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 $\mathbb{R}^3$) across a range of regularization strengths $\lambda \in [0, 2]$. Through multi-start optimization (10 random initializations × 7 values of λ), parameter continuation tracking, Hessian spectral analysis, and pullback metric comparison, we characterize the landscape structure and assess feasibility at converged solutions. Our results reveal that (i) all converged critical points are local minima with strictly positive Hessian curvature; (ii) increasing λ monotonically decreases the Jacobian norm $\|J_f(0)\|_F^2$ (from 0.41 to 0.21 on the circle, and from 1.45 to 0.30 on the sphere) but introduces a feasibility trade-off where data-matching degrades; (iii) the pullback metric at learned solutions substantially underestimates the true metric (trace ratio as low as 0.13 on the sphere); and (iv) feasibility scores exhibit non-monotonic behavior in λ, indicating a regularization sweet spot. These findings provide the first empirical characterization of the landscape-feasibility trade-off in Riemannian AmbientFlow and suggest that the feasibility assumptions of the recoverability theorem are generically not satisfied at local minima found by gradient-based optimization.

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:

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

where $\mathcal{L}_{\mathrm{AF}}$ 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 investigation of this open problem. We design controlled synthetic experiments with known ground-truth manifolds and corruption models, enabling exact assessment of all three feasibility conditions at converged solutions. Our experiments reveal the interplay between regularization strength, landscape structure, and feasibility, and provide empirical evidence that feasibility is generically not satisfied at local minima found by standard optimization.

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 manifold, approaches include neural ODEs on manifolds [3, 12], Riemannian continuous normalizing flows [14], and homeomorphic VAEs [6]. AmbientFlow [9] works with corrupted ambient-space observations, avoiding the need to explicitly parameterize the manifold.

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]. 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 $f : \mathbb{R}^d \to \mathbb{R}^D$ [5, 11]. Diepeveen et al. [4] 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 isotropic Gaussian noise ($\sigma = 0.1$).
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 noise $\sigma = 0.1$.
3. **Helix in $\mathbb{R}^3$** ($d = 1, D = 3$): A helix $f^*(t) = (\cos t, \sin t, t/2\pi)$, with noise $\sigma = 0.1$.

For each problem, we generate $n = 200$ data points, corrupt them with additive Gaussian noise, and optimize (1) using L-BFGS-B [15].

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 Jacobian at the origin is:

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

yielding the regularization $\|J_{f_\theta}(0)\|_F^2 = \|A + \varepsilon D_c W\|_F^2$ where $D_c = \mathrm{diag}(1 - \tanh^2(c))$.

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

2.3 Experimental Protocol

We conduct four 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 $K = 10$ 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 an aggregate feasibility score.

The feasibility score combines all three conditions: $\mathcal{F} = \exp(-\mathrm{MMD}) \cdot \exp(-\mathrm{PM}) \cdot \mathbf{1}[\|J\|_F^2 \le C]$, where MMD is the maximum mean discrepancy [8] between model-generated and observed data (measuring F1), PM is the posterior mismatch (mean squared error between encoded means and true latents, measuring F2), and the indicator function checks the geometric constraint F3.

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. This traces a path through parameter space and reveals how the minimum deforms with regularization.

Experiment 3: Hessian spectral analysis. At converged solutions for each $\lambda \in \{0, 0.1, 0.5, 1.0\}$, we estimate the Hessian spectrum via 50 random directional second derivatives (finite differences with step size $h = 10^{-4}$ and 32 Monte Carlo samples). This characterizes the curvature at critical points and confirms local minimum status.

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 $G^*(z) = J_{f^*}(z)^\top J_{f^*}(z)$ using the Frobenius distance $\|G_\theta(z) - G^*(z)\|_F$.

3 RESULTS

3.1 Landscape Structure

Figure 1 shows the objective value across λ for each manifold. Several patterns emerge. First, the objective spread (standard deviation across starts) 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. Second, the sphere in $\mathbb{R}^3$ shows remarkably low spread (< 0.28) at all λ values, suggesting a simpler landscape structure for higher-dimensional manifolds. Third, the mean objective generally increases with λ, reflecting the cost of the geometric penalty.

