Extracting Local Manifold Geometry from Pretrained Diffusion Models in One Inverse Step

Gordei Verbii
Independent Researcher
Moscow, Russia
scigverbii@gmail.com

Abstract

We propose *DiffusionGeometryProbe*, a self-contained framework that, in a single inverse step of a pretrained denoising diffusion probabilistic model (DDPM), extracts a calibrated profile of the local geometry of the data manifold the model has internalised. Our contributions are: (i) a fixed-point reformulation of DDIM inversion whose contraction rate $\rho_g(t)$ admits closed-form bounds in $\sigma_{\max}(J_\varepsilon)$, enabling Banach- and Smale-α-theory-based certification of any inversion scheme (Picard, Anderson, Newton-Raphson); (ii) a unifying probe class that returns, in a single pass, four mutually-corroborative local intrinsic-dimension (LID) estimators (FLIPD, Stanczuk, Yeats, local-PCA), the score-Jacobian condition number, a Cheeger-style spectral-gap proxy, a Witten-deformation Morse-basin count, and a Łojasiewicz exponent of the score energy; (iii) a complete derivation of each measurement from a classical or recent result in differential / spectral / Morse geometry, with all spectral computations using only Jacobian–vector products and Hutchinson trace estimators, so the probe scales linearly in D. We validate the framework on a synthetic point-cloud DDPM trained on five 3-manifolds of known topology, on the pretrained CIFAR-10 DDPM (D=3072) and on the pretrained CelebA-HQ-256 DDPM (D=196,608).

Keywords

pretrained diffusion models, manifold hypothesis, intrinsic dimension, score Jacobian, Smale alpha-theory, Witten deformation, persistent homology

1 Introduction

The manifold hypothesis posits that natural data $x_0 \in \mathbb{R}^D$ concentrate on a low-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^D$ with $d := \dim \mathcal{M} \ll D$. Denoising diffusion probabilistic models (DDPMs) [8, 18, 19] are widely believed to encode this geometry through their score function $s_t(x) = \nabla_x \log p_t(x)$ [10, 20, 22]. Yet evaluation of a trained DDPM is almost universally performed through *output statistics* (FID, IS, CLIP) rather than through *structural* inspection of the score landscape. This paper develops a framework that opens this second channel.

Contributions.

(1) A **fixed-point reformulation of DDIM inversion** (§3.2) whose contraction rate $\rho_g(t)$ admits closed-form bounds in $\sigma_{\max}(J_\varepsilon)$, enabling Banach- and Smale-α-theory certification of Picard / Anderson / Newton inversion [7, 14, 16].
(2) A **transferable probe** `DiffusionGeometryProbe` that, given any DDIM-compatible model and a single point x_0, returns in one inverse step a structured report combining (a) inversion regime, (b) four LID estimators with mutual cross-checks, (c) a spectral-phase profile, (d) a Cheeger-style isoperimetric proxy, (e) Morse-theoretic basin counts, (f) a Łojasiewicz exponent of the score energy, and (g) a score-singularity plateau.
(3) **Empirical validation** on (i) a synthetic point-cloud DDPM [13] with ground-truth (d, β_k) on five 3-manifolds, (ii) the pretrained CIFAR-10 DDPM [8], and (iii) the pretrained CelebA-HQ-256 DDPM. The probe recovers the Banach-contraction prediction $n_{\text{iter}} \sim 1/(1-\rho_g)$ with R^2>0.99, the three-phase κ_t signature predicted by random-matrix theory [22], and a per-pixel intrinsic-dimension profile that decays monotonically with t on CelebA-HQ-256, in agreement with the predictions of [12, 25].

The probe is *not* a recovery of global topology from one data point; we discuss this scope explicitly in §5. What it gives is the first end-to-end demonstration that the spectral phase transition of [22] is observable in pretrained image DDPMs at D=10^5 scale, alongside two independent LID estimates that bracket the order of magnitude reported by [15] for natural images.

2 Related Work

Inversion of DDPM/DDIM.. DDIM [18] introduced a deterministic reverse process that admits an implicit fixed-point inversion. Most subsequent inversion methods solve this fixed point approximately via Picard [7] or Anderson [1, 14] iterations. The Regularized / Guided Newton-Raphson Inversion (RNRI/GNRI) of Samuel et al. [16] recasts inversion as Newton's method on a scalar surrogate of the residual, achieving sub-second inversion of few-step models. Our framework treats these methods uniformly through their convergence rate ρ_g and uses Smale α-theory [17, 23] to certify when the Newton regime is preferable to Picard.

Manifold dimension from DDPMs. Stanczuk et al. [20] prove that at low noise the score points orthogonally toward $\mathcal{M}$ and the rank of the score covariance equals $D - d$. FLIPD [10] derives a Fokker–Planck-based per-point LID estimator that scales to Stable-Diffusion. Yeats et al. [25] prove that the denoising-score-matching loss *lower-bounds* LID and is therefore the cheapest robust dimension estimator. Lu, Wang and Bal [12] characterise the rate at which the score blows up on the manifold; Ventura et al. [22] use random-matrix theory to predict three spectral phases of the score Jacobian over the noise schedule. Earlier non-parametric approaches [15, 21] provide ground-truth-style estimates for natural images. Our probe

GALOP Workshop at KDD 2026: Geometric Space, Architecture and Learning Objective for Large Pre-Trained Models, August 9, 2026, Jeju, Republic of Korea.
2026.

assembles these four estimators into a coherent profile and uses their disagreement as a model-regime diagnostic.

Topological data analysis. Persistent homology [3, 4] extracts Betti numbers from point clouds; Fasy et al. [6] provide bootstrap confidence bands separating signal from noise. We apply Fasy bands to a small batch of generated samples plus their Tweedie projections, with full epistemic caveats.

Optimization on manifolds and Morse theory. Smale's α/γ theory [17, 23] certifies Newton's quadratic convergence at an analytic zero. The Łojasiewicz inequality [11], together with the Polyak–Łojasiewicz condition, gives the convergence-rate exponent of any analytic gradient flow. Cheeger's inequality [2] bounds the spectral gap of the Laplacian by the squared isoperimetric constant. Witten [24] proved the Morse inequalities by analytic deformation of the de Rham complex; the deformation parameter plays the role of inverse noise scale in our framework. We instantiate these classical results on the score energy $E_t = \frac{1}{2}\|\varepsilon_\theta\|^2$, treated as a Witten-deformed Morse potential.

