := \frac{\mathsf{k}_{y,n}(w)}{\mu_0(y)}Q(dw), \qquad \widehat{\Pi}_y := \frac{\sum_{i=1}^n \mathsf{k}_{y,n}(Y_i)\delta_{Y_i}}{\sum_{i=1}^n \mathsf{k}_{y,n}(Y_i)}, \tag{75}$$

and let C_y and $\widehat{C}_y$ denote their covariance matrices. For each fixed Q and query y, define the stable-denominator event

$$\mathcal{E}_{n,Q}(y) := \left\{\widehat{\mu}_0(y) \ge \frac{\mu_0(y)}{2}\right\}, \qquad \widehat{\mu}_0(y) := \frac{1}{n}\sum_{i=1}^n \mathsf{k}_{y,n}(Y_i). \tag{76}$$

Lemma D.3 (Self-normalized covariance concentration)**.** *Under the conditions of Lemma D.2, if $N_n = n\sigma_n^d \ge 1$, then, for $a, b \in \{T, N\}$ and $q \in \{2, 4\}$,*

$$\sup_{Q\in\mathcal{P}_d(\xi_{0,n},\tau_0,L_T,L_F)}\ \sup_{y\in\mathrm{supp}(Q)} \left(\mathbb{E}_{Q^n}\left[\mathbf{1}_{\mathcal{E}_{n,Q}(y)}\|P_a(\widehat{C}_y - C_y)P_b\|_F^q\right]\right)^{1/q} \le C\frac{r_a r_b}{\sqrt{N_n}}. \tag{77}$$

If, in addition, $N_n \ge C_{\mathrm{conc}}\bigl(1+\log(\tau_0/\sigma_n)\bigr)$ with C_{conc} sufficiently large, then, for $a, b \in \{T, N\}$ and $q \in \{2, 4\}$,

$$\sup_{Q\in\mathcal{P}_d(\xi_{0,n},\tau_0,L_T,L_F)}\ \sup_{y\in\mathrm{supp}(Q)} \left(\mathbb{E}_{Q^n}\|P_a(\widehat{C}_y - C_y)P_b\|_F^q\right)^{1/q} \le C\frac{r_a r_b}{\sqrt{N_n}}. \tag{78}$$

Proof. **Raw kernel-moment concentration.** Fix the query y; all expectations below are over the training sample. Let

$$\begin{aligned}\widehat{\mu}_0 &:= \frac{1}{n}\sum_i \mathsf{k}_{y,n}(Y_i),\\ \widehat{\mu}_a &:= \frac{1}{n}\sum_i \mathsf{k}_{y,n}(Y_i)D_a(Y_i),\\ \widehat{M}_{ab} &:= \frac{1}{n}\sum_i \mathsf{k}_{y,n}(Y_i)D_a(Y_i)D_b(Y_i)^\top,\end{aligned} \tag{79}$$

and denote their expectations without hats. Applying the scalar Rosenthal inequality entrywise and using equation 73 with kernel powers up to eight and total displacement degree up to sixteen gives, for $q \in \{2,4,8\}$, the following bounds, where $\|U\|_{L^q(F)} := (\mathbb{E}_{Q^n}\|U\|_F^q)^{1/q}$ for a random matrix U:

$$\left\|\frac{\widehat{\mu}_0-\mu_0}{\mu_0}\right\|_{L^q} \le \frac{C}{\sqrt{N_n}}, \tag{80}$$

$$\left\|\frac{\widehat{\mu}_a-\mu_a}{\mu_0}\right\|_{L^q} \le \frac{Cr_a}{\sqrt{N_n}}, \tag{81}$$

$$\left\|\frac{\widehat{M}_{ab}-M_{ab}}{\mu_0}\right\|_{L^q(F)} \le \frac{Cr_ar_b}{\sqrt{N_n}}. \tag{82}$$

For example, after division by $\mu_0 \asymp \sigma_n^d$, the variance term in the last display is $Cr_ar_bN_n^{-1/2}$ and the order-q Rosenthal term is $Cr_ar_bN_n^{-1+1/q}$, which is smaller for $N_n \ge 1$.

The stable-denominator event. Since $0 \le \mathsf{k}_{y,n} \le 1$, Bernstein's inequality and equation 72 yield

$$\mathbb{P}_{Q^n}(\mathcal{E}_{n,Q}(y)^c) \le 2\exp(-cN_n)\,. \tag{83}$$

On $\mathcal{E}_{n,Q}(y)$, use

$$\begin{aligned}P_a\widehat{C}_yP_b &= \frac{\widehat{M}_{ab}}{\widehat{\mu}_0} - \frac{\widehat{\mu}_a}{\widehat{\mu}_0}\left(\frac{\widehat{\mu}_b}{\widehat{\mu}_0}\right)^\top,\\ P_aC_yP_b &= \frac{M_{ab}}{\mu_0} - \frac{\mu_a}{\mu_0}\left(\frac{\mu_b}{\mu_0}\right)^\top.\end{aligned} \tag{84}$$

The population ratios are bounded by Cr_a and Cr_ar_b by Lemma D.2. Put $u_a = \mu_a/\mu_0$, $S_{ab} = M_{ab}/\mu_0$ and define their empirical counterparts with $\widehat{\mu}_0$ in the denominator. On $\mathcal{E}_{n,Q}(y)$, direct ratio algebra gives

$$\|\widehat{u}_a - u_a\| \le 2\left(\frac{\|\widehat{\mu}_a-\mu_a\|}{\mu_0} + \|u_a\|\frac{|\widehat{\mu}_0-\mu_0|}{\mu_0}\right), \tag{85}$$

$$\|\widehat{S}_{ab} - S_{ab}\|_F \le 2\left(\frac{\|\widehat{M}_{ab}-M_{ab}\|_F}{\mu_0} + \|S_{ab}\|_F\frac{|\widehat{\mu}_0-\mu_0|}{\mu_0}\right). \tag{86}$$

Furthermore,

$$\begin{aligned}\|P_a(\widehat{C}_y - C_y)P_b\|_F \le{}& \|\widehat{S}_{ab}-S_{ab}\|_F + \|\widehat{u}_a-u_a\|\,\|u_b\|\\ &+ \|u_a\|\,\|\widehat{u}_b-u_b\| + \|\widehat{u}_a-u_a\|\,\|\widehat{u}_b-u_b\|.\end{aligned} \tag{87}$$

Apply equation 80– equation 82 to these displays. Hölder's inequality with the order-eight bounds controls the last product, multiplied by $\mathbf{1}_{\mathcal{E}_{n,Q}(y)}$, in L^4 by Cr_ar_b/N_n, which is no larger than $Cr_ar_b/\sqrt{N_n}$. The analogous L^2 bound uses the order-four moments. This proves equation 77 for $N_n \ge 1$.

The exceptional event. For the unconditional bound, assume the stronger sample-size condition. On the exceptional event, the fixed support envelope gives $\|\widehat{C}_y - C_y\|_F \le C_{\mathrm{env}}$. Its contribution to

the L^q norm is therefore at most $C_{\mathrm{env}} \exp(-cN_n/q)$. Since $r_T = \sigma_n$ and $r_N \geq \sigma_n^2/\tau_0$, the smallest product among the four blocks is at least σ_n^4/τ_0^2. Choose C_{conc} so that $cC_{\mathrm{conc}}/q > 5$ for $q \in \{2, 4\}$. Put $N_0 = C_{\mathrm{conc}}\big(1 + \log(\tau_0/\sigma_n)\big)$. Once $N_0 \geq q/(2c)$, $N \mapsto \sqrt{N}\exp(-cN/q)$ is decreasing, and hence

$$\sup_{N \geq N_0} \sqrt{N}\exp(-cN/q) \leq \sqrt{N_0}\exp(-cN_0/q)\,.$$

The choice of exponent absorbs the logarithmic factor. Therefore, uniformly for $N_n \geq N_0$,

$$C_{\mathrm{env}} \exp(-cN_n/q) \leq C\frac{\sigma_n^4}{\tau_0^2\sqrt{N_n}} \leq C\frac{r_a r_b}{\sqrt{N_n}}. \tag{88}$$

Here $\sigma_n/\tau_0 \leq c_\sigma < 1$ and the fixed support and reach envelopes absorb the bounded-scale part and τ_0 into the constant. This proves the exceptional-event bound and completes the proof. □

D.2.3 Blockwise Hessian concentration

There exist fixed constants $0 < \sigma_0 < \tau_0$, $c_\sigma, b_0 > 0$, and $C_{\mathrm{conc}} < \infty$ such that the following conditions suffice:

$$\begin{aligned} 0 < \sigma_n \leq \sigma_0, \qquad \frac{\sigma_n}{\tau_0} \leq c_\sigma, \qquad \varepsilon_n \leq \varepsilon_0, \\ N_n \geq C_{\mathrm{conc}}\big(1 + \log(\tau_0/\sigma_n)\big), \qquad b_n := C_0\varepsilon_n \leq b_0. \end{aligned} \tag{89}$$

For an admissible representation, let P_T, P_N denote the tangent and normal projectors at the nearest-point projection of the query onto the chosen latent manifold and set

$$\widetilde{\Delta}_n := \widetilde{A}_{n,\sigma_n} - A_{Q,\sigma_n}, \qquad \delta_n := \xi_{0,n} + \frac{\sigma_n^2}{\tau_0}. \tag{90}$$

Here δ_n combines actual tube width and curvature-induced normal spread over a tangent neighborhood of radius σ_n.

Theorem D.1 (Blockwise concentration for the Gaussian-kernel LHSD estimator)**.** *Under equation 89, uniformly over the tubular class and the admissible representation,*

$$\Big(\mathbb{E}_{Q^n}\|P_T\widetilde{\Delta}_n P_T\|^2_{L^2(Q;F)}\Big)^{1/2} \lesssim N_n^{-1/2}, \tag{91}$$

$$\Big(\mathbb{E}_{Q^n}\|P_T\widetilde{\Delta}_n P_N\|^2_{L^2(Q;F)}\Big)^{1/2} \lesssim (\delta_n/\sigma_n)N_n^{-1/2}, \tag{92}$$

$$\Big(\mathbb{E}_{Q^n}\|P_N\widetilde{\Delta}_n P_N\|^2_{L^2(Q;F)}\Big)^{1/2} \lesssim (\delta_n/\sigma_n)^2 N_n^{-1/2}, \tag{93}$$

$$\Big(\mathbb{E}_{Q^n}\|\widetilde{\Delta}_n\|^4_{L^4(Q;F)}\Big)^{1/2} \lesssim N_n^{-1}. \tag{94}$$

Proof of Theorem D.1. Fix the admissible representation used to define P_T and P_N in the statement; every bound below is uniform over this choice. The conditional covariance identity equation 1 and equation 68 give, pointwise in the independent query y,

$$\widetilde{\Delta}_n(y) = -\frac{\widehat{C}_y - C_y}{\sigma_n^2}. \tag{95}$$

Apply Lemma D.3 with $q = 2$. Division by σ_n^2 gives the following blockwise root-mean-square rates:

$$\begin{aligned} TT \text{ block:} \quad & \frac{1}{\sqrt{N_n}}, \\ TN \text{ block:} \quad & \frac{\delta_n/\sigma_n}{\sqrt{N_n}}, \\ NN \text{ block:} \quad & \frac{(\delta_n/\sigma_n)^2}{\sqrt{N_n}}. \end{aligned} \tag{96}$$

Moreover,

$$\frac{\delta_n}{\sigma_n} = \frac{\xi_{0,n}}{\sigma_n} + \frac{\sigma_n}{\tau_0} \leq \sqrt{2\varepsilon_n} \lesssim \sqrt{b_n}, \tag{97}$$

so $(\delta_n/\sigma_n)^2 \lesssim b_n$. The $q = 4$ bound, summed over the four fixed-dimensional blocks, gives

$$\sup_{y\in\mathrm{supp}(Q)} \mathbb{E}_{Q^n}\|\widetilde{\Delta}_n(y)\|_F^4 \le CN_n^{-2}.$$

Finally, integrate in $y \sim Q$. Tonelli's theorem gives equation 91–equation 94. □

D.2.4 SPECTRAL TRANSFER AND RISK BOUNDS

The next lemma controls the response of the LHSD filter to the three Hessian blocks. Their respective weights are of orders b, η, and 1, where $b \asymp \varepsilon_n$ and $\eta \asymp \sqrt{\varepsilon_n}$.

Lemma D.4 (Anisotropic affine-filter transfer). *Let $P_T + P_N = I_D$ be orthogonal projectors of ranks d and $D-d$. Suppose $C_0, C_1 > 0$ are fixed and $b = C_0\eta^2$. There exists $c_{\mathrm{an}} = c_{\mathrm{an}}(C_0, C_1) > 0$ such that, if $0 < \eta \le c_{\mathrm{an}}$ and the symmetric matrix A satisfies*

$$\begin{aligned}\|P_T A P_T\|_{\mathrm{op}} + \|P_N(A - I_D)P_N\|_{\mathrm{op}} &\le C_1 b,\\ \|P_T A P_N\|_{\mathrm{op}} &\le C_1\eta.\end{aligned} \tag{98}$$

Then every symmetric H satisfies

$$|\operatorname{tr}\bar\phi_\varpi(A+H) - \operatorname{tr}\bar\phi_\varpi(A)| \le C_\varpi\big(b\|P_THP_T\|_F + \eta\|P_THP_N\|_F + \|P_NHP_N\|_F + \|H\|_F^2\big). \tag{99}$$

where C_ϖ depends only on ϖ, C_0, C_1, and the fixed ambient dimension D.

Proof. For η small, Lemma F.4 gives d eigenvalues of A in an $O(b)$ neighborhood of zero, $D-d$ eigenvalues in an $O(b)$ neighborhood of one, and a spectral gap bounded below by a numerical constant. Let Π_T and Π_N be the corresponding spectral projectors. The block eigenvector equation used in equation 216 gives

$$\|P_T - \Pi_T\|_{\mathrm{op}} = \|P_N - \Pi_N\|_{\mathrm{op}} \le C\eta. \tag{100}$$

By the Fréchet derivative formula for spectral functions [14], the derivative of $A \mapsto \operatorname{tr}\bar\phi_\varpi(A)$ at A is $H \mapsto \operatorname{tr}\big(\bar\phi'_\varpi(A)H\big)$. Since $\bar\phi'_\varpi(0) = 0$ with cubic vanishing, $\bar\phi'_\varpi(1) \ne 0$, and the two spectral clusters have radius $O(b)$, equation 100 implies, for $G = \bar\phi'_\varpi(A)$,

$$\|P_TGP_T\|_{\mathrm{op}} \le C_\varpi b, \qquad \|P_TGP_N\|_{\mathrm{op}} \le C_\varpi\eta, \qquad \|P_NGP_N\|_{\mathrm{op}} \le C_\varpi. \tag{101}$$

Indeed, the low-cluster part of the first block is $O_\varpi(b^3)$ and the leakage of the high cluster is $O_\varpi(\eta^2) = O_\varpi(b)$.

Finally, $\sup_{u\in\mathbb{R}}|\bar\phi''_\varpi(u)| < \infty$. The corresponding second-order trace-functional Taylor formula, equivalently the divided-difference formula in an eigenbasis along $A + tH$ [14], gives

$$D^2\big[\operatorname{tr}\bar\phi_\varpi\big](B)[H,H] = \sum_{i,j}(\bar\phi'_\varpi)^{[1]}(\lambda_i,\lambda_j)|(U^\top HU)_{ij}|^2, \qquad B = U\operatorname{diag}(\lambda_i)U^\top,$$

where $g^{[1]}(a,b) := (g(a) - g(b))/(a-b)$ for $a \ne b$ and $g^{[1]}(a,a) := g'(a)$. Thus, all coefficients are bounded by $\sup|\bar\phi''_\varpi|$. Hence,

$$|\operatorname{tr}\bar\phi_\varpi(A+H) - \operatorname{tr}\bar\phi_\varpi(A) - \operatorname{tr}(GH)| \le C_\varpi\|H\|_F^2. \tag{102}$$

Decompose $\operatorname{tr}(GH)$ into its TT, TN, NN blocks, apply equation 101 and the Frobenius Cauchy–Schwarz inequality, and absorb the factor two on the mixed blocks. This proves equation 99. □

The tangent, mixed, and normal blocks fluctuate at different geometric scales. The flat tangent response of the LHSD filter, together with the tangent–normal covariance structure and the spectral-projector bounds, attenuates the first two blocks so that all linear contributions have order $\varepsilon_n/\sqrt{N_n}$; the quadratic spectral remainder contributes N_n^{-1}.

Theorem D.2 (Finite-sample risk of the Gaussian-kernel LHSD estimator). *There exist $\varepsilon_{\mathrm{est}} > 0$ and $1 \le C_{\mathrm{est}} < \infty$, depending only on the fixed model parameters, such that if*

$$0 < \sigma_n < \tau_0, \qquad \varepsilon_n \le \varepsilon_{\mathrm{est}}, \qquad N_n \ge C_{\mathrm{est}},$$

then

$$\sup_{Q\in\mathcal{P}_d(\xi_{0,n},\tau_0,L_{\mathcal{T}},L_{\mathcal{F}})} \mathbb{E}_{Q^n}\|\widetilde{T}_n - T_n(Q)\|_{L^2(Q)}^2 \lesssim_\varpi \frac{\varepsilon_n^2}{N_n} + \frac{1}{N_n^2}. \tag{103}$$

Proof of Theorem D.2. Choose

$$\varepsilon_{\mathrm{est}} \le \varepsilon_0 \wedge c_\sigma^2 \wedge c_{\mathrm{an}}^2 \wedge \frac{\sigma_0^2}{\tau_0^2} \wedge \frac{b_0}{C_0}, \qquad C_{\mathrm{est}} \ge 1. \tag{104}$$

Then the assumptions of Theorem D.2 imply the conditions of Lemma D.2 and the restricted moment bound equation 77. The localized covariance-block bounds imply equation 98 with $\eta := \sqrt{\varepsilon_n}$ and $b := b_n = C_0\varepsilon_n = C_0\eta^2$, after increasing C_1 if needed. Since $\eta \asymp \sqrt{b_n}$, apply Lemma D.4 with $H = \widetilde{\Delta}_n$, multiply by $\mathbf{1}_{\mathcal{E}_{n,Q}(y)}$, and use Minkowski's inequality in the joint training–query L^2 norm. This gives

$$\begin{aligned}\sup_{Q\in\mathcal{P}_d(\xi_{0,n},\tau_0,L_{\mathcal{T}},L_{\mathcal{F}})} &\left(\mathbb{E}_{Q^n}\int \mathbf{1}_{\mathcal{E}_{n,Q}(y)}|\widetilde{T}_n(y)-T_n(Q)(y)|^2\,Q(dy)\right)^{1/2}\\ &\le C_\varpi\big(b_n e_{TT,2} + \sqrt{b_n}\,e_{TN,2} + e_{NN,2} + e_4^2\big),\end{aligned} \tag{105}$$

Here the restricted moments are

$$e_{ab,2} := \sup_{Q,\,\mathrm{rep}}\left(\mathbb{E}_{Q^n}\int \mathbf{1}_{\mathcal{E}_{n,Q}(y)}\|P_a\widetilde{\Delta}_n(y)P_b\|_F^2\,Q(dy)\right)^{1/2},$$

$$e_4 := \sup_{Q}\left(\mathbb{E}_{Q^n}\int \mathbf{1}_{\mathcal{E}_{n,Q}(y)}\|\widetilde{\Delta}_n(y)\|_F^4\,Q(dy)\right)^{1/4},$$

where the suprema range over the current tubular class and, for the block moments, its admissible representations. The event and integrands are jointly Borel in the sample and query: both kernel masses are continuous, and the projectors are continuous on the reach tube for each fixed representation. Thus Tonelli's theorem, equation 95, and equation 77 give

$$e_{TT,2} \lesssim N_n^{-1/2}, \qquad e_{TN,2} \lesssim \frac{\delta_n/\sigma_n}{\sqrt{N_n}},$$

$$e_{NN,2} \lesssim \frac{(\delta_n/\sigma_n)^2}{\sqrt{N_n}}, \qquad e_4^2 \lesssim N_n^{-1}.$$

Here the last bound follows by summing the restricted fourth moments of the four blocks and using equation 97. Substituting these bounds and $b_n = C_0\varepsilon_n$ into equation 105, and then squaring, bounds the good-event contribution by $C_\varpi(\varepsilon_n^2/N_n + N_n^{-2})$.

For the exceptional event, the affine calibration gives $-\varpi^4 < \bar{\phi}_\varpi(\lambda) \le 1$ for every real λ. Consequently, both LHSD traces lie in $(-D\varpi^4, D]$, and $|\widetilde{T}_n(y) - T_n(Q)(y)| \le D(1+\varpi^4)$. Using equation 83 pointwise in y and then Tonelli's theorem yields, uniformly over Q,

$$\mathbb{E}_{Q^n}\int \mathbf{1}_{\mathcal{E}_{n,Q}(y)^c}|\widetilde{T}_n(y)-T_n(Q)(y)|^2\,Q(dy) \le 2D^2(1+\varpi^4)^2 e^{-cN_n} \le C_\varpi N_n^{-2}.$$

Adding the two contributions proves equation 103. □

The next proposition gives a finite-sample form of Theorem 3.2.

Proposition D.3 (Minimax upper bound for finite-scale LHSD)**.** *Suppose the conditions of Theorem D.2 hold and, for fixed constants $C_\xi, C_{\mathrm{den}}, c_{\mathrm{rem}} > 0$,*

$$\xi_{0,n} \le C_\xi\frac{\sigma_n^2}{\tau_0}, \qquad \tau_0^2\Lambda_f \le C_{\mathrm{den}}, \qquad N_n\sigma_n^4 \ge c_{\mathrm{rem}}.$$

Then

$$\begin{aligned}\mathfrak{R}^{\mathrm{LHSD}}_{n,\sigma_n}(\xi_{0,n}) &\le \sup_{Q\in\mathcal{P}_d(\xi_{0,n},\tau_0,L_{\mathcal{T}},L_{\mathcal{F}})}\mathbb{E}_{Q^n}\|\widetilde{T}_n - T_n(Q)\|^2_{L^2(Q)}\\ &\lesssim_\varpi \frac{\sigma_n^4}{N_n} = \frac{\sigma_n^4}{n\sigma_n^d}.\end{aligned} \tag{106}$$

where the constant may also depend on $C_\xi, C_{\mathrm{den}}, c_{\mathrm{rem}}$ and the fixed model envelopes.

Proof of Proposition D.3. Under the conditions of Proposition D.3,

$$\varepsilon_n \le (C_\xi^2 + 1 + C_{\mathrm{den}})\frac{\sigma_n^2}{\tau_0^2}, \qquad \frac{1}{N_n^2} \le \frac{1}{c_{\mathrm{rem}}}\frac{\sigma_n^4}{N_n}.$$

Substituting these inequalities into equation 103 and absorbing the fixed factor τ_0^{-4} into the constant gives the estimator risk bound in equation 106. Taking the infimum over all estimators gives the first inequality in that display.

To obtain Theorem 3.2, fix the model parameters. The density and atlas bounds give a finite classwise Λ_f, so a fixed C_{den} can be chosen with $\tau_0^2\Lambda_f \le C_{\mathrm{den}}$. If $n^{-1/(2\alpha+d)} \le \sigma_n \to 0$ and $\xi_{0,n} \le C_\xi\sigma_n^2/\tau_0$, then $\varepsilon_n = O(\sigma_n^2/\tau_0^2) \to 0$ and, since $\alpha \ge \beta \ge 2$,

$$N_n\sigma_n^4 \ge n^{(2\alpha-4)/(2\alpha+d)} \ge 1, \qquad N_n \ge n^{2\alpha/(2\alpha+d)} \longrightarrow \infty.$$

Taking $c_{\mathrm{rem}} = 1$, all conditions above hold for sufficiently large n, including when $\alpha = 2$. Finally, substituting the empirical measure $n^{-1}\sum_{i=1}^n \delta_{Y_i}$ into Definition 3.1 gives exactly the estimator in equation 69, which proves the asserted attainment. □

D.3 Matching rates and comparison with FLIPD

Combining Propositions D.1 and D.3 gives, whenever both sets of conditions hold,

$$\mathfrak{R}^{\mathrm{LHSD}}_{n,\sigma_n}(\xi_{0,n}) \asymp_{\varpi} \frac{\sigma_n^4}{n\sigma_n^d}, \tag{107}$$

attained by the Gaussian-kernel reference estimator $\widetilde{T}_n$.

The common asymptotic regime of Theorems 3.1 and 3.2 therefore gives the matching rate whenever the curved-submodel inclusion and thin-tube conditions both hold.

For the finite-scale FLIPD field, Seo et al. [47, Theorem 3.2] establish a minimax squared-risk lower bound of order $1/(n\sigma^d)$ on a fixed regular manifold. When a full density class covered by that result is contained in our exact-support class and the scales satisfy both results, the LHSD upper bound has an additional factor σ_n^4 relative to the FLIPD lower bound. The comparison concerns two different finite-scale fields; accuracy for the integer LMD also depends on their approximation biases.

The local-sample-size condition is substantive. If the normalized Gaussian weights are dominated by one observation, the weighted covariance collapses, $\widetilde{A}_{n,\sigma_n} \approx I_D$, and $\widetilde{T}_n \approx 0$. The theorem is therefore a fixed-dimensional, sufficiently populated local-neighborhood result, not a guarantee for sparse high-dimensional neighborhoods.

D.4 Independent-query recovery of the latent dimension

Fix a scale σ_n and the tubular model class $\mathcal{P}_d(\xi_{0,n}, \tau_0, L_{\mathcal{T}}, L_{\mathcal{F}})$ on which the LMD is well-defined. Write

$$B_n := \mathfrak{B}_{\xi_{0,n}}(\sigma_n) < 1/2, \qquad R_n := \sup_{Q\in\mathcal{P}_d(\xi_{0,n},\tau_0,L_{\mathcal{T}},L_{\mathcal{F}})} \mathbb{E}_{Q^n}\|\widetilde{T}_n - T_{\sigma_n,\varpi}(Q)\|^2_{L^2(Q)}.$$

For a query $Y' \sim Q$ independent of $Y_1, \dots, Y_n$, rounding to the nearest integer, with any fixed tie-breaking rule, satisfies

$$\sup_{Q\in\mathcal{P}_d(\xi_{0,n},\tau_0,L_{\mathcal{T}},L_{\mathcal{F}})} \mathbb{P}_{Q^{n+1}}\Big\{\mathrm{round}\big(\widetilde{T}_n(Y')\big) \ne d_{\mathrm{LMD}}(Q)\Big\} \le 1 \wedge \frac{R_n}{(1/2 - B_n)^2}.$$

Proof. Fix a law Q in the stated class and write $d_Q = d_{\mathrm{LMD}}(Q)$. Incorrect rounding implies $|\widetilde{T}_n(Y') - d_Q| \ge 1/2$. Since $|T_{\sigma_n,\varpi}(Q)(Y') - d_Q| \le B_n$ almost surely,

$$\Big\{\mathrm{round}\big(\widetilde{T}_n(Y')\big) \ne d_Q\Big\} \subseteq \Big\{|\widetilde{T}_n(Y') - T_{\sigma_n,\varpi}(Q)(Y')| \ge 1/2 - B_n\Big\}.$$

Markov's inequality and independence of the query give

$$\mathbb{P}_{Q^{n+1}}\Big\{\mathrm{round}\big(\widetilde{T}_n(Y')\big) \ne d_Q\Big\} \le \frac{\mathbb{E}_{Q^n}\|\widetilde{T}_n - T_{\sigma_n,\varpi}(Q)\|^2_{L^2(Q)}}{(1/2 - B_n)^2}.$$

Taking the supremum over Q and using the trivial probability bound one proves the claim. □

Under the conditions of Theorem D.2, $R_n \lesssim_\varpi \varepsilon_n^2/N_n + N_n^{-2}$. Under the joint conditions of Theorems 3.2 and 4.2, the attaining reference estimator instead satisfies $R_n \lesssim_\varpi \sigma_n^4/(n\sigma_n^d)$ for all sufficiently large n. If $B_n \le 1/4$, then $(1/2 - B_n)^{-2} \le 16$, giving the main-text recovery bound.

E FINE-SCALE NOISE SENSITIVITY

The proof follows the chain

$$\text{normal-fiber law} \longrightarrow \text{local } (d+k)\text{-dimensional sheets} \longrightarrow A_{Q,\sigma} \approx P_N^{(d+k)} \longrightarrow T_{\sigma,\varpi} \approx d+k.$$

We first construct kernels with matched first and second moments, then analyze their component sheets. Two localization lemmas separate intersecting sheets, and the concluding proof establishes the population LHSD limits. The construction works for every codimension $D-d$. For example, when $D-d=2$, $k=0$ gives mixtures of isolated coordinate points in each normal fiber, $k=1$ gives one-dimensional circular sheets, and $k=2$ fills the normal disk.

