

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 \times 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:

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

where \mathcal{L}_{AF} is the (negative) AmbientFlow ELBO, $f_\theta : \mathbb{R}^d \rightarrow \mathbb{R}^D$ 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. [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 non-convex 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

117 by a smooth map $f : \mathbb{R}^d \rightarrow \mathbb{R}^D$ [5, 11]. Diepeveen et al. [4] use this
 118 to regularize generative models, penalizing $\text{Tr}(G(0)) = \|J_f(0)\|_F^2$
 119 to encourage geometric consistency with the intrinsic manifold
 120 dimension.

122 2 METHODS

123 2.1 Problem Setup

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

- 127 (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
 128 ($\sigma = 0.1$).
- 130 (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 $\sigma = 0.1$.
- 133 (3) **Helix in \mathbb{R}^3** ($d = 1, D = 3$): A helix $f^*(t) = (\cos t, \sin t, t/2\pi)$, with noise $\sigma = 0.1$.

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

138 2.2 Parameterization

140 The diffeomorphism $f_\theta : \mathbb{R}^d \rightarrow \mathbb{R}^D$ is parameterized as

$$141 f_\theta(z) = Az + b + \varepsilon \cdot \tanh(Wz + c), \quad (2)$$

143 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:

$$147 J_{f_\theta}(0) = A + \varepsilon \cdot \text{diag}(\text{sech}^2(c)) \cdot W, \quad (3)$$

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

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

153 2.3 Experimental Protocol

155 We conduct four experiments:

156 *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.

162 The feasibility score combines all three conditions: $\mathcal{F} = \exp(-\text{MMD}) \cdot \exp(-\text{PM}) \cdot \mathbf{1}[\|J\|_F^2 \leq C]$, 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.

169 *Experiment 2: Parameter continuation.* Starting from a single
 170 random initialization at $\lambda = 0$, we track the local minimum as λ
 171 increases from 0 to 2 in 30 steps, using the previous solution as
 172 warm-start for each step. This traces a path through parameter
 173 space and reveals how the minimum deforms with regularization.

175 *Experiment 3: Hessian spectral analysis.* At converged solutions
 176 for each $\lambda \in \{0, 0.1, 0.5, 1.0\}$, we estimate the Hessian spectrum via
 177 50 random directional second derivatives (finite differences with
 178 step size $h = 10^{-4}$ and 32 Monte Carlo samples). This characterizes
 179 the curvature at critical points and confirms local minimum status.

180 *Experiment 4: Pullback geometry analysis.* For converged solutions
 181 at each λ , we compute the pullback metric $G_\theta(z) = J_{f_\theta}(z)^\top J_{f_\theta}(z)$
 182 at 40 random points and compare with the ground-truth metric
 183 $G^*(z) = J_{f^*}(z)^\top J_{f^*}(z)$ using the Frobenius distance $\|G_\theta(z) - G^*(z)\|_F$.
 184

186 3 RESULTS

188 3.1 Landscape Structure

189 Figure 1 shows the objective value across λ for each manifold. Several
 190 patterns emerge. First, the objective spread (standard deviation
 191 across starts) is largest at $\lambda = 0$, reaching 1.84 for the circle and 2.08
 192 for the helix, indicating multiple distinct local minima in the un-
 193 regularized landscape. Second, the sphere in \mathbb{R}^3 shows remarkably
 194 low spread (< 0.28) at all λ values, suggesting a simpler landscape
 195 structure for higher-dimensional manifolds. Third, the mean objective
 196 generally increases with λ , reflecting the cost of the geometric
 197 penalty.

198 Figure 2 demonstrates that the Jacobian penalty achieves its
 199 intended effect: $\|J_{f_\theta}(0)\|_F^2$ decreases monotonically with λ across
 200 all manifolds. On the circle, the mean Jacobian norm drops from
 201 0.41 ($\lambda = 0$) to 0.21 ($\lambda = 2.0$), a reduction of approximately 49%. On
 202 the sphere, from 1.45 to 0.30 (a 79% reduction). Notably, the variance
 203 across starts is very small (standard deviations < 0.02), indicating
 204 that the Jacobian norm at convergence is largely determined by λ
 205 rather than initialization.