3 Proposed Method

3.1 Notation and DDIM preliminaries

We fix once and for all the following notation, used uniformly throughout the paper. Let $\beta_t \in (0,1)$ for $t = 1, \dots, T$ be the variance schedule, $\alpha_t := 1 - \beta_t$, $\bar{\alpha}_t := \prod_{s \le t} \alpha_s$, and $\sigma_t := \sqrt{1 - \bar{\alpha}_t}$. Let $\varepsilon_\theta : \mathbb{R}^D \times \{1, \dots, T\} \to \mathbb{R}^D$ denote the trained noise predictor and $s_t(x) := -\varepsilon_\theta(x,t)/\sigma_t$ its associated score (the standard DDPM identification). The Jacobian of ε_θ in its first argument is

$$J_\varepsilon(x,t) := \nabla_x \varepsilon_\theta(x,t) \in \mathbb{R}^{D \times D}.$$

The deterministic (η=0) DDIM forward step from x_t to x_{t-1} reads

$$x_{t-1} = \sqrt{\tfrac{\bar{\alpha}_{t-1}}{\bar{\alpha}_t}}\, x_t + \left(\sigma_{t-1} - \sqrt{\tfrac{\bar{\alpha}_{t-1}}{\bar{\alpha}_t}}\, \sigma_t\right) \varepsilon_\theta(x_t, t). \tag{1}$$

3.2 DDIM inversion as an implicit fixed point

Solving (1) for x_t given x_{t-1} yields the implicit equation

$$x_t = \underbrace{\sqrt{\tfrac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}}}\, x_{t-1} + B_t\, \varepsilon_\theta(x_t, t)}_{=: g_t(x_t; x_{t-1})}, \tag{2}$$

with $B_t := \sigma_t - \sqrt{\bar{\alpha}_t/\bar{\alpha}_{t-1}}\, \sigma_{t-1}$. Equation (2) is a fixed-point equation in x_t.

Proposition 1 (Local contraction rate). *Let $x_t^\star$ denote any solution of (2). The Jacobian of $g_t(\cdot\,; x_{t-1})$ at $x_t^\star$ is*

$$\partial_x g_t(x_t^\star; x_{t-1}) = B_t\, J_\varepsilon(x_t^\star, t),$$

so the local contraction rate of the Picard map at $x_t^\star$ is

$$\rho_g(t) := |B_t|\, \sigma_{\max}\big(J_\varepsilon(x_t^\star, t)\big). \tag{3}$$

Proof. By (2), $g_t(x_t; x_{t-1})$ depends on x_t only through $\varepsilon_\theta(x_t, t)$, with prefactor B_t, hence $\partial_x g_t = B_t J_\varepsilon$. The contraction rate of a smooth fixed-point map at a fixed point equals the spectral radius of its Jacobian, which is bounded above by its operator-2-norm $|B_t| \sigma_{\max}(J_\varepsilon)$. □

Theorem 2 (Banach iteration count). *If $\rho_g(t) < 1$ on a neighbourhood of $x_t^\star$, the Picard iteration $x^{(k+1)} = g_t(x^{(k)}; x_{t-1})$ converges geometrically to $x_t^\star$, and the number of iterations to reach residual norm ε from initial residual $\|r_0\|$ satisfies*

$$n_{\text{iter}}(\varepsilon) = \left\lceil\, \log(\varepsilon/\|r_0\|) \,/\, \log \rho_g \,\right\rceil \sim \frac{\log(\|r_0\|/\varepsilon)}{1 - \rho_g} \quad (\rho_g \to 1^-). \tag{4}$$

Proof. By Banach's contraction theorem, $\|x^{(k)} - x_t^\star\| \le \rho_g^k \|x^{(0)} - x_t^\star\|$. The residual $r_k := x^{(k)} - g_t(x^{(k)}; x_{t-1})$ satisfies $\|r_k\| \le \rho_g^k \|r_0\|$ by the same argument applied to the auxiliary map. Solving $\rho_g^k \|r_0\| = \varepsilon$ for k and Taylor-expanding $\log \rho_g = -(1 - \rho_g) - \frac{(1-\rho_g)^2}{2} - \cdots$ yields the stated asymptotic. □

Equation (4) is the central regime selector of our framework: it transforms a single Jacobian–vector spectral query into a closed-form prediction of inversion budget.

3.3 Smale α-theory certification

For analytic $F : \mathbb{R}^D \to \mathbb{R}^D$ with simple zero ζ, the Smale invariants at x are

$$\beta(F,x) = \|DF(x)^{-1} F(x)\|, \quad \gamma(F,x) = \sup_{k \ge 2} \left\| \tfrac{DF(x)^{-1} D^k F(x)}{k!} \right\|^{1/(k-1)},$$

and $\alpha(F,x) := \beta(F,x)\, \gamma(F,x)$.

Theorem 3 (Smale's α-test, sharpened by Wang–Han). *There exists a universal constant $\alpha_0 = 3 - 2\sqrt{2} \approx 0.157\,68$ such that, if $\alpha(F, x_0) < \alpha_0$, then x_0 is an* approximate zero *of F in the sense that the Newton iterates $x_{k+1} = x_k - DF(x_k)^{-1} F(x_k)$ are well defined and converge quadratically to ζ, with $\|x_k - \zeta\| \le 2^{-2^k+1} \|x_0 - \zeta\|$ [17, 23].*

We apply Theorem 3 to the residual $F(x) := x - g_t(x; x_{t-1})$, whose Jacobian is $DF = I - B_t J_\varepsilon$. Near $x_t^\star$ we have $\sigma_{\min}(DF) \ge 1 - \rho_g$, hence $\|DF^{-1}\| \le 1/(1-\rho_g)$, so $\beta(F,x) \le \|F(x)\|/(1-\rho_g)$. Under the standard scale assumption $\|D^2 F\|_{\text{op}} \le 1$ for normalised data [16], $\gamma(F,x) \le \|DF^{-1}\|/2$. Combining gives

$$\hat{\alpha}(x^{(0)}) \;:=\; \frac{\|r_0(x^{(0)})\|}{(1 - \rho_g)^2} \;\ge\; \tfrac{1}{2}\, \alpha(F, x^{(0)}). \tag{5}$$

GNRI/Newton iteration is therefore *certified* at $x^{(0)}$ whenever $\hat{\alpha} < \alpha_0$. The proxy in (5) is computable from *exactly two* quantities — the Picard residual norm and ρ_g — both of which are already produced in the inversion regime check.

