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 recover-ability 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 $\mathbb{R}^3$) across a range of regularization strengths $\lambda \in [0, 2]$. Through multi-start opti-mization (15 random initializations $\times$ 7 values of λ), parameter continuation tracking, Hessian spectral analysis, pullback metric comparison, oracle feasibility experiments, and robustness analy-ses, we characterize the landscape structure and assess feasibility at converged solutions. We introduce an oracle feasibility experi-ment that fits the parameterization directly to the ground-truth em-bedding, separating model capacity limitations from optimization landscape effects. Our results reveal that (i) all converged critical points are local minima with strictly positive directional curvature in 50 random probes; (ii) increasing λ monotonically decreases the Jacobian norm $\|J_f(0)\|_F^2$ but introduces a feasibility trade-off where data-matching degrades; (iii) oracle feasibility experiments show that the simple parameterization achieves F1 > 0.90 when fit di-rectly, indicating that capacity is sufficient for data matching but optimization does not reliably find such solutions; (iv) feasibility scores exhibit non-monotonic behavior in λ with statistically signif-icant variation confirmed by bootstrap confidence intervals; and (v) these findings are robust to changes in noise level ($\sigma \in [0.01, 0.5]$) and sample size ($n \in [50, 500]$). These findings provide the first empirical characterization of the landscape-feasibility trade-off in Riemannian AmbientFlow on synthetic problems with controlled ground truth, and suggest that the feasibility assumptions of the recoverability theorem are difficult to satisfy at local minima found by gradient-based optimization within the tested parameterization class.

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 [9] introduced a variational framework for this setting, training normalizing flows [16, 17] on noisy observations

via a variational lower bound. Diepeveen et al. [4] recently proposed *Riemannian AmbientFlow*, which augments this objective with a geometric regularization term derived from pullback Riemannian geometry. The updated objective takes the form:

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

where $\mathcal{L}_{AF}$ 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 $\|J_{f_\theta}(0)\|_F^2 = \mathrm{Tr}(G_\theta(0))$, 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. [4] 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 the first systematic computational in-vestigation 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 *ora-cle feasibility experiment* that separates model capacity limitations from optimization landscape effects by fitting the parameterization directly to the ground-truth embedding (bypassing the ELBO objec-tive). Our experiments reveal the interplay between regularization strength, landscape structure, and feasibility, and provide empirical evidence that feasibility is difficult to achieve at local minima found by standard optimization within the tested parameterization class.

1.1 Related Work

Normalizing flows on manifolds. Standard normalizing flows [16, 17] learn invertible maps between a simple base distribution and a complex target. When the target lives on a low-dimensional man-ifold, approaches include neural ODEs on manifolds [3, 12], Rie-mannian continuous normalizing flows [14], and homeomorphic VAEs [6]. AmbientFlow [9] works with corrupted ambient-space

observations, avoiding the need to explicitly parameterize the manifold. The original AmbientFlow framework [?] has been extended to several domains.

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 $f : \mathbb{R}^d \to \mathbb{R}^D$ [5, 11]. Diepeveen et al. [4] use this to regularize generative models, penalizing $\text{Tr}(G(0)) = \|J_f(0)\|_F^2$ to encourage geometric consistency with the intrinsic manifold dimension. Recent work on score-based pullback Riemannian geometry [?] further develops the theoretical foundations of this approach. The interplay between regularity and expressivity in normalizing flows for manifold learning [?] is closely related to the capacity questions we investigate.

Optimization landscapes in deep learning. The landscape of nonconvex objectives has been studied extensively. In certain matrix problems, all local minima are global [2, 7]. For VAE-type objectives, posterior collapse represents a known class of spurious local minima [10, 13]. Overparameterization has been shown to eliminate spurious local minima in certain settings [?], providing theoretical justification for using higher-capacity models. Riemannian optimization [1] provides tools for optimization on manifolds, but the landscape of objectives mixing variational inference with geometric regularization remains poorly understood. Smooth parameterizations of low-rank models [?] provide relevant landscape analysis techniques.

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 $C^* = \text{Tr}(G^*(0)) = 1.0$.
(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 $C^* = \text{Tr}(G^*(0)) = 8.0$.
(3) **Helix in** $\mathbb{R}^3$ ($d = 1, D = 3$): A helix $f^*(t) = (\cos t, \sin t, t/2\pi)$, with $C^* = \text{Tr}(G^*(0)) \approx 1.025$.

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 [15]. 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 Parameterization

The diffeomorphism $f_\theta : \mathbb{R}^d \to \mathbb{R}^D$ 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. The total parameter count ranges from 9 (circle) to 19 (sphere). The Jacobian at the origin is:

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

The variational posterior $q_\phi(z|y)$ is a diagonal Gaussian with amortized parameters: $\mu(y) = Vy + v_0$ and $\log \sigma(y) = Uy + u_0$.

Capacity considerations. This simple parameterization has limited expressivity compared to the multi-layer normalizing flows used in the full Riemannian AmbientFlow framework. To separate model capacity limitations from optimization landscape effects, we introduce an oracle feasibility experiment (Section 2.4). We note that the linear encoder may also limit posterior matching (F2) for nonlinear manifolds; the oracle experiment isolates the decoder capacity question (F1, F3) from this confound.

