YOUR GENERATIVE MODEL IS SECRETLY A DATA DIMENSIONALITY ESTIMATOR

Anonymous authors
Paper under double-blind review

ABSTRACT

Local intrinsic dimension (LID) quantifies the local degrees of freedom of the data: the independent directions in which an object can vary while remaining on the data manifold. Classical estimators infer it from nearest-neighbor statistics and can require prohibitively many samples in high dimensions. Methods such as LIDL and FLIPD instead recover it from how a generative model's density changes with the scale of added Gaussian noise. Their formulas and guarantees, however, are tied to Gaussian perturbations and to a specific model output, whereas many domains train generators with different distributions. It remains open when a model already trained for its own domain yields a justified LID estimate, and at what cost. Our main result extends this density-slope criterion beyond Gaussian noise: it gives sufficient conditions on the perturbation kernel and the local data geometry under which the logarithmic scale derivative of the perturbed density recovers LID as the noise vanishes. This derivative equals the response of the posterior-mean reconstruction of the clean point to small shifts of the noisy observation, plus a correction. The correction is required at a boundary of the support but vanishes at interior points of a smooth data manifold, where the response alone recovers LID for Gaussian and admissible non-Gaussian kernels alike. Whenever a model's output determines this reconstruction, the response is computed with a few Jacobian–vector products and no additional training. Specializing the criterion gives explicit LID estimators for diffusion, independent affine flow matching including non-Gaussian sources, generalized Brownian Schrödinger bridges and scale-conditioned normalizing flows, recovering FLIPD and Gaussian flow-matching estimators as their Gaussian special cases. On LID benchmarks with Gaussian and non-Gaussian perturbations, the response alone is competitive with the full estimate and more robust when the learned correction is inaccurate; on pretrained crystal and protein models, including ones trained with non-Gaussian noise, it tracks lattice symmetry and side-chain flexibility.

1 INTRODUCTION

The dimension of the ambient space in which an object is represented need not reflect its intrinsic complexity. An image is represented by thousands or millions of pixel values, yet its local variations may depend on far fewer factors, such as pose, lighting, or facial expression. The manifold hypothesis captures this by assuming that the data distribution is concentrated on a low-dimensional set $\mathcal{M} \subset \mathbb{R}^D$. At a regular point x, its *local intrinsic dimension* (LID) counts the independent infinitesimal directions in which the object can vary while remaining on $\mathcal{M}$. LID provides a local measure of complexity for studying representations and generalization, as well as detecting out-of-distribution observations (Kamkari et al., 2024b;a).

The data support is usually unknown, so classical estimators infer its dimension from neighborhood counts or nearest-neighbor distances (Levina & Bickel, 2004; Facco et al., 2017). Their neighborhoods must contain enough samples while remaining small enough to preserve local geometry, a difficult tradeoff in high dimensions (Tempczyk et al., 2022). One way to reduce this reliance on densely sampled neighborhoods is to learn the distribution using information from the full dataset. Generative approaches follow this route, estimating local dimension from the learned distribution.

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.

To understand these approaches, consider a clean sample $X \sim \mu$ and a noisy observation $Y = X + \varepsilon Z$, where $\varepsilon > 0$ controls the noise scale and Z is independent of X, with density K on $\mathbb{R}^D$. The conditional law of Y given X is the *perturbation kernel*, and the density ρ_ε of Y is the smoothed data density. If $Z \sim \mathcal{N}(0, I_D)$, the perturbation is Gaussian and ρ_ε is a Gaussian convolution of the data measure.

LIDL estimates dimension from how Gaussian smoothing changes the log-density with the noise scale (Tempczyk et al., 2022), and Leung et al. (2025) prove that this *density-slope criterion*, $D + \varepsilon \partial_\varepsilon \log \rho_\varepsilon(x)$, tends to d as $\varepsilon \to 0$ under regularity assumptions, for Gaussian and uniform-ball noise. FLIPD evaluates the criterion through a diffusion model's score and its divergence (Kamkari et al., 2024b), and Yeats et al. (2025) give the corresponding expression for Gaussian-source flow matching. These results establish the criterion for specific perturbation laws.

Although Gaussian perturbations work well in many applications, scientific generative modelling also involves geometries that motivate other noise laws. Molecular torsion angles are periodic, motivating wrapped-normal perturbations in Torsional Diffusion (Jing et al., 2022). Protein backbone models such as FrameDiff and RFdiffusion perturb residue orientations through Brownian motion on the rotation group $SO(3)$, alongside Gaussian translations (Yim et al., 2023; Watson et al., 2023). Crystal models such as DiffCSP and MatterGen use wrapped-normal noise for fractional atomic coordinates to respect lattice periodicity (Jiao et al., 2023; Zeni et al., 2025). These perturbations respect the geometry of their domains and, at the finite scales at which LID is evaluated, differ from additive Gaussian noise in Euclidean space; existing guarantees, stated for Gaussian or uniform-ball noise, do not cover them.

In applications, researchers often already have a generative model trained for their domain and want to estimate local data dimensionality without retraining it or fitting another model. Model families learn different quantities, such as scores, velocities (Lipman et al., 2023) or bridge drifts (De Bortoli et al., 2021; Shi et al., 2023). A useful criterion should therefore both recover LID in the small-noise limit and admit an inexpensive evaluation using quantities readily available from existing generative models. The practical question is: *when can a pretrained generative model, including one with non-Gaussian perturbations, supply a theoretically justified LID estimate without additional training, and how much computation does this require?*

We address this question by extending the density-slope criterion to a broad class of Gaussian and non-Gaussian perturbations. At a regular point of dimension d, smoothing at scale ε spreads mass over an ambient volume of order ε^D, while collecting data mass of order ε^d. The resulting density scales as ε^{d-D}. Building on this geometric principle, our master theorem establishes when its logarithmic scale derivative recovers LID beyond Gaussian smoothing (Section 2; Appendix C). The criterion describes the perturbed data law; model-specific identities then express it through the quantities a trained model provides.

The density-slope criterion decomposes into two parts, *Response* and a *correction*. Response is the Jacobian trace of the posterior mean: it measures how the conditional reconstruction of a clean sample changes when the noisy input moves. The correction is the costly part: for flow matching with a non-Gaussian source, the velocity determines the posterior mean but not the conditional mean of the kernel score that the correction requires. Single-field conversions exist for Gaussian and radial Student sources; other kernels need an additional estimator or numerical integration (Appendix K). We prove that, at regular interior points of a smooth manifold with positive continuous density, the correction vanishes as the noise scale tends to zero under the stated kernel assumptions (Corollary 2.4). Response alone therefore converges to LID. Gaussian denoiser geometry has direct precedents (Kim, 2026; Brosse & Dalalyan, 2026); our result extends this justification beyond Gaussian perturbations. The local smoothness assumption is consistent with the manifold hypothesis and empirical evidence for low-dimensional data structure (Fefferman et al., 2016; Pope et al., 2021).

This gives a *computationally efficient* estimator whenever the model supplies the posterior mean directly or through a simple conversion. At a chosen query and scale, the trace can be approximated using a small number of Hutchinson Jacobian-vector products, without additional training or integration of a sampling trajectory. In our experiments, omitting an inaccurately learned correction can also improve estimates at finite scales.

The main contributions are as follows.

- **A general kernel criterion for LID estimation.** We prove a density-slope criterion for a broad class of Gaussian and non-Gaussian perturbations arising in applications, and derive its expressions for diffusion, independent affine flow matching, Brownian Schrödinger bridges, and scale-conditioned normalizing flows, which recover FLIPD and Gaussian flow-matching estimators as special cases.
- **A computationally efficient LID estimator from an existing model.** We derive and justify a posterior-response criterion (Corollary 2.4) that extends beyond Gaussian noise and needs no auxiliary training when the model provides a compatible posterior-mean output. Benchmark experiments show that this simpler readout can retain or improve accuracy relative to the full criterion. A preliminary evaluation on real data also finds that the estimates correlate with independently computed geometric degrees of freedom.

2 A Master Theorem for Local Intrinsic Dimension

2.1 Two ways to read dimension

Let $X \sim \mu$ be data in $\mathbb{R}^D$, and fix a point x of its support with local dimension d, which describes how the probability of finding a data point in a neighborhood of x changes as its radius tends to zero. Perturb the data by independent noise with a continuously differentiable probability density K:

$$Y_\varepsilon = X + \varepsilon Z, \quad Z \perp X, \quad Z \sim K, \qquad \rho_\varepsilon = K_\varepsilon * \mu, \quad K_\varepsilon(z) = \varepsilon^{-D} K(z/\varepsilon). \tag{1}$$

Here ρ_ε is the density of the noisy observation, while ε sets the characteristic distance over which noise spreads the data. As in the introduction, the density scales as ε^{d-D} near a regular d-dimensional point; the leading factor depends on the sampling density and the kernel, while the exponent reflects the local dimension (Tempczyk et al., 2022; Kamkari et al., 2024b).

The scaling above suggests estimating $d - D$ from the relative change in density per relative change in noise width, and adding D to obtain a candidate for LID. Intuitively, fewer data directions leave more ambient directions for noise to fill, so removing that noise produces a steeper rise in density. These data directions may also be visible through reconstruction: the conditional mean $b_\varepsilon(y) = \mathbb{E}[X \mid Y_\varepsilon = y]$ estimates the clean point from the observation y. Its Jacobian $Jb_\varepsilon(y)$ describes how reconstruction responds to moving the observation; if this response is a projection onto d data directions, the Jacobian has trace d. The two candidate LID estimates are

$$\mathcal{R}^{\mathrm{dens}}_\varepsilon(y) = D + \varepsilon \partial_\varepsilon \log \rho_\varepsilon(y), \qquad \mathcal{R}^{\mathrm{resp}}_\varepsilon(y) = \operatorname{tr} Jb_\varepsilon(y). \tag{2}$$

The density-based estimate varies ε at fixed y, whereas the reconstruction-based estimate varies y at fixed ε. Both are evaluated at the data point $y = x$ as $\varepsilon \downarrow 0$.

The two estimates differ by the derivative of the log density along the reconstruction displacement $b_\varepsilon(y) - y$:

Proposition 2.1 (Decomposition of the density-based LID estimate)**.** *In* (1)*, assume finite conditional first moments and valid differentiation of the density and posterior mean under the integral. For $\varepsilon > 0$ and $\rho_\varepsilon(y) > 0$,*

$$\mathcal{R}^{\mathrm{dens}}_\varepsilon(y) = \mathcal{R}^{\mathrm{resp}}_\varepsilon(y) + \underbrace{(b_\varepsilon(y) - y)^\top \nabla \log \rho_\varepsilon(y)}_{\textit{correction}}. \tag{3}$$

It remains to find local conditions under which the density estimate tends to d and the correction vanishes; both depend on the geometry of magnified neighborhoods of x. Proofs are in Appendices B and C.

2.2 From local geometry to a class of perturbations

To analyse the readouts as noise vanishes, distances from x are measured in units of the noise scale ε. In these coordinates the kernel keeps a fixed width, while increasingly small neighborhoods of the data are enlarged to the same size. Their probability mass is multiplied by ε^{-d} to compensate for its expected decay. Assumption **(A1)** requires the resulting measures to converge to a nonzero limit, called the *tangent measure*, with finite mass in bounded regions.

For a smooth support with continuous positive sampling density, this limit is proportional to volume on a tangent plane at an interior point or a half-plane at a regular boundary. In both cases, enlarging a region about the origin by a factor a multiplies its mass by a^d. This scaling property, called *d-homogeneity*, is required by (A1) more generally.

For such a measure, the kernel's scale derivative integrates to exactly d times its mass (Lemma A.2 in Appendix A.3), so the limiting density readout is set by the tangent dimension, whatever the kernel's shape. Under (A1) and **(A2)**, which controls contributions from distant data, this gives $\mathcal{R}_\varepsilon^{\mathrm{dens}}(x) \to d$ for additive noise.

The same mechanism can apply to model perturbations that become additive locally as noise vanishes. Let $Q_\varepsilon(y \mid u)$ be the transition density from a clean point u, with observation density ρ_ε^Q and conditional-mean reconstruction B_ε^Q defining the readouts in (2). Class K requires a local kernel limit at a known noise scale ε in data coordinates, $\varepsilon^D Q_\varepsilon(x \mid x - \varepsilon z) \to K_x(z)$. Here $z = (x-u)/\varepsilon$, and K_x may depend on x. Because the readouts use derivatives, class K also requires the corresponding scale and observation derivatives to approach those of the local additive model. Definition A.1 specifies the weighted convergence and regularity conditions, including (A2).

2.3 The master theorem and its consequences

Theorem 2.2 (Master theorem for LID recovery in class K)**.** *Suppose (A1) holds at x and Q_ε belongs to class K there. Then, as $\varepsilon \downarrow 0$, for some $0 < C_x < \infty$,*

$$\rho_\varepsilon^Q(x) = \varepsilon^{d-D}(C_x + o(1)), \qquad \mathcal{R}_\varepsilon^{\mathrm{dens}}(x) \longrightarrow d. \tag{4}$$

*If reconstruction also satisfies the moment and differentiation conditions **(A3)** of Appendix A, its response and correction have finite limits, and*

$$\mathcal{R}_\varepsilon^{\mathrm{resp}}(x) + (B_\varepsilon^Q(x) - x)^\top \nabla \log \rho_\varepsilon^Q(x) \longrightarrow d. \tag{5}$$

Applying this sufficient criterion to a model requires identifying its physical noise width, verifying class K, and expressing the density or reconstruction through the available model output.

Corollary 2.3 (Tangent recovery)**.** *Under the theorem's assumptions and (A3), suppose the tangent measure is a positive constant multiple of d-dimensional volume on a plane T. Then $JB_\varepsilon^Q(x)$ converges to a rank-d projector onto T, and $\mathcal{R}_\varepsilon^{\mathrm{resp}}(x) \to d$. For a radially symmetric K_x, the projector is orthogonal.*

Corollary 2.4 (Response suffices in the interior)**.** *Suppose that near x the data have a continuous positive density on a C^1 embedded d-dimensional manifold without boundary, and that Q_ε belongs to class K and satisfies (A3) at x. Then the correction in* (5) *vanishes as $\varepsilon \downarrow 0$ and $\mathcal{R}_\varepsilon^{\mathrm{resp}}(x) \to d$, for Gaussian and admissible non-Gaussian kernels alike. A model that provides the posterior mean therefore needs no score estimate for this limit.*

The manifold assumption makes the tangent measure flat (Appendix D), so Corollary 2.3 applies; the full proof is in Appendix M.1.

The interior assumption matters. At a boundary, the absence of clean points on one side changes reconstruction's response while leaving the dimension unchanged: for a smooth boundary with Gaussian noise and the second-moment control of Corollary D.1, the response converges to $d - 2/\pi$, while the correction restores d (Appendix D.5).

3 Explicit LID Readouts for Generative Models

Table 1 turns the master theorem into formulas for the outputs that common generative models already provide; apart from the Full readout of general-kernel flow matching, which needs the kernel score s_t, every row uses outputs the model already provides. Each row lists the physical noise scale and the query at which the field is evaluated, and each model family has an appendix that derives its formula and states the conditions under which it converges to d.

Table 1: Explicit LID formulas for exact model outputs. The additive rows equal the density readout at finite noise. The bridge rows recover terminal dimension under the stated potential conditions; the current-velocity formula is asymptotically equivalent. The NF row requires a calibrated class-K density path. Notation: $s_t = \nabla_y \log p_t$ is the score and $v_t(y) = \mathbb{E}[\dot{Y}_t \mid Y_t = y]$ the flow-matching velocity. Diffusion uses $Y_t = \psi_t X + \sigma_t Z$ and affine FM $Y_t = \alpha_t X + \beta_t Z$, with $Z \perp X$ Gaussian except in the general-kernel row; $a_t = \dot{\alpha}_t/\alpha_t$ and $\kappa_t = \dot{\varepsilon}(t)/\varepsilon(t) \neq 0$. For the bridge, γ is the diffusivity, b_t^+ the forward drift and v_t the current velocity, all derived from the bridge potentials; for the NF, $p_\varepsilon = (F_\varepsilon)_\# \pi$, $u_\varepsilon = (\varepsilon \partial_\varepsilon F_\varepsilon) \circ F_\varepsilon^{-1}$ and $s_\varepsilon = \nabla \log p_\varepsilon$. Spatial derivatives hold model time fixed.

Model family	Output and calibration	LID readout	Proof
Diffusion (FLIPD)	s_t $\varepsilon = \sigma_t/\psi_t$ $y = \psi_t x$	$D + \sigma_t^2\big(\operatorname{div} s_t(y) + \|s_t(y)\|^2\big)$	App. E
Independent affine FM General kernel K	v_t, s_t $\varepsilon = \beta_t/\alpha_t$ $y = \alpha_t x$	$D + \dfrac{Da_t - \operatorname{div} v_t(y)}{\kappa_t} + \dfrac{(a_t y - v_t(y))^\top s_t(y)}{\kappa_t}$	App. F
Independent affine FM Gaussian K	v_t $\varepsilon = \beta_t/\alpha_t$ $y = \alpha_t x$	$D + \dfrac{Da_t - \operatorname{div} v_t(y)}{\kappa_t} + \dfrac{\|v_t(y) - a_t y\|^2}{\beta_t^2 \kappa_t^2}$	App. F.3
Rectified flow Gaussian source $Y_t = tX + (1-t)Z$	v_t $\varepsilon = (1-t)/t$ $y = tx, t \uparrow 1$	$t\big[D + (1-t)\operatorname{div} v_t(tx)\big] + t^2 \|v_t(tx) - x\|^2$	App. F.4
Generalized Brownian bridge Forward drift	b_t^+ $\varepsilon = \sqrt{\gamma\tau}$ $y = x, \tau = T - t$	$D + \tau \operatorname{div} b_t^+(x) + \dfrac{\tau}{\gamma}\left\|b_t^+(x)\right\|^2$	App. G
Same bridge Current velocity	v_t, s_t $\varepsilon = \sqrt{\gamma\tau}$ $y = x, \tau = T - t$	$D + 2\tau\big(\operatorname{div} v_t(x) + v_t(x)^\top s_t(x)\big)$	App. G.5
Scale-conditioned normalizing flow	$\log p_\varepsilon$ Calibrated ε Fixed query x	$D + \varepsilon \partial_\varepsilon \log p_\varepsilon(x)$ $= D - \operatorname{div} u_\varepsilon(x) - u_\varepsilon(x)^\top s_\varepsilon(x)$	App. H

The diffusion row recovers FLIPD (Kamkari et al., 2024b), the Gaussian FM row agrees with its velocity conversion (Yeats et al., 2025), and the NF density row is the differential form of LIDL (Tempczyk et al., 2022).

At flat-tangent points, the diffusion, affine-FM and forward-drift bridge formulas also have response-only versions whose corrections vanish. In the general-kernel row, the first two terms form Response and use only v_t; the last term is the correction and also needs the score s_t, which a flow-matching model does not provide for a general non-Gaussian source; radial Student sources admit a single-field closure (Appendix K). By Corollary 2.4, the correction vanishes at interior points, so the velocity alone suffices there. Derivative accuracy and trace estimation for learned outputs are treated in Appendix I.

4 Experiments

We evaluate endpoint LID readouts on learned generative fields, comparing model families and operating-scale selection.

4.1 LID readouts across generative-model families

We evaluate the theoretical readouts on LID-Benchmarks (Tempczyk et al., 2026). Its known-dimension tasks comprise seven geometries, represented both as coefficient vectors and as 28×28 images reconstructed using a Fashion-MNIST PCA basis (IDR). We also report the Arrows factor-count task. Fashion-MNIST training-size sensitivity (E1), paired transformations (E5), and additional parameterization controls are retained in Appendix U.

The Gaussian comparison includes VP/VE diffusion with FLIPD (Kamkari et al., 2024b), rectified and trigonometric flows with the equivalent FM-FLIPD readout (Leung et al., 2025; Yeats et al., 2025), and scale-conditioned RealNVP trained on Gaussian-smoothed data, whose density slope is fit over five scales (OLS5). We also evaluate radial Student t-Flow, product-logistic FM, and Brownian DSB (De Bortoli et al., 2021). The first two change the independent source noise; DSB introduces a dependent endpoint coupling. DSB still has a Gaussian Brownian reference, but dependence can make the conditional source law $p(Z \mid X)$ non-Gaussian. For a logistic source, Full (the density readout $\mathcal{R}_\varepsilon^{\mathrm{dens}}$, i.e. Response plus correction) needs the conditional mean of the kernel score, which the velocity does not determine; to measure Full itself we train a second network for this quantity (Appendix K). Response uses the velocity alone. Update budgets and backbones are matched within each representation; logistic FM and DSB train two field cores (Appendix T.2).

For known LID, a disjoint source-train holdout selects each checkpoint's scale by primary-readout MAE. Pointwise Kneedle (Satopää et al., 2011) uses the same checkpoints and candidate grids without labels. The VP, rectified-flow, logistic-FM and t-Flow Response readouts inherit Full's selected scales in both protocols. Coef. and IDR are suite-balanced relative-MAE aggregates; Arrows reports raw MAE. Evaluation uses one training seed (Appendix T.5).

Table 2: Known-LID test scores under supervised holdout selection and pointwise Kneedle; lower is better, with column minima in bold. Groups distinguish independent Gaussian perturbations from non-Gaussian sources or dependent bridge couplings. A constant estimate $\widehat{d} \equiv 1$ scores 0.485 on Coef. and IDR and 14.029 on Arrows (Appendix T.5).

	Supervised holdout selection			Pointwise Kneedle		
Model/readout	Coef. ↓	IDR ↓	Arrows ↓	Coef. ↓	IDR ↓	Arrows ↓
Gaussian perturbations						
VP diffusion + FLIPD (full)	0.204	0.383	3.762	6.894	9.329	258.662
VP diffusion (response)	0.227	0.412	6.425	6.968	**5.921**	**201.689**
VE diffusion + FLIPD (full)	0.181	1.358	123.719	0.242	11.939	620.509
Rectified flow (full)	**0.095**	0.249	**2.787**	0.213	12.116	360.918
Rectified flow (response)	0.129	**0.239**	8.411	0.231	9.195	276.386
Trigonometric flow (full)	0.206	0.268	3.511	13.847	10.141	214.287
Scale-NF (OLS5)	0.163	0.521	5.888	0.249	242.501	3058.117
Beyond Gaussian perturbations						
Brownian DSB (full)	2.833	56.150	1538.797	17.252	69.732	2815.689
t-Flow (full)	0.168	13.169	1053.584	0.232	156.694	3051.534
t-Flow (response)	0.164	1.975	244.345	**0.175**	106.305	2909.186
Logistic FM (full)	0.118	0.253	4.238	0.334	13.688	355.665
Logistic FM (response)	0.109	0.321	6.311	0.331	12.695	339.851

Gaussian models. Dropping the correction costs some accuracy under supervised selection, but not uniformly: Rectified flow (response) is more accurate than Rectified flow (full) on IDR. Under supervised selection, rectified and trigonometric flow remain competitive with the diffusion FLIPD baselines, and Scale-NF does so on Coef. and Arrows.

Non-Gaussian source with a learned correction. Under supervised selection, Logistic FM (full) is on par with the Gaussian baselines. Logistic FM (response), computed from the velocity alone, is more accurate on Coef. and less accurate on IDR and Arrows; under pointwise Kneedle it is more accurate on all three.

Failures. t-Flow (full) breaks down on IDR and Arrows, while t-Flow (response) is far more accurate under supervised selection; Section 4.2 traces this to the learned posterior mean. Brownian DSB (full) fails through learned-field and discretization errors, compounded by the finite-scale bias of the Schrödinger-weighted posterior (Appendix Q).

Pointwise Kneedle reduces accuracy across all displayed families, particularly on IDR and Arrows, showing that the supervised results do not transfer directly to this automatic selector. On IDR and Arrows every Kneedle entry is worse than the constant estimate, so these columns do not rank the readouts; label-free scale selection remains open.

4.2 Response versus Full readout

Takeaway. The correction is needed only where the posterior is asymmetric around the query, as at boundaries and creases, and it helps only if the model's posterior mean is accurate there; otherwise it adds error, and Response is the safer readout.

We observe in Table 2 that the denoiser-Jacobian trace (Response) can outperform the density-slope estimator (Full), defined in (2), despite omitting the correction in (3). This occurs for rectified flow on rendered data and more markedly for radial Student t-Flow. For Student, replacing only the predicted posterior mean with a continuous-reference value removes most of the Full error on Spiral and Funnel (Appendix P).

Corollary 2.4 establishes when Response alone suffices. Boundaries and asymmetric creases, however, can make the posterior asymmetric around the query even with Gaussian noise, producing a correction that remains nonzero as the noise scale vanishes. Both geometries shift the posterior mean, as shown by the orange arrows in Figure 1. At the illustrated boundary and crease, contraction of some axes of the purple ellipsoids reduces the posterior variance, so Response underestimates LID in the depicted limits. The squared posterior-mean displacement supplies the missing part of the posterior second moment: the correction restores the dimension and is necessary for consistency at these asymmetric points. We derive these geometric limits in Appendix M and examine the Student component substitutions and posterior-error profiles in Appendix P.

4.3 Response on pretrained scientific models

We return to the question from the introduction: can a model trained for its own domain, including one with non-Gaussian perturbations, provide useful local geometry through Response, using released weights without retraining?

Crystals. DiffCSP (Jiao et al., 2023) and OMatG (Höllmer et al., 2025) predict crystal structures conditioned on composition. We evaluate their lattice Response on all 9,046 MP-20 test crystals (Xie et al., 2022), against the symmetry-derived number of lattice degrees of freedom, including rotation. The lattice block has the same nine-dimensional ambient space for every crystal. DiffCSP uses Gaussian lattice noise; OMatG-VPSBD instead interpolates from a non-Gaussian source with lognormal lengths and uniformly sampled angles.

Proteins. FlowPacker (Lee & Kim, 2025) predicts side-chain conformations using torsional flow matching with uniform angular sources. We evaluate per-residue Response on 39 recent PDB chains (Berman et al., 2000) and 20 ATLAS chains with molecular-dynamics (MD) trajectories (Vander Meersche et al., 2024). Independent references are relative solvent accessibility (rSASA), side-chain B-factors and the circular variance of torsion angles in MD. Correlations use ranks within each residue type, keeping the number of angles fixed. Proline provides a complementary test: its ring constrains two torsions, so we measure how often its Response is lower than that of other two-torsion residues (Proline AUC).

Crystal scales are selected on one half of the test structures and evaluated on the other, in both directions. FlowPacker's scale is selected from Response curves without consulting the proxies. Full protocols, confidence intervals, random-weight controls and additional OMatG variants are in Appendix S. OMatG-VPSBD has the lowest lattice MAE of the four evaluated OMatG checkpoints. Some native kernels, such as FlowPacker's uniform arcs, are not C^1, so Theorem 2.2 does not cover them and these results are empirical associations.

Table 3 shows that Response tracks lattice symmetry and side-chain flexibility, including MD variation beyond solvent accessibility. The random-weight controls have much weaker associations. The protein measurements establish a useful finite-scale flexibility signal; they do not calibrate Response as an exact count of dimensions.

5 Conclusion

We establish a common criterion for recovering LID from the scale dependence of a perturbed data law, covering admissible non-Gaussian kernels under explicit geometric, regularity and tail assumptions. The resulting theory recovers FLIPD, places existing Gaussian FM readouts within the

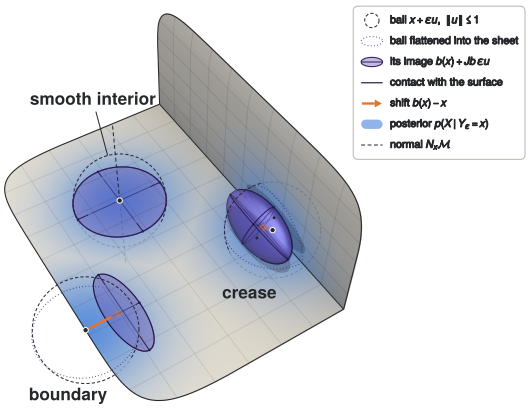

Figure 1: Gaussian posterior geometry on a two-dimensional folded sheet. The dashed ball represents small query perturbations; purple shows their linearized image under the posterior mean b_ε, whose normalized semiaxes sum to Response. Orange arrows show $b_\varepsilon(x) - x$, giving correction $\|b_\varepsilon(x) - x\|^2 / \varepsilon^2$; blue shades indicate posterior mass on the sheet. Response recovers both dimensions in the interior; at the edge and crease, the correction supplies its deficit.

Crystal lattice ($n = 9{,}046$)		
Model	MAE ↓	ρ ↑
DiffCSP	0.633	0.755
OMatG-VPSBD	0.826	0.677

FlowPacker		
Reference	PDB	ATLAS
rSASA (ρ)	0.418	0.413
B-factor (ρ)	0.422	0.384
MD variance (ρ)	—	0.330
MD | rSASA (ρ)	—	0.183
Proline AUC	0.781	0.810

Table 3: Response on released models. Crystal MAE is in dimensions; ρ denotes rank correlation (within residue type for proteins). MD | rSASA controls for solvent accessibility. Higher correlations and AUC indicate stronger agreement with the references. Protein sets contain 7,588 / 2,407 residues.

same construction, and identifies the normalizations and corrections that distinguish them. It also explains when posterior response alone recovers dimension and when asymmetric local geometry requires the Full correction.

Experiments support this beyond Gaussian perturbations: Response is comparable to Full on LID benchmarks, including a logistic-source flow, and tracks symmetry and flexibility references on released crystal and protein models. The t-Flow and DSB failures locate the remaining gap in learned posterior means and Jacobians; label-free choice of the operating scale remains open.

6 Related Work

The density-based estimate in (2) builds on the idea that adding noise makes local dimension visible through an ordinary ambient density. LIDL (Tempczyk et al., 2022) establishes the Gaussian scaling behind this idea and turns it into an estimator: fit density models at several noise levels, regress log density on log noise standard deviation, and add the ambient dimension to the fitted slope. FLIPD (Kamkari et al., 2024b) obtains the corresponding derivative from the score and its divergence, using a single diffusion model. Thus, recovering dimension from noise-dependent density, and doing so with a pretrained Gaussian model, are established results.

These constructions require a guarantee that removing the noise recovers the dimension of the original data. Leung et al. (2025) prove that the density-based estimate in (2) converges to manifold dimension under regularity and moment conditions, and establish a corresponding result for uniform-ball smoothing. Tempczyk et al. (2025) explain the common Gaussian mechanism of several generative LID estimators through the Wiener process and quantify how nonuniform sampling density

affects estimates at finite noise. Our Theorem 2.2 gives sufficient conditions on a class of perturbation kernels and local transition laws, allowing the density criterion to extend beyond these specified smoothers.

To relate this density criterion to Response, one needs the geometry of reconstruction. Alain & Bengio (2014) connect regularized reconstruction to derivatives of log density; Manor & Michaeli (2024) recover Gaussian posterior covariance and higher moments from derivatives of the denoiser. In particular, the Gaussian posterior mean satisfies $b_\varepsilon(y) = y + \varepsilon^2 s_\varepsilon(y)$, where $s_\varepsilon = \nabla \log \rho_\varepsilon$. Substituting it into FLIPD splits the density readout into Response $D + \varepsilon^2 \operatorname{div} s_\varepsilon$ and the correction $\varepsilon^2 \|s_\varepsilon\|^2$, the squared reconstruction displacement divided by ε^2 (Appendix J); this Gaussian split is already contained in FLIPD. Our decomposition (3) extends the relationship between density change, reconstruction sensitivity and displacement to general additive noise.

There are also direct precedents for recovering dimension from Gaussian reconstruction alone. Kim (2026) defines a dimension diagnostic using the denoiser Jacobian trace and its posterior-covariance interpretation. Wan et al. (2025) conjecture convergence of the denoiser Jacobian to the tangent projector, while Brosse & Dalalyan (2026) establish Gaussian small-noise expansions that give this projector limit in the smooth interior and describe its modification near boundaries and corners. The projector preserves tangent directions and removes normal ones, so its trace counts the local dimension. Our Corollary 2.4 extends this conclusion to admissible non-Gaussian kernels: at regular interior points, Response has the same LID limit as Full without estimating the correction.

Whether this cheaper estimate is available from a trained generator depends on what the model predicts. For Gaussian affine flow matching, Leung et al. (2025) already note that velocity can be converted to a score and used in FLIPD. Yeats et al. (2025) evaluate Gaussian denoising-error estimates and FLIPD on rectified flow, including Stable Diffusion 3.5. Their implicit-score-matching split of FLIPD is not Response: its score-matching part already contains part of our correction. These results already establish LID estimation from Gaussian FM velocity fields.

Prediction errors give related Gaussian quantities: divided by $\beta_t^2 \kappa_t^2$, the irreducible flow-matching error of Cha et al. (2026) equals Response; Wu & Verdú (2011) define dimension through averaged optimal reconstruction error, and Raju & Zhang (2026) attribute training-loss floors to conditional target variance. These identities do not supply our non-Gaussian Response limit.

Moving beyond Gaussian noise matters when the available generator no longer admits the same velocity-to-score conversion. For example, Pandey et al. (2024) introduce Student t-EDM and t-Flow using heavy-tailed perturbations. Our independent affine-FM formula obtains Response from velocity alone for admissible non-Gaussian sources, without fitting an additional score model. The same separation between an existing model construction and its LID justification applies to our other interfaces. Pooladian & Niles-Weed (2025) express bridge drift through a conditional terminal mean, and SoftFlow (Kim et al., 2020) conditions a normalizing-flow density on noise scale. We use these constructions to derive dimension-recovery statements; the generators and conditional-mean identities themselves are prior work.

Other geometric estimators count directions with different matrices, such as nearby score vectors (Stanczuk et al., 2024), generative transport Jacobians (Horvat & Pfister, 2022; 2024) or energy Hessians (Balcerak et al., 2025), so their finite-noise estimates need not equal the posterior Jacobian trace. Classical estimators (Levina & Bickel, 2004; Amsaleg et al., 2015; Facco et al., 2017; Ceruti et al., 2014; Gomtsyan et al., 2019) serve as independent comparators in the benchmark of Tempczyk et al. (2026); extending the theoretical scope alone does not imply an accuracy advantage.

AI USE STATEMENT

In this work, we used generative AI tools, namely the LLM coding agents OpenAI Codex and Anthropic Claude Code, to help develop and organize the theoretical framework, refine the formulation of mathematical claims and their assumptions, provide ingredients for proofs and assist in writing them, refine hypotheses about the failure modes observed in the experiments, design and review parts of the experimental protocol, implement methods, interpret results, and make the text clearer and easier to understand. We have not used generative AI tools to generate synthetic data sets or to clean and reformat data sets, and qualitative or thematic data analysis is not applicable to this work. Additionally, we used generative AI tools to create and edit software code, suggest experimental

parameters, produce figures and tables from recorded results, search for, identify and summarize related literature, format references, suggest a structure for the paper, and draft and edit its text for readability.

We have reviewed all AI-assisted work. The authors checked every theorem statement and proof; AI-drafted derivations were additionally cross-examined by independent AI sessions, and all text was checked against a written specification of the claims our theorems support. The implementation is covered by unit tests, and the benchmark metrics were recomputed from saved predictions (Appendix R). Every reference was checked against its primary source. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Guillaume Alain and Yoshua Bengio. What regularized auto-encoders learn from the data-generating distribution. *Journal of Machine Learning Research*, 15(110):3743–3773, 2014. URL https://jmlr.org/papers/v15/alain14a.html.

Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025. URL https://www.jmlr.org/papers/v26/23-1605.html.

Laurent Amsaleg, Oussama Chelly, Teddy Furon, Stéphane Girard, Michael E. Houle, Ken-ichi Kawarabayashi, and Michael Nett. Estimating local intrinsic dimensionality. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 29–38, 2015. doi: 10.1145/2783258.2783405.

Michal Balcerak, Tamaz Amiranashvili, Antonio Terpin, Suprosanna Shit, Lea Bogensperger, Sebastian Kaltenbach, Petros Koumoutsakos, and Bjoern Menze. Energy matching: Unifying flow matching and energy-based models for generative modeling. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-0291. URL https://papers.neurips.cc/paper_files/paper/2025/hash/0cbbdfb0a4098af8dc7a497a5e59aff7-Abstract-Conference.html.

Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000. doi: 10.1093/nar/28.1.235. URL https://doi.org/10.1093/nar/28.1.235.

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

Carlo Ceruti, Simone Bassis, Alessandro Rozza, Gabriele Lombardi, Elena Casiraghi, and Paola Campadelli. DANCo: Dimensionality from angle and norm concentration. *Pattern Recognition*, 47(8):2569–2581, 2014. doi: 10.1016/j.patcog.2014.02.013.

Yeonwoo Cha, Jaehoon Yoo, Semin Kim, Yunseo Park, Jinhyeon Kwon, and Seunghoon Hong. Training-free refinement of flow matching with divergence-based sampling, 2026. URL https://arxiv.org/abs/2604.04646.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34:17695–17709, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/940392f5f32a7ade1cc201767cf83e31-Abstract.html.

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

Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016. doi: 10.1090/jams/852. URL https://mitter.lids.mit.edu/publications/121_Testing_Manifold.pdf.

Marina Gomtsyan, Nikita Mokrov, Maxim Panov, and Yury Yanovich. Geometry-aware maximum likelihood estimation of intrinsic dimension. In *Proceedings of the Eleventh Asian Conference on Machine Learning*, volume 101 of *Proceedings of Machine Learning Research*, pp. 1126–1141. PMLR, 2019. URL https://proceedings.mlr.press/v101/gomtsyan19a.html.

Philipp Höllmer, Thomas Egg, Maya Martirossyan, Eric Fuemmeler, Zeren Shui, Amit Gupta, Pawan Prakash, Adrian Roitberg, Mingjie Liu, George Karypis, Mark Transtrum, Richard Hennig, Ellad B. Tadmor, and Stefano Martiniani. Open materials generation with stochastic interpolants. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 23417–23450, 2025. URL https://proceedings.mlr.press/v267/hollmer25a.html.

Christian Horvat and Jean-Pascal Pfister. Intrinsic dimensionality estimation using normalizing flows. In *Advances in Neural Information Processing Systems*, volume 35, pp. 12225–12236, 2022. doi: 10.52202/068431-0888. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/4f918fa3a7c38b2d9b8b484bcc433334-Abstract-Conference.html.

Christian Horvat and Jean-Pascal Pfister. On gauge freedom, conservativity and intrinsic dimensionality estimation in diffusion models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/d553f0e0abb80e2a60328d634583bd2e-Abstract-Conference.html.

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://arxiv.org/abs/2309.04475.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/994545b2308bbbbc97e3e687ea9e464f-Paper-Conference.pdf.

Hamidreza Kamkari, Brendan Leigh Ross, Jesse C. Cresswell, Anthony L. Caterini, Rahul Krishnan, and Gabriel Loaiza-Ganem. A geometric explanation of the likelihood OOD detection paradox. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 22908–22935. PMLR, 2024a. URL https://proceedings.mlr.press/v235/kamkari24a.html.

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, 2024b. doi: 10.52202/079017-1211.

Hyeongju Kim, Hyeonseung Lee, Woo Hyun Kang, Joun Yeop Lee, and Nam Soo Kim. SoftFlow: Probabilistic framework for normalizing flow on manifolds. In *Advances in Neural Information Processing Systems*, volume 33, pp. 16388–16397, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html.

Munsik Kim. Noise-resolved effective rank: a finite-sample diagnostic for generalization and memorization in diffusion and flow-matching models, June 2026. URL https://zenodo.org/records/20775709. Preprint, version 1.

Jin Sub Lee and Philip M. Kim. FlowPacker: protein side-chain packing with torsional flow matching. *Bioinformatics*, 41(3):btaf010, 2025. doi: 10.1093/bioinformatics/btaf010. URL https://doi.org/10.1093/bioinformatics/btaf010.

Kin Kwan Leung, Rasa Hosseinzadeh, and Gabriel Loaiza-Ganem. On convolutions, intrinsic dimension, and diffusion models. *Transactions on Machine Learning Research*, 2025. URL https://openreview.net/forum?id=xSzBf1te4s. Expert Certification.

Elizaveta Levina and Peter J. Bickel. Maximum likelihood estimation of intrinsic dimension. In *Advances in Neural Information Processing Systems*, volume 17, pp. 777–784, 2004. URL https://papers.nips.cc/paper_files/paper/2004/hash/74934548253bcab8490ebd74afed7031-Abstract.html.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Hila Manor and Tomer Michaeli. On the posterior distribution in denoising: Application to uncertainty quantification. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=adSGeugiuj.

Kushagra Pandey, Jaideep Pathak, Yilun Xu, Stephan Mandt, Michael Pritchard, Arash Vahdat, and Morteza Mardani. Heavy-tailed diffusion models, 2024. URL https://arxiv.org/abs/2410.14171. Version 2.

Aram-Alexandre Pooladian and Jonathan Niles-Weed. Plug-in estimation of Schrödinger bridges. *SIAM Journal on Mathematics of Data Science*, 7(3):1315–1336, 2025. doi: 10.1137/24M1687340.

Phillip 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. URL https://openreview.net/forum?id=XJk19XzGq2J.