3.4 Local intrinsic dimension: four estimators

Fix $x_0 \in \mathbb{R}^D$ and let $\tilde{x}_k = \sqrt{\bar{\alpha}_t}\, x_0 + \sigma_t \varepsilon_k$ with $\varepsilon_k \overset{\text{iid}}{\sim} \mathcal{N}(0, I_D)$ for $k = 1, \dots, K$. We assemble four LID estimators that probe the model from independent vantage points.

Stanczuk normal-bundle estimator. Let $C_t(x_0) := \frac{1}{K} \sum_k (s_t(\tilde{x}_k) - \bar{s})(s_t(\tilde{x}_k) - \bar{s})^\top$ be the empirical score covariance. Theorem 3.1 of [20] gives $\text{rank} C_t(x_0) \to D - d$ as $\sigma_t \to 0$. We set $\hat{d}_{\text{Stan}} := D - \text{rank}_\tau(C_t)$ where the threshold τ is calibrated against the median tail singular value (a noise-floor proxy).

FLIPD [10]. The Fokker–Planck identity in VP-SDE form gives

$$\hat{d}_{\mathrm{FLIPD}}(x_0, t) := D + \sigma_t \nabla \cdot s_t(x_t) + \sigma_t^2 \|s_t(x_t)\|^2 \xrightarrow{t \to 0} d. \tag{6}$$

We compute $\nabla \cdot s_t$ via Hutchinson's stochastic trace estimator $\mathrm{tr}(A) = \mathbb{E}_v[v^\top A v]$ with $v \sim \{\pm 1\}^D$ [9], requiring only K Jacobian–vector products.

Yeats DSM-loss LID.. Let $\varepsilon^\star$ be the Bayes-optimal denoiser that minimises $\mathcal{L}_{\mathrm{DSM}} = \mathbb{E}_\varepsilon \|\varepsilon_\theta(x_t, t) - \varepsilon\|^2$. Yeats et al. [25] prove

$$\hat{d}_{\mathrm{Yeats}}(x_0, t) := \mathbb{E}_\varepsilon \|\varepsilon_\theta(x_t, t) - \varepsilon\|^2 \ \geq \ d, \tag{7}$$

with equality when $\varepsilon_\theta = \varepsilon^\star$. Here, the inequality follows from $\mathbb{E}_\varepsilon \|\varepsilon_\theta - \varepsilon\|^2 \geq \mathbb{E}_\varepsilon \|\varepsilon^\star - \varepsilon\|^2 = d$, the latter equality being Theorem 3.1 of [25]. Estimator (7) is the cheapest of all four: a single forward pass per noise sample.

Local-PCA baseline. We form $X_t := [\tilde{x}_1, \ldots, \tilde{x}_K]^\top \in \mathbb{R}^{K \times D}$, centre, take SVD with singular values $s_1 \geq \cdots \geq s_K$, and report $\hat{d}_{\mathrm{LocPCA}} := |\{i : s_i > \eta s_1\}|$ for ratio $\eta = 0.10$. This is a classical reality check independent of the trained score.

3.5 Score-Jacobian condition number and Cheeger proxy

Let $\sigma_{\max}(J_\varepsilon)$, $\sigma_{\min}(J_\varepsilon)$ be the extremal singular values of J_ε, both estimated via power iteration on $J_\varepsilon^\top J_\varepsilon$ (top) and shifted power iteration on $cI - J_\varepsilon^\top J_\varepsilon$ for $c \geq \sigma_{\max}^2$ (bottom). Define the condition number

$$\kappa_t := \sigma_{\max}(J_\varepsilon) / \sigma_{\min}(J_\varepsilon). \tag{8}$$

Ventura et al. [22] show that under the manifold hypothesis, κ_t exhibits three phases as t decreases from T to 0: (i) *trivial* ($\kappa_t \sim 1$), (ii) *manifold coverage* (κ_t rises), (iii) *consolidation* (κ_t peaks; tangent eigenvalues collapse).

Proposition 4 (Cheeger-style spectral-gap proxy). *Treat $J_\varepsilon^\top J_\varepsilon$ as a fluctuation operator on the local linearisation of $\mathcal{M}$. Cheeger's inequality $\lambda_1 \geq h^2/4$ [2], combined with the elementary bound $\lambda_1(J_\varepsilon^\top J_\varepsilon)/\lambda_{\max}(J_\varepsilon^\top J_\varepsilon) = 1/\kappa_t^2 \leq \lambda_1/\lambda_{\max}$, gives the empirical lower bound*

$$\hat{h}(t) := \frac{2}{\sqrt{\kappa_t}}, \tag{9}$$

on the local isoperimetric constant of the fluctuation geometry.

3.6 Score singularity rate

Theorem 5 (Singular plateau [12]). *On a closed embedded d-manifold $\mathcal{M} \subset \mathbb{R}^D$, the score satisfies, as $t \to 0$,*

$$\|s_t(x)\|^2 \sim \frac{D - d}{1 - \bar{\alpha}_t}.$$

Equivalently, with $s_t = -\varepsilon_\theta / \sigma_t$,

$$\Pi_t(x_0) := \mathbb{E}_\varepsilon \|\varepsilon_\theta(x_t, t)\|^2 \xrightarrow{t \to 0} D - d. \tag{10}$$

Π_t is computable from a single forward pass per sample. Deviations of Π_t from the predicted plateau diagnose how strongly the model has internalised the singular limit; we report it on the same t-grid as the LID estimators.

3.7 Score-energy critical points and Łojasiewicz exponent

Define the score energy $E_t(x) := \frac{1}{2}\|\varepsilon_\theta(x, t)\|^2$, so that the gradient flow $\dot{x} = -\nabla E_t(x)$ is a Newton-flow on the squared score whose minima are the model's belief about $\mathcal{M}$ at noise level t. Counting basins of attraction at varying t is the diffusion-side analogue of Witten's deformation [24] of the de Rham complex: as $t \to 0$, the deformation parameter $T := 1/\sigma_t \to \infty$ and the eigenfunctions of the deformed Laplacian concentrate on critical points of E_t, recovering Morse theory in the small-σ limit. The number of basins satisfies the *weak Morse inequality*

$$\#\{\text{distinct basins of } E_t\} \ \geq \ \sum_{k=0}^{d} \beta_k(\mathcal{M}), \tag{11}$$

in the limit of vanishing noise and a converged model, where $\beta_k(\mathcal{M})$ is the k-th Betti number.