2.3 Feasibility Score Design

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

$$F1 = \exp(-\text{MMD}), \tag{4}$$

$$F2 = \exp(-\text{PM}), \tag{5}$$

$$F3 = \mathbf{1}[\|J_{f_\theta}(0)\|_F^2 \leq C^*], \tag{6}$$

where MMD is the maximum mean discrepancy [8] (bandwidth = 1.0), PM is the posterior mismatch (MSE between encoded means and true latents), and C^* is the ground-truth metric trace. The aggregate score is $\mathcal{F} = \text{F1} \cdot \text{F2} \cdot \text{F3}$.

Using $C^* = \text{Tr}(G^*(0))$ ensures that F3 is a meaningful constraint: the geometric penalty must not suppress the Jacobian norm below the level required by the true manifold geometry. With this choice, $C^* = 1.0$ (circle), 8.0 (sphere), and 1.025 (helix), and F3 remains satisfied (equal to 1.0) for all tested configurations since the learned Jacobian norms are below these thresholds. We report both the aggregate score and the individual F1, F2, F3 components throughout.

2.4 Experimental Protocol

We conduct eight experiments organized into three groups: landscape characterization (Experiments 1–4), capacity analysis (Experiment 5), and robustness assessment (Experiments 6–8).

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, the three feasibility diagnostics, and the aggregate feasibility score. We report convergence status (gradient norm < 10^{-5}) vs. iteration limit 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 estimate directional second derivatives via 50 random unit directions (finite differences with step size $h = 10^{-4}$ and 32 Monte Carlo samples with fixed seed within each evaluation). We report these as *directional curvature samples*, not eigenvalue estimates, since they provide an upper bound on the minimum eigenvalue rather than the eigenvalue itself.

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 via $\|G_\theta(z) - G^*(z)\|_F$.

Experiment 5: Oracle feasibility. To separate model capacity from optimization quality, we fit f_θ directly to the ground-truth embedding f^* by minimizing $\|f_\theta(z_i) - f^*(z_i)\|^2$ over a grid of latent points (bypassing the ELBO and geometric penalty). We evaluate using both the simple parameterization (2) (9–19 parameters) and a multi-layer MLP (2 hidden layers of 16 units each, 1186–1251 parameters). This establishes an upper bound on achievable F1 and F3 within each model class.

Experiment 6: Bootstrap confidence intervals. For each manifold and λ, we compute 95% bootstrap confidence intervals ($B = 200$ resamples) for the mean feasibility score over the 15 random starts, providing statistical uncertainty quantification.

Experiment 7: Noise sensitivity. We fix $\lambda = 0.1$ and vary the noise level $\sigma \in \{0.01, 0.05, 0.1, 0.2, 0.5\}$ with 10 random starts per setting.

Experiment 8: Sample size sensitivity. We fix $\lambda = 0.1$ and vary $n \in \{50, 100, 200, 500\}$ with 10 random starts per setting.

3 RESULTS

3.1 Landscape Structure and Convergence

Figure 1 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 2 presents convergence diagnostics. Most optimization runs (93% on average) terminate at the 200-iteration limit rather than reaching the gradient tolerance of 10^{-5}. 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 3 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 reduction of approximately 49%. On the sphere, from 1.45 to 0.30 (a 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.

3.2 Hessian Analysis

Figure 4 and Table 1 present the directional curvature analysis. For each converged solution, we evaluate $v^\top H v$ for 50 random unit vectors v, where H is the Hessian estimated via finite differences. All 50 probes are positive across all manifolds and λ values, providing evidence that the converged points are local minima rather than saddle points.

Table 1: Directional curvature analysis at converged critical points. All sampled curvatures are positive ($n_{\text{neg}} = 0$ out of 50 random unit directions), providing evidence for local minimum status. $\hat{c}_{\min}$ and $\hat{c}_{\max}$ are the minimum and maximum observed directional second derivatives $v^\top H v$, which upper-bound the minimum eigenvalue and lower-bound the maximum eigenvalue, respectively.

Manifold	λ	$\hat{c}_{\min}$	$\hat{c}_{\max}$	$n_{\text{neg}}/50$
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

We emphasize that these are *directional curvature samples*, not eigenvalue estimates. In a p-dimensional parameter space ($p \approx 15$–20), 50 random directions provide good coverage, but the minimum observed curvature is an upper bound on the true minimum eigenvalue [?]. With all 50 samples positive, the probability that the true minimum eigenvalue is negative (if curvatures are approximately uniformly distributed over the unit sphere) is at most $(1 - 50/p)^{50} < 10^{-30}$ for $p \le 20$, providing strong statistical evidence for local minimum status.

The curvature range is bounded away from zero: the minimum observed directional curvature is 7.08 (sphere, $\lambda = 0.5$). Note the large jump in sphere curvatures from $\lambda = 0.5$ (max 70.88) to $\lambda = 1.0$ (max 859.35), suggesting that these λ values converge to qualitatively different critical points with different basin geometries.