Avinash Raju and Kai Zhang. The loss floor of denoising score matching: Fisher geometry from Schrödinger bridges, 2026. URL https://arxiv.org/abs/2608.23916. Preprint, version 1.

Ville Satopää, Jeannie Albrecht, David Irwin, and Barath Raghavan. Finding a "Kneedle" in a haystack: Detecting knee points in system behavior. In *31st International Conference on Distributed Computing Systems Workshops*, pp. 166–171, 2011.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion Schrödinger bridge matching. In *Advances in Neural Information Processing Systems*, volume 36, pp. 62183–62223, 2023. doi: 10.52202/075280-2717. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/c428adf74782c2092d254329b6b02482-Abstract-Conference.html.

Leon Simon. Introduction to geometric measure theory, 2018. URL https://math.stanford.edu/~lms/ntu-gmt-text.pdf. Lecture notes, 2018 revision. Stanford University.

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

Piotr Tempczyk, Rafał Michaluk, Łukasz Garncarek, Przemysław Spurek, Jacek Tabor, and Adam Golinski. LIDL: Local intrinsic dimension estimation using approximate likelihood. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 21205–21231. PMLR, 2022. URL https://proceedings.mlr.press/v162/tempczyk22a.html.

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. doi: 10.1609/aaai.v39i19.34299. URL https://ojs.aaai.org/index.php/AAAI/article/view/34299.

Piotr Tempczyk, Dominik Filipiak, Łukasz Garncarek, Ksawery Smoczyński, and Adam Kurpisz. Why we need new benchmarks for local intrinsic dimension estimation. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=ZEf03Uunvk.

Matthew Z. Tien, Austin G. Meyer, Dariya K. Sydykova, Stephanie J. Spielman, and Claus O. Wilke. Maximum allowed solvent accessibilites of residues in proteins. *PLOS ONE*, 8(11):e80635, 2013. doi: 10.1371/journal.pone.0080635. URL https://doi.org/10.1371/journal.pone.0080635.

Yann Vander Meersche, Gabriel Cretin, Aria Gheeraert, Jean-Christophe Gelly, and Tatiana Galochkina. ATLAS: protein flexibility description from atomistic molecular dynamics simulations. *Nucleic Acids Research*, 52(D1):D384–D392, 2024. doi: 10.1093/nar/gkad1084. URL https://doi.org/10.1093/nar/gkad1084.

Zhengchao Wan, Qingsong Wang, Gal Mishne, and Yusu Wang. Elucidating flow matching ODE dynamics via data geometry and denoisers. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 62020–62083. PMLR, 2025. URL https://proceedings.mlr.press/v267/wan25e.html.

Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, et al. De novo design of protein structure and function with RFdiffusion. *Nature*, 620:1089–1100, 2023. doi: 10.1038/s41586-023-06415-8. URL https://www.nature.com/articles/s41586-023-06415-8.

Yihong Wu and Sergio Verdú. MMSE dimension. *IEEE Transactions on Information Theory*, 57(8): 4857–4879, 2011. doi: 10.1109/TIT.2011.2158905.

Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. In *International Conference on Learning Representations*, 2022. URL https://arxiv.org/abs/2110.06197.

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 *Third Workshop on Structured Probabilistic Inference and Generative Modeling at NeurIPS*, 2025. URL https://arxiv.org/abs/2510.12975.

Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. SE(3) diffusion model with application to protein backbone generation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 40001–40039. PMLR, 2023. URL https://proceedings.mlr.press/v202/yim23a.html.

Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Zilong Wang, Aliaksandra Shysheya, Jonathan Crabbé, Shoko Ueda, et al. A generative model for inorganic materials design. *Nature*, 639:624–632, 2025. doi: 10.1038/s41586-025-08628-5. URL https://www.nature.com/articles/s41586-025-08628-5.

Part I: Theoretical Derivations

Derivations Supporting Sections 2–3

A Assumptions and the Local Form of the Perturbation

The master theorem separates the geometry of the data from the behavior of the perturbation. Assumption (A1) specifies the geometry that remains when neighborhoods of the query point are magnified; (A2) ensures that kernel-weighted integrals retain that geometry in the limit; and (A3) justifies the observation derivatives used by reconstruction. We first state these assumptions and explain their analytic consequences, then define class K and prove the homogeneous Euler identity. Appendix A.4 completes the additive case and gives an explicit example showing why numerical local dimension alone is insufficient.

A.1 Data, tangent measures, and analytic assumptions

Let μ be a probability measure on $\mathbb{R}^D$ and let $x \in \operatorname{supp}\mu$. For a closed ball $B(x,r)$, the numerical local dimension is

$$d(x) = \lim_{r\downarrow 0} \frac{\log \mu(B(x,r))}{\log r},$$

when this limit exists. The theorem uses a finite exponent $d \geq 0$ and a stronger condition on the rescaled measures, rather than assuming this numerical limit alone.

(A1) Tangent structure. Define the magnified data measure by

$$\nu_{x,\varepsilon} = \varepsilon^{-d}(T_{x,\varepsilon})_{\#}\mu \longrightarrow \nu_x, \qquad T_{x,\varepsilon}(u) = \frac{u-x}{\varepsilon}, \qquad \nu_x(aA) = a^d \nu_x(A), \quad a > 0. \tag{6}$$

The convergence is against continuous compactly supported functions, along the whole family $\varepsilon \downarrow 0$, and the limit is nonzero and finite on bounded sets. The last equality is d-homogeneity. It is also forced by full convergence with this normalization: if $S_a(z) = az$, then

$$\nu_{x,a\varepsilon} = a^{-d}(S_{1/a})_{\#}\nu_{x,\varepsilon}.$$

Passing to the limit against compactly supported continuous functions gives $\nu_x = a^{-d}(S_{1/a})_{\#}\nu_x$, which is the stated homogeneity relation.

This assumption identifies d with the numerical local dimension. Indeed, let $c_x = \nu_x(B(0,1))$. Local finiteness and nonzero homogeneity imply $0 < c_x < \infty$ and

$$\nu_x(B(0,r)) = c_x r^d \qquad (r > 0).$$

For $d > 0$, continuity of this expression in r implies that every sphere centered at the origin has zero ν_x-mass. For $d = 0$, the mass of every such ball equals c_x; continuity of measures as $r \downarrow 0$ gives $\nu_x(\{0\}) = c_x$, so $\nu_x = c_x\delta_0$, and the spheres again have zero mass. Locally weak convergence therefore gives convergence of the mass of the unit ball and hence

$$\mu(B(x,r)) = r^d(c_x + o(1)).$$

Taking logarithms proves $d(x) = d$. In the case $d = 0$, continuity of $\mu(B(x,r))$ as $r \downarrow 0$ also shows that $\mu(\{x\}) = c_x > 0$. No upper bound $d \leq D$ is imposed by (A1).

The reflected measures are $\check{\nu}_{x,\varepsilon}(A) = \nu_{x,\varepsilon}(-A)$ and $\check{\nu}_x(A) = \nu_x(-A)$. Reflection is needed because the kernel is evaluated at the displacement from a clean point to the observation, namely $(x-u)/\varepsilon$, whereas $T_{x,\varepsilon}$ uses $(u-x)/\varepsilon$.

Perturbation densities and reconstruction. For a transition density $Q_\varepsilon(y \mid u)$, the observation density and conditional-mean reconstruction are

$$\rho_\varepsilon^Q(y) = \int Q_\varepsilon(y \mid u)\,\mu(\mathrm{d}u), \qquad B_\varepsilon^Q(y) = \frac{\int uQ_\varepsilon(y \mid u)\,\mu(\mathrm{d}u)}{\rho_\varepsilon^Q(y)}. \tag{7}$$

The reconstruction is defined where the denominator is positive and the numerator has a finite absolute first moment. The readouts in (2) use these two quantities. In the additive case, $Q_\varepsilon(y \mid u) = \varepsilon^{-D} K((y-u)/\varepsilon)$ with a nonnegative C^1 probability density K, and the notation reduces to ρ_ε and b_ε.

(A2) Weighted tail control. For a nonnegative C^1 kernel K, set

$$W_K(z) = K(z) + \|z\|\, K(z) + \|\nabla K(z)\| + \|z\|\, \|\nabla K(z)\|\,. \tag{8}$$

For some $\varepsilon_0 > 0$, require

$$\lim_{R\to\infty} \sup_{0<\varepsilon<\varepsilon_0} \int_{\|z\|>R} W_K(-z)\, \nu_{x,\varepsilon}(\mathrm{d}z) = 0, \tag{9}$$

together with $0 < \int K \,\mathrm{d}\check{\nu}_x < \infty$. Equivalently, the integral in (9) is taken against $\check{\nu}_{x,\varepsilon}$ with weight $W_K(z)$. For a class-K family, this assumption concerns its limiting kernel K_x.

The four terms control the density, the first moment needed by reconstruction, the score numerator, and the position–gradient moment needed by the reconstruction Jacobian. The term $\|z\|\,\|\nabla K(z)\|$ also controls the scale derivative. Condition (9) prevents data that move arbitrarily far away under magnification from contributing a nonvanishing amount to any of these integrals. Positivity of the limiting kernel mass separately ensures that the limiting ratios have a nonzero denominator.

We record explicitly why (A1) and (A2) allow these integrals to converge. If f is continuous and $|f| \le C_f W_K$, then

$$\int f \,\mathrm{d}\check{\nu}_{x,\varepsilon} \longrightarrow \int f \,\mathrm{d}\check{\nu}_x.$$

To prove this, write $\tau(R)$ for the supremum of the weighted tails in (9). Local convergence and nonnegative compactly supported cutoffs in $\{\|z\| > R\}$ give

$$\int_{\|z\|>R} W_K(z)\, \check{\nu}_x(\mathrm{d}z) \le \tau(R).$$

Here the bound first holds on a compact annulus and then on the whole exterior region by increasing the annuli and their cutoffs. Since W_K is continuous and the measures have uniformly bounded mass on each fixed compact set, this also establishes integrability of W_K for the limiting measure. Now choose $\chi_R \in C_c^1$ with $0 \le \chi_R \le 1$, equal to one on $B(0,R)$ and zero outside $B(0,2R)$. The integral of $\chi_R f$ converges by (A1), while the two remaining tails are bounded by $C_f\tau(R)$. First let $\varepsilon \downarrow 0$ and then $R \to \infty$. The same argument applies componentwise to vector and matrix integrands.

Checking the tails for common kernels. For a local kernel $K \in C_c^1$, the weight has compact support and (9) is automatic. A second useful sufficient condition is rapid decay: for constants $C, c, \vartheta > 0$ and an integer $m_K \ge 0$,

$$|K(z)| + \|\nabla K(z)\| \le C(1+\|z\|)^{m_K} e^{-c\|z\|^\vartheta}. \tag{10}$$

The Gaussian satisfies this bound. To check uniform integrability under (A1), use the ball-mass asymptotic proved above to choose $r_0, C_0 > 0$ such that $\mu(B(x,r)) \le C_0 r^d$ for $r \le r_0$. If $\varepsilon R \le r_0$, this gives $\nu_{x,\varepsilon}(B(0,R)) \le C_0 R^d$; otherwise, the probability bound $\mu(B(x,\varepsilon R)) \le 1$ gives

$$\nu_{x,\varepsilon}(B(0,R)) \le \varepsilon^{-d} \le r_0^{-d} R^d.$$

Thus the magnified measures have a common polynomial bound on ball masses. On the annuli $2^j R < \|z\| \le 2^{j+1}R$, (10) bounds the weighted tail, uniformly in ε, by a constant times

$$\sum_{j=0}^{\infty} (1 + 2^{j+1}R)^{d+m_K+1} e^{-c(2^j R)^\vartheta},$$

which tends to zero as $R \to \infty$. Nondegeneracy of $\int K \,\mathrm{d}\check{\nu}_x$ must still be checked; it holds, for example, when K is strictly positive.

Compact support and rapid decay are sufficient examples, not restrictions in the definition of (A2). Kernels with polynomial tails require a separate check of (9) and the limiting kernel mass. Conversely, local regularity of the data density does not by itself justify differentiation at a fixed positive noise width; that operation concerns the original μ-integrals and has its own regularity requirements.

(A3) Reconstruction regularity. At the positive noise widths and observations where reconstruction is used, its conditional first moment is finite. Writing $N_\varepsilon(y) = \int u Q_\varepsilon(y \mid u)\, \mu(\mathrm{d}u)$, assume that ρ_ε^Q and N_ε are differentiable in the observation and that differentiation under their integrals is valid, with absolutely convergent integrals:

$$\nabla_y \rho_\varepsilon^Q(y) = \int \nabla_y Q_\varepsilon(y \mid u)\, \mu(\mathrm{d}u), \qquad J_y N_\varepsilon(y) = \int u \nabla_y Q_\varepsilon(y \mid u)^\top\, \mu(\mathrm{d}u).$$

These are the identities needed by the quotient rule for JB_ε^Q. A sufficient way to verify them is to dominate $(1 + \|u\|)\, \|\nabla_y Q_\varepsilon(y \mid u)\|$ by a μ-integrable function uniformly for y in a neighborhood of the observation, together with the stated finite moments. This domination is a sufficient check, rather than an additional required form of (A3). In particular, Lebesgue integrability of ∇K alone is not a substitute for integrability against the data measure and its first moment. The tangent integration by parts used in the flat case follows from (A2); its cutoff justification is given in Appendix C.

If a separate assertion involves second moments, it additionally requires the corresponding weighted control. For additive noise, uniform integrability of $\|z\|^2 K(z)$ gives convergence of the rescaled posterior second moment. For a general class-K family, the first-moment error condition below does not provide this conclusion: one must also control the second-moment contribution of its density error. The master theorem as stated does not require that additional posterior-moment conclusion.

Charts for the geometric examples. At an interior point of a C^1 embedded manifold, translate x to zero and choose orthonormal matrices $U \in \mathbb{R}^{D\times d}$ and $N \in \mathbb{R}^{D\times(D-d)}$ spanning the tangent and normal spaces. A local graph chart is

$$F(u) = Uu + \eta(u), \qquad \eta(0) = 0, \qquad D\eta(0) = 0, \tag{11}$$

where $\eta(u)$ takes values in the normal space. Its area factor is $J_F(u) = \sqrt{\det(DF(u)^\top DF(u))}$, so $J_F(u) \to 1$. For a C^2 chart, Taylor's formula also gives $\eta(u) = \mathcal{O}(\|u\|^2)$. Appendix D uses this chart to prove (A1) for a continuous positive sampling density and to identify its flat tangent measure.

A.2 Class K and normal-form errors

The additive model supplies the local form to which more general perturbations are compared. The width ε must be known in data coordinates: replacing it by an arbitrary model time changes the scale derivative in the readout. At a fixed query x, a limiting profile K_x may depend on x. It is required to be nonnegative and C^1, with (A2), but no additional ambient normalization of this local profile is imposed.

For a proposed profile K_x, define the errors in transition density, scale derivative, and observation derivative by

$$\begin{aligned} e_\varepsilon^{(0)}(z) &:= \varepsilon^D Q_\varepsilon(x \mid x - \varepsilon z) - K_x(z), \\ e_\varepsilon^{(s)}(z) &:= \varepsilon^D [DQ_\varepsilon + \varepsilon \partial_\varepsilon Q_\varepsilon](x \mid x - \varepsilon z) + z^\top \nabla K_x(z), \\ e_\varepsilon^{(y)}(z) &:= \varepsilon^{D+1} \nabla_y Q_\varepsilon(y \mid x - \varepsilon z)\big|_{y=x} - \nabla K_x(z). \end{aligned} \tag{12}$$

In the scale derivative, the observation and clean point are held fixed before the substitution $u = x - \varepsilon z$. Differentiating the substituted expression with z held fixed would instead move the clean point and would not give the derivative used by the readout.

Definition A.1 (Class-K noise model)**.** A family Q_ε belongs to class K at x relative to μ if its physical scale $\varepsilon \downarrow 0$ is known, there is a nonnegative C^1 profile K_x satisfying (A2), and the errors in (12) satisfy

$$\int (1 + \|z\|) \left(|e_\varepsilon^{(0)}(z)| + |e_\varepsilon^{(s)}(z)| + \left\| e_\varepsilon^{(y)}(z) \right\| \right) \check{\nu}_{x,\varepsilon}(\mathrm{d}z) \longrightarrow 0. \tag{13}$$

For μ-almost every u, the map $\varepsilon \mapsto Q_\varepsilon(x \mid u)$ is differentiable. Around every positive width in its domain, $|\partial_\varepsilon Q_\varepsilon(x \mid u)|$ is bounded by a μ-integrable majorant independent of the width in that neighborhood. Observation differentiation of the density and first moment is governed by (A3).

Equivalently, the required normal form is

$$\begin{aligned} \varepsilon^D Q_\varepsilon(x \mid x - \varepsilon z) &\longrightarrow K_x(z), \\ \varepsilon^D (DQ_\varepsilon + \varepsilon \partial_\varepsilon Q_\varepsilon)(x \mid x - \varepsilon z) &\longrightarrow -z^\top \nabla K_x(z), \\ \varepsilon^{D+1} \nabla_y Q_\varepsilon(y \mid x - \varepsilon z)\big|_{y=x} &\longrightarrow \nabla K_x(z), \end{aligned} \tag{14}$$

where all three arrows mean the weighted convergence in (13), not a separate pointwise assumption. The first line controls the density and reconstruction mass; the second controls the density readout; and the third controls the score and reconstruction response. The factor $1 + \|z\|$ covers both zeroth and first moments of these errors. The stated local domination in width justifies

$$\varepsilon \partial_\varepsilon \rho_\varepsilon^Q(x) = \int \varepsilon \partial_\varepsilon Q_\varepsilon(x \mid u)\, \mu(\mathrm{d}u).$$

For an additive family, direct differentiation at fixed y, u gives

$$(DQ_\varepsilon + \varepsilon \partial_\varepsilon Q_\varepsilon)(y \mid u) = -\varepsilon^{-D} z^\top \nabla K(z), \qquad \nabla_y Q_\varepsilon(y \mid u) = \varepsilon^{-D-1} \nabla K(z), \quad z = \frac{y-u}{\varepsilon}.$$

All three errors therefore vanish identically. Membership still requires (A2) and legitimate scale differentiation. For C_c^1 kernels, or kernels satisfying (10), the scale derivative at widths in a compact subinterval of $(0, \infty)$ is bounded as a function of u, which supplies the required majorant because μ is a probability measure.

A.3 Homogeneity and the scale slope

Lemma A.2 (Homogeneous Euler identity)**.** *For a nonzero locally finite d-homogeneous measure ν and a nonnegative C^1 kernel K such that either $\int K(az)\, \nu(\mathrm{d}z)$ can be differentiated under the integral at $a = 1$ or $\int (K(z) + \|z\|\, \|\nabla K(z)\|)\, \nu(\mathrm{d}z) < \infty$,*

$$\int -z^\top \nabla K(z)\, \nu(\mathrm{d}z) = d \int K(z)\, \nu(\mathrm{d}z). \tag{15}$$

We prove Lemma A.2, which converts homogeneous tangent geometry into the value d of the density readout. Let ν be a nonzero locally finite d-homogeneous measure, and suppose first that differentiation under the following integral is permitted:

$$\Lambda(a) = \int K(az)\, \nu(\mathrm{d}z), \qquad a > 0.$$

The pushforward of ν under $z \mapsto az$ is $a^{-d}\nu$, so

$$\Lambda(a) = a^{-d} \Lambda(1).$$

Differentiating at $a = 1$ yields

$$\int z^\top \nabla K(z)\, \nu(\mathrm{d}z) = -d \int K(z)\, \nu(\mathrm{d}z),$$

which is (15). For compactly supported C^1 kernels, differentiation is justified by a bounded integrand on a fixed compact set. For rapidly decaying kernels, a bound of the form (10), uniform for a near one, is integrable against the polynomial ball growth $\nu(B(0,R)) = R^d \nu(B(0,1))$.

The identity also holds under precisely the integrability needed in the master theorem, without requiring such a pointwise majorant over dilations. Suppose

$$\int (K(z) + \|z\|\, \|\nabla K(z)\|)\, \nu(\mathrm{d}z) < \infty.$$

Choose $\chi \in C_c^1$ with $0 \le \chi \le 1$, equal to one on $B(0,1)$ and zero outside $B(0,2)$, and put $\chi_R(z) = \chi(z/R)$. Apply the compact-support identity to $K\chi_R$ and expand its derivative:

$$\int -\chi_R(z) z^\top \nabla K(z)\, \nu(\mathrm{d}z) - \int K(z) z^\top \nabla \chi_R(z)\, \nu(\mathrm{d}z) = d \int K(z) \chi_R(z)\, \nu(\mathrm{d}z).$$

The first and last integrals converge by absolute integrability. The remaining term satisfies

$$\left| \int K(z) z^\top \nabla \chi_R(z)\, \nu(\mathrm{d}z) \right| \le 2 \left\| \nabla \chi \right\|_\infty \int_{\|z\|>R} K(z)\, \nu(\mathrm{d}z) \longrightarrow 0.$$

Letting $R \to \infty$ proves the identity. The weighted-convergence argument above shows that (A1)–(A2) supply this integrability for $\nu = \check{\nu}_x$ and $K = K_x$. Thus the Euler step used by the master theorem is justified throughout the stated weighted class, including when compact support or rapid decay is not assumed.

A.4 The role of homogeneity and possible relaxations

For additive noise, the density readout admits an exact expression in terms of the magnified data. We derive this expression before taking a limit, use it to identify the role of homogeneous tangents, and then construct a probability measure whose numerical LID exists although its density readout does not converge. This also distinguishes the full tangent assumption in the theorem from a sufficient subsequential variant available in the additive case.

The exact density readout. For the additive model, changing from u to the reflected displacement $z = (x-u)/\varepsilon$ in the sense of a pushforward measure gives

$$\int f(z)\, \check{\nu}_{x,\varepsilon}(\mathrm{d}z) = \varepsilon^{-d} \int f\left(\frac{x-u}{\varepsilon}\right) \mu(\mathrm{d}u).$$

Consequently,

$$\rho_\varepsilon(x) = \int \varepsilon^{-D} K\left(\frac{x-u}{\varepsilon}\right) \mu(\mathrm{d}u) = \varepsilon^{d-D} \int K\, \mathrm{d}\check{\nu}_{x,\varepsilon}.$$

To obtain the readout, differentiate the original convolution at a fixed query x, under the scale-differentiation condition of Definition A.1. At a fixed clean point u,

$$\varepsilon \partial_\varepsilon \left[\varepsilon^{-D} K\left(\frac{x-u}{\varepsilon}\right) \right] = \varepsilon^{-D} \left[-DK(z) - z^\top \nabla K(z) \right], \qquad z = \frac{x-u}{\varepsilon}.$$

Thus $D\rho_\varepsilon(x) + \varepsilon \partial_\varepsilon \rho_\varepsilon(x)$ is the convolution with $-z^\top \nabla K(z)$ in place of $K(z)$. Dividing by the positive density yields the two exact identities

$$\rho_\varepsilon(x) = \varepsilon^{d-D} \int K\, \mathrm{d}\check{\nu}_{x,\varepsilon}, \qquad \mathcal{R}_\varepsilon^{\mathrm{dens}}(x) = \frac{\int -z^\top \nabla K(z)\, \check{\nu}_{x,\varepsilon}(\mathrm{d}z)}{\int K(z)\, \check{\nu}_{x,\varepsilon}(\mathrm{d}z)}. \tag{16}$$

This derivation neither differentiates an asymptotic remainder nor divides by the kernel at points where it may vanish.

Under (A1)–(A2), the weighted-convergence argument in Appendix A applies to both integrands. The denominator tends to $C_x = \int K\, \mathrm{d}\check{\nu}_x > 0$, and the Euler identity gives

$$\int -z^\top \nabla K(z)\, \check{\nu}_{x,\varepsilon}(\mathrm{d}z) \longrightarrow \int -z^\top \nabla K(z)\, \check{\nu}_x(\mathrm{d}z) = dC_x.$$

Therefore $\rho_\varepsilon(x) = \varepsilon^{d-D}(C_x + o(1))$ and $\mathcal{R}_\varepsilon^{\mathrm{dens}}(x) \to d$. Assumption (A1) also gives $d(x) = d$, as proved above, so this is recovery of the numerical local dimension. The argument uses control of both the density and its scale derivative; the density asymptotic alone does not justify differentiation of its remainder.

A sufficient condition with nonunique tangents. The additive readout can converge without a unique tangent limit. Suppose that the family $\nu_{x,\varepsilon}$ is relatively compact for locally weak convergence as $\varepsilon \downarrow 0$, that every nonzero subsequential limit is d-homogeneous, and that (9) holds. Suppose also that

$$\liminf_{\varepsilon \downarrow 0} \int K\, \mathrm{d}\check{\nu}_{x,\varepsilon} > 0.$$

Take any sequence of widths tending to zero and pass to a subsequence with a locally weak limit ν. Weighted tail control gives convergence of both integrals in (16) along that subsequence. The

denominator bound implies $\int K \, \mathrm{d}\check{\nu} > 0$, so ν is nonzero. Homogeneity and the cutoff proof of the Euler identity then show that the ratio tends to d. If the full readout failed to converge to d, some sequence would stay a fixed positive distance from d, contradicting the convergent subsequence just obtained. This is a separate sufficient condition for the additive ratio; the master theorem retains its stated full-limit assumption (A1).

A probability measure with LID but no limiting readout. Consider $x = 0$ in $\mathbb{R}^2$. Let $X = (T, 0)$, where $T \in (0, 1)$ has density

$$p(t) = \frac{4}{5}\left(1 + \frac{1}{2}\cos(\log t)\right), \qquad 0 < t < 1.$$

This is strictly positive, and it integrates to one because

$$\int_0^1 \cos(\log t)\, \mathrm{d}t = \left[\frac{t}{2}\big(\cos(\log t) + \sin(\log t)\big)\right]_0^1 = \frac{1}{2}.$$

For $0 < r < 1$, its ball mass at the origin is

$$\mu(B(0, r)) = \frac{4}{5} r \left[1 + \frac{1}{4}\big(\cos(\log r) + \sin(\log r)\big)\right].$$

The bracket lies between $1 - \sqrt{2}/4$ and $1 + \sqrt{2}/4$. Its logarithm is therefore bounded, and division by $\log r$ proves $d(0) = 1$.

Use the strictly positive standard Gaussian kernel $K(z) = (2\pi)^{-1} e^{-\|z\|^2/2}$. The measure has compact support, so the convolution and its scale and observation derivatives are legitimate at every positive width. Substituting $t = \varepsilon s$ gives

$$\rho_\varepsilon(0) = \frac{2}{5\pi}\varepsilon^{-1} I_\varepsilon, \qquad I_\varepsilon = \int_0^{1/\varepsilon} e^{-s^2/2}\left[1 + \frac{1}{2}\cos(\log\varepsilon + \log s)\right] \mathrm{d}s.$$

To evaluate the oscillation that survives smoothing, set

$$A = \int_0^\infty e^{-s^2/2}\, \mathrm{d}s = \sqrt{\frac{\pi}{2}}, \qquad M = \int_0^\infty e^{-s^2/2} s^{\mathrm{i}}\, \mathrm{d}s = 2^{(\mathrm{i}-1)/2}\Gamma\left(\frac{1+\mathrm{i}}{2}\right),$$

where $\mathrm{i}^2 = -1$. The second equality follows by the substitution $v = s^2/2$ in the defining gamma integral. In particular, $M \neq 0$: the gamma reflection identity gives

$$|M|^2 = \frac{1}{2}\Gamma\left(\frac{1+\mathrm{i}}{2}\right)\Gamma\left(\frac{1-\mathrm{i}}{2}\right) = \frac{\pi}{2\cosh(\pi/2)} > 0.$$

Define the positive periodic function

$$F(\theta) = A + \frac{1}{2}\operatorname{Re}(e^{\mathrm{i}\theta} M).$$

Its integral representation gives $A/2 \le F(\theta) \le 3A/2$. Moreover,

$$|I_\varepsilon - F(\log\varepsilon)| \le \frac{3}{2}\int_{1/\varepsilon}^\infty e^{-s^2/2}\, \mathrm{d}s \longrightarrow 0.$$

The derivative must also be checked. With $L = 1/\varepsilon$, direct differentiation of the integral, including its moving endpoint, gives

$$\varepsilon\partial_\varepsilon I_\varepsilon = \frac{1}{2}\operatorname{Re}\left(\mathrm{i}e^{\mathrm{i}\log\varepsilon}\int_0^L e^{-s^2/2} s^{\mathrm{i}}\, \mathrm{d}s\right) - \frac{3}{2} L e^{-L^2/2}.$$

The endpoint coefficient is $3/2$ because $\log\varepsilon + \log L = 0$. Subtracting $F'(\log\varepsilon)$ bounds the error by

$$\frac{1}{2}\int_L^\infty e^{-s^2/2}\, \mathrm{d}s + \frac{3}{2} L e^{-L^2/2} \longrightarrow 0.$$

Since F is bounded away from zero, the exact density formula now yields

$$\mathcal{R}_\varepsilon^{\mathrm{dens}}(0) = 2 + \varepsilon\partial_\varepsilon \log \rho_\varepsilon(0) = 1 + \frac{F'(\log \varepsilon)}{F(\log \varepsilon)} + o(1).$$

Write $M = |M|e^{\mathrm{i}\phi}$. Along $\varepsilon_n = e^{-\phi-2\pi n}$, the readout tends to 1, whereas along $\widetilde{\varepsilon}_n = e^{\pi/2-\phi-2\pi n}$ it tends to

$$1 - \frac{|M|}{2A} = 1 - \frac{1}{2\sqrt{\cosh(\pi/2)}} \neq 1.$$

The density readout therefore has no limit, despite $d(0) = 1$.

The failure is in tangent homogeneity, not in the Gaussian tails or differentiation. For a compactly supported continuous f,

$$\int f \,\mathrm{d}\nu_{0,\varepsilon} = \frac{4}{5}\int_0^{1/\varepsilon} f((s,0))\left[1 + \frac{1}{2}\cos(\log\varepsilon + \log s)\right] \mathrm{d}s.$$

Along a sequence with fixed phase $\log\varepsilon \bmod 2\pi = \theta$, this converges to the measure on the positive axis with density $\frac{4}{5}[1 + \frac{1}{2}\cos(\theta + \log s)]$. Dilation changes the phase by the logarithm of the dilation factor, so these limits are not 1-homogeneous and differ across phases. Meanwhile, all the magnified measures are bounded by $6/5$ times length measure on the positive axis; the reflected measures have the corresponding bound on the negative axis. The restriction of W_K to either axis is a polynomial times a Gaussian, and hence its tails vanish uniformly in ε. Thus the weighted tail part of (A2) holds, and the kernel masses stay bounded away from zero for small widths, but (A1) fails. The example establishes the claimed distinction between an existing numerical LID and the geometric conditions needed by the readout.

B Proof of the Density Decomposition

Proposition 2.1 relates the change in density with noise width to the response of the posterior-mean reconstruction. The relation is exact at each positive width; it does not require a tangent measure or a small-noise limit. Its proof uses the same posterior distribution to express both derivatives, after which the correction appears when a cross moment is separated into a covariance and a product of means.

Fix $\varepsilon > 0$ and an observation y with $\rho_\varepsilon(y) > 0$ in the additive model (1). Write

$$Z_y = \frac{y - X}{\varepsilon}, \qquad \mathbb{E}[\,\cdot \mid y] = \mathbb{E}[\,\cdot \mid Y_\varepsilon = y].$$

Bayes' formula identifies the conditional law used below: for every posterior-integrable function h,

$$\mathbb{E}[h(Z_y) \mid y] = \frac{1}{\rho_\varepsilon(y)}\int h\left(\frac{y-u}{\varepsilon}\right)\varepsilon^{-D}K\left(\frac{y-u}{\varepsilon}\right)\mu(\mathrm{d}u).$$

We use the finite conditional first moments and valid differentiation under the integral assumed in the proposition. In particular, observation differentiation applies to both the density and the vector first-moment numerator $N_\varepsilon(y) = \int uK_\varepsilon(y-u)\,\mu(\mathrm{d}u)$ of $b_\varepsilon = N_\varepsilon/\rho_\varepsilon$; the differentiated integrands are integrable with respect to μ. This is the weighted integrability needed for the posterior cross moments below, rather than an assumption on the unconditional moments of X.

The kernel need not be strictly positive. Define

$$\psi_K(z) = \begin{cases} \nabla K(z)/K(z), & K(z) > 0, \\ 0, & K(z) = 0. \end{cases}$$

Since K is nonnegative and C^1, every zero of K is a minimum and satisfies $\nabla K = 0$. Hence $K\psi_K = \nabla K$ everywhere, while the set where $K = 0$ has zero posterior mass. The chosen extension therefore introduces no division by zero and does not affect any expectation.

Expressing the two derivatives through the posterior. In the convolution integrand, set $z = (y-u)/\varepsilon$. Differentiation in width holds y and u fixed, whereas differentiation in the observation holds ε and u fixed. Thus

$$\varepsilon\partial_\varepsilon\big[\varepsilon^{-D}K(z)\big] = \varepsilon^{-D}\big[-DK(z) - z^\top\nabla K(z)\big],$$
$$\nabla_y\big[\varepsilon^{-D}K(z)\big] = \varepsilon^{-D-1}\nabla K(z).$$

The assumed derivative exchanges let us integrate these identities. Dividing the first by $\rho_\varepsilon(y)$ and adding D gives

$$D + \varepsilon\partial_\varepsilon \log\rho_\varepsilon(y) = \mathbb{E}[-Z_y^\top\psi_K(Z_y) \mid y]. \tag{17}$$

Dividing the observation derivative by the same density gives

$$\nabla \log\rho_\varepsilon(y) = \varepsilon^{-1}\mathbb{E}[\psi_K(Z_y) \mid y]. \tag{18}$$

These expectations are finite: after multiplication by the posterior weight K, their absolute integrands are bounded by $\|z\|\,\|\nabla K(z)\|$ and $\|\nabla K(z)\|$, respectively, whose corresponding integrals are among those required for the valid derivative exchanges.

To differentiate reconstruction, take the Jacobian of $N_\varepsilon/\rho_\varepsilon$, with convention $(Jb)_{ij} = \partial_{y_j} b_i$. The numerator derivative is

$$JN_\varepsilon(y) = \int u\,\nabla_y K_\varepsilon(y-u)^\top\,\mu(\mathrm{d}u) = \varepsilon^{-1}\rho_\varepsilon(y)\mathbb{E}[X\psi_K(Z_y)^\top \mid y].$$

The quotient rule and (18) therefore yield

$$\begin{aligned} Jb_\varepsilon(y) &= \varepsilon^{-1}\left(\mathbb{E}[X\psi_K(Z_y)^\top \mid y] - \mathbb{E}[X \mid y]\mathbb{E}[\psi_K(Z_y) \mid y]^\top\right) \\ &= \varepsilon^{-1}\operatorname{Cov}(X, \psi_K(Z_y) \mid y) = -\operatorname{Cov}(Z_y, \psi_K(Z_y) \mid y). \end{aligned} \tag{19}$$

The last equality uses $X = y - \varepsilon Z_y$ under the conditional law. Integrability of the differentiated vector numerator controls the cross moment involving X; together with integrability of the density gradient it also controls the cross moment involving Z_y. No second moment of X is needed for this general-kernel identity.

Obtaining the correction. Taking the trace in (19) and expanding the covariance gives

$$\operatorname{tr} Jb_\varepsilon(y) = \mathbb{E}[-Z_y^\top\psi_K(Z_y) \mid y] + \mathbb{E}[Z_y \mid y]^\top\mathbb{E}[\psi_K(Z_y) \mid y].$$

The first term is $\mathcal{R}_\varepsilon^{\mathrm{dens}}(y)$ by (17). For the second, the posterior mean and (18) give

$$b_\varepsilon(y) - y = -\varepsilon\mathbb{E}[Z_y \mid y],$$
$$(b_\varepsilon(y) - y)^\top\nabla\log\rho_\varepsilon(y) = -\mathbb{E}[Z_y \mid y]^\top\mathbb{E}[\psi_K(Z_y) \mid y].$$

Consequently,

$$\mathcal{R}_\varepsilon^{\mathrm{dens}}(y) = \mathcal{R}_\varepsilon^{\mathrm{resp}}(y) + (b_\varepsilon(y) - y)^\top\nabla\log\rho_\varepsilon(y),$$

which proves Proposition 2.1, with the correction sign as stated in (3). The correction is a directional derivative of log density at y, evaluated along $b_\varepsilon(y) - y$; it is not the finite change in log density between these two points. Its sign is not fixed for a general kernel.

Gaussian specialization and the small-noise question. For the standard Gaussian kernel, $\psi_K(z) = -z$. Equations (18) and (19) become

$$b_\varepsilon(y) = y + \varepsilon^2\nabla\log\rho_\varepsilon(y),$$
$$Jb_\varepsilon(y) = \operatorname{Cov}(Z_y \mid y) = \varepsilon^{-2}\operatorname{Cov}(X \mid Y_\varepsilon = y).$$

The required posterior moments are finite under Gaussian smoothing of a probability measure. In this case the correction is the nonnegative quantity

$$(b_\varepsilon(y) - y)^\top\nabla\log\rho_\varepsilon(y) = \frac{\|b_\varepsilon(y) - y\|^2}{\varepsilon^2} = \varepsilon^2\|\nabla\log\rho_\varepsilon(y)\|^2.$$

Writing $s_\varepsilon = \nabla\log\rho_\varepsilon$, differentiation of the Gaussian reconstruction identity gives $\operatorname{tr} Jb_\varepsilon = D + \varepsilon^2\operatorname{div} s_\varepsilon$. The decomposition is therefore

$$\mathcal{R}_\varepsilon^{\mathrm{dens}} = D + \varepsilon^2\big(\operatorname{div} s_\varepsilon + \|s_\varepsilon\|^2\big),$$

the FLIPD expression. This specialization uses Gaussian smoothness; the general-kernel proof above uses only first derivatives.

None of these finite-width identities alone establishes convergence to LID. Appendix C proves the density and corrected-response limits, and shows why the correction vanishes on a flat tangent even for an admissible non-Gaussian kernel. At a boundary it need not vanish, as the Gaussian example in Appendix D.5 makes explicit.

C Proof of the Master Theorem and the Adapter Principle

The master theorem concerns the exact density and posterior mean of a class-K perturbation. We first show that its density and scale derivative are governed by the same homogeneous tangent measure, which proves both assertions in (4). We then differentiate the posterior mean before taking the small-noise limit, retaining all terms that contribute to the correction in (5). Finally, when the tangent is a plane, integration by parts identifies the limiting Jacobian as the projector in Corollary 2.3.

Passing from the original data to the tangent measure. Fix x satisfying (A1), and use the reflected magnified measures from Appendix A. Their defining change of variables is

$$\int f(z)\,\check{\nu}_{x,\varepsilon}(\mathrm{d}z) = \varepsilon^{-d}\int f\left(\frac{x-u}{\varepsilon}\right)\,\mu(\mathrm{d}u). \tag{20}$$

Reflection preserves local weak convergence and d-homogeneity, so the full family $\check{\nu}_{x,\varepsilon}$ converges locally weakly to the nonzero measure $\check{\nu}_x$. Write $K = K_x$ within this proof. The errors in (12) give the exact representations

$$\begin{aligned}
\varepsilon^D Q_\varepsilon(x \mid x-\varepsilon z) &= K(z) + e_\varepsilon^{(0)}(z),\\
\varepsilon^D(DQ_\varepsilon + \varepsilon\partial_\varepsilon Q_\varepsilon)(x \mid x-\varepsilon z) &= -z^\top \nabla K(z) + e_\varepsilon^{(s)}(z),\\
\varepsilon^{D+1}\nabla_y Q_\varepsilon(y \mid x-\varepsilon z)\big|_{y=x} &= \nabla K(z) + e_\varepsilon^{(y)}(z).
\end{aligned}$$

The scale derivative in the second line is taken with both arguments of Q_ε fixed, before the substitution for u. Class K requires

$$\int (1+\|z\|)\left(|e_\varepsilon^{(0)}| + |e_\varepsilon^{(s)}| + \left\|e_\varepsilon^{(y)}\right\|\right)\,\mathrm{d}\check{\nu}_{x,\varepsilon} \longrightarrow 0.$$

We shall repeatedly integrate these representations. To justify each limit, let F be any continuous scalar, vector, or matrix function bounded in norm by a constant multiple of W_K in (8). Multiplying F by a continuous cutoff supported in a ball makes its integral converge by (A1). The omitted tail is uniformly small by (A2); the same tail bound for the limiting measure follows by compactly supported cutoffs and monotone convergence. Sending the cutoff radius to infinity therefore gives

$$\int F\,\mathrm{d}\check{\nu}_{x,\varepsilon} \longrightarrow \int F\,\mathrm{d}\check{\nu}_x.$$

This applies to K, zK, ∇K, $z\nabla K^\top$, and $z^\top\nabla K$, since all are continuous and controlled by W_K. It also shows that their limiting integrals are absolutely finite.