E.1 MATCHED-MOMENT NORMAL-NOISE CONSTRUCTION

Let $\omega_d := \mathcal{H}^d(\mathbb{S}^d)$ and choose $R > 0$ so that $f_- \le (\omega_d R^d)^{-1} \le f_+$. Such an R exists because $R \mapsto (\omega_d R^d)^{-1}$ ranges over $(0,\infty)$. Set $\tau_0 := 3R/4$ and write

$$\mathbb{R}^D = \mathbb{R}^{d+1} \times \mathbb{R}^{D-d-1}.$$

Let

$$\mathcal{M} = \mathbb{S}^d(R) \times \{0\}^{D-d-1} \subset \mathbb{R}^{d+1} \times \mathbb{R}^{D-d-1}, \tag{108}$$

and let $X = (RU, 0)$, where U is uniform on the unit sphere $\mathbb{S}^d$. Then $\mathcal{M}$ is compact, has no boundary, and $\operatorname{reach}(\mathcal{M}) = R$. At $x = (Ru, 0)$, define the linear map $U_x : \mathbb{R}^{D-d} \to N_x\mathcal{M} \subset \mathbb{R}^D$ by $U_x z = (z_1 u, z_2, \dots, z_{D-d})$. Its columns form an orthonormal normal frame:

$$U_x e_1 = (u, 0), \qquad U_x e_j = (0, e_{j-1}), \quad j = 2, \dots, D-d. \tag{109}$$

For $D-d=1$, only the first vector is present. Consider arbitrary positive sequences $(\xi_{0,n})$ and (σ_n) satisfying $\xi_{0,n} \le \tau_0/3$ and $\sigma_n/\xi_{0,n} \to 0$.

Conditional noise kernels. In normal coordinates, put

$$r_n := \xi_{0,n}\sqrt{\frac{D-d}{D-d+2}}. \tag{110}$$

This radius is chosen so that an isotropic sphere mixture of radius r_n has the same covariance $\xi_{0,n}^2 I_{D-d}/(D-d+2)$ as the uniform distribution on the filled $(D-d)$-ball of radius $\xi_{0,n}$. The idea is to choose each conditional noise probability measure as a mixture of k-spheres in the normal fiber. Locally this creates $(d+k)$-dimensional sheets, while averaging over coordinate subspaces restores the same isotropic covariance for every k. For $0 \le k < D-d$, let

$$\mathcal{I}_{k+1} := \{S \subset [D-d] : |S| = k+1\}, \qquad E_S := \operatorname{span}\{e_j : j \in S\},$$

write $\omega_k := \mathcal{H}^k(\mathbb{S}^k)$ for the surface area of the unit k-sphere, where $\mathcal{H}^k$ is k-dimensional Hausdorff measure ($\mathcal{H}^0$ is counting measure), and define the probability measure

$$\nu_{S,n} := \frac{\mathcal{H}^k \restriction \{z \in E_S : \|z\| = r_n\}}{\omega_k r_n^k}.$$

Then set

$$\kappa_n^{(k)} := \pi_k \sum_{S \in \mathcal{I}_{k+1}} \nu_{S,n}, \qquad \pi_k := \binom{D-d}{k+1}^{-1}. \tag{111}$$

For $k=0$, normalized $\mathcal{H}^0$ gives mass $1/2$ to each point in $\{-r_n e_j, r_n e_j\}$. For $k = D-d$, set

$$\kappa_n^{(D-d)} := \operatorname{Unif}(\mathbb{B}_{D-d}(0, \xi_{0,n})). \tag{112}$$

For every n and $k \in [D-d]_0$, define

$$K_{x,n}^{(k)} := (U_x)_{\#}\kappa_n^{(k)}.$$

Here $(U_x)_{\#}$ denotes pushforward: for every Borel set $A \subset \mathbb{R}^D$, $K_{x,n}^{(k)}(A) = \kappa_n^{(k)}(\{z : U_x z \in A\})$. The frame in equation 109 depends smoothly on $x \in \mathcal{M}$. Hence, for every bounded Borel function $\psi : \mathbb{R}^D \to \mathbb{R}$,

$$x \longmapsto \int_{\mathbb{R}^D} \psi(v)\, K_{x,n}^{(k)}(dv) = \int_{\mathbb{R}^{D-d}} \psi(U_x z)\, \kappa_n^{(k)}(dz)$$

is Borel measurable. Thus, $x \mapsto K_{x,n}^{(k)}$ is a Borel probability kernel supported on $N_x\mathcal{M}$.

For every n, k, take $\zeta_n^{(k)} \sim \kappa_n^{(k)}$ independently of X and set $\xi_n^{(k)} := U_X \zeta_n^{(k)}$. Its conditional law given $X = x$ is the kernel $K_{x,n}^{(k)}$ defined above. Every $\kappa_n^{(k)}$ is centrally symmetric and supported in $\overline{\mathbb{B}_{D-d}(0, \xi_{0,n})}$. Since $U_x^\top U_x = I_{D-d}$, the map U_x is an isometry and preserves norms, so

$$\mathbb{E}(\xi_n^{(k)} \mid X = x) = U_x \mathbb{E}\zeta_n^{(k)} = 0, \qquad \|\xi_n^{(k)}\| = \|U_X \zeta_n^{(k)}\| = \|\zeta_n^{(k)}\| \le \xi_{0,n} \quad \text{a.s.} \tag{113}$$

For a fixed $S \in \mathcal{I}_{k+1}$, let $P_{E_S} = \sum_{j \in S} e_j e_j^\top$ be the orthogonal projector onto E_S. Sign symmetry makes the cross moments zero, while symmetry among the $k+1$ active coordinates and $\|z\|^2 = r_n^2$ give

$$\int z z^\top \, d\nu_{S,n}(z) = \frac{r_n^2}{k+1} P_{E_S}. \tag{114}$$

Since each coordinate belongs to $\binom{D-d-1}{k}$ of the $\binom{D-d}{k+1}$ subsets,

$$\begin{aligned} \binom{D-d}{k+1}^{-1} \sum_{S \in \mathcal{I}_{k+1}} P_{E_S} &= \frac{k+1}{D-d} I_{D-d}, \\ \mathrm{Cov}(\kappa_n^{(k)}) = \frac{r_n^2}{D-d} I_{D-d} &= \frac{\xi_{0,n}^2}{D-d+2} I_{D-d}, \qquad 0 \le k < D-d. \end{aligned} \tag{115}$$

The factors $1/(k+1)$ and $(k+1)/(D-d)$ cancel, so this covariance is independent of k. For the uniform ball, radial integration gives $\mathbb{E}\|z\|^2 = (D-d)\xi_{0,n}^2/(D-d+2)$; rotational symmetry therefore gives the same covariance for $k = D-d$. Since U_x is a linear isometry from $\mathbb{R}^{D-d}$ onto $N_x\mathcal{M}$,

$$\begin{aligned} \mathrm{Cov}(\xi_n^{(k)} \mid X = x) &= U_x \,\mathrm{Cov}(\kappa_n^{(k)}) U_x^\top \\ &= \frac{\xi_{0,n}^2}{D-d+2} U_x U_x^\top = \frac{\xi_{0,n}^2}{D-d+2} P_{N_x\mathcal{M}}. \end{aligned} \tag{116}$$

Equations equation 113 and equation 116 prove equation 3. For $k < D-d$, the actual support radius is $r_n < \xi_{0,n}$; the common quantity is the prescribed almost-sure upper bound.

Since $\mathcal{M}$ and its uniform latent density are fixed, choose a finite geometric envelope $L_{\mathcal{T},0}$ containing its atlas, support, and volume bounds, and choose $L_{\mathcal{F},0} \ge (\omega_d R^d)^{-1}$. The density bounds f_-, f_+ hold by the choice of R, and the constant density has vanishing positive-order derivatives in every chart. These choices are independent of n, k, and the two scale sequences. With $\tau_0 = 3R/4$,

$$2\xi_{0,n} \le R/2 < \tau_0, \qquad \mathrm{reach}(\mathcal{M}) = R \ge \tau_0.$$

Consequently, $Q_n^{(k)} \in \mathcal{P}_d(\xi_{0,n}, \tau_0, L_{\mathcal{T},0}, L_{\mathcal{F},0})$, and Theorem 2.1 gives $d_{\mathrm{LMD}}(Q_n^{(k)}; \xi_{0,n}, \tau_0, L_{\mathcal{T},0}, L_{\mathcal{F},0}) = d$. This proves equation 4.

E.2 Geometry and localization on component sheets

The tubular map

$$F(x, z) := x + U_x z, \qquad x \in \mathcal{M}, \quad z \in \mathbb{B}_{D-d}(0, R), \tag{117}$$

is one-to-one: writing $F((Ru, 0), z) = (w^{(1)}, w^{(2)})$, its inverse is $u = w^{(1)}/\|w^{(1)}\|$, $z_1 = \|w^{(1)}\| - R$, and $(z_2, \ldots, z_{D-d}) = w^{(2)}$, since $R + z_1 > 0$. Here z records the $D-d$ normal coordinates in the frame U_x. For each $S \in \mathcal{I}_{k+1}$ define

$$\mathcal{N}_{S,n} := F\big(\mathcal{M} \times \{z \in E_S : \|z\| = r_n\}\big). \tag{118}$$

Let $\mu_{\mathcal{M}} = \mathrm{vol}_d(\mathcal{M})^{-1}\,\mathrm{vol}_d \restriction \mathcal{M}$ be the uniform probability distribution of X on $\mathcal{M}$. For $0 \le k < D-d$ define the component probability measure

$$Q_{S,n} := F_{\#}(\mu_{\mathcal{M}} \otimes \nu_{S,n}), \qquad Q_n^{(k)} = \pi_k \sum_{S \in \mathcal{I}_{k+1}} Q_{S,n}. \tag{119}$$

The domain defining each component is the product of the d-dimensional base and a k-sphere, so its dimension is $p = d + k$. Recall that r_n in equation 110 is the fiber-sphere radius, with $r_n \asymp \xi_{0,n}$ and $r_n < R/4$.

Lemma E.1 (Scale-relative geometry of the component sheets)**.** *Put $p = d + k$. Each $\mathcal{N}_{S,n}$ is a compact, possibly disconnected, embedded p-manifold and*

$$\mathrm{reach}(\mathcal{N}_{S,n}) \ge c_{d,D-d} r_n. \tag{120}$$

The component probability measure $Q_{S,n}$ has a density $h_{S,n}$ with respect to p-dimensional volume satisfying

$$\begin{aligned} cR^{-d} r_n^{-k} \le h_{S,n} \le CR^{-d} r_n^{-k}, \\ \|\nabla \log h_{S,n}\| \le C/R. \end{aligned} \tag{121}$$

If $\sigma_n / r_n \to 0$, then

$$\sup_{y \in \mathcal{N}_{S,n}} \|A_{Q_{S,n},\sigma_n}(y) - P_{N_y \mathcal{N}_{S,n}}\|_{\mathrm{op}} \le C \frac{\sigma_n}{r_n} = o(1). \tag{122}$$

All constants are independent of n and S. The subscript in $c_{d,D-d}$ allows dependence on the fixed dimensions; any dependence on k can be absorbed by taking the minimum over its finitely many values $0 \le k < D - d$.

Proof. **Embedding and reach.** Use the explicit coordinates in equation 109. If $x = (Ru, 0)$, then

$$F(x, z) = \big((R + z_1)u, z_2, \ldots, z_{D-d}\big). \tag{123}$$

We write $F(u, z)$ for $F((Ru, 0), z)$. For $\dot{u} \perp u$ and $\dot{z} \in E_S \cap z^{\perp}$, differentiation gives $DF_{(u,z)}(\dot{u}, \dot{z}) = ((R + z_1)\dot{u} + \dot{z}_1 u, \dot{z}_2, \ldots, \dot{z}_{D-d})$. The cross term vanishes because $\dot{u} \perp u$, hence

$$\|DF_{(u,z)}(\dot{u}, \dot{z})\|^2 = (R + z_1)^2 \|\dot{u}\|^2 + \|\dot{z}\|^2. \tag{124}$$

Since $R + z_1 \ge R - r_n > 3R/4$, this norm vanishes only when $(\dot{u}, \dot{z}) = 0$, proving immersion. It also gives uniform metric bounds after scaling the base and fiber coordinates by R and r_n, respectively. Moreover, for any two domain points,

$$\begin{aligned} &\|F(u, z) - F(u', z')\|^2 \\ &\quad = (z_1 - z_1')^2 + (R + z_1)(R + z_1')\|u - u'\|^2 + \|(z_2, \ldots, z_{D-d}) - (z_2', \ldots, z_{D-d}')\|^2 \\ &\quad \asymp R^2 \|u - u'\|^2 + \|z - z'\|^2. \end{aligned} \tag{125}$$

Thus, the immersion is an embedding.

Put $\Delta u = u' - u$, $\Delta z = z' - z$, $a = P_{u^{\perp}} \Delta u$, and $b = P_{E_S \cap z^{\perp}} \Delta z$. Thus a and b are the tangent projections of the base and fiber increments, respectively. The identities $\langle u, \Delta u \rangle = -\|\Delta u\|^2/2$ and $\langle z, \Delta z \rangle = -\|\Delta z\|^2/2$ give

$$\Delta u - a = -\frac{\|\Delta u\|^2}{2} u, \qquad \Delta z - b = -\frac{\|\Delta z\|^2}{2 r_n^2} z,$$

and the preceding differential formula, evaluated at (a, b), gives

$$DF_{(u,z)}(a, b) = \big((R + z_1)a + b_1 u, b_2, \ldots, b_{D-d}\big).$$

Direct subtraction in equation 123 therefore gives

$$\begin{aligned} &\mathrm{dist}\big(F(u', z') - F(u, z), T_{F(u,z)} \mathcal{N}_{S,n}\big) \\ &\le C \left(R\|\Delta u\|^2 + \frac{\|\Delta z\|^2}{r_n} + \|\Delta u\|\,\|\Delta z\| \right) \quad \text{(subtract } DF_{(u,z)}(a, b)\text{)} \\ &\le \frac{C}{r_n} \left(R^2 \|\Delta u\|^2 + \|\Delta z\|^2 \right) \quad (r_n \le R/4\text{, Young's inequality}) \\ &\le \frac{C}{r_n} \|F(u', z') - F(u, z)\|^2 \quad \text{(by equation 125)}, \end{aligned} \tag{126}$$

Federer's secant characterization

$$\operatorname{reach}(\mathcal{N}_{S,n}) = \inf_{\substack{a,b\in\mathcal{N}_{S,n}\\ a\neq b}} \frac{\|a-b\|^2}{2\operatorname{dist}(b-a, T_a\mathcal{N}_{S,n})}$$

proves equation 120. For $k=0$, the same argument applies within each of the two sheets, while their mutual separation is $2r_n$: for $S=\{1\}$ they are concentric spheres of radii $R\pm r_n$; for $S=\{j\}$, $j>1$, they are translates by $\pm r_n e_{j-1}$ in an orthogonal direction. A point at distance less than r_n cannot have nearest points on both sheets, so their union also has reach at least cr_n.

Density and volume bounds. Recall that $\omega_j := \mathcal{H}^j(\mathbb{S}^j)$ is the unit j-sphere's surface area; thus $\omega_j r^j$ is its area at radius r, and $\omega_0 = 2$. Relative to the product of Riemannian volume on $\mathcal{M}$ and surface measure on $\{z\in E_S : \|z\| = r_n\}$, the Jacobian of F is $J_F(x,z) = (1+z_1/R)^d$. Since the base and fiber probability measures are normalized volume measures, the component density is exactly

$$h_{S,n}(F(x,z)) = \frac{(1+z_1/R)^{-d}}{\operatorname{vol}_d(\mathcal{M})\,\omega_k r_n^k}. \tag{127}$$

This proves the two-sided density bound. Moreover,

$$\log h_{S,n}(F(x,z)) = \text{const} - d\log(1+z_1/R).$$

To control the intrinsic gradient of z_1, write an ambient point as $w=(w^{(1)},w^{(2)})\in\mathbb{R}^{d+1}\times\mathbb{R}^{D-d-1}$ and use the extension $\zeta(w) := \|w^{(1)}\| - R$. On the sheet, $\zeta(F(x,z)) = z_1$ and $\|w^{(1)}\| = R + z_1 \geq 3R/4$. Since $\|\nabla_{\mathcal{N}_{S,n}}\zeta\| \leq \|\nabla\zeta\| = 1$, the chain rule applied to $-d\log(1+\zeta/R)$ gives $\|\nabla \log h_{S,n}\| \leq C/R$, proving equation 121. For $k=0$, the gradient is understood componentwise.

Consequently, there exist fixed $c_0, c_1 \in (0,1)$ and $c, C > 0$, independent of n and S, for which the following two volume bounds hold. The same metric and density estimates give, for $y\in\mathcal{N}_{S,n}$ and $0 < t \leq c_0 r_n$,

$$cR^{-d}r_n^{-k}t^p \leq Q_{S,n}(\mathbb{B}(y,t)) \leq CR^{-d}r_n^{-k}t^p. \tag{128}$$

They also give, for every $a\in\mathbb{R}^D$ and $0 < t \leq c_1 R$,

$$Q_{S,n}(\mathbb{B}(a,t)) \leq CR^{-d}t^d\left(1\wedge\left(\frac{t}{r_n}\right)^k\right). \tag{129}$$

Indeed, the bi-Lipschitz bound in equation 125 traps the preimage of $\mathbb{B}(F(x,z),t)$, for $t\leq c_0 r_n$, between product caps of radii comparable to t/R and t/r_n. Normalized spherical-cap bounds give $(t/R)^d(t/r_n)^k$. For larger t, the fiber cap saturates at one; if $\mathbb{B}(a,t)$ meets the sheet, recentering at an intersection point and enlarging the radius to $2t$ gives equation 129.

Gaussian localization on one sheet. For later use, the density of $Q_{S,n} * \mathcal{N}(0,\sigma^2 I_D)$ is

$$q_{S,n,\sigma}(w) = (2\pi\sigma^2)^{-D/2}\int_{\mathcal{N}_{S,n}} \exp\left(-\frac{\|w-z\|^2}{2\sigma^2}\right) h_{S,n}(z)\,d\operatorname{vol}_p(z). \tag{130}$$

Fix $y\in\mathcal{N}_{S,n}$ and put

$$I_{S,n,\sigma}(y) := \int \exp\left(-\frac{\|z-y\|^2}{2\sigma^2}\right) dQ_{S,n}(z).$$

For $\sigma \leq c_0 r_n$, equation 128 yields

$$I_{S,n,\sigma}(y) \geq \exp(-1/2)\,Q_{S,n}(\mathbb{B}(y,\sigma)) \geq cR^{-d}r_n^{-k}\sigma^p. \tag{131}$$

The volume bound equation 129 and total mass one imply $Q_{S,n}(\mathbb{B}(y,t)) \leq CR^{-d}r_n^{-k}t^p$ for all $t>0$: for $t > c_1 R$, the right-hand side is bounded below by a fixed positive constant because $r_n \leq R/4$. On the annulus $j\sigma \leq \|z-y\| < (j+1)\sigma$, division by equation 131 therefore bounds the weighted integral by $C_\ell(1+j)^{p+\ell}e^{-j^2/2}$. Summing from $j=0$ gives the moment bound; restricting to $j\geq\lfloor c_0 r_n/\sigma\rfloor$ gives the tail bound, using

$$\sum_{j\geq J}(1+j)^a e^{-cj^2} \leq C_{a,c}(1+J)^a e^{-c'J^2}, \qquad J\geq 0,$$

absorbs these polynomials into the Gaussian tail. Hence, for every fixed integer $\ell \geq 0$,

$$\frac{1}{I_{S,n,\sigma}(y)} \int \left(1 + \frac{\|z-y\|}{\sigma}\right)^{\ell} \exp\left(-\frac{\|z-y\|^2}{2\sigma^2}\right) dQ_{S,n}(z) \leq C_\ell, \tag{132}$$

$$\frac{1}{I_{S,n,\sigma}(y)} \int_{\{\|z-y\|>c_0 r_n\}} \left(1 + \frac{\|z-y\|}{\sigma}\right)^{\ell} \exp\left(-\frac{\|z-y\|^2}{2\sigma^2}\right) dQ_{S,n}(z) \leq C_\ell \exp\left(-c_\ell \frac{r_n^2}{\sigma^2}\right). \tag{133}$$

These bounds are uniform in n, S, and y.

Let $T = T_y\mathcal{N}_{S,n}$ and $N = T^\perp$. After decreasing the fixed c_0 if necessary, the reach bound provides a graph representation

$$z = y + v + H_y(v), \qquad v \in T, \qquad H_y(v) \in N,$$

on a tangent ball of radius $c_0 r_n$, with $H_y(0) = 0$ and $DH_y(0) = 0$. Indeed, reach at least cr_n bounds curvature by C/r_n; on a sufficiently small graph this gives $\|D^2 H_y\| \leq C/r_n$. Integrating once and twice along $s \mapsto sv$ yields

$$\|H_y(v)\| \leq C\frac{\|v\|^2}{r_n}, \qquad \|DH_y(v)\|_{\mathrm{op}} \leq C\frac{\|v\|}{r_n}. \tag{134}$$

In these coordinates, the density is

$$\ell_y(v) = h_{S,n}(y + v + H_y(v))\sqrt{\det(I_T + DH_y(v)^\top DH_y(v))}.$$

The graph path from 0 to v has length at most $C\|v\|$, so equation 121 bounds the log-density change by $C\|v\|/R$. Also, $\log\det(I+B) \leq \operatorname{tr} B$ for $B \succeq 0$ bounds the log-Jacobian by $C\|v\|^2/r_n^2$. Their sum is at most $C\|v\|/r_n$ when $\|v\| \leq c_0 r_n$, giving

$$\left|\log \frac{\ell_y(v)}{\ell_y(0)}\right| \leq C\frac{\|v\|}{r_n}, \qquad \|v\| \leq c_0 r_n. \tag{135}$$

Set $a = \sigma/r_n$, $v = \sigma t$, and define

$$\Theta_{y,\sigma}(t) := \frac{\ell_y(\sigma t)}{\ell_y(0)} \exp\left(-\frac{\|H_y(\sigma t)\|^2}{2\sigma^2}\right)$$

on the rescaled graph domain. Density variation contributes $O(a\|t\|)$, the graph Jacobian contained in ℓ_y contributes $O(a^2\|t\|^2)$, and

$$\frac{\|H_y(\sigma t)\|^2}{\sigma^2} \leq Ca^2\|t\|^4.$$

Indeed, on the graph domain,

$$\begin{aligned} \left|\frac{\ell_y(\sigma t)}{\ell_y(0)} - 1\right| &\leq Ca\|t\| \exp(Ca\|t\|), \\ \left|\exp\left(-\frac{\|H_y(\sigma t)\|^2}{2\sigma^2}\right) - 1\right| &\leq Ca^2\|t\|^4. \end{aligned} \tag{136}$$

Together with equation 133, these estimates imply, for every fixed ℓ,

$$\int (1+\|t\|)^\ell |\Theta_{y,\sigma}(t) - 1| \exp\left(-\frac{\|t\|^2}{2}\right) dt \leq C_\ell a. \tag{137}$$

Here $\Theta_{y,\sigma}$ may be set to zero outside the rescaled graph domain; the resulting Gaussian tail is $O\big(\exp\big(-c/a^2\big)\big)$ and is absorbed by the right-hand side.

Let

$$\Pi_{S,n,\sigma,y}(dz) := \frac{\exp\big(-\|z-y\|^2/(2\sigma^2)\big)\, Q_{S,n}(dz)}{I_{S,n,\sigma}(y)} \tag{138}$$

be the Gaussian-localized probability measure associated with $Q_{S,n}$ at y. Let $Z \sim \Pi_{S,n,\sigma,y}$ and put

$$V := \sigma^{-1} P_T(Z - y), \qquad W := \sigma^{-1} P_N(Z - y).$$

Let γ_T be standard Gaussian measure on T and define $\mathcal{Z}_{y,\sigma} := \int \Theta_{y,\sigma}\, d\gamma_T$. Taking $\ell = 4$ in equation 137 gives

$$|\mathcal{Z}_{y,\sigma} - 1| \le Ca, \qquad \int (1 + \|t\|^4) \left| \mathcal{Z}_{y,\sigma}^{-1} \Theta_{y,\sigma}(t) - 1 \right| d\gamma_T(t) \le Ca. \tag{139}$$

On the graph-chart contribution, the normalized pushforward probability measure of V has density $\mathcal{Z}_{y,\sigma}^{-1}\Theta_{y,\sigma}$ with respect to γ_T. The complement of the chart has Gaussian-localized probability, together with its fourth weighted moment, bounded by $C\exp\bigl(-c/a^2\bigr)$ by equation 133. Hence, the moments under the full Gaussian-localized probability measure differ from those under the normalized graph-chart probability measure by at most $C\exp\bigl(-c/a^2\bigr)$. Consequently,

$$\|\mathbb{E}V\| \le Ca, \qquad \|\mathbb{E}(VV^\top) - P_T\|_{\mathrm{op}} \le Ca, \qquad \mathbb{E}\|V\|^4 \le C. \tag{140}$$

More precisely, equation 139 and the chart-tail bound give $d_{\mathrm{TV}}(\mathcal{L}(V), \mathcal{N}(0, P_T)) \le Ca$. Here $\mathcal{N}(0, P_T)$ is the ambient, degenerate Gaussian supported on T; the Gaussian approximation is asymptotic as $a = \sigma/r_n \to 0$, not an exact distributional identity at positive σ.

Tangent and normal covariance blocks. On the chart, $\|W\| = \sigma^{-1}\|H_y(\sigma V)\| \le Ca\|V\|^2$; the localized-tail bound handles its complement. Hence,

$$\mathbb{E}\|W\|^2 \le Ca^2. \tag{141}$$

Writing $\Sigma_{S,n,\sigma}(y) := \mathrm{Cov}_{\Pi_{S,n,\sigma,y}}(Z)$, we have $Z - y = \sigma(V + W)$ because $P_T + P_N = I_D$. The rule $\mathrm{Cov}(AZ, BZ) = A\,\mathrm{Cov}(Z)B^\top$ and $P_T(Z - y) = \sigma V$, $P_N(Z - y) = \sigma W$ give

$$P_T\Sigma_{S,n,\sigma}P_T = \sigma^2\,\mathrm{Cov}(V), \quad P_T\Sigma_{S,n,\sigma}P_N = \sigma^2\,\mathrm{Cov}(V, W), \quad P_N\Sigma_{S,n,\sigma}P_N = \sigma^2\,\mathrm{Cov}(W).$$

Therefore, equation 140– equation 141 and Cauchy–Schwarz give

$$\left\|\sigma^{-2}P_T\Sigma_{S,n,\sigma}(y)P_T - P_T\right\|_{\mathrm{op}} \le Ca, \tag{142}$$

$$\left\|\sigma^{-2}P_T\Sigma_{S,n,\sigma}(y)P_N\right\|_{\mathrm{op}} \le Ca, \tag{143}$$

$$\left\|\sigma^{-2}P_N\Sigma_{S,n,\sigma}(y)P_N\right\|_{\mathrm{op}} \le Ca^2. \tag{144}$$

Differentiating equation 130 gives

$$\frac{\sigma\nabla q_{S,n,\sigma}(y)}{q_{S,n,\sigma}(y)} = \mathbb{E}_\Pi\frac{Z - y}{\sigma}, \qquad \frac{\sigma^2\nabla^2 q_{S,n,\sigma}(y)}{q_{S,n,\sigma}(y)} = \mathbb{E}_\Pi\frac{(Z - y)(Z - y)^\top}{\sigma^2} - I_D.$$

Taking norms and applying equation 132 with $\ell = 1, 2$ gives

$$\frac{\sigma^j\|\nabla^j q_{S,n,\sigma}(y)\|_{\mathrm{op}}}{q_{S,n,\sigma}(y)} \le C, \qquad j = 1, 2. \tag{145}$$

Finally, the conditional covariance identity equation 1 yields

$$A_{Q_{S,n},\sigma}(y) = I_D - \sigma^{-2}\Sigma_{S,n,\sigma}(y).$$

Since $I_D = P_T + P_N$, the three block estimates in equation 144 imply

$$\sup_{y \in \mathcal{N}_{S,n}} \left\|A_{Q_{S,n},\sigma_n}(y) - P_{N_y\mathcal{N}_{S,n}}\right\|_{\mathrm{op}} \le C\frac{\sigma_n}{r_n}.$$

This proves equation 122. □

A spectral-to-LHSD bound. Set

$$\Lambda_\varpi := \sup_{\lambda\in\mathbb{R}} |\bar{\phi}'_\varpi(\lambda)| < \infty, \qquad K_\varpi := D(1+\varpi^4).$$

For any symmetric $D \times D$ matrices A and B, Weyl's inequality and the scalar mean-value theorem give

$$|\operatorname{tr}\bar{\phi}_\varpi(A) - \operatorname{tr}\bar{\phi}_\varpi(B)| \le D\Lambda_\varpi \|A-B\|_{\mathrm{op}}. \tag{146}$$

Indeed, after arranging both spectra increasingly, Weyl's inequality bounds each paired eigenvalue difference by $\|A-B\|_{\mathrm{op}}$, and the scalar Lipschitz bound applies termwise. Moreover, the calibrated filter satisfies $\bar{\phi}_\varpi(\lambda) = \varpi^4(1-\lambda^4)/(\varpi^4+\lambda^4) \in (-\varpi^4, 1]$. Summing over the D eigenvalues gives $-D\varpi^4 < \operatorname{tr}\bar{\phi}_\varpi(A) \le D$. Hence,

$$|\operatorname{tr}\bar{\phi}_\varpi(A) - r| \le K_\varpi, \qquad r \in [0, D]. \tag{147}$$

Lemma E.2 (Tubes around coordinate-sphere intersections)**.** *Fix $0 \le k < D-d$. For $U \in \mathcal{I}_{k+1}$ and $r > 0$, write $\mathbb{S}_U(r) := \{z \in E_U : \|z\| = r\}$ and let $\nu_{U,r}$ be its normalized k-dimensional surface probability measure. There exist $c_{D-d,k} \in (0,1)$, $C_{D-d,k} < \infty$, and $C_{d,D-d,k} < \infty$ such that, uniformly over distinct $S, T \in \mathcal{I}_{k+1}$, $r > 0$, and $0 < \delta \le c_{D-d,k} r$, one has*

$$\nu_{S,r}(\{z : \operatorname{dist}(z, \mathbb{S}_T(r)) \le \delta\}) \le C_{D-d,k}\frac{\delta}{r}, \tag{148}$$

Consequently,

$$Q_{S,n}(\{y \in \mathcal{N}_{S,n} : \operatorname{dist}(y, \mathcal{N}_{T,n}) \le \delta\}) \le C_{d,D-d,k}\frac{\delta}{r_n}. \tag{149}$$

For $k = 0$, both sets are empty when $\delta \le c_{D-d,0} r$.