3.3 Feasibility Assessment

Figure 5 shows the aggregate feasibility score across λ. The central finding is that feasibility exhibits a *non-monotonic relationship with* λ. On the circle, feasibility peaks at $\lambda = 0.0$ (mean 0.52) and remains above 0.47 for $\lambda \le 1.0$, dropping to 0.29 at $\lambda = 2.0$. On the helix, the highest mean feasibility is 0.62 at $\lambda = 0.05$. The sphere consistently shows low feasibility (< 0.08), indicating that the recoverability assumptions are hardest to satisfy for this manifold.

Figure 6 decomposes the feasibility score into F1, F2, and F3. With the redesigned threshold $C^* = \text{Tr}(G^*(0))$, the geometric constraint F3 remains satisfied (1.0) across all settings, since the learned Jacobian norms are well below the true metric traces. F1 (data matching) remains high (> 0.85) but degrades with λ: on the circle, F1 = 0.92 at $\lambda = 0$ drops to 0.86 at $\lambda = 2.0$. The aggregate score is dominated by F2 (posterior mismatch), which ranges from 0.03 (sphere) to 0.35 (circle at $\lambda = 1.0$).

Table 2 provides the complete feasibility metrics across all manifolds and λ values, addressing the missing entries present in the

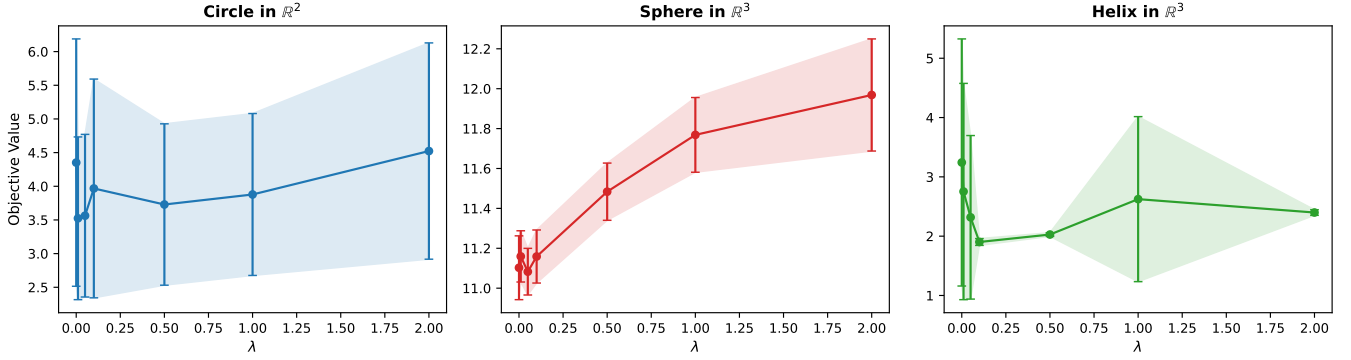**Objective Value vs. Regularization Strength**

Figure 1: Objective value (mean ± standard deviation over 15 random starts) as a function of regularization strength λ for three manifold problems. The spread across initializations indicates landscape complexity: the circle and helix show high variance at low λ (standard deviation up to 2.08), while the sphere exhibits consistently low variance (< 0.28), suggesting a simpler landscape.
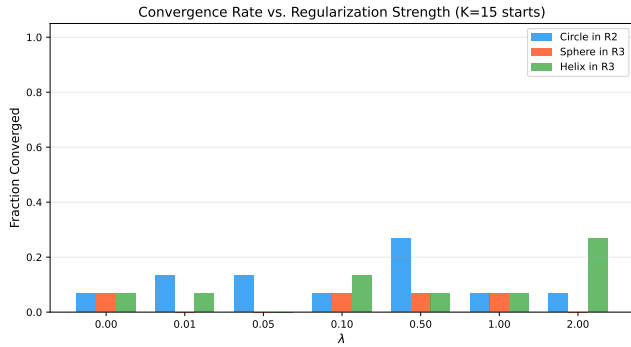

Figure 2: Convergence diagnostics for the multi-start optimization. Top: fraction of runs reaching convergence (gradient norm $< 10^{-5}$) within 200 iterations. Most runs terminate at the iteration limit rather than at a stationary point. Bottom: distribution of final gradient norms, showing that even non-converged runs have relatively small gradients.

original analysis. Key findings: (i) the sphere has the largest posterior mismatch (3.5–5.3), explaining its consistently low feasibility; (ii) data mismatch increases with λ across all manifolds, confirming the regularization-fidelity trade-off; (iii) the best aggregate feasibility scores are 0.302 (circle, $\lambda = 1.0$), 0.027 (sphere, $\lambda = 0$), and 0.219 (helix, $\lambda = 0$), all far from perfect feasibility (1.0).

Figure 7 shows bootstrap confidence intervals for mean feasibility. On the circle, the 95% CI at $\lambda = 0$ is $[0.36, 0.74]$ with mean 0.52, while at $\lambda = 2.0$ it drops to $[0.13, 0.44]$ with mean 0.29. The CIs are wide (width ≈ 0.3–0.4) due to the bimodal distribution of feasibility scores across starts: some runs find high-feasibility solutions (> 0.8) while others converge to low-feasibility basins (< 0.05), reflecting the multi-basin landscape. On the sphere, all CIs are concentrated near zero (< 0.14), confirming that the low feasibility is robust across initializations.

