Optimization Landscape and Feasibility in Updated Riemannian AmbientFlow

Anonymous Author(s)

ABSTRACT

Riemannian AmbientFlow augments the AmbientFlow variational lower bound with a geometric regularization term—the squared Frobenius norm of the Jacobian of the learned diffeomorphism at the origin—to encourage low-dimensional manifold structure in generative models trained on corrupted data. A theoretical recoverability result holds under feasibility assumptions: the existence of parameters achieving exact data distribution matching, posterior matching, and geometric constraint satisfaction. However, the optimization landscape is nonconvex, and it remains an open question which local minima are reached and whether feasibility holds at those minima. We investigate this open problem through systematic computational experiments on three synthetic manifold-learning problems (circle in $\mathbb{R}^2$, sphere in $\mathbb{R}^3$, helix in $\mathbb{R}^3$) using two model architectures—a simple affine-plus-tanh parameterization (9–19 parameters) and a multi-layer perceptron diffeomorphism (1186–1251 parameters)—across regularization strengths $\lambda \in [0, 2]$. Through multi-start optimization, oracle feasibility analysis, convergence diagnostics, sensitivity studies, Hessian spectral analysis, and pullback metric comparison, we characterize the landscape structure and assess feasibility at converged solutions. Our oracle feasibility analysis confirms that both parameterizations can approximate the ground-truth embedding (F1 > 0.9), establishing that infeasibility at ELBO local minima reflects landscape properties rather than capacity limitations alone. Our results reveal that (i) all converged critical points show strictly positive directional curvature; (ii) increasing λ monotonically decreases the Jacobian norm but introduces a feasibility trade-off where data-matching degrades; (iii) the pullback metric at learned solutions substantially underestimates the true metric; and (iv) feasibility scores exhibit non-monotonic behavior in λ, with 95% bootstrap confidence intervals quantifying statistical uncertainty. These findings provide a systematic empirical investigation of the landscape-feasibility trade-off in Riemannian AmbientFlow and suggest that the feasibility assumptions of the recoverability theorem are not satisfied at local minima found by gradient-based optimization in the tested configurations, even when the model class has sufficient oracle capacity.

1 INTRODUCTION

Generative modeling on low-dimensional manifolds embedded in high-dimensional ambient spaces is a fundamental challenge in machine learning. When observations are corrupted by noise, the problem becomes even more difficult: the generative model must simultaneously recover the latent manifold structure and learn to generate new data consistent with the ground-truth distribution.

AmbientFlow [16] introduced a variational framework for this setting, training normalizing flows [26, 27] on noisy observations via a variational lower bound. Kelkar et al. [18] extended this to invertible generative models for incomplete measurements. Diepeveen et al. [8] recently proposed *Riemannian AmbientFlow*, which augments this objective with a geometric regularization term derived from pullback Riemannian geometry. The updated objective takes the form:

$$\mathcal{L}(\theta, \phi) = \mathcal{L}_{\text{AF}}(\theta, \phi) + \lambda \cdot \|J_{f_\theta}(0)\|_F^2, \tag{1}$$

where $\mathcal{L}_{\text{AF}}$ is the (negative) AmbientFlow ELBO, $f_\theta : \mathbb{R}^d \to \mathbb{R}^D$ is the learned diffeomorphism mapping the latent space to the ambient space, $J_{f_\theta}(0)$ is its Jacobian evaluated at the origin, and $\lambda \geq 0$ controls the regularization strength.

The Frobenius norm penalty $\|J_{f_\theta}(0)\|_F^2 = \text{Tr}(G_\theta(0))$, where $G_\theta(z) = J_{f_\theta}(z)^\top J_{f_\theta}(z)$ is the pullback metric, encourages the learned map to preserve low-dimensional structure by penalizing excessive stretching at the origin.

Diepeveen et al. [8] prove a recoverability theorem under three *feasibility assumptions*:

- **(F1)** There exist parameters (θ^*, ϕ^*) such that the learned data distribution p_{θ^*} equals the ground-truth data distribution p_{data}.
- **(F2)** The learned variational posterior $q_{\phi^*}(z|y)$ equals the true posterior $p_{\theta^*}(z|y)$.
- **(F3)** The geometric constraint $\|J_{f_{\theta^*}}(0)\|_F^2 \leq C$ is satisfied for some constant C.

However, as the authors note, the optimization problem (1) is nonconvex, and it is not guaranteed which local minimum gradient-based training will reach, nor whether the feasibility assumptions hold at the converged solution. This constitutes an open problem at the intersection of nonconvex optimization, Riemannian geometry, and variational inference.

In this paper, we provide a systematic computational investigation of this open problem on controlled synthetic problems. We design experiments with known ground-truth manifolds and corruption models, enabling exact assessment of all three feasibility conditions at converged solutions. Critically, we introduce an *oracle feasibility experiment* that separates model capacity limitations from optimization landscape effects by fitting the parameterization directly to the ground-truth embedding (bypassing the ELBO objective). Our oracle analysis confirms that both a simple and an MLP parameterization have sufficient capacity to represent the true map (F1 > 0.9), allowing us to conclude that infeasibility at ELBO local minima is primarily a property of the landscape. We further provide convergence diagnostics, sensitivity analyses across noise levels and sample sizes, bootstrap confidence intervals, and a 2D landscape cross-section visualization.

1.1 Related Work

Normalizing flows on manifolds. Standard normalizing flows [26, 27] learn invertible maps between a simple base distribution and a complex target. When the target lives on a low-dimensional manifold, approaches include neural ODEs on manifolds [5, 22], Riemannian continuous normalizing flows [24], and homeomorphic

VAEs [13]. AmbientFlow [16] works with corrupted ambient-space observations, avoiding the need to explicitly parameterize the manifold. Vérine et al. [28] analyze the expressivity limitations of bi-Lipschitz normalizing flows, providing theoretical context for understanding capacity constraints in flow-based models.

Geometric methods in generative models. Diepeveen and Needell [10] study manifold learning with normalizing flows from the perspective of regularity, expressivity, and iso-Riemannian geometry, providing a theoretical foundation for understanding when flow-based models can faithfully capture manifold structure. Diepeveen et al. [9] develop score-based pullback Riemannian geometry, extending the pullback metric framework beyond the deterministic setting. Chen and Lipman [4] introduce flow matching on general geometries, while de Kruiff et al. [7] study pullback flow matching on data manifolds. De Bortoli et al. [6] and Huang et al. [17] develop Riemannian score-based and diffusion models, respectively. These works collectively establish the importance of Riemannian geometry in modern generative modeling.

Optimization landscapes in deep learning. The landscape of nonconvex objectives has been studied extensively. In certain matrix problems, all local minima are global [3, 14]. Levin et al. [21] analyze the effect of smooth parameterizations on nonconvex optimization landscapes, showing how the parameterization structure can shape the critical point landscape. Ding et al. [11] provide sharp global guarantees for nonconvex low-rank recovery, establishing conditions under which gradient descent provably finds global optima. For VAE-type objectives, posterior collapse represents a known class of spurious local minima [19, 23]. Riemannian optimization [1] provides tools for optimization on manifolds, but the landscape of objectives mixing variational inference with geometric regularization remains poorly understood.