Proof. Put $J = S \cap T$, and let $P_{E_J} z = \sum_{j\in J} z_j e_j$ be the projection onto their shared coordinate subspace. If $J = \varnothing$, the two coordinate spheres are separated by $\sqrt{2}r$. Otherwise, $\max_{w\in\mathbb{S}_T(r)}\langle z, w\rangle = r\|P_{E_J} z\|$. Expanding $\|z-w\|^2 = 2r^2 - 2\langle z, w\rangle$ gives the first equality below; the second uses $r^2 - \|P_{E_J} z\|^2 = \|P_{E_S \cap E_J^\perp} z\|^2$:

$$\begin{aligned} \operatorname{dist}(z, \mathbb{S}_T(r))^2 &= 2r\big(r - \|P_{E_J} z\|\big) \\ &= \frac{2r}{r + \|P_{E_J} z\|}\|P_{E_S\cap E_J^\perp} z\|^2. \end{aligned} \tag{150}$$

Thus, the event in equation 148 lies in a $C\delta$-tube around a proper coordinate subsphere. If Z is uniform on $\mathbb{S}^k$ and $q = |S \setminus T|$, then the squared norm of these q coordinates has the $\mathrm{Beta}(q/2, (k+1-q)/2)$ distribution. Its lower tail bounds the normalized measure by $C(\delta/r)^q \le C\delta/r$. If an embedded point lies within δ of $\mathcal{N}_{T,n}$, the lower bound in equation 125 places its fiber coordinate within $C\delta$ of $\mathbb{S}_T(r_n)$. Integrating the first estimate over the base proves equation 149. □

Lemma E.3 (Gaussian localization between separated sheets)**.** *Put $p = d + k$. There are constants $c, C > 0$, depending only on the fixed construction parameters, such that, whenever $S \ne T$, $0 < \sigma \le c r_n$, and $y \in \mathcal{N}_{S,n}$ satisfies $\operatorname{dist}(y, \mathcal{N}_{T,n}) \ge \ell\sigma$ with $1 \le \ell \le c r_n/\sigma$, then, for $j = 0, 1, 2$,*

$$\frac{\sigma^j \|\nabla^j q_{T,n,\sigma}(y)\|_{\mathrm{op}}}{q_{S,n,\sigma}(y)} \le C\left(1 + \left(\frac{r_n}{\sigma}\right)^{D+4}\right)\exp\big(-c\ell^2\big). \tag{151}$$

For $j = 0$ the norm denotes absolute value, and for $j = 1$ it denotes the Euclidean norm.

Proof. For $j \le 2$,

$$\sigma^j \|\nabla^j \varphi_\sigma(v)\|_{\mathrm{op}} \le C\sigma^{-D}(1 + \|v\|/\sigma)^2 \exp\left(-\frac{\|v\|^2}{2\sigma^2}\right). \tag{152}$$

First, equation 129 implies the global bound

$$Q_{T,n}(\mathbb{B}(y,t)) \le C R^{-d} r_n^{-k} t^p, \qquad t > 0. \tag{153}$$

For $t \le c_1 R$, this follows from $1 \wedge (t/r_n)^k \le (t/r_n)^k$. For $t > c_1 R$, use total mass one and $r_n \le R/4$: the quantity $R^{-d} r_n^{-k} t^p$ is bounded below by a fixed positive constant. Thus the constant in equation 153 is independent of n.

Decompose $\mathcal{N}_{T,n}$ into shells $j_0\sigma \le \|z-y\| < (j_0+1)\sigma$. The separation assumption and the Gaussian derivative envelope give

$$\sigma^j \|\nabla^j q_{T,n,\sigma}(y)\|_{\mathrm{op}} \le C\sigma^{-D} \sum_{j_0 \ge \lfloor \ell \rfloor} (1+j_0)^2 \exp\!\big(-j_0^2/2\big) \cdot Q_{T,n}(\mathbb{B}(y,(j_0+1)\sigma)). \tag{154}$$

Insert equation 153 and divide by $q_{S,n,\sigma}(y) \ge cR^{-d} r_n^{-k} \sigma^{p-D}$ from equation 131. The factors $R^{-d} r_n^{-k} \sigma^{p-D}$ cancel, yielding

$$\frac{\sigma^j \|\nabla^j q_{T,n,\sigma}(y)\|_{\mathrm{op}}}{q_{S,n,\sigma}(y)} \le C \sum_{j_0 \ge \lfloor \ell \rfloor} (1+j_0)^{p+2} \exp(-j_0^2/2) \le C \exp(-c\ell^2).$$

The last inequality absorbs the fixed-degree polynomial into the Gaussian tail and is uniform in n, r_n, and σ. This also implies equation 151. □

E.3 Fine-scale LHSD limits

We combine the matched-moment construction with the preceding localization bounds to prove the main-text nonuniversality result.

Theorem 4.1 (restated) (Different LHSD limits under matched noise moments)**.** *Fix $d \in [D-1]$. There exist fixed parameters $(\tau_0, L_{\mathcal{T},0}, L_{\mathcal{F},0})$ and a common latent variable X with a smooth positive density on a compact connected boundaryless d-dimensional manifold $\mathcal{M} \subset \mathbb{R}^D$ with the following property. For any positive sequences $(\xi_{0,n})$ and (σ_n) with $0 < \xi_{0,n} \le \tau_0/3$ and $\sigma_n/\xi_{0,n} \to 0$, there are models $Q_n^{(k)} := \mathcal{L}(X + \xi_n^{(k)})$, $k \in [D-d]_0$, whose Borel conditional normal-noise kernels satisfy, for every n, k, and $x \in \mathcal{M}$,*

$$\mathbb{E}(\xi_n^{(k)} \mid X = x) = 0, \qquad \|\xi_n^{(k)}\| \le \xi_{0,n} \text{ a.s.}, \qquad \mathrm{Cov}(\xi_n^{(k)} \mid X = x) = \frac{\xi_{0,n}^2}{D-d+2} P_{N_x\mathcal{M}}.$$

For every n and k,

$$Q_n^{(k)} \in \mathcal{P}_d(\xi_{0,n}, \tau_0, L_{\mathcal{T},0}, L_{\mathcal{F},0}), \qquad d_{\mathrm{LMD}}(Q_n^{(k)}) = d.$$

Their LHSD fields nevertheless satisfy

$$\|T_{\sigma_n,\varpi}(Q_n^{(k)}) - (d+k)\|_{L^2(Q_n^{(k)})} \longrightarrow 0, \qquad k \in [D-d]_0.$$

Proof of Theorem 4.1. The construction above proves equation 3–equation 4. It remains to prove the population LHSD limits.

Step 1: Analyze the surface-sheet cases $0 \le k < D-d$. Injectivity of F gives

$$\mathcal{N}_{S,n} \cap \mathcal{N}_{T,n} = F\big(\mathcal{M} \times \{z \in E_S \cap E_T : \|z\| = r_n\}\big). \tag{155}$$

For $S \ne T$, the right-hand side has dimension at most $d+k-1$. Write $\mathbb{S}_S(r) = \{w \in E_S : \|w\| = r\}$. For the coordinate spheres used here, orthogonal projection gives

$$\mathrm{dist}(z, \mathbb{S}_T(r_n)) = \mathrm{dist}(z, \mathbb{S}_{S\cap T}(r_n)), \qquad z \in \mathbb{S}_S(r_n), \tag{156}$$

whenever the intersection is nonempty; if it is empty, the two spheres are separated by at least r_n. The global bi-Lipschitz estimate equation 125 transfers these comparisons to the embedded sheets. For all sufficiently large n, let

$$\ell_n := \sqrt{C_{\mathrm{loc}} \log(r_n/\sigma_n)} \tag{157}$$

and put $r_n^{\mathrm{bad}} = \ell_n \sigma_n$. Since $\sigma_n/r_n \to 0$, we have $r_n^{\mathrm{bad}}/r_n \to 0$. We discard the finitely many earlier indices, which do not affect the asserted limits. Define the bad set exactly by

$$\mathcal{B}_n^{(k)} := \bigcup_{S \in \mathcal{I}_{k+1}} \left\{ y \in \mathcal{N}_{S,n} : \min_{\substack{T \in \mathcal{I}_{k+1} \\ T \ne S}} \mathrm{dist}(y, \mathcal{N}_{T,n}) \le r_n^{\mathrm{bad}} \right\}, \tag{158}$$

where the minimum over an empty set is $+\infty$. This is a Borel subset of $\mathrm{supp}(Q_n^{(k)})$. Applying Lemma E.2 componentwise, summing over the fixed finite index set, and using equation 119 give

$$Q_n^{(k)}(\mathcal{B}_n^{(k)}) \le C\frac{r_n^{\mathrm{bad}}}{r_n} = C\ell_n\frac{\sigma_n}{r_n} \longrightarrow 0. \tag{159}$$

For $k = 0$, distinct coordinate spheres are separated by a constant multiple of r_n, so $\mathcal{B}_n^{(0)}$ is empty for all sufficiently large n.

Localize to the unique sheet away from intersections. Set

$$\mathcal{G}_n^{(k)} := \mathrm{supp}(Q_n^{(k)}) \setminus \mathcal{B}_n^{(k)}.$$

Every $y \in \mathcal{G}_n^{(k)}$ belongs to exactly one component. Define

$$S_n(y) := \text{the unique } S \in \mathcal{I}_{k+1} \text{ such that } y \in \mathcal{N}_{S,n}. \tag{160}$$

Every $T \neq S_n(y)$ is at distance greater than r_n^{bad} from y. Put

$$R_{S,n,\sigma} := \sum_{\substack{T\in\mathcal{I}_{k+1}\\ T\neq S}} q_{T,n,\sigma}.$$

Derivatives of a Gaussian density through order two are bounded by a degree-two polynomial times a Gaussian density. Applying this bound on annuli centered at y, using equation 129 for every competing component and equation 131 for the component through y, gives, for $j = 0, 1, 2$,

$$\begin{aligned}\eta_{j,n} &:= \sup_{y\in\mathcal{G}_n^{(k)}} \frac{\sigma_n^j \|\nabla^j R_{S_n(y),n,\sigma_n}(y)\|_{\mathrm{op}}}{q_{S_n(y),n,\sigma_n}(y)}\\ &\le C\left(1+\left(\frac{r_n}{\sigma_n}\right)^{D+4}\right)\exp\!\left(-c\ell_n^2\right) = o(1),\end{aligned} \tag{161}$$

after choosing $cC_{\mathrm{loc}} > D + 4$: with $b_n = r_n/\sigma_n \to \infty$, the bound is $C(1 + b_n^{D+4})b_n^{-cC_{\mathrm{loc}}} = o(1)$. Indeed, Lemma E.3 applies to every $T \neq S_n(y)$ with $\ell = \ell_n$; summing its bound over the fixed finite index set gives the display. In equation 161, the norm means absolute value when $j = 0$, Euclidean norm when $j = 1$, and operator norm when $j = 2$.

For the fixed query y, write $S = S_n(y)$. The mixture density is proportional to $q_{S,n,\sigma_n} + R_{S,n,\sigma_n}$. Therefore, the normalized derivative bounds equation 145, the three bounds in equation 161, and the identity

$$\nabla^2 \log q = q^{-1}\nabla^2 q - q^{-2}\nabla q \nabla q^\top$$

imply

$$\sup_{y\in\mathcal{G}_n^{(k)}} \left\| A_{Q_n^{(k)},\sigma_n}(y) - A_{Q_{S_n(y),n},\sigma_n}(y)\right\|_{\mathrm{op}} \le C \max_{0\le j\le 2}\eta_{j,n} = o(1). \tag{162}$$

Consequently, insert $A_{Q_{S_n(y),n},\sigma_n}(y)$ between the mixture matrix and the normal projector. The triangle inequality, the preceding localization bound, and Lemma E.1 give

$$\begin{aligned}\varepsilon_{n,k} &:= \sup_{y\in\mathcal{G}_n^{(k)}} \left\| A_{Q_n^{(k)},\sigma_n}(y) - P_{N_y\mathcal{N}_{S_n(y),n}}\right\|_{\mathrm{op}}\\ &\le C\frac{\sigma_n}{r_n} + C\max_{0\le j\le 2}\eta_{j,n} \longrightarrow 0.\end{aligned} \tag{163}$$

This is the required projector limit for $k < D - d$, with $\mathcal{N}_{n,y}^{(k)} = \mathcal{N}_{S_n(y),n}$. The projector in this display has nullity $d + k$. Since $\mathrm{tr}\,\bar{\phi}_\varpi(P_{N_y\mathcal{N}_{S_n(y),n}}) = d + k$, equation 146 yields

$$\sup_{y\in\mathcal{G}_n^{(k)}} |T_{\sigma_n,\varpi}(Q_n^{(k)})(y) - (d+k)| \le D\Lambda_\varpi\varepsilon_{n,k} \longrightarrow 0. \tag{164}$$

Finally, equation 147 on the bad set and equation 159 give

$$\|T_{\sigma_n,\varpi}(Q_n^{(k)}) - (d+k)\|^2_{L^2(Q_n^{(k)})} \le D^2\Lambda_\varpi^2\varepsilon_{n,k}^2 + K_\varpi^2 Q_n^{(k)}(\mathcal{B}_n^{(k)}) \longrightarrow 0. \tag{165}$$

This proves equation 5 for $k < D - d$.

Step 2: Analyze the filled-fiber case $k = D - d$. The tubular coordinate map for equation 108 is

$$F(u, z) = \big((R + z_1)u, z_2, \ldots, z_{D-d}\big), \qquad u \in \mathbb{S}^d, \quad z \in \mathbb{B}_{D-d}(0, R/2). \tag{166}$$

Let

$$\mathcal{T} := F(\mathbb{S}^d \times \mathbb{B}_{D-d}(0, R/2)), \qquad \Omega_n := F(\mathbb{S}^d \times \mathbb{B}_{D-d}(0, \xi_{0,n})).$$

The map F is a diffeomorphism onto the open tubular chart $\mathcal{T}$. Relative to Riemannian volume on $\mathcal{M}$ and Lebesgue measure in z, its Jacobian is

$$J(z) = \left(1 + \frac{z_1}{R}\right)^d. \tag{167}$$

Thus, $\operatorname{supp}(Q_n^{(D-d)}) = \overline{\Omega}_n = F(\mathbb{S}^d \times \overline{\mathbb{B}_{D-d}(0, \xi_{0,n})})$, and $Q_n^{(D-d)}$ has the following Lebesgue density almost everywhere:

$$q_n(F(u, z)) = C_n \left(1 + \frac{z_1}{R}\right)^{-d} \mathbf{1}_{\Omega_n}(F(u, z)), \tag{168}$$

where C_n is a normalizing constant. On every fixed smaller tubular chart, its smooth factor satisfies

$$\|\nabla \log q_n\| \le \frac{C}{R}, \qquad \|\nabla^2 \log q_n\|_{\mathrm{op}} \le \frac{C}{R^2}. \tag{169}$$

For all sufficiently large n, put

$$a_n := \frac{\sigma_n}{\xi_{0,n}}, \qquad \ell_n := \sqrt{C_{\mathrm{loc}} \log(1/a_n)}, \tag{170}$$

where C_{loc} is a sufficiently large fixed constant. Discarding the finitely many earlier indices does not affect any limit below. For a sufficiently large fixed C_1, define the normal boundary layer and its complement by

$$\begin{aligned} \mathcal{B}_n &:= \left\{ F(u, z) \in \operatorname{supp}(Q_n^{(D-d)}) : \operatorname{dist}(z, \partial \mathbb{B}_{D-d}(0, \xi_{0,n})) \le C_1 \ell_n \sigma_n \right\}, \\ \mathcal{G}_n^{(D-d)} &:= \operatorname{supp}(Q_n^{(D-d)}) \setminus \mathcal{B}_n. \end{aligned} \tag{171}$$

Since the normal coordinate is uniform on $\mathbb{B}_{D-d}(0, \xi_{0,n})$, for all sufficiently large n,

$$\begin{aligned} Q_n^{(D-d)}(\mathcal{B}_n) &\le 1 - \left(1 - \frac{C_1 \ell_n \sigma_n}{\xi_{0,n}}\right)^{D-d} \\ &\le C(D - d)\ell_n a_n \longrightarrow 0. \end{aligned} \tag{172}$$

Compare with a smooth interior extension. On $\mathcal{T}$, define the smooth interior factor

$$h_n^\circ(F(u, z)) := C_n \left(1 + \frac{z_1}{R}\right)^{-d}, \qquad \|z\| < R/2.$$

Choose a cutoff $\chi \in C_c^\infty(\mathcal{T})$ with $0 \le \chi \le 1$, equal to one on $F(\mathbb{S}^d \times \overline{\mathbb{B}_{D-d}(0, R/3)})$, set $\widetilde{h}_n := \chi h_n^\circ$ on $\mathcal{T}$, and extend it by zero. Such a cutoff exists because the closed $R/3$-tube is compactly contained in the open $R/2$-tube. Choosing it at scale R gives $\|\nabla^j \chi\| \le C R^{-j}$ for $j = 1, 2$. It preserves the density near $\overline{\Omega}_n$ while allowing derivatives to pass through convolution without boundary terms. The product rule then gives $\widetilde{h}_n \in C_c^2(\mathbb{R}^D)$ and

$$0 \le \widetilde{h}_n \le C C_n, \qquad \|\nabla \widetilde{h}_n\| \le C C_n / R, \qquad \|\nabla^2 \widetilde{h}_n\|_{\mathrm{op}} \le C C_n / R^2. \tag{173}$$

Since $\xi_{0,n} \le R/4$, $\widetilde{h}_n = h_n^\circ$ on a fixed neighborhood of $\overline{\Omega}_n$.

Let

$$\begin{aligned} p_n &:= q_{Q_n^{(D-d)}, \sigma_n} = \varphi_{\sigma_n} * (h_n^\circ \mathbf{1}_{\Omega_n}), \\ \widetilde{p}_n &:= \varphi_{\sigma_n} * \widetilde{h}_n. \end{aligned} \tag{174}$$

The global bi-Lipschitz estimate for F, together with the choice of C_1, shows that every $y \in \mathcal{G}_n^{(D-d)}$ has Euclidean distance at least $c\ell_n\sigma_n$ from Ω_n^c. Moreover, for some fixed $c_{\mathrm{int}} > 0$, $\mathbb{B}_D(y, c_{\mathrm{int}}\sigma_n) \subset \Omega_n$ for all sufficiently large n. Since $h_n^\circ \asymp C_n$ on this ball, Gaussian integration gives

$$p_n(y) \wedge \widetilde{p}_n(y) \geq cC_n, \qquad y \in \mathcal{G}_n^{(D-d)}. \tag{175}$$

The difference $\widetilde{h}_n - h_n^\circ \mathbf{1}_{\Omega_n}$ vanishes on Ω_n and is bounded by CC_n. Therefore, Gaussian derivative tails through order two imply

$$\max_{0\leq j\leq 2} \sup_{y\in\mathcal{G}_n^{(D-d)}} \frac{\sigma_n^j \|\nabla^j(p_n - \widetilde{p}_n)(y)\|_{\mathrm{op}}}{C_n} \leq C\eta_n, \qquad \eta_n := (1+\ell_n^{D+4})\exp\bigl(-c\ell_n^2\bigr). \tag{176}$$

Here the norm is absolute value for $j = 0$, Euclidean norm for $j = 1$, and operator norm for $j = 2$. To verify the display, use that $\nabla^j \varphi_{\sigma_n}$ is a polynomial of degree at most j times a Gaussian density and integrate over $\{v : \|v\| \geq c\ell_n\sigma_n\}$.

Since $\widetilde{h}_n$ is compactly supported, derivatives can be moved from the Gaussian kernel to $\widetilde{h}_n$. Hence, on $\mathcal{G}_n^{(D-d)}$,

$$\widetilde{p}_n \geq cC_n, \qquad \|\nabla\widetilde{p}_n\| \leq CC_n/R, \qquad \|\nabla^2\widetilde{p}_n\|_{\mathrm{op}} \leq CC_n/R^2. \tag{177}$$

Set $e_n := p_n - \widetilde{p}_n$. On $\mathcal{G}_n^{(D-d)}$,

$$|e_n| \leq CC_n\eta_n, \qquad \|\nabla e_n\| \leq CC_n\eta_n/\sigma_n, \qquad \|\nabla^2 e_n\|_{\mathrm{op}} \leq CC_n\eta_n/\sigma_n^2. \tag{178}$$

For all sufficiently large n, $\eta_n \leq c/2$, so equation 175 bounds both denominators away from zero. In particular,

$$|p_n^{-1} - \widetilde{p}_n^{-1}| \leq C\eta_n/C_n, \qquad |p_n^{-2} - \widetilde{p}_n^{-2}| \leq C\eta_n/C_n^2. \tag{179}$$

Writing $p_n = \widetilde{p}_n + e_n$, expanding the inverse factors and outer products, and using equation 177 give

$$\sigma_n^2 \left\|p_n^{-1}\nabla^2 p_n - \widetilde{p}_n^{-1}\nabla^2\widetilde{p}_n\right\|_{\mathrm{op}} \leq C\eta_n, \tag{180}$$

$$\sigma_n^2 \left\|p_n^{-2}\nabla p_n\nabla p_n^\top - \widetilde{p}_n^{-2}\nabla\widetilde{p}_n\nabla\widetilde{p}_n^\top\right\|_{\mathrm{op}} \leq C\eta_n. \tag{181}$$

For example, each term after expansion contains one of e_n, ∇e_n, or $\nabla^2 e_n$ and the remaining normalized factors are bounded by 1, σ_n/R, or $(\sigma_n/R)^2$. Moreover,

$$\sigma_n^2\|\nabla^2 \log\widetilde{p}_n\|_{\mathrm{op}} \leq C(\sigma_n/R)^2. \tag{182}$$

Combining these bounds with

$$\nabla^2 \log p_n = p_n^{-1}\nabla^2 p_n - p_n^{-2}\nabla p_n\nabla p_n^\top \tag{183}$$

gives

$$\begin{aligned}\varepsilon_{n,D-d} &:= \sup_{y\in\mathcal{G}_n^{(D-d)}} \|A_{Q_n^{(D-d)},\sigma_n}(y)\|_{\mathrm{op}} \\ &\leq C\left(\frac{\sigma_n}{R}\right)^2 + C\eta_n \longrightarrow 0.\end{aligned} \tag{184}$$

This proves the required interior Hessian limit after writing the boundary layer $\mathcal{B}_n$ as $\mathcal{B}_n^{(D-d)}$. Indeed,

$$\eta_n = \left[1 + \{C_{\mathrm{loc}}\log(1/a_n)\}^{(D+4)/2}\right] a_n^{cC_{\mathrm{loc}}} \longrightarrow 0.$$

Any fixed $C_{\mathrm{loc}} > 0$ suffices for this limit; the larger choice above also meets the surface-sheet requirement $cC_{\mathrm{loc}} > D+4$ (using the smaller tail constant c if necessary). The normalizing constant C_n is not subject to a largeness condition. Also,

$$\frac{\sigma_n}{R} = \frac{\sigma_n}{\xi_{0,n}}\frac{\xi_{0,n}}{R} \leq \frac{1}{4}\frac{\sigma_n}{\xi_{0,n}} \longrightarrow 0.$$

Pass from the interior Hessian limit to LHSD. Finally, $\operatorname{tr}\bar{\phi}_\varpi(0_D) = D$. Applying equation 146 on the good set and equation 147 on the bad set yields

$$\begin{aligned}&\|T_{\sigma_n,\varpi}(Q_n^{(D-d)}) - D\|^2_{L^2(Q_n^{(D-d)})} \\ &\quad\leq D^2\Lambda_\varpi^2\varepsilon_{n,D-d}^2 + K_\varpi^2 Q_n^{(D-d)}(\mathcal{B}_n) \longrightarrow 0.\end{aligned} \tag{185}$$

This proves equation 5 for $k = D - d$ and completes the proof. □

F POPULATION LHSD APPROXIMATION AND SHARPNESS

This section proves Theorems 4.2 and 4.3. The key covariance structure is

$$\frac{\Sigma_{Q,\sigma}}{\sigma^2} = \begin{pmatrix} I_d + E & G \\ G^\top & F \end{pmatrix}, \qquad E, F = O(\varepsilon), \qquad G = O(\sqrt{\varepsilon}).$$

A direct operator-norm comparison would therefore give only an $O(\sqrt{\varepsilon})$ projector error. The tangent and normal clusters are separated by a unit spectral gap, so the mixed block affects their eigenvalues through a Schur-complement term of order $G^2 = O(\varepsilon)$. Retaining this block structure is what yields the sharper spectral rate.

Throughout this section, C denotes a finite constant depending only on the fixed parameters of Assumptions 2.1–2.2; its value may change from line to line. The final ε_0 is the minimum of the finitely many smallness thresholds imposed below.

F.1 GAUSSIAN-LOCALIZATION IDENTITIES

Let $Y \sim Q$, let $G \sim \mathcal{N}(0, I_D)$ be independent, and put $W_\sigma = Y + \sigma G$ and $q_\sigma := q_{Q,\sigma}$. For every y with $q_\sigma(y) > 0$, define the Gaussian-localized probability measure

$$\Pi^y_{Q,\sigma}(dw) := \frac{\varphi_\sigma(y - w)}{q_\sigma(y)}\, dQ(w),$$

and its mean and covariance by

$$\mu_{Q,\sigma}(y) := \int w\, \Pi^y_{Q,\sigma}(dw), \qquad \Sigma_{Q,\sigma}(y) := \int (w - \mu_{Q,\sigma}(y))(w - \mu_{Q,\sigma}(y))^\top \Pi^y_{Q,\sigma}(dw).$$

Differentiation under the Gaussian convolution gives

$$\nabla \log q_\sigma(y) = \frac{\mu_{Q,\sigma}(y) - y}{\sigma^2}, \tag{186}$$

$$\nabla^2 \log q_\sigma(y) = -\frac{I_D}{\sigma^2} + \frac{\Sigma_{Q,\sigma}(y)}{\sigma^4}. \tag{187}$$

Indeed, both identities follow by differentiating $q_\sigma(y) = \int \varphi_\sigma(y - w) dQ(w)$. Consequently,

$$A_{Q,\sigma}(y) = I_D - \sigma^{-2}\Sigma_y, \qquad \Sigma_y := \Sigma_{Q,\sigma}(y), \tag{188}$$

which also proves equation 1.

F.2 LOCAL GEOMETRY AND GAUSSIAN-WEIGHTED COVARIANCE

The regularity classes in Appendix B.2 give the following uniform local graph representation.

Lemma F.1 (Uniform local geometry under positive reach)**.** *Let $\mathcal{M} \in \mathcal{C}^{\beta+1}_{d, L_\mathcal{T}, \tau_0}(\mathbb{R}^D)$. There exist fixed constants $c_{\mathrm{gr}}, c_{\mathrm{sep}} \in (0,1)$ and $C_{\mathrm{geo}} < \infty$ such that the following holds. For every $x \in \mathcal{M}$, put $P = P_T(x)$ and $P^\perp = P_N(x)$. There are an open neighborhood $\mathcal{U}_x \subset \mathcal{M}$ and a $C^{\beta+1}$ map*

$$h_x : \mathbb{B}_{T_x\mathcal{M}}(0, c_{\mathrm{gr}}\tau_0) \longrightarrow N_x\mathcal{M}$$

with the following properties.