Introduce the moments needed for density and reconstruction:

$$\begin{aligned}
C_\varepsilon &= \int (K + e_\varepsilon^{(0)})\,\mathrm{d}\check{\nu}_{x,\varepsilon}, & M_\varepsilon &= \int z(K + e_\varepsilon^{(0)})\,\mathrm{d}\check{\nu}_{x,\varepsilon},\\
G_\varepsilon &= \int (\nabla K + e_\varepsilon^{(y)})\,\mathrm{d}\check{\nu}_{x,\varepsilon}, & H_\varepsilon &= \int z(\nabla K + e_\varepsilon^{(y)})^\top\,\mathrm{d}\check{\nu}_{x,\varepsilon}.
\end{aligned} \tag{21}$$

The weighted error condition controls $e_\varepsilon^{(0)}$ and $ze_\varepsilon^{(0)}$ in the first two integrals, and $e_\varepsilon^{(y)}$ and $z(e_\varepsilon^{(y)})^\top$ in the last two. Together with the cutoff argument, it proves

$$(C_\varepsilon, M_\varepsilon, G_\varepsilon, H_\varepsilon) \longrightarrow (C, M, G, H),$$

where

$$C = \int K \, \mathrm{d}\check{\nu}_x = C_x, \qquad M = \int zK \, \mathrm{d}\check{\nu}_x,$$
$$G = \int \nabla K \, \mathrm{d}\check{\nu}_x, \qquad H = \int z\nabla K^\top \, \mathrm{d}\check{\nu}_x.$$

By (A2), $0 < C < \infty$, so $C_\varepsilon > 0$ for all sufficiently small ε. These are first moments and mixed first-derivative moments; convergence of posterior second moments is not needed.

The density and its scale response. Apply (20) to the exact observation density in (7). The first class-K representation gives

$$\rho_\varepsilon^Q(x) = \int Q_\varepsilon(x \mid u) \, \mu(\mathrm{d}u) = \varepsilon^{d-D} C_\varepsilon. \tag{22}$$

Thus the density asymptotic follows from $C_\varepsilon \to C$. Its scale derivative requires a separate argument: convergence of C_ε alone would not control its derivative. The local domination in Definition A.1 permits differentiation of the original density integral in ε. Only after taking this derivative do we use the magnified coordinates, obtaining

$$\begin{aligned} D\rho_\varepsilon^Q(x) + \varepsilon\partial_\varepsilon\rho_\varepsilon^Q(x) &= \int (DQ_\varepsilon + \varepsilon\partial_\varepsilon Q_\varepsilon)(x \mid u) \, \mu(\mathrm{d}u) \\ &= \varepsilon^{d-D} \int [-z^\top \nabla K(z) + e_\varepsilon^{(s)}(z)] \, \check{\nu}_{x,\varepsilon}(\mathrm{d}z). \end{aligned}$$

Division by (22) gives

$$D + \varepsilon\partial_\varepsilon \log \rho_\varepsilon^Q(x) = \frac{\int [-z^\top \nabla K(z) + e_\varepsilon^{(s)}(z)] \, \check{\nu}_{x,\varepsilon}(\mathrm{d}z)}{C_\varepsilon}. \tag{23}$$

The weighted error bound removes $e_\varepsilon^{(s)}$, and the cutoff argument passes the remaining integral to $\check{\nu}_x$. Since this measure is d-homogeneous, the Euler identity (15), justified in Appendix A.3, supplies its value:

$$C_\varepsilon \longrightarrow C > 0, \qquad \int [-z^\top \nabla K + e_\varepsilon^{(s)}] \, \mathrm{d}\check{\nu}_{x,\varepsilon} \longrightarrow \int -z^\top \nabla K \, \mathrm{d}\check{\nu}_x = dC. \tag{24}$$

Equations (22) and (23) now prove $\rho_\varepsilon^Q(x) = \varepsilon^{d-D}(C_x + o(1))$ and $\mathcal{R}_\varepsilon^{\mathrm{dens}}(x) \to d$, precisely (4).

The limiting posterior and reconstruction displacement. The same change of variables identifies the posterior distribution that will determine the response. For every bounded continuous f,

$$\mathbb{E}\left[f\left(\frac{x - X}{\varepsilon} \right) \middle| Y_\varepsilon = x \right] = \frac{\int f(z)(K(z) + e_\varepsilon^{(0)}(z)) \, \check{\nu}_{x,\varepsilon}(\mathrm{d}z)}{C_\varepsilon}. \tag{25}$$

The $e_\varepsilon^{(0)}$ term is bounded by $\|f\|_\infty \int |e_\varepsilon^{(0)}| \, \mathrm{d}\check{\nu}_{x,\varepsilon}$ and vanishes. The remaining numerator converges by the cutoff argument applied to fK. Because $C_\varepsilon \to C > 0$, the conditional law converges weakly to the probability measure

$$\Pi_x(\mathrm{d}z) = C_x^{-1} K_x(z) \check{\nu}_x(\mathrm{d}z). \tag{26}$$

The same reasoning applies to continuous f of at most linear growth: fK is controlled by W_K, and the error term is controlled by the factor $1 + \|z\|$ in class K. In particular, the first moments converge and

$$B_\varepsilon^Q(x) = x - \varepsilon \frac{M_\varepsilon}{C_\varepsilon}, \qquad \frac{B_\varepsilon^Q(x) - x}{\varepsilon} \longrightarrow -m_x, \quad m_x = \frac{M}{C} = \mathbb{E}_{\Pi_x} Z.$$

The minus sign follows from the chosen displacement $z = (x - u)/\varepsilon$.

Differentiating reconstruction and identifying the correction. For this part assume (A3). We use its valid observation differentiation for both ρ_ε^Q and the vector numerator $N_\varepsilon(y) = \int u Q_\varepsilon(y \mid u)\, \mu(\mathrm{d}u)$ in $B_\varepsilon^Q = N_\varepsilon/\rho_\varepsilon^Q$. The first-moment and derivative integrals are finite by these assumptions. With the Jacobian convention $(JB)_{ij} = \partial_{y_j} B_i$, the third class-K representation gives

$$\begin{aligned}\nabla\rho_\varepsilon^Q(x) &= \varepsilon^{d-D-1} G_\varepsilon,\\ N_\varepsilon(x) &= \varepsilon^{d-D}(xC_\varepsilon - \varepsilon M_\varepsilon),\\ JN_\varepsilon(x) &= \varepsilon^{d-D-1}(xG_\varepsilon^\top - \varepsilon H_\varepsilon).\end{aligned}$$

Here the observation derivative is taken before setting $y = x$; the substitution $u = x - \varepsilon z$ is only a way to evaluate the resulting integrals. The quotient rule consequently gives

$$\begin{aligned}JB_\varepsilon^Q(x) &= \frac{JN_\varepsilon(x)}{\rho_\varepsilon^Q(x)} - \frac{N_\varepsilon(x)\nabla\rho_\varepsilon^Q(x)^\top}{(\rho_\varepsilon^Q(x))^2}\\ &= \frac{xG_\varepsilon^\top - \varepsilon H_\varepsilon}{\varepsilon C_\varepsilon} - \frac{(xC_\varepsilon - \varepsilon M_\varepsilon)G_\varepsilon^\top}{\varepsilon C_\varepsilon^2}.\end{aligned}$$

The terms containing x cancel, leaving the exact identity

$$JB_\varepsilon^Q(x) = -\frac{H_\varepsilon}{C_\varepsilon} + \frac{M_\varepsilon G_\varepsilon^\top}{C_\varepsilon^2}. \tag{27}$$

All four quantities on the right have already been shown to converge, including the mixed moment H_ε. Thus this step needs no convergence of the posterior as a function of observations in a neighborhood, nor a second derivative of K.

To express the limit in posterior terms, define $\psi_{K_x} = \nabla K_x / K_x$ where $K_x > 0$ and set it to zero elsewhere. Nonnegativity and C^1 regularity imply $\nabla K_x = 0$ on its zero set, so $K_x \psi_{K_x} = \nabla K_x$ everywhere. Moreover, (A2) gives

$$\mathbb{E}_{\Pi_x} \|\psi_{K_x}(Z)\| < \infty, \qquad \mathbb{E}_{\Pi_x}\big[\|Z\|\, \|\psi_{K_x}(Z)\|\big] < \infty.$$

Consequently the following covariance is well-defined, with $g_x = G/C = \mathbb{E}_{\Pi_x}\psi_{K_x}(Z)$:

$$\begin{aligned}JB_\varepsilon^Q(x) &\longrightarrow -\frac{H}{C} + \frac{MG^\top}{C^2}\\ &= -\mathbb{E}_{\Pi_x}[Z\psi_{K_x}(Z)^\top] + m_x g_x^\top = -\operatorname{Cov}_{\Pi_x}(Z, \psi_{K_x}(Z)) =: A_x^Q.\end{aligned} \tag{28}$$

Taking the trace and using the Euler identity gives

$$\operatorname{tr} A_x^Q = -\frac{1}{C}\int z^\top \nabla K_x(z)\, \check{\nu}_x(\mathrm{d}z) + m_x^\top g_x = d + m_x^\top g_x. \tag{29}$$

We have therefore established the matrix and trace limits

$$JB_\varepsilon^Q(x) \longrightarrow A_x^Q = -\operatorname{Cov}_{\Pi_x}(Z, \psi_{K_x}(Z)), \qquad \operatorname{tr} A_x^Q - m_x^\top g_x = d. \tag{30}$$

The correction has the opposite limiting sign. Indeed, the density derivative above and the posterior displacement give

$$\begin{aligned}\varepsilon\nabla \log \rho_\varepsilon^Q(x) &= \frac{G_\varepsilon}{C_\varepsilon} \longrightarrow g_x,\\ (B_\varepsilon^Q(x) - x)^\top \nabla \log \rho_\varepsilon^Q(x) &= -\frac{M_\varepsilon^\top G_\varepsilon}{C_\varepsilon^2} \longrightarrow -m_x^\top g_x.\end{aligned}$$

Both limits are finite, and their sum is d by (29). This proves (5) and completes Theorem 2.2. For a general class-K perturbation the relation is asymptotic; the exact finite-width decomposition in Appendix B applies to the additive model.

Flat tangents and tangent recovery. For Corollary 2.3, assume $\check{\nu}_x = c\mathcal{H}^d \lfloor_T$ with $c > 0$. Choose orthonormal coordinates $T \oplus T^\perp$, so that $T = \mathbb{R}^d \times \{0\}$. Since $Z \sim \Pi_x$ lies in T, its mean m_x is tangent and every normal row of $A_x^Q = -\operatorname{Cov}_{\Pi_x}(Z, \psi_{K_x}(Z))$ is zero. To determine the tangent columns, it remains to integrate first derivatives of K_x along the plane.

These integrations by parts are justified by the integrability in (A2). In tangent coordinates choose $\chi_R \in C_c^1(T)$ equal to one on the ball of radius R, zero outside the ball of radius $2R$, and with $\|\nabla_T \chi_R\| \le C_0/R$. For $i, j \le d$, ordinary integration by parts with compact support gives

$$\int_T \chi_R \partial_j K_x \,\mathrm{d}\mathcal{H}^d = -\int_T K_x \partial_j \chi_R \,\mathrm{d}\mathcal{H}^d,$$
$$\int_T \chi_R z_i \partial_j K_x \,\mathrm{d}\mathcal{H}^d = -\delta_{ij} \int_T \chi_R K_x \,\mathrm{d}\mathcal{H}^d - \int_T z_i K_x \partial_j \chi_R \,\mathrm{d}\mathcal{H}^d.$$

The first right-hand side tends to zero because K_x is integrable on T. The last term is bounded by a constant times $\int_{\|z\|>R} K_x \,\mathrm{d}\mathcal{H}^d$ and also tends to zero. The left-hand sides converge by integrability of ∇K_x and $\|z\|\,\|\nabla K_x\|$ on T. Hence

$$\int_T \partial_j K_x \,\mathrm{d}\mathcal{H}^d = 0, \qquad \int_T z_i \partial_j K_x \,\mathrm{d}\mathcal{H}^d = -\delta_{ij} \int_T K_x \,\mathrm{d}\mathcal{H}^d.$$

The first identity says that every tangent component of g_x is zero. Thus $g_x \perp T$ and $m_x^\top g_x = 0$. The second identity makes the tangent block of $-\mathbb{E}_{\Pi_x}[Z\psi_{K_x}(Z)^\top]$ equal to I_d; the tangent columns of $m_x g_x^\top$ vanish. Together with the zero normal rows, this proves

$$JB_\varepsilon^Q(x) \longrightarrow A_x^Q = \begin{pmatrix} I_d & B \\ 0 & 0 \end{pmatrix}, \qquad (A_x^Q)^2 = A_x^Q, \qquad \operatorname{tr} A_x^Q = \operatorname{rank} A_x^Q = d. \tag{31}$$

Its range is T, and it acts as the identity on T; it is therefore a projector onto the tangent plane. The block B may map a normal input to a tangent output, so orthogonality does not follow for an arbitrary kernel. The trace nevertheless converges to d, and the correction tends to zero because $m_x^\top g_x = 0$.

For a radially symmetric C^1 kernel, reflect across any normal coordinate hyperplane. The reflection fixes every $z \in T$ and leaves K_x unchanged, so differentiating this symmetry in the reflected normal direction gives $\partial_j K_x(z) = -\partial_j K_x(z) = 0$ for $j > d$. Thus every normal component of ψ_{K_x} vanishes on T wherever $K_x > 0$; our extension also vanishes at its zeros. Every normal column of the covariance matrix is consequently zero, giving $B = 0$ and $A_x^Q = P_T$. This uses only first derivatives and completes Corollary 2.3.

Appendix D verifies the flat tangent at an interior point of an embedded C^1 manifold with continuous positive sampling density. On a half-plane, the tangent integration by parts can instead contain a boundary contribution, so the correction need not vanish; Appendix D.5 evaluates the Gaussian case.

C.1 Conversion to model coordinates and time

The readouts in Section 3 use the coordinates and time variable exposed by each model. To connect them to the master theorem, the query must first be expressed in data coordinates and the model clock converted to physical noise width. The following result records these two changes of variables; the separate model appendices then express the resulting derivatives through scores, velocities, or drifts.

Let $y = \Phi_t(\bar{y})$ be a known local diffeomorphism from data coordinates to model coordinates, and let p_t^y denote the model density. Its pullback to data coordinates is

$$\bar{p}_t(\bar{y}) = p_t^y(\Phi_t(\bar{y}))\,|\det D\Phi_t(\bar{y})|.$$

The determinant accounts for the change in ambient volume. Write $\varepsilon(t) > 0$ for the physical width, assume it is differentiable, and set $\kappa_t = \dot{\varepsilon}(t)/\varepsilon(t) \neq 0$. Derivatives in t below hold the data query $\bar{y}$ fixed; the model query $\Phi_t(\bar{y})$ can therefore move.

Corollary C.1 (Adapter principle)**.** *Suppose $\varepsilon(t) \to 0$, and let ρ_ε^Q satisfy the assumptions of Theorem 2.2. If, near the query,*

$$\bar{p}_t(\bar{y}) = w_t(\bar{y}) \rho^Q_{\varepsilon(t)}(\bar{y}), \qquad w_t(\bar{y}) > 0, \qquad \frac{\partial_t \log w_t(x)}{\kappa_t} \longrightarrow 0, \tag{32}$$

with the indicated derivatives defined, then

$$D + \frac{\partial_t \log \bar{p}_t(x)}{\kappa_t} \longrightarrow d. \tag{33}$$

If reconstruction in data coordinates satisfies $J\bar{B}_t(x) - JB^Q_{\varepsilon(t)}(x) \to 0$, then, under (A3) and a flat tangent, $J\bar{B}_t(x)$ converges to the tangent projector of Corollary 2.3 and $\operatorname{tr} J\bar{B}_t(x) \to d$.

Proof and the moving query. Taking the logarithm in (32) and using $\dot{\varepsilon} = \kappa_t \varepsilon$ gives the exact chain rule

$$D + \frac{\partial_t \log \bar{p}_t(x)}{\kappa_t} = \big[D + \varepsilon \partial_\varepsilon \log \rho^Q_\varepsilon(x)\big]_{\varepsilon=\varepsilon(t)} + \frac{\partial_t \log w_t(x)}{\kappa_t}. \tag{34}$$

The first term tends to d by the master theorem; the second tends to zero by hypothesis. For reconstruction, the assumed Jacobian difference tends to zero and the exact posterior Jacobian has the asserted projector limit. Taking traces proves the response conclusion.

To make explicit what the density derivative contains, differentiation of the pullback at fixed x yields

$$\begin{aligned}\partial_t \log \bar{p}_t(x) = {} & \partial_t \log p^y_t(y)|_{y=\Phi_t(x)} + \partial_t \Phi_t(x)^\top \nabla_y \log p^y_t(\Phi_t(x)) \\ & + \partial_t \log |\det D\Phi_t(x)|.\end{aligned}$$

Thus both motion of the model query and change of volume enter the adapter. When a transition density Q^y_t is available and its posterior mean estimates the clean point in data coordinates, the corresponding objects are

$$\bar{Q}_t(\bar{y} \mid u) = Q^y_t(\Phi_t(\bar{y}) \mid u)|\det D\Phi_t(\bar{y})|, \qquad \bar{B}_t(\bar{y}) = B^y_t(\Phi_t(\bar{y})). \tag{35}$$

Class K is checked on this pulled-back channel. The Jacobian in the density cancels in the posterior quotient, which explains the second identity.

For example, an affine family $Y_t = c_t X + b_t Z$ with positive coefficients and independent $Z \sim K$ gives $\Phi_t(\bar{y}) = c_t \bar{y}$, $\bar{p}_t(\bar{y}) = c_t^D p^y_t(c_t \bar{y}) = \rho_{b_t/c_t}(\bar{y})$, and $w_t = 1$. This covers the coordinate conversion in both the diffusion and independent affine FM rows of Table 1. Each appendix derives the further identity needed to evaluate the width derivative using the available model field.

Finally, a regular change of clock $t = \eta(s)$ multiplies both $\partial_t \log \bar{p}_t(x)$ and κ_t by $\eta'(s)$ when expressed as derivatives in s. Their ratio is unchanged. The model time can therefore be reparametrized, while the calibration to a known physical noise width remains necessary.

D Geometric Consequences of the Master Theorem

The master theorem assumes a limiting geometry rather than a particular kind of support. This appendix verifies that geometry for smooth manifolds, rectifiable measures, and several kinds of intersections, and then computes what changes at a boundary. For a smooth interior point, the calculation below also proves the density and reconstruction limits directly from a local chart. Its purpose is to separate the geometric input—which data directions survive magnification—from the analytic conditions that allow the kernel and its derivatives to detect those directions.

Throughout the smooth-manifold argument below, x is an interior point of an embedded C^1 manifold $\mathcal{M}$ of dimension d, and, in an ambient neighborhood of x,

$$\mu = q\,\mathcal{H}^d \restriction_{\mathcal{M}}, \qquad q \text{ is continuous at } x, \qquad 0 < q(x) < \infty.$$

The measure μ is the probability law of the data. It may have other support outside this neighborhood. The direct calculations concern the additive family (1); the geometric conclusions also apply to class K through Theorem 2.2 and Corollary 2.3.

D.1 The tangent measure and the numerical local dimension

Translate x to the origin and choose orthonormal matrices $U \in \mathbb{R}^{D\times d}$ and $N \in \mathbb{R}^{D\times(D-d)}$ spanning $T = T_x\mathcal{M}$ and its orthogonal complement. The graph representation of an embedded C^1 manifold gives, after restricting the neighborhood,

$$F(v) = Uv + Nh(v), \qquad h(0) = 0, \qquad Dh(0) = 0, \qquad J_F(v) = \sqrt{\det(I_d + Dh(v)^\top Dh(v))}.$$

Here v ranges over an open neighborhood V of zero in $\mathbb{R}^d$. In particular, $J_F(v) \to 1$, $h(v) = o(\|v\|)$, and $\|F(v)\| \geq \|v\|$. This is the chart (11) with $\eta = Nh$.

Let f be continuous with support in an ambient ball of radius R. For sufficiently small ε, every data point contributing to $f(u/\varepsilon)$ lies in this chart. Changing variables $v = \varepsilon z$ in the area integral yields

$$\int f \, \mathrm{d}\nu_{0,\varepsilon} = \int_{V/\varepsilon} f\left(Uz + N\frac{h(\varepsilon z)}{\varepsilon}\right) q(F(\varepsilon z)) J_F(\varepsilon z) \, \mathrm{d}z.$$

The integrand vanishes unless $\|z\| \leq R$. On this fixed ball, $h(\varepsilon z)/\varepsilon \to 0$ uniformly, and the other two factors converge uniformly to $q(0)$ and 1. They are also uniformly bounded there. Dominated convergence therefore gives

$$\nu_{x,\varepsilon} \longrightarrow q(x)\mathcal{H}^d \restriction_T .$$

This proves (A1) along the full family of scales. The limiting measure is nonzero, locally finite, and d-homogeneous because d-dimensional volume on T scales by a^d under dilation by a.

The same chart identifies d with the numerical LID used in Appendix A. Apply the preceding change of variables to the indicator of $B(0, r)$ with $v = rz$. The indicator converges to that of the unit ball in $\mathbb{R}^d$ except on its boundary, which has zero d-dimensional volume; its support is contained in $\|z\| \leq 1$. Hence, with ω_d denoting the volume of that unit ball,

$$\mu(B(x, r)) = q(x)\omega_d r^d + o(r^d), \qquad \frac{\log \mu(B(x, r))}{\log r} \longrightarrow d. \tag{36}$$

Thus the exponent supplied by the tangent measure is the local dimension of the original data law, not a dimension introduced by smoothing. For $d = 0$, the same assertion is the atomic case: the local mass tends to the positive atom at x, and the logarithmic ratio tends to zero.

Tail control at these points. For completeness, the kernel classes of Appendix A admit a direct tail check here. Equation (36) gives $\mu(B(x, r)) \leq Cr^d$ for small r. Since μ is finite, increasing C gives this bound for every $r > 0$, and consequently

$$\nu_{x,\varepsilon}(B(0, R)) \leq CR^d \quad \text{for every } R > 0 \text{ and } \varepsilon > 0. \tag{37}$$

For a compactly supported kernel, W_K in (8) has compact support, so its tail condition is immediate. For a rapidly decaying kernel, $W_K(z) \leq C_1(1 + \|z\|)^p e^{-c\|z\|^\vartheta}$ for a finite p. Splitting the region $\|z\| > R$ into shells $n < \|z\| \leq n + 1$ and using (37) bounds its weighted mass, uniformly in ε, by a constant times

$$\sum_{n \geq \lfloor R \rfloor} (n + 2)^{p+d} e^{-cn^\vartheta},$$

which tends to zero. The same argument applies after multiplying K by $\|z\|^2$, when second moments are required. The positive limiting kernel mass in (A2) remains necessary: a nonnegative kernel can vanish on the whole tangent plane. In what follows,

$$c_{K,T} := \int_T K(z) \, \mathcal{H}^d(\mathrm{d}z) = \int_{\mathbb{R}^d} K(-Uz) \, \mathrm{d}z > 0.$$

D.2 Direct density and scale-derivative limits

The chart first gives an asymptotic density. The scale derivative must then be computed from the original convolution; it cannot be inferred by differentiating an unspecified $o(1)$ term.

Choose a fixed ambient neighborhood of zero in which the chart represents all of μ. Its contribution to the density is

$$\rho_\varepsilon(0) = \varepsilon^{d-D} \int_{V/\varepsilon} K\left(-Uz - N\frac{h(\varepsilon z)}{\varepsilon}\right) q(F(\varepsilon z)) J_F(\varepsilon z) \, \mathrm{d}z + \mathrm{Tail}_\varepsilon, \tag{38}$$

where, if necessary, the chart domain is restricted to that neighborhood and $\mathrm{Tail}_\varepsilon$ contains the remaining data. The following cutoff argument justifies this limit and all the moment limits used below without a uniform estimate on derivatives of q.

Write

$$I_\varepsilon(f) := \int f(z)\, \check{\nu}_{0,\varepsilon}(\mathrm{d}z).$$

The reflected tangent measure is again $q(0)\mathcal{H}^d \lfloor_T$. If f is continuous and $|f| \le C_f W_K$, insert a continuous cutoff χ_R equal to one on $B(0,R)$ and zero outside $B(0,2R)$. Local weak convergence proves convergence of $I_\varepsilon(\chi_R f)$. The omitted part is bounded uniformly by $C_f \int_{\|z\|>R} W_K \,\mathrm{d}\check{\nu}_{0,\varepsilon}$, which tends to zero by (A2). The corresponding limiting tail also tends to zero: apply local weak convergence to bounded outer cutoffs of $(1-\chi_R)W_K$ and then let their supports increase. Therefore

$$I_\varepsilon(f) \longrightarrow q(0) \int_T f \,\mathrm{d}\mathcal{H}^d \quad \text{whenever } f \text{ is continuous and } |f| \le C_f W_K. \tag{39}$$

For vector and matrix integrands this statement is applied entrywise. It covers K, zK, ∇K, $z\nabla K^\top$, and $-DK - z^\top \nabla K$.

In particular, the exact identity $\varepsilon^{D-d}\rho_\varepsilon(0) = I_\varepsilon(K)$ and (39) give

$$\varepsilon^{D-d}\rho_\varepsilon(x) \longrightarrow q(x) c_{K,T}. \tag{40}$$

This also controls the remainder in (38): points outside the fixed chart neighborhood have magnified distance at least δ/ε for some $\delta > 0$, so their normalized contribution is bounded by a tail in (A2) and tends to zero.

For the derivative, define the signed kernel

$$(\mathcal{D}K)(z) = -DK(z) - z^\top \nabla K(z).$$

Differentiating before rescaling, as permitted by the differentiation assumptions of Proposition 2.1, gives

$$\varepsilon \partial_\varepsilon \rho_\varepsilon(x) = \int \varepsilon^{-D} (\mathcal{D}K)((x-u)/\varepsilon)\, \mu(\mathrm{d}u). \tag{41}$$

For the two stated kernel classes, this differentiation has an explicit local domination: on any compact interval of positive widths, the derivative is bounded as a function of u by a compactly supported bounded function or by a polynomial times a decaying exponential. Either is bounded and thus integrable against the finite measure μ. Applying (39) to $\mathcal{D}K$ now reduces the derivative limit to an integral on T.

To evaluate that integral, put $k(v) = K(Uv)$. The plane restrictions of K, ∇K, and $\|z\|\,\|\nabla K\|$ are integrable, by the kernel conditions above or the limiting form of (A2). Integration by parts with a smooth cutoff $\chi_R(v)$ gives

$$\int \chi_R \partial_j k = -\int k\, \partial_j \chi_R,$$
$$\int \chi_R v_i \partial_j k = -\delta_{ij} \int \chi_R k - \int v_i k\, \partial_j \chi_R.$$

Choose $|\nabla \chi_R| \le C/R$ supported in $R \le \|v\| \le 2R$. The first right-hand remainder is bounded by $(C/R)\int_{\|v\|>R} k$, and the second by $2C \int_{\|v\|>R} k$; both vanish. Taking $R \to \infty$ therefore proves

$$\int \partial_j k = 0, \qquad \int v_i \partial_j k = -\delta_{ij} c_{K,T}. \tag{42}$$

Summing the second identity over $i = j$ yields

$$\int_T \mathcal{D}K \,\mathrm{d}\mathcal{H}^d = (d-D) c_{K,T}.$$

Consequently,

$$\varepsilon^{D-d} \varepsilon \partial_\varepsilon \rho_\varepsilon(x) \longrightarrow q(x)(d-D) c_{K,T}.$$

Dividing by (40), whose limit is positive, proves $\mathcal{R}_\varepsilon^{\mathrm{dens}}(x) \to d$. Only the kernel was differentiated; continuity of q at x was sufficient for its entire role in the argument.

D.3 Direct reconstruction and Jacobian limits

The derivative in the response is an observation derivative at the query point. It is enough to differentiate the posterior quotient there and pass its four integrals to the limit. No convergence of the denoisers in a whole neighborhood is needed.

Continue to set $x = 0$, and define the rescaled reconstruction wherever its denominator is positive by

$$H_\varepsilon(a) := \frac{b_\varepsilon(\varepsilon a)}{\varepsilon} = \frac{\int wK(a-w)\,\nu_{0,\varepsilon}(\mathrm{d}w)}{\int K(a-w)\,\nu_{0,\varepsilon}(\mathrm{d}w)}. \tag{43}$$

For comparison, the reconstruction associated with constant density on the tangent plane is

$$H_0(a) = U\frac{\int vK(a-Uv)\,\mathrm{d}v}{\int K(a-Uv)\,\mathrm{d}v}. \tag{44}$$

For the local and rapidly decaying kernel classes, its numerator and denominator are continuously differentiable near zero, with differentiation justified by their compact support or decaying majorants; the denominator is positive there because its value at zero is $c_{K,T} > 0$.

The quotient defining H_ε may be differentiated at zero under (A3). This assumption concerns both the density denominator and the vector first-moment numerator. Set

$$C_\varepsilon = I_\varepsilon(K), \qquad M_\varepsilon = I_\varepsilon(zK), \qquad G_\varepsilon = I_\varepsilon(\nabla K), \qquad L_\varepsilon = I_\varepsilon(z\nabla K^\top).$$

The sign change $z = -w$ in (43) shows that its numerator at zero is $-M_\varepsilon$, its numerator derivative is $-L_\varepsilon$, and its denominator derivative is G_ε. Thus the quotient rule and the chain rule give the exact identities

$$Jb_\varepsilon(0) = JH_\varepsilon(0) = -\frac{L_\varepsilon}{C_\varepsilon} + \frac{M_\varepsilon G_\varepsilon^\top}{C_\varepsilon^2}, \qquad \frac{b_\varepsilon(0)}{\varepsilon} = -\frac{M_\varepsilon}{C_\varepsilon}, \qquad \varepsilon\nabla\log\rho_\varepsilon(0) = \frac{G_\varepsilon}{C_\varepsilon}. \tag{45}$$

Every integral on the right converges by (39). The factor $q(0)$ cancels from the ratios. Writing

$$c = c_{K,T}, \quad M = \int_T zK\,\mathrm{d}\mathcal{H}^d, \quad G = \int_T \nabla K\,\mathrm{d}\mathcal{H}^d, \quad L = \int_T z\nabla K^\top\,\mathrm{d}\mathcal{H}^d,$$

we obtain $Jb_\varepsilon(x) \to -L/c + MG^\top/c^2$. For the local and rapidly decaying kernel classes, differentiating (44) identifies this limit with $JH_0(0)$ as well; the pointwise moment argument itself only uses (A2) and (A3).

It remains to identify this matrix. The vector M lies in T, and every normal row of L is zero. The first identity in (42) says that G is orthogonal to T, while the second says that the tangent-to-tangent block of L is $-cI_d$. In coordinates $T \oplus T^\perp$, the limit therefore has the form

$$A = \begin{pmatrix} I_d & B \\ 0 & 0 \end{pmatrix}.$$

Direct multiplication gives $A^2 = A$; its image is T because all outputs lie in T and A fixes every vector in T. It consequently has rank and trace d, proving the projector and response conclusions of Corollary 2.3 in the chart setting. The last two identities of (45) also give

$$(b_\varepsilon(x) - x)^\top\nabla\log\rho_\varepsilon(x) \longrightarrow -\frac{M^\top G}{c^2} = 0.$$

The correction vanishes because its two limiting vectors lie in orthogonal subspaces, not merely because the unscaled reconstruction displacement tends to zero.

If K is radially symmetric, its derivative in every normal direction vanishes on T; hence the normal columns of L vanish, and $G = 0$ by (42). Thus $B = 0$ and $A = P_T$ is the orthogonal projector. For a general admissible kernel, B can be nonzero, but the trace is still d.

These calculations establish both readouts for the additive model without any differentiability assumption on the sampling density beyond continuity at the query. For a class-K family, the same chart verifies the flat tangent required by Theorem 2.2 and Corollary 2.3; its kernel-matching and differentiation assumptions then provide the corresponding limits for ρ_ε^Q and B_ε^Q.

D.4 RECTIFIABLE MEASURES AND INTERSECTIONS OF FINITELY MANY COMPONENTS

The smooth chart provided a flat tangent, but the master theorem only needs the tangent measure itself. This distinction matters for data supported on nonsmooth sets and for mixtures: an exceptional point of the support can still have a well-defined homogeneous tangent, although that tangent need not be a single plane.

Rectifiable data. Let $E \subset \mathbb{R}^D$ be $\mathcal{H}^d$-measurable and countably d-rectifiable: up to an $\mathcal{H}^d$-null set, it is covered by countably many Lipschitz images of subsets of $\mathbb{R}^d$. Suppose

$$\mu = \theta \mathcal{H}^d \lfloor_E, \qquad 0 < \theta(x) < \infty \text{ on } E,$$

where θ is measurable and the weighted measure is locally finite. Zero-density sets can be discarded without changing μ; in this paper μ is a probability measure. The weighted tangent theorem for rectifiable measures states that, at μ-almost every x, there is a d-plane $T_x E$ such that

$$\varepsilon^{-d}(T_{x,\varepsilon})_{\#}\mu \longrightarrow \theta(x)\mathcal{H}^d \lfloor_{T_x E} \quad \text{locally weakly as } \varepsilon \downarrow 0. \tag{46}$$

The precise result used here is the weighted version in Simon (2018, Chapter 3, Definition 1.7 and Theorem 1.9). Its assumptions are countable rectifiability and local finiteness of $\theta\mathcal{H}^d \lfloor_E$; it does not require the unweighted area of E to be locally finite.

The density factor in (46) has a direct explanation on a piece with locally finite area. At an area-density point where E has tangent plane $T_x E$ and θ has its Lebesgue value, for $\operatorname{supp} f \subset B(0,R)$,

$$\begin{aligned}&\left|\varepsilon^{-d}\int_E f\left(\frac{u-x}{\varepsilon}\right)(\theta(u)-\theta(x))\,\mathcal{H}^d(\mathrm{d}u)\right| \\ &\qquad \le \|f\|_\infty \varepsilon^{-d}\int_{E\cap B(x,\varepsilon R)} |\theta(u)-\theta(x)|\,\mathrm{d}\mathcal{H}^d \longrightarrow 0.\end{aligned}$$

Indeed, the average absolute difference tends to zero, while the area of the ball divided by ε^d stays bounded. Multiplication of the unweighted tangent limit by $\theta(x)$ then gives the weighted limit. The cited weighted theorem supplies this conclusion also when finite weighted mass, rather than locally finite unweighted area, is available.

At each point of (46), approximation of the unit-ball indicator by continuous cutoffs gives

$$\mu(B(x,r)) = \theta(x)\omega_d r^d + o(r^d),$$

because the limiting plane measure gives zero mass to the unit sphere. Hence the numerical LID is d. This also gives the ball-growth bound (37). The shell argument of Appendix D proves weighted tail control for the local and rapidly decaying kernel classes, while positivity of the kernel mass on $T_x E$ must still be checked. More generally, one can impose (A2) directly for the chosen kernel. The cutoff argument (39) then upgrades (46) to convergence of the kernel integrals and their required moments.

It follows from Theorem 2.2 and Corollary 2.3 that, wherever the channel satisfies class K and reconstruction satisfies (A3), both readouts converge to d at these points and the reconstruction Jacobian converges to a projector onto $T_x E$. These are almost-everywhere statements for the weighted data law. They do not assign a tangent plane or a response limit to every corner or self-intersection of the set; such individual points require their own local analysis.

A component separated from the others. Consider a finite mixture $\mu = \sum_{j=1}^m \pi_j \mu_j$, with $\pi_j > 0$, $\sum_j \pi_j = 1$, and probability measures μ_j. If a neighborhood of $x \in \operatorname{supp}\mu_i$ meets no other component, every compactly supported test function in the magnified coordinates sees only $\pi_i\mu_i$ once ε is sufficiently small. Thus the mixture has the same tangent dimension as μ_i, with its tangent measure multiplied by π_i. The kernel-tail condition excludes any surviving weighted contribution from the other components. Both the numerical LID and the readout conclusions are therefore those of component i, subject to the same geometric and analytic assumptions as before.

Components of different dimensions meeting at the query. For a precise intersection statement, suppose the finite set of components through x have positive continuous densities q_j on embedded

C^1 manifolds of dimensions d_j near x. Components absent from a neighborhood of x can be handled as in the preceding paragraph. Define

$$d_* := \min_j d_j, \qquad I_* := \{j : d_j = d_*\}, \qquad \nu_j := q_j(x)\mathcal{H}^{d_j} \lfloor_{T_x\mathcal{M}_j} .$$

The smooth calculation proves $\varepsilon^{-d_j}(T_{x,\varepsilon})_{\#}\mu_j \to \nu_j$ and $\mu_j(B(x,r)) = q_j(x)\omega_{d_j} r^{d_j} + o(r^{d_j})$. For a compactly supported continuous f,

$$\int f \,\mathrm{d}\big[\varepsilon^{-d_*}(T_{x,\varepsilon})_{\#}\mu\big] = \sum_j \pi_j \varepsilon^{d_j - d_*} \int f \,\mathrm{d}\big[\varepsilon^{-d_j}(T_{x,\varepsilon})_{\#}\mu_j\big]$$
$$\longrightarrow \sum_{j \in I_*} \pi_j \int f \,\mathrm{d}\nu_j.$$

Each integral in the finite sum converges and is bounded; the factors with $d_j > d_*$ tend to zero. The mixture thus has the full, nonzero, d_*-homogeneous tangent

$$\nu_* := \sum_{j \in I_*} \pi_j \nu_j. \tag{47}$$

Similarly, adding the ball-mass expansions gives

$$\mu(B(x,r)) = r^{d_*}\left(\omega_{d_*} \sum_{j \in I_*} \pi_j q_j(x) + o(1)\right).$$

The constant is positive, so the numerical LID at this intersection is d_*. Under the master theorem's analytic assumptions, the density readout consequently converges to d_*. This argument uses a finite mixture with positive fixed weights and regular component densities; it makes no claim for accumulating strata or arbitrary infinite mixtures.

The response is determined by ν_* as well: higher-dimensional components have disappeared from the limiting posterior. If the minimum-dimensional tangent is one plane, including the case of a unique minimum-dimensional component, Corollary 2.3 gives response limit d_*. If several distinct minimum-dimensional planes remain, their sum need not be flat, so a projector conclusion requires further analysis. There is no ambiguity in the numerical LID defined here; the distinction concerns the shape of the limiting reconstruction response.

Homogeneous unions and the role of kernel symmetry. The density conclusion also applies to any specified nonzero homogeneous tangent, including a conical union of equal-dimensional pieces, whenever (A2) and the class-K conditions hold. Under (A3), Appendix C gives the exact limiting criterion

$$\lim_{\varepsilon \downarrow 0} \mathcal{R}^{\mathrm{resp}}_{\varepsilon}(x) = d + m_x^{\top} g_x, \qquad \lim_{\varepsilon \downarrow 0} \text{correction} = -m_x^{\top} g_x.$$

Here m_x and g_x are the posterior mean displacement and mean kernel score defined there. If both the tangent measure and K_x are invariant under $z \mapsto -z$, the limiting posterior is symmetric, so its integrable mean m_x is zero. Symmetry of the support alone is not sufficient for this argument when the kernel is asymmetric.

Even a correct trace need not imply a tangent projector when the limiting measure is not flat. For example, take equal tangent weights on the two coordinate lines in $\mathbb{R}^2$ and standard Gaussian noise. The limiting posterior is the equal mixture of centered one-dimensional standard Gaussians on those lines. Its covariance is

$$\tfrac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \tfrac{1}{2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \tfrac{1}{2} I_2.$$

By the Gaussian covariance identity this is the response-matrix limit. Its trace is the numerical LID, 1, but it has rank 2 and its square is not itself. This is why the flat-tangent condition in Corollary 2.3 is needed for the matrix conclusion.

What the geometric conclusions require. Table 4 summarizes the resulting cases. All readout entries assume the relevant kernel/channel and differentiation conditions; geometry alone does not replace them. A numerical local dimension without a full homogeneous tangent does not invoke the theorem: the explicit example in Appendix A.4 has a well-defined numerical dimension and an oscillating density readout. No blanket assertion about all fractal or self-similar supports is needed for this distinction.

Table 4: Geometric consequences under the stated analytic assumptions. The response is the trace of the posterior-mean Jacobian. A corrected response has the same limiting value as the density readout.

Geometry at x	Density readout	Response readout
Smooth manifold, interior point	d	d; matrix limit is a tangent projector
Countably d-rectifiable data	d at almost every point	d at the same regular points, under (A3)
Smooth boundary, Gaussian noise	d	$d - 2/\pi$; correction tends to $2/\pi$
General homogeneous tangent	d	$d + m_x^\top g_x$; correction tends to $-m_x^\top g_x$
Finite intersection of regular components	Smallest active dimension d_*	Determined by the minimum-dimensional tangent; d_* if it is flat
Numerical local dimension alone	Not guaranteed	Not guaranteed

D.5 Gaussian boundary bias and tangent cones

At an interior point, the limiting clean-data posterior extends across the whole tangent plane. A boundary removes one side of that posterior, moving its mean inward and reducing its variance in the missing direction. For Gaussian perturbations, the reconstruction response is exactly the posterior variance in noise units, while the correction is the squared posterior displacement in the same units. We now prove the boundary claim used in Section 2, including the geometric premise, and then show that the same calculation applies to any measurable tangent cone with positive Gaussian mass.