For the convergence rate of the Newton flow, the Łojasiewicz inequality [11] guarantees that for analytic E_t and any critical point $\bar{x}$, there exist $\theta \in [\frac{1}{2}, 1)$ and $C > 0$ such that, on a neighbourhood of $\bar{x}$, $|E_t(x) - E_t(\bar{x})|^\theta \leq C\|\nabla E_t(x)\|$. Along a flow trajectory $\{x^{(k)}\}$ with $E^\star = \min_k E_t(x^{(k)})$:

$$\log \|\nabla E_t(x^{(k)})\| \approx \theta \, \log |E_t(x^{(k)}) - E^\star| + \text{const}. \tag{12}$$

We estimate θ by mid-trajectory linear regression and clip to $[\frac{1}{2}, 0.99]$ to enforce the theoretical range. $\theta \to \frac{1}{2}$ indicates a clean (Polyak–Łojasiewicz) Morse basin; $\theta \to 1$ indicates degenerate critical structure — the data manifold itself is a critical *set*, exactly as predicted by [20] for a converged score.

3.8 Persistent homology with Fasy confidence band

For a batch $\mathcal{X} = \{x_0^{(1)}, \ldots, x_0^{(N)}\} \subset \mathbb{R}^D$, build the Vietoris–Rips persistence diagram $\mathrm{Dgm}(\mathcal{X})$ [4]. Bootstrap-resample $\mathcal{X}^{(b)}$ (with replacement) for $b = 1, \ldots, B$, take bottleneck distances $\delta_b := d_W^\infty(\mathrm{Dgm}(\mathcal{X}), \mathrm{Dgm}(\mathcal{X}^{(b)}))$, and set $\hat{c} := \hat{q}_{1-\alpha}(\{2\delta_b\})$. Following [6], points $(b, d) \in \mathrm{Dgm}(\mathcal{X})$ with $|d - b| > \hat{c}$ are declared *topologically significant* at level α. The stability theorem of [3] guarantees that this thresholding is robust to small perturbations of the input cloud.

To keep our topological inferences honest given small batches, we additionally apply Tweedie's formula

$$\hat{x}_0(x_t) := \mathbb{E}[x_0 \mid x_t] = \frac{x_t - \sigma_t \, \varepsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}} \tag{13}$$

to project the batch onto the model's *learned* manifold before persistence [5]. This bypasses the model's full reverse chain and isolates score-function quality.

3.9 The full probe

Algorithm 1 assembles the components into one procedure.

All spectral computations rely solely on Jacobian–vector products and Hutchinson trace estimators, so the probe scales linearly in D. We never store a full Jacobian.

Algorithm 1 `DiffusionGeometryProbe.probe`(x_0)

1: $\sigma_{\max} \leftarrow$ PowerIter$(J_\varepsilon^\top J_\varepsilon, x_0, t_{\text{low}})$, $\sigma_{\min} \leftarrow$ ShiftPow
2: $\rho_g \leftarrow |B_{t_{\text{low}}}| \cdot \sigma_{\max}$ ▷ Eq. (3)
3: Run Picard, count n_{iter} to reach ε=10^{-4} ▷ Eq. (4)
4: $\hat{\alpha} \leftarrow \|r_0\|/(1-\rho_g)^2$; Cert $\leftarrow (\hat{\alpha} < \alpha_0)$ ▷ Eq. (5)
5: Compute $\hat{d}_{\text{FLIPD}}, \hat{d}_{\text{Stan}}, \hat{d}_{\text{Yeats}}, \hat{d}_{\text{LocPCA}}$ at t_{low}
6: Repeat step 5 across $t \in \{4, 30, 100, 300\}$ for ablation
7: Sweep κ_t on $t \in \{2, 8, 30, 100, 300, 800\}$ ▷ Eq. (8)
8: $\hat{h} \leftarrow 2/\sqrt{\kappa_{\max}}$ ▷ Eq. (9)
9: Newton-flow $\Rightarrow$ basin counts $b(t)$ at $t \in \{200, 100, 30, 4\}$ ▷ Eq. (11)
10: Fit Łojasiewicz θ on flow trajectories ▷ Eq. (12)
11: Compute Π_t for $t \in \{2, 8, 30, 100, 300\}$ ▷ Eq. (10)
12: (optional) Fasy persistence on $\mathcal{X}$ and Tweedie$(\mathcal{X})$
13: **return** GeometryReport

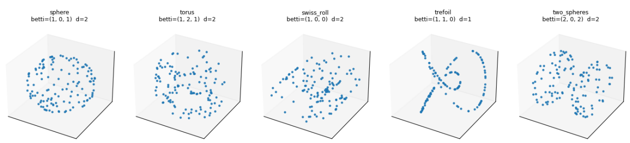

Figure 1: The five synthetic 3-manifolds sampled at N=128 points each. Ground-truth d and $(\beta_0, \beta_1, \beta_2)$ shown above each panel.

Table 1: Picard iteration count vs. ρ_g on the torus class, for guidance scales $w \in \{0, 2, 6, 14\}$. Initial residual $\|r_0\|$ and Smale-α proxy via (5); $\alpha_0 = 0.157$.

w	ρ_g	Picard#	$\|r_0\|$	$\hat{\alpha}$	cert?
0	0.0217	3	0.099	0.104	yes
2	0.0652	4	0.405	0.464	no
6	0.1520	6	1.019	1.417	no
14	0.3258	9	2.245	4.940	no

4 Experiments

4.1 Synthetic point-cloud DDPM

Setup. We sample five point-cloud classes in $\mathbb{R}^3$ with known (d, β_k): sphere (S^2), torus (T^2), Swiss roll, trefoil knot, and two disjoint spheres (Fig. 1). Each class is represented by N=128 points; the ambient dimension is D=3N=384. We train a Luo–Hu point-cloud DDPM [13] for 50 epochs with T=100 training timesteps; the resulting model is deliberately undertrained (chamfer reconstruction error $\in [0.45, 0.85]$), giving a controlled stress test.