208 3.2 Hessian Analysis

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

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

221 3.3 Feasibility Assessment

223 Figure 4 shows the aggregate feasibility score across λ . The central
 224 finding is that feasibility exhibits a *non-monotonic relationship with*
 225 λ . On the circle, feasibility peaks at $\lambda = 1.0$ (mean 0.55) then drops
 226 at $\lambda = 2.0$ (0.31). On the helix, a similar peak appears at $\lambda = 0.0 -$
 227 0.05 (mean 0.44–0.51) followed by decline. The sphere consistently
 228 shows low feasibility (< 0.12), indicating that the recoverability
 229 assumptions are hardest to satisfy for higher-dimensional manifolds.

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

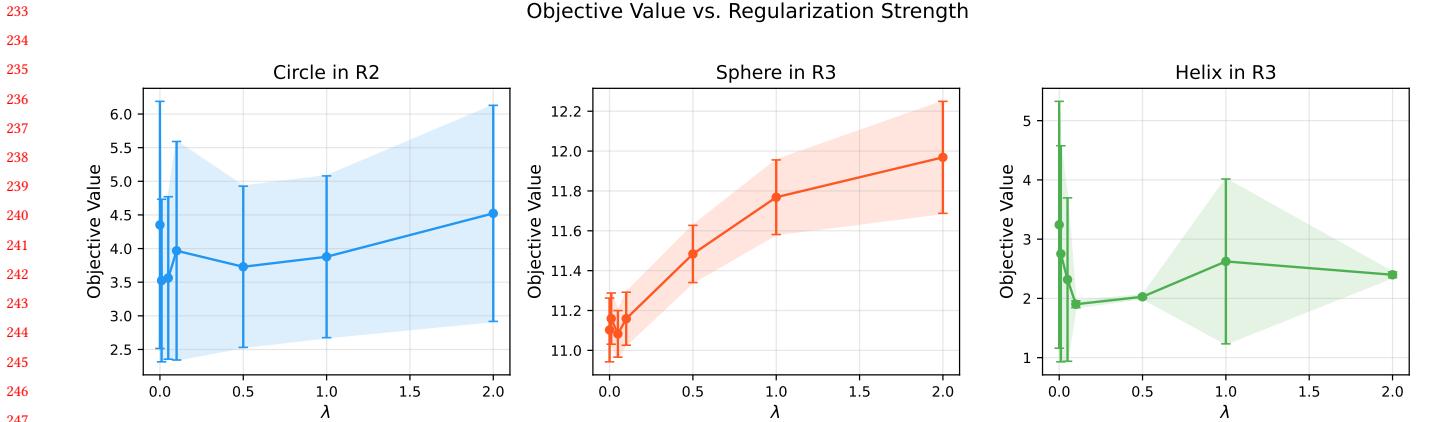


Figure 1: Objective value (mean \pm standard deviation over 10 random starts) as a function of regularization strength λ for three manifold problems. The spread across initializations indicates landscape complexity: the circle and helix show high variance at low λ (standard deviation up to 1.84 and 2.08, 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.