(i) **Graph representation.** *One has*

$$\mathcal{U}_x = \{z_x(v) := x + v + h_x(v) : \|v\| < c_{\mathrm{gr}}\tau_0\}, \qquad h_x(0) = 0, \qquad Dh_x(0) = 0, \tag{189}$$

$$\begin{aligned} \|h_x(v)\| &\le C_{\mathrm{geo}} \frac{\|v\|^2}{\tau_0}, \quad \|Dh_x(v)\|_{\mathrm{op}} \le C_{\mathrm{geo}} \frac{\|v\|}{\tau_0}, \\ |J_x(v) - 1| &\le C_{\mathrm{geo}} \frac{\|v\|^2}{\tau_0^2}, \qquad\qquad J_x(v) := \det\big(I_d + Dh_x(v)^\top Dh_x(v)\big)^{1/2}. \end{aligned} \tag{190}$$

(ii) **Tangent stability and sheet separation.** *For every admissible v,*

$$\|P_T(x) - P_T(z_x(v))\|_{\mathrm{op}} \le C_{\mathrm{geo}} \frac{\|v\|}{\tau_0}, \qquad \mathcal{M} \cap \mathbb{B}_D(x, c_{\mathrm{sep}}\tau_0) \subset \mathcal{U}_x. \tag{191}$$

(iii) **Normal-space rotation.** *For every admissible v and every $u \in N_{z_x(v)}\mathcal{M}$,*

$$\|Pu\| \le C_{\mathrm{geo}} \frac{\|v\|}{\tau_0}\|u\|, \qquad \|P^{\perp}u\| \le \|u\|. \tag{192}$$

Proof of Lemma F.1. **Step 1: Curvature and tangent rotation.** The reach inequality equation 19 gives the required quadratic control of normal secants. Let II_p be the second fundamental form of $\mathcal{M}$ at p. For $a \in T_p\mathcal{M}$, choose a C^2 curve γ in $\mathcal{M}$ with $\gamma(0) = p$ and $\gamma'(0) = a$. Then

$$P_N(p)(\gamma(t) - p) = \frac{t^2}{2}\,\mathrm{II}_p(a, a) + o(t^2), \qquad \|\gamma(t) - p\|^2 = t^2\|a\|^2 + o(t^2).$$

Dividing equation 19 by $t^2/2$ and letting $t \to 0$ yields $\|\,\mathrm{II}_p(a, a)\| \le \tau_0^{-1}\|a\|^2$. Polarization, applied after pairing with each unit normal vector, gives

$$\|\,\mathrm{II}_p(a, b)\| \le \tau_0^{-1}\|a\|\,\|b\|, \qquad a, b \in T_p\mathcal{M}. \tag{193}$$

We next obtain a chordwise tangent estimate without presupposing a common chart. Fix $q \in \mathcal{M}$, put $r = \|q - x\| \le \tau_0$, and let $a \in T_q\mathcal{M}$ be a unit vector. Since the compact manifold $\mathcal{M}$ is complete, let γ be the unit-speed geodesic with $\gamma(0) = q$ and $\gamma'(0) = a$. By equation 193,

$$\gamma(r) = q + ra + R_r, \qquad \|R_r\| \le \frac{r^2}{2\tau_0}.$$

Applying equation 19 at x to q and $\gamma(r)$ gives

$$\begin{aligned} r\|P^{\perp}a\| &\le \|P^{\perp}(\gamma(r) - x)\| + \|P^{\perp}(q - x)\| + \|R_r\| \\ &\le \frac{\|\gamma(r) - x\|^2}{2\tau_0} + \frac{r^2}{\tau_0} \le C\frac{r^2}{\tau_0}, \end{aligned}$$

since $\|\gamma(r) - x\| \le 2r$. Hence, by the principal-angle identity for two subspaces of the same dimension,

$$\|P_T(q) - P_T(x)\|_{\mathrm{op}} \le C\frac{\|q - x\|}{\tau_0}, \qquad \|q - x\| \le \tau_0. \tag{194}$$

Step 2: Uniform graph construction. Choose $r_* = c_*\tau_0$ with $c_* > 0$ so small that the right-hand side of equation 194 is at most $1/4$ throughout $\mathbb{B}_D(x, r_*)$. The value $1/4$ is only a convenient fixed margin; any constant strictly smaller than $1/2$ would work. The tangent projection

$$\Psi_x : \mathcal{M} \longrightarrow T_x\mathcal{M}, \qquad \Psi_x(q) := P(q - x),$$

has injective derivative on $\mathcal{M} \cap \mathbb{B}_D(x, r_*)$. It is also one-to-one on that set. Indeed, if q, z belong to this set and $\Psi_x(q) = \Psi_x(z)$, then $w = z - q \in N_x\mathcal{M}$. Thus,

$$\|P_T(q)w\| \le \|P_T(q) - P\|_{\mathrm{op}}\|w\| \le \frac{1}{4}\|w\|, \qquad \|P_N(q)w\| \ge \frac{3}{4}\|w\|.$$

The second constant is simply $3/4 = 1 - 1/4$, since $w = P_T(q)w + P_N(q)w$ and hence $\|P_N(q)w\| \ge \|w\| - \|P_T(q)w\|$. On the other hand, equation 19 at q gives $\|P_N(q)w\| \le \|w\|^2/(2\tau_0)$. If $w \ne 0$, these two inequalities imply $\|w\| \ge 3\tau_0/2$, contrary to $\|w\| \le 2r_* < 3\tau_0/2$. Therefore $w = 0$, so $q = z$; this explicitly proves that Ψ_x is injective on $\mathcal{M} \cap \mathbb{B}_D(x, r_*)$.

Let $\mathcal{U}_x^{\circ}$ denote the connected piece of $\mathcal{M} \cap \mathbb{B}_D(x, r_*)$ that contains x. The preceding derivative and injectivity statements imply that $\Psi_x|_{\mathcal{U}_x^{\circ}}$ is an injective local diffeomorphism. Moreover, $\mathcal{U}_x^{\circ}$ is relatively open because manifolds are locally path connected, and it is relatively closed because every connected component is closed. Fix $\|v\| < r_*/2$ and put

$$I_v := \{t \in [0, 1] : tv \in \Psi_x(\mathcal{U}_x^{\circ})\}.$$

This set is nonempty and relatively open. If $t_j \in I_v$, $t_j \to t$, and $q_j = \Psi_x^{-1}(t_j v)$, compactness gives a subsequence converging to $q \in \overline{\mathcal{U}_x^\circ}$. If $\|q - x\| = r_*$, then equation 19 gives

$$\|tv\|^2 = \|P(q-x)\|^2 \geq r_*^2 - \frac{r_*^4}{4\tau_0^2} > \frac{r_*^2}{4},$$

contrary to $\|tv\| < r_*/2$. Thus $q \in \mathcal{U}_x^\circ$, so I_v is also closed and hence equals $[0,1]$. Therefore,

$$\mathbb{B}_{T_x\mathcal{M}}(0, r_*/2) \subset \Psi_x(\mathcal{U}_x^\circ).$$

Let $c_{\mathrm{gr}} = r_*/(4\tau_0)$ and define z_x as the inverse of Ψ_x on $\mathbb{B}_{T_x\mathcal{M}}(0, c_{\mathrm{gr}}\tau_0)$; then $z_x(v) = x + v + h_x(v)$ for a $C^{\beta+1}$ map h_x into $N_x\mathcal{M}$.

For $z = z_x(v)$, orthogonality and equation 19 give

$$\|h_x(v)\| \leq \frac{\|v\|^2 + \|h_x(v)\|^2}{2\tau_0}.$$

Since $z \in \mathbb{B}_D(x, r_*)$ and $r_* \leq \tau_0/4$, the quadratic term on the right can be absorbed, yielding $\|h_x(v)\| \leq C\|v\|^2/\tau_0$. After decreasing c_{gr} if necessary, this also gives $\|z_x(v) - x\| \leq 2\|v\|$. Combining this with equation 194 and the elementary relation between the slope of a graph and its largest principal angle yields

$$\|Dh_x(v)\|_{\mathrm{op}} \leq C\frac{\|v\|}{\tau_0}, \qquad \|P_T(x) - P_T(z_x(v))\|_{\mathrm{op}} \leq C\frac{\|v\|}{\tau_0}.$$

The determinant identity defining J_x then gives $|J_x(v) - 1| \leq C\|v\|^2/\tau_0^2$.

Step 3: Sheet separation. Choose $c_{\mathrm{sep}} < c_{\mathrm{gr}}/2$. If $q \in \mathcal{M} \cap \mathbb{B}_D(x, c_{\mathrm{sep}}\tau_0)$, then $v = \Psi_x(q)$ lies in the domain of z_x, and the graph bound places $z_x(v)$ in $\mathbb{B}_D(x, r_*)$. Since $\Psi_x(q) = \Psi_x(z_x(v))$, the injectivity proved above gives $q = z_x(v)$. This proves the inclusion in equation 191.

Step 4: Rotation of normal perturbations. Finally, if $u \in N_{z_x(v)}\mathcal{M}$, then $P_T(z_x(v))u = 0$, and hence

$$\|Pu\| \leq \|P_T(x) - P_T(z_x(v))\|_{\mathrm{op}}\|u\| \leq C\frac{\|v\|}{\tau_0}\|u\|.$$

The inequality $\|P^\perp u\| \leq \|u\|$ completes the proof. □

For this subsection, put

$$\varepsilon_{\xi_0,\sigma} := \underbrace{\left(\frac{\xi_0}{\sigma}\right)^2}_{\text{tubular noise}} + \underbrace{\left(\frac{\sigma}{\tau_0}\right)^2}_{\text{curvature}} + \underbrace{\sigma^2\Lambda_f}_{\text{density variation}}. \tag{195}$$

Thus the total error separates into tubular-noise, curvature, and density- variation contributions.

The reference $(1 + \sigma\langle g_x, \cdot\rangle)\gamma_d$ below is a signed first-order expansion of the rescaled tangent law, not itself a probability measure. Its total mass is one because the linear Gaussian moment vanishes.

Lemma F.2 (Normalized tangent–Gaussian comparison)**.** *Suppose $\varepsilon_{\xi_0,\sigma} \leq \varepsilon_0$. Fix $y \in \mathrm{supp}(Q)$, put $x = \pi_{\mathcal{M}}(y)$ and $P = P_{T_x\mathcal{M}}$, and identify $T_x\mathcal{M}$ with $\mathbb{R}^d$ through an arbitrary orthonormal frame. For every nonnegative Borel function H on the normal bundle, define*

$$\overline{\Pi}_{y,\sigma}(H) := \frac{1}{f(x)(2\pi\sigma^2)^{d/2}} \int_{\mathcal{M}} f(z) \int_{N_z\mathcal{M}} H(z,u) \exp\left(-\frac{\|y - (z+u)\|^2}{2\sigma^2}\right) K_z(du)\, d\,\mathrm{vol}_d(z). \tag{196}$$

Let $\mathcal{Z}_{y,\sigma} := \overline{\Pi}_{y,\sigma}(1)$, $\Pi_{y,\sigma} := \overline{\Pi}_{y,\sigma}/\mathcal{Z}_{y,\sigma}$, and set

$$\mathcal{T}_x(z,u) := \frac{P(z-x)}{\sigma}, \qquad g_x := \nabla_{\mathcal{M}} \log f(x).$$

The functional $\overline{\Pi}_{y,\sigma}$ is an unnormalized finite measure on the normal bundle. Since its integrand is nonnegative and $0 < \mathcal{Z}_{y,\sigma} < \infty$, the normalized object $\Pi_{y,\sigma}$ is a probability measure. Then the following conclusions hold.

(i) **Localization.** *Under $\Pi_{y,\sigma}$, the pushforward by $(z,u)\mapsto z+u$ equals the Gaussian-localized probability measure $\Pi^{y}_{Q,\sigma}$.*

(ii) **Tangent Gaussian comparison.** *Let γ_d be standard Gaussian measure on $\mathbb{R}^d$ and define*

$$\|\mu\|_{\mathrm{TV},r} := \int_{\mathbb{R}^d} (1+\|t\|^r)\, d|\mu|(t).$$

For every integer $0\le r\le 8$, the finite signed reference measure $(1+\sigma\langle g_x,\,\cdot\,\rangle)\gamma_d$ obeys

$$\big\|(\mathcal{T}_x)_{\#}\overline{\Pi}_{y,\sigma} - \big(1+\sigma\langle g_x,\,\cdot\,\rangle\big)\gamma_d\big\|_{\mathrm{TV},r} \le C_r\varepsilon_{\xi_0,\sigma}. \tag{197}$$

In particular, after reducing ε_0 if necessary,

$$|\mathcal{Z}_{y,\sigma}-1|\le C\varepsilon_{\xi_0,\sigma}, \qquad \frac{1}{2}\le \mathcal{Z}_{y,\sigma}\le \frac{3}{2}. \tag{198}$$

(iii) **Normalized moments.** *If $\ell\ge 1$, $\psi:\mathbb{R}^d\to\mathbb{R}^\ell$ is Borel, and $\|\psi(t)\|\le 1+\|t\|^r$, then*

$$\left\|\mathbb{E}_{\Pi_{y,\sigma}}\psi(\mathcal{T}_x) - \int_{\mathbb{R}^d}\psi(t)\big(1+\sigma\langle g_x,t\rangle\big)d\gamma_d(t)\right\| \le C_r\varepsilon_{\xi_0,\sigma}. \tag{199}$$

(iv) **Off-chart tail.** *If $\mathcal{U}_x$ is the local graph neighborhood in equation 189, then*

$$\overline{\Pi}_{y,\sigma}\big((1+\|\mathcal{T}_x\|^r)\mathbf{1}_{\{z\notin\mathcal{U}_x\}}\big) \le C_r\exp\left(-c\frac{\tau_0^2}{\sigma^2}\right). \tag{200}$$

All constants are uniform in Q, y, and the Borel kernel $x\mapsto K_x$.

Proof. **Step 1: Local coordinates.** The closed normal disk bundle $\{(z,u): z\in\mathcal{M},\ u\in N_z\mathcal{M},\ \|u\|\le\xi_0\}$ is compact, and the support of Q is contained in the image of this bundle under $(z,u)\mapsto z+u$. Hence, $y=z+u$ for some such (z,u). Since $\varepsilon_{\xi_0,\sigma}\le\varepsilon_0$ implies $2\xi_0<\tau_0$, the nearest-point projection is unique and equals $x:=z=\pi_{\mathcal{M}}(y)$. Thus, writing $y=x+\eta$, one has $\eta\in N_x\mathcal{M}$ and $\|\eta\|\le\xi_0$. Choose the graph neighborhood $\mathcal{U}_x$ from Lemma F.1. Thus, with $r_\star=c_{\mathrm{gr}}$ and $c_\star=c_{\mathrm{sep}}$, it contains $z_x(\mathbb{B}_{T_x\mathcal{M}}(0,r_\star\tau_0))$, while its complement in $\mathcal{M}$ is at Euclidean distance at least $c_\star\tau_0$ from x.

On the graph, put $z=z_x(\sigma t)$ and define

$$\Delta(t,u) := \|y-(z_x(\sigma t)+u)\|^2-\sigma^2\|t\|^2.$$

The change of variables $v=\sigma t$ in equation 196 shows that the local pushforward under $\mathcal{T}_x$ has density $R_{y,\sigma}$ with respect to γ_d, where

$$R_{y,\sigma}(t) := \mathbf{1}_{\{\sigma t\in V_x\}}\frac{f(z_x(\sigma t))J_x(\sigma t)}{f(x)}\cdot\int\exp\left(-\frac{\Delta(t,u)}{2\sigma^2}\right)K_{z_x(\sigma t)}(du). \tag{201}$$

Here V_x is the tangent domain of the graph. Writing $h=h_x(\sigma t)$, orthogonal decomposition at x gives

$$P(y-(z_x(\sigma t)+u)) = -\sigma t - Pu, \qquad P^\perp(y-(z_x(\sigma t)+u)) = \eta - h - P^\perp u.$$

Consequently,

$$\Delta(t,u) = 2\sigma\langle t,Pu\rangle + \|Pu\|^2 + \|\eta-h-P^\perp u\|^2.$$

Moreover, equation 190–equation 192 give $\|Pu\|\le C\xi_0\sigma\|t\|/\tau_0$, $\|h\|\le C\sigma^2\|t\|^2/\tau_0$, and $\|\eta\|\vee\|P^\perp u\|\le\xi_0$. Applying these bounds to the exact display above gives, pointwise for every admissible u and with $a_-:=(-a)\vee 0$,

$$\frac{\Delta(t,u)_-}{\sigma^2}\le C\frac{\xi_0}{\tau_0}\|t\|^2, \tag{202}$$

$$\frac{|\Delta(t,u)|}{\sigma^2}\le C\left(\left(\frac{\xi_0}{\sigma}\right)^2+\frac{\xi_0}{\tau_0}\|t\|^2+\left(\frac{\sigma}{\tau_0}\right)^2\|t\|^4\right). \tag{203}$$

Step 2: Comparison with the tangent Gaussian measure. Since $\xi_0/\tau_0 \le \varepsilon_{\xi_0,\sigma}/2$, choosing ε_0 small makes the perturbed integrand, after multiplication by the standard Gaussian density, bounded by $C\exp\bigl(-3\|t\|^2/8\bigr)$. Therefore, $|\exp(-a)-1| \le |a|\,(1+\exp(-a))$ and equation 203 imply, for every fixed $r \le 10$,

$$\int_{\{\sigma t \in V_x\}} (1+\|t\|^r)\left|\int \exp\left(-\frac{\Delta(t,u)}{2\sigma^2}\right) K_{z_x(\sigma t)}(du) - 1\right| d\gamma_d(t) \le C_r \varepsilon_{\xi_0,\sigma}. \tag{204}$$

This bound is pointwise in u before integration and depends only on the support constraint for K_z. Writing $v = \sigma t$, intrinsic Taylor expansion along $\mathcal{M}$ and the graph coordinate bounds give

$$\frac{f(z_x(v))}{f(x)} = 1 + \langle g_x, v\rangle + O\bigl((\Lambda_f + \tau_0^{-2})\|v\|^2\bigr), \qquad J_x(v) = 1 + O\left(\frac{\|v\|^2}{\tau_0^2}\right).$$

Therefore,

$$\left|\frac{f(z_x(\sigma t))J_x(\sigma t)}{f(x)} - 1 - \sigma\langle g_x, t\rangle\right| \le C\sigma^2(\Lambda_f + \tau_0^{-2})\|t\|^2 \le C\varepsilon_{\xi_0,\sigma}\|t\|^2 \tag{205}$$

on the graph domain. The density–volume ratio is uniformly bounded there. Indeed, the coordinate correction in the linear density term is bounded by $C\|g_x\|\|v\|^2/\tau_0$, which is absorbed using $\|g_x\|/\tau_0 \le (\|g_x\|^2 + \tau_0^{-2})/2$. Combining this display with equation 204 proves the weighted-variation estimate on the local graph. To see the combination explicitly, set

$$a(t) := \frac{f(z_x(\sigma t))J_x(\sigma t)}{f(x)}, \qquad k(t) := \int \exp\left(-\frac{\Delta(t,u)}{2\sigma^2}\right) K_{z_x(\sigma t)}(du).$$

Then the exact local density in equation 201 is $a(t)k(t)$, and

$$\bigl|a(t)k(t) - (1+\sigma\langle g_x,t\rangle)\bigr| \le \bigl|a(t) - (1+\sigma\langle g_x,t\rangle)\bigr|k(t) + |1+\sigma\langle g_x,t\rangle|\,|k(t)-1|.$$

After multiplication by $(1+\|t\|^r)$ and integration against γ_d, the first term is $O(\varepsilon_{\xi_0,\sigma})$ by equation 205, the preceding Gaussian domination of $k(t)$, and Gaussian moment integrability. The second term is $O(\varepsilon_{\xi_0,\sigma})$ by equation 204, using its $r+1$ version to absorb the linear factor. Consequently,

$$\int (1+\|t\|^r)\bigl|a(t)k(t) - (1+\sigma\langle g_x,t\rangle)\bigr|\, d\gamma_d(t) \le C_r\varepsilon_{\xi_0,\sigma}$$

on the local graph. The Gaussian tail outside the graph domain is absorbed into the same bound, which gives the claimed weighted-variation conclusion.

Step 3: Off-chart localization. It remains to control the contribution from $\mathcal{M}\setminus\mathcal{U}_x$. If $z \notin \mathcal{U}_x$, then, for sufficiently small ε_0 and every admissible u,

$$\|y-(z+u)\| \ge \|z-x\| - \|y-x\| - \|u\| \ge c_\star\tau_0 - 2\xi_0 \ge \frac{c_\star\tau_0}{2}.$$

The fixed support and volume envelopes bound the left-hand side of equation 200 by

$$C(1+\sigma^{-d-r})\exp\left(-c\frac{\tau_0^2}{\sigma^2}\right) \le C_r\exp\left(-c'\frac{\tau_0^2}{\sigma^2}\right). \tag{206}$$

The last inequality uses the fixed positive value of τ_0 and $\sigma/\tau_0 \le \sqrt{\varepsilon_0}$. The Gaussian mass outside the tangent coordinate domain obeys the same bound. Since $\exp\bigl(-c\tau_0^2/\sigma^2\bigr) \le C_r(\sigma/\tau_0)^2 \le C_r\varepsilon_{\xi_0,\sigma}$, this proves equation 197.

Step 4: Normalization. Taking total mass in equation 197 and using $\int\langle g_x,t\rangle d\gamma_d(t) = 0$ proves the first part of equation 198. More explicitly, choose the smallness threshold so that $C\varepsilon_0 \le 1/2$. Then $|\mathcal{Z}_{y,\sigma} - 1| \le C\varepsilon_{\xi_0,\sigma} \le 1/2$, which gives $1/2 \le \mathcal{Z}_{y,\sigma} \le 3/2$. Finally, divide by $\mathcal{Z}_{y,\sigma}$ in equation 197. The reference weighted moments are uniformly bounded since $\sigma\|g_x\| \le \sqrt{\varepsilon_{\xi_0,\sigma}}$. This gives equation 199. □

Lemma F.3 (Gaussian-localized covariance blocks)**.** *Suppose that $\varepsilon_{\xi_0,\sigma} \le \varepsilon_0$ for a sufficiently small constant ε_0. For every $y \in \mathrm{supp}(Q)$, let $x = \pi_{\mathcal{M}}(y)$, which is well-defined since $2\xi_0/\tau_0 \le \varepsilon_{\xi_0,\sigma} < 1$. Relative to $T_x\mathcal{M} \oplus N_x\mathcal{M}$, the Gaussian-localized covariance satisfies*

$$\frac{\Sigma_y}{\sigma^2} = \begin{pmatrix} I_d + E_y & G_y \\ G_y^\top & F_y \end{pmatrix}, \tag{207}$$

with the uniform bounds

$$\|E_y\|_{\mathrm{op}} \le C\varepsilon_{\xi_0,\sigma}, \qquad \|F_y\|_{\mathrm{op}} \le C\varepsilon_{\xi_0,\sigma}, \qquad \|G_y\|_{\mathrm{op}} \le C\sqrt{\varepsilon_{\xi_0,\sigma}}. \tag{208}$$

Only the support condition on K_x enters these bounds.

Proof. Use the notation of Lemma F.2. Under $\Pi_{y,\sigma}$ write $W = z + u$ and $\mathcal{T} = P(z - x)/\sigma$. Applying equation 199 coordinatewise with $\psi(t) = t$ and $\psi(t) = tt^\top$ gives

$$\mathbb{E}_{\Pi_{y,\sigma}}\mathcal{T} = \sigma\nabla_{\mathcal{M}} \log f(x) + O(\varepsilon_{\xi_0,\sigma}), \tag{209}$$

$$\mathbb{E}_{\Pi_{y,\sigma}}(\mathcal{T}\mathcal{T}^\top) = I_d + O(\varepsilon_{\xi_0,\sigma}). \tag{210}$$

The linear correction in the second display vanishes since every third moment of a centered Gaussian is zero. Since $\sigma^2\|\nabla_{\mathcal{M}} \log f(x)\|^2 \le \varepsilon_{\xi_0,\sigma}$, subtracting the outer product of the first display yields

$$\mathrm{Cov}_{\Pi_{y,\sigma}}(\mathcal{T}) = I_d + O(\varepsilon_{\xi_0,\sigma}). \tag{211}$$

The same comparison with $r = 4$ gives $\mathbb{E}_{\Pi_{y,\sigma}}\|\mathcal{T}\|^q \le C_q$ for $1 \le q \le 4$.

On the local graph event, write $z = x + \sigma\mathcal{T} + h_x(\sigma\mathcal{T})$ and define, before projecting,

$$r_T := \frac{Pu}{\sigma}, \qquad r_N := \frac{h_x(\sigma\mathcal{T}) + P^\perp u}{\sigma}.$$

Then the full rescaled displacement decomposes as

$$\frac{W - x}{\sigma} = \mathcal{T} + r_T + r_N,$$

where $\mathcal{T} + r_T \in T_x\mathcal{M}$ and $r_N \in N_x\mathcal{M}$. Equations equation 189 and equation 192 therefore give

$$\frac{P(W - x)}{\sigma} = \mathcal{T} + r_T, \qquad \|r_T\| \le C\frac{\xi_0}{\tau_0}\|\mathcal{T}\|, \tag{212}$$

$$\frac{P^\perp(W - x)}{\sigma} = r_N, \qquad \|r_N\| \le C\left(\underbrace{\frac{\xi_0}{\sigma}}_{\text{tubular noise}} + \underbrace{\frac{\sigma}{\tau_0}\|\mathcal{T}\|^2}_{\text{curvature}}\right). \tag{213}$$

Let $\mathcal{O}_x := \{z \notin \mathcal{U}_x\}$ denote the off-chart event. The tail bound equation 200, the normalizer lower bound equation 198, and the fixed support envelope control its contribution. For the tangent remainder, retain the noise factor: globally $r_T = Pu/\sigma$, so

$$\mathbb{E}_{\Pi_{y,\sigma}}\left[\|r_T\|^2 \mathbf{1}_{\mathcal{O}_x}\right] \le C\left(\frac{\xi_0}{\sigma}\right)^2 \exp\left(-c\frac{\tau_0^2}{\sigma^2}\right) \le C\left(\frac{\xi_0}{\tau_0}\right)^2.$$

For the normal remainder, the fixed support envelope gives $\|r_N\| \le C/\sigma$ on this event. Its squared-moment contribution is at most $C\sigma^{-2}\exp(-c\tau_0^2/\sigma^2) \le C(\sigma/\tau_0)^2$, with constants depending on the fixed envelopes. Combining these bounds with the local estimates proves the two remainder bounds below, including the case $\xi_0 = 0$. Consequently,

$$\mathbb{E}_{\Pi_{y,\sigma}}\|r_T\|^2 \le C\left(\frac{\xi_0}{\tau_0}\right)^2, \qquad \mathbb{E}_{\Pi_{y,\sigma}}\|r_N\|^2 \le C\left(\left(\frac{\xi_0}{\sigma}\right)^2 + \left(\frac{\sigma}{\tau_0}\right)^2\right) \le C\varepsilon_{\xi_0,\sigma}. \tag{214}$$

Since $\xi_0/\tau_0 \le \varepsilon_{\xi_0,\sigma}/2$, equations equation 211–equation 214 and covariance Cauchy–Schwarz imply

$$\left\|\mathrm{Cov}_{\Pi^y_{Q,\sigma}}\left(\frac{PW}{\sigma}\right) - I_d\right\|_{\mathrm{op}} \le C\varepsilon_{\xi_0,\sigma},$$

$$\left\|\mathrm{Cov}_{\Pi^y_{Q,\sigma}}\left(\frac{P^\perp W}{\sigma}\right)\right\|_{\mathrm{op}} \le C\varepsilon_{\xi_0,\sigma},$$

$$\left\|\mathrm{Cov}_{\Pi^y_{Q,\sigma}}\left(\frac{PW}{\sigma}, \frac{P^\perp W}{\sigma}\right)\right\|_{\mathrm{op}} \le C\sqrt{\varepsilon_{\xi_0,\sigma}}.$$

These are exactly equation 207–equation 208. All bounds in Lemma F.2 are uniform over measurable Borel kernels K_z, so the same is true here. □

F.3 Spectral separation and intermediate-scale recovery

The following abstract block matrix is the tangent–normal form needed for $A_{Q,\sigma}(y)$ below: E and F are the small within-block errors, while G is the tangent–normal coupling. Isolating this form makes the second-order effect of G explicit.