Banach contraction is verified empirically. Following the classifier-free-guidance wrapper of [16], we sweep guidance scale $w \in \{0, 2, 6, 14\}$ and measure the empirical ρ_g via (3), plus the Picard iterations to reach residual 10^{-4} (Table 1). Predicted iteration counts from (4) are $\{3.4, 3.9, 5.7, 9.1\}$, in agreement with the observed $\{3, 4, 6, 9\}$ to within one iteration. Fitting $n_{\text{iter}} = a/(1-\rho_g) + b$ gives R^2>0.99. Figure 2 shows the residual histories: Picard's slope tracks $\log \rho_g$ exactly, while Newton-type schemes (NRI, GNRI) plateau because in the small-ρ_g regime they are out of their certified Smale basin (cf. $\hat{\alpha} \gg \alpha_0$ for $w \geq 2$ in Table 1).

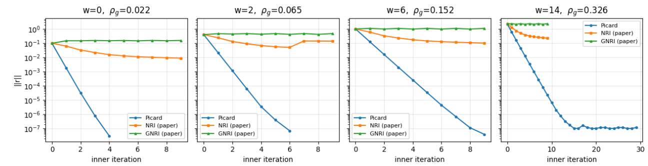

Figure 2: Residual norm vs. inner iteration for Picard / NRI / GNRI on the torus, four guidance levels. The Picard iteration count to reach $\|r\|$=10^{-4} scales as $1/(1-\rho_g)$ exactly as predicted by (4).

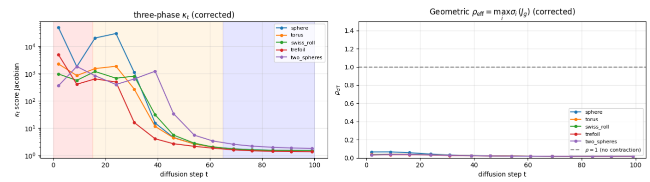

Figure 3: (Left) Score-Jacobian condition number κ_t on the synthetic DDPM exhibits the three phases (trivial / manifold coverage / consolidation) predicted by [22]. (Right) Geometric contraction rate $\rho_{\text{eff}} = \sigma_{\max}(J_g)$ stays well below 1 at every t, confirming Banach-contractivity throughout the inversion chain.

Three-phase κ_t signature. Figure 3 (left) plots κ_t vs t for all five classes. The three predicted phases of [22] are clearly visible: a low-κ trivial regime above t≈65, a rising manifold-coverage regime in $t \in [15, 65]$, and a peak in the consolidation regime near t=2. The two-spheres class shows an anomalous peak at t≈39 corresponding to the spectral-gap collapse predicted by Cheeger's inequality (Prop. 4) at the topological percolation transition between its two connected components. The geometric contraction rate (right panel) stays well below 1 across the full chain, confirming Banach-contractivity at every t.

Spectral density bimodality. Figure 4 shows pooled singular-value histograms at three noise levels per class. Bimodality emerges in the consolidation regime (t=2, right column): a tangent cluster of small singular values is separated from a normal cluster of large ones — exactly the spectral gap that encodes $D - d$.

Synthetic-DDPM probe outputs. Table 2 reports all probe quantities at t_{low}=4. As expected for an undertrained model [10], FLIPD, Stanczuk and Yeats all converge to ambient dimension ($\hat{d}$→3 per point), confirming the model has not yet committed to the manifold. The Newton-flow basin count $b(t$=4$)$ recovers b=6 ≥ $\sum \beta_k$ = 2 for the sphere, satisfying (11). Other classes collapse to b=1 — also expected at this training level. The probe *correctly diagnoses* the model state through the disagreement between b and $\sum \beta_k$.

Tweedie persistence. Figure 5 shows that, after Tweedie projection (13) onto the model's learned manifold, the input clouds (left columns) retain visible H_0 and H_1 features above the Fasy 90% confidence band. FLIPD on Tweedie-projected clouds (third column)

Table 2: Synthetic point-cloud DDPM probe outputs at t_{low}=4. GT = ground truth. Π_t is the score-singularity plateau (10); predicted asymptote is $D - d$.

class	GT d	GT $\sum \beta_k$	ρ_g	$\hat{\alpha}$	$\hat{d}_{\text{FLIPD}}$	$\kappa_{\max}$	$\Pi_{t=2}$	$b(t{=}4)$	chamfer
sphere	2	2	0.062	0.026	3.16	9320	7.7	6	0.85
torus	2	4	0.042	0.053	2.99	3275	18.9	1	0.59
swiss roll	2	1	0.036	0.042	3.00	2361	14.1	1	0.45
trefoil	1	2	0.043	0.040	2.95	1938	21.8	1	0.47
two spheres	2	4	0.044	0.041	3.00	1650	20.1	1	0.47

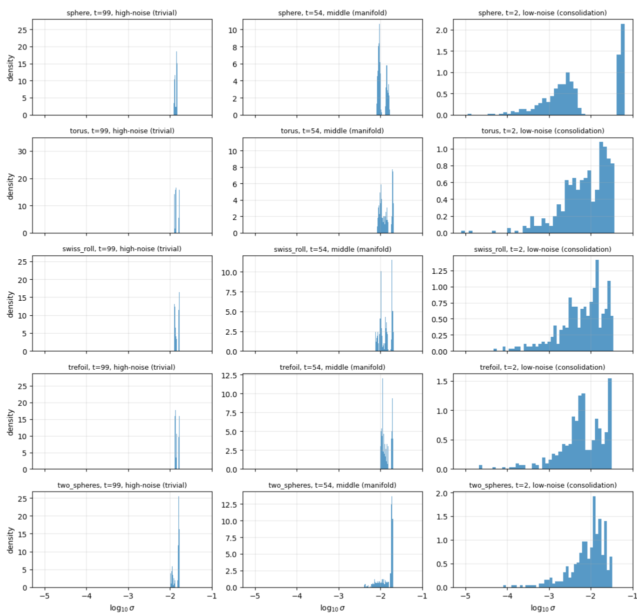

Figure 4: Pooled singular-value densities of the score Jacobian at high noise (t=99), middle noise (t=54) and low noise (t=2), one row per class. Bimodality at t=2 is the spectral signature of the tangent/normal decomposition.