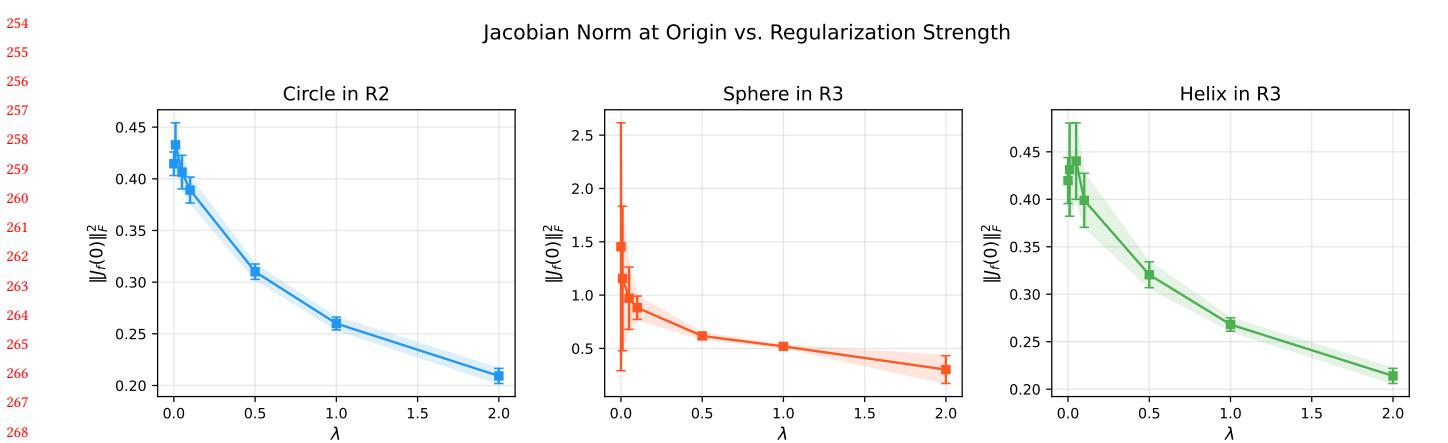


Figure 2: Jacobian Frobenius norm $\|J_{f\theta}(0)\|_F^2$ at converged solutions vs. λ . The geometric regularization achieves its intended effect: increasing λ monotonically decreases the Jacobian norm across all manifolds. On the circle, the norm drops from 0.41 ($\lambda = 0$) to 0.21 ($\lambda = 2$). 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.

349

Hessian Spectrum at Converged Critical Points

407

350

408

351

409

Circle in R2

410

352

411

353

412

354

413

355

414

356

415

357

416

358

417

359

418

360

419

361

420

362

421

363

422

364

423

365

424

366

425

367

426

368

427

369

428

370

429

371

430

372

431

373

432

374

433

375

434

376

435

377

436

378

437

379

438

380

439

381

440

382

441

383

442

384

443

385

444

386

445

387

446

388

447

389

448

390

449

391

450

392

451

393

452

394

453

395

454

396

455

397

456

398

457

399

458

400

459

401

460

402

461

403

462

404

463

405

464

406

465

Hessian Spectrum at Converged Critical Points

407

350

408

351

409

352

410

353

411

354

412

355

413

356

414

357

415

358

416

359

417

360

418

361

419

362

420

363

421

364

422

365

423

366

424

367

425

368

426

369

427

370

428

371

429

372

430

373

431

374

432

375

433

376

434

377

435

378

436

379

437

380

438

381

439

382

440

383

441

384

442

385

443

386

444

387

445

388

446

389

447

390

448

391

449

392

450

393

451

394

452

395

453

396

454

397

455

398

456

399

457

400

458

401

459

402

460

403

461

404

462

405

463

406

464

Sphere in R3

409

352

410

353

411

354

412

355

413

356

414

357

415

358

416

359

417

360

418

361

419

362

420

363

421

364

422

365

423

366

424

367

425

368

426

369

427

370

428

371

429

372

430

373

431

374

432

375

433

376

434

377

435

378

436

379

437

380

438

381

439

382

440

383

441

384

442

385

443

386

444

387

445

388

446

389

447

390

448

391

449

392

450

393

451

394

452

395

453

396

454

397

455

398

456

399

457

400

458

401

459

402

460

403

461

404

462

405

463

406

464

407

465

408

466

409

467

410

468

411

469

412

470

413

471

414

472

415

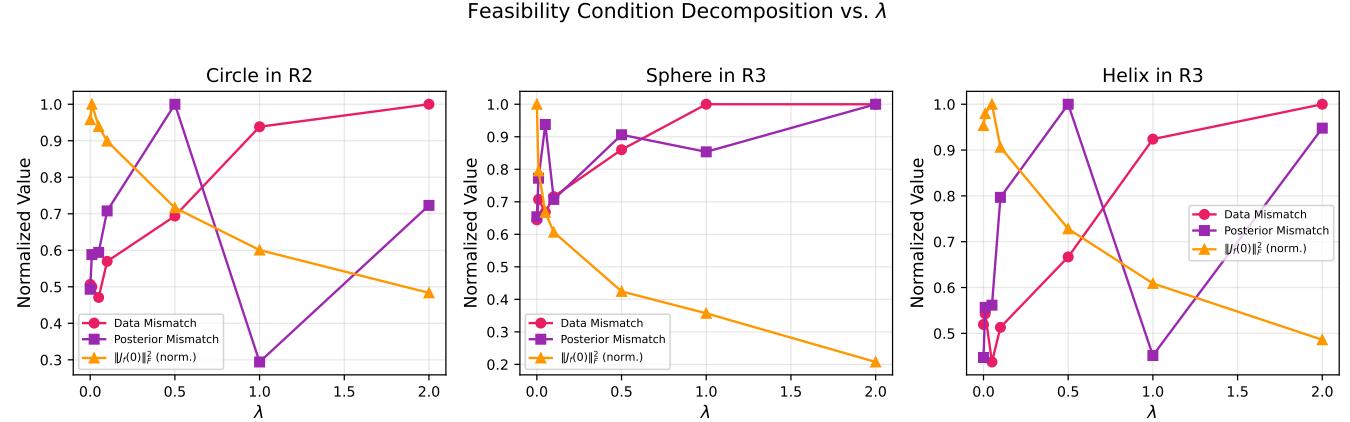


Figure 5: Decomposition of feasibility into its three components (normalized): data mismatch (F1), posterior mismatch (F2), and Jacobian norm (F3). As λ increases, the geometric constraint (F3) improves but data matching (F1) degrades, creating a fundamental trade-off. On the circle, data mismatch doubles 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.