Lemma F.4 (Separated block spectra)**.** *Let*

$$B = \begin{pmatrix} E & G \\ G^\top & I_{D-d} - F \end{pmatrix} \tag{215}$$

be symmetric, where $E \in \mathbb{R}^{d\times d}$, $F \in \mathbb{R}^{(D-d)\times(D-d)}$, and $G \in \mathbb{R}^{d\times(D-d)}$. Suppose $\|E\|_{\mathrm{op}} \vee \|F\|_{\mathrm{op}} \le a$ and $\|G\|_{\mathrm{op}} \le b$. If $a + b \le 1/8$, then the eigenvalues $\lambda_1(B) \le \cdots \le \lambda_D(B)$ satisfy

$$\max_{1\le j\le d} |\lambda_j(B)| \ \vee \ \max_{d<j\le D} |\lambda_j(B) - 1| \le C_{\mathrm{blk}}(a + b^2),$$

where $C_{\mathrm{blk}} > 0$ is a numerical constant.

Proof. Weyl's inequality first places the spectrum in the union of an $(a + b)$-neighborhood of zero and an $(a + b)$-neighborhood of one. These intervals are disjoint, and continuity along the path that replaces (E, F, G) by (tE, tF, tG) shows that they contain d and $D - d$ eigenvalues, respectively.

Let λ belong to the cluster around zero and write a corresponding unit eigenvector as (u, v). The two block rows of $B(u, v)^\top = \lambda(u, v)^\top$ are

$$Eu + Gv = \lambda u, \qquad G^\top u + (I_{D-d} - F)v = \lambda v.$$

Since $I_{D-d} - F - \lambda I_{D-d}$ is invertible with inverse norm at most 2,

$$v = -(I_{D-d} - F - \lambda I_{D-d})^{-1} G^\top u, \qquad \|v\| \le 2b\|u\|. \tag{216}$$

Substitution in the first block equation gives

$$\lambda u = \left[E - G(I_{D-d} - F - \lambda I_{D-d})^{-1} G^\top\right] u. \tag{217}$$

If $R_\lambda := (I_{D-d} - F - \lambda I_{D-d})^{-1}$, then

$$\|GR_\lambda G^\top\|_{\mathrm{op}} \le \|G\|_{\mathrm{op}} \, \|R_\lambda\|_{\mathrm{op}} \, \|G^\top\|_{\mathrm{op}} \le b \cdot 2 \cdot b = 2b^2.$$

Thus the mixed block enters through a quadratic Schur-complement term, and $|\lambda| \le a + 2b^2$. The same argument applied to $I_D - B$ gives $|\lambda - 1| \le a + 2b^2$ for the upper cluster. □

The technical spectral statement used in the main theorem is the following. Suppose $\varepsilon(\xi_0, \sigma) \le \varepsilon_0$ for a sufficiently small constant $\varepsilon_0 > 0$. For each $Q \in \mathcal{P}_d(\xi_0, \tau_0, L_{\mathcal{T}}, L_{\mathcal{F}})$, fix any admissible representation $(\mathcal{M}, f, K)$. Uniformly over Q, the chosen representation, and Q-almost every y,

$$\left\| A_{Q,\sigma}(y) - P_{N_{\pi_{\mathcal{M}}(y)}\mathcal{M}} \right\|_{\mathrm{op}} = O\left(\sqrt{\varepsilon(\xi_0, \sigma)}\right), \tag{218}$$

and, with the eigenvalues ordered increasingly,

$$\max_{1\le j\le d} |\lambda_j(A_{Q,\sigma}(y))| = O(\varepsilon(\xi_0, \sigma)), \tag{219}$$

$$\max_{d<j\le D} |\lambda_j(A_{Q,\sigma}(y)) - 1| = O(\varepsilon(\xi_0, \sigma)). \tag{220}$$

Theorem 4.2 (restated) (Recovery at intermediate scales)**.** *There exists $\varepsilon_0 > 0$, depending only on the fixed model parameters, such that if $\varepsilon(\xi_0, \sigma) \le \varepsilon_0$, then $2\xi_0 < \tau_0$ and the LMD is well-defined. For every $Q \in \mathcal{P}_d(\xi_0, \tau_0, L_{\mathcal{T}}, L_{\mathcal{F}})$ and Q-almost every y, the eigenvalues of $A_{Q,\sigma}(y)$ can be indexed so that*

$$\max_{1\le j\le d} |\lambda_j(y)| \ \vee \ \max_{d<j\le D} |\lambda_j(y) - 1| \lesssim \varepsilon(\xi_0, \sigma).$$

Consequently,

$$\sup_{Q\in\mathcal{P}_d(\xi_0,\tau_0,L_{\mathcal{T}},L_{\mathcal{F}})} \|T_{\sigma,\varpi}(Q) - d_{\mathrm{LMD}}(Q)\|_{L^\infty(Q)} \lesssim_\varpi d\varepsilon(\xi_0, \sigma)^4 + (D - d)\varepsilon(\xi_0, \sigma).$$

Proof of Theorem 4.2. **Spectral cluster bounds.** Combining equation 188 and equation 207 gives

$$A_{Q,\sigma}(y) = \begin{pmatrix} -E_y & -G_y \\ -G_y^\top & I_{D-d} - F_y \end{pmatrix}. \tag{221}$$

This is exactly the abstract form equation 215 with $E = -E_y$, $G = -G_y$, and $F = F_y$; the signs do not affect the operator-norm bounds. The block bounds immediately imply

$$\|A_{Q,\sigma}(y) - P^\perp\|_{\mathrm{op}} \le C\sqrt{\varepsilon_{\xi_0,\sigma}}, \tag{222}$$

which is equation 218. In Lemma F.4, take $a = C\varepsilon_{\xi_0,\sigma}$ and $b = C\sqrt{\varepsilon_{\xi_0,\sigma}}$. Hence $a + b^2 = O(\varepsilon_{\xi_0,\sigma})$, so the displacement of both spectral clusters is at most $C\varepsilon_{\xi_0,\sigma}$. This proves equation 219–equation 220 and equation 6.

Exact-support specialization. When $\xi_0 = 0$, $\varepsilon_{\xi_0,\sigma} = (\sigma/\tau_0)^2 + \sigma^2\Lambda_f \to 0$ as $\sigma \downarrow 0$. For all sufficiently small σ, the preceding lemmas apply. Since $\mathrm{supp}(Q) = \mathcal{M}$ in this case, equation 207 and equation 222 give

$$\sup_{x\in\mathcal{M}} \left\|\sigma^{-2}\Sigma_{Q,\sigma}(x) - P_T(x)\right\|_{\mathrm{op}} \longrightarrow 0, \qquad \sup_{x\in\mathcal{M}} \|A_{Q,\sigma}(x) - P_N(x)\|_{\mathrm{op}} \longrightarrow 0. \tag{223}$$

Affine-calibrated spectral filter.

Let $b = C\varepsilon_{\xi_0,\sigma} < 1/2$ be the spectral bound just proved. Since

$$1 - \bar{\phi}_\varpi(\lambda) = \frac{(1+\varpi^4)\lambda^4}{\varpi^4 + \lambda^4} \le C_\varpi|\lambda|^4, \qquad |\bar{\phi}_\varpi(\lambda)| \le L_\varpi|\lambda - 1|, \tag{224}$$

where $L_\varpi = \sup_{|\lambda-1|\le 1/2} |\bar{\phi}'_\varpi(\lambda)|$; the second inequality follows from the mean-value theorem on $|\lambda - 1| \le b$. Therefore, the tangent and normal clusters give, pointwise on $\mathrm{supp}(Q)$,

$$|T_{\sigma,\varpi}(Q)(y) - d| \le C_\varpi d b^4 + (D-d)L_\varpi b. \tag{225}$$

Furthermore,

$$\frac{2\xi_0}{\tau_0} = 2\frac{\xi_0}{\sigma}\frac{\sigma}{\tau_0} \le \left(\frac{\xi_0}{\sigma}\right)^2 + \left(\frac{\sigma}{\tau_0}\right)^2 \le \varepsilon_{\xi_0,\sigma}. \tag{226}$$

After choosing $\varepsilon_0 < 1$, LMD identifiability follows from Theorem 2.1. Since equation 225 is pointwise on $\mathrm{supp}(Q)$ and $b \lesssim \varepsilon_{\xi_0,\sigma}$, it gives $d\varepsilon^4_{\xi_0,\sigma} + (D-d)\varepsilon_{\xi_0,\sigma}$ up to a filter-dependent constant. This proves equation 7. □

F.4 Sharpness of the population approximation bias

This subsection proves Theorem 4.3 by exhibiting a spherical probability measure for which the normalized radial variance has the expansion

$$\frac{\mathrm{Var}(\text{radial displacement})}{\sigma^2} = \left(\frac{\xi_0}{\sigma}\right)^2 + \frac{d}{2}\left(\frac{\sigma}{R}\right)^2 + o\big((\xi_0/\sigma)^2 + (\sigma/R)^2\big).$$

The first term comes from the two normal shells and the second from the quadratic radial displacement of a curved sphere. We derive this expansion below and then translate the resulting radial eigenvalue into the LHSD lower bound and optimal scale.

Choose the constant c_{loc} appearing in Theorem 4.3 small enough that

$$2c_{\mathrm{loc}}^2 < 1, \qquad c_{\mathrm{loc}}^2(2 + C_{\mathrm{den}}) \le \varepsilon_0, \tag{227}$$

where ε_0 is the threshold in Theorem 4.2. Throughout this subsection, τ_0 and all model envelopes are held fixed; all limits are taken as $\xi_0, \sigma \downarrow 0$ with $u = \xi_0/\sigma \to 0$ and $v = \sigma/\tau_0 \to 0$. Put $R = \tau_0$ and embed $\mathcal{M}_\circ = \mathbb{S}^d(R)$ in the first $d+1$ coordinates of $\mathbb{R}^D$. Let Θ be uniform on the unit sphere, let S be an independent Rademacher variable, and define

$$W = (R + S\xi_0)\Theta \in \mathbb{R}^{d+1} \times \{0\}^{D-d-1}. \tag{228}$$

The induced probability measure is admissible for the tubular class, with latent point $X = R\Theta$ and centered normal perturbation $S\xi_0\Theta$. Define $Q_{\xi_0,R} := \mathcal{L}(W)$. Also set

$$u := \xi_0/\sigma, \qquad v := \sigma/R, \qquad \kappa_\varpi := -\bar{\phi}'_\varpi(1) = \frac{4\varpi^4}{1+\varpi^4} > 0. \tag{229}$$

Let P_T be the orthogonal projector onto $\mathrm{span}\{e_2, \dots, e_{d+1}\}$. In covariance displays we identify this range with $\mathbb{R}^d$ through the ordered basis $(e_2, \dots, e_{d+1})$.

Lemma F.5 (Sphere localization expansion)**.** *For $s \in \{-1, 1\}$, let $y_s := (R + s\xi_0)e_1$ be the query point on the observed sphere of radius $R + s\xi_0$. As $(u, v) \to (0, 0)$, uniformly over these two query points,*

$$\frac{1}{\sigma^2} P_T \Sigma_{Q_{\xi_0,R},\sigma}(y_s) P_T = \big(1 + O(u^2 + v^2)\big) I_d, \tag{230}$$

$$\frac{1}{\sigma^2} e_1^\top \Sigma_{Q_{\xi_0,R},\sigma}(y_s) e_1 = u^2 + \frac{d}{2} v^2 + o(u^2 + v^2). \tag{231}$$

All tangent–radial covariance blocks vanish, and the localized covariance in the remaining $D-d-1$ coordinates is zero.

Proof. **Step 1: Exact spherical coordinates.** Let γ_d denote the standard Gaussian probability measure on $\mathbb{R}^d$, let

$$\nu := \frac{1}{2} \sum_{s' \in \{-1,1\}} \delta_{s'} \otimes \gamma_d,$$

and write $\mathbb{E}_0$ for expectation under ν. Condition on $S' = s'$ and use the north-pole chart with rescaled tangent coordinate $z = \sigma t$. Set

$$a_{s'} := 1 + s'uv, \qquad \mathcal{D}_{s',v} := \{t \in \mathbb{R}^d : v\|t\| < a_{s'}\}.$$

The northern candidate point has coordinates

$$W_T = \sigma t, \qquad W_N = R a_{s'} \sqrt{1 - \frac{v^2\|t\|^2}{a_{s'}^2}}. \tag{232}$$

Its exact radial displacement from the query, divided by σ, is

$$\Delta_{s,u,v}(s', t) := (s' - s)u + \frac{a_{s'}}{v} \left(\sqrt{1 - \frac{v^2\|t\|^2}{a_{s'}^2}} - 1 \right). \tag{233}$$

After removing a common multiplicative factor, the exact northern-hemisphere likelihood ratio with respect to ν is

$$\begin{aligned} L_{s,u,v}(s', t) := \mathbf{1}_{\mathcal{D}_{s',v}}(t)\, a_{s'}^{-d} \left(1 - \frac{v^2\|t\|^2}{a_{s'}^2}\right)^{-1/2} \\ \times \exp\left(-\frac{1}{2}\Delta_{s,u,v}(s', t)^2\right). \end{aligned} \tag{234}$$

For the southern candidate define

$$\Delta^-_{s,u,v}(s', t) := (s' - s)u + \frac{a_{s'}}{v} \left(-\sqrt{1 - \frac{v^2\|t\|^2}{a_{s'}^2}} - 1 \right), \tag{235}$$

$$L^-_{s,u,v}(s', t) := \mathbf{1}_{\mathcal{D}_{s',v}}(t)\, a_{s'}^{-d} \left(1 - \frac{v^2\|t\|^2}{a_{s'}^2}\right)^{-1/2} \exp\left(-\frac{1}{2}\Delta^-_{s,u,v}(s', t)^2\right). \tag{236}$$

Thus, for every polynomial P, let

$$\mathcal{R}_P := \mathbb{E}_0\big[P(S', T) L^-_{s,u,v}(S', T)\big]. \tag{237}$$

Step 2: Likelihood expansion and tail control. Set

$$\mathcal{K}_v := \{t : \|t\| \le (4v)^{-1}\}, \qquad \eta_{u,v} := u^2 + uv + v^2.$$

For sufficiently small u, v, $\mathcal{K}_v \subset \mathcal{D}_{s',v}$ for both signs. Taylor's formula gives, uniformly in s, s' and $t \in \mathcal{K}_v$,

$$\begin{aligned} &\Delta_{s,u,v}(s',t) = (s'-s)u - \frac{v\|t\|^2}{2a_{s'}} + r_{s,u,v}(s',t), \\ &|r_{s,u,v}(s',t)| \le Cv^3\|t\|^4. \end{aligned} \tag{238}$$

Using $|a_{s'}^{-d} - 1| \le Cuv$, the Taylor bound for $(1-r)^{-1/2}$ on $0 \le r \le 1/4$, and $|\exp(-z) - 1| \le |z|\exp(|z|)$ yields

$$|L_{s,u,v}(s',t) - 1| \le C\eta_{u,v}(1 + \|t\|^8)\exp\big(c_\star\|t\|^2\big), \qquad c_\star < \frac{1}{4}, \tag{239}$$

on $\mathcal{K}_v$.

On the northern branch with $t \notin \mathcal{K}_v$, the tangential distance from the query is at least $R/4$. Moreover, polar coordinates and the substitution $r = a_{s'}v^{-1}\sqrt{1-q}$ give, for every fixed $p \le 16$,

$$\int_{\mathcal{K}_v^c \cap \mathcal{D}_{s',v}} (1 + \|t\|^p)\left(1 - \frac{v^2\|t\|^2}{a_{s'}^2}\right)^{-1/2} \exp\left(-\frac{\|t\|^2}{2}\right) dt \le C_p \exp\left(-\frac{c}{v^2}\right). \tag{240}$$

On the southern branch, $|\Delta^-_{s,u,v}| \ge c/v$ once uv is sufficiently small. Hence, the same exponential bound holds for its likelihood contribution; any factors v^{-q} from the southern radial displacement are absorbed after decreasing c. Thus, if $P(s',t)$ is a polynomial of degree at most sixteen, then

$$\int_{\mathcal{K}_v^c} |P(s',t)|\big(1 + L_{s,u,v}(s',t)\big)d\nu(s',t) + |\mathcal{R}_P| \le C_P \exp\left(-\frac{c}{v^2}\right). \tag{241}$$

Moreover, for every fixed $p \le 16$ and $q \le 4$,

$$\mathbb{E}_0\big[\big(1 + \|T\|^p + |\Delta^-_{s,u,v}(S',T)|^q\big)\, L^-_{s,u,v}(S',T)\big] \le C_{p,q}\exp\left(-\frac{c}{v^2}\right). \tag{242}$$

Indeed, $|\Delta^-_{s,u,v}| \le C/v$ on its domain; the resulting power of v^{-1} is absorbed by decreasing c in equation 242. Consequently,

$$\big|\mathbb{E}_0\big(P(S',T)\big(L_{s,u,v}(S',T) - 1\big)\big)\big| \le C_P\eta_{u,v} + C_P\exp\left(-\frac{c}{v^2}\right). \tag{243}$$

Step 3: Normalized moment comparison. Define the exact normalizer

$$Z_{s,u,v} := \mathbb{E}_0\big[L_{s,u,v}(S',T) + L^-_{s,u,v}(S',T)\big]. \tag{244}$$

Taking $P = 1$ gives, for all sufficiently small u, v,

$$|Z_{s,u,v} - 1| \le C\eta_{u,v} + C\exp\left(-\frac{c}{v^2}\right), \qquad Z_{s,u,v} \ge \frac{1}{2}. \tag{245}$$

For any pair of branch observables $F = (F^+, F^-)$, define its exact Gaussian-localized expectation by

$$\mathbb{E}_{s,u,v}F := \frac{\mathbb{E}_0\big[F^+(S',T)L_{s,u,v}(S',T) + F^-(S',T)L^-_{s,u,v}(S',T)\big]}{Z_{s,u,v}}. \tag{246}$$

For a branch-invariant polynomial P, this becomes

$$\mathbb{E}_{s,u,v}P = \frac{\mathbb{E}_0(P(S',T)L_{s,u,v}(S',T)) + \mathcal{R}_P}{Z_{s,u,v}}.$$

If $\mu_P = \mathbb{E}_0 P(S',T)$, subtraction gives

$$\mathbb{E}_{s,u,v}P - \mathbb{E}_0 P = \frac{\mathbb{E}_0((P-\mu_P)(L_{s,u,v}-1)) + \mathcal{R}_P - \mu_P\mathcal{R}_1}{Z_{s,u,v}}.$$

Hence, for every polynomial P of degree at most sixteen,

$$|\mathbb{E}_{s,u,v}P(S',T) - \mathbb{E}_0 P(S',T)| \le C_P\eta_{u,v} + C_P\exp\Big(-\frac{c}{v^2}\Big), \tag{247}$$

uniformly in $s \in \{-1,1\}$.

Step 4: Tangent and radial covariances. The normalized measure defining $\mathbb{E}_{s,u,v}$ is invariant under $t \mapsto Ot$ for $O \in O(d)$, so its tangent mean and every tangent–radial covariance vanish. Applying equation 247 to $tt^\top$ gives equation 230. Define the exact rescaled radial displacement by

$$H_{s,u,v} := \frac{\langle W, e_1\rangle - (R + s\xi_0)}{\sigma}.$$

On the northern branch, $H_{s,u,v} = \Delta_{s,u,v}$; on the southern branch, $H_{s,u,v} = \Delta^-_{s,u,v}$. Moreover, equation 238 and $a_{s'}^{-1} = 1 + O(uv)$ give, on $\mathcal{K}_v$,

$$\begin{aligned} H_{s,u,v}(s',t) &= (s'-s)u - \frac{v}{2}\|t\|^2 + r^H_{s,u,v}(s',t), \\ |r^H_{s,u,v}(s',t)| &\le C\left(uv^2\|t\|^2 + v^3\|t\|^4\right). \end{aligned} \tag{248}$$

Under the reference product measure, S' and T are independent and

$$\mathrm{Var}(S'u) = u^2, \qquad \mathrm{Var}\left(\frac{v}{2}\|T\|^2\right) = \frac{v^2}{4}\,2d = \frac{d}{2}v^2, \qquad \mathrm{Cov}(S', \|T\|^2) = 0. \tag{249}$$

The remainder estimate in equation 248, together with equation 241 and the normalizer lower bound equation 245, gives, for $1 \le q \le 4$,

$$\left(\mathbb{E}_{s,u,v}\left|H_{s,u,v} - \left((S'-s)u - \frac{v}{2}\|T\|^2\right)\right|^q\right)^{1/q} \le C_q(uv^2 + v^3) + C_q\exp\big(-c/v^2\big). \tag{250}$$

By equation 247, the change of measure alters the variance of the leading displacement by at most $C(u^2+v^2)(u^2+uv+v^2) + C\exp\big(-c/v^2\big)$. By equation 250 and Cauchy–Schwarz, replacing the leading displacement by the exact one adds at most $C(u+v)(uv^2+v^3) + C\exp\big(-c/v^2\big)$. These two error bounds are $o(u^2+v^2)$. This proves equation 231. The last assertion follows since W is supported on the first $d+1$ coordinates. □

Theorem 4.3 (restated) (Sharp scale bias tradeoff)**.** *Assume $\tau_0^2\Lambda_f \le C_{\mathrm{den}}$ for a fixed dimensionless constant $0 < C_{\mathrm{den}} < \infty$, and that the model class contains the uniform law on $\mathcal{M}_\circ = \mathbb{S}^d(\tau_0)\times\{0\}^{D-d-1}$ and the observed laws obtained by adding the spherical normal perturbations constructed in Appendix F.4. There exists $c_{\mathrm{loc}} \in (0, 1/\sqrt{2})$ such that, whenever $\xi_0/\sigma \le c_{\mathrm{loc}}$ and $\sigma/\tau_0 \le c_{\mathrm{loc}}$,*

$$\mathfrak{B}_{\xi_0}(\sigma) \asymp_\varpi \left(\frac{\xi_0}{\sigma}\right)^2 + \left(\frac{\sigma}{\tau_0}\right)^2.$$

Consequently, as $\xi_0 \downarrow 0$ with τ_0 fixed,

$$\inf_{\substack{\sigma>0:\ \xi_0/\sigma \le c_{\mathrm{loc}} \\ \sigma/\tau_0 \le c_{\mathrm{loc}}}} \mathfrak{B}_{\xi_0}(\sigma) \asymp_\varpi \frac{\xi_0}{\tau_0}.$$

A scale satisfying $\sigma^2 \asymp \xi_0\tau_0$ attains this infimum up to constant factors.

Proof of Theorem 4.3. **Step 1: A matching spherical lower bound.** Rotational invariance lets us apply Lemma F.5 at every point on either observed sphere. The d tangent eigenvalues of $A_{Q_{\xi_0,R},\sigma}$ are $O(u^2+v^2)$, its radial eigenvalue equals

$$1 - u^2 - \frac{d}{2}v^2 + o(u^2+v^2), \tag{251}$$

and its remaining $D-d-1$ eigenvalues equal one. Since $1 - \bar\phi_\varpi(\lambda) = O_\varpi(\lambda^4)$ near zero and $\bar\phi_\varpi(1-h) = \kappa_\varpi h + O_\varpi(h^2)$, we obtain

$$T_{\sigma,\varpi}(Q_{\xi_0,R})(y) - d = \kappa_\varpi\left(u^2 + \frac{d}{2}v^2\right) + o(u^2+v^2) \tag{252}$$

uniformly on $\mathrm{supp}(Q_{\xi_0,R})$. For $s \in \{-1,1\}$, define the remainder at $y_s = (R + s\xi_0)e_1$ by

$$R_s(u,v) := T_{\sigma,\varpi}(Q_{\xi_0,R})(y_s) - d - \kappa_\varpi \left(u^2 + \frac{d}{2}v^2\right). \tag{253}$$

Equation 252 is uniform in the two values of s, so $\max_s |R_s(u,v)|/(u^2+v^2) \to 0$ as $(u,v) \to (0,0)$. After reducing c_{loc} if necessary, it follows that

$$\max_{s\in\{-1,1\}} |R_s(u,v)| \le \frac{1}{2}\kappa_\varpi \left(u^2 + \frac{d}{2}v^2\right) \tag{254}$$

throughout $0 \le u, v \le c_{\mathrm{loc}}$. By equation 227, $2\xi_0/R = 2uv \le u^2+v^2 \le 2c_{\mathrm{loc}}^2 < 1$, so Theorem 2.1 gives $d_{\mathrm{LMD}}(Q_{\xi_0,R}) = d$. Equation 252 supplies the lower bound in equation 8; the upper bound is Theorem 4.2, using $\sigma^2\Lambda_f \le C_{\mathrm{den}}v^2$ by equation 23.

Step 2: Optimal population scale. Finally, hold τ_0 fixed and set

$$\delta := \frac{\xi_0}{\tau_0}, \qquad t := \frac{\sigma^2}{\xi_0\tau_0}.$$

Then

$$u^2 = \frac{\delta}{t}, \qquad v^2 = \delta t. \tag{255}$$

Therefore, equation 8 gives

$$c_\varpi \delta \left(t \vee t^{-1}\right) \le \mathfrak{B}_{\xi_0}(\sigma) \le C_\varpi \delta \left(t + t^{-1}\right). \tag{256}$$

Since $t \vee t^{-1} \ge 1$, the left-hand side gives a lower bound of order δ for every admissible scale. Conversely, $t = 1$, equivalently $\sigma^2 = \xi_0\tau_0$, is admissible for all sufficiently small δ, and the right-hand side is then of order δ. This proves equation 9. □

G Experimental Details

G.1 Evaluation setup and reported statistics

This appendix extends the experiments in Section 5 with a numerical population reference for the linear model, local-tube references for the curved models, a main-text DiT–reference comparison, and comparisons of noise response, full-scale RMSE, sample-size sensitivity, signed errors, and Hessian spectra, together with the full-scale numerical RMSE table. Table 2 lists the twenty supplementary figures. Table 3 reproduces the main-text RMSE values for convenient comparison with the scale-dependent error curves.

Synthetic models. We use three generators with $D = 256$ and $R = 4$. Let $\iota_k : \mathbb{R}^k \to \mathbb{R}^D$ denote zero-padding after the first k coordinates. Before a common orthogonal rotation, the linear model has $U \sim \mathcal{N}(0, I_d)$ and $F_{\mathrm{lin}}(U) = \iota_d(RU)$; thus each tangent coordinate has variance R^2. For the sphere and squiggly models, $U \sim \mathrm{Unif}(\mathbb{S}^d)$, where $\mathbb{S}^d$ is the unit sphere in $\mathbb{R}^{d+1}$, and

$$F_{\mathrm{sph}}(U) = \iota_{d+1}(RU), \qquad F_{\mathrm{sq}}(U) = \iota_{d+1}(Rf(U)U), \qquad f(U) = 1 + 0.15\sin(6U_1)\sin(6U_2).$$

Observations are generated as

$$Y = X + \xi, \qquad X = OF_g(U), \qquad \xi = ON_g(U)Z, \qquad Z \sim \mathrm{Unif}\big(B^{D-d}(0,\tau)\big),$$

with Z independent of U. The columns of $N_g(U)$ form an orthonormal basis of the normal space at $F_g(U)$ before rotation. For the linear model they are the last $D-d$ coordinate directions. For the sphere they consist of the embedded radial unit vector U and the last $D-d-1$ coordinate directions. For the squiggly model, the first normal is the embedded unit vector

$$\nu(U) = \frac{U - \nabla_{\mathbb{S}^d} f(U)/f(U)}{\|U - \nabla_{\mathbb{S}^d} f(U)/f(U)\|}, \qquad \nabla_{\mathbb{S}^d} f(U) = (I_{d+1} - UU^\top)\nabla f(U),$$

and the remaining normals are the extra coordinate directions. Thus squiggly noise follows the surface normal, which need not be radial. The supplied configuration uses seed 2026; the rotation seed

depends only on this seed and D. Consequently, the same fixed rotation is shared across benchmark geometries and conditions with $D = 256$.

The squiggly law is the pushforward of uniform sphere sampling, not uniform surface-area sampling on the deformed sphere. The linear model is a noncompact flat baseline outside the compact-manifold assumptions. The experimental radius τ equals the theoretical noise bound ξ_0, not the reach lower bound τ_0. The scale R equals the sphere's reach; no such identification is made for squiggly.

Conditions and diffusion scales. The supplementary figures use $d \in \{5, 10, 50, 100\}$, while the numerical RMSE table retains the main-text dimensions $d \in \{10, 50, 100\}$. The noise comparisons use $\tau/R \in \{0, 0.01, 0.05, 0.1\}$ and $n \in \{20{,}000, 100{,}000, 200{,}000\}$. Unless a figure varies these quantities, $n = 200{,}000$ and $\tau/R = 0.05$. Each condition uses $m = 128$ evaluation queries. The spectral-filter parameter is $\varpi = 0.05$. Scale-dependent plots use the ten absolute scales

$$\sigma \in \{0.004, 0.008, 0.016, 0.028, 0.04, 0.08, 0.16, 0.28, 0.40, 0.60\},$$

corresponding to $\sigma/R \in [0.001, 0.15]$. Sample-size and spectral comparisons use the same absolute scale $\sigma = 0.04$ for all sample sizes, as a common scale for examining empirical performance across conditions.