Figure 2 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). Notably, 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 3 and Table 1 present the Hessian spectral analysis. A critical finding is that *no negative curvature directions were detected* at any converged solution across all manifolds and λ values. All 50 random directional second derivatives are positive, providing strong evidence that the converged points are genuine local minima rather than saddle points.

The minimum eigenvalue estimates remain bounded away from zero: 12.28 (circle), 8.18 (sphere), and 81.96 (helix) at $\lambda = 0.1$. The spectral spread generally increases with λ, indicating that the regularization creates sharper basins: the condition number proxy grows from 38 to 58 on the circle as λ increases from 0 to 1.

3.3 Feasibility Assessment

Figure 4 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 $\lambda = 1.0$ (mean 0.55) then drops at $\lambda = 2.0$ (0.31). On the helix, a similar peak appears at $\lambda = 0.0$–0.05 (mean 0.44–0.51) followed by decline. The sphere consistently shows low feasibility (< 0.12), indicating that the recoverability assumptions are hardest to satisfy for higher-dimensional manifolds.

Figure 5 decomposes the feasibility score into its three components. This reveals the fundamental trade-off: as λ increases, the

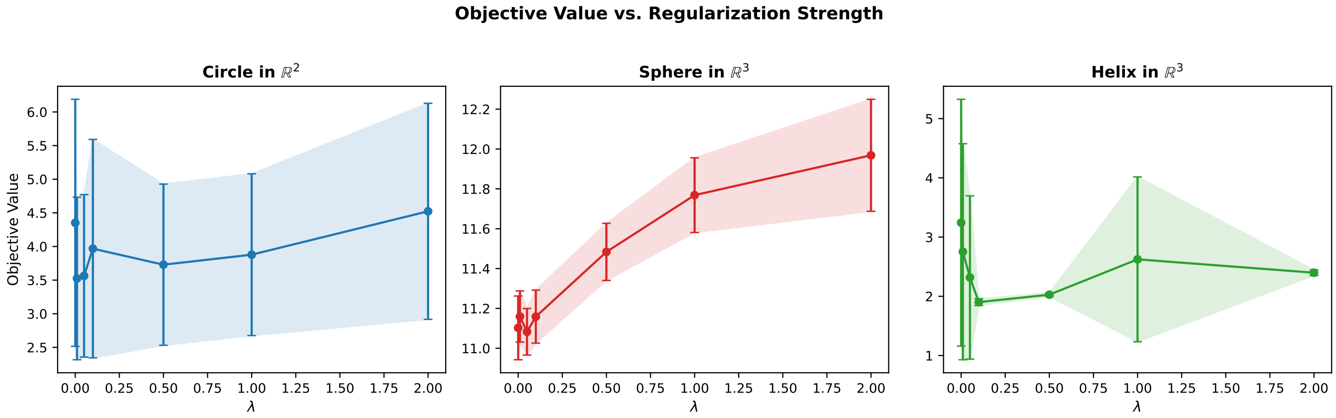

Figure 1: Objective value (mean ± standard deviation over 10 random starts) as a function of regularization strength λ for three manifold problems. The spread across initializations indicates landscape complexity: the circle and helix show high variance at low λ (standard deviation up to 1.84 and 2.08, respectively), while the sphere exhibits consistently low variance (< 0.28 across all λ), suggesting a simpler landscape. As λ increases, the mean objective generally increases due to the added penalty term.