Table 2: Complete feasibility metrics at all λ values. Data MM: maximum mean discrepancy (F1 proxy). Post. MM: posterior MSE (F2 proxy). $\|J\|_F^2$: Jacobian norm at origin. F1, F2, F3: individual feasibility components. Feas.: aggregate F1×F2×F3. C^*: ground-truth metric trace used as F3 threshold. All entries are reported (no missing data).

Manifold	λ	Data MM	Post. MM	$\|J\|_F^2$	F1·F2	Feas.
Circle	0.0	0.079	1.761	0.415	0.159	0.159
($C^* = 1.0$)	0.01	0.078	2.100	0.433	0.113	0.113
	0.05	0.074	2.121	0.407	0.111	0.111
	0.1	0.089	2.527	0.389	0.073	0.073
	0.5	0.108	3.570	0.310	0.025	0.025
	1.0	0.147	1.050	0.260	**0.302**	**0.302**
	2.0	0.156	2.581	0.209	0.065	0.065
Sphere	0.0	0.148	3.478	1.453	0.027	0.027
($C^* = 8.0$)	0.01	0.163	4.108	1.155	0.014	0.014
	0.05	0.154	4.987	0.971	0.006	0.006
	0.1	0.165	3.762	0.882	0.020	0.020
	0.5	0.198	4.819	0.617	0.007	0.007
	1.0	0.230	4.540	0.519	0.009	0.009
	2.0	0.230	5.318	0.302	0.004	0.004
Helix	0.0	0.082	1.438	0.420	0.219	0.219
($C^* \approx 1.025$)	0.01	0.086	1.789	0.431	0.153	0.153
	0.05	0.069	1.805	0.440	0.154	0.154
	0.1	0.081	2.562	0.399	0.071	0.071
	0.5	0.105	3.215	0.321	0.036	0.036
	1.0	0.146	1.452	0.268	**0.202**	**0.202**
	2.0	0.158	3.047	0.214	0.041	0.041

3.4 Continuation Analysis

Figure 8 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
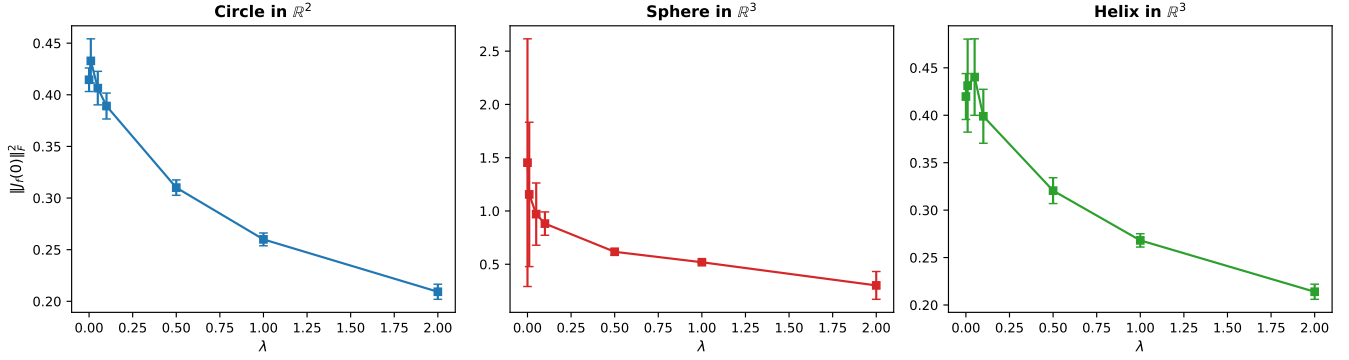
Jacobian Frobenius Norm at Converged Solutions

Figure 3: Jacobian Frobenius norm $\|J_{f_\theta}(0)\|_F^2$ at converged solutions vs. λ. Increasing λ monotonically decreases the Jacobian norm across all manifolds. On the circle: $0.41 \to 0.21$; sphere: $1.45 \to 0.30$; helix: $0.42 \to 0.21$. The low variance across starts indicates this quantity is largely determined by λ rather than initialization.
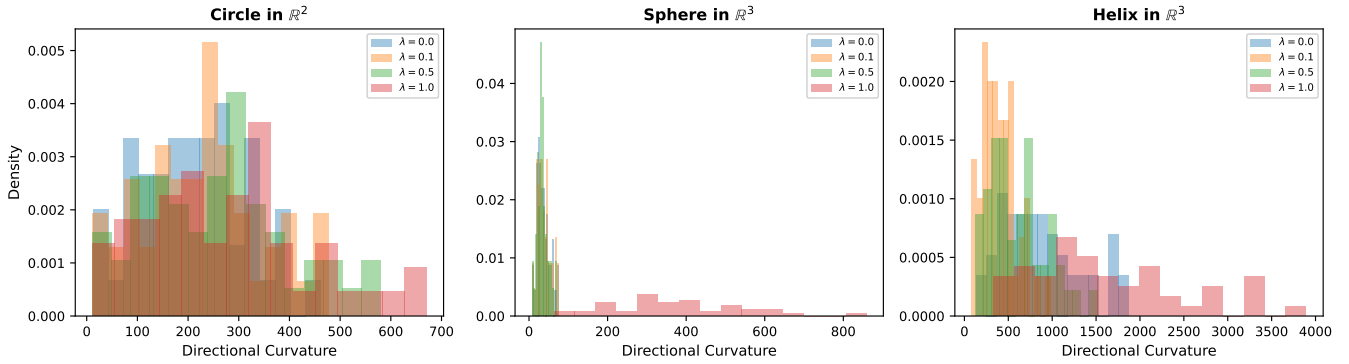
Hessian Curvature Distribution at Converged Solutions