DiT architecture and training. The vector DiT splits the $D = 256$ input coordinates into 32 patches of size 8 and uses eight transformer blocks, hidden width 256, eight attention heads, and an MLP expansion ratio of four. Noise levels are encoded by a 256-dimensional sinusoidal embedding of $\log(\sigma/R)$. The implementation uses additive corruption $W = Y + \sigma G$, corresponding to $a_t = 1$ and $s_t = \sigma$ below, and predicts G directly from inputs preconditioned by $(R^2 + \sigma^2)^{-1/2}$. Training minimizes unweighted noise-prediction MSE, with $\log(\sigma/R)$ sampled uniformly between $\log(0.0005)$ and $\log(20)$. The supplied training configuration uses 100 epochs, batch size 512, AdamW with constant learning rate 10^{-4} and weight decay 0.01, and gradient-norm clipping at 1. The EMA decay is warm-started and capped at 0.9999. After each epoch, the EMA model is evaluated on 10,000 independent validation observations with fixed validation corruptions; the checkpoint with the lowest overall validation MSE is used for estimation. With this epoch-based schedule, increasing the training sample size also increases the number of optimization steps.

Baseline estimators. We use the LIDL and FLIPD estimators discussed in the related-work section, evaluated using the same trained DiT as LHSD. The following settings specify their implementation in our additive-noise convention; they are not a claim that every numerical setting matches the original papers' defaults. Write $\epsilon_\theta(y, \sigma)$ for the predicted standard Gaussian noise and $s_\theta(y, \sigma) = -\epsilon_\theta(y, \sigma)/\sigma$ for the score. FLIPD is evaluated as

$$\widehat{d}_{\mathrm{FLIPD},\sigma}(y) = D + \sigma^2\big(\operatorname{div} s_\theta(y, \sigma) + \|s_\theta(y, \sigma)\|^2\big) = D - \sigma \operatorname{tr} \nabla_y \epsilon_\theta(y, \sigma) + \|\epsilon_\theta(y, \sigma)\|^2.$$

The full noise-predictor Jacobian is computed by automatic differentiation, and its diagonal sum is used without a stochastic trace approximation. Derivatives are with respect to the original input coordinates and include the input preconditioning described above.

For LIDL, we use a two-point central slope in log scale, rather than a multi-point regression. With $h = 0.2$ and $\sigma_\pm = \sigma e^{\pm h}$, the estimate is

$$\widehat{d}_{\mathrm{LIDL},\sigma}(y) = D + \frac{\widehat{\ell}_{\sigma_+}(y) - \widehat{\ell}_{\sigma_-}(y)}{2h},$$

where $\widehat{\ell}_s(y)$ is the probability-flow estimate of the log density at scale s. With integration variable $u = \log s$, we integrate the velocity $v_\theta(x, u) = e^u \epsilon_\theta(x, e^u)$ from the query to the terminal scale $s_{\max} = 20R$, accumulating the integral of $\operatorname{div} v_\theta$. The starting log density is the terminal log density plus this integral. The terminal density is approximated by a Gaussian with the training-data mean and diagonal covariance equal to the coordinatewise training-data variances plus $s_{\max}^2$. Each integration uses 64 uniform steps in log scale with fourth-order Runge–Kutta. Divergences use eight Rademacher Hutchinson probes, held fixed across integration stages and shared between the two LIDL endpoints. These choices introduce ODE, trace-estimation, terminal-density, and finite-difference errors in addition to score-learning error.

Table 2: **Coverage of the complete figure set.** Each figure contains a grid of geometries and latent dimensions.

Diagnostic	Conditions	Figures
Noise response	Four noise levels at each of three sample sizes	4–6
Reference noise response	Four noise levels; common reference curves for all sample sizes	7
Main-text comparison	DiT LHSD with the linear population reference and curved-model local-tube references; $\tau/R = 0.05, 0.1$	8
Full-scale RMSE	$n = 200{,}000$; $\tau/R = 0, 0.01, 0.05, 0.1$	9–12
Sample-size sensitivity	Three sample sizes at $\sigma = 0.04$; four noise levels	13–16
Signed-error distributions	Three sample sizes; $\tau/R = 0.05$	17–19
Mean sorted spectra	$\sigma = 0.04$; $\tau/R = 0, 0.05, 0.1$; three sample sizes	20–22
Filter-width sensitivity	$n = 200{,}000$, $\sigma = 0.04$; eight noise levels	23

Learned Hessian and calibration. Within each experimental condition, LIDL, FLIPD, and LHSD are computed from the same trained DiT model. Under $Z_t = a_t Y + s_t G$, with $a_t, s_t > 0$ and independent $G \sim \mathcal{N}(0, I_D)$, the effective smoothing scale is $\sigma_t = s_t/a_t$. If p_t is the density of Z_t, the population noise predictor satisfies $\epsilon_*(z,t) = -s_t \nabla_z \log p_t(z)$. Consequently,

$$s_t \nabla_z \epsilon_*(a_t y, t) = -s_t^2 \nabla_z^2 \log p_t(a_t y) = A_{Q,\sigma_t}(y).$$

For a trained predictor, we use

$$\begin{aligned} \widehat{A}_{\theta,t}(y) &= s_t \,\mathrm{sym}(\nabla_z \epsilon_\theta(a_t y, t)) \,, \\ \widehat{T}_{\theta,t}(y) &= \mathrm{tr}\, \bar{\phi}_\varpi\Big(\widehat{A}_{\theta,t}(y)\Big) \,, \end{aligned} \tag{257}$$

where $\mathrm{sym}(B) = (B + B^\top)/2$. As in Section 3.2,

$$\phi_\varpi(\lambda) = \frac{\varpi^4}{\varpi^4 + \lambda^4}, \qquad r_\varpi = \phi_\varpi(1), \qquad \bar{\phi}_\varpi(\lambda) = \frac{\phi_\varpi(\lambda) - r_\varpi}{1 - r_\varpi}.$$

Thus $\bar{\phi}_\varpi(0) = 1$ and $\bar{\phi}_\varpi(1) = 0$, and

$$\widehat{T}_{\theta,t}(y) = \frac{\mathrm{tr}\, \phi_\varpi\Big(\widehat{A}_{\theta,t}(y)\Big) - D r_\varpi}{1 - r_\varpi}.$$

All LHSD dimension estimates and reference counts in the main text and this appendix use this affine calibration with $\varpi = 0.05$. For the linear model, the reference is a numerical population reference; for the sphere and squiggly models, the references are local-tube approximations. This includes the main-text Figure 2 comparison reproduced in Figure 8, the supplementary dimension and error plots, and Table 3. The spectral figures show the eigenvalues of the learned normalized Hessian, before applying the spectral filter.

Reported statistics. For an estimate $\widehat{d}_\sigma(y_j)$ at query y_j, define the signed error $e_j(\sigma) = \widehat{d}_\sigma(y_j) - d$. The query mean and root mean squared error are

$$\overline{\widehat{d}}(\sigma) = \frac{1}{m} \sum_{j=1}^m \widehat{d}_\sigma(y_j), \qquad \mathrm{RMSE}(\sigma) = \left[\frac{1}{m} \sum_{j=1}^m e_j(\sigma)^2 \right]^{1/2}.$$

The noise-response figures display $\overline{\widehat{d}}(\sigma)/d$, while the signed-error figures show the distribution of $e_j(\sigma)$. RMSE measures deviations across queries without cancellation between positive and negative errors.

G.2 JOINT DEPENDENCE ON NOISE AND SAMPLE SIZE

Figures 4–6 compare the four noise levels at each training sample size. With the larger training sets, many normalized estimates remain close to one over an intermediate range of scales. The location and extent of this range vary with noise level, dimension, and geometry.

At $n = 20{,}000$, accurate dimension estimation is less consistent across scales and geometries. Several low-dimensional cases underestimate d over a broad range, while high-dimensional linear cases retain substantial overestimation to larger scales than their larger-sample counterparts. These observations show that the useful scale range depends on the training sample size as well as the underlying geometry and noise.

G.3 NUMERICAL AND LOCAL-TUBE REFERENCES

Reference targets and query averaging. The form of the reference depends on the geometry. For the linear model, let Q_{lin} denote the observation law of $Y = X + \xi$, including the bounded normal noise. The linear reference is the population functional

$$T^{\mathrm{ref}}_{\mathrm{lin},\sigma,\varpi}(y) = T_{\sigma,\varpi}(Q_{\mathrm{lin}})(y) = \operatorname{tr} \bar{\phi}_{\varpi}\left(I_D - \frac{\Sigma_{Q_{\mathrm{lin}},\sigma}(y)}{\sigma^2}\right).$$

Here $\Sigma_{Q_{\mathrm{lin}},\sigma}(y)$ is the covariance under Q_{lin}, reweighted by $\exp(-\|v-y\|^2/(2\sigma^2))$ and normalized to unit mass.

For the sphere and squiggly models, the displayed references are local-tube approximations evaluated at each query. We denote these values by

$$T^{\mathrm{LT}}_{g,\sigma,\varpi}(y), \qquad g \in \{\mathrm{sph}, \mathrm{sq}\}.$$

These references approximate the curved generator locally and do not numerically integrate the full population law of the corresponding curved generator.

For each condition and scale, the relevant reference is evaluated at the same $m = 128$ queries used for the DiT. The displayed curve is

$$\overline{T}^{\mathrm{ref}}_{g,\sigma,\varpi} = \frac{1}{m}\sum_{j=1}^{m} T^{\mathrm{ref}}_{g,\sigma,\varpi}(y_j),$$

where

$$T^{\mathrm{ref}}_{g,\sigma,\varpi}(y) = \begin{cases} T^{\mathrm{ref}}_{\mathrm{lin},\sigma,\varpi}(y), & g = \mathrm{lin}, \\ T^{\mathrm{LT}}_{g,\sigma,\varpi}(y), & g \in \{\mathrm{sph}, \mathrm{sq}\}. \end{cases}$$

The spectral filter and affine calibration are applied separately at each query before averaging. This differs from applying the nonlinear filter to a query-averaged spectrum. The plotted line is a finite-query average rather than an exact expectation over all possible queries. The reference calculations do not depend on the training sample size.

Linear numerical population reference. In canonical coordinates, Q_{lin} factors into $\mathcal{N}(0, R^2 I_d)$ in the tangent coordinates and the uniform law on $B^{D-d}(0,\tau)$ in the normal coordinates. The Gaussian-weighted tangent mean and covariance are analytic. The normal component requires the mass, first moment, and second moment of a Gaussian centered at the query's normal component, restricted to this ball. These quantities are evaluated using noncentral chi-square cumulative probabilities with degrees of freedom $D-d$, $D-d+2$, and $D-d+4$. The resulting weighted covariance yields the normalized Hessian and the affine-calibrated LHSD. This is a special-function evaluation, not a Monte Carlo average over generated training observations. Computations use float64 and log-domain evaluations, with a Poisson–gamma series fallback for extreme lower-tail probabilities. The implementation includes checks against the noiseless analytic case and independent one-dimensional quadrature for codimension one. These checks do not constitute a certified error bound for every benchmark condition, and an analytic reduction does not eliminate floating-point error.

Curved-model local-tube references. For the sphere and squiggly models, the reference calculation uses a local-tube approximation at each evaluation query. The curved geometry is represented locally by the corresponding tangent and normal tube geometry, together with the bounded normal-noise radius τ. The approximation is evaluated separately at each query and diffusion scale before applying the spectral filter and affine calibration.

For the sphere, the local-tube construction uses the local tangent and normal geometry of the radius-R sphere. For the squiggly model, it uses the local geometry and normal direction induced by the radial deformation $U \mapsto Rf(U)U$. Thus, the squiggly reference reflects the location-dependent local geometry, but it does not integrate over the complete pushforward distribution of the deformed sphere.

The local-tube references should therefore not be interpreted as exact population LHSD values for the full curved generators. Their numerical evaluation error is distinct from the approximation error introduced by replacing the full curved generator with its local-tube model.

Interpretation of the reference plots. Figure 7 displays the query-averaged reference counts divided by d for comparisons with $n = 20{,}000$, $100{,}000$, and $200{,}000$. Figure 8 reproduces the main-text Figure 2 comparison at $n = 200{,}000$ and $\tau/R \in \{0.05, 0.1\}$.

The learned and reference counts use the same affine-calibrated spectral filter with $\varpi = 0.05$. For the linear model, the difference between the learned estimate and the reference reflects score-learning and numerical reference errors. For the sphere and squiggly models, the difference may additionally include local-tube approximation error. Accordingly, agreement with a curved-model reference should be interpreted as agreement with the corresponding local-tube calculation rather than with an exact numerical integral of the full generator distribution.

G.4 Full-scale errors at all noise levels

Figures 9–12 compare LIDL, FLIPD, and affine-calibrated LHSD over the full scale grid at all four noise levels. LHSD achieves substantially smaller errors over an intermediate range in many conditions. The most accurate scales nevertheless depend on the condition: some low-noise, low-dimensional sphere cases already have small LHSD error at the smallest evaluated scales.

A lower RMSE than competing methods does not necessarily imply a small absolute estimation error. For the linear model with $d = 100$ and $\tau/R = 0.05$, Table 3 gives LHSD RMSE 104.312 at $\sigma = 0.016$, 1.012 at $\sigma = 0.04$, and 0.043 at $\sigma = 0.16$. This example illustrates why the comparison with baselines must be considered alongside the magnitude of the error.

The higher-dimensional curved models remain vulnerable to coarse-scale underestimation across the noise settings. At $\tau/R = 0.05$, $\sigma = 0.40$, and $d = 100$, Table 3 reports LHSD RMSE 0.239 for the linear model, compared with 99.902 and 99.986 for the sphere and squiggly models. LIDL or FLIPD is more accurate in several such coarse-scale conditions. Thus, the full scale grid identifies both the range of empirical accuracy and the conditions under which it deteriorates.

G.5 Training sample-size sensitivity at a common scale

Figures 13–16 compare $n = 20{,}000, 100{,}000, 200{,}000$ at the common scale $\sigma = 0.04$ for each noise level. At this scale and $n = 200{,}000$, LHSD has lower RMSE than both baselines in all evaluated geometry–dimension combinations at the four noise levels.

Larger training sets often yield smaller LHSD errors, particularly in the high-dimensional linear cases, but the observed changes are not uniformly monotone. For example, at $\tau/R = 0.05$, the squiggly $d = 5$ error increases across the displayed sample sizes, while the $d = 50$ error is non-monotone. Each condition uses one model seed. The results therefore provide a descriptive comparison of individual trained models at different sample sizes, without separately quantifying variability across training runs.

G.6 Signed errors and variation across queries

Figures 17–19 show the signed-error distributions at $\tau/R = 0.05$ for all three sample sizes. With the smaller training set, substantial positive errors persist in high-dimensional conditions. The linear model also exhibits greater query-to-query variation near the transition from overestimation to smaller errors. With the larger training sets, many intermediate-scale LHSD errors move close to zero. At coarse scales, errors in the higher-dimensional sphere and squiggly models remain strongly negative, indicating systematic underestimation.

G.7 Mean sorted spectra across noise levels

Figures 20–22 display the mean sorted eigenvalues of the learned normalized Hessian at $\sigma = 0.04$ for three noise levels and three training sample sizes. Eigenvalues are sorted separately at each query and then averaged across queries. In the ideal projector geometry, d eigenvalues are zero and the remaining $D - d$ are one. The affine-calibrated filter maps these ideal tangent and normal contributions to one and zero, respectively.

The larger-sample spectra generally exhibit clearer separation near rank d, providing a spectral interpretation of low LHSD error at the shared scale. Increasing the noise level tends to lower the remaining spectral bulk at this fixed scale. In the high-dimensional linear cases, eigenvalues beyond rank d can remain far below one even with $n = 200{,}000$.

G.8 Sensitivity to the spectral-filter width

Figure 23 examines the choice of the common spectral-filter width $\varpi = 0.05$. We fix the learned Hessian spectra, $n = 200{,}000$, and the absolute diffusion scale $\sigma = 0.04$, and vary only the filter width. The figure covers the three geometries, $d \in \{10, 50, 100\}$, and all eight saved noise levels.

Across these conditions, widths near 0.05 generally place the normalized affine-calibrated count close to one. Substantially smaller widths suppress tangent contributions in many panels, while larger widths admit more non-tangent contributions, most visibly in the high-dimensional linear cases. As a descriptive aggregate check, the root mean squared deviation from one across the 72 displayed conditions is 0.0328 near $\varpi = 0.042$ and 0.0332 at $\varpi = 0.05$ for the plotted mean-spectrum approximation. This broad low-error region supports using 0.05 as a common empirical working width rather than tuning a different width for each condition.

The sensitivity curves apply the nonlinear filter to the saved query-mean eigenvalues. They are therefore not the query averages of LHSD at the alternative widths. The crosses at $\varpi = 0.05$ are the saved query-averaged affine-calibrated LHSD values. This diagnostic supports the stability of the chosen width, but does not establish that 0.05 minimizes query-level RMSE or is optimal for every condition.

Latent manifold	$\sigma = 0.004$	$\sigma = 0.008$	$\sigma = 0.016$	$\sigma = 0.028$	$\sigma = 0.040$
$d = 10$, linear	241.990 / 240.955 / **107.370**	225.798 / 224.135 / **3.129**	162.760 / 166.283 / **0.045**	83.926 / 93.221 / **0.006**	47.192 / 55.244 / **0.012**
$d = 10$, sphere	243.164 / 242.165 / **39.103**	225.659 / 224.350 / **0.418**	159.243 / 157.544 / **0.014**	79.006 / 82.828 / **0.004**	40.223 / 47.546 / **0.015**
$d = 10$, squiggly	243.045 / 242.153 / **40.030**	226.474 / 224.363 / **1.416**	165.484 / 158.020 / **1.013**	92.191 / 83.554 / **0.996**	53.732 / 48.828 / **0.961**
$d = 50$, linear	**204.930** / 204.945 / 205.809	201.686 / 201.699 / **172.363**	189.284 / 189.338 / **15.002**	161.706 / 161.795 / **0.446**	132.178 / 132.264 / **0.055**
$d = 50$, sphere	202.422 / 202.972 / **93.986**	187.160 / 191.894 / **1.396**	128.199 / 148.117 / **0.023**	62.293 / 86.257 / **0.081**	35.185 / 52.408 / **0.266**
$d = 50$, squiggly	203.672 / 202.952 / **95.139**	194.087 / 191.997 / **2.223**	154.485 / 148.253 / **0.969**	94.355 / 86.158 / **0.789**	58.390 / 51.574 / **0.704**
$d = 100$, linear	**155.752** / 155.754 / 155.998	154.992 / **154.981** / 155.471	151.960 / 151.914 / **104.312**	144.072 / 143.926 / **10.256**	133.198 / 132.912 / **1.012**
$d = 100$, sphere	154.319 / 154.387 / **128.966**	148.298 / 148.945 / **4.936**	122.852 / 125.776 / **0.065**	78.670 / 83.519 / **0.131**	49.613 / 53.710 / **0.318**
$d = 100$, squiggly	154.690 / 154.364 / **125.773**	150.021 / 148.793 / **4.891**	129.786 / 125.054 / **0.976**	90.853 / 82.367 / **0.521**	61.043 / 52.385 / **0.351**

Latent manifold	$\sigma = 0.080$	$\sigma = 0.160$	$\sigma = 0.280$	$\sigma = 0.400$	$\sigma = 0.600$
$d = 10$, linear	14.827 / 17.600 / **0.011**	6.016 / 6.547 / **0.007**	3.647 / 3.457 / **0.002**	3.140 / 2.735 / **0.001**	3.135 / 2.693 / **0.280**
$d = 10$, sphere	9.134 / 12.784 / **0.294**	2.702 / **2.052** / 5.494	2.160 / **0.931** / 9.165	2.100 / **0.948** / 9.455	2.044 / **1.106** / 7.652
$d = 10$, squiggly	16.556 / 16.027 / **0.188**	4.752 / 4.362 / **0.164**	2.177 / 1.041 / **0.732**	2.210 / **0.956** / 1.298	2.843 / 1.941 / **1.311**
$d = 50$, linear	63.716 / 63.677 / **0.040**	21.566 / 21.198 / **0.058**	8.809 / 8.041 / **0.033**	5.019 / 3.555 / **0.038**	3.559 / 1.204 / **0.621**
$d = 50$, sphere	16.429 / 20.744 / **2.397**	16.604 / 17.102 / **9.126**	18.276 / 20.155 / **5.979**	**13.988** / 17.060 / 46.408	**5.490** / 8.758 / 49.933
$d = 50$, squiggly	20.385 / 17.350 / **0.546**	6.530 / **3.985** / 21.356	2.745 / **2.200** / 49.532	**2.458** / 5.056 / 49.959	**5.828** / 11.147 / 49.995
$d = 100$, linear	91.906 / 91.188 / **0.012**	41.051 / 40.378 / **0.043**	16.389 / 15.692 / **0.070**	8.456 / 7.367 / **0.239**	3.982 / **1.448** / 2.722
$d = 100$, sphere	18.494 / 20.045 / **1.258**	**9.632** / 11.028 / 29.794	3.993 / **3.630** / 98.384	**6.987** / 9.099 / 99.902	**21.966** / 31.417 / 99.990
$d = 100$, squiggly	21.896 / 15.924 / **13.984**	6.078 / **1.752** / 96.179	**4.302** / 13.006 / 99.915	**9.935** / 25.361 / 99.986	**23.983** / 46.692 / 99.997

Table 3: **Full-scale RMSE at $\xi_0/R = 0.05$.** Each entry reports LIDL / FLIPD / affine-calibrated LHSD for $n = 200{,}000$, $D = 256$, $R = 4$. The LHSD filter width is $\varpi = 0.05$. The upper and lower blocks cover the ten diffusion scales in order. Bold indicates the lowest RMSE in each cell. The numerical values reproduce the main-text RMSE table for reference.

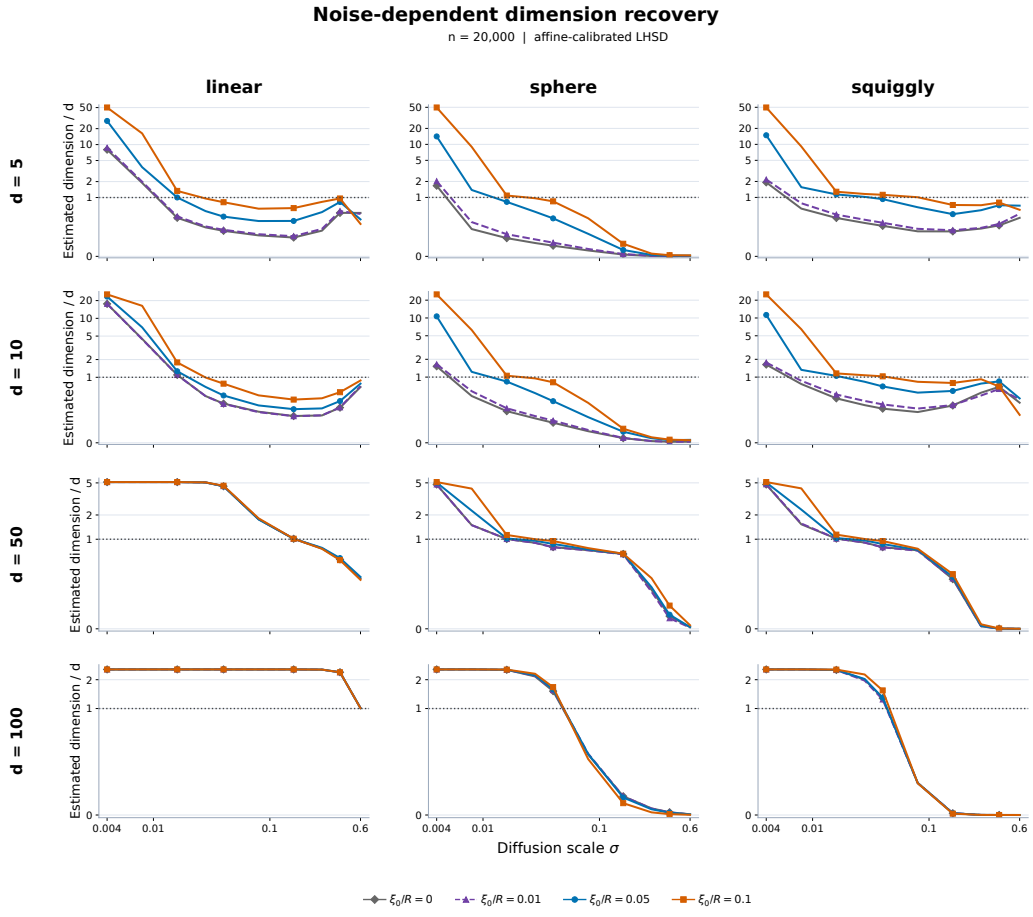


Figure 4: **Noise response with $n = 20{,}000$.** Affine-calibrated LHSD divided by d, with $\varpi = 0.05$. Curves compare $\xi_0/R = 0, 0.01, 0.05, 0.1$; the 0.01 curve is dashed. The horizontal reference is one.

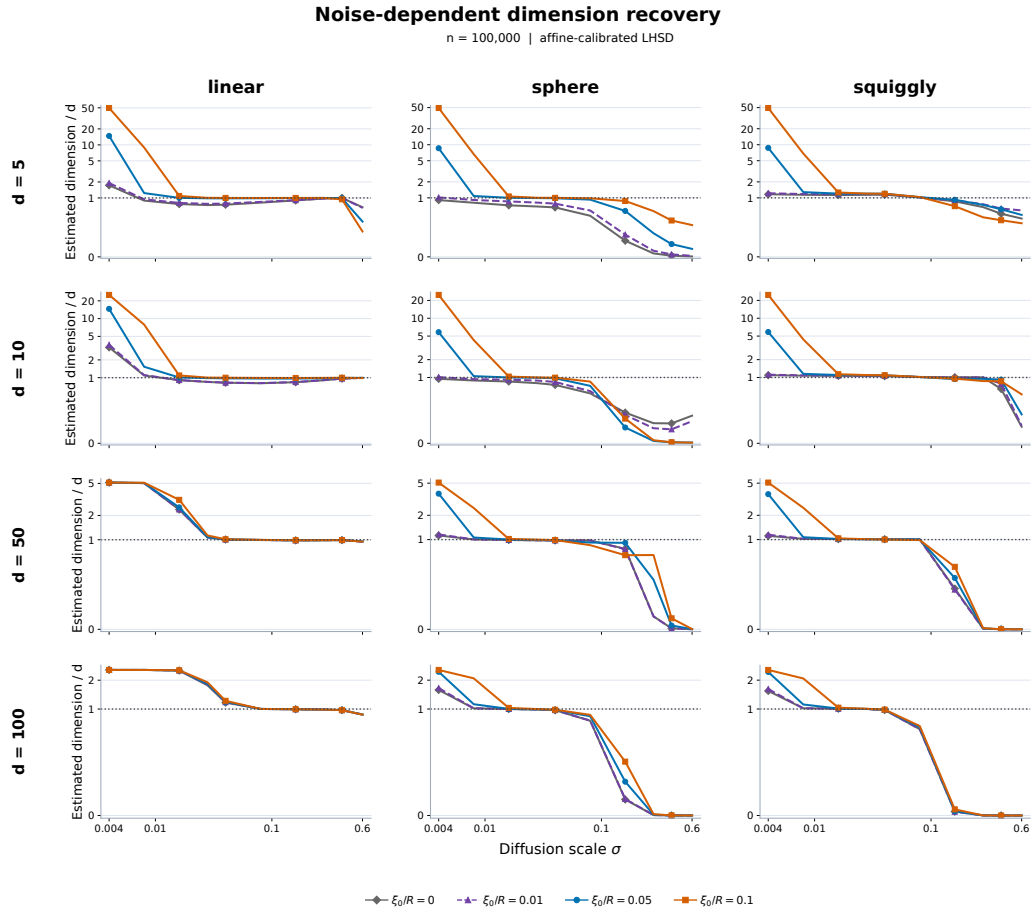


Figure 5: **Noise response with** $n = 100{,}000$**.** Affine-calibrated LHSD divided by d, with $\varpi = 0.05$.