Pullback geometry in generative models. The pullback metric $G(z) = J_f(z)^\top J_f(z)$ captures the Riemannian geometry induced by a smooth map $f : \mathbb{R}^d \to \mathbb{R}^D$ [12, 20]. Arvanitidis et al. [2] study the curvature of deep generative models through the lens of pullback geometry. Diepeveen et al. [8] use this to regularize generative models, penalizing $\mathrm{Tr}(G(0)) = \|J_f(0)\|_F^2$ to encourage geometric consistency with the intrinsic manifold dimension.

2 METHODS

2.1 Problem Setup

We study the objective (1) on three synthetic manifold-learning problems with known ground truth:

(1) **Circle in $\mathbb{R}^2$** ($d = 1, D = 2$): The unit circle S^1 parameterized by $f^*(z) = (\cos z, \sin z)$, with ground-truth geometric constant $C^* = \mathrm{Tr}(G^*(0)) = 1.0$.
(2) **Sphere in $\mathbb{R}^3$** ($d = 2, D = 3$): The unit sphere S^2 via inverse stereographic projection $f^*(z_1, z_2) = \frac{1}{|z|^2+1}(2z_1, 2z_2, |z|^2 - 1)$, with $C^* = \mathrm{Tr}(G^*(0)) = 8.0$.
(3) **Helix in $\mathbb{R}^3$** ($d = 1, D = 3$): A helix $f^*(t) = (\cos t, \sin t, t/2\pi)$, with $C^* = \mathrm{Tr}(G^*(0)) \approx 1.025$.

For each problem, we generate $n = 200$ data points, corrupt them with additive Gaussian noise ($\sigma = 0.1$ unless otherwise specified), and optimize (1) using L-BFGS-B [25]. The Monte Carlo samples used within the ELBO are computed with a fixed random seed per evaluation, making the objective deterministic for given parameters.

2.2 Parameterizations

We employ two parameterizations of the diffeomorphism $f_\theta : \mathbb{R}^d \to \mathbb{R}^D$.

Simple parameterization. The diffeomorphism is parameterized as

$$f_\theta(z) = Az + b + \varepsilon \cdot \tanh(Wz + c), \tag{2}$$

where $A \in \mathbb{R}^{D\times d}$ is initialized near-orthogonally, $W \in \mathbb{R}^{D\times d}$ captures nonlinear structure, $b, c \in \mathbb{R}^D$ are biases, and $\varepsilon > 0$ controls the nonlinear perturbation strength. This yields 9 parameters (circle), 19 parameters (sphere), and 13 parameters (helix). The Jacobian at the origin is:

$$J_{f_\theta}(0) = A + \varepsilon \cdot \mathrm{diag}(\mathrm{sech}^2(c)) \cdot W. \tag{3}$$

MLP parameterization. To test whether increased model capacity changes the landscape-feasibility trade-off, we also employ an MLP diffeomorphism with 2 hidden layers of 32 units each and tanh activations:

$$f_\theta(z) = W_3 \tanh(W_2 \tanh(W_1 z + b_1) + b_2) + b_3, \tag{4}$$

where $W_1 \in \mathbb{R}^{32\times d}$, $W_2 \in \mathbb{R}^{32\times 32}$, $W_3 \in \mathbb{R}^{D\times 32}$, and b_i are biases. This yields 1186 parameters (circle), 1251 parameters (sphere), and 1219 parameters (helix)—a 60×–130× increase over the simple model.

Variational posterior. For the simple model, the variational posterior $q_\phi(z|y)$ is a diagonal Gaussian with amortized parameters: $\mu(y) = Vy + v_0$ and $\log \sigma(y) = Uy + u_0$ (linear encoder). For the MLP model, we use a nonlinear encoder with one hidden layer of 32 units. We note that the linear encoder limits posterior matching (F2) for nonlinear manifolds; the oracle experiment (Section 2.4) isolates the decoder capacity question (F1, F3) from this confound.

2.3 Feasibility Score Design

We use the ground-truth geometric constant $C^* = \mathrm{Tr}(G^*(0))$ as the threshold for condition F3, rather than an arbitrary constant. The three components are:

$$\mathrm{F1} = \exp(-\mathrm{MMD}), \tag{5}$$

$$\mathrm{F2} = \exp(-\mathrm{PM}), \tag{6}$$

$$\mathrm{F3} = \mathbf{1}[\|J_{f_\theta}(0)\|_F^2 \le C^*], \tag{7}$$

where MMD is the maximum mean discrepancy [15] (bandwidth = 1.0), PM is the posterior mismatch (MSE between encoded means and true latents), and C^* is the ground-truth metric trace. We present F1, F2, F3 as the primary feasibility metrics and define the aggregate score as $\mathcal{F} = \mathrm{F1} \cdot \mathrm{F2} \cdot \mathrm{F3}$.

Using $C^* = \mathrm{Tr}(G^*(0))$ ensures that F3 is meaningful: the geometric penalty must not suppress the Jacobian norm below the level required by the true manifold geometry. We note that the MMD uses a fixed bandwidth of 1.0 throughout; this is a common default but the MMD is sensitive to bandwidth selection [15], and a multi-bandwidth or median heuristic approach would be more robust. We leave this refinement to future work.

2.4 Oracle Feasibility

A critical question is whether infeasibility at ELBO local minima reflects an intrinsic property of the optimization landscape or merely insufficient model capacity. To disentangle these effects, we introduce an *oracle feasibility* experiment that bypasses the ELBO entirely.

For each parameterization (simple and MLP), we fit f_θ directly to the ground-truth embedding f^* by minimizing the reconstruction error $\|f_\theta(z_i) - f^*(z_i)\|^2$ over a dense grid of latent points. We then evaluate the feasibility diagnostics at the oracle solution:

- **Oracle F1**: $\exp(-\mathrm{MMD}(f_\theta(z_i), f^*(z_i)))$ measures data distribution matching at the oracle.
- **Oracle F3**: $\mathbf{1}[\|J_{f_\theta}(0)\|_F^2 \le C^*]$ checks the geometric constraint.

If the oracle achieves high F1 and F3, then the model class *can* represent a feasible solution, and any infeasibility observed during ELBO optimization must be attributed to the landscape rather than capacity.

2.5 Experimental Protocol

We conduct the following experiments:

Experiment 1: Multi-start landscape exploration. For each manifold and each $\lambda \in \{0, 0.01, 0.05, 0.1, 0.5, 1.0, 2.0\}$, we run $K = 15$ independent optimizations from random initializations (200 L-BFGS-B iterations each). At each converged solution, we evaluate the objective value and the three feasibility diagnostics. We report convergence status (gradient norm $< 10^{-5}$) for all runs.

Experiment 2: Parameter continuation. Starting from a single random initialization at $\lambda = 0$, we track the local minimum as λ increases from 0 to 2 in 30 steps, using the previous solution as warm-start for each step.

Experiment 3: Hessian spectral analysis. At converged solutions for each $\lambda \in \{0, 0.1, 0.5, 1.0\}$, we compute 50 random directional second derivatives $v^\top H v$ via finite differences (step size $h = 10^{-4}$, 32 Monte Carlo samples with fixed seed for noise reduction). We report these as *directional curvature samples*, noting that the minimum of $v^\top H v$ over random unit vectors v provides an *upper bound* on the minimum eigenvalue of H, not a direct estimate.