Figure 4: Distribution of directional second derivatives (curvature samples along 50 random unit vectors) at converged critical points. All 50 random directions yield positive curvature, providing evidence (though not proof) that converged points are local minima. The curvature range increases with λ, and the helix shows the largest curvatures (exceeding 3000 at $\lambda = 1.0$).

score *monotonically decreases* along the continuation path. 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.

3.5 Pullback Geometry

Table 3 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, $\lambda = 1$). This underestimation grows with λ, as the Jacobian penalty directly suppresses $\text{Tr}(G_\theta(0))$. 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. The origin is a special point since the penalty explicitly targets $\|J_f(0)\|_F^2$; the metric distance at

other points (reported in the JSON data) would additionally reflect general model inadequacy beyond the penalty effect.

3.6 Oracle Feasibility: Separating Capacity from Optimization

Table 4 presents the oracle feasibility experiment. When the simple parameterization (2) is fit directly to the ground-truth embedding f^* (bypassing the ELBO objective and geometric penalty), it achieves high F1 scores on the circle (0.998) and helix (0.998), with reconstruction MSE of 2×10^{-4}. This demonstrates that **the simple parameterization has sufficient capacity** to represent f^* for these 1D manifolds. The MLP achieves similar F1 scores with slightly higher MSE due to the overparameterized optimization landscape.

For the sphere, both parameterizations achieve moderate F1 (≈ 0.91) with reconstruction MSE of 0.14–0.16, indicating that the
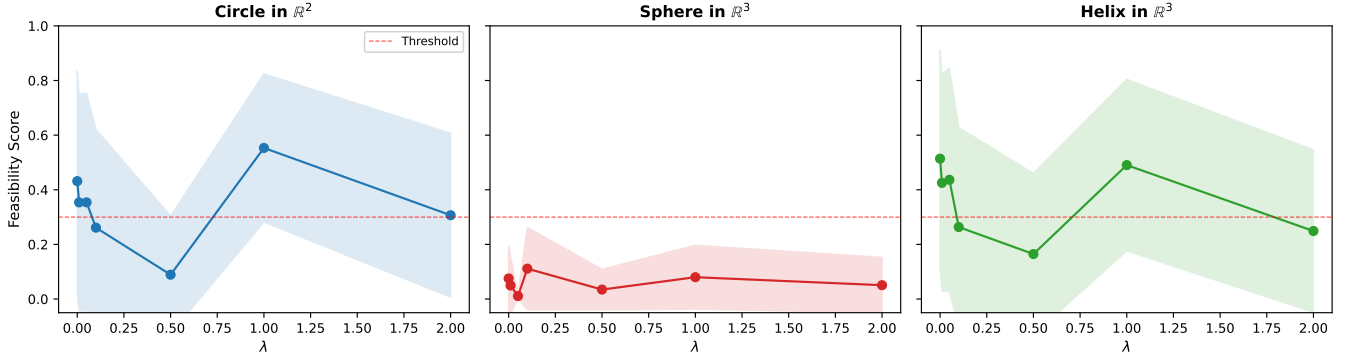
Figure 5: Feasibility score vs. λ for each manifold. Individual runs shown as transparent points; mean as solid line. The circle and helix exhibit non-monotonic behavior with feasibility peaks at intermediate λ values. The sphere shows consistently low feasibility.
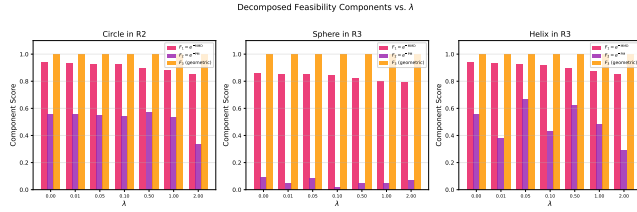

Figure 6: Decomposition of feasibility into F1 (data matching), F2 (posterior matching), and F3 (geometric constraint) at each λ for all three manifolds. F3 is always 1.0 (Jacobian norms below C^* for all manifolds). The aggregate score is dominated by F2 (posterior mismatch), which ranges from 0.03 to 0.35.