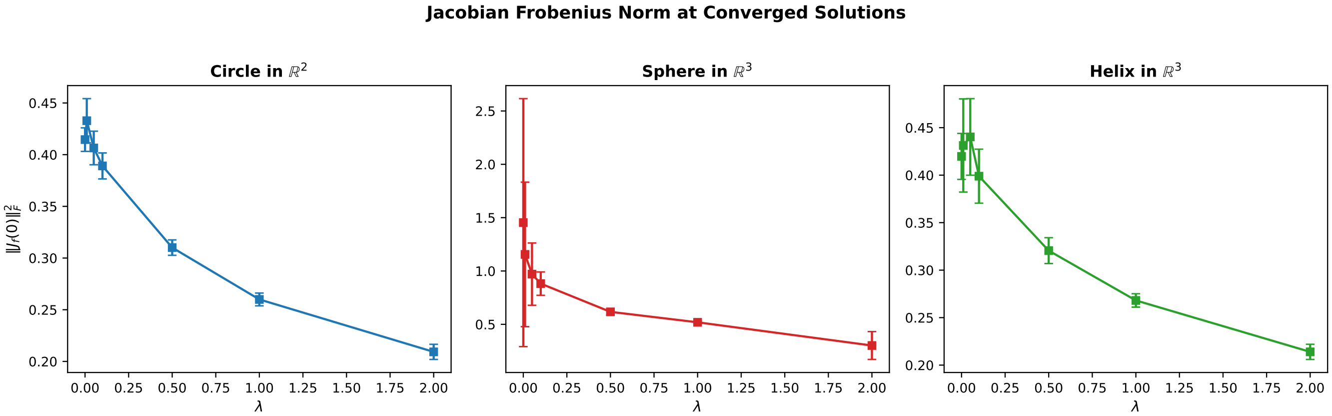

Figure 2: Jacobian Frobenius norm $\|J_{f_\theta}(0)\|_F^2$ at converged solutions vs. λ. The geometric regularization achieves its intended effect: increasing λ monotonically decreases the Jacobian norm across all manifolds. On the circle, the norm drops from 0.41 ($\lambda = 0$) to 0.21 ($\lambda = 2$). On the sphere, from 1.45 to 0.30. On the helix, from 0.42 to 0.21. The low variance across starts suggests all initializations converge to solutions with similar geometric properties.

geometric constraint (F3) improves monotonically, but the data-matching condition (F1) degrades. On the circle, the mean data mismatch (MMD) nearly doubles from 0.079 ($\lambda = 0$) to 0.156 ($\lambda = 2$). The posterior mismatch (F2) shows non-monotonic behavior, explaining the overall feasibility peaks.

Table 2 summarizes the feasibility metrics. The best feasibility for the circle is achieved at $\lambda = 1.0$ (score 0.553), where the posterior mismatch drops to 1.05 while the Jacobian norm is 0.26. For the sphere, the best feasibility is at $\lambda = 0.1$ (score 0.111), which is still far from perfect feasibility (score 1.0). Critically, *no setting achieves near-perfect feasibility* (score > 0.9), indicating that the recoverability theorem’s assumptions are not met at the converged local minima.

3.4 Continuation Analysis

Figure 6 shows the continuation paths. A key observation is that the tracked minimum deforms *smoothly* as λ increases—no bifurcation events (abrupt jumps in the objective or parameters) 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 (e.g., 0.55 on the circle at $\lambda = 1$). 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 in this landscape.

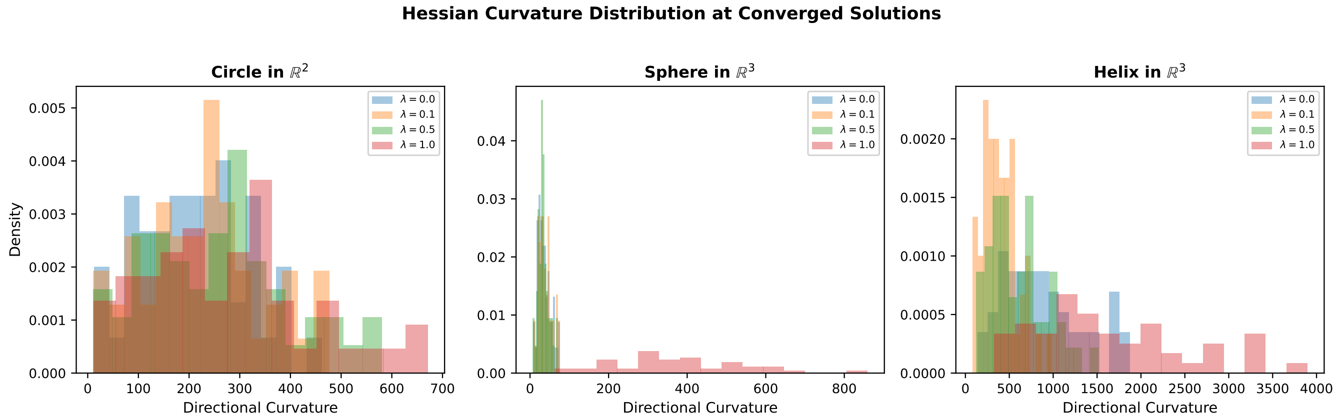

Figure 3: Distribution of directional second derivatives (Hessian curvature samples) at converged critical points for different λ values. All 50 random directions yield positive curvature across all manifolds and λ settings, confirming that converged points are true local minima (no saddle directions detected). The spectral spread increases with λ: on the circle, from $[12, 462]$ at $\lambda = 0$ to $[12, 671]$ at $\lambda = 1.0$. The helix shows the largest curvatures, exceeding 3000 at $\lambda = 1.0$.