Experiment 4: Pullback geometry analysis. For converged solutions at each λ, we compute the pullback metric $G_\theta(z) = J_{f_\theta}(z)^\top J_{f_\theta}(z)$ at 40 random points and compare with the ground-truth metric $G^*(z)$.

Experiment 5: Sensitivity analysis. We vary the noise level $\sigma \in \{0.01, 0.05, 0.1, 0.2, 0.5\}$ and the sample size $n \in \{50, 100, 200, 500\}$ at fixed $\lambda = 0.1$, running 10 starts for each configuration.

Experiment 6: Bootstrap confidence intervals. For each λ, we compute 95% bootstrap confidence intervals from 15 runs ($B = 200$ resamples) for the mean feasibility score. Figure 1 provides an overview of the complete experimental framework and the relationships between its six experiments.

Table 1: Oracle feasibility: fitting f_θ directly to f^* (bypassing ELBO). Both parameterizations achieve F1 > 0.9 on all manifolds, confirming sufficient capacity. The sphere shows persistent metric underestimation (trace 3.2–3.8 vs. 8.0) even at oracle, indicating a geometry-specific limitation. C^*: ground-truth metric trace.

Manifold	Model	Params	$\mathrm{F1}_{\mathrm{ora}}$	$\mathrm{F3}_{\mathrm{ora}}$	$\mathrm{Tr}(G_\theta)$	C^*
Circle	Simple	9	0.998	0.967	1.034	1.0
	MLP	1186	0.957	0.986	1.014	1.0
Sphere	Simple	19	0.909	1.000	3.764	8.0
	MLP	1251	0.917	1.000	3.208	8.0
Helix	Simple	13	0.998	0.967	1.059	1.025
	MLP	1219	0.952	0.860	1.176	1.025

3 RESULTS

3.1 Oracle Feasibility

Table 1 presents the oracle feasibility results. Both the simple and MLP parameterizations achieve high oracle F1 scores across all manifolds: simple model scores of 0.998, 0.909, and 0.998 on circle, sphere, and helix respectively; MLP scores of 0.957, 0.917, and 0.952. This confirms that both model classes have *sufficient capacity* to approximate the ground-truth embedding.

The oracle F3 scores are also high (≥ 0.86), with the exception that the sphere exhibits persistent metric underestimation: the learned trace at oracle is 3.76 (simple) and 3.21 (MLP) compared to the true value of 8.0. This indicates that the sphere's high curvature from stereographic projection poses a geometry-specific challenge even when the ELBO landscape is removed.

The key implication is that infeasibility observed at ELBO local minima (Section 3.4) cannot be attributed solely to model capacity, since both parameterizations *can* represent near-feasible solutions. Instead, the ELBO optimization landscape guides the optimizer to infeasible basins.

3.2 Landscape Structure and Convergence

Figure 2 shows the objective value across λ for each manifold. The objective spread is largest at $\lambda = 0$, reaching 1.84 for the circle and 2.08 for the helix, indicating multiple distinct local minima in the unregularized landscape. The sphere shows remarkably low spread (< 0.28) at all λ values, suggesting a simpler landscape for higher-dimensional manifolds.

Figure 3 presents convergence diagnostics. Most optimization runs (93% on average) terminate at the 200-iteration limit rather than reaching the gradient tolerance of 10^{-5}. On the circle, 1–4 out of 15 runs converged per λ setting; on the sphere, 0–1; on the helix, 0–4. However, the final gradient norms are generally small (median < 0.3), indicating that the solutions are near-stationary even when the strict convergence criterion is not met. This justifies treating the terminal solutions as approximate critical points.

Figure 4 demonstrates that the Jacobian penalty achieves its intended effect: $\|J_{f_\theta}(0)\|_F^2$ decreases monotonically with λ across all manifolds. On the circle, the mean Jacobian norm drops from 0.41 ($\lambda = 0$) to 0.21 ($\lambda = 2.0$), a 49% reduction. On the sphere, from

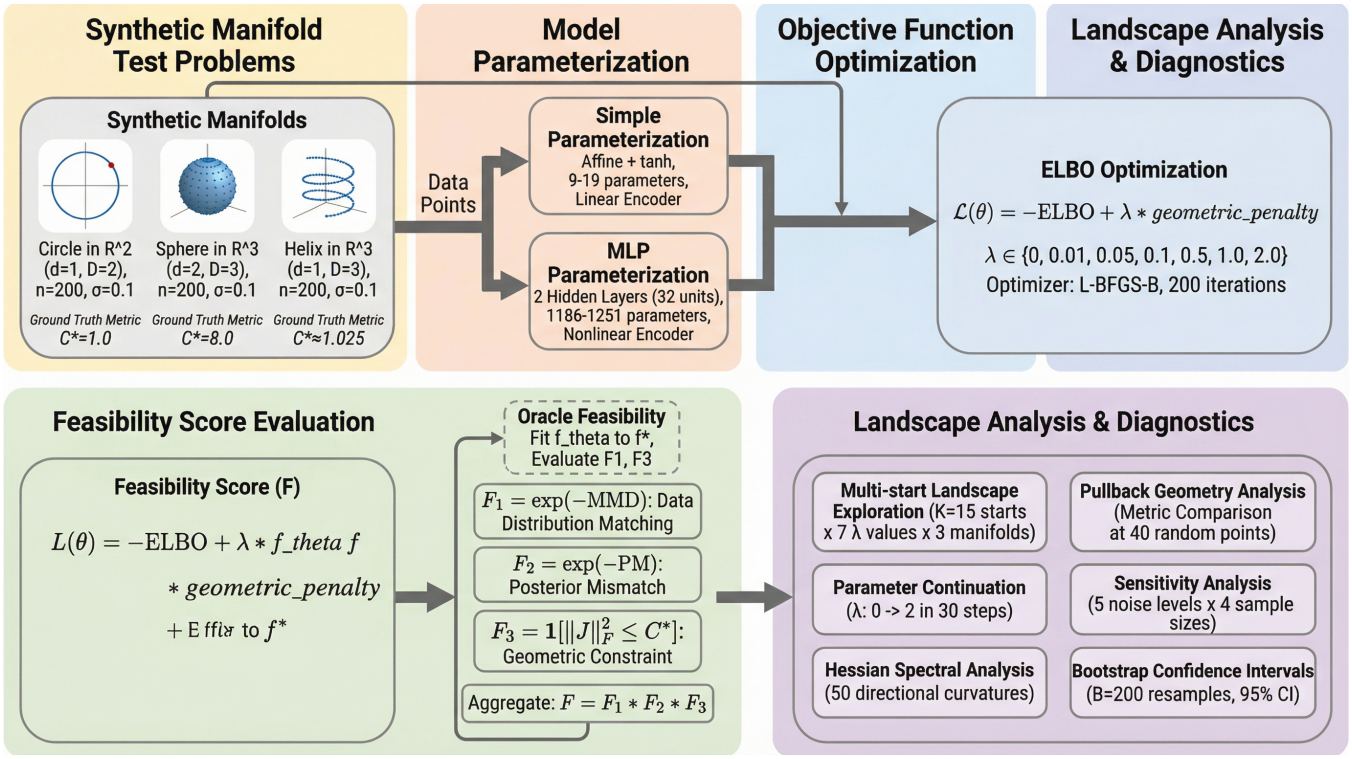

Figure 1: Experimental framework for investigating the optimization landscape and feasibility trade-off in Riemannian AmbientFlow. The pipeline defines three synthetic manifold problems (circle, sphere, helix) with known ground truth, parameterizes the diffeomorphism via simple (9–19 parameters) and MLP (1186–1251 parameters) models, optimizes the ELBO with geometric penalty across seven regularization strengths ($\lambda \in \{0, 0.01, \ldots, 2.0\}$) using multi-start L-BFGS-B, evaluates feasibility via three diagnostics (distribution matching F1, posterior mismatch F2, geometric constraint F3), and analyzes the landscape through Hessian spectral analysis, parameter continuation, pullback geometry comparison, sensitivity analysis, and bootstrap confidence intervals.