Why a smooth boundary gives a half-space tangent. Suppose that near x, $\mu = q\mathcal{H}^d \lfloor_{\mathcal{M}}$, where $\mathcal{M}$ is an embedded C^1 manifold with boundary, x is a boundary point, $d \geq 1$, and q is continuous at x with $q(x) > 0$. Choose a boundary chart on a neighborhood of zero in $\{v_1 \geq 0\} \subset \mathbb{R}^d$ such that

$$F(0) = x, \qquad DF(0) = U, \qquad U^\top U = I_d.$$

Such coordinates choose the first tangent direction inward and the remaining ones along the boundary. The C^1 property gives $F(v) = x + Uv + o(\|v\|)$, uniformly as $v \to 0$ in the half-space, and $J_F(v) \to 1$. After shrinking the chart, $\|F(v) - x\| \geq \|v\|/2$.

For a compactly supported continuous f, changing variables $v = \varepsilon z$ in the chart gives

$$\int f \, \mathrm{d}\nu_{x,\varepsilon} = \int_{\{z_1 \geq 0\}} f\left(\frac{F(\varepsilon z) - x}{\varepsilon}\right) q(F(\varepsilon z)) J_F(\varepsilon z) \, \mathrm{d}z,$$

with the rescaled chart domain understood. If $\operatorname{supp} f \subset B(0, R)$, the integrand is zero unless $\|z\| \leq 2R$, so dominated convergence proves

$$\nu_{x,\varepsilon} \longrightarrow q(x)\mathcal{H}^d \lfloor_C, \qquad C = U\{z : z_1 \geq 0\} \subset T_x\mathcal{M}. \tag{48}$$

Replacing f by the unit-ball indicator is valid by the same argument: its limiting discontinuity lies on a sphere of zero d-dimensional area. Consequently,

$$\mu(B(x, r)) = \tfrac{1}{2} q(x) \omega_d r^d + o(r^d).$$

The numerical LID is therefore d, and the local ball-growth bound holds. Since μ is finite, (37) and its shell proof apply; in particular they give the Gaussian-weighted second-moment condition below. Thus a regular smooth boundary with positive continuous density supplies the geometric and tail premises of the following corollary.

Corollary D.1 (Boundary bias)**.** *Consider the additive family* (1) *with the standard Gaussian kernel. Suppose the normalized blow-ups* (6) *converge to a positive constant multiple of $\mathcal{H}^d$ on a half-space in a d-dimensional plane, $d \geq 1$. Under Gaussian-weighted second-moment uniform integrability* (50)*,*

$$\mathcal{R}_\varepsilon^{\mathrm{resp}}(x) \longrightarrow d - \frac{2}{\pi}, \qquad \frac{\|b_\varepsilon(x) - x\|^2}{\varepsilon^2} \longrightarrow \frac{2}{\pi}, \qquad \mathcal{R}_\varepsilon^{\mathrm{dens}}(x) \longrightarrow d. \tag{49}$$

The result follows from a posterior limit on cones. We state that limit first so that the half-space calculation can be separated from the convergence argument.

Theorem D.2 (Gaussian decomposition on a tangent cone)**.** *For the additive standard Gaussian family, suppose*

$$\nu_{x,\varepsilon} \longrightarrow q_x \mathcal{H}^d \lceil_C \quad \textit{locally weakly}, \qquad q_x > 0,$$

where C is a measurable cone in a d-dimensional plane T: $aC = C$ for every $a > 0$. Assume its Gaussian mass is positive, $\alpha_C := \mathbb{P}(W \in C) > 0$ for $W \sim \mathcal{N}(0, I_T)$, and assume

$$\lim_{R\to\infty} \sup_{0<\varepsilon<\varepsilon_0} \int_{\|w\|>R} (1 + \|w\|^2) e^{-\|w\|^2/2} \nu_{x,\varepsilon}(\mathrm{d}w) = 0. \tag{50}$$

Let W_C have the conditional law of W given $W \in C$, regarded as a vector in $\mathbb{R}^D$. Then

$$\frac{\mathrm{Cov}(X \mid Y_\varepsilon = x)}{\varepsilon^2} \longrightarrow \mathrm{Cov}(W_C), \tag{51}$$

$$\frac{x - b_\varepsilon(x)}{\varepsilon} \longrightarrow -\mathbb{E}W_C, \tag{52}$$

and

$$\mathrm{tr}\,\mathrm{Cov}(W_C) + \|\mathbb{E}W_C\|^2 = \mathbb{E}\,\|W_C\|^2 = d. \tag{53}$$

In particular, the response tends to $d - \|\mathbb{E}W_C\|^2$, the correction tends to $\|\mathbb{E}W_C\|^2$, and the density readout tends to d.

No convexity assumption is needed. Positive Gaussian mass ensures that the conditional law is defined and that the limiting density normalization is nonzero. The specified tangent is a measure on a cone, not merely the existence of a set-theoretic tangent cone; both the mass distribution and its convergence are part of the hypothesis.

Proof of the posterior limit. Use the rescaled clean-point position $w = (X - x)/\varepsilon$. Its sign is opposite to the displacement $Z_x = (x - X)/\varepsilon$ used in Appendix B; this is why C is a cone for $\nu_{x,\varepsilon}$ rather than its reflection. The Gaussian likelihood at the observation x weights w by $e^{-\|w\|^2/2}$. Its exact posterior law is therefore

$$\mathbb{P}\left(\frac{X - x}{\varepsilon} \in \mathrm{d}w \,\middle|\, Y_\varepsilon = x\right) = \frac{e^{-\|w\|^2/2}\nu_{x,\varepsilon}(\mathrm{d}w)}{A_\varepsilon}, \qquad A_\varepsilon := \int e^{-\|w\|^2/2}\,\nu_{x,\varepsilon}(\mathrm{d}w). \tag{54}$$

The factors $\varepsilon^{d-D}(2\pi)^{-D/2}$ cancel between numerator and denominator.

Insert a continuous cutoff equal to one on $B(0, R)$ and supported in $B(0, 2R)$. Local weak convergence gives the limit of the cutoff integral, and (50) bounds the omitted part uniformly in ε. Letting $R \to \infty$ proves

$$A_\varepsilon \longrightarrow q_x \int_C e^{-\|w\|^2/2}\,\mathcal{H}^d(\mathrm{d}w) = q_x(2\pi)^{d/2}\alpha_C > 0.$$

The same cutoff argument applies to each entry of w and $ww^\top$, since their absolute values are bounded by a constant times $1 + \|w\|^2$. It also applies to bounded continuous test functions. Dividing their limiting integrals by the positive limiting denominator shows that the posterior in (54) converges to W_C, with convergence of its first and second moments.

Let $m_C = \mathbb{E}W_C$ and $S_C = \mathbb{E}[W_C W_C^\top]$. The posterior mean of w is $(b_\varepsilon(x) - x)/\varepsilon$, and its covariance is $\mathrm{Cov}(X \mid Y_\varepsilon = x)/\varepsilon^2$. Their limits are therefore m_C and $S_C - m_C m_C^\top$, proving the first two assertions, including the sign in the displacement formula.

The same positive normalization yields the density asymptotic

$$\rho_\varepsilon(x) = \varepsilon^{d-D}\left(q_x(2\pi)^{(d-D)/2}\alpha_C + o(1)\right).$$

Also, the limiting measure gives zero mass to the unit sphere, so continuous inner and outer cutoffs give $r^{-d}\mu(B(x, r)) \to q_x\mathcal{H}^d(C \cap B(0, 1)) > 0$. This identifies the exponent d with numerical LID. Neither density asymptotic will be differentiated to obtain the readout; its limit follows from the exact Gaussian identities below.

Why a cone preserves the radial second moment. For $d \geq 1$, write a standard Gaussian vector on T as $W = R\Theta$, with $R \geq 0$ and Θ on the unit sphere of T. In polar coordinates the Gaussian measure factors into a radial density proportional to $r^{d-1}e^{-r^2/2}$ and uniform spherical measure. Thus R and Θ are independent. Except at the origin, which has zero Gaussian mass, membership in C depends only on Θ because $aC = C$. Conditioning on C consequently leaves the radial law unchanged, and

$$\mathbb{E}\,\|W_C\|^2 = \frac{\int_0^\infty r^{d+1}e^{-r^2/2}\,\mathrm{d}r}{\int_0^\infty r^{d-1}e^{-r^2/2}\,\mathrm{d}r} = d.$$

For the last equality, integrate $r^d(re^{-r^2/2})$ by parts; the boundary term $r^d e^{-r^2/2}$ is zero at both endpoints. If $d = 0$, the Gaussian on $T = \{0\}$ is concentrated at zero and the same second-moment conclusion is immediate. The identity $\mathbb{E}\,\|W_C\|^2 = \mathrm{tr}(S_C - m_C m_C^\top) + \|m_C\|^2$ now proves (53).

From posterior moments to the readouts. All the finite-width Gaussian identities of Appendix B apply here. To check their regularity, take a compact interval of positive widths around ε and a bounded neighborhood of the observation. The Gaussian kernel, its scale and observation derivatives, and their products with u or $uu^\top$ are uniformly bounded functions of the clean point u over these widths and observations. They are therefore integrable against the probability measure μ, even if its unconditioned moments are infinite. Differentiation under the integral and the posterior moments needed below are valid.

The identities give

$$\begin{aligned} Jb_\varepsilon(x) &= \varepsilon^{-2}\,\mathrm{Cov}(X \mid Y_\varepsilon = x), \\ \nabla \log \rho_\varepsilon(x) &= \varepsilon^{-2}(b_\varepsilon(x) - x), \\ \mathcal{R}^{\mathrm{dens}}_\varepsilon(x) &= \varepsilon^{-2}\mathbb{E}[\|X - x\|^2 \mid Y_\varepsilon = x]. \end{aligned}$$

Taking the trace in the first line and the posterior limits proved above yields the response limit $\mathrm{tr}\,\mathrm{Cov}(W_C)$. The correction is exactly $\|b_\varepsilon(x) - x\|^2/\varepsilon^2$ and tends to $\|m_C\|^2$. The last line and the radial second moment give the density-readout limit d. This proves Theorem D.2.

Proof of the boundary corollary. Let $n \in T$ be the inward unit vector perpendicular to the boundary within the tangent plane. Choose an orthonormal basis of T with first vector n, so the half-space is $\{w_1 \geq 0\}$. The first coordinate of W_C is a standard normal conditioned to be nonnegative, with density $\sqrt{2/\pi}\,e^{-t^2/2}$ for $t \geq 0$. Its first two moments are

$$\begin{aligned} \mathbb{E}(W_C)_1 &= \sqrt{\frac{2}{\pi}}\int_0^\infty te^{-t^2/2}\,\mathrm{d}t = \sqrt{\frac{2}{\pi}}, \\ \mathbb{E}[(W_C)_1^2] &= \sqrt{\frac{2}{\pi}}\int_0^\infty t^2e^{-t^2/2}\,\mathrm{d}t = 1. \end{aligned}$$

The second integral follows by integration by parts from $\int_0^\infty e^{-t^2/2}\,\mathrm{d}t = \sqrt{\pi/2}$. Conditioning only this coordinate leaves the remaining $d-1$ independent coordinates centered standard Gaussian. In ambient coordinates, therefore,

$$\mathbb{E}W_C = \sqrt{\frac{2}{\pi}}\,n, \qquad \mathrm{Cov}(W_C) = P_T - \frac{2}{\pi}nn^\top. \tag{55}$$

Its trace is $d - 2/\pi$, and the squared norm of its mean is $2/\pi$. Substitution into Theorem D.2 proves all three limits of Corollary D.1. Notice that the limiting response matrix is not a projector: its eigenvalue in the inward direction is $1 - 2/\pi$, while the other tangent eigenvalues are one.

Corners and the meaning of the deficit. For an orthant constraining k orthogonal coordinates in a d-plane, the same independent-coordinate calculation gives

$$\|\mathbb{E}W_C\|^2 = \frac{2k}{\pi}, \qquad \mathrm{tr}\,\mathrm{Cov}(W_C) = d - \frac{2k}{\pi}.$$

A half-line has $d = k = 1$; a quadrant at a corner of a surface has $d = k = 2$. For a general cone, the deficit is determined by its conditional Gaussian mean rather than by a count of faces. It is always nonnegative and is positive precisely when that mean is nonzero.

These conclusions concern the posterior-mean response $\operatorname{tr} Jb_\varepsilon$ and any model expression already proved to equal that response. They are not a statement about the divergence of an arbitrary field. Keeping the Gaussian correction restores the full density readout at a boundary or corner. The deficit persists for the exact posterior, so its origin is the local support geometry rather than an error in learning the reconstruction.

E Diffusion Models

The diffusion readout uses the score of a Gaussian forward marginal. Its derivation has two steps: express the Gaussian density readout through the score, then convert from the model's coordinates to the data coordinates used by the master theorem. The same calculation also gives formulas for models that predict the noise or the clean point.

E.1 The forward marginal and its physical scale

Let $X \sim \mu$ and

$$Y_t = \psi_t X + \sigma_t Z, \qquad Z \sim \mathcal{N}(0, I_D), \qquad Z \perp X, \qquad \psi_t, \sigma_t > 0.$$

Write p_t for the density of Y_t and $s_t(y) = \nabla_y \log p_t(y)$ for its exact score. Dividing the observation by ψ_t gives the additive channel

$$\overline{Y}_t = X + \varepsilon(t) Z, \qquad \varepsilon(t) = \frac{\sigma_t}{\psi_t}.$$

Thus the native query corresponding to a data point x is $y_t = \psi_t x$. The change of variables gives the exact density relation

$$p_t(y) = \psi_t^{-D} \rho_{\varepsilon(t)}(y/\psi_t). \tag{56}$$

All observation derivatives below hold t fixed; the substitution $y = y_t$ is made after differentiation. The readout uses no time derivative of the score field.

Corollary E.1 (Diffusion formula)**.** *Under (A1)–(A2) for the standard Gaussian kernel, as $\varepsilon(t) \to 0$,*

$$\mathcal{R}^{\mathrm{dens}}_{\varepsilon(t)}(x) = D + \sigma_t^2 \big(\operatorname{div}_y s_t(\psi_t x) + \|s_t(\psi_t x)\|^2\big) \longrightarrow d. \tag{57}$$

The identity holds at every positive width. The limit is the Gaussian instance of the master theorem, and the expression is FLIPD (Kamkari et al., 2024b).

E.2 Derivation through reconstruction and change of coordinates

First work in data coordinates with $\overline{Y} = X + \varepsilon Z$. Write $s_\varepsilon = \nabla \log \rho_\varepsilon$ and $b_\varepsilon(\bar{y}) = \mathbb{E}[X \mid \overline{Y} = \bar{y}]$. The Gaussian likelihood satisfies

$$\nabla_{\bar{y}} \log K_\varepsilon(\bar{y} - u) = \frac{u - \bar{y}}{\varepsilon^2}.$$

Differentiating the convolution and dividing by $\rho_\varepsilon(\bar{y})$ therefore gives the conditional mean of $(X - \bar{y})/\varepsilon^2$. Differentiating the first-moment numerator as in Appendix B then gives the conditional covariance. These are the identities

$$b_\varepsilon(\bar{y}) = \bar{y} + \varepsilon^2 s_\varepsilon(\bar{y}), \tag{58}$$

$$Jb_\varepsilon(\bar{y}) = \varepsilon^{-2} \operatorname{Cov}(X \mid \overline{Y} = \bar{y}). \tag{59}$$

They hold for any probability law μ. Indeed, in a bounded observation neighborhood and a compact interval of positive widths, the Gaussian kernel and all derivatives used here, multiplied by any of the required first or second moments of u, are bounded functions of u. They provide integrable majorants against μ, and the posterior denominator is positive.

Differentiating (58) and taking the trace gives $\operatorname{tr} Jb_\varepsilon = D + \varepsilon^2 \operatorname{div} s_\varepsilon$. The reconstruction displacement is $b_\varepsilon - \bar{y} = \varepsilon^2 s_\varepsilon$, so the correction in Proposition 2.1 is $\varepsilon^2 \|s_\varepsilon\|^2$. Consequently,

$$\mathcal{R}^{\mathrm{dens}}_\varepsilon(\bar{y}) = D + \varepsilon^2 \big(\operatorname{div} s_\varepsilon(\bar{y}) + \|s_\varepsilon(\bar{y})\|^2\big) \tag{60}$$

$$= \frac{\operatorname{tr} \operatorname{Cov}(X \mid \overline{Y} = \bar{y})}{\varepsilon^2} + \frac{\|\bar{y} - b_\varepsilon(\bar{y})\|^2}{\varepsilon^2}. \tag{61}$$

The full density readout is the posterior second moment about the query in units of ε^2; the response is its variance component.

It remains to express (60) through the native score. At fixed t, differentiating (56) with respect to $\bar{y}$ gives

$$s_{\varepsilon(t)}(\bar{y}) = \psi_t s_t(\psi_t \bar{y}),$$
$$J_{\bar{y}} s_{\varepsilon(t)}(\bar{y}) = \psi_t^2 J_y s_t(\psi_t \bar{y}).$$

Set $\bar{y} = x$ and use $\varepsilon(t)^2 \psi_t^2 = \sigma_t^2$ in (60); this proves the finite-width identity (57). The normalized channel is additive Gaussian and has zero class-K errors. Its scale differentiation is legitimate by the same Gaussian domination, and (A1)–(A2) give the limit d through Theorem 2.2.

E.3 Noise prediction, clean-point prediction, and the response alone

Two equivalent model outputs are the exact noise prediction and clean-point prediction,

$$n_t(y) = \mathbb{E}[Z \mid Y_t = y], \qquad r_t(y) = \mathbb{E}[X \mid Y_t = y].$$

Posterior averaging of the native likelihood score gives $s_t = -n_t/\sigma_t$, while the conditional mean of $Y_t = \psi_t X + \sigma_t Z$ gives

$$r_t(y) = \frac{y - \sigma_t n_t(y)}{\psi_t} = \frac{y + \sigma_t^2 s_t(y)}{\psi_t} = b_{\varepsilon(t)}(y/\psi_t).$$

Hence, at $y_t = \psi_t x$,

$$\mathcal{R}^{\mathrm{dens}}_{\varepsilon(t)}(x) = D - \sigma_t \operatorname{div}_y n_t(y_t) + \|n_t(y_t)\|^2, \tag{62}$$

$$= \psi_t \operatorname{tr} J_y r_t(y_t) + \frac{\|\psi_t r_t(y_t) - y_t\|^2}{\sigma_t^2}. \tag{63}$$

The factor ψ_t in the second formula follows from differentiating $b_{\varepsilon(t)}(x) = r_t(\psi_t x)$ with respect to the data-coordinate query.

If the tangent measure is flat, Gaussian regularity supplies (A3), and Corollary 2.3 gives

$$D + \sigma_t^2 \operatorname{div}_y s_t(y_t) = D - \sigma_t \operatorname{div}_y n_t(y_t) = \psi_t \operatorname{tr} J_y r_t(y_t) \longrightarrow d.$$

The nonnegative correction in (57) then tends to zero. For other homogeneous tangents the full readout retains that correction; at a regular smooth boundary its limit is $2/\pi$, as proved in Appendix D.5.

F Independent Affine Flow Matching

For an independent affine interpolant, the velocity determines the posterior-mean reconstruction. This provides a response readout for a general admissible source kernel. A Gaussian source additionally makes the score a function of the same velocity, giving the full density readout from velocity alone. We derive both statements, then specialize them to rectified flow and relate the response to conditional flow-matching risk.

F.1 The interpolant and its normalized channel

Let $X \sim \mu$, let $Z \perp X$ have a nonnegative C^1 density K, and consider

$$Y_t = \alpha_t X + \beta_t Z, \qquad v_t^*(y) = \mathbb{E}[\dot{\alpha}_t X + \dot{\beta}_t Z \mid Y_t = y].$$

The schedules are differentiable and satisfy $\alpha_t, \beta_t > 0$ at the times used. Write p_t for the density of Y_t, and define

$$\varepsilon(t) = \frac{\beta_t}{\alpha_t}, \qquad a_t = \frac{\dot{\alpha}_t}{\alpha_t}, \qquad \kappa_t = \frac{\dot{\varepsilon}(t)}{\varepsilon(t)} = \frac{\dot{\beta}_t}{\beta_t} - a_t \neq 0. \tag{64}$$

The normalized observation $\bar{y} = y/\alpha_t$ belongs to the additive channel $X + \varepsilon(t)Z$, so

$$p_t(y) = \alpha_t^{-D} \rho_{\varepsilon(t)}(y/\alpha_t), \qquad y_t = \alpha_t x$$

is the native query corresponding to a fixed data point x. All spatial derivatives below are with respect to y at fixed t unless indicated otherwise, and formulas are evaluated where $p_t(y) > 0$. The finite-width identities require finite conditional first moments and legitimate scale and observation differentiation as specified in Appendices A and B. For the Gaussian source these operations are valid for every probability law μ; for a general K they remain part of the assumptions.

Corollary F.1 (Affine flow matching)**.** *Under (A1)–(A3), if $\varepsilon(t) \to 0$ and the tangent at x is flat,*

$$\mathcal{R}_t^{\mathrm{FM}}(x) := D + \frac{Da_t - \operatorname{div}_y v_t^*(\alpha_t x)}{\kappa_t} \longrightarrow d. \tag{65}$$

For a general homogeneous tangent, under (A1)–(A2) and the stated finite-width regularity, the density readout is, with $y = \alpha_t x$ and $s_t = \nabla_y \log p_t$,

$$\mathcal{R}_{\varepsilon(t)}^{\mathrm{dens}}(x) = \mathcal{R}_t^{\mathrm{FM}}(x) + \frac{(a_t y - v_t^*(y))^\top s_t(y)}{\kappa_t} \longrightarrow d. \tag{66}$$

F.2 Reconstruction, velocity, and the correction

Fix a time and abbreviate the conditional means in normalized coordinates by

$$h(\bar{y}) = \mathbb{E}[X \mid X + \varepsilon(t)Z = \bar{y}] = b_{\varepsilon(t)}(\bar{y}), \qquad m(\bar{y}) = \mathbb{E}[Z \mid X + \varepsilon(t)Z = \bar{y}].$$

Taking the conditional mean of the normalized observation gives

$$h(\bar{y}) = \bar{y} - \varepsilon(t) m(\bar{y}). \tag{67}$$

The velocity is the conditional mean of the path derivative, and hence

$$\begin{aligned} v_t^*(y) &= \dot{\alpha}_t h(\bar{y}) + \dot{\beta}_t m(\bar{y}) \\ &= a_t y + \big(\dot{\beta}_t - a_t \beta_t\big) m(\bar{y}) = a_t y + \beta_t \kappa_t m(\bar{y}). \end{aligned} \tag{68}$$

Eliminating m with (67) proves the reconstruction identity

$$v_t^*(y) = a_t y + \alpha_t \kappa_t \big(\bar{y} - b_{\varepsilon(t)}(\bar{y})\big), \qquad \bar{y} = y/\alpha_t. \tag{69}$$

In particular, the posterior mean can be read directly from velocity:

$$\mathbb{E}[X \mid Y_t = y] = \frac{(a_t + \kappa_t) y - v_t^*(y)}{\alpha_t \kappa_t}.$$

The condition $\kappa_t \neq 0$ is needed for this inversion. When the physical noise width is instantaneously constant, the velocity identity reduces to $v_t^*(y) = a_t y$ and does not reveal the reconstruction.

Differentiate (69) in the native observation. Since $J_y \bar{y} = \alpha_t^{-1} I_D$, the two coordinate factors cancel and

$$J_y v_t^*(y) = a_t I_D + \kappa_t \big(I_D - J b_{\varepsilon(t)}(\bar{y})\big).$$

Solving for the reconstruction Jacobian and taking its trace gives

$$\operatorname{tr} J b_{\varepsilon(t)}(y/\alpha_t) = D + \frac{Da_t - \operatorname{div}_y v_t^*(y)}{\kappa_t}. \tag{70}$$

This identity holds for every source kernel satisfying the stated regularity. At $y = \alpha_t x$, its left-hand side is the response readout in Section 2. The flat-tangent conclusion of Corollary 2.3 therefore proves (65); the matrix identity likewise recovers the limiting tangent projector, which may be oblique for a nonradial kernel.

To identify the full density readout, differentiate the density change of variables at fixed t:

$$\nabla_{\bar{y}} \log \rho_{\varepsilon(t)}(\bar{y}) = \alpha_t s_t(y).$$

The displacement in (69) is

$$b_{\varepsilon(t)}(\bar{y}) - \bar{y} = \frac{a_t y - v_t^*(y)}{\alpha_t \kappa_t}.$$

Their inner product is $(a_t y - v_t^*(y))^\top s_t(y)/\kappa_t$. Substitution into the exact additive decomposition of Proposition 2.1, together with (70), proves the equality in (66). The density conclusion of Theorem 2.2 gives its limit d. This proof differentiates only in the observation and the physical noise width; it does not differentiate the velocity along the moving query y_t.

The independence assumption identifies the normalized interpolant with the additive channel. The formulas here concern that channel and the exact conditional velocity defined above; a different coupling or a nonlinear interpolant requires its own relation to the master theorem.

F.3 Gaussian velocity formulas and conditional risk

Now let $Z \sim \mathcal{N}(0, I_D)$. At a native observation y, the conditional likelihood score is

$$\nabla_y \log p_t(y \mid X) = -\frac{y - \alpha_t X}{\beta_t^2}.$$

Posterior averaging gives $s_t(y) = -m(\bar{y})/\beta_t$. Inserting (68) yields

$$s_t(y) = \frac{a_t y - v_t^*(y)}{\beta_t^2 \kappa_t}. \tag{71}$$

Thus the correction in (66) can also be evaluated from velocity. At $y = \alpha_t x$,

$$\mathcal{R}^{\mathrm{dens}}_{\varepsilon(t)}(x) = \mathcal{R}^{\mathrm{FM}}_t(x) + \frac{\left\| v_t^*(y) - a_t y \right\|^2}{\beta_t^2 \kappa_t^2}. \tag{72}$$

The sign of κ_t does not change the sign of this Gaussian correction. Equivalently, using the reconstruction displacement,

$$\mathcal{R}^{\mathrm{dens}}_{\varepsilon(t)}(x) = D + \frac{D a_t - \operatorname{div}_y v_t^*(\alpha_t x)}{\kappa_t} + \frac{\left\| x - b_{\varepsilon(t)}(x) \right\|^2}{\varepsilon(t)^2}. \tag{73}$$

This is the velocity form of FLIPD; the corresponding connection for Gaussian flow matching was studied by Yeats et al. (2025).

The response component has a second interpretation in terms of the conditional regression problem defining flow matching. Let

$$U_t = \dot{\alpha}_t X + \dot{\beta}_t Z, \qquad \ell_t^*(y) = \mathbb{E}\left[\left\| U_t - v_t^*(y) \right\|^2 \,\middle|\, Y_t = y \right].$$

The quantity ℓ_t^* is the conditional risk at the conditional mean, rather than an error of an approximate velocity model.

Theorem F.2 (Gaussian covariance and conditional-risk representations)**.** *For the independent affine Gaussian family at a positive width with $\kappa_t \neq 0$,*

$$\frac{\operatorname{Cov}(X \mid Y_t = y)}{\varepsilon(t)^2} = I_D + \frac{a_t I_D - J_y v_t^*(y)}{\kappa_t}, \tag{74}$$

$$\frac{\ell_t^*(y)}{\beta_t^2 \kappa_t^2} = D + \frac{D a_t - \operatorname{div}_y v_t^*(y)}{\kappa_t}. \tag{75}$$

Proof. The observation $Y_t = y$ is equivalent to $X + \varepsilon(t) Z = y/\alpha_t$. By (59), the left-hand side of (74) is $J b_{\varepsilon(t)}(y/\alpha_t)$. The matrix form of (70) proves the first equality.

Conditional on $Y_t = y$, eliminate $Z = (y - \alpha_t X)/\beta_t$ from the target velocity:

$$U_t = \left(\dot{\alpha}_t - \alpha_t \frac{\dot{\beta}_t}{\beta_t} \right) X + \frac{\dot{\beta}_t}{\beta_t} y = -\alpha_t \kappa_t X + \frac{\dot{\beta}_t}{\beta_t} y.$$

Subtracting its conditional mean gives $U_t - v_t^*(y) = -\alpha_t \kappa_t (X - \mathbb{E}[X \mid Y_t = y])$. Therefore

$$\ell_t^*(y) = \alpha_t^2 \kappa_t^2 \operatorname{tr} \operatorname{Cov}(X \mid Y_t = y).$$

Divide by $\beta_t^2 \kappa_t^2$ and use $\varepsilon(t) = \beta_t/\alpha_t$; the trace of (74) proves (75). All the conditional moments are finite by Gaussian smoothing, without an unconditional second-moment assumption on X.

For any deterministic prediction $w \in \mathbb{R}^D$ at this observation, expansion of the square gives

$$\mathbb{E}[\left\| U_t - w \right\|^2 \mid Y_t = y] = \ell_t^*(y) + \left\| v_t^*(y) - w \right\|^2,$$

because the conditional mean of $U_t - v_t^*(y)$ is zero. This proves that ℓ_t^* is the irreducible conditional risk. Its normalized value equals the response in (75); under a flat tangent and (A1)–(A3), this value tends to d at $y = \alpha_t x$, and the matrix in (74) tends to the orthogonal tangent projector.

F.4 RECTIFIED FLOW

With a standard Gaussian source, use the convention in which the path moves from noise to data:

$$Y_t = tX + (1-t)Z, \qquad 0 < t < 1, \qquad v_t^*(y) = \mathbb{E}[X - Z \mid Y_t = y], \qquad t \uparrow 1.$$

Here

$$\alpha_t = t, \qquad \beta_t = 1 - t, \qquad a_t = \frac{1}{t}, \qquad \kappa_t = -\frac{1}{t(1-t)}, \qquad \varepsilon(t) = \frac{1-t}{t}, \qquad y_t = tx.$$

Substituting these coefficients into Theorem F.2 gives

$$t\big[I_D + (1-t)J_y v_t^*(y)\big] = \frac{\mathrm{Cov}(X \mid Y_t = y)}{((1-t)/t)^2}, \tag{76}$$

$$t\big[D + (1-t)\,\mathrm{div}_y\, v_t^*(y)\big] = t^2 \ell_t^*(y). \tag{77}$$

At $y = tx$ and a flat tangent satisfying the theorem's assumptions, the matrix in (76) converges to P_T and its trace converges to d.

The full readout follows by retaining the Gaussian displacement correction. The path identity $X = Y_t + (1-t)(X - Z)$ gives, after conditioning,

$$\mathbb{E}[X \mid Y_t = y] = y + (1-t)v_t^*(y).$$

At $y = tx$, this is $b_{\varepsilon(t)}(x) = tx + (1-t)v_t^*(tx)$, so

$$\frac{\left\|x - b_{\varepsilon(t)}(x)\right\|^2}{\varepsilon(t)^2} = t^2 \left\|v_t^*(tx) - x\right\|^2.$$

Combining this with the response term proves

$$\mathcal{R}^{\mathrm{dens}}_{\varepsilon(t)}(x) = t\big[D + (1-t)\,\mathrm{div}_y\, v_t^*(tx)\big] + t^2 \left\|v_t^*(tx) - x\right\|^2. \tag{78}$$

The full expression converges to d under (A1)–(A2), including the homogeneous boundary geometries covered by the master theorem. Its response component alone has that limit on a flat tangent. These formulas use the native query tx and the specified noise-to-data schedule; both enter the derivation.

G GENERALIZED BROWNIAN SCHRÖDINGER BRIDGES

The bridge formulas in Section 3 follow from a Gaussian posterior representation of the terminal point. The measure entering that representation is the terminal potential measure, which need not be a probability measure. We first specify the bridge construction and justify its exact reconstruction identities, then show that its potential and terminal data law have the same local geometry. This proves the forward-drift formula. Passing to current velocity requires an additional regularity condition on the opposite potential.

G.1 BRIDGE CONSTRUCTION AND TERMINAL POTENTIAL

Fix $T > 0$ and a constant diffusivity $\gamma > 0$, and write $\tau = T - t$ for $0 < t < T$. Let F, G be nonzero nonnegative locally finite Radon measures such that

$$\pi^\star(\mathrm{d}x_0, \mathrm{d}x_T) = h_{\gamma T}(x_T - x_0)\, F(\mathrm{d}x_0)\, G(\mathrm{d}x_T), \qquad h_a(z) = \frac{e^{-\|z\|^2/(2a)}}{(2\pi a)^{D/2}}, \tag{79}$$

has total mass one. The generalized bridge is the mixture of Brownian bridges of diffusivity γ with endpoint law $\pi^\star$. This construction accommodates singular endpoint laws; it does not assume that their path-space relative entropy with respect to Brownian motion is finite (Pooladian & Niles-Weed, 2025). We assume this factorized coupling exists, rather than asserting such a representation for an arbitrary dependent coupling.

Define the heat potentials and terminal law by

$$\widehat{\varphi}_t = h_{\gamma t} * F, \qquad \varphi_t = h_{\gamma\tau} * G, \qquad p_t = \widehat{\varphi}_t \varphi_t, \qquad \mu_T = \widehat{\varphi}_T G. \tag{80}$$

At the queries under consideration these potentials are finite and positive. Their differentiations below are understood in a terminal time neighborhood where the corresponding heat integrals and differentiated integrals exist. A concrete sufficient check for G is

$$\int h_{a_*}(x-z)\,G(\mathrm{d}z) < \infty \quad \text{for some } a_* > 0.$$

For widths $a < a_*$, locally in a and in y near x, the Gaussian kernel and its first two observation derivatives, its variance derivative, and its products with z or $zz^\top$ are bounded by a constant multiple of $h_{a_*}(x-z)$. The bound follows by absorbing the polynomial derivative factors and bounded shifts of y into the strictly slower Gaussian decay at variance a_*. It justifies differentiation under the integral and finite conditional second moments at all sufficiently small positive widths. The analogous integrability is assumed for F whenever its heat derivatives are used. Endpoint continuity of $\widehat{\varphi}_T$ alone will not be used to infer bounded endpoint derivatives.

To obtain the joint law needed for reconstruction, condition a Brownian bridge on its endpoints. Its density at time t is

$$\frac{h_{\gamma t}(y-x_0)h_{\gamma\tau}(x_T-y)}{h_{\gamma T}(x_T-x_0)}.$$

Multiplication by (79) cancels the denominator, and integration over x_0 gives

$$\mathbb{P}(X_t \in \mathrm{d}y, X_T \in \mathrm{d}z) = \widehat{\varphi}_t(y)h_{\gamma\tau}(z-y)\,\mathrm{d}y\,G(\mathrm{d}z). \tag{81}$$

Tonelli's theorem and the heat semigroup identity show that its marginal density is p_t in (80), with unit integral, and its terminal marginal is μ_T. In particular, the exact terminal posterior is

$$\Pi_t(\mathrm{d}z \mid y) = \frac{h_{\gamma\tau}(z-y)\,G(\mathrm{d}z)}{\varphi_t(y)}, \qquad B_t(y) = \mathbb{E}[X_T \mid X_t = y] = \int z\,\Pi_t(\mathrm{d}z \mid y). \tag{82}$$

The posterior is normalized even if $G(\mathbb{R}^D) = \infty$. It is a Gaussian kernel reweighting of G, not the posterior of an independent-noise channel with prior μ_T.

For a finite history ending at $X_t = y$, the same cancellation gives the conditional transition density $h_{\gamma(u-t)}(z-y)\varphi_u(z)/\varphi_t(y)$ for $t < u < T$. Its integral is one by the heat semigroup identity, and it depends on the history only through y. Thus the factorized mixture is Markov.

Use the forward and backward drift conventions

$$b_t^+ = \gamma\nabla\log\varphi_t, \qquad b_t^- = -\gamma\nabla\log\widehat{\varphi}_t, \qquad v_t = \tfrac{1}{2}(b_t^+ + b_t^-), \qquad s_t = \nabla\log p_t.$$

For the factorized mixture these are the usual drift fields. For example, the forward drift of a Brownian bridge pinned at $X_T = z$ is $(z-y)/\tau$; averaging over (82) gives $(B_t(y)-y)/\tau = \gamma\nabla\log\varphi_t(y)$, as also verified directly below.

G.2 Terminal LID formulas

Corollary G.1 (Terminal bridge estimator)**.** *In the factorized Brownian setting above, suppose $\widehat{\varphi}_T$ is positive and continuous at x, and the locally finite measure G satisfies (A1)–(A2) for the standard Gaussian kernel. Assume its convolutions are finite and positive at the query, its conditional second moments are finite, and heat differentiation and two observation derivatives may be taken under the integrals as described above. Then d is the LID of μ_T at x, and, as $t \uparrow T$,*

$$D + \tau\,\mathrm{div}\,b_t^+(x) + \frac{\tau}{\gamma}\left\|b_t^+(x)\right\|^2 \longrightarrow d. \tag{83}$$

If $\partial_t \log\widehat{\varphi}_t(x)$ remains bounded, then

$$D + 2\tau\big(\mathrm{div}\,v_t(x) + v_t(x)^\top s_t(x)\big) \longrightarrow d. \tag{84}$$

For a flat tangent and the stated Gaussian second-moment control, $D + \tau\,\mathrm{div}\,b_t^+(x) \to d$. If $b_t^-(x)$ and $Jb_t^-(x)$ also remain bounded, then $D + 2\tau\,\mathrm{div}\,v_t(x) \to d$.

The physical width in this corollary is $\varepsilon = \sqrt{\gamma\tau}$, and every displayed readout is evaluated at the fixed terminal data point x. The full forward expression uses only b_t^+ and its divergence. A flat tangent permits omission of its nonnegative squared-drift correction. The full current-velocity expression instead uses both v_t and the marginal score s_t.

G.3 Proof through exact reconstruction and heat identities

Write $a = \gamma\tau$. Differentiation of the Gaussian likelihood in (82) gives

$$\nabla_y h_a(z-y) = \frac{z-y}{a}\, h_a(z-y).$$

Integrating and dividing by $\varphi_t(y)$ therefore yields

$$\nabla \log \varphi_t(y) = \frac{\mathbb{E}[X_T - y \mid X_t = y]}{\gamma\tau}. \tag{85}$$

For the Jacobian, differentiate the ratio defining B_t. With the convention $(JB)_{ij} = \partial_{y_j} B_i$, the numerator derivative is $a^{-1}\int z(z-y)^\top h_a(z-y)\, G(\mathrm{d}z)$. Subtracting the denominator term gives

$$\begin{aligned} JB_t(y) &= a^{-1}\left(\mathbb{E}[X_T(X_T-y)^\top \mid X_t = y] - B_t(y)\mathbb{E}[X_T - y \mid X_t = y]^\top\right) \\ &= \frac{\mathrm{Cov}(X_T \mid X_t = y)}{\gamma\tau}. \end{aligned} \tag{86}$$

Consequently,

$$B_t(y) = y + \tau b_t^+(y), \qquad JB_t(y) = I_D + \tau J b_t^+(y) = \frac{\mathrm{Cov}(X_T \mid X_t = y)}{\gamma\tau}. \tag{87}$$

These are exact identities of the terminal posterior (Pooladian & Niles-Weed, 2025); no endpoint limit has been taken.

Taking the trace identifies the response, while the posterior mean identifies the correction:

$$\operatorname{tr} JB_t(y) = D + \tau \operatorname{div} b_t^+(y), \qquad \frac{\|B_t(y) - y\|^2}{\gamma\tau} = \frac{\tau}{\gamma}\left\|b_t^+(y)\right\|^2.$$

Their sum is the posterior second moment about the observation, since variance plus squared mean equals that second moment. The same sum has a density interpretation. The identity $\Delta\varphi/\varphi = \Delta\log\varphi + \|\nabla\log\varphi\|^2$ and the backward heat equation give

$$D + \tau \operatorname{div} b_t^+(y) + \frac{\tau}{\gamma}\left\|b_t^+(y)\right\|^2 = D + \gamma\tau\frac{\Delta\varphi_t(y)}{\varphi_t(y)}, \tag{88}$$

$$= D - 2\tau\partial_t \log\varphi_t(y), \tag{89}$$

where $\partial_t\varphi_t = -(\gamma/2)\Delta\varphi_t$ follows by differentiating $h_{\gamma(T-t)}$. Denoting the full forward expression by $d_{\mathrm{SB}}^{\mathrm{full}}(y,t)$, we have proved

$$d_{\mathrm{SB}}^{\mathrm{full}}(y,t) = D - 2\tau\partial_t\log\varphi_t(y) = \frac{\mathbb{E}[\|X_T - y\|^2 \mid X_t = y]}{\gamma\tau}. \tag{90}$$

The corresponding conditional regression risk provides another exact representation of the response. For $\ell_t^{\mathrm{BM}}(y) = \mathbb{E}[\left\|(X_T - y)/\tau - b_t^+(y)\right\|^2 \mid X_t = y]$, the regression target has conditional mean $b_t^+(y)$ by (87). Therefore its squared residual has mean $\operatorname{tr}\mathrm{Cov}(X_T \mid X_t = y)/\tau^2$, and

$$\frac{\tau}{\gamma}\ell_t^{\mathrm{BM}}(y) = D + \tau \operatorname{div} b_t^+(y). \tag{91}$$

This is a conditional-risk identity; it does not replace the pointwise derivative conditions by a training-loss guarantee.