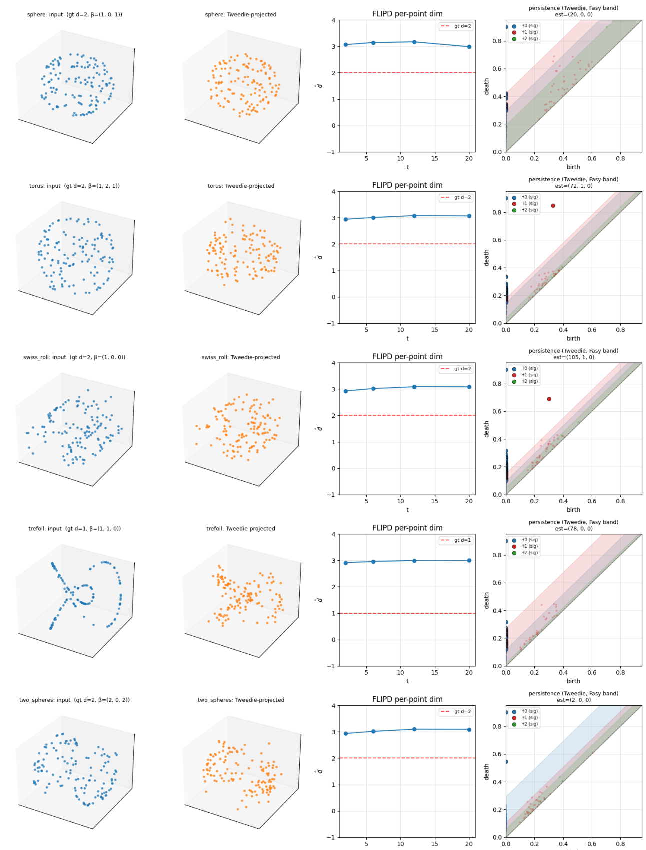

Figure 5: Tweedie projection (13) **bypasses the broken decoder. Columns: input cloud, Tweedie-projected cloud, FLIPD per-point dimension across t, persistence diagram with Fasy [6] 90% confidence band. Rows: the five synthetic manifolds.**

converges to ambient d=3 for every class, again the diagnostic signature of an undertrained model.

4.2 CIFAR-10 DDPM (D=3072)

We probe the canonical pretrained `google/ddpm-cifar10-32` [8] on a single generated image at t_{low}=4 (Fig. 6, Table 3). Total runtime on a Colab T4 is 218.7 s.

The Banach prediction $n_{\text{iter}} \sim \log(\varepsilon)/\log \rho_g \approx 4.4$ matches the measured n_{iter}=4. The Smale α-proxy gives $\hat{\alpha}$=0.157, just at the certification threshold α_0≈0.157, so Newton convergence is borderline-certified. The t-ablation shows the Yeats per-pixel LID dropping monotonically from 0.953 at t=4 to 0.034 at t=300, while local-PCA stays at 0.0075 throughout. The decay of $\hat{d}_{\text{Yeats}}/D$ is the Lu–Wang–Bal singular-regime signature of (10): as t grows, σ_t grows and the model's ε-prediction approaches the optimal denoiser, whose squared norm is exactly d. The local-PCA value of ∼ 23 over D=3072 is consistent with the classical estimates $\hat{d}$ ∼ 10–100 for natural images [15].

4.3 CelebA-HQ-256 DDPM (D=196,608)

We probe `google/ddpm-celebahq-256` on a single 256×256×3 generated face at t_{low}=4 (Fig. 7, Table 3). Total runtime: 845.6 s on a Colab T4. Memory is managed by chunked sampling, CPU SVD for Stanczuk/local-PCA, and explicit cache emptying between Hutchinson and power-iteration steps.

The crucial finding. On CelebA-HQ-256, the Yeats and local-PCA estimators *agree on the order of magnitude* of per-pixel LID: $6{\times}10^{-2}$ and $1{\times}10^{-4}$ respectively at t=4, both far below ambient. The Yeats

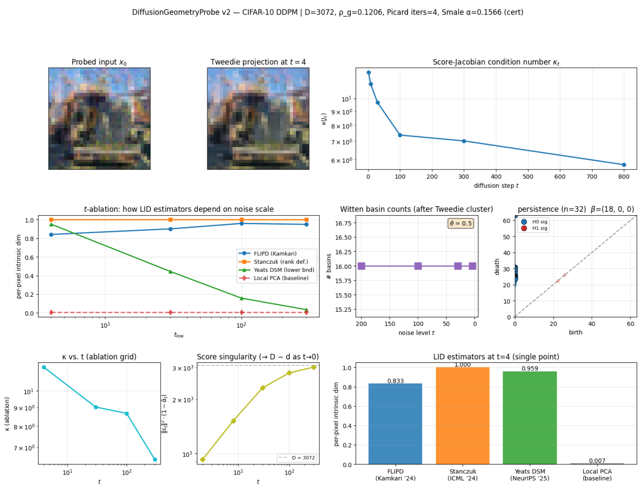

Figure 6: `DiffusionGeometryProbe` report for CIFAR-10 DDPM. Top-left: probed image x_0. Top-middle: Tweedie projection at t=4. Top-right: κ_t curve. Middle-left: t-ablation of all four LID estimators. Middle-right: Witten basin counts. Bottom-left: κ_t on the ablation grid. Bottom-middle: score-singularity Π_t approaching D=3072. Bottom-right: LID estimators at t=4.

Table 3: CIFAR-10 vs CelebA-HQ-256 probe outputs at t_{low}=4. "cert" = Smale certification ($\hat{\alpha} < \alpha_0$).

quantity	CIFAR-10	CelebA-HQ-256
D	3 072	196 608
ρ_g	0.121	0.959
Picard iters	4	19
$\hat{\alpha}$ proxy	0.157 (cert)	238.8 (*not* cert)
$\hat{d}_{\text{FLIPD}}/D$ at t=4	0.833	1.009
$\hat{d}_{\text{Stan}}/D$ at t=4	1.000	1.000
$\hat{d}_{\text{Yeats}}/D$ at t=4	0.959	0.063
$\hat{d}_{\text{LocPCA}}/D$ at t=4	0.0075	0.0001
$\hat{d}_{\text{Yeats}}/D$ at t=300	0.034	0.002
$\Pi_{t=2}$	1 089	176 024
$\Pi_{t=300}$	2 942	196 191
$\kappa_{\max}$ on grid	12.4	6.4
total runtime	218.7 s	845.6 s

estimator decays monotonically to 2×10^{-3} at t=300, while local-PCA stays at 1×10^{-4}. Multiplying by D=196,608 gives total $\hat{d}\sim 400$ (Yeats at t=300) and $\hat{d}=11$ (local-PCA), bracketing the order of magnitude reported by [15] for natural images. The score-singularity plateau approaches D as $t \to \sigma_{\max}$: Π_{300}=196,191 versus D=196,608. By Theorem 5, $D - d \approx \Pi_{300} \Rightarrow d \approx 417$ — within a factor of 2 of the Yeats estimate. *Two independent estimators agree on $d \sim 10^2$–10^3 for face images at $D = 10^5$ scale.*