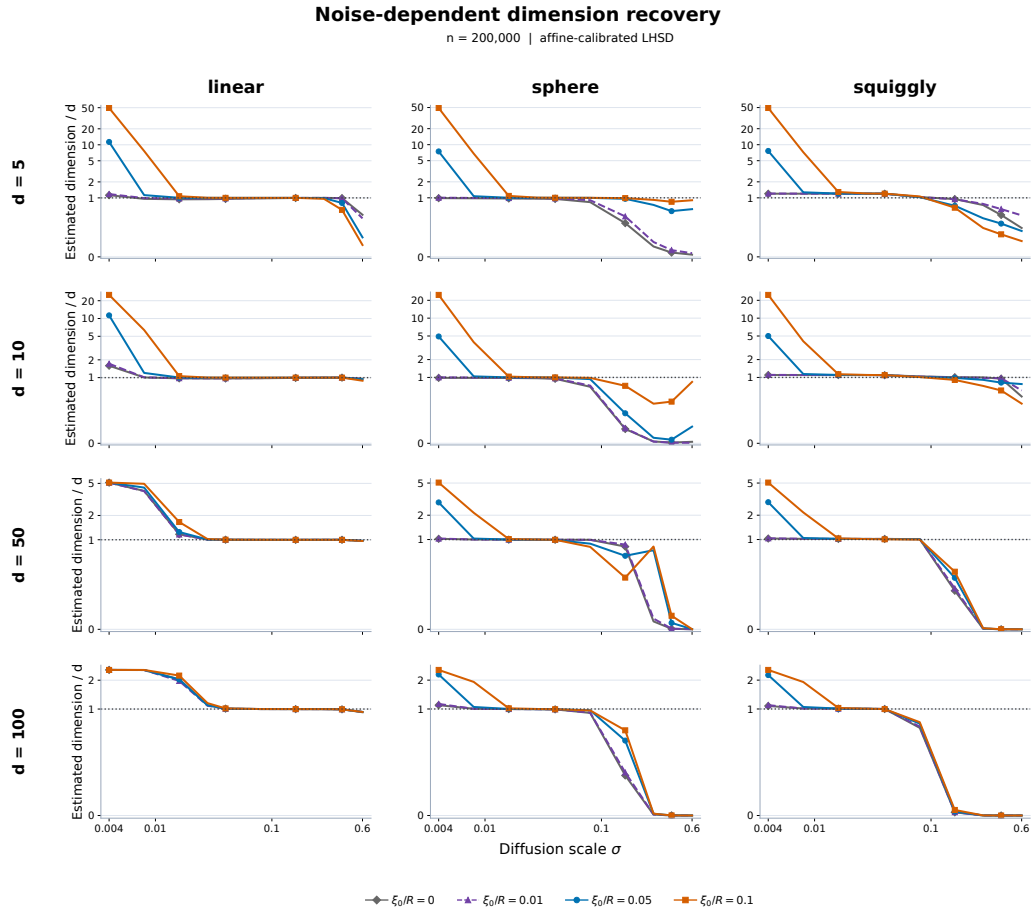


Figure 6: **Noise response with** $n = 200{,}000$**.** Affine-calibrated LHSD divided by d, with $\varpi = 0.05$.

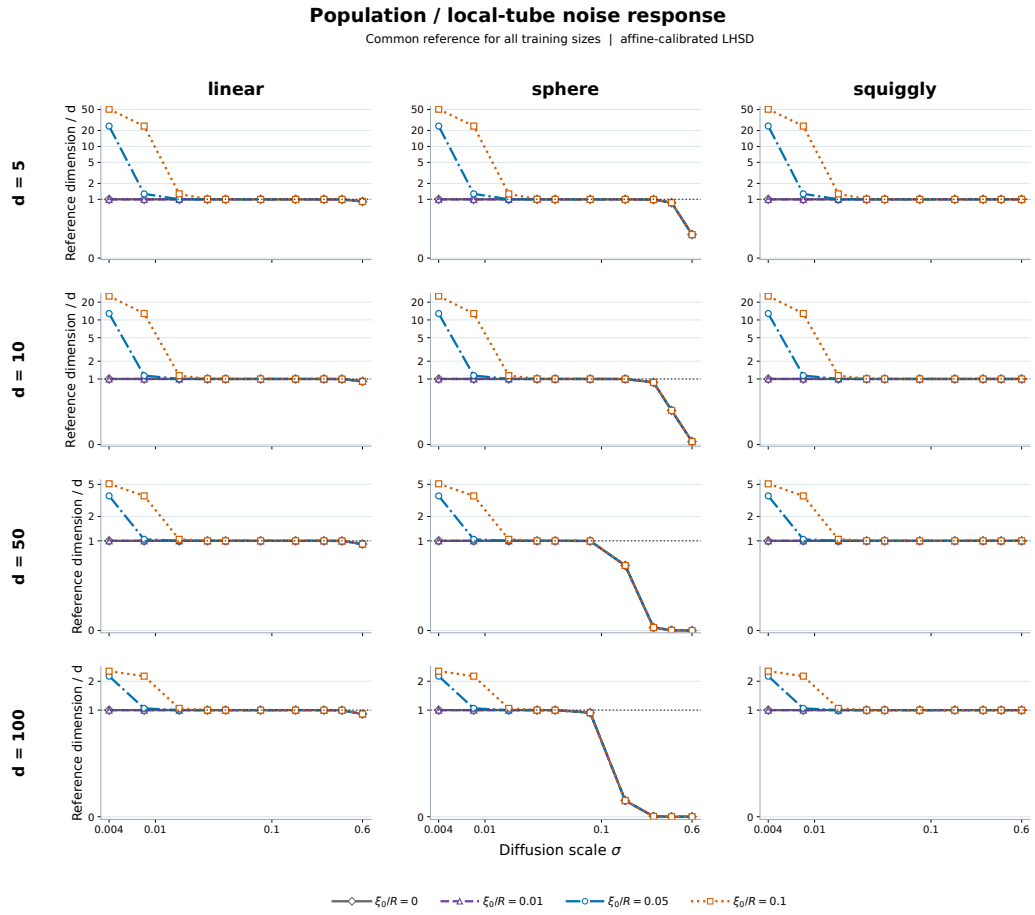


Figure 7: **Reference noise response common to all training sample sizes.** Affine-calibrated reference count divided by d, for four noise levels. The linear model uses a numerical population reference, whereas the sphere and squiggly models use local-tube references. At each scale, the affine-calibrated reference values are evaluated at the same 128 queries as the DiT estimates and then averaged. The references are independent of the training sample size and apply to comparisons with $n = 20{,}000$, $100{,}000$, and $200{,}000$.

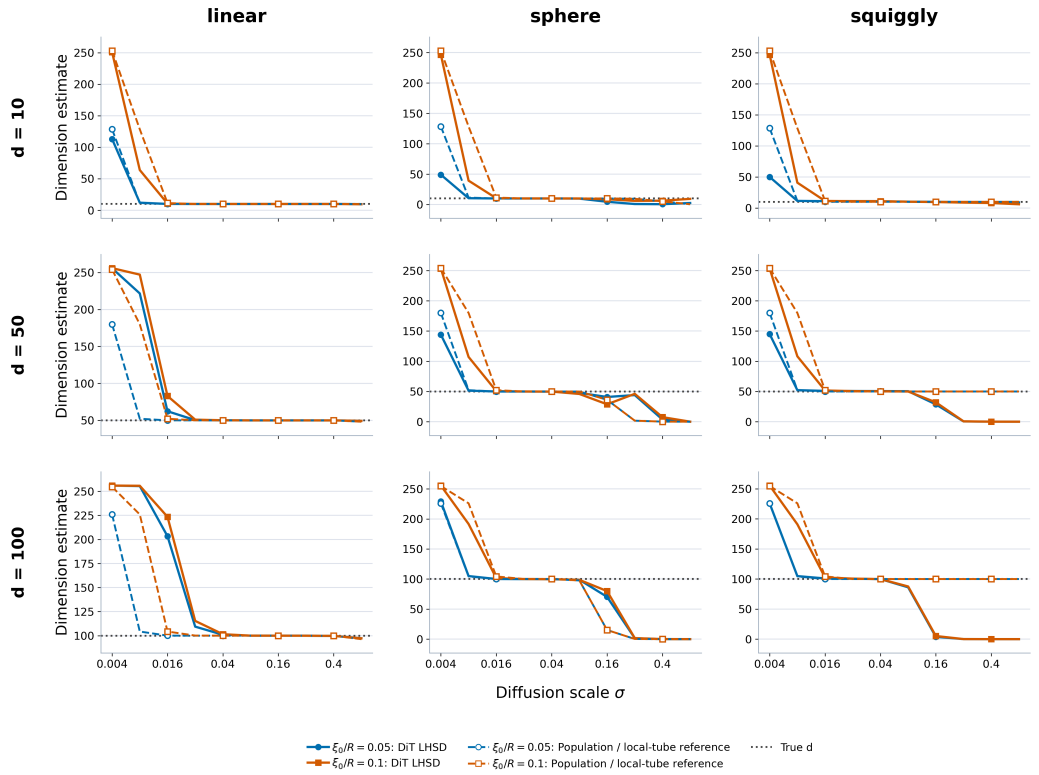


Figure 8: **Main-text comparison of affine-calibrated DiT LHSD and references.** Reproduced from the main-text Figure 2, with $n = 200{,}000$ and $\varpi = 0.05$. Columns indicate geometry and rows indicate latent dimension $d = 10, 50, 100$. Blue and orange denote $\xi_0/R = 0.05$ and 0.1, respectively. Solid curves with filled circles show query-averaged, affine-calibrated DiT LHSD. Dashed curves with open squares show the numerical population reference for the linear model and local-tube references for the sphere and squiggly models, using the same affine calibration. All references are evaluated at the same queries as the DiT estimates and then averaged. Dotted lines indicate the latent dimension d. Horizontal positions are equally spaced for the evaluated scales.

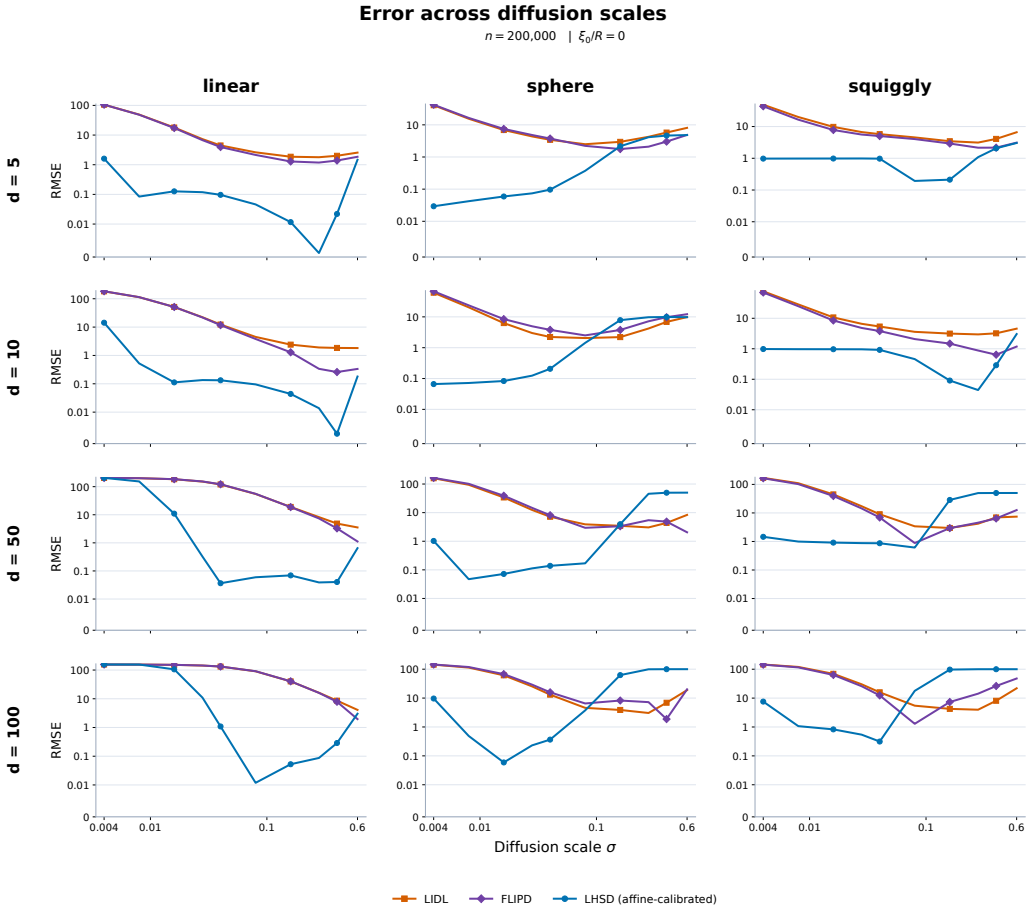


Figure 9: **Full-scale RMSE at $\xi_0/R = 0$.** LIDL, FLIPD, and affine-calibrated LHSD with $n = 200{,}000$; LHSD uses $\varpi = 0.05$.

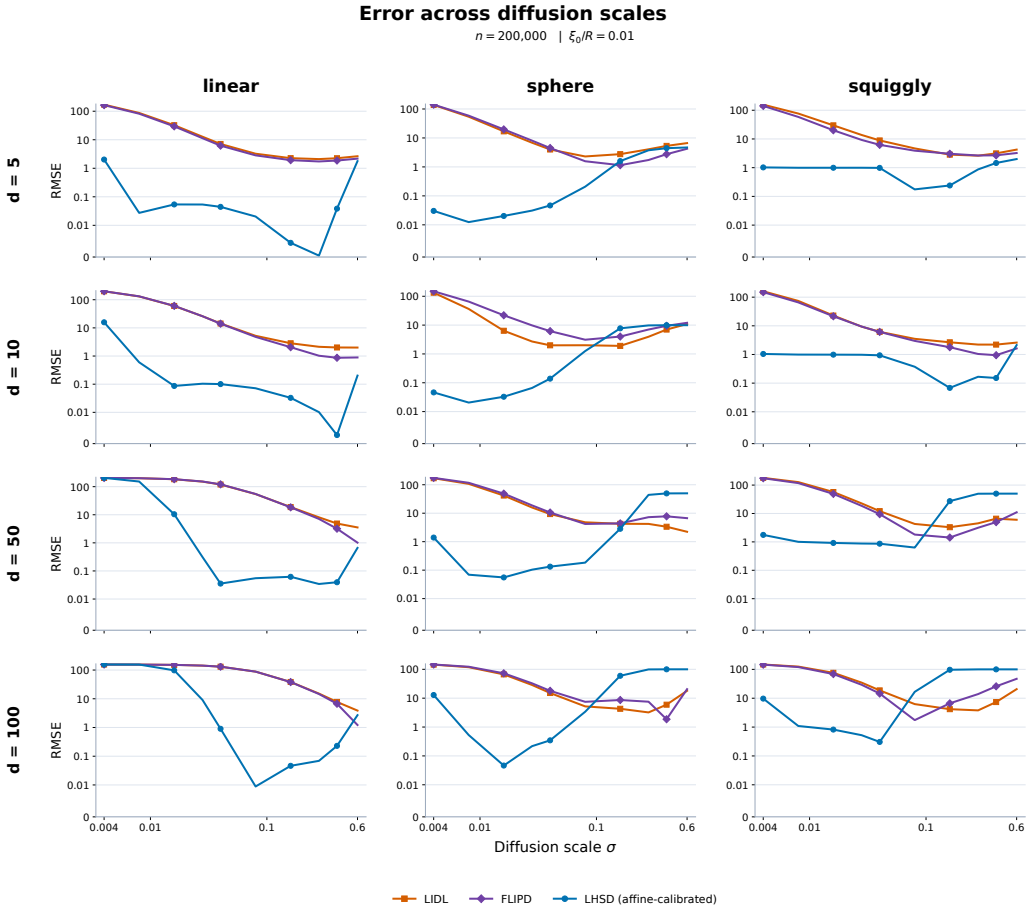


Figure 10: **Full-scale RMSE at $\xi_0/R = 0.01$.** LIDL, FLIPD, and affine-calibrated LHSD with $n = 200{,}000$; LHSD uses $\varpi = 0.05$.

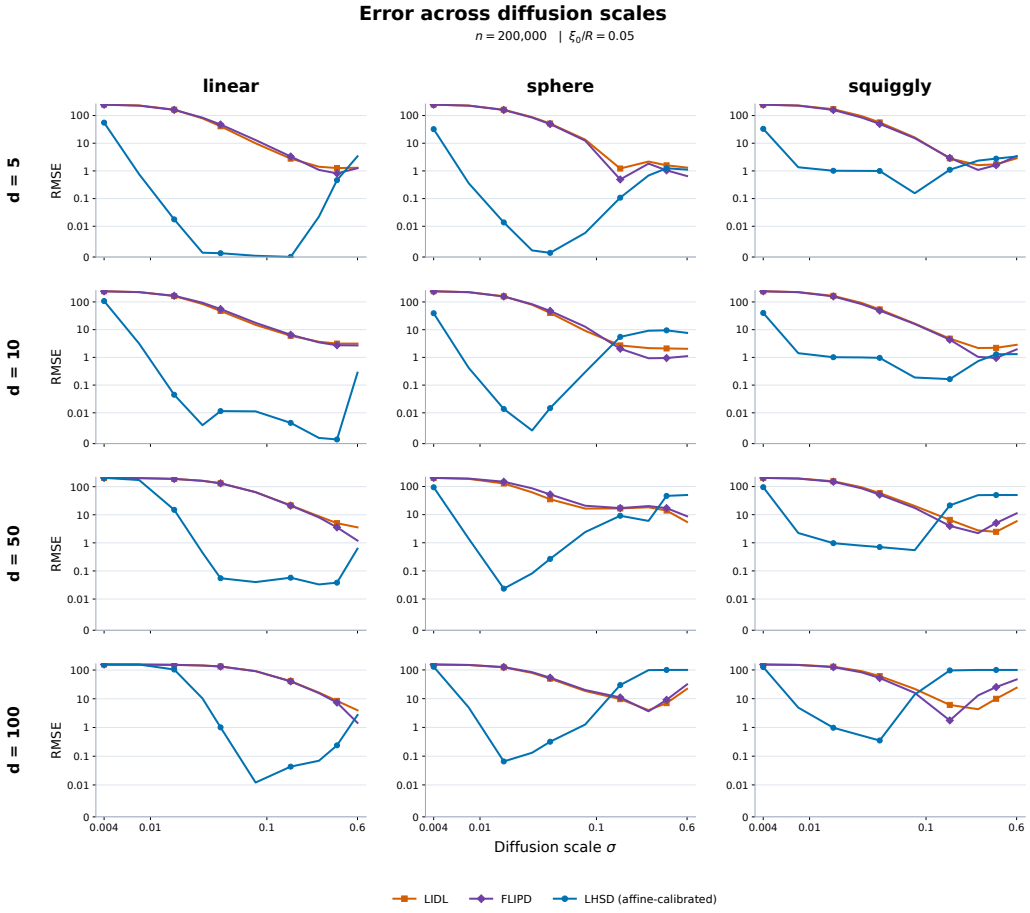


Figure 11: **Full-scale RMSE at $\xi_0/R = 0.05$.** LIDL, FLIPD, and affine-calibrated LHSD with $n = 200{,}000$; LHSD uses $\varpi = 0.05$.

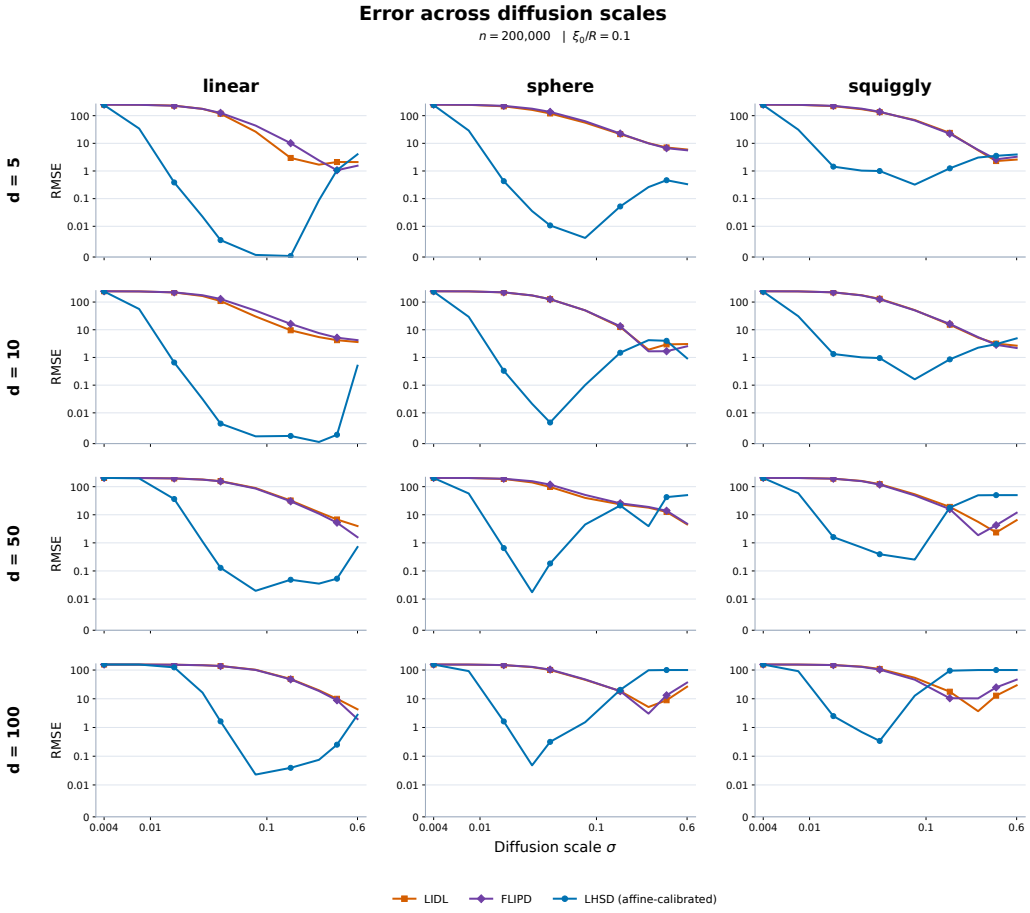


Figure 12: **Full-scale RMSE at $\xi_0/R = 0.1$.** LIDL, FLIPD, and affine-calibrated LHSD with $n = 200{,}000$; LHSD uses $\varpi = 0.05$.

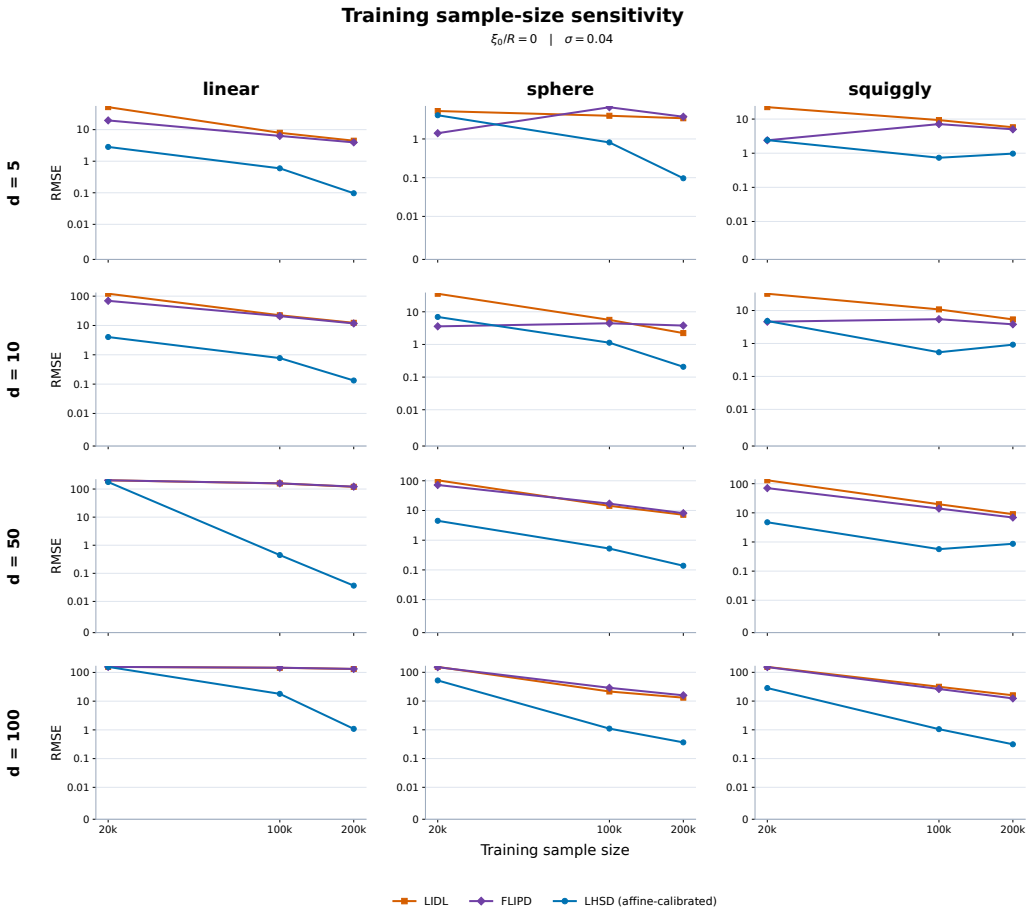


Figure 13: **Sample-size sensitivity at $\xi_0/R = 0$.** RMSE for $n = 20{,}000, 100{,}000, 200{,}000$ at the same absolute scale $\sigma = 0.04$. LHSD is affine-calibrated with $\varpi = 0.05$.

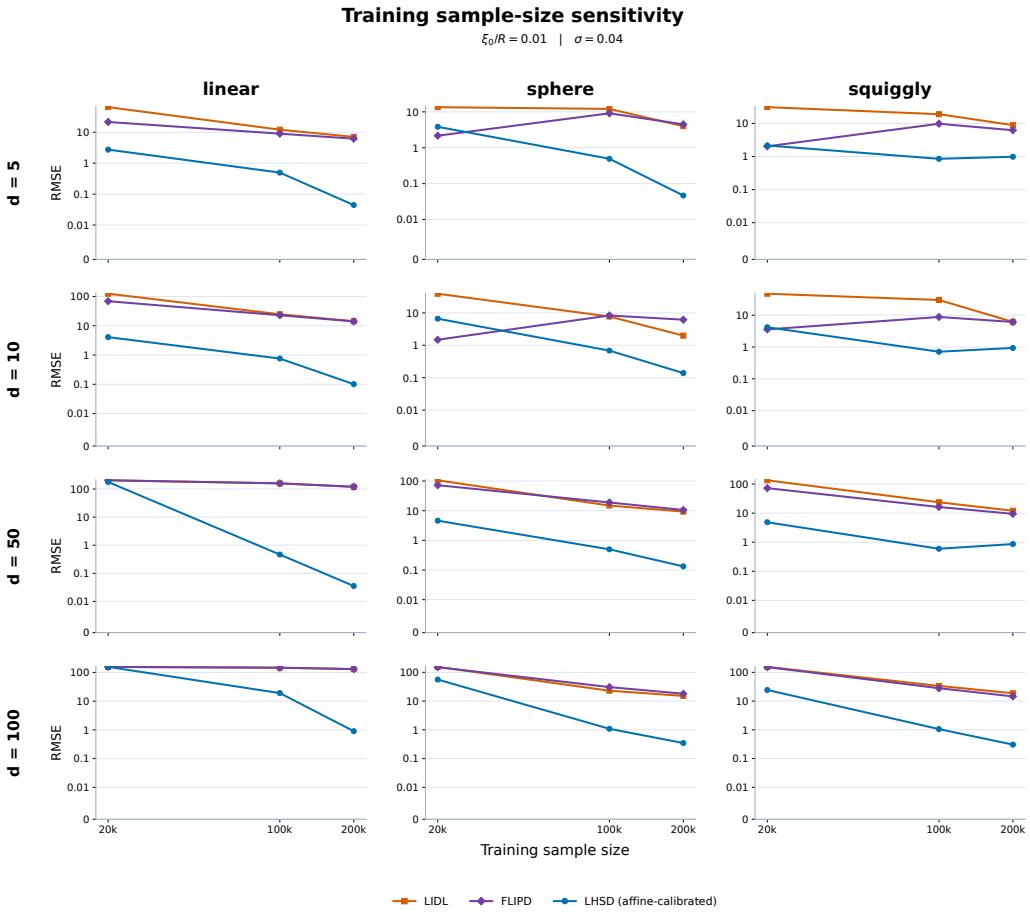


Figure 14: **Sample-size sensitivity at $\xi_0/R = 0.01$.** RMSE for $n = 20{,}000, 100{,}000, 200{,}000$ at the same absolute scale $\sigma = 0.04$. LHSD is affine-calibrated with $\varpi = 0.05$.