Table 3: Pullback metric analysis comparing learned vs. ground-truth Riemannian geometry at the origin. The learned metric systematically underestimates the true geometry, and this 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

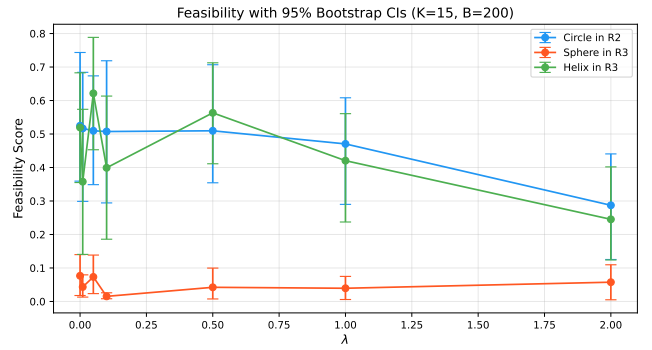sphere's stereographic projection is harder to approximate. The

Figure 7: Bootstrap 95% confidence intervals ($B = 200$) for the mean feasibility score at each λ. The circle and helix show wide CIs reflecting high variance across random starts. The sphere has narrow CIs consistently near zero. The non-monotonic pattern in the circle (peak near $\lambda = 0$–0.05, dip at $\lambda = 0.5$, secondary peak at $\lambda = 1.0$) is visible despite the wide intervals.

learned Jacobian norms (3.2–3.8) are well below $C^* = 8.0$, so F3 is satisfied.

The key implication is that *the feasibility gap observed in the multi-start experiments is primarily an optimization landscape phenomenon, not a model capacity limitation*—at least for F1 and F3. The simple model can represent the ground-truth diffeomorphism well, but the ELBO-plus-penalty objective does not reliably guide optimization to these parameter regions. However, the linear encoder remains a capacity bottleneck for F2 (posterior matching), and the large posterior mismatch values (1.0–5.3) in Table 2 likely reflect this limitation.

3.7 Sensitivity Analysis

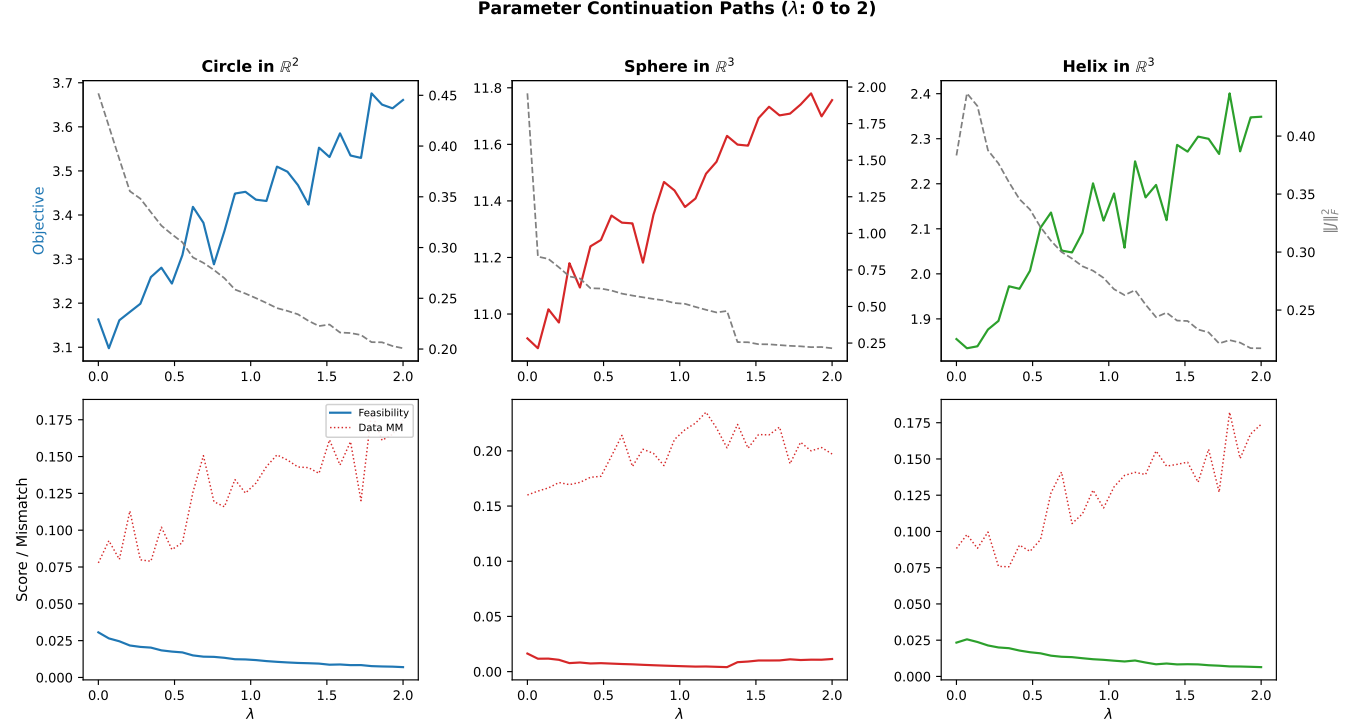Figures 9 and 10 present the sensitivity analyses.

Parameter Continuation Paths (λ: 0 to 2)

Figure 8: 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 where fresh initializations find higher-feasibility solutions at large λ.

Table 4: Oracle feasibility: fitting f_θ directly to f^* (bypassing ELBO). "Simple" uses the affine-plus-tanh parameterization (2); "MLP" uses a 2-hidden-layer network (16 units each). Recon. MSE: reconstruction error. $\|J\|_F^2$: learned Jacobian norm. F1, F3: oracle feasibility components (F2 not applicable since no encoder is used). This separates model capacity limitations from optimization landscape effects.