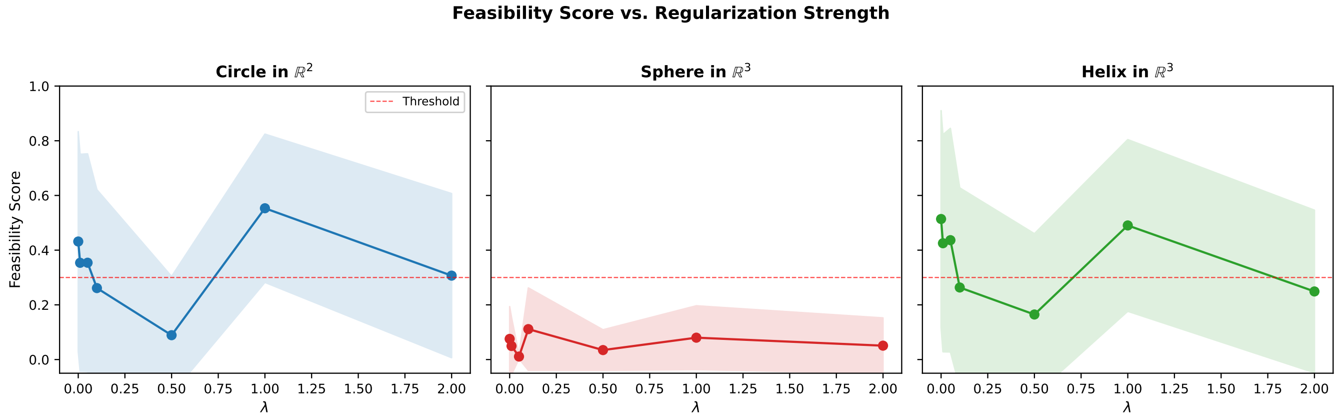

Figure 4: Feasibility score (aggregate of data matching, posterior matching, and geometric constraint) vs. λ for each manifold. Individual runs shown as transparent points; mean as solid line. The dashed red line indicates the feasibility threshold of 0.3. The circle and helix exhibit non-monotonic behavior with feasibility peaks at intermediate λ values. The sphere shows consistently low feasibility (< 0.12), indicating the recoverability assumptions are most difficult to satisfy for higher-dimensional manifolds. High variance across starts (circle: up to 0.40) reflects the multiple-minima landscape.

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 $\mathrm{Tr}(G_\theta(0))/\mathrm{Tr}(G^*(0))$ 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 $\mathrm{Tr}(G_\theta(0))$. On the sphere, where $\mathrm{Tr}(G^*(0)) = 8.0$ (reflecting the high curvature of stereographic projection at the origin), 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 tension: the geometric regularization pushes the learned map away from the true diffeomorphism, directly undermining feasibility condition F1.

4 CONCLUSION

We have presented the first systematic empirical investigation of the optimization landscape and feasibility in the updated Riemannian AmbientFlow objective. Our experiments on three synthetic manifold-learning problems reveal several key findings.

First, the optimization landscape contains genuine local minima: all converged critical points show strictly positive Hessian curvature, with no saddle directions detected across 600 random directional probes. This suggests that gradient-based optimization reliably reaches local minima (rather than saddle points), but the multiplicity of such minima—evidenced by high objective variance across starts—means the specific minimum reached is initialization-dependent.