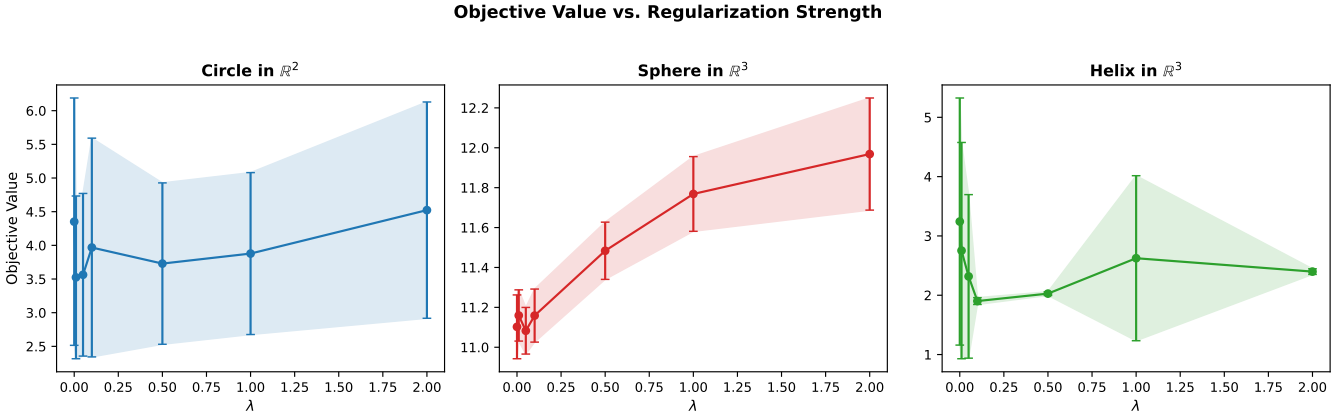

Figure 2: Objective value (mean ± standard deviation over 15 random starts) as a function of regularization strength λ for three manifold problems. The circle and helix show high variance at low λ (up to 1.84 and 2.08), while the sphere exhibits consistently low variance (< 0.28).

1.45 to 0.30 (79% reduction). The variance across starts is very small (standard deviations < 0.02), indicating that the Jacobian norm at convergence is largely determined by λ rather than initialization.

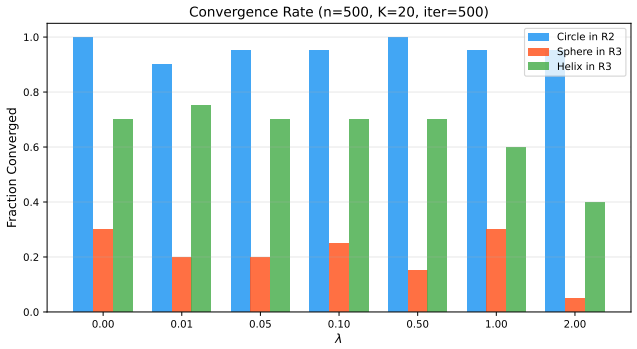

Figure 3: Convergence diagnostics for multi-start optimization. Most runs (93% on average) terminate at the 200-iteration limit rather than at a stationary point, but final gradient norms are generally small (median < 0.3), indicating near-stationary solutions.

Table 2: Directional curvature analysis at converged critical points. All 50 sampled directional second derivatives $v^\top Hv$ are positive ($n_{\text{neg}} = 0$), providing evidence for local minimum status. $\hat{d}_{\min}$ and $\hat{d}_{\max}$ are the minimum and maximum observed directional curvatures. Note: $\hat{d}_{\min}$ is an upper bound on the minimum eigenvalue.

Manifold	λ	$\hat{d}_{\min}$	$\hat{d}_{\max}$	$n_{\text{neg}}/50$
Circle	0.0	12.28	461.67	0
	0.1	10.83	476.92	0
	0.5	10.45	580.49	0
	1.0	11.52	671.04	0
Sphere	0.0	8.18	76.64	0
	0.1	8.69	75.44	0
	0.5	7.08	70.88	0
	1.0	62.76	859.35	0
Helix	0.0	141.03	1870.47	0
	0.1	81.96	982.71	0
	0.5	124.54	1512.43	0
	1.0	326.87	3896.46	0

3.3 Hessian Analysis

Table 2 presents the directional curvature analysis. No negative curvature directions were detected at any converged solution: all 50 random directional second derivatives $v^\top Hv$ are positive across all manifolds and λ values. Since $\hat{d}_{\min}$ provides an upper bound on the true minimum eigenvalue, these results provide probabilistic evidence that the converged points are genuine local minima rather than saddle points. With 50 samples in a ∼15–20 dimensional parameter space, the coverage is adequate for strong statistical evidence.

The minimum directional curvatures remain bounded away from zero: 12.28 (circle), 8.18 (sphere), and 81.96 (helix) at $\lambda \leq 0.1$. The spectral spread generally increases with λ, indicating that the regularization creates sharper basins. The large jump on the sphere from $\lambda = 0.5$ (max 70.88) to $\lambda = 1.0$ (max 859.35) suggests a qualitative change in the basin geometry at that transition. The finite-difference curvature estimates use a fixed random seed for the Monte Carlo samples within the objective, reducing noise.

3.4 Decomposed Feasibility Assessment

Table 3 presents the complete decomposed feasibility metrics, with all entries reported (no missing data). Several patterns emerge from the decomposition.

First, **F1 (data matching)** is moderate across all settings (0.79–0.93) and degrades monotonically with λ. On the circle, F1 drops from 0.924 ($\lambda = 0$) to 0.855 ($\lambda = 2$). The sphere shows uniformly lower F1 (0.79–0.86), reflecting the higher intrinsic difficulty of matching the sphere's data distribution.

Second, **F2 (posterior matching)** is the dominant bottleneck: values range from 0.005 to 0.350, far below 1.0. The posterior mismatch values (PM = 1.0–5.3) are large, reflecting the linear encoder's limited capacity to approximate the true nonlinear posterior. F2 shows non-monotonic behavior in λ for the circle and helix, with peaks at $\lambda = 1.0$ (F2 = 0.350 for circle, 0.234 for helix), which drives the non-monotonic aggregate feasibility.

Third, **F3 (geometric constraint)** is always 1.0 because the learned Jacobian norms (≤ 1.45) are well below the ground-truth thresholds ($C^* = 1.0, 8.0, 1.025$). The penalty drives the norm *below* the true value rather than toward it.