G.4 Transferring terminal geometry to the potential

The target dimension belongs to μ_T, whereas the Gaussian convolution above uses G. The relation $\mu_T = \widehat{\varphi}_T G$ resolves this distinction locally. Positivity and continuity of $\widehat{\varphi}_T$ at x imply that its value is bounded above and away from zero on a sufficiently small neighborhood. In fact,

$$\frac{\mu_T(B(x,r))}{G(B(x,r))} \longrightarrow \widehat{\varphi}_T(x) \quad (r \downarrow 0),$$

so the two numerical local dimensions coincide whenever either exists.

The same reweighting preserves the stronger tangent assumption. If μ_T has normalized blow-ups $\nu_{x,\varepsilon}$ with limit ν_x, put $\omega = 1/\widehat{\varphi}_T$ near x. For every continuous compactly supported f,

$$\varepsilon^{-d} \int f((z-x)/\varepsilon)\, G(\mathrm{d}z) = \int f(u)\omega(x+\varepsilon u)\, \nu_{x,\varepsilon}(\mathrm{d}u) \longrightarrow \omega(x) \int f(u)\, \nu_x(\mathrm{d}u). \tag{92}$$

The convergence follows because $\omega(x+\varepsilon u) \to \omega(x)$ uniformly on the support of f, and locally convergent measures have bounded mass there. Reversing the roles of G and μ_T proves the converse. Thus the homogeneity exponent, flat tangent, and cone geometry are preserved, with only the tangent measure's positive prefactor changed.

This is a local assertion. The reciprocal weight can amplify distant mass, so the weighted tail condition must be checked for G itself. Write $\nu^G_{x,\varepsilon} = \varepsilon^{-d}(T_{x,\varepsilon})_{\#}G$. For the standard Gaussian, $\|\nabla K(z)\| = \|z\|\, K(z)$, so (A2) is equivalent, up to constant factors, to

$$\lim_{R\to\infty} \sup_{0<\varepsilon<\varepsilon_0} \int_{\|z\|>R} (1+\|z\|^2) e^{-\|z\|^2/2}\, \nu^G_{x,\varepsilon}(\mathrm{d}z) = 0$$

and positive finite limiting kernel mass. In particular, the second-moment control used by the flat and cone conclusions is already present in Gaussian (A2). Unit total mass of G is not assumed.

The cutoff arguments and exact integral identities of Appendices B–C apply to this locally finite measure: they require finite kernel-weighted integrals, a positive denominator, and valid differentiation, all specified above. The probability-measure hypothesis in their convenient Gaussian tail checks is replaced here by the stated checks for G. The posterior itself is the probability measure (82).

Now $\varphi_{T-\varepsilon^2/\gamma} = K_\varepsilon * G$ with $\varepsilon = \sqrt{\gamma\tau}$. At fixed y the chain rule is

$$\varepsilon\partial_\varepsilon \log \varphi_{T-\varepsilon^2/\gamma}(y) = -2\tau\partial_t \log \varphi_t(y).$$

The density branch of the master theorem, applied through its integral proof to G, and (90) prove (83). If the tangent is flat, the Gaussian response matrix tends to the orthogonal projector P_{T_x}, proving $D + \tau \operatorname{div} b^+_t(x) \to d$ and $\tau\gamma^{-1} \left\|b^+_t(x)\right\|^2 \to 0$. For a cone tangent, let W_C be the conditional Gaussian on that cone as in Theorem D.2. The same posterior limit gives

$$D + \tau \operatorname{div} b^+_t(x) \longrightarrow \operatorname{tr}\operatorname{Cov}(W_C), \tag{93}$$

$$\frac{\tau}{\gamma} \left\|b^+_t(x)\right\|^2 \longrightarrow \left\|\mathbb{E}W_C\right\|^2. \tag{94}$$

Their sum is d; for a half-space the two limits are $d - 2/\pi$ and $2/\pi$. The full formula therefore retains the boundary correction.

G.5 Current velocity and the opposite potential

Current velocity is associated with the actual marginal p_t, not just φ_t. Their density slopes differ exactly by

$$D - 2\tau\partial_t \log p_t(x) = d^{\mathrm{full}}_{\mathrm{SB}}(x,t) - 2\tau\partial_t \log \widehat{\varphi}_t(x). \tag{95}$$

The final term vanishes if $\partial_t \log \widehat{\varphi}_t(x)$ is bounded; more generally its product with τ tending to zero suffices.

To express the marginal slope through model fields, the heat equations give

$$\partial_t p_t = \frac{\gamma}{2}\big(\varphi_t \Delta\widehat{\varphi}_t - \widehat{\varphi}_t\Delta\varphi_t\big), \qquad p_t v_t = \frac{\gamma}{2}\big(\widehat{\varphi}_t\nabla\varphi_t - \varphi_t\nabla\widehat{\varphi}_t\big).$$

The cross-gradient terms cancel when the second expression is differentiated in space. Hence

$$b^+_t - b^-_t = \gamma s_t, \qquad v_t = \frac{b^+_t + b^-_t}{2}, \qquad \partial_t p_t + \operatorname{div}(p_t v_t) = 0. \tag{96}$$

Dividing the continuity equation by the positive marginal density proves the exact identity

$$D - 2\tau\partial_t \log p_t(x) = D + 2\tau\big(\operatorname{div} v_t(x) + v_t(x)^\top s_t(x)\big). \tag{97}$$

Together with (95), this proves (84).

For the response-only branch, the drift relation gives

$$I_D + 2\tau J v_t(x) = I_D + \tau J b_t^+(x) + \tau J b_t^-(x) \longrightarrow P_{T_x} \tag{98}$$

under a flat tangent and bounded $Jb_t^-(x)$. Taking the trace proves the remaining response assertion in Corollary G.1. The factor two is essential: omitting it gives the matrix limit $(I_D + P_{T_x})/2$ and trace $(D + d)/2$.

The exact quadratic relation also explains the correction in the current formula:

$$2\tau v_t^\top s_t = \frac{\tau}{\gamma}\big(\|b_t^+\|^2 - \|b_t^-\|^2\big).$$

Thus bounded $b_t^-(x)$ makes the current correction have the same limit as the forward correction, including at a boundary. At finite time the full current and forward expressions differ by the regular-factor term in (95).

If the data are instead at the initial endpoint, exchange F and G and reverse time. Provided the corresponding positivity, tangent, weighted-tail, and derivative assumptions hold for F and the opposite potential φ_0, the inward field is $-b_t^-$. With physical width $\sqrt{\gamma t}$, the response and full expressions are respectively

$$D - t \operatorname{div} b_t^-(x), \qquad D - t \operatorname{div} b_t^-(x) + \frac{t}{\gamma}\|b_t^-(x)\|^2. \tag{99}$$

They have the same flat and homogeneous-tangent scopes. The current response is $D - 2t \operatorname{div} v_t(x)$ if the opposite forward Jacobian is bounded; at a source boundary it retains the same Gaussian variance deficit.

G.6 Time-dependent scalar diffusivity

The exact Brownian calculation extends to a deterministic scalar time change. Let the reference process satisfy $\mathrm{d}R_t = \sqrt{g_t}\,\mathrm{d}W_t$, where g is deterministic, nonnegative, and integrable on $[0, T]$, and is continuous and positive near T, with $g_T > 0$. Its residual variance and calibrated coefficient are

$$\Gamma_t = \int_t^T g_u \,\mathrm{d}u, \qquad c_t = \frac{\Gamma_t}{g_t}. \tag{100}$$

Use the corresponding factorized bridge construction, with total variance $\int_0^T g_u \,\mathrm{d}u$ in the endpoint coupling. Then $\varphi_t = h_{\Gamma_t} * G$ and $b_t^+ = g_t \nabla \log \varphi_t$. Gaussian posterior differentiation, now with variance Γ_t, gives

$$B_t(y) = y + c_t b_t^+(y), \tag{101}$$

$$D + c_t \operatorname{div} b_t^+(y) = \frac{\operatorname{tr}\operatorname{Cov}(X_T \mid X_t = y)}{\Gamma_t}, \tag{102}$$

$$D + c_t \operatorname{div} b_t^+(y) + \frac{\Gamma_t}{g_t^2}\|b_t^+(y)\|^2 = D - 2c_t\partial_t \log \varphi_t(y). \tag{103}$$

For the last equality, $\partial_t \varphi_t = -(g_t/2)\Delta\varphi_t$ and $c_t g_t = \Gamma_t$. The scale is $\varepsilon = \sqrt{\Gamma_t}$, so $\varepsilon\partial_\varepsilon = -2c_t\partial_t$ at a fixed observation. The same assumptions on G therefore give the full limit d, and a flat tangent gives the response limit d.

Define $b_t^- = -g_t \nabla \log \widehat{\varphi}_t$, with $\widehat{\varphi}_t = h_{\int_0^t g_u \,\mathrm{d}u} * F$, $v_t = (b_t^+ + b_t^-)/2$, and $s_t = \nabla \log p_t$. The potential heat equations again imply continuity, giving

$$D - 2c_t\partial_t \log p_t(y) = D + 2c_t\big(\operatorname{div} v_t(y) + v_t(y)^\top s_t(y)\big). \tag{104}$$

Its difference from (103) is $-2c_t\partial_t \log \widehat{\varphi}_t(y)$. Accordingly, current full recovery requires this term to vanish at x, and current response recovery at a flat tangent requires $c_t J b_t^-(x) \to 0$. Bounded opposite-potential derivatives are sufficient because $c_t \sim T - t$. These conditions are part of the time-change extension, not consequences of choosing a schedule alone.

G.7 A CONDITIONAL UNIFORMLY ELLIPTIC EXTENSION

For a general reference diffusion the heat factorization is no longer an exact additive Gaussian channel. The following statement records the differential identity and the additional endpoint hypothesis that would transfer it to LID. It is conditional and is not an extra model class covered by the Brownian table rows.

Proposition G.2 (Conditional formula for a uniformly elliptic bridge)**.** *Let a nonexplosive reference diffusion have transition density $r(t, y; T, z)$ and generator*

$$\mathcal{L}_t^R f = (b_t^R)^\top \nabla f + \tfrac{1}{2} \operatorname{tr}(a_t^R \nabla^2 f),$$

with symmetric uniformly elliptic coefficients of class $C_x^2 C_t^1$ near (T, x). Let $\Phi_t(y) = \int r(t, y; T, z)\, G(\mathrm{d}z) > 0$ be a finite terminal potential satisfying the backward Kolmogorov equation, with all displayed differentiations under the integral valid. Set $c_t^R = a_t^R \nabla \log \Phi_t$ and $\tau = T - t$. Then

$$\begin{aligned} D - 2\tau \partial_t \log \Phi_t(x) = D + \tau \operatorname{div} c_t^R(x) + \tau (2b_t^R - \operatorname{div} a_t^R)^\top (a_t^R)^{-1} c_t^R(x) \\ + \tau (c_t^R(x))^\top (a_t^R(x))^{-1} c_t^R(x), \end{aligned} \tag{105}$$

where $(\operatorname{div} a)_j = \sum_i \partial_i a_{ij}$ and all coefficients are evaluated at (t, x).

Assume additionally that G has a nonzero d-homogeneous full tangent at x and that the kernels $L_\varepsilon(y, z) = r(T - \varepsilon^2, y; T, z)$ satisfy the three normal-form error conditions (12)*–*(13) *and the scale derivative exchange, with G in place of μ and limiting profile*

$$K_x(w) = \frac{\exp[-\frac{1}{2} w^\top (a_T^R(x))^{-1} w]}{(2\pi)^{D/2} \det(a_T^R(x))^{1/2}}.$$

Require (A2) for this profile. If b_t^R and $\operatorname{div} a_t^R$ are bounded near (T, x), then

$$D + \tau \operatorname{div} c_t^R(x) + \tau (c_t^R(x))^\top (a_t^R(x))^{-1} c_t^R(x) \longrightarrow d. \tag{106}$$

For a flat tangent the quadratic term tends to zero, so the divergence expression alone has the same limit.

To prove the identity, write $h = \log \Phi$. The backward equation gives

$$-2\partial_t h = 2(b_t^R)^\top \nabla h + \operatorname{tr}(a_t^R \nabla^2 h) + (\nabla h)^\top a_t^R \nabla h.$$

Since $\operatorname{div}(a_t^R \nabla h) = (\operatorname{div} a_t^R)^\top \nabla h + \operatorname{tr}(a_t^R \nabla^2 h)$, substitution of $\nabla h = (a_t^R)^{-1} c_t^R$ proves (105).

For the limit, the imposed weighted normal form permits the same integral proof as Appendix C, with $\varepsilon = \sqrt{\tau}$: $D - 2\tau \partial_t \log \Phi_t(x) \to d$ and $\sqrt{\tau} \nabla \log \Phi_t(x) \to g_x$, where g_x is the mean limiting kernel score. This proof does not require L_ε to integrate to one in y; it uses its positive integrals, derivative exchanges, and weighted errors. That distinction matters because a reference transition density is normalized in its terminal variable z. Smooth uniformly elliptic coefficients alone are not a verification of the imposed weighted normal form.

The gradient limit and bounded coefficients give $c_t^R(x) = \mathcal{O}(\tau^{-1/2})$. Uniform ellipticity bounds $(a_t^R)^{-1}$, so the linear reference term in (105) is $\mathcal{O}(\sqrt{\tau})$ and vanishes. This proves (106). On a flat tangent, the centered anisotropic Gaussian profile is even, so its kernel-weighted tangent mean is zero; its score is $-(a_T^R(x))^{-1} w$, hence $g_x = 0$. It follows that $\sqrt{\tau} c_t^R(x) \to 0$, which removes the quadratic term. To identify d with a bridge terminal marginal, the same positive continuous terminal reweighting used in (92) is also required.

G.8 SUBSTITUTING LEARNED FIELDS AND CHOOSING THE ENDPOINT SCALE

The formulas above concern exact fields. If a learned forward drift is $b_{\theta,t}^+ = b_t^+ + e_t^+$, the exact plug-in error in the full expression is

$$\widehat{d}_{\theta,t}^{\mathrm{SB},+}(x) - d_{\mathrm{SB}}^{\mathrm{full}}(x, t) = \tau \operatorname{div} e_t^+(x) + \frac{2\tau}{\gamma} b_t^+(x)^\top e_t^+(x) + \frac{\tau}{\gamma} \left\| e_t^+(x) \right\|^2. \tag{107}$$

For $v_{\theta,t} = v_t + e_{v,t}$ and $s_{\theta,t} = s_t + e_{s,t}$, the current full expression has error

$$2\tau \left[\operatorname{div} e_{v,t} + e_{v,t}^\top s_t + v_t^\top e_{s,t} + e_{v,t}^\top e_{s,t}\right](x). \tag{108}$$

Under the posterior first-moment limit, $b_t^+(x) = \mathcal{O}(\tau^{-1/2})$. Thus $\tau \operatorname{div} e_t^+(x) \to 0$ and $\sqrt{\tau} e_t^+(x) \to 0$ are sufficient for the forward full error to vanish. If the opposite drift is bounded, then $v_t(x), s_t(x) = \mathcal{O}(\tau^{-1/2})$; sufficient current conditions are $\tau \operatorname{div} e_{v,t}(x) \to 0$, $\sqrt{\tau} e_{v,t}(x) \to 0$, and $\sqrt{\tau} e_{s,t}(x) \to 0$. For either response-only expression, only its scaled divergence error enters. A small field-prediction loss does not by itself establish these derivative conditions. Trace-estimation error is treated separately in Appendix I.

The physical width must also resolve the intended data geometry. In the translation-invariant flat tangent model, convolving the terminal d-plane measure with ambient Gaussian variance δ^2 gives a normal Gaussian profile. At x on the plane, its further convolution of variance $\gamma\tau$ has $b_t^+(x) = 0$ and $\operatorname{div} b_t^+(x) = -\gamma(D-d)/(\delta^2 + \gamma\tau)$. Both forward expressions then equal

$$d_{\text{eff}}(\tau, \delta) = D - (D-d)\frac{\gamma\tau}{\delta^2 + \gamma\tau}. \tag{109}$$

This calculation illustrates the noncommuting limits: fixed $\delta > 0$ yields D as $\tau \downarrow 0$, whereas $\delta^2/(\gamma\tau) \to 0$ yields d.

A fixed empirical terminal factor is atomic and gives zero at an isolated atom in the infinitesimal limit. A population interpretation consequently needs a regime in which local kernel averages approximate the population while the physical width remains below the scale of geometric variation; such a statistical limit is not supplied by the present theorem. Finally, in the unsmoothed flat-tangent model ($\delta = 0$), evaluating the full expression at $x + \varepsilon z_N$ for a fixed normal vector gives $d + \|z_N\|^2$. The scale derivative still holds the observation fixed before this evaluation. The theorem and the table use the data point x itself.

H Normalizing Flows: Density Readouts and a Separate Transport Criterion

A normalizing flow supplies a density through an invertible change of variables, but its architecture does not specify how that density approaches the data law. The NF row of Section 3 applies when a scale-conditioned family represents a class-K density path. We first prove its fixed-query likelihood formula and its expression through a transport velocity. We then explain why a single regular flow cannot have lower-dimensional infinitesimal geometry, and state a separate, more restrictive transport criterion under which a velocity divergence alone recovers dimension. This latter result, called class T below, is not a consequence of the class-K master theorem.

H.1 The class-K likelihood formula at a fixed data point

Let π be a fixed base probability density, let F_ε be a family of diffeomorphisms, and write $p_\varepsilon = (F_\varepsilon)_\# \pi$. The parameter ε is the physical perturbation width in data coordinates. The query x is held fixed when ε varies. Define the scale velocity and density score by

$$u_\varepsilon(y) = (\varepsilon \partial_\varepsilon F_\varepsilon)(F_\varepsilon^{-1}(y)), \qquad s_\varepsilon(y) = \nabla_y \log p_\varepsilon(y). \tag{110}$$

The velocity identities require the corresponding local regularity. We assume differentiation of the weak pushforward identity is valid, and use the pointwise formulas where p_ε is positive and continuously differentiable in width and observation and u_ε is continuously differentiable in observation. The Jacobian calculations below additionally use continuous mixed derivatives of F_ε and the equality $D_z \partial_\varepsilon F_\varepsilon = \partial_\varepsilon D_z F_\varepsilon$. These assumptions concern evaluation of the model interface; they do not establish class-K membership.

Corollary H.1 (Normalizing flows along a calibrated path)**.** *Suppose (A1) holds at x and $p_\varepsilon = \rho_\varepsilon^Q$ for a class-K channel. Under the pushforward differentiation and local regularity just stated, at positive widths with $p_\varepsilon(x) > 0$,*

$$D + \varepsilon \partial_\varepsilon \log p_\varepsilon(x) = D - \operatorname{div} u_\varepsilon(x) - u_\varepsilon(x)^\top s_\varepsilon(x) \longrightarrow d. \tag{111}$$

The derivative on the left holds the data point x fixed.

The equality is a continuity identity for the density path; the limit is supplied by Theorem 2.2. In particular, this is a differential implementation of the calibrated density idea used by LIDL (Tempczyk et al., 2022), rather than an additional claim about arbitrary NF architectures. The density branch retains the homogeneous-tangent scope of the master theorem, including admissible boundaries and cones.

Proof of the corollary. For a smooth compactly supported test function ϕ, the pushforward identity and its assumed differentiation give

$$\begin{aligned}\varepsilon\partial_\varepsilon \int \phi(y)p_\varepsilon(y)\,\mathrm{d}y &= \varepsilon\partial_\varepsilon \int \phi(F_\varepsilon(z))\pi(z)\,\mathrm{d}z \\ &= \int \nabla\phi(F_\varepsilon(z))^\top \varepsilon\partial_\varepsilon F_\varepsilon(z)\,\pi(z)\,\mathrm{d}z \\ &= \int \nabla\phi(y)^\top u_\varepsilon(y)p_\varepsilon(y)\,\mathrm{d}y.\end{aligned}$$

Integration by parts, with no boundary term because ϕ has compact support, proves

$$\varepsilon\partial_\varepsilon p_\varepsilon + \operatorname{div}(p_\varepsilon u_\varepsilon) = 0 \tag{112}$$

as a distributional identity. Under the stated local regularity its residual is continuous, hence the identity holds pointwise. Expanding the product divergence and dividing by the positive density gives

$$\varepsilon\partial_\varepsilon \log p_\varepsilon = -\operatorname{div} u_\varepsilon - u_\varepsilon^\top \nabla \log p_\varepsilon,$$

which proves the equality in (111). Since $p_\varepsilon = \rho_\varepsilon^Q$, the left-hand side plus D is exactly the density readout of Section 2. Conditions (A1) and class K therefore imply convergence to d by Theorem 2.2. The proof uses no flat-tangent assumption and no reconstruction interface.

A model clock and direct likelihood evaluation. If the family is indexed by a differentiable time t with $\varepsilon(t) \to 0$, put $\kappa_t = \partial_t \log \varepsilon(t) \neq 0$, $F_t = F_{\varepsilon(t)}$, and $v_t = (\partial_t F_t) \circ F_t^{-1}$. The chain rule gives $v_t = \kappa_t u_{\varepsilon(t)}$. Because κ_t depends on time but not on the observation,

$$D + \frac{\partial_t \log p_t(x)}{\kappa_t} = D - \frac{\operatorname{div} v_t(x) + v_t(x)^\top \nabla \log p_t(x)}{\kappa_t} \longrightarrow d. \tag{113}$$

Thus an arbitrary model clock cannot replace the known physical scale without the factor κ_t.

An NF that evaluates likelihood need not construct a separate velocity to use the first expression in (111). With encoder $f_\varepsilon = F_\varepsilon^{-1}$, change of variables gives $\log p_\varepsilon(x) = \log \pi(f_\varepsilon(x)) + \log|\det D_x f_\varepsilon(x)|$. Differentiating at fixed x, where the base score and mixed derivatives exist, and using the derivative of log determinant gives

$$\begin{aligned}D + \varepsilon\partial_\varepsilon \log p_\varepsilon(x) = D &+ \nabla \log \pi(f_\varepsilon(x))^\top \varepsilon\partial_\varepsilon f_\varepsilon(x) \\ &+ \operatorname{tr}\left[(D_x f_\varepsilon(x))^{-1}\varepsilon D_x(\partial_\varepsilon f_\varepsilon)(x)\right].\end{aligned} \tag{114}$$

Both terms are necessary in general. The base point $f_\varepsilon(x)$ moves when the density is differentiated at a fixed data point.

H.2 Fixed queries, moving characteristics, and transport choice

It is tempting to retain only a determinant derivative or a velocity divergence. Those quantities have a different interpretation along a moving characteristic. Fix a latent point z with $\pi(z) > 0$ and let $x_\varepsilon = F_\varepsilon(z)$. The chain rule and (112) imply

$$\begin{aligned}\varepsilon\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\log p_\varepsilon(x_\varepsilon) &= \varepsilon(\partial_\varepsilon \log p_\varepsilon)(x_\varepsilon) + u_\varepsilon(x_\varepsilon)^\top s_\varepsilon(x_\varepsilon) \\ &= -\operatorname{div} u_\varepsilon(x_\varepsilon).\end{aligned}$$

This is a material derivative, whereas the master theorem uses a fixed observation.

The same distinction follows directly from the NF Jacobian. Differentiating (110) in the observation and evaluating at $F_\varepsilon(z)$ gives

$$Ju_\varepsilon(F_\varepsilon(z)) = (\varepsilon\partial_\varepsilon DF_\varepsilon(z))DF_\varepsilon(z)^{-1}.$$

Taking the trace and applying Jacobi's formula yields

$$\varepsilon\partial_\varepsilon \log|\det DF_\varepsilon(z)| = \operatorname{tr}\big((\varepsilon\partial_\varepsilon DF_\varepsilon(z))DF_\varepsilon(z)^{-1}\big) = \operatorname{div} u_\varepsilon(F_\varepsilon(z)). \tag{115}$$

Consequently,

$$D + \varepsilon\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\log p_\varepsilon(x_\varepsilon) = D - \operatorname{div} u_\varepsilon(x_\varepsilon) = D - \varepsilon\partial_\varepsilon \log|\det DF_\varepsilon(z)|. \tag{116}$$

These exact equalities alone give no LID limit: a characteristic need not stay at the fixed query, and omitting its score contribution changes the statistic.

The full fixed-query expression depends only on the marginal density path. To verify this directly, let u_ε and $\widetilde{u}_\varepsilon$ be two sufficiently regular velocities producing the same p_ε, and put $w_\varepsilon = \widetilde{u}_\varepsilon - u_\varepsilon$. Subtracting their continuity equations gives

$$\operatorname{div}(p_\varepsilon w_\varepsilon) = 0, \qquad \operatorname{div} w_\varepsilon + w_\varepsilon^\top s_\varepsilon = 0$$

where $p_\varepsilon > 0$. The change in the two terms on the right of (111) therefore cancels exactly. Neither term is separately invariant in general.

The fixed-query density statistic is also unchanged by an ε-independent smooth change of data coordinates $y = h(x)$ with nonzero Jacobian determinant. In those coordinates, $\widetilde{p}_\varepsilon(h(x)) = p_\varepsilon(x)/|\det Dh(x)|$; differentiation in ε at fixed x, equivalently at fixed $h(x)$, leaves the scale derivative unchanged. This argument concerns a fixed coordinate map. A scale-dependent coordinate map requires the coordinate and clock corrections of Appendix C.1.

H.3 Why a fixed regular flow remains full-dimensional

The class-K corollary concerns a path of densities. It does not identify the infinitesimal dimension of one regular, positive-density NF with that of a singular data law.

Proposition H.2 (The invertibility obstruction)**.** *Let F be an invertible C^1 change of variables defining an NF, with a C^1 inverse near x. If π is positive and continuous near $F^{-1}(x)$, then $F_\#\pi$ has a positive continuous density near x and its pointwise LID is D. If a differentiable flow $\dot{F}_t = v_t \circ F_t$, $F_0 = I_D$, generated by a spatially C^1 field on $\mathbb{R}^D$ exists on $[0,T]$ and*

$$\int_0^T \|Jv_t\|_{L^\infty}\,\mathrm{d}t < \infty, \tag{117}$$

then F_T is bi-Lipschitz on its flow domain and the pushforward of an absolutely continuous measure remains absolutely continuous. In particular, it cannot be supported on a lower-dimensional manifold.

Proof. The single inverse branch supplied by the invertible NF gives

$$p(x) = \pi(F^{-1}(x))|\det DF(F^{-1}(x))|^{-1}. \tag{118}$$

The determinant is nonzero and continuous, so p is positive and continuous near x. Uniform continuity at x implies

$$\mu(B(x,r)) = p(x)\omega_D r^D + o(r^D),$$

and taking the logarithmic ratio proves LID D. The invertibility qualification matters: regularity of one branch of a noninjective map would not control contributions from other preimages.

For the continuous flow, set $L(t) = \|Jv_t\|_{L^\infty}$. Two trajectories $a_t = F_t(a)$ and $b_t = F_t(b)$ obey $\|v_t(a_t) - v_t(b_t)\| \le L(t)\,\|a_t - b_t\|$. Applying the differential inequalities for their distance, or their integral Grönwall form, gives

$$e^{-\int_0^t L(s)\,\mathrm{d}s}\,\|a-b\| \le \|F_t(a) - F_t(b)\| \le e^{\int_0^t L(s)\,\mathrm{d}s}\,\|a-b\|\,.$$

The lower bound excludes merging trajectories and bounds the inverse on the image; the upper bound makes the map Lipschitz. A Lipschitz map in $\mathbb{R}^D$ sends Lebesgue-null sets to Lebesgue-null sets, as follows by covering such a set by cubes of arbitrarily small total volume and bounding the diameters of their images. Apply this fact to the inverse map: the preimage of a null subset of the image is null. An absolutely continuous initial law therefore assigns zero mass to that preimage, proving absolute continuity of its pushforward. This proves the proposition.

H.4 A SEPARATE CRITERION FOR CALIBRATED TRANSPORT DEGENERATION

A score-free transport statistic can nevertheless be valid when the family itself has a controlled geometric degeneration. The next theorem states sufficient conditions for that different mechanism. Its geometric limit, calibration, and anchoring are assumptions; they are not inferred from the likelihood path or supplied by Section 2.

Theorem H.3 (Projector of a calibrated singular flow)**.** *Let F_ε be diffeomorphisms with the mixed-derivative regularity required for* (110) *and* (115)*. Suppose $\pi(z_0) > 0$, $F_\varepsilon(z_0) = x$ for every ε, and $\mu_\varepsilon = (F_\varepsilon)_\#(\pi\,\mathrm{d}z) \Rightarrow \mu_0$. Assume there are fixed neighborhoods $U \ni z_0$ and $V \ni x$ such that $F_\varepsilon \to F_0$ in $C^1(U)$, F_0 has constant rank d on U, $\mathcal{M} = F_0(U)$ is embedded in V, and*

$$\mu_0 \restriction_V = q_0 \mathcal{H}^d \restriction_{\mathcal{M}\cap V}, \qquad q_0 \text{ is continuous}, \qquad q_0(x) > 0.$$

Suppose further that a fixed orthogonal matrix $Q = [Q_T, Q_N]$, partitioned into d and $m = D - d$ columns, satisfies

$$DF_\varepsilon(z_0)Q = [G_\varepsilon,\ \varepsilon A_\varepsilon], \qquad G_\varepsilon \to G, \quad A_\varepsilon \to A, \qquad \varepsilon\partial_\varepsilon G_\varepsilon \to 0, \quad \varepsilon\partial_\varepsilon A_\varepsilon \to 0, \tag{119}$$

with $[G, A]$ invertible. A family satisfying these conditions is called a class-T transport *at x. Then $Ju_\varepsilon(x)$ and $I_D - Ju_\varepsilon(x)$ converge to complementary oblique projectors of ranks m and d. The numerical LID of μ_0 at x is d, and*

$$D - \operatorname{div} u_\varepsilon(x) \longrightarrow d, \qquad D - \frac{\operatorname{div} v_t(x)}{\kappa_t} \longrightarrow d, \tag{120}$$

where the second expression uses $F_t = F_{\varepsilon(t)}$, $\varepsilon(t) \to 0$, $v_t = (\partial_t F_t)\circ F_t^{-1}$, and $\kappa_t = \partial_t \log \varepsilon(t) \neq 0$.

The partition appears in the sufficient assumptions; the trace statistic does not require it as an input. The factor ε specifies a common, known first-order transverse rate. The fixed-point condition is equally substantive: differentiating $F_\varepsilon(z_0) = x$ gives $u_\varepsilon(x) = 0$, so this query is an anchored characteristic. At that query the fixed and material continuity identities coincide. The theorem does not allow the score term to be dropped at an arbitrary moving or unanchored query of a class-K family.

Proof of the transport theorem. The C^1 limit and the anchor imply $F_0(z_0) = x$ and

$$DF_0(z_0) = [G, 0]Q^\top.$$

The constant-rank hypothesis identifies $\operatorname{range}(G) = T_x\mathcal{M}$. The positive continuous density of μ_0 on this embedded manifold gives $\mu_0(B(x, r)) = q_0(x)\omega_d r^d + o(r^d)$ by the chart argument in Appendix D; hence d is the numerical LID of the limiting law. This geometric step is separate from the matrix calculation.

Put $B_\varepsilon = [G_\varepsilon, A_\varepsilon]$ and $B = [G, A]$. Since B is invertible, $B_\varepsilon^{-1} \to B^{-1}$. At the fixed latent point, the chain rule gives

$$\begin{aligned} Ju_\varepsilon(x) &= (\varepsilon\partial_\varepsilon DF_\varepsilon(z_0))DF_\varepsilon(z_0)^{-1} \\ &= \big[\varepsilon\partial_\varepsilon G_\varepsilon, \varepsilon A_\varepsilon + \varepsilon^2\partial_\varepsilon A_\varepsilon\big][G_\varepsilon, \varepsilon A_\varepsilon]^{-1} \\ &= \big[\varepsilon\partial_\varepsilon G_\varepsilon, A_\varepsilon + \varepsilon\partial_\varepsilon A_\varepsilon\big]B_\varepsilon^{-1} \longrightarrow [0, A]B^{-1}. \end{aligned} \tag{121}$$

The third line factors the common diagonal matrix $\operatorname{diag}(I_d, \varepsilon I_m)$ out of both products, so no uncontrolled inverse factor remains. Set

$$P_{\mathrm{col}}^{\mathrm{obl}} := [0, A]B^{-1} = B\operatorname{diag}(0_d, I_m)B^{-1}, \qquad P_{\tan}^{\mathrm{obl}} := I_D - P_{\mathrm{col}}^{\mathrm{obl}}. \tag{122}$$

Conjugation preserves products and rank, so the first matrix squares to itself, has rank m, image $\operatorname{span}(A)$ and kernel $\operatorname{span}(G)$. The complementary matrix has image $T_x\mathcal{M}$ and rank d. They are orthogonal projectors precisely when $\operatorname{span}(A) = T_x\mathcal{M}^\perp$; otherwise they are oblique. Taking traces in (121) gives the first limit in (120). Finally $v_t = \kappa_t u_{\varepsilon(t)}$, with a spatially constant κ_t, gives the second limit. This proves Theorem H.3.

A spectral reading under the same geometric premise. One can also count separated singular directions without supplying their number to an estimator. Retain the fixed-neighborhood geometric assumptions above, and let $J_\varepsilon = DF_\varepsilon(z_0)$. Suppose d singular values are bounded away from zero and their left singular subspace tends to $T_x\mathcal{M}$, while the others are $\mathcal{O}(\varepsilon)$. For $\eta_\varepsilon \downarrow 0$ with $\varepsilon/\eta_\varepsilon \to 0$, define

$$R_{\varepsilon,\eta} = J_\varepsilon J_\varepsilon^\top (J_\varepsilon J_\varepsilon^\top + \eta_\varepsilon^2 I_D)^{-1}. \tag{123}$$

In the left singular basis its eigenvalues are $\sigma_i(\varepsilon)^2/(\sigma_i(\varepsilon)^2 + \eta_\varepsilon^2)$. They tend to one in the first group, since its singular values have a positive lower bound, and to zero in the second, since $\sigma_i(\varepsilon)/\eta_\varepsilon = \mathcal{O}(\varepsilon/\eta_\varepsilon) \to 0$. Together with convergence of the first left subspace, this proves $R_{\varepsilon,\eta} \to P_{T_x\mathcal{M}}$ in matrix norm and $\operatorname{tr} R_{\varepsilon,\eta} \to d$. The geometric premise is what identifies this number with data LID. At each positive width the exact rank of the invertible NF Jacobian is still D.

H.5 Why the transport hypotheses cannot be omitted

Calibration and determinant exponents. Take a smooth positive base density. In $\mathbb{R}^2$, $F_\varepsilon = \operatorname{diag}(1, \varepsilon^2)$ converges to a line-supported law, but $D - \varepsilon\partial_\varepsilon \log|\det DF_\varepsilon| = 2 - 2 = 0$. The physical transverse width is $\delta = \varepsilon^2$; using $\delta\partial_\delta$ gives the correct value 1. More generally, the two families in $\mathbb{R}^3$

$$\operatorname{diag}(1, 1, \varepsilon^2), \qquad \operatorname{diag}(1, \varepsilon, \varepsilon) \tag{124}$$

both have determinant ε^2, but their limiting ranks are two and one. For a standard Gaussian base these limits have positive densities on the corresponding plane and line, so their numerical dimensions are indeed different. A determinant exponent counts rates of compression, not the number of collapsing directions without a calibration assumption.

Pointwise pinching without a geometric collapse. A small Jacobian at one latent point does not establish a lower-dimensional limiting distribution. An explicit one-dimensional example makes the missing hypothesis visible. Choose $h \in C_c^\infty(\mathbb{R})$ with $0 \le h \le 1$ and $h(0) = 1$, and put $\Psi(s) = \int_0^s h(a)\,\mathrm{d}a$. The function Ψ is bounded. For $0 < \varepsilon < 1$, define

$$F_\varepsilon(z) = z - (1-\varepsilon)\varepsilon\Psi(z/\varepsilon).$$

Then $F_\varepsilon(0) = 0$ and $F_\varepsilon'(z) = 1 - (1-\varepsilon)h(z/\varepsilon) \ge \varepsilon > 0$. Its limits at $z \to \pm\infty$ are $\pm\infty$, so it is a smooth diffeomorphism. Moreover, $\sup_z |F_\varepsilon(z) - z| \le \varepsilon \|\Psi\|_\infty \to 0$. For a smooth positive base density, dominated convergence against bounded continuous test functions therefore gives $(F_\varepsilon)_\# \pi \Rightarrow \pi$. The limiting law has LID 1 at zero. Nevertheless $F_\varepsilon'(0) = \varepsilon$, so the pointwise determinant statistic is $1 - \varepsilon\partial_\varepsilon \log F_\varepsilon'(0) = 0$ at every width. There is no C^1 convergence to the identity on a fixed neighborhood of zero: its derivative at zero tends to zero instead of one. The fixed-neighborhood condition in class T rules out exactly this type of shrinking exceptional region.

Changing the latent transport while preserving the density path. If H_ε preserves the base measure, then F_ε and $F_\varepsilon \circ H_\varepsilon$ have the same marginal densities but can have different local Jacobians and characteristics. Such reparameterizations are not merely formal. For a smooth positive π on $\mathbb{R}^D$, $D \ge 2$, a smooth compactly supported χ, and a constant skew-symmetric matrix J_0, let

$$w(z) = \pi(z)^{-1} J_0 \nabla\chi(z).$$

This is a smooth compactly supported field, and $\operatorname{div}(\pi w) = \operatorname{tr}(J_0\nabla^2\chi) = 0$ because the Hessian is symmetric. Its flow H_s exists for every finite s. Differentiating along that flow gives

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\pi(H_s(z))\det DH_s(z)\right] = \operatorname{div}(\pi w)(H_s(z))\det DH_s(z) = 0.$$

Thus H_s preserves π, and taking $H_\varepsilon = H_{a(\varepsilon)}$ supplies scale-dependent examples. These maps can change singular values: in two dimensions take a centered Gaussian base, $J_0 = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, and $\chi(z) = (z_1^2 - z_2^2)/2$ near zero, with a smooth compact cutoff. Then $w(0) = 0$ and $Jw(0) = \pi(0)^{-1}\left(\begin{smallmatrix} 0 & -1 \\ -1 & 0 \end{smallmatrix}\right)$, so $DH_s(0) = \exp(sJw(0))$ has singular values $e^{s/\pi(0)}$ and $e^{-s/\pi(0)}$ while preserving the base law. The full density statistic remains unchanged by the continuity calculation above. A chosen reparameterization can, however, destroy the fixed anchor or the calibrated matrix limit required by class T; the transport theorem is conditional on those properties of the represented transport.

Boundary points. The constant-rank theorem maps an open latent neighborhood to a local manifold patch without boundary at the image of a regular point. Hence the class-T geometric assumptions describe regular interior points. A half-line endpoint can instead arise from a latent boundary, an escaping latent preimage, or a critical limiting map such as $z \mapsto z^2$, but those cases do not satisfy the stated regular fixed-neighborhood setup. The class-K density result and its boundary correction remain applicable when their own assumptions hold; their boundary scope cannot be transferred to class T solely because both formulas involve a flow.

H.6 What a learned NF must approximate

The preceding statements concern exact paths and derivatives. Their NF-specific error identities are simple. If $\log \widehat{p}_\varepsilon = \log p_\varepsilon + e_\varepsilon$ and both width derivatives exist, subtraction at the same fixed query gives

$$\widehat{d}_\varepsilon^{\,\mathrm{fixed}} - d_\varepsilon^{\,\mathrm{fixed}} = \varepsilon \partial_\varepsilon e_\varepsilon(x), \qquad d_\varepsilon^{\,\mathrm{fixed}} = D + \varepsilon \partial_\varepsilon \log p_\varepsilon(x). \tag{125}$$

For the separate class-T velocity statistic, using the same calibrated κ_t and writing $\widehat{v}_t = v_t + e_{v,t}$ gives the exact error

$$\left(D - \frac{\operatorname{div} \widehat{v}_t(x)}{\kappa_t}\right) - \left(D - \frac{\operatorname{div} v_t(x)}{\kappa_t}\right) = -\frac{\operatorname{div} e_{v,t}(x)}{\kappa_t}. \tag{126}$$

These formulas state which derivatives must be controlled; likelihood accuracy at isolated scales or a small field-regression loss does not itself supply that control. The general separation of finite-scale bias, learned-field error, and trace estimation is treated in Appendix I. A fixed empirical target is atomic, so interpreting a finite-scale model readout as population manifold LID additionally requires an appropriate sampling regime, rather than taking an infinitesimal limit of that fixed empirical law.

I From Exact Readouts to Learned Estimates

The formulas in Table 1 describe exact model outputs. Their evaluation with a learned field and a stochastic divergence estimate introduces two further errors, in addition to the finite-scale difference between the exact readout and LID. We separate these terms for the affine response formula, then prove sufficient conditions for consistency. The model appendices give the corresponding field-error identities for their other readouts.