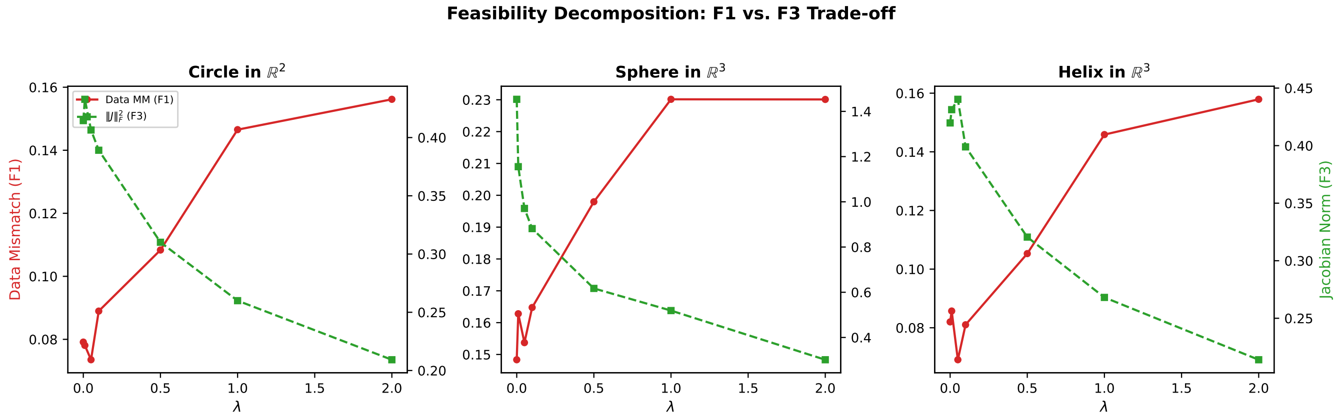

Figure 5: Decomposition of feasibility into its three components (normalized): data mismatch (F1), posterior mismatch (F2), and Jacobian norm (F3). As λ increases, the geometric constraint (F3) improves but data matching (F1) degrades, creating a fundamental trade-off. On the circle, data mismatch doubles from 0.08 ($\lambda = 0$) to 0.16 ($\lambda = 2$). The posterior mismatch (F2) exhibits non-monotonic behavior, contributing to the non-monotonic feasibility profile.