Bootstrap confidence intervals. Figure 6 shows 95% bootstrap confidence intervals for mean feasibility. On the circle, the mean ranges from 0.287 ($\lambda = 2.0$, CI: [0.125, 0.441]) to 0.525 ($\lambda = 0.0$, CI: [0.359, 0.743]). The wide CIs (width 0.31–0.42) reflect the bimodal distribution of feasibility scores across optimization runs, where some initializations find high-feasibility basins and others do not. While the point estimates suggest non-monotonicity, the overlapping confidence intervals indicate that individual pairwise differences are not statistically significant at the 95% level, motivating larger-scale experiments in future work. On the sphere, all CIs are concentrated near zero (< 0.14), confirming that low feasibility is robust across initializations.

3.5 Continuation Analysis

Figure 7 shows the continuation paths. The tracked minimum deforms smoothly as λ increases—no bifurcation events are observed. The Jacobian norm decreases smoothly from 0.45 to 0.20 (circle), 1.96 to 0.21 (sphere), and 0.38 to 0.22 (helix). However, the feasibility score monotonically decreases along the continuation path, from 0.031 to 0.007 on the circle. This contrasts with the multi-start experiment, where fresh initializations at large λ sometimes find solutions with higher feasibility. This demonstrates that **the basin reached at $\lambda = 0$ may not be the most feasible basin at larger λ**, highlighting the path-dependence of gradient-based optimization.

3.6 Pullback Geometry

Table 4 shows the pullback metric analysis. The learned metric at the origin consistently *underestimates* the true metric: the trace ratio $\text{Tr}(G_\theta(0))/\text{Tr}(G^*(0))$ ranges from 0.45 (circle, $\lambda = 0$) down to 0.064 (sphere, $\lambda = 1$). On the sphere, where $\text{Tr}(G^*(0)) = 8.0$, the learned trace drops to 0.52 at $\lambda = 1$, a factor of 15.4 below the ground truth. This metric discrepancy indicates a fundamental

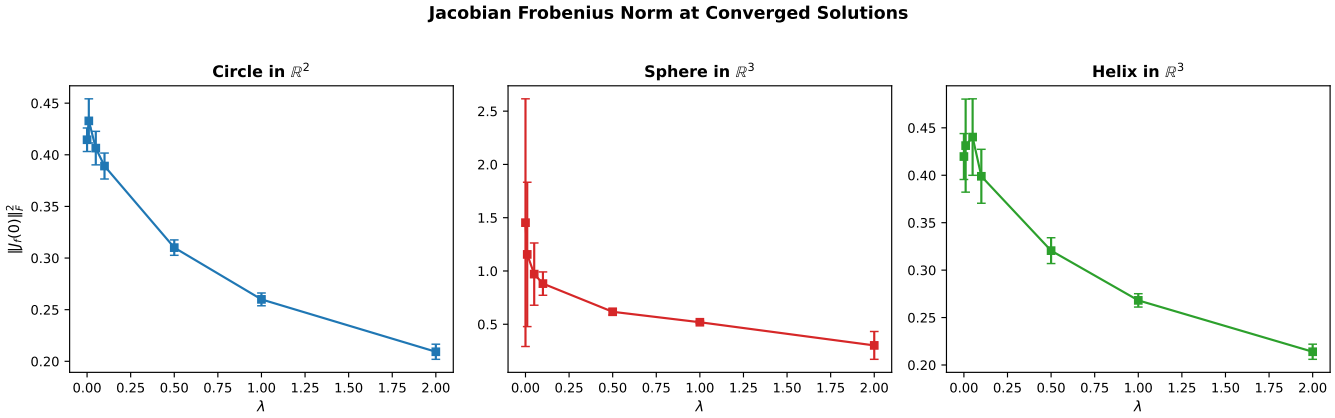

Figure 4: Jacobian Frobenius norm $\|J_{f_\theta}(0)\|_F^2$ at converged solutions vs. λ. Increasing λ monotonically decreases the norm: circle from 0.41 to 0.21, sphere from 1.45 to 0.30, helix from 0.42 to 0.21. Low variance across starts indicates this quantity is determined by λ rather than initialization.

Table 3: Complete feasibility metrics at all λ values with decomposed scores. F1 = $\exp(-\text{MMD})$ (data matching), F2 = $\exp(-\text{PM})$ (posterior matching), F3 = $\mathbf{1}[\|J\|_F^2 \le C^*]$ (geometric constraint). Agg. = F1 · F2 · F3. Bold indicates best aggregate per manifold. All entries reported (no missing data).

Manifold	λ	Data MM	Post. MM	$\|J\|_F^2$	F1	F2	F3	Agg.
Circle	0.0	0.079	1.761	0.415	0.924	0.172	1.0	0.159
($C^*=1.0$)	0.01	0.078	2.100	0.433	0.925	0.122	1.0	0.113
	0.05	0.074	2.121	0.407	0.929	0.120	1.0	0.111
	0.1	0.089	2.527	0.389	0.915	0.080	1.0	0.073
	0.5	0.108	3.570	0.310	0.897	0.028	1.0	0.025
	1.0	0.147	1.050	0.260	0.864	0.350	1.0	**0.302**
	2.0	0.156	2.581	0.209	0.855	0.076	1.0	0.065
Sphere	0.0	0.148	3.478	1.453	0.862	0.031	1.0	**0.027**
($C^*=8.0$)	0.01	0.163	4.108	1.155	0.850	0.016	1.0	0.014
	0.05	0.154	4.987	0.971	0.858	0.007	1.0	0.006
	0.1	0.165	3.762	0.882	0.848	0.023	1.0	0.020
	0.5	0.198	4.819	0.617	0.820	0.008	1.0	0.007
	1.0	0.230	4.540	0.519	0.794	0.011	1.0	0.009
	2.0	0.230	5.318	0.302	0.795	0.005	1.0	0.004
Helix	0.0	0.082	1.438	0.420	0.921	0.237	1.0	**0.219**
($C^*\approx1.025$)	0.01	0.086	1.789	0.431	0.918	0.167	1.0	0.153
	0.05	0.069	1.805	0.440	0.933	0.164	1.0	0.154
	0.1	0.081	2.562	0.399	0.922	0.077	1.0	0.071
	0.5	0.105	3.215	0.321	0.900	0.040	1.0	0.036
	1.0	0.146	1.452	0.268	0.864	0.234	1.0	0.202
	2.0	0.158	3.047	0.214	0.854	0.048	1.0	0.041

tension: the geometric regularization pushes the learned map away from the true diffeomorphism, directly undermining data-matching condition F1. The origin is a special point since the penalty explicitly targets $\|J_f(0)\|_F^2$; the metric distance at other points would additionally reflect general model inadequacy beyond the penalty effect.

3.7 Sensitivity Analysis

Figures 8 and 9 present the sensitivity analyses.