I.1 Finite-scale, model, and trace errors

Use the independent affine schedule of Appendix F, with $y_t = \alpha_t x$, $\varepsilon(t) = \beta_t/\alpha_t$, $a_t = \dot{\alpha}_t/\alpha_t$, and $\kappa_t = \dot{\varepsilon}(t)/\varepsilon(t) \neq 0$. Let v_t^* be the exact velocity, $v_{\theta,t}$ its learned approximation, and $e_t = v_{\theta,t} - v_t^*$. Define

$$d^*(x,t) := D + \frac{Da_t - \operatorname{div} v_t^*(y_t)}{\kappa_t}, \qquad \widehat{d}_{\theta,L}(x,t) := D + \frac{Da_t - \widehat{\tau}_L}{\kappa_t}, \tag{127}$$

where $\widehat{\tau}_L$ estimates $\operatorname{div} v_{\theta,t}(y_t)$. The exact identity (70) makes d^* the posterior response at scale $\varepsilon(t)$. Adding and subtracting the learned divergence gives

$$\begin{aligned} \widehat{d}_{\theta,L}(x,t) - d = \underbrace{d^*(x,t) - d}_{\text{finite-scale error}} - \underbrace{\frac{\operatorname{div} e_t(y_t)}{\kappa_t}}_{\text{model error}} \\ \underbrace{- \frac{\widehat{\tau}_L - \operatorname{div} v_{\theta,t}(y_t)}{\kappa_t}}_{\text{trace-estimation error}}. \end{aligned} \tag{128}$$

The first term vanishes at the flat-tangent points of Corollary 2.3. The second shows which approximation property is needed from the learned field:

$$\frac{|\operatorname{div} e_t(\alpha_t x)|}{|\kappa_t|} \longrightarrow 0. \tag{129}$$

This is a condition on the divergence error in physical-scale units. It is sufficient for the response readout, without requiring every entry of the learned Jacobian to converge.

For the full Gaussian rectified-flow formula (78), substituting $v_\theta = v^* + e$ changes the deterministic readout by exactly

$$t(1-t)\operatorname{div} e_t(tx) + 2t^2\,\langle v_t^*(tx) - x, e_t(tx)\rangle + t^2\,\|e_t(tx)\|^2\,. \tag{130}$$

Indeed, the divergence term is linear in the field, while the correction is obtained by expanding $t^2\,\|v_t^*(tx) + e_t(tx) - x\|^2 - t^2\,\|v_t^*(tx) - x\|^2$. Stochastic divergence estimation adds $t(1-t)[\widehat{\tau}_L - \operatorname{div} v_{\theta,t}(tx)]$. Thus consistency of the full readout requires control of these additional value-error terms as well as divergence error. Their exact form depends on the model formula being evaluated.

I.2 Stochastic trace estimation and consistency

At a fixed time and query, let $A = Jv_{\theta,t}(y_t)$. Conditional on the learned field, take independent probes $\xi_\ell \sim \mathcal{N}(0, I_D)$, independent of the training randomness, and use

$$\widehat{\tau}_L = \frac{1}{L}\sum_{\ell=1}^{L} \xi_\ell^\top A \xi_\ell.$$

Since the skew-symmetric part has zero quadratic form, the statistic depends only on $S = (A + A^\top)/2$. Diagonalize $S = O\operatorname{diag}(\lambda_1, \dots, \lambda_D)O^\top$. The coordinates of $O^\top \xi_\ell$ remain independent standard Gaussians, so

$$\mathbb{E}[\xi_\ell^\top A \xi_\ell \mid A] = \sum_i \lambda_i = \operatorname{tr} A, \qquad \operatorname{Var}[\xi_\ell^\top A \xi_\ell \mid A] = 2\sum_i \lambda_i^2.$$

Here each squared Gaussian has mean one and variance two, and different coordinates have zero covariance. Independence of the probes therefore gives

$$\operatorname{Var}[\widehat{\tau}_L \mid A] = \frac{2}{L}\,\|\operatorname{sym} A\|_F^2\,, \qquad \operatorname{sym} A = \tfrac{1}{2}(A + A^\top). \tag{131}$$

Corollary I.1 (Sufficient plug-in consistency)**.** *Let t_n satisfy $\varepsilon(t_n) \to 0$ and $\kappa_{t_n} \neq 0$, and let x be a flat-tangent point satisfying the assumptions of Corollary F.1. Use L_n Gaussian probes, independent of the learned field and conditionally independent of one another. If*

$$\frac{\operatorname{div} e_{t_n}(\alpha_{t_n} x)}{\kappa_{t_n}} \xrightarrow{\mathbb{P}} 0, \qquad \frac{\|\operatorname{sym} Jv_{\theta_n,t_n}(\alpha_{t_n} x)\|_F}{\sqrt{L_n}|\kappa_{t_n}|} \xrightarrow{\mathbb{P}} 0, \tag{132}$$

then $\widehat{d}_{\theta_n,L_n}(x, t_n) \xrightarrow{\mathbb{P}} d$.

Proof. The exact response tends to d by Corollary 2.3 and (70); the first condition in (132) removes the model error. By (131), the conditional variance of the normalized trace error is

$$V_n = \frac{2\,\|\operatorname{sym} Jv_{\theta_n,t_n}(\alpha_{t_n} x)\|_F^2}{L_n \kappa_{t_n}^2} \xrightarrow{\mathbb{P}} 0.$$

Conditional Chebyshev gives, for every $\delta > 0$,

$$\mathbb{P}\left(\left|\frac{\widehat{\tau}_{L_n} - \operatorname{div} v_{\theta_n,t_n}(\alpha_{t_n} x)}{\kappa_{t_n}}\right| > \delta\right) \le \mathbb{E}\min(1, V_n/\delta^2) \longrightarrow 0.$$

To justify the last limit without a moment assumption on the trained field, for every $\eta > 0$ the expectation is at most $\eta/\delta^2 + \mathbb{P}(V_n > \eta)$. First take $n \to \infty$ and then $\eta \downarrow 0$. All three terms of (128) now vanish, proving the corollary.

Which matrix to probe. For the exact Gaussian affine model, if $a_t/\kappa_t \to 0$, then (74) and tangent recovery give $Jv_t^*(y_t)/\kappa_t \to P_N = I_D - P_T$. Estimating this normal trace and subtracting it from D has asymptotic variance $2(D-d)/L$ at a fixed probe count L. Probing the full response matrix $I_D + (a_t I_D - Jv_t^*(y_t))/\kappa_t \to P_T$ instead has asymptotic variance $2d/L$. Both statements follow by substituting the limiting orthogonal projector into (131). For learned fields they additionally require the corresponding normalized Jacobian matrix error to vanish; the divergence condition alone does not provide this matrix limit.

Finally, small prediction error does not ensure the derivative condition. In one dimension, $e_n(y) = n^{-1}\sin(n^2 y)$ tends uniformly to zero and its squared regression error is at most n^{-2} under any probability law, while $e_n'(0) = n$. A guarantee for the readouts therefore needs the stated derivative-level control, in addition to accuracy of the field values.

J Connections to Prior LID Estimators

This appendix collects the mathematical comparisons summarized in Section 6: density estimators, denoiser traces, conditional FM error, denoising loss and spectral diagnostics. It includes their coordinate conversions, exact equalities, sufficient geometric conditions for the LID limits and learned-model qualifications, so the comparisons can be read together.

Except for the final non-Gaussian discussion, $X \sim \mu$, $Z \sim \mathcal{N}(0, I_D)$ independently, $Y_\varepsilon = X + \varepsilon Z$, and $b_\varepsilon(y) = \mathbb{E}_{X\sim p_\varepsilon(\cdot|y)}[X]$ is the exact posterior mean. Write

$$F_\varepsilon(y) = \mathcal{R}_\varepsilon^{\mathrm{dens}}(y), \qquad R_\varepsilon(y) = \mathcal{R}_\varepsilon^{\mathrm{resp}}(y), \qquad C_\varepsilon(y) = \frac{\|b_\varepsilon(y) - y\|^2}{\varepsilon^2}. \tag{133}$$

All finite-width identities below require the indicated conditional moments, derivatives and integrals to exist. They do not require a manifold assumption. For the limiting statements, a concrete sufficient setting is a compact embedded C^2 manifold without boundary, with positive reach and positive continuous density; Corollary J.2 proves the limits in that setting. More general local conditions are those of the main theorem: a smooth interior point has a tangent plane, and the weighted kernel integrals and derivatives must converge to those of its local tangent model. This is the meaning of the smooth-interior condition in the main-text comparison.

J.1 Density estimators: LIDL, FLIPD and Gaussian FM-FLIPD

The conditional Gaussian likelihood has score $-(y - X)/\varepsilon^2$. Averaging this likelihood score over $X \sim p_\varepsilon(\cdot \mid y)$, and differentiating the posterior-mean quotient, respectively, give

$$b_\varepsilon(y) = y + \varepsilon^2 \nabla \log \rho_\varepsilon(y), \qquad Jb_\varepsilon(y) = \frac{\mathrm{Cov}_{X\sim p_\varepsilon(\cdot|y)}(X)}{\varepsilon^2}. \tag{134}$$

For the second equality, each matrix entry follows from $\partial_j b_{\varepsilon,i} = \mathrm{Cov}_{X\sim p_\varepsilon(\cdot|y)}(X_i, X_j)/\varepsilon^2$. The Gaussian kernel also satisfies $\partial_\varepsilon K_\varepsilon = \varepsilon \Delta K_\varepsilon$. Convolving with μ and dividing by ρ_ε gives $\varepsilon \partial_\varepsilon \log \rho_\varepsilon = \varepsilon^2(\mathrm{div}\, s_\varepsilon + \|s_\varepsilon\|^2)$, where $s_\varepsilon = \nabla \log \rho_\varepsilon$.

For $s_\varepsilon = \nabla \log \rho_\varepsilon$, differentiating $b_\varepsilon(y) = y + \varepsilon^2 s_\varepsilon(y)$ yields

$$R_\varepsilon = D + \varepsilon^2 \,\mathrm{div}\, s_\varepsilon, \qquad F_\varepsilon = D + \varepsilon^2\big(\mathrm{div}\, s_\varepsilon + \|s_\varepsilon\|^2\big). \tag{135}$$

The second expression is FLIPD (Kamkari et al., 2024b), and the first shows $F_\varepsilon = R_\varepsilon + C_\varepsilon$. For an independent affine FM path, define

$$Y_t = \alpha_t X + \beta_t Z, \qquad \varepsilon = \beta_t/\alpha_t, \qquad a_t = \dot\alpha_t/\alpha_t, \qquad \kappa_t = \dot\varepsilon/\varepsilon \neq 0,$$

with $\alpha_t, \beta_t > 0$. Let $D_t(y) = \mathbb{E}_{X\sim p_t(\cdot|y)}[X]$ and $v_t^*(y) = \mathbb{E}_{(X,Z)\sim p_t(\cdot,\cdot|y)}[\dot\alpha_t X + \dot\beta_t Z]$. Eliminating $Z = (y - \alpha_t X)/\beta_t$ at fixed y gives

$$v_t^*(y) = (a_t + \kappa_t)y - \alpha_t \kappa_t D_t(y). \tag{136}$$

For Gaussian noise, the native likelihood score gives $D_t(y) = (y + \beta_t^2 s_t(y))/\alpha_t$, so $s_t(y) = (a_t y - v_t^*(y))/(\beta_t^2 \kappa_t)$. Since $b_\varepsilon(x) = D_t(\alpha_t x)$, substitution yields, at $y = \alpha_t x$,

$$R_\varepsilon(x) = D + \frac{D a_t - \mathrm{div}_y\, v_t^*(y)}{\kappa_t}, \tag{137}$$

$$F_\varepsilon(x) = R_\varepsilon(x) + \frac{\|v_t^*(y) - a_t y\|^2}{\beta_t^2 \kappa_t^2}. \tag{138}$$

This is the Gaussian FM-FLIPD construction described by Leung et al. (2025) and evaluated by Yeats et al. (2025). For the rectified path $\alpha_t = t$, $\beta_t = 1 - t$, it becomes

$$R_\varepsilon(x) = t[D + (1 - t)\,\mathrm{div}\, v_t^*(tx)], \qquad F_\varepsilon(x) = R_\varepsilon(x) + t^2 \left\|v_t^*(tx) - x\right\|^2.$$

These are estimator identities. Our benchmark's rectified and trigonometric Full rows implement this same functional, but our checkpoints, architectures, trace budgets and scale selection differ from the cited experimental protocols.

For a differentiable fitted reconstruction $\widehat{b}_\varepsilon$, define $\widehat{s}_\varepsilon = (\widehat{b}_\varepsilon - y)/\varepsilon^2$. Then the same algebra gives

$$D + \varepsilon^2(\operatorname{div} \widehat{s}_\varepsilon + \|\widehat{s}_\varepsilon\|^2) = \operatorname{div} \widehat{b}_\varepsilon + \frac{\left\|\widehat{b}_\varepsilon - y\right\|^2}{\varepsilon^2} =: \widehat{F}_\varepsilon. \tag{139}$$

The equality persists for the fitted fields, but the fitted quantity is not automatically the scale derivative of the true density.

LIDL (Tempczyk et al., 2022) approximates the derivative using density estimates at several widths. With $h = \log \varepsilon$ and $g_x(h) = \log \rho_{e^h}(x)$, its two-width version is

$$D + \frac{g_x(h_2) - g_x(h_1)}{h_2 - h_1} = \frac{1}{h_2 - h_1} \int_{h_1}^{h_2} F_{e^h}(x)\, \mathrm{d}h.$$

A least-squares slope over more widths is likewise a weighted average of pairwise secant slopes. Hence the population regression shares the LID limit when the entire noise range tends to zero under the density theorem, but need not equal the derivative at a chosen finite width. Density-approximation error is an additional issue.

J.2 Denoiser trace: Response after coordinate normalization

Let $D_t(y) = \mathbb{E}_{X \sim p_t(\cdot|y)}[X]$ reconstruct the clean endpoint from $Y_t = \alpha_t X + \beta_t Z$. Since $b_\varepsilon(x) = D_t(\alpha_t x)$, the chain rule gives

$$R_\varepsilon(x) = \alpha_t \operatorname{tr} J_y D_t(\alpha_t x), \qquad \varepsilon = \beta_t/\alpha_t. \tag{140}$$

This is the coordinate normalization relating the denoiser trace in Kim (2026) to Response; for rectified flow the multiplier is t. The same identity holds for a fitted reconstruction with the same change of coordinates. This correspondence concerns the trace defined in that preprint, not its thresholded eigenvalue counts. The latter are discussed separately below.

The Gaussian geometric interpretation also has direct precedents: Wan et al. (2025) conjecture convergence of the denoiser Jacobian to a tangent projector, and Brosse & Dalalyan (2026) establish Gaussian expansions giving that limit in the smooth interior. The projector preserves the d tangent directions and removes the normal ones, so its trace is d. The proof below supplies a sufficient smooth setting for the limits used in this comparison; our extension to admissible non-Gaussian kernels is a separate consequence of the main theorem.

J.3 Conditional FM prediction error: Response for the optimal field

Different clean-point/noise pairs can give the same intermediate observation y at time t. Each pair supplies its own training target velocity. The optimal predictor of that target from (y, t) is its conditional mean, so a nonzero squared error can remain solely because those targets differ. It is this minimum conditional prediction error, rather than model approximation error, that enters the correspondence.

The conditional FM error considered by Cha et al. (2026) concerns the velocity target $U_t = \dot{\alpha}_t X + \dot{\beta}_t Z$. Let $p_t(\cdot, \cdot \mid y)$ denote the conditional law of (X, Z) given $Y_t = y$. For the exact field $v_t^*(y) = \mathbb{E}_{(X,Z)\sim p_t(\cdot,\cdot|y)}[U_t]$, define

$$\ell_t^*(y) = \mathbb{E}_{(X,Z)\sim p_t(\cdot,\cdot|y)}\left[\|U_t - v_t^*(y)\|^2\right].$$

Eliminating Z at fixed y gives $U_t - v_t^*(y) = -\alpha_t \kappa_t (X - D_t(y))$. The Gaussian posterior covariance identity then yields

$$\frac{\ell_t^*(\alpha_t x)}{\beta_t^2 \kappa_t^2} = \frac{\operatorname{tr} \operatorname{Cov}_{X \sim p_t(\cdot|\alpha_t x)}(X)}{\varepsilon^2} = R_\varepsilon(x). \tag{141}$$

The last equality uses the covariance identity (134) in the normalized coordinates. Cha et al. use the conditional error for sampling; the displayed normalization identifies it with our dimension readout. Replacing v_t^* by a learned field inside the conditional squared loss adds $\|\widehat{v}_t(y) - v_t^*(y)\|^2$ to $\ell_t^*(y)$, so that loss is not automatically the fitted divergence readout.

J.4 Denoising loss: a Gaussian-noise average of Full and Response

Yeats et al. (2025) also estimate LID from Gaussian denoising loss, independently of their FM-FLIPD experiments. At a fixed clean query x, their population statistic can be written as

$$L_\varepsilon(x) = \mathbb{E}_{Z\sim\mathcal{N}(0,I_D)}\left[\frac{\|b_\varepsilon(x+\varepsilon Z)-x\|^2}{\varepsilon^2}\right]. \tag{142}$$

To see its equality with the noise-prediction loss, write the predictor as $\widehat{Z}(y) = (y - b_\varepsilon(y))/\varepsilon$. Then $Z - \widehat{Z}(x+\varepsilon Z) = (b_\varepsilon(x+\varepsilon Z) - x)/\varepsilon$. Yeats et al. motivate dimension recovery through the surviving tangent-noise components and give a small-noise averaged-loss bound. The next identity expresses this established estimator in our notation by Gaussian integration by parts; it does not introduce a new denoising estimator.

Proposition J.1 (Finite-width denoising-loss identity)**.** *Fix x and $\varepsilon > 0$. Suppose the reconstruction and its derivatives permit Gaussian integration by parts under $y = x + \varepsilon Z$, with finite displayed expectations and vanishing boundary terms. Then*

$$L_\varepsilon(x) = \mathbb{E}_{Z\sim\mathcal{N}(0,I_D)}\left[F_\varepsilon(x+\varepsilon Z) + R_\varepsilon(x+\varepsilon Z) - D\right]. \tag{143}$$

The same equality holds for a differentiable fitted reconstruction, using the algebraic readouts $\widehat{R}_\varepsilon = \operatorname{div}\widehat{b}_\varepsilon$ and $\widehat{F}_\varepsilon = \widehat{R}_\varepsilon + \left\|\widehat{b}_\varepsilon - y\right\|^2/\varepsilon^2$.

Proof. Put $r_\varepsilon(y) = b_\varepsilon(y) - y$. Expanding the square in (142) and using $\mathbb{E}_{Z\sim\mathcal{N}(0,I_D)}\|Z\|^2 = D$ gives

$$L_\varepsilon(x) = \mathbb{E}_{Z\sim\mathcal{N}(0,I_D)}\left[\frac{\|r_\varepsilon(x+\varepsilon Z)\|^2}{\varepsilon^2} + \frac{2Z^\top r_\varepsilon(x+\varepsilon Z)}{\varepsilon} + D\right].$$

Gaussian integration by parts replaces the middle term's expectation by twice the expectation of $\operatorname{div} r_\varepsilon(x+\varepsilon Z) = R_\varepsilon(x+\varepsilon Z) - D$. Therefore $L_\varepsilon(x) = \mathbb{E}_{Z\sim\mathcal{N}(0,I_D)}[C_\varepsilon(x+\varepsilon Z) + 2R_\varepsilon(x+\varepsilon Z) - D]$, proving the result. This calculation applies to any reconstruction satisfying the stated analytic conditions. □

The average in (143) holds the clean point fixed and varies the observation. It differs from the fixed-observation posterior expectation in (141), and is generally neither $F_\varepsilon(x)$ nor $R_\varepsilon(x)$. For the exact posterior mean, total expectation nevertheless gives

$$\mathbb{E}_{X\sim\mu}[L_\varepsilon(X)] = \mathbb{E}_{Y\sim\rho_\varepsilon}[R_\varepsilon(Y)], \tag{144}$$

whenever these averages are finite: both sides equal the mean posterior squared reconstruction error divided by ε^2.

This distinguishes the local readouts from MMSE dimension (Wu & Verdú, 2011), which uses the globally averaged optimal reconstruction error in a high-signal-to-noise limit. In the Gaussian channel, its finite-noise precursor is precisely the average in (144), not the value at an individual query. Likewise, the conditional score-target variance underlying the loss-floor analysis of Raju & Zhang (2026) has trace $R_\varepsilon(y)/\varepsilon^2$: the conditional score target is $-(y - X)/\varepsilon^2$, whose covariance is $\operatorname{Cov}_{X\sim p_\varepsilon(\cdot|y)}(X)/\varepsilon^4$. Spatial or time integration changes this local quantity into an averaged objective.

Yeats et al. also compare FLIPD with the implicit score-matching expression $I_\varepsilon(y) = \varepsilon^2(\operatorname{div} s_\varepsilon(y) + \|s_\varepsilon(y)\|^2/2)$. Direct substitution shows

$$D + I_\varepsilon(y) = R_\varepsilon(y) + \tfrac{1}{2}C_\varepsilon(y), \qquad F_\varepsilon(y) = D + I_\varepsilon(y) + \tfrac{1}{2}C_\varepsilon(y).$$

Thus that score-matching expression contains half the correction; it is not Response alone.

J.5 Geometric conditions and approximation error

Corollary J.2 (Pointwise LID limit of Gaussian denoising loss)**.** *Let $\mathcal{M} \subset \mathbb{R}^D$ be a compact embedded C^2 manifold of dimension d, without boundary and with positive reach, meaning that points in a fixed neighborhood of $\mathcal{M}$ have a unique nearest point on it. Let $\mu = q\mathcal{H}^d \lfloor_{\mathcal{M}}$, where q is continuous and strictly positive. For the exact Gaussian posterior mean and every $x \in \mathcal{M}$,*

$$\frac{b_\varepsilon(x+\varepsilon Z)-x}{\varepsilon} \longrightarrow P_{T_x\mathcal{M}} Z \quad \text{in } L^2(Z\sim\mathcal{N}(0,I_D)), \qquad L_\varepsilon(x) \longrightarrow d. \tag{145}$$

In the same setting, $Jb_\varepsilon(x) \to P_{T_x\mathcal{M}}$ and $R_\varepsilon(x), F_\varepsilon(x) \to d$.

Proof. Fix x and put $T = T_x\mathcal{M}$. Parametrize a neighborhood of x as a graph over T, with vanishing normal derivative at x, and rescale its coordinates by ε. For each fixed z, the posterior quotient becomes a Gaussian-weighted integral over this rescaled graph, giving

$$\frac{b_\varepsilon(x+\varepsilon z)-x}{\varepsilon} \longrightarrow \frac{\int_T u \exp(-\|z-u\|^2/2)\,\mathcal{H}^d(\mathrm{d}u)}{\int_T \exp(-\|z-u\|^2/2)\,\mathcal{H}^d(\mathrm{d}u)} = P_T z.$$

Continuity of q cancels its value at x from the ratio. The rescaled chart tends to T, and Gaussian decay controls its moment integrals; mass outside a fixed chart neighborhood is exponentially small for fixed z. To pass this pointwise limit through the noise expectation, we now give a uniform bound and control the remaining tail.

Choose a tube radius r smaller than half the reach. For y in this tube, write $p = \pi_{\mathcal{M}}(y)$ for its nearest point. The reach bound on normal chords implies constants $c_1, c_2 > 0$, uniform in y and $u \in \mathcal{M}$, such that

$$c_1\|u-p\|^2 \le \|y-u\|^2 - \|y-p\|^2 \le c_2\|u-p\|^2.$$

Compactness, the local volume bounds of a smooth manifold, and the positive lower and upper bounds on q therefore give

$$\int_{\mathcal{M}} e^{-\|y-u\|^2/(2\varepsilon^2)} q(u)\,\mathcal{H}^d(\mathrm{d}u) \ge c\varepsilon^d e^{-\operatorname{dist}(y,\mathcal{M})^2/(2\varepsilon^2)},$$
$$\int_{\mathcal{M}} \|u-p\|\, e^{-\|y-u\|^2/(2\varepsilon^2)} q(u)\,\mathcal{H}^d(\mathrm{d}u) \le C\varepsilon^{d+1} e^{-\operatorname{dist}(y,\mathcal{M})^2/(2\varepsilon^2)}.$$

Taking their ratio shows $\|b_\varepsilon(y)-p\| \le C'\varepsilon$. When $y = x+\varepsilon z$ and $\varepsilon\|z\| < r$, the nearest-point property gives $\|p-x\| \le 2\varepsilon\|z\|$, hence

$$\frac{\|b_\varepsilon(x+\varepsilon z)-x\|}{\varepsilon} \le C' + 2\|z\|.$$

Its square is integrable under $Z\sim\mathcal{N}(0,I_D)$. Dominated convergence therefore applies on the growing sets $\{z : \varepsilon\|z\| < r\}$.

On the complement, $b_\varepsilon(y)$ lies in the convex hull of $\mathcal{M}$, so

$$\mathbb{E}_{Z\sim\mathcal{N}(0,I_D)}\left[\frac{\|b_\varepsilon(x+\varepsilon Z)-x\|^2}{\varepsilon^2}\mathbf{1}_{\{\|Z\|\ge r/\varepsilon\}}\right] \le \frac{\operatorname{diam}(\mathcal{M})^2}{\varepsilon^2}\mathbb{P}_{Z\sim\mathcal{N}(0,I_D)}(\|Z\|\ge r/\varepsilon) \longrightarrow 0.$$

The projected Gaussian second moment on the same tail also vanishes. This proves the L^2 limit, and $\mathbb{E}_{Z\sim\mathcal{N}(0,I_D)}\|P_T Z\|^2 = \operatorname{tr} P_T = d$ gives the loss limit.

For the clean-query statements, apply the same local-chart calculation to the first two posterior moments of $(X-x)/\varepsilon$ conditional on $Y_\varepsilon = x$. Gaussian decay controls these polynomially weighted integrals. Their limits are the mean zero and covariance P_T of the standard Gaussian restricted to T. Equation (134) gives $Jb_\varepsilon(x) \to P_T$, while $C_\varepsilon(x) = \|(b_\varepsilon(x)-x)/\varepsilon\|^2 \to 0$. Taking the trace and using $F_\varepsilon = R_\varepsilon + C_\varepsilon$ proves the remaining claims. □

These are sufficient conditions for the averaged limit, rather than an assertion that the clean-query Response theorem alone implies it. Along $y = x + \varepsilon Z$, the same L^2 limit gives

$$\mathbb{E}_{Z\sim\mathcal{N}(0,I_D)}[C_\varepsilon(x+\varepsilon Z)] \longrightarrow D - d.$$

Thus the correction does not vanish on typical corrupted queries: their normal distance is of order ε. Equation (143) consequently has the limiting balance $2d + (D-d) - D = d$, consistent with the vanishing correction at the clean interior point.

For a learned reconstruction, define

$$a_\varepsilon(x) = \left(\mathbb{E}_{Z\sim\mathcal{N}(0,I_D)} \frac{\left\| \widehat{b}_\varepsilon(x+\varepsilon Z) - b_\varepsilon(x+\varepsilon Z) \right\|^2}{\varepsilon^2} \right)^{1/2}.$$

The triangle inequality gives $|\sqrt{\widehat{L}_\varepsilon(x)} - \sqrt{L_\varepsilon(x)}| \le a_\varepsilon(x)$, so $a_\varepsilon(x) \to 0$ transfers the LID limit. This controls reconstruction values in a noise neighborhood; Response instead requires control of the reconstruction's derivatives at its query. Finally, (143) uses Gaussian integration by parts and is not asserted unchanged for arbitrary non-Gaussian noise.

For completeness, sufficient clean-query approximation conditions can be stated in the same notation. If $\mathrm{tr}(J\widehat{b}_\varepsilon(x) - Jb_\varepsilon(x)) \to 0$, then $\widehat{R}_\varepsilon(x) \to d$. For $\widehat{F}_\varepsilon(x) \to d$ in the smooth setting above, it is sufficient in addition that $\left\| \widehat{b}_\varepsilon(x) - b_\varepsilon(x) \right\| / \varepsilon \to 0$. These conditions concern the model fields; a numerical trace estimate must also have vanishing estimation error for a consistency statement. None follows solely from a small globally averaged training loss.

J.6 Spectral diagnostics and the boundary of the equivalences

Trace and eigenvalue counts differ at finite width: eigenvalues $(0.8, 0.4)$ sum to 1.2, whereas a threshold of $1/2$ counts one direction. Kim's thresholded denoiser statistic therefore is not its trace. Under the exact Gaussian projector limit established above, the eigenvalues of the normalized denoiser Jacobian approach d ones and $D-d$ zeros. Any fixed threshold strictly between zero and one then eventually counts d directions. This does not identify the two finite-noise statistics or establish a learned spectral algorithm's error guarantees.

Other methods form different matrices. Normal Bundle (Stanczuk et al., 2024) samples nearby score vectors and estimates the number of normal directions from their singular values. The error-bundle variant of Yeats et al. (2025) instead forms a Gram matrix of noise-prediction errors: its trace is the empirical denoising loss, but its count of non-negligible eigenvalues is another statistic. Energy Matching (Balcerak et al., 2025) counts near-zero eigenvalues of an energy Hessian, while ID-NF/ID-DM (Horvat & Pfister, 2022; 2024) examine singular values of generative transport Jacobians. The posterior-Jacobian trace identities above do not establish finite-width equality for these other matrices or their thresholding procedures. Our Full and Response benchmark rows do not evaluate these spectral counts.

J.7 What changes beyond Gaussian noise

The reconstruction–velocity identity (136), and hence (137), use the independent affine path but not a Gaussian source. For admissible non-Gaussian sources, the main theorem therefore retains the Response limit at a smooth interior point and allows its computation from velocity alone. The Gaussian score conversion and nonnegative squared-displacement correction need not survive: a general Full readout also needs the conditional mean of the kernel score. Student t-EDM and t-Flow (Pandey et al., 2024) are existing generators to which this distinction applies; their model construction is prior work, whereas the general-kernel LID guarantee is the extension made here.

The bridge conditional-mean construction of Pooladian & Niles-Weed (2025) and the noise-conditioned normalizing-flow family of SoftFlow (Kim et al., 2020) likewise provide model interfaces, rather than prior instances of all the LID results in this paper. Their use as ingredients should be distinguished from the exact estimator equivalences proved above. For density-conditioned normalizing flows, the relation to the differential LIDL criterion is already contained in the first subsection of this appendix.

DERIVATIONS SUPPORTING SECTION 4

K FULL READOUTS FOR NON-GAUSSIAN AFFINE FLOW MATCHING

Affine flow matching always provides a posterior-mean interface, but a general source law need not provide the marginal score from that mean at one query. This appendix derives the extra regression used for logistic FM, the Gaussian and radial-Student formulas that avoid it, and the scope of this distinction. All identities below concern population conditional means; substituting learned fields introduces estimation error.

K.1 TWO CONDITIONAL MEANS REQUIRED BY FULL

Let $X \sim \mu$ and $Z \sim K$ be independent, with a differentiable source density K. Set $Y_t = \alpha_t X + \beta_t Z$ and $\lambda = \beta_t/\alpha_t$, as in Corollary F.1. For a canonical observation $r = Y_t/\alpha_t$, write $\Pi_{\lambda,r}$ for the conditional law of Z given $X + \lambda Z = r$. Define the *negative* kernel score $\psi_K(z) = -\nabla_z \log K(z)$ and the two conditional means

$$m_\lambda(r) = \mathbb{E}_{Z\sim\Pi_{\lambda,r}}[Z] = \frac{r - b_\lambda(r)}{\lambda}, \qquad a_\lambda(r) = \mathbb{E}_{Z\sim\Pi_{\lambda,r}}[\psi_K(Z)]. \tag{146}$$

This sign convention is opposite to $\psi_K = \nabla \log K$ in Appendix B. Differentiation of the convolution gives

$$\begin{aligned} \nabla_r \log \rho_\lambda(r) = -\frac{a_\lambda(r)}{\lambda}, \qquad R_\lambda(r) = \operatorname{tr} J b_\lambda(r), \\ F_\lambda(r) := D + \partial_{\log\lambda} \log \rho_\lambda(r) = R_\lambda(r) + m_\lambda(r)^\top a_\lambda(r). \end{aligned} \tag{147}$$

The first identity requires integrable kernel derivatives; the second is (3). Direct scale differentiation also gives $F_\lambda(r) = \mathbb{E}_{Z\sim\Pi_{\lambda,r}}[Z^\top \psi_K(Z)]$.

The FM velocity supplies the first conditional mean because

$$v_t^*(\alpha_t r) = \dot\alpha_t b_\lambda(r) + \dot\beta_t m_\lambda(r) = a_t \alpha_t r + \alpha_t \kappa_t (r - b_\lambda(r)). \tag{148}$$

Here a_t and κ_t are the schedule coefficients from (64), distinct from the vector a_λ above. Consequently one velocity field determines b_λ and Response whenever $\kappa_t \neq 0$. It does not in general determine a_λ by a pointwise function of m_λ:

$$\mathbb{E}_{Z\sim\Pi_{\lambda,r}}[\psi_K(Z)] \neq \psi_K\big(\mathbb{E}_{Z\sim\Pi_{\lambda,r}}[Z]\big)\,.$$

For example, the one-dimensional logistic score is proportional to $\tanh(z/(2s))$. A point mass at zero and a distribution assigning probabilities $2/3, 1/3$ to $-s, 2s$ have the same mean zero, but their mean scores are respectively zero and $[-2\tanh(1/2) + \tanh(1)]/(3s) \neq 0$. Both can arise as conditional noise laws for suitable discrete data priors, since the logistic density is everywhere positive.

This is a limitation of a local mean-only conversion, not an impossibility of obtaining density information from an entire velocity field. Integrating the continuity equation along its characteristics recovers a model density, and differentiating that construction can supply a score, at additional ODE and derivative cost. Separate regression is another route, consistent with the distinct velocity and score objectives in stochastic-interpolant models (Albergo et al., 2025). Nor is two a mandatory number of network instances: shared architectures with multiple outputs are possible. Our benchmark uses two copies of the prescribed field backbone.

K.2 THE PRODUCT-LOGISTIC IMPLEMENTATION

Each source coordinate has unit-variance logistic density

$$k(z) = \frac{1}{4s}\operatorname{sech}^2\Big(\frac{z}{2s}\Big), \quad s = \frac{\sqrt{3}}{\pi}, \qquad K(z) = \prod_{j=1}^{D} k(z_j), \qquad [\psi_K(z)]_j = \frac{1}{s}\tanh\Big(\frac{z_j}{2s}\Big)\,. \tag{149}$$

For $Y_t = tX + (1-t)Z$, two separately parameterized fields regress $X - Z$ and $\psi_K(Z)$. With $t \sim \mathcal{U}(0,1)$ independent of $X \sim \mu$ and $Z \sim K$, their joint objective is

$$\mathbb{E}_{\substack{X\sim\mu,\, Z\sim K,\, t\sim\mathcal{U}(0,1)\\ X,\, Z,\, t \text{ independent}}}\left[\frac{\|v_\theta(Y_t,t)-(X-Z)\|^2}{D}+\frac{\|h_\phi(Y_t,t)-\psi_K(Z)\|^2}{D}\right]. \tag{150}$$

Squared-loss minimization gives the two corresponding conditional means. Both fields see the same corruption draw; their estimation errors need not be statistically independent. At $t = (1+\lambda)^{-1}$,

$$b_{\theta,\lambda}(r) = tr + (1-t)v_\theta(tr,t), \qquad \widehat{F}_\lambda(r) = \operatorname{tr} J_r b_{\theta,\lambda}(r) + \frac{(r-b_{\theta,\lambda}(r))^\top h_\phi(tr,t)}{\lambda}. \tag{151}$$

The Jacobian includes the native-input rescaling by t. The implementation encodes time as $999t$ for both fields. It uses coordinate-mean MSE, sums the two losses, and updates both cores on each training step. Thus each core receives the benchmark's update budget, while parameter count and training computation exceed those of a single-core FM. Inverse-CDF noise sampling clips the uniform variate to $[10^{-7}, 1-10^{-7}]$ to avoid infinite floating-point values; the identities describe the ideal continuous logistic law. The main row uses the same data partitions, normalization, checkpoint-selection rule and readout-scale protocols as the other benchmark rows. It uses neither an auxiliary density model nor an ODE-based density evaluation.

K.3 Why Gaussian and radial Student admit one-field readouts

Gaussian: affine kernel score. For $K = \mathcal{N}(0, I_D)$, $\psi_K(z) = z$, so $a_\lambda = m_\lambda$ and $F_\lambda = R_\lambda + \|r - b_\lambda\|^2/\lambda^2$. A non-isotropic Gaussian similarly has an affine score and a known quadratic-form correction. More generally, if the expectation of a continuous function ψ_K were determined by the mean for every finite-support probability law on $\mathbb{R}^D$, point masses and two-point mixtures imply the Jensen identity, forcing ψ_K to be affine. For a positive, normalizable smooth density on all of $\mathbb{R}^D$, an affine score integrates to a Gaussian log-density. This is the scope of the Gaussian mean-only exception; it does not rule out identities using derivatives or additional conditional moments.

Radial Student: a scale-derivative identity. Use the unit-coordinate-variance multivariate Student kernel

$$K(z) \propto \left(1+\frac{\|z\|^2}{c}\right)^{-(D+\nu)/2}, \quad c = \nu - 2 > 0, \qquad \psi_K(z) = \frac{D+\nu}{c+\|z\|^2}z. \tag{152}$$

The score is nonlinear, but its radial factor satisfies

$$z^\top \psi_K(z) = D + \nu - c\frac{D+\nu}{c+\|z\|^2}. \tag{153}$$

At fixed canonical r, differentiating the posterior mean gives

$$\dot{b}_\lambda(r) := \partial_{\log\lambda} b_\lambda(r) = -\lambda \operatorname{Cov}_{Z\sim\Pi_{\lambda,r}}\big(Z, Z^\top\psi_K(Z)\big), \tag{154}$$

$$\frac{\dot{b}_\lambda}{\lambda} = ca_\lambda - (D+\nu-F_\lambda)m_\lambda. \tag{155}$$

The covariance is between the vector Z and its scalar score contraction, under the explicitly defined conditional noise law. Substituting the second identity into (147), with $e = r - b_\lambda$ and $h = \|e\|^2/(c\lambda^2)$, yields

$$F_\lambda = \frac{R_\lambda + (D+\nu)h + e^\top\dot{b}_\lambda/(c\lambda^2)}{1+h}, \qquad C_\lambda := F_\lambda - R_\lambda = \frac{(D+\nu-R_\lambda)h + e^\top\dot{b}_\lambda/(c\lambda^2)}{1+h}. \tag{156}$$

This is the implemented Student readout: one posterior field, its spatial trace, and one scale derivative, rather than a second learned score field. The identity relies on a single shared radial scale for the entire noise vector. A product of independent univariate Student coordinates does not satisfy this scalar radial relation; the same formula must not be used for that different source law.

Other kernels and the scope of the claim. Gaussian and Student are not the only imaginable derivative-based cases. For example, the smooth compactly supported radial density $K(z) \propto (1 - \|z\|^2/c)_+^\eta$, $\eta > 1$, has $\psi_K(z) = 2\eta z/(c - \|z\|^2)$ inside its support. The same covariance calculation gives, wherever differentiation is justified and $h < 1$,

$$F_\lambda = \frac{R_\lambda + 2\eta h - e^\top \dot{b}_\lambda/(c\lambda^2)}{1-h}, \qquad h = \frac{\|e\|^2}{c\lambda^2}. \tag{157}$$

We do not benchmark this kernel. Its vanishing boundary weight and integrable derivatives are essential; this argument does not extend to uniform noise by setting $\eta = 0$. Gamma or beta labels alone also do not establish the needed identities for additive corruption: their support and kernel-boundary terms require separate treatment. An exponential-family posterior identity generally concerns the natural parameter of that observation model; it does not automatically recover the clean additive signal or make Student a Gaussian exponential-family special case. Our claim is therefore the availability of the displayed Gaussian and radial-Student formulas, and the use of an additional conditional-score regression for logistic FM, not a classification of all possible one-network estimators.