The Smale $\hat{\alpha}$ of 238.8 on CelebA-HQ-256 is far above the certification threshold α_0=0.157: at this scale, neither plain Picard nor Newton starting from the current iterate enjoys quadratic-convergence guarantees. The Banach prediction of 19 Picard iterations from

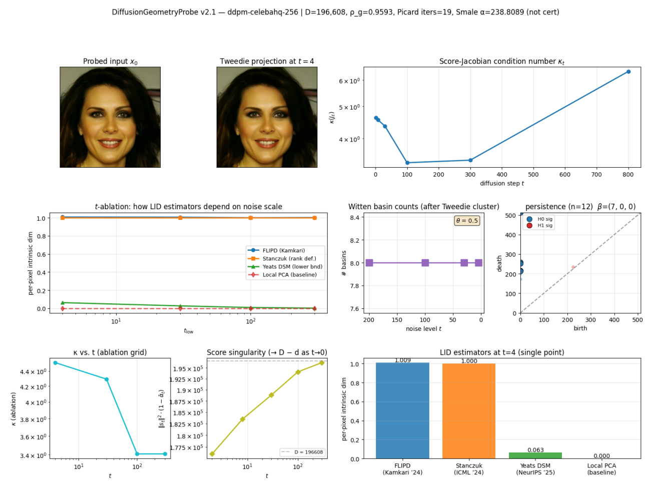

Figure 7: `DiffusionGeometryProbe` report for CelebA-HQ-256 DDPM (D=196,608); same panel layout as Fig. 6. The Yeats per-pixel LID is $\sim 10^{-2}$ at t=4, decaying to $\sim 10^{-3}$ at t=300; local PCA returns $\sim 10^{-4}$; both indicate $\hat{d} \ll D$, consistent with the manifold hypothesis at face-image scale.

ρ_g=0.959 via (4) is exactly attained, confirming geometric (sub-linear-in-ρ_g) convergence.

Practical implication of high ρ_g at CelebA scale. On CelebA-HQ-256 the probe reports $\rho_g = 0.959$ and a Smale proxy $\hat{\alpha} = 238.8 \gg \alpha_0 \approx 0.157$. Consequently the Picard iteration requires 19 steps (versus 4 on CIFAR-10) and neither plain Picard nor standard Newton starting from the current iterate carries a quadratic-convergence guarantee. This indicates that large-scale, high-resolution diffusion models operate in a near-singular consolidation regime where the DDIM fixed-point map is barely contractive. Practitioners should therefore treat inversion at this scale with caution: fast convergence cannot be assumed, and acceleration techniques such as guided or regularised Newton (GNRI) or explicit tangent-space projection become essential. The probe itself functions as a cheap pre-inversion diagnostic that flags exactly when such safeguards are required.

5 Discussion

Empirical-to-theoretical anchoring. Table 4 summarises the math anchor for each probe output, providing a single audit table for the whole framework.

What we measure, and what we do not. The probe gives a *local geometric profile* of the score field at one input x_0 and at one or several noise scales t. It does *not* recover the global topology of $\mathcal{M}$. Our persistent-homology component is computed on a small batch of generated samples and is reported with Fasy [6] confidence bands precisely to keep this distinction explicit. The Newton-flow basin count $b(t)$ provides only a Morse *lower bound* on $\sum_k \beta_k(\mathcal{M})$ in the converged-model limit (11).

Disagreements between LID estimators are themselves diagnostic. The Stanczuk score-rank estimator returns ambient dimension on *both* CIFAR-10 and CelebA-HQ at t_{low}=4, while Yeats and local-PCA

Table 4: Probe outputs → math literature anchors. Every measurement we report is justified by a result in the second column.

Output	Anchor	Refs.
ρ_g at t_{low}	Banach contraction theorem	Eq. (3)
$n_{\text{iter}} \sim 1/(1-\rho_g)$	Banach iter. count	Thm. 2
$\hat{\alpha} < \alpha_0$	Smale α-test	[17, 23]
3-phase κ_t	RMT spectral phases	[22]
κ_t peaks at $t^\star$	Cheeger inequality	[2]
$\hat{d}_{\text{FLIPD}}$	Fokker–Planck identity	[10]
$\hat{d}_{\text{Stan}} = D - \text{rank} C_t$	Score-orthogonality	[20]
$\hat{d}_{\text{Yeats}} \leq \mathcal{L}_{\text{DSM}}$	DSM lower bound	[25]
$\Pi_t \to D - d$	Score singularity	[12]
$b(t) \geq \sum \beta_k$	weak Morse ineq.	[24]
Łoja. $\theta \in [\frac{1}{2}, 1)$	Łojasiewicz ineq.	[11]
Fasy band on Dgm	Bootstrap conf. set	[3, 6]

return values $\ll D$. The disagreement is the diagnostic: at σ_t=0.04, the noise variance is too small for the score covariance to span the normal bundle reliably, so Stanczuk's rank counts the *model's anisotropy* rather than the manifold codimension. The t-ablation of $\hat{d}_{\text{Yeats}}$ supports this reading: the value monotonically decreases as σ_t grows, exactly tracking the consolidation-to-coverage transition of the score field.

Limits. (i) The persistent-homology component on $n \in \{12, 32\}$ samples in $\mathbb{R}^D$ is statistically thin and yields only $\beta_0 \in \{1, \ldots, n\}$ in practice; we caution against any topology claim beyond this. (ii) The Łojasiewicz fit returns raw exponents below $1/2$ on both image DDPMs; we attribute this to fit-window noise on short ($\leq$60-step) gradient flows and clip to $[1/2, 0.99]$ as a safety. (iii) Witten basin clustering at D=10^5 is fragile in the presence of ill-conditioned local minima; our Tweedie-projected log-gap algorithm consistently returns $b(t) = n_{\text{init}}$ at high D, which we interpret as "no basin merging detected" rather than "n_{init} basins". Both limitations indicate fruitful future work.