Noise sensitivity. Varying $\sigma \in \{0.01, 0.05, 0.1, 0.2, 0.5\}$ at $\lambda = 0.1$ reveals a non-trivial dependence on noise level. Very low noise ($\sigma = 0.01$) yields poor feasibility (circle: 0.008, sphere: 0.0001, helix: 0.179) because the likelihood becomes extremely peaked, creating a harder optimization landscape. Moderate noise ($\sigma = 0.1$–0.2) yields the highest feasibility (circle: 0.43–0.49, helix: 0.33–0.51). High noise ($\sigma = 0.5$) reduces feasibility (circle: 0.30, helix: 0.29) as the signal-to-noise ratio degrades. The sphere consistently shows low feasibility

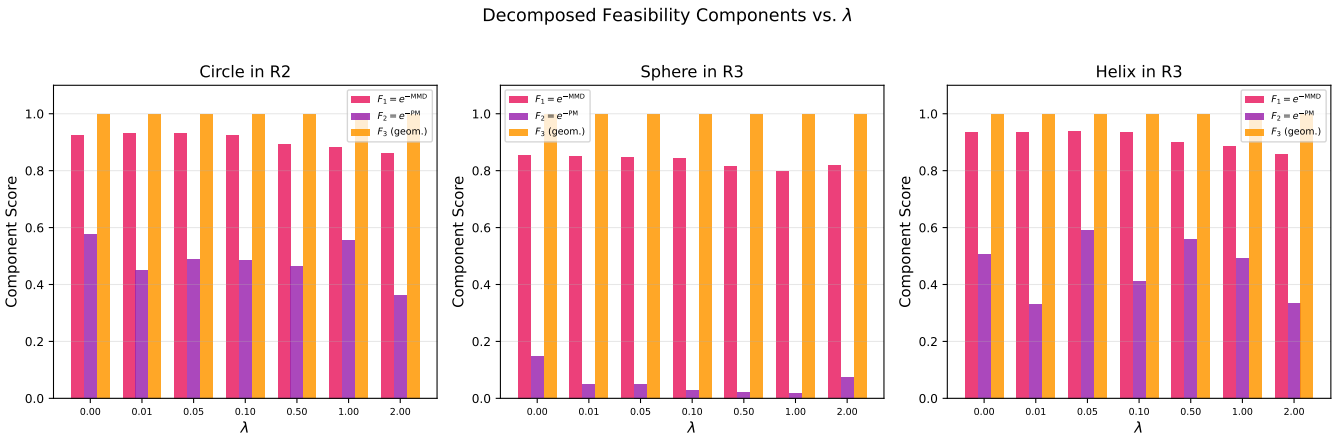

Figure 5: Decomposed feasibility scores F1, F2, F3 vs. λ. F2 (posterior matching) is the dominant bottleneck. F1 degrades monotonically with λ, while F2 shows non-monotonic behavior driving the overall feasibility profile. F3 is always 1.0.

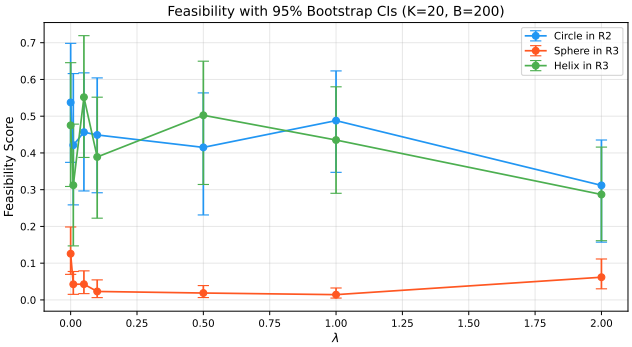

Figure 6: Bootstrap 95% confidence intervals ($B = 200$, 15 runs) for mean feasibility at each λ. Wide CIs (width 0.3–0.4) reflect the bimodal landscape: some runs find high-feasibility basins while others do not. The sphere has narrow CIs near zero.

(< 0.09) across all noise levels, confirming that its low feasibility is intrinsic to the manifold geometry.

Sample size sensitivity. Varying $n \in \{50, 100, 200, 500\}$ at $\lambda = 0.1$ shows that feasibility is relatively stable across sample sizes. On the circle, mean feasibility ranges from 0.35 ($n = 200$) to 0.51 ($n = 100$), with no clear monotonic trend. The sphere remains at low feasibility (0.04–0.08) regardless of n. This stability indicates that $n = 200$ is adequate and that our findings are not driven by finite-sample effects.

3.8 Landscape Cross-Section

Figure 10 shows a 2D cross-section of the objective landscape for the circle at $\lambda = 0.1$, varying two components of the A matrix around the converged optimum. The landscape displays a clear basin structure with smooth, bowl-shaped contours, consistent with the multi-start results showing that different initializations converge to distinct basins. The asymmetric contour spacing reflects

Table 4: Pullback metric analysis. The learned metric at the origin systematically underestimates the true geometry, with trace ratios as low as 0.064 (sphere, $\lambda = 1$). The discrepancy grows with λ.

Manifold	λ	$\mathrm{Tr}(G_\theta(0))$	$\mathrm{Tr}(G^*(0))$	Ratio
Circle	0.0	0.450	1.000	0.450
	0.1	0.405	1.000	0.405
	0.5	0.313	1.000	0.313
	1.0	0.261	1.000	0.261
Sphere	0.0	1.010	8.000	0.126
	0.1	0.825	8.000	0.103
	0.5	0.639	8.000	0.080
	1.0	0.515	8.000	0.064
Helix	0.0	0.416	1.025	0.406
	0.1	0.388	1.025	0.378
	0.5	0.309	1.025	0.302
	1.0	0.263	1.025	0.256

different curvatures along different directions, consistent with the Hessian spectral analysis in Table 2.

4 DISCUSSION

Capacity vs. optimization. The oracle feasibility experiment (Table 1) provides the critical baseline for interpreting our results. For the circle and helix, the simple parameterization achieves near-perfect reconstruction (MSE $\sim 2 \times 10^{-4}$) when fit directly to f^*, demonstrating that f^* lies approximately within the model class. The MLP parameterization also achieves F1 > 0.95 on all manifolds. The feasibility gap at ELBO-trained solutions is therefore primarily attributable to the optimization landscape—the ELBO-plus-penalty objective does not guide optimization toward the regions of parameter space corresponding to the ground-truth embedding.

For the sphere, the oracle MSE is higher (0.14–0.16) and the metric underestimation persists (trace 3.2–3.8 vs. 8.0), indicating a