L DSB: Posterior Error and Finite-Scale Bias Derivations

Use the terminal-data orientation of Corollary G.1. Write $\tau = T - t$, $\lambda^2 = \gamma\tau$ (so λ is the width ε of Table 1), and $\varepsilon_{\mathrm{SB}} = \gamma T$ for the bridge's entropic regularization. For terminal law μ and static Schrödinger potential g, the weighted endpoint measure is $G(\mathrm{d}z) = e^{g(z)/\varepsilon_{\mathrm{SB}}}\mu(\mathrm{d}z)$. The exact posterior is

$$\pi_\lambda(\mathrm{d}z \mid x) = \frac{\exp\!\Big(-\frac{\|z-x\|^2}{2\lambda^2} + \frac{g(z)}{\varepsilon_{\mathrm{SB}}}\Big)\,\mu(\mathrm{d}z)}{\int \exp\!\Big(-\frac{\|u-x\|^2}{2\lambda^2} + \frac{g(u)}{\varepsilon_{\mathrm{SB}}}\Big)\,\mu(\mathrm{d}u)}. \tag{158}$$

Let $q_\lambda^*(x) = \int z\,\pi_\lambda(\mathrm{d}z \mid x)$ and denote its Response and Full readouts by R_λ^* and F_λ^*. Differentiating the posterior with respect to x gives

$$R_\lambda^* = \frac{\operatorname{tr}\operatorname{Cov}_{\pi_\lambda}(Z)}{\lambda^2}, \qquad C_\lambda^* = \frac{\|q_\lambda^* - x\|^2}{\lambda^2}, \qquad F_\lambda^* = R_\lambda^* + C_\lambda^* = \frac{\mathbb{E}_{\pi_\lambda}\|Z - x\|^2}{\lambda^2}. \tag{159}$$

Thus posterior moments suffice to evaluate Full without differentiating a learned drift. Removing $g/\varepsilon_{\mathrm{SB}}$ from (158) gives an untilted Gaussian-corruption reference on the same data law; this isolates the effect of the Schrödinger weight at fixed x and λ.

For a learned posterior $q_{\theta,\lambda} = q_\lambda^* + \delta_\lambda$, expansion of its Full readout yields

$$F_{\theta,\lambda} - d = \underbrace{\operatorname{tr} J\delta_\lambda + \frac{2(q_\lambda^* - x)^\top\delta_\lambda + \|\delta_\lambda\|^2}{\lambda^2}}_{\text{learned-posterior error}} + \underbrace{F_\lambda^* - d}_{\text{finite-scale bias}}. \tag{160}$$

This is (107) in posterior coordinates, with the population bias displayed separately. Approximate trace computation adds its own numerical error. Small posterior value error alone does not control either its Jacobian error or the terms divided by λ^2.

The endpoint theorem makes the second term vanish under its assumptions, but supplies no model-independent scale at which it becomes small. To see the role of the bridge weight, consider an interior point of an affine d-plane with positive smooth density p and write $w = e^{g/\varepsilon_{\mathrm{SB}}}p$. Under the usual local Taylor and tail bounds, Gaussian convolution within that plane gives, at fixed $\varepsilon_{\mathrm{SB}}$,

$$q_\lambda^* - x = \lambda^2\nabla_M \log w(x) + o(\lambda^2), \tag{161}$$

$$F_\lambda^* - d = \lambda^2\left(\Delta_M \log w(x) + \|\nabla_M \log w(x)\|^2\right) + o(\lambda^2). \tag{162}$$

The displacement contribution alone is $C_\lambda^* = \lambda^2\|\nabla_M \log w(x)\|^2 + o(\lambda^2)$. Since $\nabla_M \log w = \nabla_M \log p + \nabla_M g/\varepsilon_{\mathrm{SB}}$, strong reweighting can delay the small-bias regime. The Laplacian term also matters, so the entire bias cannot be identified with the positive correction alone. This affine calculation illustrates the mechanism; curvature and boundaries add finite-scale effects on the benchmark geometries.

M RESPONSE AND CORRECTION: POPULATION GEOMETRY

This appendix establishes when the population correction is geometrically necessary. Learned-model diagnostics are given in Appendix P. We use $R_\lambda = \operatorname{tr} Jb_\lambda$ for Response, $F_\lambda = D + \partial_{\log\lambda} \log \rho_\lambda$ for Full, and $C_\lambda = F_\lambda - R_\lambda$ for the entire correction in (3). For Gaussian noise, $C_\lambda = \|b_\lambda - x\|^2 / \lambda^2$; for radial Student it is the different expression in (156).

M.1 REGULAR POINTS AND ASYMMETRIC TANGENT GEOMETRY

Proof of Corollary 2.4. By Appendix D, (A1) holds at x with the flat tangent measure $q(x)\mathcal{H}^d \lfloor_{T_x\mathcal{M}}$, where q is the density on $\mathcal{M}$. Corollary 2.3 therefore gives $\mathcal{R}^{\mathrm{resp}}_\varepsilon(x) \to d$, while (5) gives $\mathcal{R}^{\mathrm{resp}}_\varepsilon(x) + (B^Q_\varepsilon(x) - x)^\top \nabla \log \rho^Q_\varepsilon(x) \to d$. Subtracting, the correction vanishes. For additive noise it equals C_λ with $\lambda = \varepsilon$ by Proposition 2.1. Equivalently, the limiting posterior displacement mean lies in $T_x\mathcal{M}$, whereas the mean log-kernel gradient is orthogonal to $T_x\mathcal{M}$ by tangent integration by parts. Their inner product therefore vanishes, even when the kernel is not symmetric. For Gaussian noise the last limit is also $\|b_\lambda(x) - x\| / \lambda \to 0$. □

The conclusion is pointwise under the stated positive-density and tail conditions. It is a statement about dimension, not recovery of the whole data distribution. It neither assumes nor establishes that real image distributions are smooth manifolds without boundary. A smooth-interior model justifies considering Response; proximity to an edge, a change of stratum, or failure to resolve the local geometry can still matter at the available scales. Smooth curvature itself does not create a nonzero limiting correction.

Reading the geometric illustration. Figure 1 uses a Gaussian kernel and uniform local surface area on a sheet in $\mathbb{R}^3$. Its three glyphs use, respectively, a plane, a half-plane, and two equally weighted perpendicular half-planes meeting along a line. The blue shading represents the conditional law of the clean point X given $X + \lambda Z = x$, with $Z \sim \mathcal{N}(0, I_3)$ independent of X. The posterior mean is $b_\lambda(x)$. The dashed ball shows possible small changes of the observation; its purple linearized image is $b_\lambda(x) + Jb_\lambda(x)\lambda u$, $\|u\| \le 1$. These semiaxes are λ times the eigenvalues of Jb_λ, not posterior standard deviations. Their sum divided by λ is Response. The orange shift gives the Gaussian correction through its squared length divided by λ^2. Dotted outlines are visual references obtained by flattening the ball into the displayed sheet plane or planes; they are not an additional estimator.

Let $W = (X - x)/\lambda$ under this conditional law in the local conical limit, and call its limiting distribution Π^W_x. Then $Jb_\lambda \to \operatorname{Cov}_{W\sim\Pi^W_x}(W)$ and $C_\lambda \to \left\|\mathbb{E}_{W\sim\Pi^W_x}[W]\right\|^2$. Gaussian radial integration on each two-dimensional sheet gives $\mathbb{E}_{W\sim\Pi^W_x}[\|W\|^2] = 2$, yielding the following values.

Table 5: Population limits depicted in Figure 1; all three local supports have dimension two.

Location	Eigenvalues of Jb_λ	Response	Correction
Interior	$1, 1, 0$	2	0
Boundary	$1, 1 - 2/\pi, 0$	$2 - 2/\pi$	$2/\pi$
Right-angle crease	$1, 1/2, 1/2 - 1/\pi$	$2 - 1/\pi$	$1/\pi$

At the crease, the hinge coordinate is standard Gaussian; the perpendicular coordinate is an equally weighted half-normal on either of two orthogonal rays. Its mean has squared norm $1/\pi$, and its two covariance eigenvalues are $1/2$ and $1/2 - 1/\pi$. This proves the crease row without requiring a single tangent plane there. The boundary row is Corollary D.1. The illustration magnifies these local limits for visibility; the outer rims of the drawn sheet are not extra boundaries in the local posterior calculation. A singularity need not produce a correction: a centrally symmetric tangent configuration can have zero posterior mean. The relevant feature is posterior asymmetry, not merely the absence of smoothness.

Relation to response-only diagnostics. The comparisons with earlier estimators, including denoiser traces, FM-FLIPD, denoising loss and spectral counts, are collected with their normalizations and proofs in Appendix J.

Part II: Experimental Supplement

Ablations and Extended Analyses

N Extended Benchmark Analyses

The September 16 native-task campaign supplies selected predictions, source-train holdout curves, and known-LID test curves. We retain all completed trainings and failed selections; preflight runs are excluded. The main table contains eight generative-model recipes and dependent Response readouts for VP diffusion, rectified flow, t-Flow and logistic FM under both known-LID selection protocols. The four author-defined controls, E1/E5 measurements, and complete dependent readouts are reported in Appendix U.

N.1 All primary and dependent readouts

The original campaign has 429 training cells (11 recipes times 39 conditions), including 44 generated calibration cells. Its export has 1,656 test-summary records: supervised, pointwise and historical-selector records on known-LID data, plus reference-selected E1/E5 records and dependent readouts. Of these, 120 records represent failed selection, corresponding to two readouts for 60 physical cells. We independently replayed all 1,536 available summary records from saved predictions and checked 1,067 prediction-artifact hashes against the completion manifests. The subsequent logistic-FM extension adds 15 known-LID training cells; its protocol and separate export are specified in Appendices K and T.

Response and Full use the same checkpoint and selected scale, so their differences concern the readout rather than separate training or tuning. For example, rectified flow changes from 0.249 (Full) to 0.239 (Response) on rendered data but from 0.095 to 0.129 on coefficients. For t-Flow, Response reduces the rendered and Arrows errors while both readouts retain the failed E1/E5 reference selections. These comparisons condition on Full's scale choices; they do not measure how a separately tuned Response estimator would perform. No uniformly preferable readout or causal explanation follows from this campaign.

N.2 Known-LID conditions

Tables 12 to 15 report raw test MAE for every canonical condition, and Table 16 for the generated calibration conditions; they must not be confused with the relative, suite-balanced aggregates. Geometric conditions have 1,000 test queries each; Arrows has 10,000. Bias, RMSE, prediction quantiles, selected scales and selection status remain in the accompanying `unified_results.csv`. No test-error threshold was used to exclude a method or cell.

Location-dependent slices require additional geometry metadata; the archived query predictions alone do not identify boundary distance or tangent spaces. Arrows retains nominal factor-count targets and is excluded from the strict constant-target aggregate.

N.3 Sensitivity to training-set size

The E1 reference scale is selected on the full-data model and reused at all 13 training-set sizes. The same 5,000 test images are evaluated in every condition. Table 17 shows absolute means at the two endpoints as well as the aggregate mean-drift score. The log-noise posterior control has the smallest primary-readout drift score (0.109), but its mean changes from 136.278 to 193.553; without true Fashion-MNIST LID this is a stability observation, not an accuracy claim. The score uses all 13 sizes, so the two endpoint means alone do not determine it.

DSB's reference mean peaks too late to leave the required three-point descending segment. Student and NF return a boundary knee, which the reference rule rejects. These failures propagate to each method's 13 sample-size cells. Paired query arrays are retained for successful measurements, but no bootstrap intervals or distributional drift analysis are claimed here.

N.4 Invariance and equivariance

Each transformation reuses the downscaled Fashion-MNIST reference scale. The errors in Table 18 compare paired predicted changes against 0, 4 or 8, rather than comparing two unrelated marginal means. The same reference selection failures affect DSB, Student and NF across all seven E5 cells. Saved query identifiers, class-label order and pinned canonical-file pairing contracts were checked; the raw benchmark images are not part of this archive.

The best available primary invariance score is 10.830 (rectified flow), and the best equivariance score is 32.121 (log-noise velocity control). Both exceed the constant-estimator diagnostics of 0 and 5.708, respectively. In particular, successful scale selection does not establish the expected response to a dimension-changing transformation.

N.5 Coefficient space versus high-dimensional rendering

The two representations use separate checkpoints and separately selected scales. Their difference combines ambient dimension, input representation, architecture and operating scale; it is not a causal representation ablation.

N.6 Scale curves and the admissible output range

All 429 original cells retain 29-scale source-train holdout curves. All 209 original known-LID cells additionally retain per-query test curves for the primary readout and, for fields, Response. E1/E5 retain selected-point predictions when reference selection succeeds; full test curves for these unknown-LID cells are not included. The historical narrow-grid selector is archived separately and does not enter the current primary tables.

All primary selected predictions in the 165 canonical known-LID cells are finite. The minimum fraction inside $[0, D]$, taken separately over cells, is 4.7% overall (DSB on coefficient-space Crescent). It is 100% for RF, trigonometric flow and Student, 96.4% for VP diffusion, and 6.0% for NF. These are range checks, not LID-accuracy guarantees: Student can remain in the ambient range while having very large error. No predictions are clipped to $[0, D]$ in the reported error tables.

N.7 NF density-slope readouts

One sigma-conditioned RealNVP is trained per task, rather than separate flows for each noise level. OLS5 is primary and chooses the shared scale. The new export also contains the exact fixed-point autograd derivative, recorded as `fixed_likelihood`; it is a LID statistic, not a raw likelihood used as a dimension estimate.

OLS5 evaluates the same model at $\lambda \exp(.05k)$, $k = -2, \ldots, 2$, fitting log density against log scale and adding D. The trained sigma support is approximately $[.003535, 70.731]$, so the full stencil fits around all centers in $[1/256, 64]$. The archive retains the resulting per-query slopes, not the five separate likelihood values used in each fit. A new stencil would therefore require inference with the saved weights, not retraining.

O Automatic Operating-Scale Selection

The current campaign separates label-free reference-mean selection for E1/E5 from pointwise selection on known-LID data. The latter is evaluated on the same fitted checkpoints and the same test queries as supervised source-train holdout selection. Neither automatic rule uses LID labels.

O.1 Implemented knee rule and candidate grids

The primary automatic protocol is `pointwise_kneedle_model_logspace_v4`. It applies `kneed` 0.8.6 to each primary-readout curve against $\log \lambda$, with sensitivity 1, convex shape, decreasing direction, offline operation and linear interpolation. Both this selector and supervised holdout selection use the same 29 model-specific geometric candidates (Appendix T.4). The comparison therefore

holds candidate-grid coverage fixed within each model, although different models have different supported scale windows.

If no knee is detected, the pointwise protocol returns ambient dimension. A detected grid endpoint is retained and counted as a boundary selection. Dependent Response uses the same indices and fallback rule as Full. The reference-mean E1/E5 selector is different: it starts at the mean curve's global maximum, requires at least three remaining points, and rejects a boundary knee. Its failure leaves the whole reference group unselected, rather than substituting ambient dimension.

The archive also retains a historical FLIPD-grid diagnostic constructed from 22 reference VP times in $(.05, .5)$, restricted to native support when necessary. It is not pooled with the current 29-scale comparison.

O.2 Coverage and selection stability

Across the original 11 primary methods and 15 canonical known-LID cells, all 264,000 query evaluations return a knee: no ambient-dimension fallback occurs. This does not imply an interior or accurate choice. Endpoint selections range from 0% for VP diffusion to 76.904% for Student and 75.017% for NF. Percentages count query evaluations, not independent training runs or cells. Perturbation stability and training-seed uncertainty remain unmeasured.

O.3 Gap to the controlled protocol

Pointwise Kneedle has higher error for each coefficient aggregate, rendered aggregate and Arrows MAE in Table 20. This holds for all 12 primary methods, including the later logistic-FM extension, and for the four dependent Response readouts of Table 2, which add no independent checkpoint or scale decisions. Supervised Arrows scores are in Tables 2 and 10. The E1/E5 reference selector is unchanged; its scores and the four author-defined controls are retained only in the appendix tables. The deterioration is not explained by no-knee fallback or unequal candidate-grid coverage within a model. Even VP, with no boundary selections, deteriorates.

This is a comparison of supervised dataset-level selection with a label-free pointwise rule, not a significance test or a result about all possible automatic selectors. It establishes that these learned curves do not make the implemented pointwise rule an adequate substitute for supervised holdout selection. It does not identify the cause of each curve's error.

P Student: Learned Correction and Posterior-Mean Diagnostics

The population geometry is developed in Appendix M. Here we test whether the fitted Student model estimates the correction accurately, using fixed-checkpoint component substitutions and a posterior-error profile.

P.1 Posterior-value error amplified by Full

We reuse the actual radial Student $\nu = 5$ checkpoints from the benchmark, with 128,000 training updates and seed zero. The diagnostic uses the first 32 source-training holdout queries of Spiral and Funnel/Exp images ($D = 784$, nominal $d = 1$ and 2). The weights, normalization, queries and scale are fixed while one component of the readout is replaced by its continuous-posterior reference. The reference integrates the known continuous generator with the radial Student likelihood, using an independent Gamma/Gaussian quadrature, rather than a finite-sample nearest neighbor or training-point posterior.

Table 21 reports MAE relative to nominal d, averaged over the same 32 queries. The primary scale is the archived Full-selected $\lambda = 64$, retained despite its unresolved upper-boundary status. Additional scales were chosen post hoc from the saved test Response curves to investigate the strongest readout contrasts; they are diagnostic operating points, not validated deployment choices. Response uses 128 fixed Rademacher trace probes, and the scale derivative uses JVP. Reference refinement and direct density-derivative checks are documented with the saved diagnostic.

Writing the Student Full correction in terms of the native noise predictor h_θ makes the amplification explicit. With $s = \lambda/\sqrt{5/3}$, $t = (1+s)^{-1}$ and $b_\lambda(x) = x - s h_\theta(tx, 1-t)$, let $\dot{h}_\theta = \partial_{\log\lambda} h_\theta(tx, 1-t)$ at fixed x. Equation (156) becomes

$$C_{\theta,\lambda} = \frac{(D + 4 - R_{\theta,\lambda}) \left\|h_\theta\right\|^2 - h_\theta^\top \dot{h}_\theta}{5 + \left\|h_\theta\right\|^2}. \tag{163}$$

When $R_{\theta,\lambda} \ll D$ and the predicted noise norm is small, the coefficient on that squared norm is approximately $D/5$. The component substitutions identify posterior-value error as the source of most of the extra Full error in these conditions: replacing only the posterior mean greatly reduces it, whereas replacing only the scale derivative does not. Such a hybrid may not correspond to a single differentiable posterior field and is not a proposed estimator. Residual Response error and population finite-scale bias remain; an accidentally small nominal MAE does not prove either component is correct.

P.2 Posterior error versus distance from the support

To measure posterior accuracy directly, we additionally evaluate the same two checkpoints at their original Full-selected $\lambda = 64$. For each of the same 32 holdout points x_i, draw two fixed unit vectors orthogonal to the entire affine PCA rendering span and include their negatives. Set $r_{ij} = x_i + \delta n_{ij}$, $j = 1, \ldots, 4$, keeping the conditioning scale fixed. Since every data point lies in that affine span, $\mathrm{dist}(r_{ij}, M) = \delta$ exactly. This probes a controlled set of normal directions, not all directions in the manifold's normal space or the distribution of ordinary training corruptions.

For a radial Student kernel, orthogonality gives $\left\|r_{ij} - X\right\|^2 = \left\|x_i - X\right\|^2 + \delta^2$ for every clean X. Consequently the reference posterior mean at (r_{ij}, λ) equals the mean at $(x_i, \lambda_{\mathrm{eff}})$, where

$$\lambda_{\mathrm{eff}}^2 = \lambda^2 + \frac{\delta^2}{\nu - 2}. \tag{164}$$

This is an exact reduction of the population integral. We evaluate the remaining continuous integral numerically and check quadrature refinement. The measured error, in the checkpoint's normalized data coordinates, is

$$\mathrm{RMSE}_b(\delta) = \left[\frac{1}{32 \cdot 4 \cdot 784} \sum_{i=1}^{32} \sum_{j=1}^{4} \left\|b_{\theta,\lambda}(r_{ij}) - b_\lambda^*(r_{ij})\right\|^2\right]^{1/2}, \tag{165}$$

where b_λ^* denotes the continuous-posterior reference. The zero-distance value repeats each clean query four times and adds no independent observations. We keep negative and positive directions paired throughout. No weights or scales are refitted, and no off-manifold Full value is interpreted as the manifold's LID.

At zero distance, RMSE is 0.708 for Spiral and 0.774 for Funnel. It remains almost unchanged through $\delta/\lambda = 0.3$, falls slightly at intermediate distances, and rises to 1.604 and 2.009 at $\delta/\lambda = 30$ (Figure 2). Thus the posterior mean is imperfect on the support, but these profiles do not show a special error peak there. Doubling both quadrature orders changes reference values by less than 7×10^{-14} in per-coordinate RMS on four queries per task at each distance.

The distance profile does not identify a cause in training coverage, optimization or model capacity. The component substitutions above, rather than a comparison of clean and noisy prediction losses, establish the sensitivity of Full to the measured value error.

What the training objective controls. With an absolutely continuous corruption kernel at a positive scale, the noisy observation law has an ambient density. A lower-dimensional data manifold therefore has zero probability under that law, although for our positive kernels it lies inside its support. Small average regression error on noisy inputs alone cannot guarantee accurate posterior values on the manifold, and it supplies no corresponding Jacobian guarantee (Appendix I). This mismatch between the training objective and pointwise LID evaluation is relevant to both diffusion and FM models. It does not identify the cause of the errors in these two Student fits: their distance profiles do not demonstrate a special loss of accuracy at clean inputs.

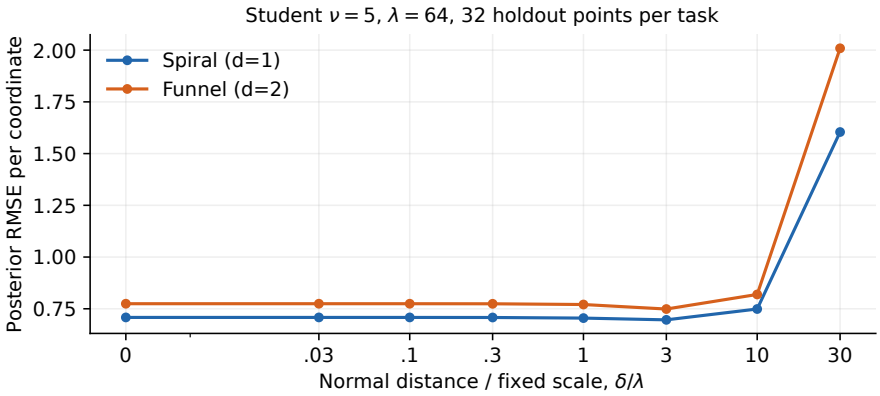

Figure 2: Posterior-mean RMSE versus exact normal distance in units of the fixed model scale, for the benchmark Student $\nu = 5$ image checkpoints. Each curve uses 32 fixed holdout points and two normal directions with both signs; the horizontal scale is linear near zero and logarithmic beyond .03. The target is the continuous posterior at the displaced observation, not the original clean point.

Q DSB: Learned-Field Error and Finite-Scale Bias

The large DSB errors in Table 2 raise two distinct questions: does the learned field approximate the bridge posterior, and would that posterior itself give an accurate dimension estimate at the evaluated scales? The decomposition in Appendix L separates these questions analytically. Here we compare saved neural readouts with numerical evaluations of the population posterior formula. The comparison is a diagnostic at fixed checkpoints, not a new model or scale-selection protocol.

Q.1 Fixed-checkpoint comparison with posterior references

Figure 3 uses two previously trained seed-0 checkpoints, each with 128,000 updates. Within each case, the neural and posterior curves use the same clean test queries and the checkpoint's normalization. We report the median Full LID and $\mathrm{MAE}(\lambda) = n^{-1}\sum_{i=1}^{n} |F_\lambda(x_i) - d|$, where d is the benchmark's nominal LID. These are errors relative to d, not pointwise distances between the learned and reference readouts. All saved scales are shown; no scale is selected on these query subsets.

Benchmark IPF-DSB. The first case is the actual Brownian DSB checkpoint from the main campaign on Spiral coefficients ($D = 30$, $d = 1$), restricted to the first 256 test queries. Its discretization has $\gamma = 2$, $T = .2$, and $\varepsilon = .4$, with accessible physical scales $\sqrt{.02} \le \lambda \le \sqrt{.4}$. We reuse the 29 saved neural scales and exact coordinate traces. The posterior reference solves the static entropic transport problem between 20,000 sampled training points and 20,000 standard Gaussian points, using 3,000 Sinkhorn iterations. Posterior moments are evaluated directly as finite weighted sums at seven scales in the same interval. These archived reference summaries have two-decimal precision. Their medians use the lower middle observation, which we also use for the corresponding neural summaries.

DSBM endpoint diagnostic. The second case uses a separately trained DSBM checkpoint (Shi et al., 2023) on Exp images ($D = 784$, $d = 2$), restricted to the first 64 test queries, with $\gamma = T = \varepsilon = 1$. This is not the DSB row in Table 2. It uses two width-8 U-Nets and iterative Markovian fitting: 32,000 initial bridge-matching updates per direction, six alternating 8,000-update stages, and 16,000 final dataward updates, with EMA. Its native orientation puts data at time zero, so the dataward posterior is $q_{\theta,\lambda}(x) = x + \lambda^2 b_\theta(x, \lambda^2)$. We reuse 29 neural scales from .01 to approximately 1, with 64 fixed Rademacher probes for each trace.

To probe smaller scales without reducing the data law to isolated atoms, we use the known continuous Exp generator

$$c(u, \vartheta) = (u - 4,\ 3e^{-u}\sin\vartheta,\ 3e^{-u}\cos\vartheta), \qquad u \sim \mathcal{U}[0, 8], \quad \vartheta \sim \mathcal{U}[0, 2\pi].$$

Its linear PCA rendering preserves coefficient-space distances up to the checkpoint's scalar normalization. The reference uses a 20,000-point Sinkhorn potential after 2,000 iterations, continuously extended by its entropic c-transform. At each query and scale, importance-weighted integration uses 100,000 generator samples in a local parameter box covering the image-space ball of radius 10λ. The reference grid extends down to 10^{-5}; we do not extrapolate the neural curve beyond its saved range. Both this reference and the finite-support Spiral reference approximate (158) numerically. They are not closed-form population solutions or certified error bounds: empirical potentials, local truncation, and Monte Carlo integration introduce approximation error. The small-scale Exp error levels off near the untilted reference's numerical floor. At coarse scales, strong tilting can also place mass outside the local integration box; no bound on that omitted mass is available.

What the comparison establishes. For the benchmark IPF-DSB, negative neural Full estimates already rule out finite-scale population bias as the sole explanation: the exact moment readout in (159) is nonnegative. The reference also develops large MAE toward the coarser end of the grid, despite a moderate median, revealing heterogeneous finite-scale bias. This case does not reach a small-scale regime that demonstrates endpoint recovery.

The separate Exp diagnostic exposes that regime. With the Schrödinger weight, the numerical reference has MAE 6.29 at $\lambda = .1$, falls to $.087$ at $.01$, and reaches approximately $.030$ at $.003$. The untilted errors at the first two scales are $.21$ and $.040$, respectively, so the weight substantially changes the usable scales. Meanwhile the learned model has MAE about 763 at $.01$ on the same 64 queries. Thus approaching the endpoint reduces posterior-reference bias while the learned field becomes inaccurate. This supports the finite-scale explanation without treating it as the only cause of the original DSB failure, or implying that a better scale selector can recover an accurate Full estimate from these checkpoints.

Q.2 Where the benchmark field loses the endpoint geometry

Further frozen-checkpoint diagnostics localize part of the learned-field error on Spiral coefficients. The data lie in a two-dimensional affine plane in $D = 30$, so an exact terminal posterior has zero response in the 28 directions orthogonal to that plane. This gives a derivative target without estimating the posterior along the spiral. Write the implemented readout as $B_{\theta,k}(x) = x + k\,\mathrm{inc}_\theta(x, T - k\Delta t)$, where inc_θ is the learned one-step dataward increment, $\Delta t = .01$, $T = .2$, and $\lambda^2 = 2k\Delta t$. For seven fixed orthonormal directions n_j in the normal space, define

$$g_k(x) = -\frac{k}{7}\sum_{j=1}^{7} n_j^\top J_x \mathrm{inc}_\theta(x, T - k\Delta t) n_j. \tag{166}$$

Correct normal collapse requires $g_k = 1$; the mean response of $B_{\theta,k}$ in these directions is $1 - g_k$.

We evaluate these directional derivatives by JVP on the first 64 test queries, then repeat after adding Gaussian offsets restricted to the normal space, with standard deviation λ per normal coordinate. The checkpoint, clean queries, directions and time condition are unchanged within each comparison. At $k = 1, 2, 4$, median g_k across clean queries is $1.33, 2.71, 5.26$, versus $.90, .93, .94$ at the displaced queries. The negative normal response at clean inputs thus reflects inaccurate derivatives of the learned field. These offsets are synthetic Gaussian probes, not samples from the learned bridge's training marginals. A separate sweep of radius and time on 256 fixed queries finds that, near the data plane, the normal increment per unit displacement changes little over $k = 1, 2, 4$ even though its correct magnitude scales as $1/k$. This is evidence of inadequate time dependence in this fitted field.

There is also an approximation in converting a learned discrete increment into a terminal posterior. A diagnostic simulation of 4,096 trajectories under the frozen backward network gives mean squared normal coordinate $.949$ at step 20, compared with $.4$ for the Brownian reference. A linear-Gaussian approximation using the measured chain moments predicts normal gain 1.48 for the ideal one-step regression target at that step. Thus even accurate regression of that approximate target need not make $x + k\,\mathrm{inc}_\theta$ an exact endpoint posterior. This calculation is a diagnostic approximation, not an exact solution of the learned chain. Together with the posterior-reference comparison, it identifies field and discretization errors in addition to finite-scale bridge bias; it neither establishes a single cause across the campaign nor reports a repaired DSB estimator.

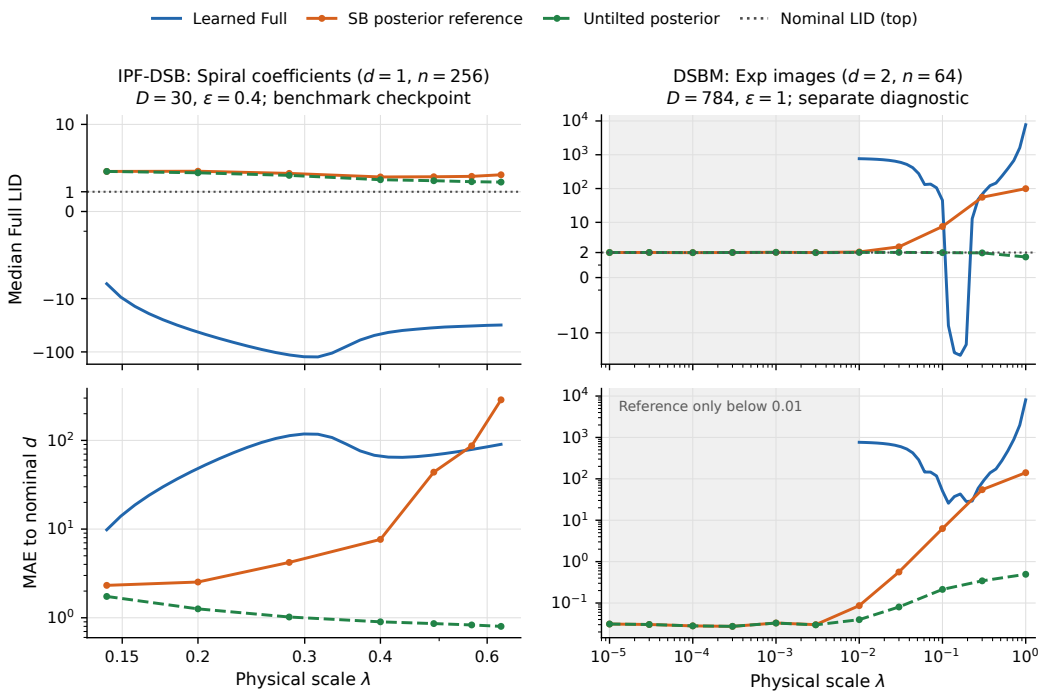

Figure 3: DSB Full LID (top) and MAE to nominal dimension (bottom) versus physical scale, for fixed checkpoints and matched test-query subsets. The SB posterior reference retains the Schrödinger weight; the untilted reference removes it. Both are numerical evaluations of the posterior moment formula. Lines connect saved values on each method's original grid; the shaded region has no neural evaluations. The upper vertical axes are symmetric-logarithmic to retain negative neural estimates; the lower ones are logarithmic.

R Verification and Scope of the Benchmark Evidence

R.1 Saved-prediction verification

The archive records checkpoint, source, split, and prediction hashes. We recomputed all available summary metrics from saved query predictions, replayed supervised selections from holdout MAE curves, and checked automatic predictions against their recorded curve indices. The audit covers the dependent Response readouts as well as the primary readouts; failed reference selections remain explicit missing measurements. This verifies numerical accounting from the retained predictions, without repeating training, model inference, or the Kneedle detector.

R.2 Gaussian readout equivalence

For a Gaussian affine channel, converting the same native field to a canonical posterior or to a score gives algebraically equivalent Full readouts (Equation (60)). These are two implementations of one functional, not independent estimators. The campaign records Full and Response without a separate FM-to-score row, so its tables provide no additional empirical parity measurement.

The derivative in the reported Response is taken in canonical data coordinates, including the chain rule through the native input (Appendix T.2). The archive does not supply a campaign-wide exact-trace or precision ablation. Consequently, aggregate errors alone cannot separate field approximation, derivative estimation, and finite-scale population bias.

R.3 Non-Gaussian endpoint kernels

For independent affine FM, the endpoint kernel is the source law after normalizing Y_t by α_t. A nonlinear data distribution therefore does not make Gaussian-source rectified flow a non-Gaussian-kernel method. The radial Student t-Flow and product-logistic FM rows supply the non-Gaussian sources in this benchmark. Their Response is the posterior Jacobian trace; the t-Flow Full readout uses the Student-specific identity in Appendix T.2, and the logistic Full readout uses the second, kernel-score network of Appendix K. The Gaussian displacement correction cannot be substituted for either.

The theorem concerns population endpoint limits under its stated assumptions. These learned finite-scale comparisons test the implemented recipes and selectors; poor accuracy does not by itself refute the population limit or identify a training mechanism.

S Pretrained Crystal and Protein Models

This appendix specifies the experiments in Section 4.3. All readouts use released model weights and clean experimental structures as queries. No model is trained or fine-tuned for LID estimation. The references are computed independently of model predictions: crystal symmetry describes admissible lattice deformations, while protein structure and MD provide measures of side-chain flexibility.

S.1 Crystal models, readout and symmetry reference

We use the DiffCSP `mp_csp/last.ckpt` release (Jiao et al., 2023) and the four OMatG MP-20-CSP checkpoints (Höllmer et al., 2025) listed in Table 6. The latter are distributed in the authors' public OMatG/MP-20-CSP repository. All models predict positions and lattice vectors conditional on atomic composition. Evaluation uses all 9,046 structures in the MP-20 test split (Xie et al., 2022). Each model retains its own published preprocessing: DiffCSP uses reduced cells; OMatG uses its configured fractional-coordinate conversion without Niggli reduction.

Only the lattice block enters the main comparison. Write the lattice as $L \in \mathbb{R}^{3\times3}$ and let $\widehat{b}_L$ be its reconstructed clean value. We compute $R_L = \operatorname{tr} J_{r_L}\widehat{b}_L$ by differentiating its nine output coordinates exactly, holding the other query coordinates fixed. Here r_L is the canonical input obtained by dividing the native noisy lattice by its clean-data coefficient. For DiffCSP, $L_t = c_t L + s_t Z_L$ with $Z_L \sim \mathcal{N}(0, I_9)$, giving $r_L = L_t/c_t$ and $\widehat{b}_L = r_L - (s_t/c_t)\widehat{Z}_L$. The other channel uses wrapped Gaussian perturbations of fractional positions. Thus DiffCSP provides the Gaussian lattice reference in this experiment.

For OMatG-VPSBD and Linear-ODE, the lattice path is $L_t = tL + (1-t)L_0$. The source L_0 is sampled using lognormal cell lengths and uniform cell angles, then converted to a lattice matrix. Its law is non-Gaussian and does not become Gaussian as $t \to 1$ under canonical scaling. The network velocity gives $\widehat{b}_L = L_t + (1-t)\widehat{v}_L$, differentiated with respect to $r_L = L_t/t$. The VPSBD name describes the positional channel; its lattice still follows this linear non-Gaussian path. VESBD-ODE and Linear-SDE-Gamma additionally use a Gaussian interpolant term $\gamma(t)Z_L$ and a trained noise head. Their reconstructions use both existing heads; no extra head is fitted here.

The reference is $d_{\mathrm{sym}} = \min(g+3, 9)$, where g counts the free cell-metric parameters: $6, 4, 3, 2, 2, 2, 1$ for triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal and cubic systems, respectively. The extra three account for lattice rotations in the models' equivariant matrix representation. The space group is extracted with spglib at tolerance 10^{-3}, with 10^{-5} and 10^{-2} used for sensitivity checks. This is a geometric reference for the symmetry family, rather than an assumption that each learned conditional distribution fills that entire family. Keeping the ambient lattice dimension fixed at nine avoids the atom-count confound of a full-coordinate trace.

Scale selection and metrics. We split structures by even and odd indices, select a scale on one half by minimizing the mean of $\log(1 + |R_L - d_{\mathrm{sym}}|/d_{\mathrm{sym}})$, and evaluate it on the other half; we then reverse the roles and pool the held-out predictions. Both halves select the same scale for each model. DiffCSP uses its 18-step grid and selects step 3, corresponding to canonical lattice scale 0.01732 and positional scale 0.00505. OMatG uses $1 - t \in$

$\{0.001, 0.002, 0.005, 0.01, 0.02, 0.03, 0.05, 0.07, 0.1, 0.15, 0.2, 0.3, 0.5\}$. MAE is the mean absolute difference from d_{sym}; Spearman correlation uses average ranks for ties. The final column gives the fraction with absolute error below one dimension. These are supervised scale-selection results, not a label-free calibration procedure.

Table 6: Lattice Response on 9,046 MP-20 test structures. All OMatG variants evaluated in the saved study are shown. The two random-weight controls use 400 structures and the corresponding trained-model scale.

Model	Selected scale	MAE ↓	ρ ↑	$\lvert\text{error}\rvert < 1$
DiffCSP	step 3	0.633	0.755	80.7%
OMatG VPSBD-ODE	$1 - t = 0.05$	0.826	0.677	70.5%
OMatG VESBD-ODE	$1 - t = 0.03$	0.905	0.724	62.8%
OMatG Linear-ODE	$1 - t = 0.30$	1.065	0.333	56.4%
OMatG Linear-SDE-Gamma	$1 - t = 0.005$	1.241	0.544	51.5%
DiffCSP, random weights	step 3	3.579	0.098	4.8%
OMatG, random weights	$1 - t = 0.05$	3.126	−0.046	4.8%

VPSBD-ODE is the OMatG variant with the lowest held-out lattice MAE among those evaluated. This choice is a comparison among measured variants, not a separately validated model-selection step. Changing variants also changes loss weights, so their differences do not isolate a causal effect of the source law. The positional block is excluded: native denoising checks fail for the VPSBD and VESBD position heads, and Linear-ODE and Linear-SDE-Gamma often give negative positional traces at clean queries. The positive lattice results therefore do not establish accurate full-structure LID estimates. Saved checks confirm the coordinate conversions through finite differences and the lattice rotation identity to numerical precision.

S.2 FlowPacker data, readout and independent references

We use the authors' `cluster.pth` checkpoint with its exponential-moving- average weights (Lee & Kim, 2025). FlowPacker conditions on sequence and backbone and predicts side-chain torsion velocities. Its training path starts from independent uniform angles and follows the shortest angular path to a clean conformation. For an ordinary 2π-periodic torsion, the conditional perturbation is a uniform arc of half-width $(1 - t)\pi$; the authors' encoding of π-symmetric torsions is retained. In locally unwrapped coordinates the reconstruction is $\widehat{b}_\chi(\chi_t, t) = \chi_t + (1 - t)\widehat{v}_\chi(\chi_t, t)$. The residue-level Response sums the diagonal Jacobian entries for its active torsions. We estimate those entries with 16 independent sign probes per torsion index, as opposed to a single trace over the entire protein. Three-protein exact-diagonal checks give mean absolute residue-trace differences of 0.020–0.025 at $1 - t = 0.1$ and 0.3.

The recent-PDB set contains 39 chains and 7,588 residues with torsions. Selection requires release on or after 1 March 2025, X-ray resolution at most 1.8 Å, a single protein chain of 120–350 residues, at least 95% complete side chains, and removal of chains with over 90% sequence identity to another selected chain. The ATLAS set contains 20 chains and 2,407 residues with torsions, selected at 80–200 residues, resolution at most 2.0 Å, no reconstructed crystal regions and mean RMSF below 2.5 Å, with one chain per UniProt entry. Both sets' identifiers are absent from FlowPacker's supplied training list; this does not exclude training homologues. Protein structures come from PDB (Berman et al., 2000); ATLAS supplies three 100-ns MD replicates per chain (Vander Meersche et al., 2024). The model is queried at the crystal structure, while MD supplies an independent reference.

References and statistics. rSASA is the Shrake–Rupley solvent-accessible surface area divided by the residue-type maximum of Tien et al. (2013). The B-factor reference averages crystallographic B-factors over side-chain atoms beyond C_β and standardizes them within each chain. For each torsion, MD variability is $1 - |\frac{1}{3000}\sum_{f=1}^{3000}\exp(ik\chi_f)|$, using 1,000 frames per replicate and $k = 2$ for π-symmetric torsions ($k = 1$ otherwise); residue values average over their torsions. These measure packing exposure, crystallographic displacement and simulated angular variability.