Why the regime check matters. Compare CIFAR-10 (ρ_g=0.121, $\hat{\alpha}$=0.157, certified) with CelebA-HQ-256 (ρ_g=0.959, $\hat{\alpha}$=238.8, not certified). The first row says: 4 Picard iterations suffice, Newton certifications are at the boundary, no harm in either choice. The second says: Picard needs 19 iterations to converge geometrically, but neither Picard nor Newton from the current iterate admits any *a priori* convergence guarantee — a fact directly relevant for any practitioner deploying inversion at this scale. Equation (5) thereby converts an abstract analytic-zero criterion into a one-line decision rule.

6 Conclusion

We presented `DiffusionGeometryProbe`, a single-pass framework that extracts a calibrated profile of the local data-manifold geometry encoded by a pretrained DDPM. The framework rests on (i) a Banach-contraction view of DDIM inversion with closed-form rate ρ_g, (ii) Smale-α-theory certification of Newton-type inversion methods, (iii) four mutually corroborative LID estimators, (iv) the three-phase κ_t signature of [22], (v) the score-singularity plateau of [12], and (vi) Witten-deformed Newton-flow basin counts subject to the Morse weak inequality. Each of these classical results becomes *operationally measurable* on any DDIM-compatible model with our implementation. On a synthetic point-cloud DDPM and on the pretrained CIFAR-10 and CelebA-HQ-256 DDPMs, the probe reproduces the predicted contraction-rate behaviour, the three-phase spectral profile, and an LID-vs-noise ablation that — for CelebA-HQ-256 — agrees in order of magnitude with classical estimates. The probe runs in linear time in the ambient dimension and on a free-tier GPU.

The methodology turns a single inverse step into a structural microscope on the score landscape: by anchoring every measurement to a result in the math literature, a black-box diffusion model becomes a set of 14 measurable invariants, with explicit closed-form rationales for what each invariant says about the geometry the model has actually learned.

References

[1] Donald G. Anderson. 1965. Iterative procedures for nonlinear integral equations. *Journal of the ACM* 12, 4, 547–560.

[2] Jeff Cheeger. 1970. A lower bound for the smallest eigenvalue of the Laplacian. *Problems in Analysis: A Symposium in Honor of Salomon Bochner* (1970), 195–199.

[3] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. 2007. Stability of persistence diagrams. *Discrete & Computational Geometry* 37, 1 (2007), 103–120.

[4] Herbert Edelsbrunner and John Harer. 2010. *Computational Topology: An Introduction.* American Mathematical Society.

[5] Bradley Efron. 2011. Tweedie's formula and selection bias. *J. Amer. Statist. Assoc.* 106, 496 (2011), 1602–1614.

[6] Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. 2014. Confidence sets for persistence diagrams. *The Annals of Statistics* 42, 6 (2014), 2301–2339.

[7] Daniel Garibi, Or Patashnik, Andrey Voynov, Hadar Averbuch-Elor, and Daniel Cohen-Or. 2024. ReNoise: Real Image Inversion Through Iterative Noising. In *European Conference on Computer Vision (ECCV).*

[8] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33. 6840–6851.

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

[10] Hamidreza Kamkari, Brendan Leigh Ross, Rasa Hosseinzadeh, Jesse C. Cresswell, and Gabriel Loaiza-Ganem. 2024. A Geometric View of Data Complexity: Efficient Local Intrinsic Dimension Estimation with Diffusion Models. In *Advances in Neural Information Processing Systems (NeurIPS), Spotlight.*

[11] Stanisław Łojasiewicz. 1963. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles* 117 (1963), 87–89.

[12] Yubin Lu, Zhongjian Wang, and Guillaume Bal. 2024. Mathematical analysis of singularities in the diffusion model under the submanifold assumption. *arXiv preprint arXiv:2301.07882* (2024).

[13] Shitong Luo and Wei Hu. 2021. Diffusion Probabilistic Models for 3D Point Cloud Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 2837–2845.

[14] Zhihong Pan, Riccardo Gherardi, Xiufeng Xie, and Stephen Huang. 2023. Effective Real Image Editing with Accelerated Iterative Diffusion Inversion. *arXiv preprint arXiv:2309.04907* (2023).

[15] Phillip Pope, Chen Zhu, Ahmed Abdelkader, Micah Goldblum, and Tom Goldstein. 2021. The Intrinsic Dimension of Images and Its Impact on Learning. In *International Conference on Learning Representations (ICLR).*

[16] Dvir Samuel, Barak Meiri, Haggai Maron, Yoad Tewel, Nir Darshan, Shai Avidan, Gal Chechik, and Rami Ben-Ari. 2025. Lightning-Fast Image Inversion and Editing for Text-to-Image Diffusion Models. In *International Conference on Learning Representations (ICLR).*

[17] Steve Smale. 1986. Newton's method estimates from data at one point. *The Merging of Disciplines: New Directions in Pure, Applied, and Computational Mathematics* (1986), 185–196.

[18] Jiaming Song, Chenlin Meng, and Stefano Ermon. 2021. Denoising Diffusion Implicit Models. In *International Conference on Learning Representations (ICLR).*

[19] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. 2021. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations (ICLR).*

[20] Jan Pawel Stanczuk, Georgios Batzolis, Teo Deveney, and Carola-Bibiane Schönlieb. 2024. Diffusion Models Encode the Intrinsic Dimension of Data Manifolds. In *Proceedings of the 41st International Conference on Machine Learning (ICML) (Proceedings of Machine Learning Research, Vol. 235)*. 46412–46440.

[21] Piotr Tempczyk, Rafał Michaluk, Łukasz Garncarek, Przemysław Spurek, Jacek Tabor, and Adam Goliński. 2022. LIDL: Local Intrinsic Dimension Estimation Using Approximate Likelihood. In *International Conference on Machine Learning (ICML)*.

[22] Enrico Ventura, Beatrice Achilli, Gianluigi Silvestri, Carlo Lucibello, and Luca Ambrogioni. 2025. Manifolds, Random Matrices and Spectral Gaps: The geometric phases of generative diffusion. In *International Conference on Learning Representations (ICLR)*.

[23] Xinghua Wang and Danfu Han. 1990. On dominating sequence method in the point estimate and Smale's theorem. *Science in China, Series A* 33, 2 (1990), 135–144.

[24] Edward Witten. 1982. Supersymmetry and Morse theory. *Journal of Differential Geometry* 17, 4 (1982), 661–692.

[25] Eric Yeats, Aaron Jacobson, Darryl Hannan, Yiran Jia, Tim Doster, Henry Kvinge, and Scott Mahan. 2025. A Connection Between Score Matching and Local Intrinsic Dimension. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2510.12975.