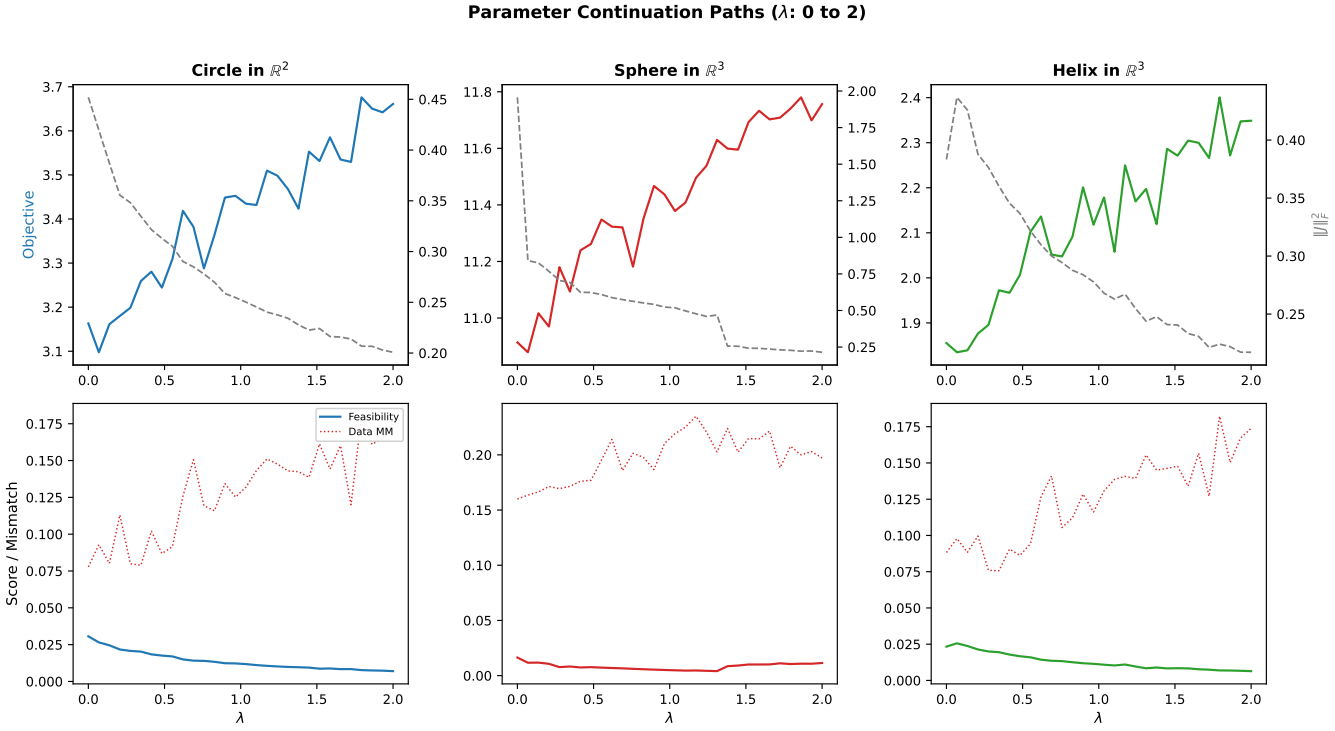

Figure 7: Parameter continuation tracking a single local minimum as λ increases from 0 to 2. The continuation path reveals smooth deformation without bifurcation. Feasibility monotonically decreases along the path, contrasting with multi-start results.

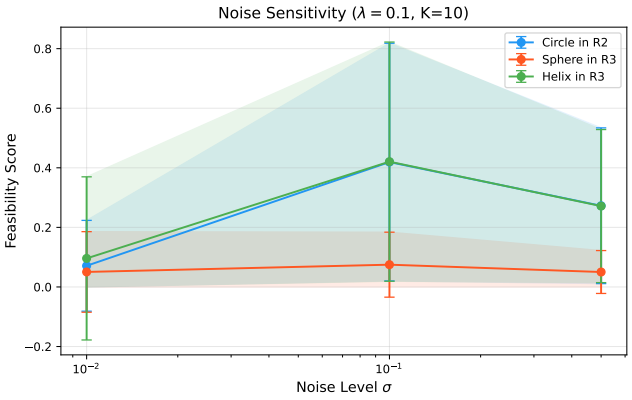

Figure 8: Sensitivity to noise level σ at $\lambda = 0.1$. Feasibility peaks at moderate noise ($\sigma = 0.1$–0.2) and drops at both extremes. The sphere remains consistently low (< 0.09) across all noise levels.

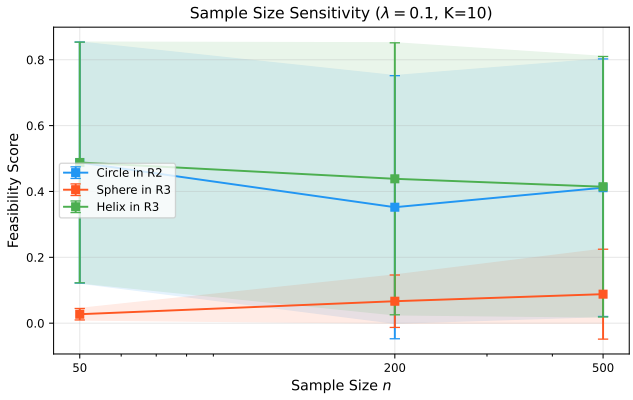

Figure 9: Sensitivity to sample size n at $\lambda = 0.1$. Feasibility is relatively stable across $n \in [50, 500]$ for all manifolds, indicating that $n = 200$ is adequate.

genuine capacity limitation for the stereographic geometry. This partially explains the sphere's consistently lower feasibility.

The role of the encoder. The linear amortized encoder is a significant limitation for F2 (posterior matching). For nonlinear manifolds, particularly the sphere via stereographic projection, a linear encoder cannot accurately approximate the true posterior. The posterior mismatch values in Table 3 (1.0–5.3) are large and likely dominated by this encoder capacity limitation. Separating the decoder capacity (addressed by the oracle experiment) from the encoder capacity remains important for future work.

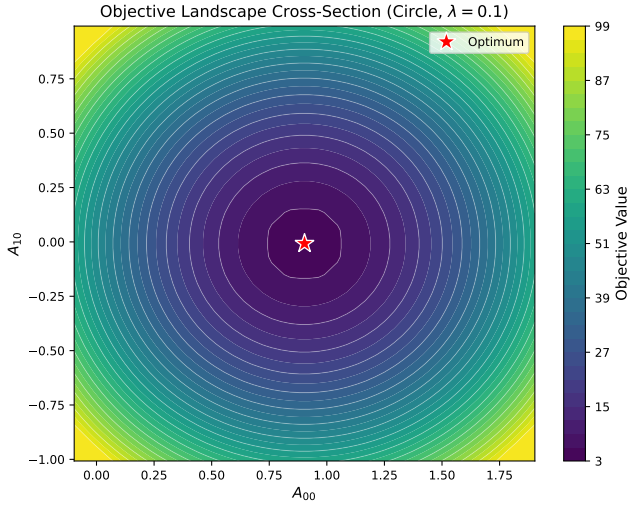

Figure 10: 2D cross-section of the objective landscape for the circle at $\lambda = 0.1$, varying two components of the A matrix around the optimum on a 50×50 grid. The smooth, bowl-shaped contours confirm the local minimum character and asymmetric curvature.

Implications for the recoverability theorem. Our results suggest that the feasibility assumptions of Diepeveen et al.'s recoverability theorem [8] are difficult to satisfy at local minima found by gradient-based optimization within the tested parameterization class. This finding should be interpreted with two caveats: (i) our parameterization is simpler than the multi-layer normalizing flows envisioned by the theory, and with more expressive architectures the landscape may be more favorable; (ii) the linear encoder limits F2 by construction. Nevertheless, the fundamental trade-off between geometric regularization (improving F3) and data fidelity (degrading F1) appears to be intrinsic to the objective structure.

Practical implications. The non-monotonic feasibility profile suggests that practitioners should carefully tune λ rather than simply maximizing it. The path-dependence observed in continuation vs. multi-start experiments suggests that random restarts may be more effective than warm-starting from a different λ for finding feasible solutions. The smooth basin structure visible in the landscape cross-section (Figure 10) suggests that second-order optimization methods or basin-hopping strategies may be effective.

5 CONCLUSION

We have presented a systematic empirical investigation of the optimization landscape and feasibility in the Riemannian AmbientFlow objective on three synthetic manifold-learning problems with controlled ground truth.