For each reference, we convert Response and reference values to percentile ranks separately within each amino-acid type with at least 20 valid residues, then compute Pearson correlation of the pooled

ranks. This controls the number and identity of torsions. For partial correlation, we regress the rSASA rank out of both ranked variables within each type before pooling. Proline AUC is the probability that a proline has lower Response than a comparison residue, with half credit for ties, averaged equally over the eight other two-torsion types. Confidence intervals resample whole proteins 1,000 times and recompute the statistic, preserving residue clustering.

Operating scale. The recorded rule selects the grid point where mean Response divided by the number of torsions, averaged over residues of the recent-PDB set, is closest to 0.5. It uses model outputs alone and selects $1-t = 0.05$, a 9° half-width for ordinary torsions; this scale is transferred to ATLAS. The grid is $\{0.002, 0.005, 0.01, 0.02, 0.035, 0.05, 0.075, 0.1, 0.15, 0.2, 0.3, 0.5\}$. Random-weight controls use the same scale. A secondary fixed scale $1 - t = 0.1$ gives the same correlation signs but weaker associations.

Table 7: FlowPacker at $1-t = 0.05$. Trained-model entries include 95% protein-bootstrap confidence intervals. Random-weight columns give point estimates on the same proteins. A dash means that no MD reference is available for that set.

	Recent PDB		ATLAS	
Reference	Trained	Random	Trained	Random
rSASA (ρ)	.418 [.391, .443]	.054	.413 [.355, .465]	.075
B-factor (ρ)	.422 [.386, .459]	.032	.384 [.325, .442]	.047
MD variance (ρ)	—	—	.330 [.287, .376]	.005
MD | rSASA (ρ)	—	—	.183 [.141, .232]	−.022
B-factor | rSASA (ρ)	—	—	.238 [.178, .291]	.012
Proline AUC	.781 [.749, .812]	.329	.810 [.756, .860]	.321

Interpretation and numerical limits. These measurements test whether the learned conditional distribution carries useful information about physical flexibility. They do not supply a true LID label for each residue: the model learns from crystal conformations, whereas MD describes thermal variation in solution. In particular, the proline ranking does not demonstrate a one-dimensional plateau. At the selected scale, about 1% of residues have negative estimated Response; at $1 - t = 0.1$ this rises to about 22% in the recent-PDB set, indicating model or readout error. The reported correlations retain these values without clipping. The uniform angular kernel remains non-Gaussian at finite scales, but its compact support and periodic coordinates require care in applying smooth Euclidean kernel assumptions. This study validates the practical trace construction and its associations, rather than every hypothesis of an asymptotic theorem for these native channels.

The saved implementation uses pure-PyTorch replacements for unavailable scatter/graph extensions. Native denoising and packing checks are retained; FlowPacker's scatter operations produce small forward-pass nondeterminism, so finite-difference checks use steps large enough to resolve it. Artifact records include checkpoint identifiers, scale grids, per-query predictions, proxy tables and analysis scripts. All table point estimates were replayed from those saved values before inclusion in the manuscript.

EXPERIMENTAL SETUPS AND REPRODUCIBILITY

T EXPERIMENTAL PROTOCOL AND METHODS

This appendix specifies the protocol required to reproduce the headline results. The main text retains only the details needed to interpret those results.

T.1 BENCHMARK TASKS AND REPRESENTATIONS

Table 8: Role of each LID-Benchmarks condition (Tempczyk et al., 2026) in our evaluation. A known dimension is shown only when it is part of the task contract.

Block	Condition	Target	Property tested
Known LID	Spiral, coefficients / rendered	$d = 1$	nearby turns and change of representation
	Uniform, coefficients / rendered	$d = 20$	proximity to a boundary
	Arrows	nominal factor count	factor complexity of the renderer
	Funnel, coefficients / rendered	$d = 2$	nonuniform local scale
	Crescent, coefficients / rendered	$d = 3$	curvature and thickness
	Gaussian4 and Sphere4, coefficients / rendered	$d = 5$	density and curvature
	Spaghetti, coefficients / rendered	$d = 1$	strong curvature and folding
Additional diagnostics	Gaussian-PCA / Sphere-PCA, coefficients / rendered	$d = 20/19$	simple calibration conditions outside the canonical headline set
E1	Fashion-MNIST with decreasing training-set size	unknown	sensitivity to training-set size
E5	paired Fashion-MNIST transformations	$\Delta d \in \{0, 4, 8\}$	invariance and equivariance

For conditions available in two representations, the same underlying objects are evaluated both in coefficient space and after high-dimensional rendering. We therefore treat the representations as paired conditions, not as independent datasets. The generated E3/E4 conditions (Gaussian-PCA and Sphere-PCA) are used only as diagnostics and are excluded from the canonical headline aggregate.

The loader applies three explicit corrections to inconsistent labels in the published archive: Funnel is set to $d = 2$, Crescent to $d = 3$, and Spaghetti to $d = 1$ on every split. The archived "Sphere4" condition is the code-level S^5 construction with target $d = 5$, rather than the S^4 description in the benchmark prose. We retain that artifact contract. Arrows is treated as a factor-count stress test with construction target $6k$ for k arrows; because rasterization, quantization, and overlap invalidate a literal smooth-support interpretation, it is reported separately from the strict known-LID aggregate.

The September 16 export contains all 429 native-task cells, including the 44 generated E3/E4 cells. Manifests record input hashes, normalization and partition identities. Saved targets agree with the corrected dimensions. The raw benchmark inputs are not bundled, so their bytes and construction cannot be independently reconstructed from the prediction archive alone.

T.2 MODELS, READOUTS, AND COMPUTATIONAL INTERFACES

The original campaign's producing revision is `c227b69`, protocol `native_tasks_points_scalar_nf_pfgm_disabled_v13`. The archive identifies 159 producing source files. Native network outputs and objectives are given below. Let X denote a normalized clean observation, $Z \sim N(0, I)$ independent Gaussian noise, and F_θ the raw network. Except for diffusion's coordinate-summed loss, squared regression losses are averaged over coordinates and examples.

The controls sample $\log \lambda$ uniformly on $[\log(1/256), \log 64]$. Their role is to compare output and input parameterizations with the native learning recipes; they are not additional published algorithms. VE uses the same Gaussian draw for its corruption and negative-noise target, correcting the inconsistent forward corruption in the pinned FLIPD code; this is a mathematical correction, not reproduction of that code defect.

For DSB (De Bortoli et al., 2021), the source is a standard Gaussian and the terminal distribution is the data. The Brownian diffusion variance is 2, with 20 transitions of length .01 and horizon .2.

Table 9: Implemented learning tasks. Output conversion is used only for LID evaluation. The four controls are our parameterizations, not additional published baseline methods. Shared optimization choices are stated below.

Method	Network input and conditioning	Target and weighting
VP diffusion	$a(t)X + b(t)Z$, $t \sim U[0,1)$; $a = e^{-(.1t+9.95t^2)/2}$, $b = \sqrt{1-a^2}$	$-Z$; coordinate-summed MSE
VE diffusion	$X + .01(5000)^t Z$, $t \sim U[0,1)$	$-Z$; coordinate-summed MSE
Rectified flow	$tX + (1-t)Z$, $t \sim U[.001, 1)$; condition $999t$	$X - Z$; unweighted MSE
Trigonometric flow	$aX + bZ$, $a = \sin(\pi t/2)$, $b = \cos(\pi t/2)$; $t \sim U[0,1)$	$(\pi/2)(bX - aZ)$; unweighted MSE
t-Flow	$tX + (1-t)\eta$, $\eta = Z/\sqrt{U/5}$, $U \sim \chi^2_5$; $t \sim U[0,1)$, condition $1-t$	η; unweighted MSE, one radial draw per vector
Logistic FM	$tX + (1-t)Z_L$, product-logistic Z_L of unit coordinate variance; $t \sim U[0,1)$, condition $999t$	$X - Z_L$ and $\psi_K(Z_L)$; sum of two unweighted MSEs, separate network cores (Appendix K)
Brownian DSB	Discrete bridge trajectories, native time; two increment networks	Alternating IPF regression; details below
Scale-NF	One invertible model conditioned on scalar $\log\sigma$; observations $X + \sigma Z$	Exact conditional negative log likelihood; log-uniform σ
Log-noise velocity control	$X + \lambda Z$, condition $\log\lambda$	λZ; λ^{-2} MSE
Log-noise posterior control	$X + \lambda Z$, condition $\log\lambda$	X; λ^{-2} MSE
Rectified posterior control	$(X + \lambda Z)/(1+\lambda)$, condition $\log\lambda$	X; λ^{-2} MSE
VP posterior control	$(X + \lambda Z)/\sqrt{1+\lambda^2}$, condition $\log\lambda$	X; λ^{-2} MSE

Twenty complete IPF rounds use 40 alternating half-passes and two full-size networks. Given the opposite frozen transition mean $m(x,t) = x + F_{\text{opposite}}(x,t)$, draw $x' = m(x,t) + \sqrt{2\Delta t}\, Z$ and regress the active increment at $(x', T - t - \Delta t)$ onto $m(x,t) - m(x',t)$. The first projection uses $m(x,t) = x$. Fresh trajectories replace a finite trajectory cache. The active network and optimizer moments are reset at each half-pass. The final pair is retained, with no EMA across IPF objectives. The resulting model is a finite-discretization Brownian DSB. General bridge factors need not define an independent convolution kernel.

The original 429 cells completed 128,000 AdamW updates, batch size 256, learning rate $2 \cdot 10^{-4}$, weight decay 10^{-6} and gradient clipping at 1. The final 8,000 steps use cosine learning-rate decay to 1% of the initial rate. Fields use the same architecture within a representation: a width-512 point MLP with four residual blocks for coordinate data, or a width-8 U-Net for images. Point inputs have no spatial Fourier expansion. NF uses RealNVP coupling conditioners with a scalar $\log\sigma$ embedding, no Fourier conditioning, and parameter count within 10% of one shared field. Its explicit input scaling by $\sqrt{1+\sigma^2}$ is retained. DSB contains two field cores and has additional trajectory cost; equal update budgets do not imply equal FLOPs.

The subsequent logistic-source FM campaign adds 15 canonical known-LID cells, with the same data, partitions, normalization, field backbones, 128,000-update budget, scale grids and selectors. It contains two separately parameterized field cores, trained jointly on velocity and kernel-score targets; Appendix K gives the objectives and Full readout. Each core receives the full update budget. E1/E5 and the generated calibration tasks were not run for this model, and their absent results are distinguished from failed selection in the tables.

Except for DSB, candidate checkpoints are evaluated every 500 steps by native held-out loss, including EMA candidates with decay .999 after step 120,000. The best candidate is retained; 148 of 429 selected checkpoints precede step 128,000, although every training reaches that step. These are shared benchmark optimization choices, not a claim to reproduce each paper's entire architecture, optimizer and dataset recipe.

For Gaussian affine channels, native outputs are converted to the canonical posterior b_λ only at evaluation. For $Y_t = \alpha_t X + \beta_t Z$, the conversion evaluates the native field at $Y_t = \alpha_t x$ and uses $\lambda = \beta_t/\alpha_t$, as derived in Section 3. The spatial derivative is with respect to the canonical query x, including this input rescaling. Response is $R = \operatorname{tr} J_x b_\lambda$; Full adds $\|b_\lambda - x\|^2/\lambda^2$. Thus the two readouts share learned fields but measure different finite-scale quantities. DSB uses its dataward drift b_+ and $\tau = \lambda^2/2$ to form $q_\lambda(x) = x + \tau b_+(x, T-\tau)$, then evaluates the terminal-factor response

and correction from the bridge interface. For t-Flow, λ is coordinate RMS: if $s = (1-t)/t$ then $\lambda = s\sqrt{5/3}$. Its noise predictor gives the canonical posterior $b_\lambda(x) = x - sF_\theta(tx, 1-t)$; Response differentiates this same posterior with respect to x. Its Full readout uses the Student density identity, not the Gaussian squared-displacement formula. With $e = x - b_\lambda$, $A = (5-2)\lambda^2$ and $h = \|e\|^2/A$, it is

$$\widehat{d}_{\text{full}} = \frac{R + (D+5)h + e^\top \partial_{\log\lambda} b_\lambda / A}{1+h}.$$

The scale derivative holds x fixed and differentiates through both the native input and its conditioning. The t-Flow Response rows in Tables 2 and 20 reuse this model and Full's operating points; they introduce neither a new fit nor Response-specific scale selection. Radial Student noise uses one shared random scale per vector, whereas the Gaussian-source flows retain Gaussian endpoint kernels. Scale-NF uses an OLS slope of $\log p_\lambda(x)$ against $\log\lambda$ over five scales, plus D. Its exact fixed-point autograd slope is retained as a dependent diagnostic. A singular CNF is not evaluated here. The Gaussian qualifier refers to the perturbation $X + \lambda Z$ used to train this density family, not to the distribution represented by RealNVP. Its density-slope readout follows the LIDL construction (Tempczyk et al., 2022).

T.3 Interpretation of the benchmark readouts

Appendix J gives the estimator identities, coordinate normalizations and comparisons with prior methods. Our benchmark uses its own checkpoints, architectures, trace budgets and scale selection; estimator equivalence does not mean reproduction of a cited experimental protocol. The Full and Response rows evaluate scalar readouts, rather than the separate spectral counts discussed in that appendix. The dependent Response rows are retained in Table 11.

T.4 Splits, checkpoints, and operating scales

A separate checkpoint is trained for every model–task–representation cell. The source training split is partitioned into optimizer-fit and held-out subsets. Native checkpoint loss and operating-scale selection use held-out source-training data; neither uses test labels. Normalization subtracts the optimizer-fit mean and divides by a global per-coordinate RMS. Consequently, equal normalized λ in separately fitted E1/E5 cells does not imply equal noise width in the original pixel units.

For known-LID cells, the selected scale minimizes primary-readout MAE over the source-train holdout; ties within 10^{-12} take the smaller scale. Response inherits Full's scale without separate tuning. For E1 and E5, Kneedle is applied to the reference model's held-out mean curve after its global maximum; a detected interior knee is frozen for all dependent cells. A failed reference selection remains a failure for the whole comparison group, with no replacement scale.

Both supervised and pointwise automatic known-LID selection use the same 29 geometric candidates within each model:

VP:	$[1/256, 152.1669703]$,
VE:	$[.01, 50]$,
RF:	$[1/256, 999]$,
DSB:	$[\sqrt{.02}, \sqrt{.4}]$,
trigonometric, Student, logistic, NF, controls:	$[1/256, 64]$.

There is no adaptive grid extension. Of 209 physical known-LID cells, 38 select a grid endpoint; 32 of these belong to the 165 canonical cells. All remain in the tables. Six failed reference selections (DSB, Student and NF on both E1 and E5) propagate to 60 physical cells. The remaining 331 physical cells have an interior selected scale.

T.5 Metrics and aggregation

For a known-LID condition c with query-level target d_{ci}, the primary error summary for method m is

$$\text{MAE}_{mc} = \frac{1}{n_c}\sum_{i=1}^{n_c} \left|\widehat{d}_{mci} - d_{ci}\right|. \tag{167}$$

For constant-target conditions, the corresponding relative error is

$$r_{mc} = \frac{\mathrm{MAE}_{mc}}{|d_c|}. \tag{168}$$

Coefficient-space and high-dimensional rendered results are reported separately. Let $\mathcal{S}$ be the five canonical strict known-LID suites (Spiral; Uniform; Funnel; Crescent; and Gaussian4, Sphere4 and Spaghetti, which form one suite) and $\mathcal{C}_{s,r}$ their conditions in representation r. The corresponding headline aggregate is

$$R_{m,r} = \mathrm{expm1}\left[\frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\frac{1}{|\mathcal{C}_{s,r}|}\sum_{c\in\mathcal{C}_{s,r}}\log(1+r_{mc})\right]. \tag{169}$$

Thus every suite receives equal weight, while its variants are averaged within the suite. Arrows is excluded from $R_{m,r}$ and reported as raw MAE. We also report signed bias and RMSE. An additional validity statistic is the minimum, over canonical known-LID cells, of the test-query fraction whose estimate is finite and lies in $[0, D]$; it is a worst-cell diagnostic rather than a weighted accuracy score.

E1 measures the stability of the mean prediction relative to a model trained on the complete training set. Let $n_0 = 50{,}000$, $x_n = \log_2(n_0/n)$, and $q_n = |\bar{d}_n - \bar{d}_{n_0}|/|\bar{d}_{n_0}|$. The reported score is

$$E1_m = \exp\left\{\frac{1}{x_{\max}-x_{\min}}\int_{x_{\min}}^{x_{\max}}\log(1+q_n)\,\mathrm{d}x\right\} - 1, \tag{170}$$

with the integral evaluated by the trapezoidal rule. E1 is not an absolute-accuracy metric: a constant estimator has score zero. It is therefore reported separately and does not enter a composite quality score.

For a paired transformation g in E5, we compute

$$\mathrm{MAE}_g^{\Delta} = \frac{1}{n}\sum_{i=1}^{n}\left|\left(\widehat{d}(g(x_i)) - \widehat{d}(x_i)\right) - \Delta_g\right|. \tag{171}$$

Transformations with $\Delta_g = 0$ and $\Delta_g > 0$ form separate invariance and equivariance groups G_0 and G_+. For either group G, the displayed score is

$$E5_{m,G} = \exp\left\{\frac{1}{|G|}\sum_{g\in G}\log(1+\mathrm{MAE}_{m,g}^{\Delta})\right\} - 1. \tag{172}$$

We do not construct a single ordinal leaderboard from Known LID, E1, and E5: the tasks test different properties and admit different trivial solutions. For reference, a constant estimate $\widehat{d} \equiv 1$ scores 0.485 on either strict aggregate, 14.029 on Arrows, and 0/0/5.708 on E1/E5 invariance/equivariance. The last value is $\sqrt{(1+4)(1+8)} - 1$.

The campaign uses training seed 0 and reports point estimates without confidence intervals. Query-sampling uncertainty and retraining uncertainty are therefore not quantified; the table supports descriptive comparisons of the fitted recipes and declared selectors.

The displayed constant baseline is computed from the known targets. Saved predictions permit direct admissible-range checks. Across the 165 canonical known-LID primary cells, every selected prediction is finite, but the worst fraction inside $[0, D]$ is only 4.7% (DSB, coefficient-space Crescent). Finite output is therefore not a calibration guarantee.

T.6 Numerical evaluation and computational cost

Field evaluation, automatic differentiation, and readout algebra use fp32, with TF32 disabled; final predictions are stored as fp64 arrays. Vector conditions use exact full-ambient traces, while image conditions use 64 Rademacher probes. A probe v has independent entries in $\{-1, 1\}$, and the trace estimate averages $v^\top J_x b_\lambda v$ over probes. The Student Full readout additionally requires a scale derivative; NF OLS5 requires five density evaluations per candidate scale, while the dependent NF autograd readout uses fp64. Equal update counts do not equalize training or inference cost. No common-hardware timing or peak-memory comparison is reported.

U COMPLETE METRIC TABLES

This appendix collects the metric tables for the analyses in Appendices N to P. Datasets, models, selection rules, and metric definitions are specified in Appendix T. Per-condition test MAE for every readout of Table 2 is listed for supervised holdout selection in Tables 12 and 13 and for pointwise Kneedle in Tables 14 and 15.

U.1 PRIMARY AND DEPENDENT READOUTS

Table 10: The 11 primary-readout recipes of the September 16 native campaign and the later logistic-source row, seed 0. Known-LID uses supervised source-train holdout selection; E1/E5 use a frozen reference-mean Kneedle scale. Lower is better. Dashes denote failed reference selection, not zero error; n/a marks the E1/E5 scores of the logistic row, whose run covered the known-LID cells only. All boundary selections are retained.

Model/readout	Coef. ↓	IDR ↓	Arrows ↓	E1 ↓	E5 $\Delta d = 0$	E5 $\Delta d > 0$
VP diffusion + FLIPD (full)	0.204	0.383	3.762	0.249	14.281	45.452
VE diffusion + FLIPD (full)	0.181	1.358	123.719	0.139	45.130	606.375
Rectified flow (full)	0.095	0.249	2.787	0.273	10.830	35.436
Trigonometric flow (full)	0.206	0.268	3.511	0.282	13.550	43.100
Brownian DSB (full)	2.833	56.150	1538.797	–	–	–
t-Flow (full)	0.168	13.169	1053.584	–	–	–
Scale-NF (OLS5)	0.163	0.521	5.888	–	–	–
Log-noise velocity control (full)	0.091	1.862	37.055	0.726	21.560	32.121
Log-noise posterior control (full)	0.139	0.491	3.365	0.109	16.566	100.024
Rectified posterior control (full)	0.134	0.367	2.240	0.132	45.868	802.933
VP posterior control (full)	0.131	0.325	2.069	0.218	44.850	817.939
Logistic FM (full)	0.118	0.253	4.238	n/a	n/a	n/a

Table 11: Dependent response readouts from the same field checkpoints and the same Full-selected scales as the primary table. NF has no response row. No response-specific scale tuning is performed. Lower is better; dashes denote failed reference selection and n/a marks scores the logistic run did not measure.

Model/readout	Coef. ↓	IDR ↓	Arrows ↓	E1 ↓	E5 $\Delta d = 0$	E5 $\Delta d > 0$
VP diffusion (response)	0.227	0.412	6.425	0.500	5.740	21.528
VE diffusion (response)	0.186	1.413	72.222	0.100	43.240	548.976
Rectified flow (response)	0.129	0.239	8.411	0.378	6.031	18.579
Trigonometric flow (response)	0.239	0.253	13.283	0.555	7.194	17.017
Brownian DSB (response)	3.113	19.149	438.878	–	–	–
t-Flow (response)	0.164	1.975	244.345	–	–	–
Log-noise velocity control (response)	0.090	1.540	33.835	3.289	14.031	17.168
Log-noise posterior control (response)	0.129	0.602	14.496	0.167	12.537	68.125
Rectified posterior control (response)	0.119	0.526	12.494	0.183	37.215	785.419
VP posterior control (response)	0.125	0.547	10.603	0.385	34.653	792.000
Logistic FM (response)	0.109	0.321	6.311	n/a	n/a	n/a

U.2 KNOWN-LID CONDITIONS AND CALIBRATION CONTROLS

Table 12: Canonical known-LID coefficients test MAE (lower is better), at each checkpoint's supervised holdout-selected scale. Each geometric condition has 1,000 test queries; Arrows, where shown, has 10,000 and nominal factor-count targets. All boundary selections are retained. The logistic-source rows (Full and Response) were measured later with the same benchmark code, data, partitions, backbone, budget, grid and selector. Response rows reuse the scales selected for Full; rows shown in Table 2 aggregate to its supervised columns, and the others to Tables 10 and 11.

Model/readout	Spiral	Uniform	Funnel	Crescent	Gaussian4	Sphere4	Spaghetti
Primary readouts							
VP diffusion + FLIPD (full)	0.440	0.147	0.641	0.541	1.179	0.412	0.045
VE diffusion + FLIPD (full)	0.864	0.069	0.282	0.092	0.103	0.121	0.095
Rectified flow (full)	0.203	0.074	0.332	0.221	0.303	0.120	0.043
Trigonometric flow (full)	0.531	0.180	0.553	0.193	1.812	1.254	0.050
Brownian DSB (full)	14.124	0.951	7.745	7.600	9.105	7.321	3.020
t-Flow (full)	0.570	0.139	0.402	0.182	0.259	0.260	0.144
Scale-NF (OLS5)	0.637	0.014	0.294	0.085	0.160	0.521	0.173
Log-noise velocity control (full)	0.420	0.016	0.084	0.014	0.046	0.057	0.103
Log-noise posterior control (full)	0.595	0.029	0.259	0.001	0.029	0.222	0.137
Rectified posterior control (full)	0.560	0.032	0.283	0.002	0.029	0.317	0.079
VP posterior control (full)	0.546	0.029	0.262	0.001	0.029	0.295	0.103
Logistic FM (full)	0.332	0.079	0.347	0.112	0.115	0.131	0.169
Dependent Response readouts at the Full-selected scales							
VP diffusion (response)	0.433	0.280	0.668	0.686	1.677	0.601	0.065
VE diffusion (response)	0.836	0.090	0.327	0.104	0.131	0.236	0.094
Rectified flow (response)	0.168	0.079	0.373	0.286	0.521	0.223	0.505
Trigonometric flow (response)	0.496	0.270	0.474	0.338	2.345	1.574	0.417
Brownian DSB (response)	8.498	1.137	10.920	11.625	12.889	11.221	3.454
t-Flow (response)	0.527	0.274	0.388	0.141	0.196	0.887	0.098
Log-noise velocity control (response)	0.405	0.019	0.095	0.013	0.040	0.075	0.099
Log-noise posterior control (response)	0.481	0.102	0.272	0.002	0.038	0.454	0.158
Rectified posterior control (response)	0.427	0.055	0.297	0.001	0.040	0.544	0.094
VP posterior control (response)	0.319	0.053	0.275	0.001	0.037	0.533	0.549
Logistic FM (response)	0.245	0.095	0.397	0.127	0.131	0.175	0.172

Table 13: Canonical known-LID rendered test MAE (lower is better), at each checkpoint's supervised holdout-selected scale. Each geometric condition has 1,000 test queries; Arrows, where shown, has 10,000 and nominal factor-count targets. All boundary selections are retained. The logistic-source rows (Full and Response) were measured later with the same benchmark code, data, partitions, backbone, budget, grid and selector. Response rows reuse the scales selected for Full; rows shown in Table 2 aggregate to its supervised columns, and the others to Tables 10 and 11.

Model/readout	Spiral	Uniform	Funnel	Crescent	Gaussian4	Sphere4	Spaghetti	Arrows
Primary readouts								
VP diffusion + FLIPD (full)	1.298	0.786	0.125	0.415	0.068	0.310	3.983	3.762
VE diffusion + FLIPD (full)	2.153	18.948	3.123	2.220	6.803	6.653	2.430	123.719
Rectified flow (full)	0.356	0.993	0.440	0.289	3.354	3.396	0.461	2.787
Trigonometric flow (full)	0.565	1.941	0.519	0.253	2.079	2.316	0.317	3.511
Brownian DSB (full)	87.597	392.580	66.465	260.354	264.068	323.917	385.325	1538.797
t-Flow (full)	23.395	81.833	28.677	41.214	34.118	42.969	111.117	1053.584
Scale-NF (OLS5)	1.856	3.184	0.770	0.749	1.484	1.141	0.794	5.888
Log-noise velocity control (full)	1.868	7.392	2.518	9.130	13.537	9.970	12.816	37.055
Log-noise posterior control (full)	1.084	1.787	1.031	1.157	0.814	1.891	1.316	3.365
Rectified posterior control (full)	1.186	1.514	0.656	0.559	1.109	1.535	0.341	2.240
VP posterior control (full)	1.071	1.883	0.446	0.587	1.229	1.310	0.197	2.069
Logistic FM (full)	0.456	1.702	0.633	0.236	2.454	2.920	0.108	4.238
Dependent Response readouts at the Full-selected scales								
VP diffusion (response)	1.002	4.951	0.200	0.514	0.201	0.255	3.830	6.425
VE diffusion (response)	2.133	26.235	1.361	3.175	6.161	6.045	6.051	72.222
Rectified flow (response)	0.256	3.187	0.174	0.218	2.821	3.810	0.847	8.411
Trigonometric flow (response)	0.257	5.987	0.360	0.343	1.514	1.771	0.689	13.283
Brownian DSB (response)	46.872	179.001	23.766	38.640	58.809	105.450	209.221	438.878
t-Flow (response)	3.073	25.627	1.553	12.530	10.951	3.438	2.764	244.345
Log-noise velocity control (response)	1.842	7.181	2.376	4.815	8.644	9.931	12.559	33.835
Log-noise posterior control (response)	1.071	8.546	1.439	1.139	0.772	1.834	1.163	14.496
Rectified posterior control (response)	1.155	8.981	0.712	1.120	1.337	1.605	0.719	12.494
VP posterior control (response)	1.038	10.956	0.626	1.247	1.470	1.819	0.954	10.603
Logistic FM (response)	0.212	6.791	0.771	0.257	2.629	3.716	0.676	6.311

Table 14: Canonical known-LID coefficients test MAE per condition under pointwise Kneedle selection, on the same checkpoints and 29-scale model-specific grids as the supervised tables (lower is better). Each geometric condition has 1,000 test queries. Response rows reuse the pointwise indices and fallbacks selected for Full. Rows shown in Table 2 aggregate to its Pointwise Kneedle columns, and the primary rows to Table 20. The logistic-source rows come from the later run.

Model/readout	Spiral	Uniform	Funnel	Crescent	Gaussian4	Sphere4	Spaghetti
Primary readouts							
VP diffusion + FLIPD (full)	35.082	1.118	30.724	39.081	18.003	13.877	1.478
VE diffusion + FLIPD (full)	1.012	0.069	0.381	0.096	0.103	0.133	0.616
Rectified flow (full)	0.592	0.544	0.399	0.355	1.062	0.258	0.341
Trigonometric flow (full)	135.593	3.462	78.851	57.429	40.119	24.851	2.137
Brownian DSB (full)	120.186	4.856	46.735	107.112	96.127	97.287	7.220
t-Flow (full)	0.236	0.218	0.819	0.249	0.584	0.828	1.526
Scale-NF (OLS5)	0.808	0.898	0.389	0.210	0.255	0.741	0.658
Log-noise velocity control (full)	1.008	0.017	0.108	0.014	0.099	0.346	0.241
Log-noise posterior control (full)	1.024	1.266	0.378	0.010	0.047	1.033	0.237
Rectified posterior control (full)	1.023	1.412	0.417	0.012	0.039	1.073	0.355
VP posterior control (full)	1.027	1.288	0.361	0.010	0.048	0.985	0.371
Logistic FM (full)	1.152	0.624	0.639	0.551	0.119	1.117	0.439
Dependent Response readouts							
VP diffusion (response)	35.367	1.172	31.007	39.360	18.566	14.001	1.562
VE diffusion (response)	1.006	0.090	0.439	0.102	0.131	0.148	0.635
Rectified flow (response)	0.719	0.481	0.419	0.340	1.317	0.237	0.273
Trigonometric flow (response)	135.843	3.512	79.234	58.268	40.287	25.083	2.258
Brownian DSB (response)	116.494	5.454	58.204	119.291	115.426	115.732	3.403
t-Flow (response)	0.240	0.482	0.395	0.230	0.734	0.545	1.006
Log-noise velocity control (response)	1.007	0.017	0.099	0.013	0.091	0.248	0.185
Log-noise posterior control (response)	1.000	0.003	0.376	0.001	0.016	0.917	0.199
Rectified posterior control (response)	1.000	0.004	0.421	0.001	0.010	0.950	0.322
VP posterior control (response)	1.000	0.003	0.364	0.001	0.016	0.878	0.315
Logistic FM (response)	1.125	0.602	0.650	0.552	0.138	1.129	0.423

Table 15: Canonical known-LID rendered test MAE per condition under pointwise Kneedle selection, on the same checkpoints and 29-scale model-specific grids as the supervised tables (lower is better). Each geometric condition has 1,000 test queries; Arrows has 10,000 and nominal factor-count targets. Response rows reuse the pointwise indices and fallbacks selected for Full. Rows shown in Table 2 aggregate to its Pointwise Kneedle columns, and the primary rows to Table 20. The logistic-source rows come from the later run.

Model/readout	Spiral	Uniform	Funnel	Crescent	Gaussian4	Sphere4	Spaghetti	Arrows
Primary readouts								
VP diffusion + FLIPD (full)	16.320	46.859	32.989	19.971	20.592	37.477	79.131	258.662
VE diffusion + FLIPD (full)	9.880	106.474	25.312	38.475	56.655	65.543	124.073	620.509
Rectified flow (full)	31.855	32.352	42.059	36.686	42.514	34.326	48.722	360.918
Trigonometric flow (full)	26.290	20.627	24.981	31.842	62.071	40.895	61.685	214.287
Brownian DSB (full)	92.544	521.319	83.548	393.229	266.059	336.633	510.940	2815.689
t-Flow (full)	203.125	764.615	382.039	738.601	700.218	758.243	783.225	3051.534
Scale-NF (OLS5)	784.102	769.845	782.967	781.262	780.747	780.574	788.181	3058.117
Log-noise velocity control (full)	17.778	97.405	18.741	23.587	33.816	43.567	124.242	139.344
Log-noise posterior control (full)	40.902	219.709	33.163	55.550	89.162	127.223	228.483	1693.394
Rectified posterior control (full)	51.868	203.299	39.964	65.720	89.494	124.650	245.330	1607.480
VP posterior control (full)	44.368	237.073	36.036	63.644	89.161	120.401	208.359	1568.487
Logistic FM (full)	29.783	45.252	45.634	40.949	49.006	46.072	66.299	355.665
Dependent Response readouts								
VP diffusion (response)	9.421	33.570	19.982	11.632	11.320	23.947	62.284	201.689
VE diffusion (response)	6.029	82.064	18.216	30.851	41.773	49.720	100.925	554.439
Rectified flow (response)	26.664	21.393	31.779	27.225	29.517	24.244	34.754	276.386
Trigonometric flow (response)	24.756	8.693	19.496	18.760	45.643	26.714	45.524	130.577
Brownian DSB (response)	56.897	373.109	57.796	240.126	86.892	154.887	399.215	2609.317
t-Flow (response)	95.078	709.926	209.269	555.776	517.810	572.816	724.624	2909.186
Log-noise velocity control (response)	16.764	75.436	16.376	17.128	25.401	33.096	101.523	98.459
Log-noise posterior control (response)	27.983	171.784	27.391	48.768	69.249	106.118	184.579	1217.773
Rectified posterior control (response)	34.526	161.118	33.268	57.422	69.417	102.795	201.798	1145.909
VP posterior control (response)	31.265	191.029	30.221	56.167	70.109	99.965	164.414	1145.101
Logistic FM (response)	28.775	38.748	41.454	39.678	43.976	41.658	60.915	339.851

Table 16: Generated Gaussian-PCA ($d = 20$) and Sphere-PCA ($d = 19$) calibration conditions, excluded from the canonical aggregate. Test MAE at the supervised holdout-selected scale, 1,000 queries per cell; lower is better.

Model/readout	Gaussian coef.	Gaussian IDR	Sphere coef.	Sphere IDR
VP diffusion + FLIPD (full)	0.052	2.325	0.052	1.211
VE diffusion + FLIPD (full)	0.024	5.986	0.070	20.163
Rectified flow (full)	0.033	1.598	0.048	1.018
Trigonometric flow (full)	0.120	2.058	0.059	0.528
Brownian DSB (full)	0.762	341.671	0.756	361.394
t-Flow (full)	0.123	84.896	0.054	76.566
Scale-NF (OLS5)	0.025	3.428	0.072	4.271
Log-noise velocity control (full)	0.013	1.642	0.070	12.980
Log-noise posterior control (full)	0.038	1.431	0.038	2.702
Rectified posterior control (full)	0.033	4.244	0.037	0.855
VP posterior control (full)	0.036	3.290	0.028	0.723

U.3 TRAINING-SET SIZE, INVARIANCE, AND EQUIVARIANCE

Table 17: E1 at the frozen reference-selected scale: mean LID for 50,000 and 12 training examples, each evaluated on the same 5,000 test images, and normalized drift AUC (lower is more stable, not necessarily more accurate). Dashes mark failed reference selection.

Model/readout	λ	Mean, $n = 50{,}000$	Mean, $n = 12$	Drift AUC
VP diffusion + FLIPD (full)	0.771	114.553	269.695	0.249
VE diffusion + FLIPD (full)	0.018	466.311	806.235	0.139
Rectified flow (full)	1.975	46.857	104.031	0.273
Trigonometric flow (full)	1.414	62.749	196.153	0.282
Brownian DSB (full)	–	–	–	–
t-Flow (full)	–	–	–	–
Scale-NF (OLS5)	–	–	–	–
Log-noise velocity control (full)	4.000	24.653	31.142	0.726
Log-noise posterior control (full)	1.414	136.278	193.553	0.109
Rectified posterior control (full)	2.000	99.917	176.065	0.132
VP posterior control (full)	2.000	80.694	259.185	0.218

Table 18: E5 paired LID-change MAE on 5,000 test pairs per transformation; lower is better. Upscale, zero padding and powers .25/4 have expected change zero; Pad +4/+8 have expected change 4/8. Each method reuses its downscaled-reference scale. Dashes denote failed reference selection.

Model/readout	Upscale	Pad +0	Power .25	Power 4	Pad +4	Pad +8
VP diffusion + FLIPD (full)	12.674	55.409	6.131	8.913	52.991	38.966
VE diffusion + FLIPD (full)	17.093	597.153	10.488	35.422	603.302	609.463
Rectified flow (full)	12.116	44.272	3.277	6.711	41.601	30.163
Trigonometric flow (full)	8.823	61.055	5.898	9.658	42.651	43.554
Brownian DSB (full)	–	–	–	–	–	–
t-Flow (full)	–	–	–	–	–	–
Scale-NF (OLS5)	–	–	–	–	–	–
Log-noise velocity control (full)	28.521	57.839	8.537	14.637	54.375	18.810
Log-noise posterior control (full)	8.832	105.429	5.565	12.859	117.153	85.378
Rectified posterior control (full)	30.759	796.616	11.359	14.411	804.326	801.542
VP posterior control (full)	30.228	813.862	10.527	14.066	833.781	802.397

U.4 NORMALIZING-FLOW READOUTS

Table 19: Dependent NF density-slope readouts on the same checkpoints and OLS5-selected scales. Both return $D + \partial_{\log \lambda} \log p_\lambda(x)$, using a five-scale fit or an autograd derivative. Lower is better; reference selection fails for both Fashion-MNIST groups.

Model/readout	Coef. ↓	IDR ↓	Arrows ↓	E1 ↓	E5 $\Delta d = 0$	E5 $\Delta d > 0$
NF OLS5	0.163	0.521	5.888	–	–	–
NF autograd	0.165	0.506	5.946	–	–	–

U.5 AUTOMATIC SCALE SELECTION

Table 20: Supervised holdout selection (sup.) versus pointwise Kneedle (auto), on identical 29-scale model-specific grids and checkpoints. Strict scores are relative-MAE aggregates; Arrows is raw MAE. Lower is better. Each dependent Response row inherits the selected indices of its Full row. Fallback and boundary percentages are query-weighted over 24,000 evaluations per recipe, including Arrows; each Response row shares the selections of its Full row. Boundary means a selected grid endpoint; no queries are discarded.

Model/readout	Coef. sup.	Coef. auto	IDR sup.	IDR auto	Arrows auto	Fallback %	Boundary %
VP diffusion + FLIPD (full)	0.204	6.894	0.383	9.329	258.662	0.000	0.000
VP diffusion (response)	0.227	6.968	0.412	5.921	201.689	0.000	0.000
VE diffusion + FLIPD (full)	0.181	0.242	1.358	11.939	620.509	0.000	25.492
Rectified flow (full)	0.095	0.213	0.249	12.116	360.918	0.000	11.175
Rectified flow (response)	0.129	0.231	0.239	9.195	276.386	0.000	11.175
Trigonometric flow (full)	0.206	13.847	0.268	10.141	214.287	0.000	2.983
Brownian DSB (full)	2.833	17.252	56.150	69.732	2815.689	0.000	61.896
t-Flow (full)	0.168	0.232	13.169	156.694	3051.534	0.000	76.904
t-Flow (response)	0.164	0.175	1.975	106.305	2909.186	0.000	76.904
Scale-NF (OLS5)	0.163	0.249	0.521	242.501	3058.117	0.000	75.017
Log-noise velocity control (full)	0.091	0.187	1.862	10.645	139.344	0.000	25.017
Log-noise posterior control (full)	0.139	0.241	0.491	23.236	1693.394	0.000	18.304
Rectified posterior control (full)	0.134	0.255	0.367	25.875	1607.480	0.000	17.850
VP posterior control (full)	0.131	0.248	0.325	24.783	1568.487	0.000	17.400
Logistic FM (full)	0.118	0.334	0.253	13.688	355.665	0.000	8.283
Logistic FM (response)	0.109	0.331	0.321	12.695	339.851	0.000	8.283

U.6 STUDENT COMPONENT SUBSTITUTIONS

Table 21: Student image-checkpoint component substitutions on 32 fixed holdout queries per task: MAE to nominal LID, lower is better. The two rows at $\lambda = 64$ use the original Full-selected boundary scale; † marks a post-hoc Response-selected diagnostic scale. Only the named quantity is replaced; the final column replaces all components.

Task	λ	Response	Full	Full, $b \leftarrow b^*$	Full, $\dot{b} \leftarrow \dot{b}^*$	Full, all reference
Spiral	64	3.091	21.355	2.514	21.376	0.936
Spiral	$22.63^\dagger$	0.736	40.524	0.905	40.557	0.905
Funnel	64	1.553	28.729	1.438	28.756	0.587
Funnel	$45.25^\dagger$	0.083	36.147	0.121	36.178	0.543