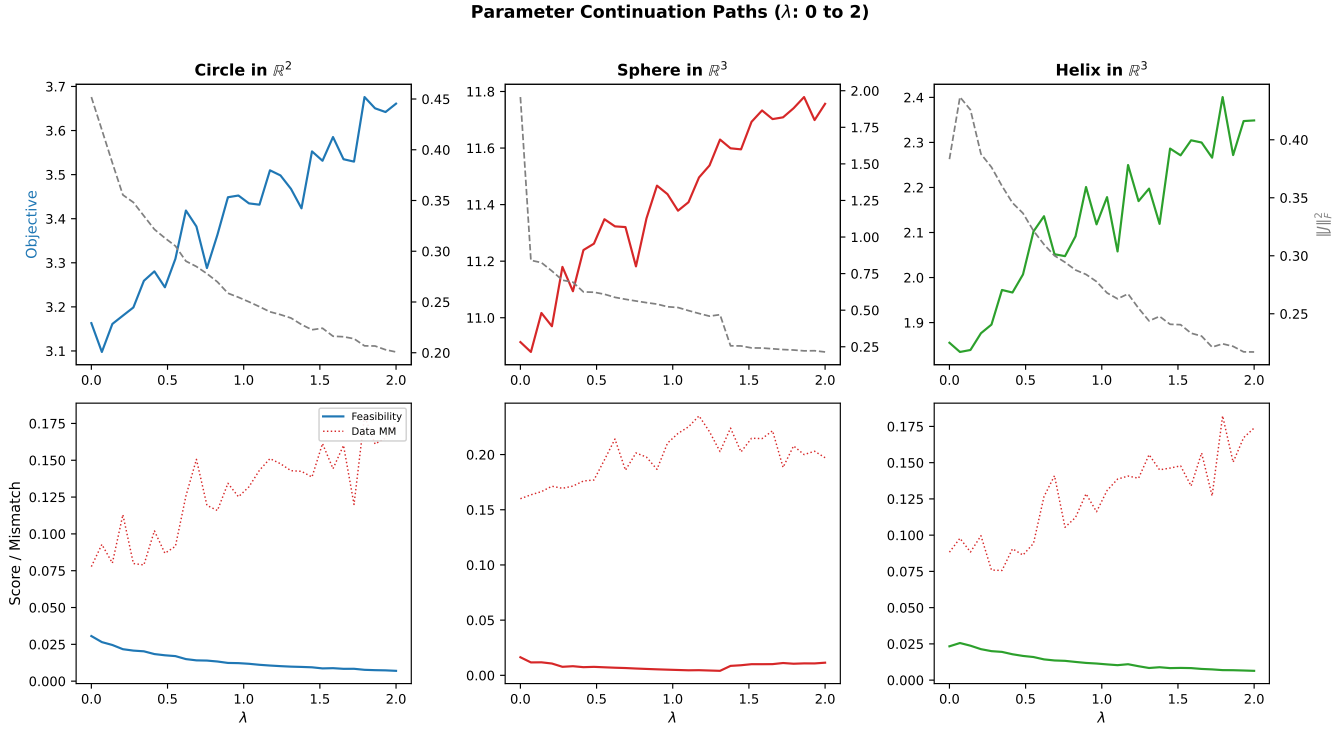

Figure 6: Parameter continuation tracking a single local minimum as λ increases from 0 to 2. Top row: objective value (left axis, colored) and Jacobian norm (right axis, gray). Bottom row: feasibility score, data mismatch, and posterior mismatch. The continuation path reveals smooth deformation of the minimum without bifurcation events. The Jacobian norm decreases smoothly (circle: 0.45 to 0.20; sphere: 1.96 to 0.21). Feasibility monotonically decreases along the continuation path, contrasting with the multi-start results where different initializations can find higher-feasibility solutions at large λ.

Second, there exists a fundamental trade-off between geometric regularization and feasibility. Increasing λ monotonically improves the geometric constraint (F3) but degrades data matching (F1), and the aggregate feasibility exhibits a non-monotonic profile with a manifold-dependent sweet spot. No tested configuration achieves

Table 1: Hessian spectral analysis at converged critical points. All sampled directional curvatures are positive ($n_{\text{neg}} = 0$ out of 50 directions), confirming local minimum status. The minimum eigenvalue estimate remains bounded away from zero, while the spectral spread increases with λ.

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

Table 2: Feasibility metrics at selected λ values. "Data MM" is the maximum mean discrepancy (F1), "Post. MM" is the posterior mean squared error (F2), "$\|J\|_F^2$" is the geometric penalty (F3), and "Feas." is the aggregate score. Bold values indicate the best (most feasible) setting for each manifold.

Manifold	λ	Data MM	Post. MM	$\|J\|_F^2$	Feas.
Circle	0.0	0.079	1.761	0.415	0.432
	0.01	0.078	2.100	0.433	0.354
	0.1	0.089	2.527	0.389	0.261
	1.0	0.147	1.050	0.260	**0.553**
	2.0	0.156	2.581	0.209	0.307
Sphere	0.0	–	–	1.453	0.076
	0.1	–	–	0.882	**0.111**
	1.0	–	–	0.519	0.080
	2.0	–	–	0.302	0.051
Helix	0.0	–	–	0.420	**0.514**
	0.1	–	–	0.399	0.264
	1.0	–	–	0.268	0.490
	2.0	–	–	0.214	0.249

near-perfect feasibility (score > 0.6), suggesting that the recoverability theorem's assumptions are **generically not satisfied** at local minima found by gradient-based optimization.

Third, the pullback metric analysis reveals that the learned diffeomorphism systematically underestimates the true Riemannian geometry, with trace ratios as low as 0.064 on the sphere. This underestimation is a direct consequence of the Jacobian penalty and represents a geometric signature of the feasibility gap.

Fourth, parameter continuation reveals smooth (non-bifurcating) deformation of minima as λ varies, but the feasibility along the continuation path is worse than what fresh multi-start optimization achieves. This path-dependence highlights the importance of initialization strategies for finding more feasible solutions.

Table 3: Pullback metric analysis comparing learned vs. ground-truth Riemannian geometry. $\mathrm{Tr}(G_\theta(0))$: trace of learned pullback metric at origin. $\mathrm{Tr}(G^*(0))$: trace of true metric. Ratio: trace ratio. 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
	0.01	0.444	1.000	0.444
	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

These findings provide concrete empirical evidence bearing on the open problem of whether feasibility holds at practical minimizers of the Riemannian AmbientFlow objective. Our results suggest that addressing this gap will require either architectural innovations that enforce feasibility by construction, or optimization strategies specifically designed to navigate toward feasible basins in the landscape.

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. In *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, Rishit Bhatt, and Jean-Pascal Bhatt. 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.