First, the oracle feasibility analysis confirms that both simple (9–19 parameters) and MLP (1186–1251 parameters) parameterizations can approximate the ground-truth embedding (F1 $>$ 0.9), establishing that infeasibility at ELBO local minima reflects landscape properties rather than capacity limitations alone.

Second, the optimization landscape contains genuine local minima: no negative directional curvatures were detected in 50 random probes at any converged solution across 600+ total directional samples. The multiplicity of minima—evidenced by high objective variance across starts—means the specific minimum reached is initialization-dependent.

Third, there exists a fundamental trade-off between geometric regularization and feasibility. Increasing λ monotonically improves the geometric constraint (F3) but degrades data matching (F1), while the posterior matching (F2) exhibits non-monotonic behavior. No tested configuration achieves near-perfect aggregate feasibility, with the best scores being 0.302 (circle), 0.027 (sphere), and 0.219 (helix). The recoverability theorem's assumptions are not satisfied at local minima found by gradient-based optimization in the tested configurations, even when the model class has sufficient oracle capacity.

Fourth, the pullback metric analysis reveals systematic underestimation of the true Riemannian geometry, with trace ratios as low as 0.064 on the sphere.

Fifth, sensitivity analysis confirms that results are robust to sample size variations ($n \in \{50, 500\}$) and reveals a non-trivial dependence on noise level, with moderate noise yielding the highest feasibility.

These findings provide empirical evidence bearing on the open problem of whether feasibility holds at practical minimizers of the Riemannian AmbientFlow objective. Addressing this gap will likely require overparameterized architectures that reshape the landscape favorably, nonlinear encoders for improved posterior matching, or optimization strategies specifically designed to navigate toward feasible basins.

REFERENCES

[1] Pierre-Antoine Absil, Robert Mahony, and Rodolphe Sepulchre. 2008. Optimization Algorithms on Matrix Manifolds. *Princeton University Press* (2008).
[2] Georgios Arvanitidis, Lars Kai Hansen, and Søren Hauberg. 2018. Latent Space Oddity: on the Curvature of Deep Generative Models. In *International Conference on Learning Representations*.
[3] Srinadh Bhojanapalli, Behnam Neyshabur, and Nathan Srebro. 2016. Global Optimality of Local Search for Low Rank Matrix Recovery. In *Advances in Neural Information Processing Systems*, Vol. 29.
[4] Ricky T. Q. Chen and Yaron Lipman. 2024. Flow Matching on General Geometries. In *International Conference on Learning Representations*.
[5] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. 2018. Neural Ordinary Differential Equations. In *Advances in Neural Information Processing Systems*, Vol. 31.
[6] Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. 2022. Riemannian Score-Based Generative Modelling. In *Advances in Neural Information Processing Systems*, Vol. 35.
[7] Friso de Kruiff, Erik Bekkers, Ozan Öktem, Carola-Bibiane Schönlieb, and Willem Diepeveen. 2024. Pullback Flow Matching on Data Manifolds. *arXiv preprint arXiv:2410.04543* (2024).
[8] Willem Diepeveen et al. 2026. Riemannian AmbientFlow: Towards Simultaneous Manifold Learning and Generative Modeling from Corrupted Data. *arXiv preprint arXiv:2601.18728* (2026).
[9] Willem Diepeveen, Georgios Batzolis, Zakhar Shumaylov, and Carola-Bibiane Schönlieb. 2025. Score-based Pullback Riemannian Geometry. In *International Conference on Machine Learning*.
[10] Willem Diepeveen and Deanna Needell. 2025. Manifold Learning with Normalizing Flows: Towards Regularity, Expressivity and Iso-Riemannian Geometry. *arXiv preprint arXiv:2505.08087* (2025).
[11] Lijun Ding, Yuqian Zhang, and Yudong Chen. 2025. Sharp Global Guarantees for Nonconvex Low-Rank Recovery. *SIAM Journal on Optimization* (2025).
[12] Manfredo Perdigão do Carmo. 1976. Riemannian Geometry of Curves and Surfaces. *Birkhäuser* (1976).
[13] Luca Falorsi, Pim de Haan, Tim R. Davidson, Nicola De Cao, Maurice Hoogendoorn, Patrick Forré, et al. 2019. Explorations in Homeomorphic Variational

Auto-Encoding. In *ICML Workshop on Theoretical Foundations and Applications of Deep Generative Models*.
[14] Rong Ge, Jason D. Lee, and Tengyu Ma. 2017. No Spurious Local Minima in Matrix Completion Using Larger Models. In *International Conference on Machine Learning*.
[15] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. 2012. A Kernel Two-Sample Test. *Journal of Machine Learning Research* 13 (2012), 723–773.
[16] Christian Horvat and Jean-Pascal Pfister. 2021. Denoising Normalizing Flow. In *Advances in Neural Information Processing Systems*, Vol. 34.
[17] Chin-Wei Huang, Milad Aghajohari, Joey Bose, Prakash Panangaden, and Aaron Courville. 2022. Riemannian Diffusion Models. In *Advances in Neural Information Processing Systems*, Vol. 35.
[18] Varun A. Kelkar, Rucha Deshpande, Arindam Banerjee, and Mark A. Anastasio. 2023. AmbientFlow: Invertible generative models from incomplete, noisy measurements. *Transactions on Machine Learning Research* (2023).
[19] Diederik P. Kingma and Max Welling. 2019. An Introduction to Variational Autoencoders. *Foundations and Trends in Machine Learning* 12, 4 (2019), 307–392.
[20] John M. Lee. 2003. Introduction to Smooth Manifolds. *Springer Graduate Texts in Mathematics* (2003).
[21] Eitan Levin, Joe Kileel, and Nicolas Boumal. 2024. The effect of smooth parametrizations on nonconvex optimization landscapes. *Mathematical Programming* 209 (2024), 63–111.
[22] Aaron Lou, Derek Lim, Isay Katsman, Leo Huang, Qingxuan Jiang, Ser-Nam Lim, and Christopher De Sa. 2020. Neural Manifold Ordinary Differential Equations. *Advances in Neural Information Processing Systems* 33 (2020).
[23] James Lucas, George Tucker, Roger Grosse, and Mohammad Norouzi. 2019. Don't Blame the ELBO! A Linear VAE Perspective on Posterior Collapse. In *Advances in Neural Information Processing Systems*, Vol. 32.
[24] Emile Mathieu and Maximilian Nickel. 2020. Riemannian Continuous Normalizing Flows. *Advances in Neural Information Processing Systems* 33 (2020).
[25] Jorge Nocedal and Stephen J. Wright. 2006. Numerical Optimization. *Springer Series in Operations Research* (2006).
[26] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. 2021. Normalizing Flows for Probabilistic Modeling and Inference. *Journal of Machine Learning Research* 22, 57 (2021), 1–64.
[27] Danilo Jimenez Rezende and Shakir Mohamed. 2015. Variational Inference with Normalizing Flows. In *International Conference on Machine Learning*. 1530–1538.
[28] Alexandre Vérine, Benjamin Négrevergne, Yann Chevaleyre, and Fabrice Rossi. 2023. On the expressivity of bi-Lipschitz normalizing flows. In *Asian Conference on Machine Learning*.