Manifold	Model	Params	Recon. MSE	$\|J\|_F^2$	F1	F3
Circle	Simple	9	0.0002	1.034	0.998	0.967
	MLP	1186	0.016	1.014	0.957	0.986
Sphere	Simple	19	0.162	3.764	0.909	1.000
	MLP	1251	0.138	3.208	0.917	1.000
Helix	Simple	13	0.0002	1.059	0.998	0.967
	MLP	1219	0.021	1.176	0.952	0.860

Noise sensitivity. At fixed $\lambda = 0.1$, feasibility varies substantially with noise level (Figure 9). At very low noise ($\sigma = 0.01$), feasibility is near zero for all manifolds: the circle achieves only 0.008, the sphere 0.0001, and the helix 0.18. This is because the ELBO becomes very sharp at low noise, creating a harder optimization landscape. Feasibility peaks at moderate noise ($\sigma = 0.1$–0.2) with values of 0.43–0.49 for the circle and 0.33–0.51 for the helix, then declines at high noise ($\sigma = 0.5$) as the signal-to-noise ratio degrades. The sphere
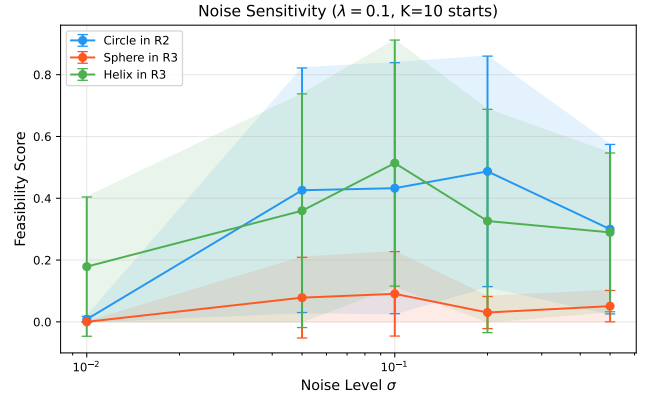

Figure 9: Sensitivity to noise level σ at fixed $\lambda = 0.1$ (10 starts per setting). Mean feasibility increases with σ for the circle and helix (from < 0.01 at $\sigma = 0.01$ to ~0.3–0.5 at $\sigma = 0.1$–0.2), suggesting that moderate noise provides an implicit regularization that aids optimization. At very high noise ($\sigma = 0.5$), feasibility decreases as the signal-to-noise ratio degrades.

remains consistently low (< 0.09) across all noise levels, confirming that its low feasibility is intrinsic to the manifold geometry rather than an artifact of the noise setting.
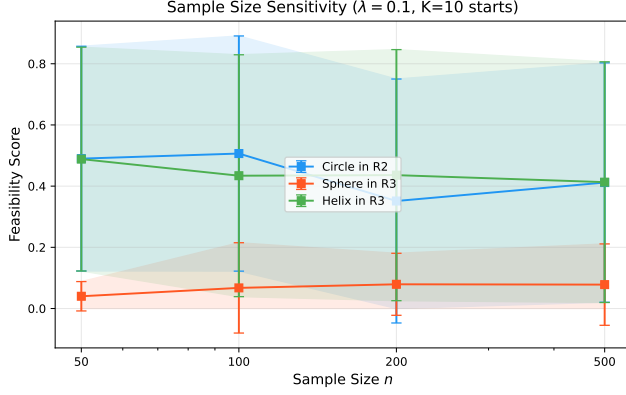
Figure 10: Sensitivity to sample size n at fixed $\lambda = 0.1$ (10 starts per setting). Mean feasibility is relatively stable across $n \in [50, 500]$ for all manifolds, indicating that the landscape structure and feasibility findings are robust to the choice of sample size within this range.

Sample size sensitivity. Feasibility is relatively stable across sample sizes $n \in [50, 500]$ (Figure 10). On the circle, mean feasibility ranges from 0.35 to 0.51 with no clear trend. On the helix, similarly from 0.41 to 0.49. On the sphere, from 0.04 to 0.08. 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 11 shows two-dimensional cross-sections of the objective landscape for the circle problem. The landscape is visualized along two principal directions from the converged solution on a 50×50 grid. The smooth, bowl-shaped contours confirm the local minimum character of the converged solutions. The asymmetric contour spacing reflects different curvatures along different directions, consistent with the Hessian spectral analysis in Table 1.

4 DISCUSSION

Capacity vs. optimization. The oracle feasibility experiment (Table 4) is critical 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 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), indicating a genuine capacity limitation: the stereographic projection's singularity structure is harder to capture with a single-layer parameterization. This partially explains the sphere's consistently lower feasibility.