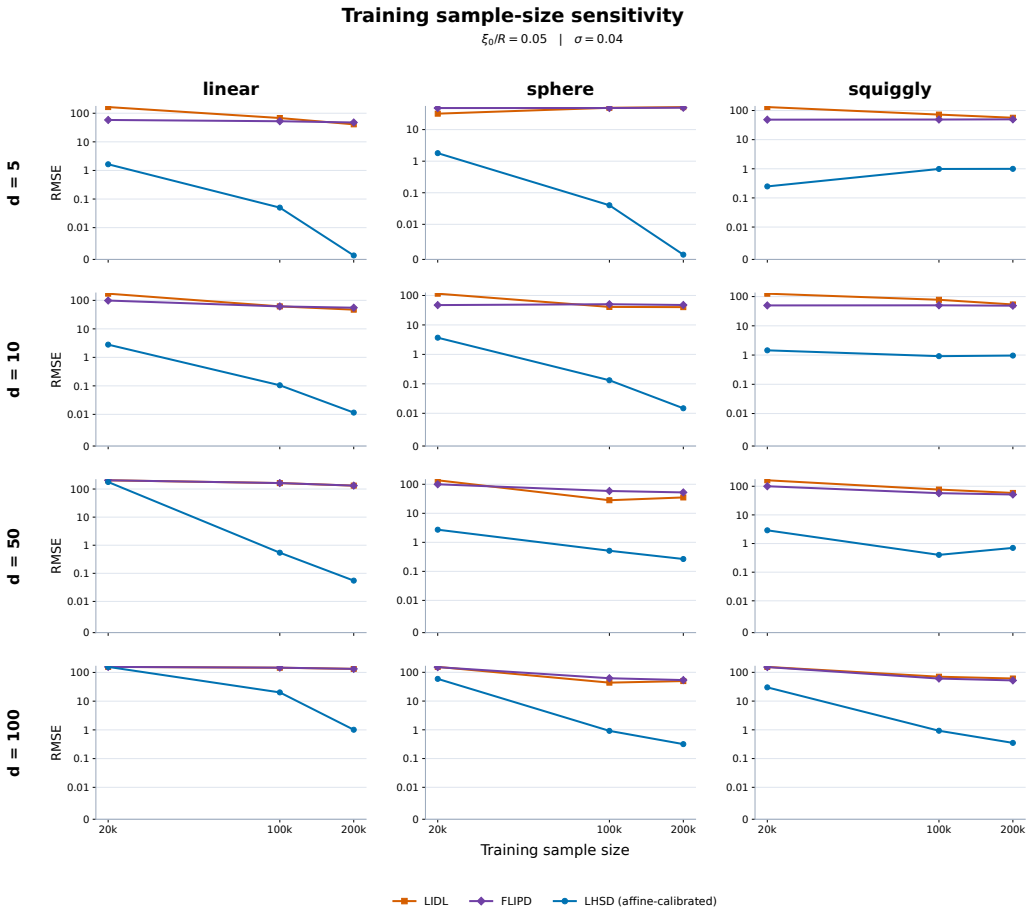


Figure 15: **Sample-size sensitivity at $\xi_0/R = 0.05$.** RMSE for $n = 20{,}000, 100{,}000, 200{,}000$ at the same absolute scale $\sigma = 0.04$. LHSD is affine-calibrated with $\varpi = 0.05$.

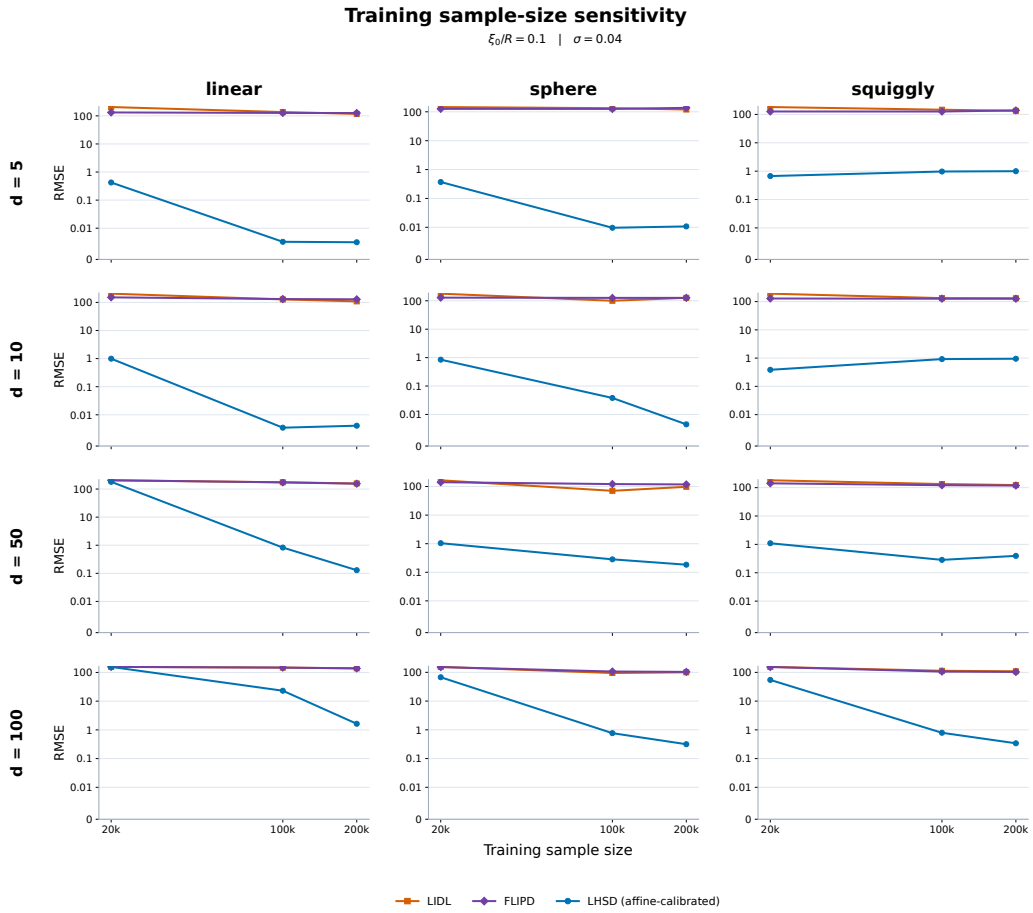


Figure 16: **Sample-size sensitivity at $\xi_0/R = 0.1$.** RMSE for $n = 20{,}000, 100{,}000, 200{,}000$ at the same absolute scale $\sigma = 0.04$. LHSD is affine-calibrated with $\varpi = 0.05$.

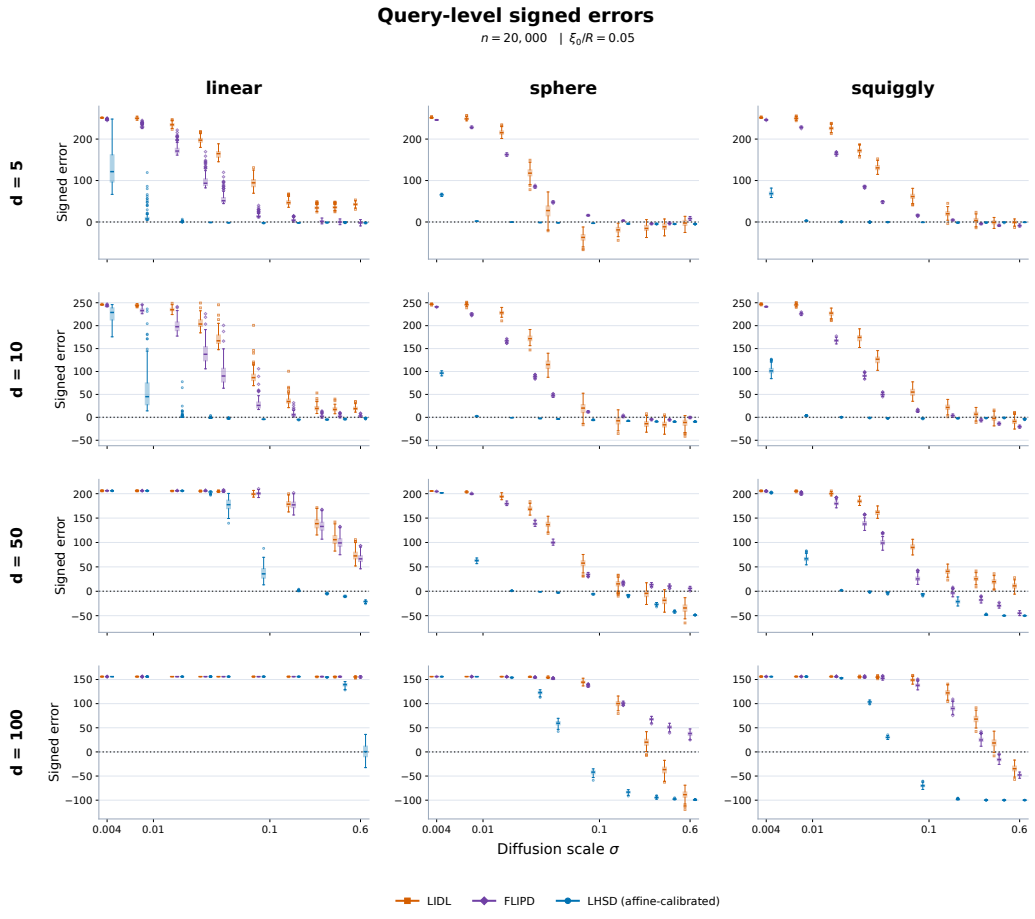


Figure 17: **Query-level signed errors with** $n = 20{,}000$**.** Distributions of $\widehat{d} - d$ over the queries, with $\xi_0/R = 0.05$ and $\varpi = 0.05$ for affine-calibrated LHSD. Positive and negative errors indicate overestimation and underestimation, respectively.

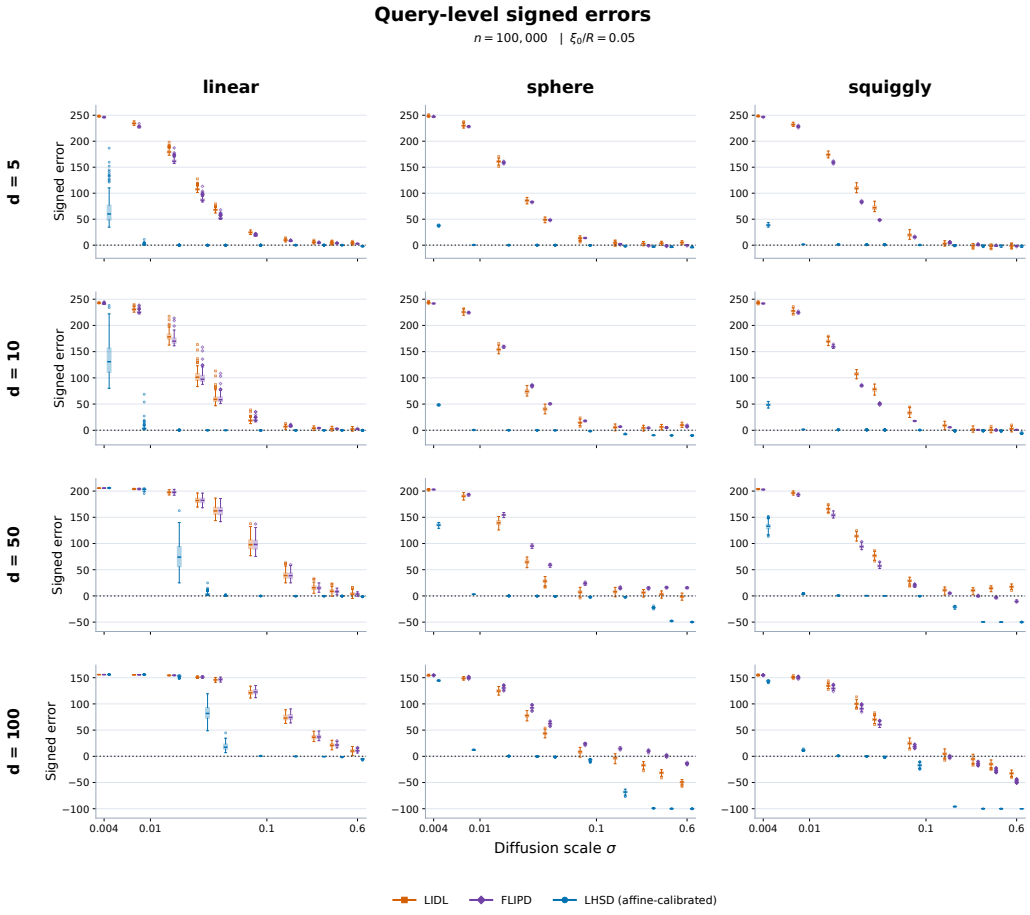


Figure 18: **Query-level signed errors with $n = 100{,}000$.** Distributions of $\widehat{d} - d$ over the queries, with $\xi_0/R = 0.05$ and $\varpi = 0.05$ for affine-calibrated LHSD.

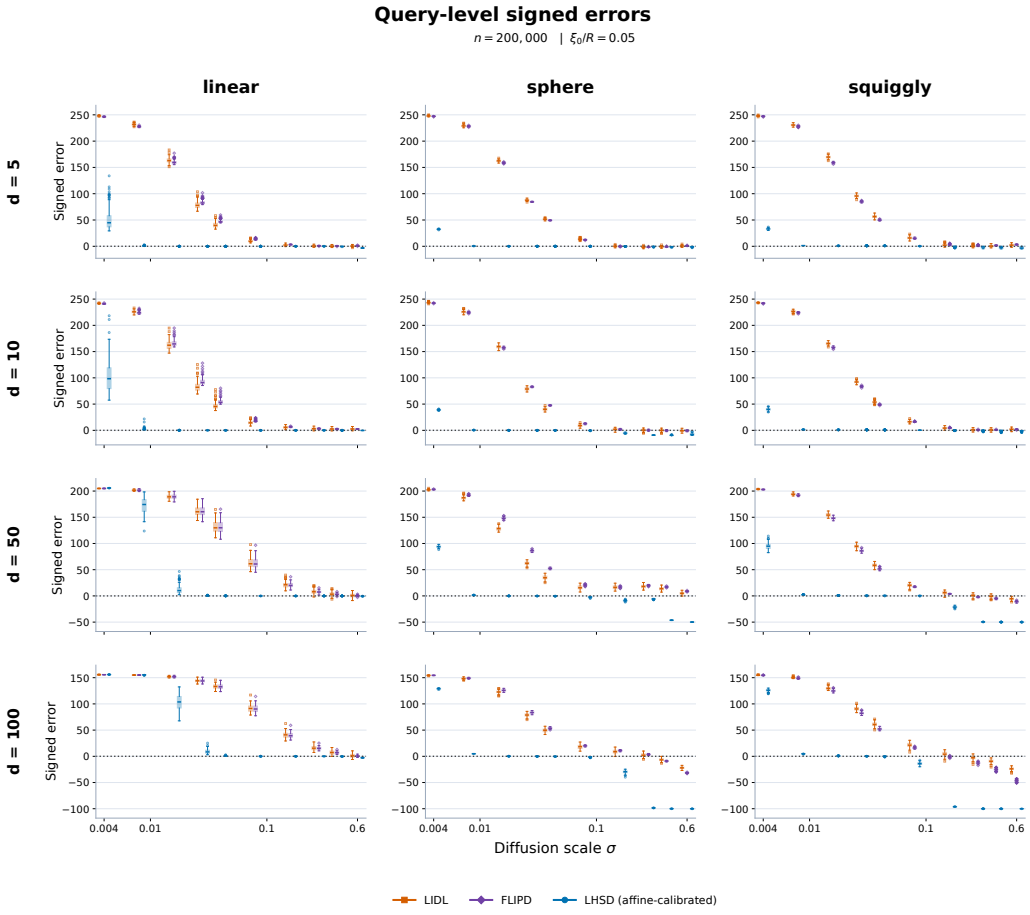


Figure 19: **Query-level signed errors with $n = 200{,}000$.** Distributions of $\widehat{d} - d$ over the queries, with $\xi_0/R = 0.05$ and $\varpi = 0.05$ for affine-calibrated LHSD.

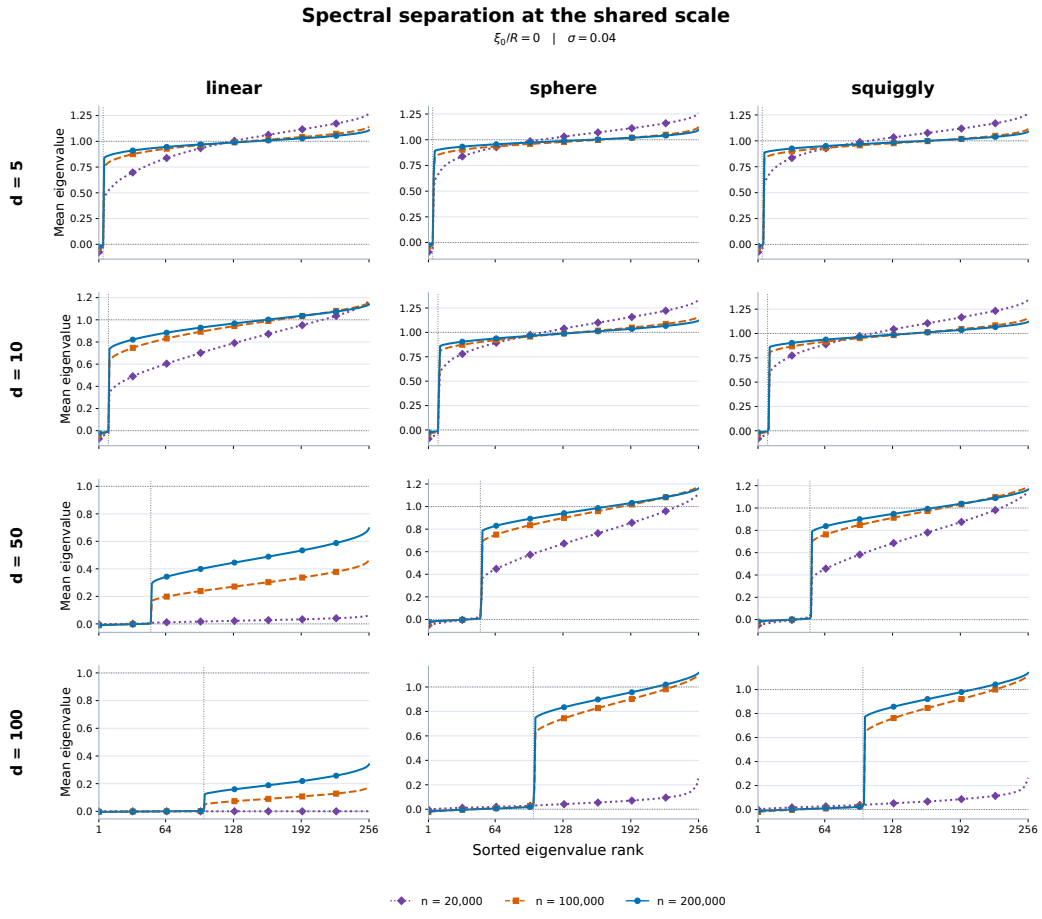


Figure 20: **Mean sorted Hessian spectra at $\xi_0/R = 0$.** Query-mean sorted eigenvalues of the learned normalized Hessian at $\sigma = 0.04$, comparing $n = 20{,}000, 100{,}000, 200{,}000$. Vertical guides mark rank d and horizontal guides mark eigenvalues zero and one.

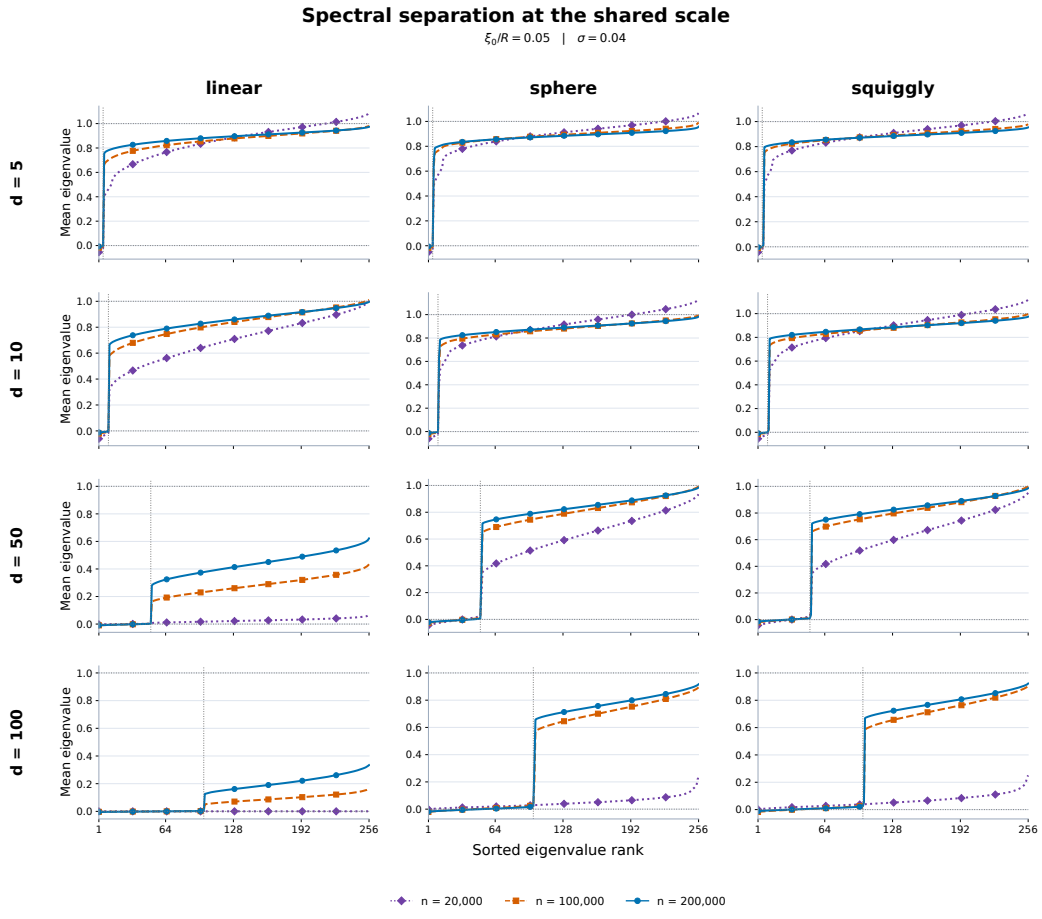


Figure 21: **Mean sorted Hessian spectra at $\xi_0/R = 0.05$.** Query-mean sorted eigenvalues of the learned normalized Hessian at $\sigma = 0.04$, comparing $n = 20{,}000, 100{,}000, 200{,}000$. Vertical guides mark rank d and horizontal guides mark eigenvalues zero and one.

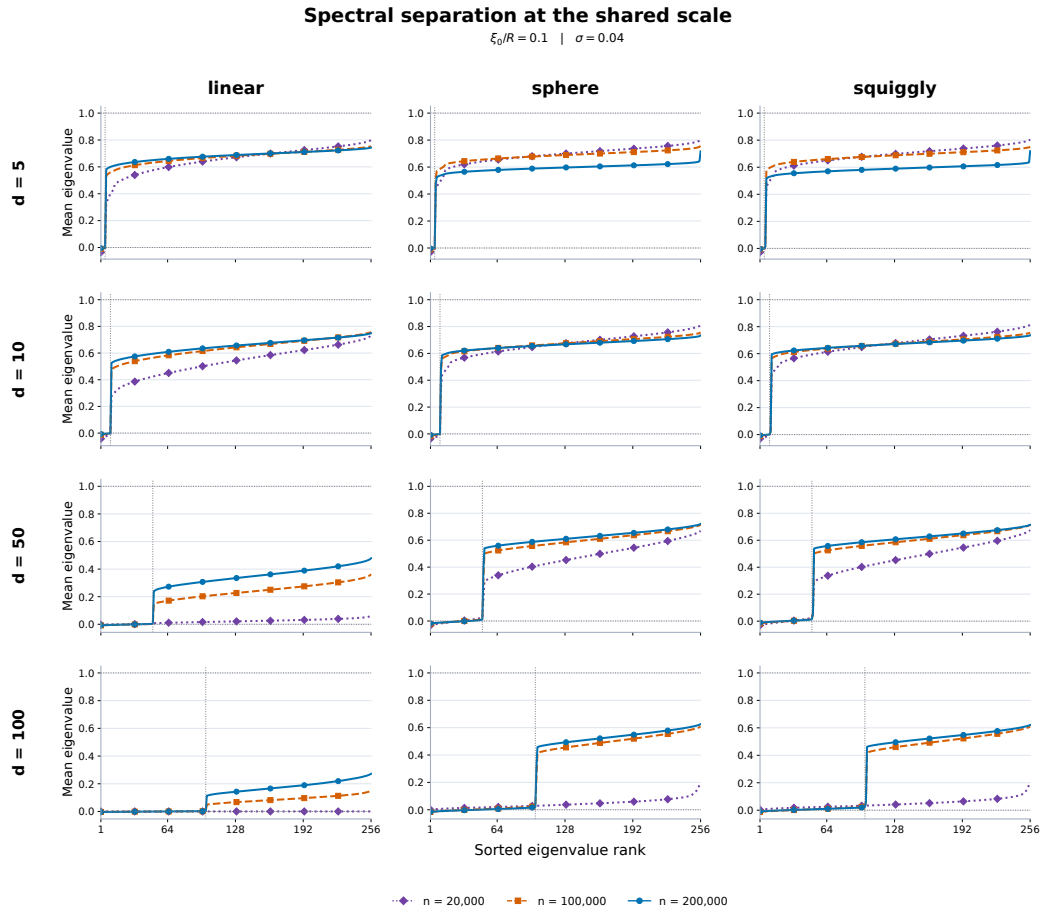


Figure 22: **Mean sorted Hessian spectra at $\xi_0/R = 0.1$.** Query-mean sorted eigenvalues of the learned normalized Hessian at $\sigma = 0.04$, comparing $n = 20{,}000, 100{,}000, 200{,}000$. Vertical guides mark rank d and horizontal guides mark eigenvalues zero and one.

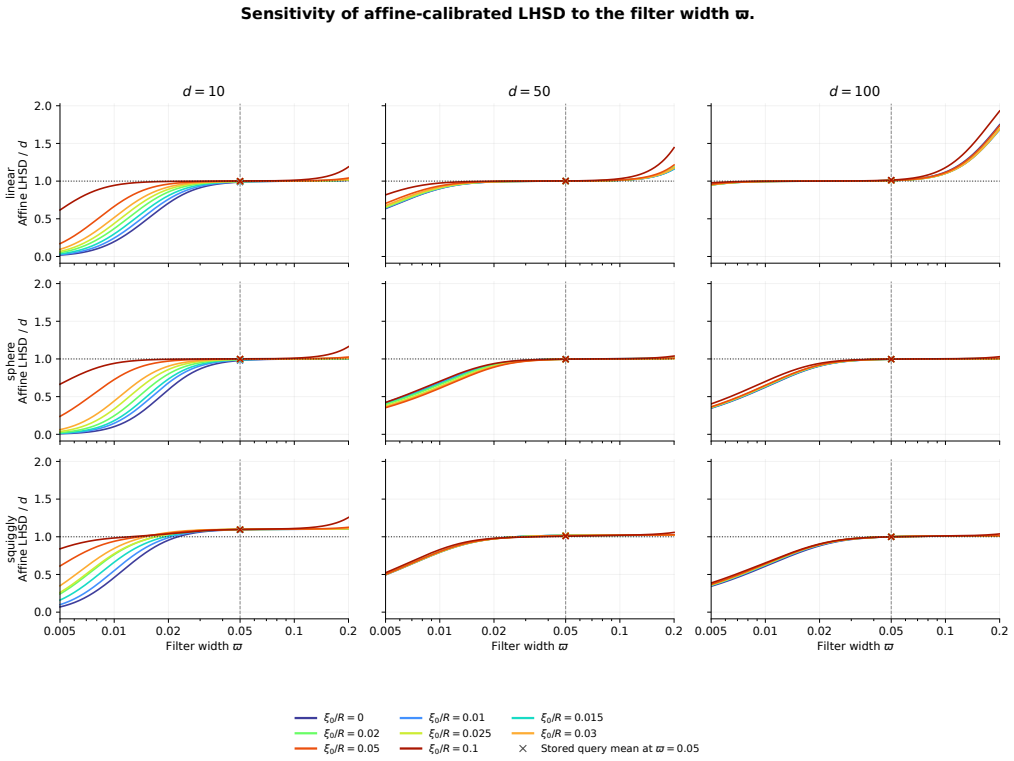


Figure 23: **Sensitivity of affine-calibrated LHSD to the filter width ϖ.** With $n = 200{,}000$ and $\sigma = 0.04$, rows indicate geometry and columns indicate $d = 10, 50, 100$. The eight colors correspond to $\xi_0/R \in \{0, 0.01, 0.015, 0.02, 0.025, 0.03, 0.05, 0.1\}$. Curves apply the affine filter to the saved mean eigenvalues and are mean-spectrum approximations; crosses mark the actual saved query means at $\varpi = 0.05$. The dotted horizontal line denotes the true dimension after normalization by d, and the dashed vertical line marks the width used in the main experiments.