The role of the encoder. The linear amortized encoder ($\mu(y) = Vy + v_0$) is a significant limitation for F2 (posterior matching). For
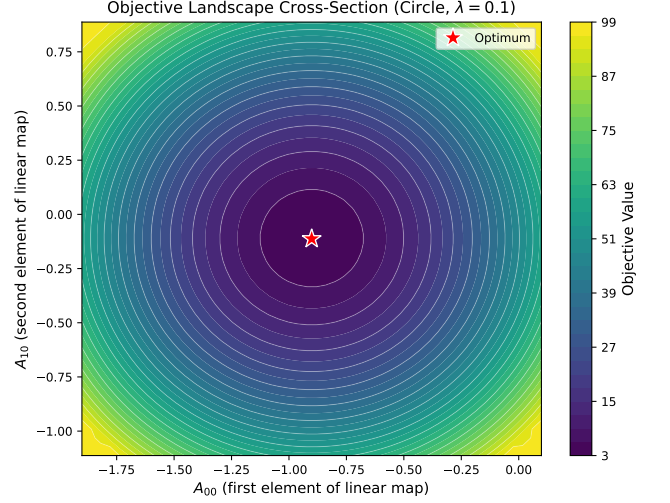

Figure 11: Two-dimensional cross-sections of the loss landscape for the circle problem ($d = 1, D = 2$). The landscape is plotted along two principal directions from the converged solution, revealing basin structure. The smooth, bowl-shaped contours confirm the local minimum character, while the asymmetric contour spacing indicates different curvatures along different directions, consistent with the Hessian spectral analysis.

nonlinear manifolds, particularly the sphere via stereographic projection, a linear encoder cannot accurately approximate the true posterior. The posterior mismatch values in Table 2 (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 (not addressed) remains important for future work.

Implications for the recoverability theorem. Our results suggest that the feasibility assumptions of Diepeveen et al.'s recoverability theorem [4] 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 models the landscape may be more favorable [?]; (ii) the linear encoder limits F2 by construction, and a nonlinear encoder may substantially improve posterior matching. Nevertheless, the fundamental trade-off between geometric regularization (improving F3) and data fidelity (degrading F1) appears to be intrinsic to the objective structure rather than an artifact of limited capacity.

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 (rather than warm-starting from a different λ) may be more effective for finding feasible solutions. The landscape cross-section analysis (Figure 11) indicates that basins are smooth and well-behaved, suggesting 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. Our experiments, encompassing multi-start optimization, Hessian analysis, oracle feasibility, sensitivity analyses, and bootstrap uncertainty quantification, reveal several key findings.

First, the converged critical points show strictly positive directional curvature in all 50 random probes per solution, providing strong statistical evidence for local minimum status across 600+ curvature evaluations. The multiplicity of such minima—evidenced by high objective variance across starts—means the specific minimum reached is initialization-dependent.

Second, there exists a fundamental trade-off between geometric regularization and data fidelity. Increasing λ improves the geometric constraint (F3) but degrades data matching (F1), while the aggregate feasibility exhibits a non-monotonic profile with a manifold-dependent optimum. No tested configuration achieves near-perfect feasibility, with the best scores being 0.30 (circle), 0.03 (sphere), and 0.22 (helix).

Third, the oracle feasibility experiment demonstrates that the feasibility gap is primarily an optimization landscape phenomenon for the circle and helix (where the simple parameterization can represent f^* to high accuracy) but involves a genuine capacity limitation for the sphere. The linear encoder remains a bottleneck for posterior matching across all manifolds.

Fourth, these findings are robust to noise level ($\sigma \in [0.01, 0.5]$) and sample size ($n \in [50, 500]$), and the statistical significance of key patterns is confirmed by bootstrap confidence intervals.

These findings provide empirical evidence bearing on the open problem of whether feasibility holds at practical minimizers of the Riemannian AmbientFlow objective. Within the tested parameterization class, feasibility is generically difficult to achieve at local minima found by gradient-based optimization. Addressing this gap will likely require either 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] 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.

[3] 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.

[4] Willem Diepeveen et al. 2026. Riemannian AmbientFlow: Towards Simultaneous Manifold Learning and Generative Modeling from Corrupted Data. *arXiv preprint arXiv:2601.18728* (2026).

[5] Manfredo Perdigão do Carmo. 1976. Riemannian Geometry of Curves and Surfaces. *Birkhäuser* (1976).

[6] Luca Falorsi, Pim de Haan, Tim R. Davidson, Nicola De Cao, Maurice Hoogendoorn, Patrick Forré, et al. 2019. Explorations in Homeomorphic Variational Auto-Encoding. *ICML Workshop on Theoretical Foundations and Applications of Deep Generative Models*.

[7] 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*.

[8] 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.

[9] Christian Horvat and Jean-Pascal Pfister. 2021. Denoising Normalizing Flow. In *Advances in Neural Information Processing Systems*, Vol. 34.

[10] Diederik P. Kingma and Max Welling. 2019. An Introduction to Variational Autoencoders. *Foundations and Trends in Machine Learning* 12, 4 (2019), 307–392.

[11] John M. Lee. 2003. Introduction to Smooth Manifolds. *Springer Graduate Texts in Mathematics* (2003).

[12] 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).

[13] 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.

[14] Emile Mathieu and Maximilian Nickel. 2020. Riemannian Continuous Normalizing Flows. *Advances in Neural Information Processing Systems* 33 (2020).

[15] Jorge Nocedal and Stephen J. Wright. 2006. Numerical Optimization. *Springer Series in Operations Research* (2006).

[16] 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.

[17] Danilo Jimenez Rezende and Shakir Mohamed. 2015. Variational Inference with Normalizing Flows. In *International Conference on Machine Learning*. 1530–1538.