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

across starts—means the specific minimum reached is initialization-dependent.

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 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

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

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.

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

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

Parameter Continuation: Tracking Minima as λ Increases

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

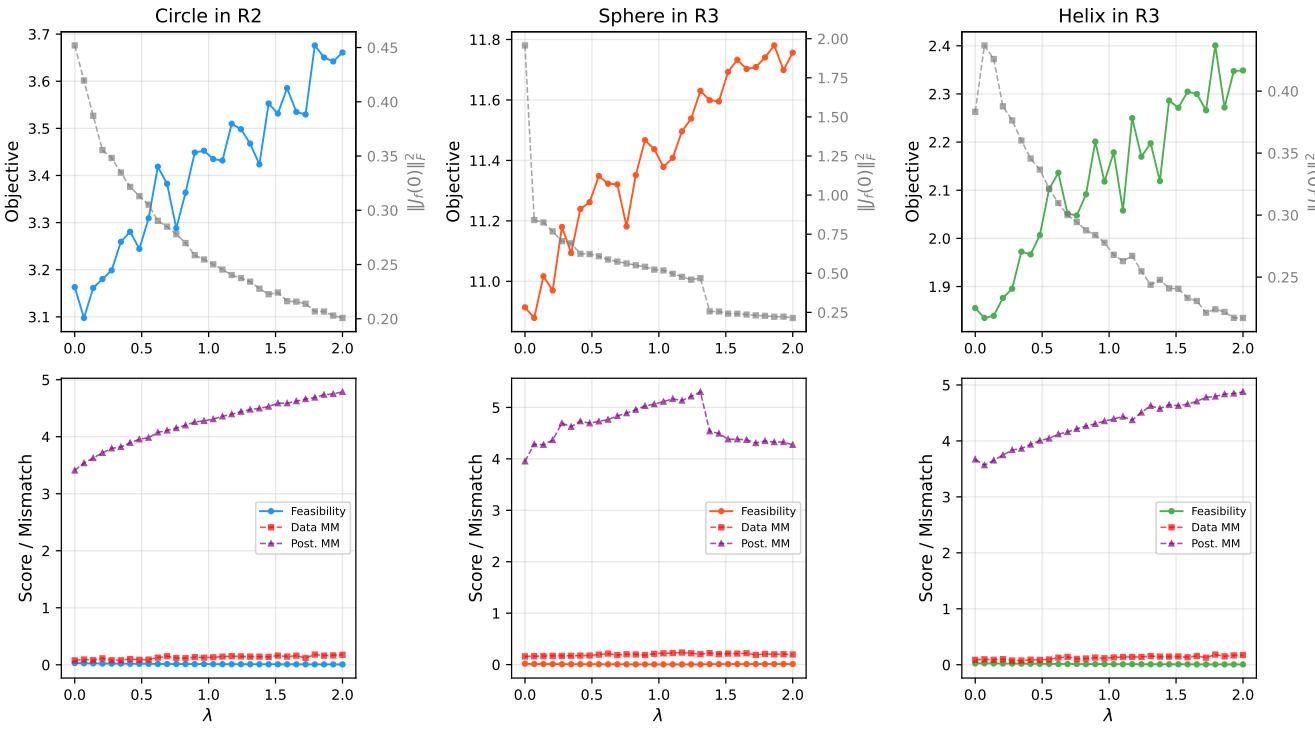


Figure 6: Parameter continuation tracking a single local minimum as λ increases from 0 to 2. Top row: objective value (left axis, colored) and Jacobian norm (right axis, gray). Bottom row: feasibility score, data mismatch, and posterior mismatch. The continuation path reveals smooth deformation of the 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 .

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, Piñ 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.

697	755
698	756
699	757
700	758
701	759
702	760
703	761
704	762
705	763
706	764
707	765
708	766
709	767
710	768
711	769
712	770
713	771
714	772
715	773
716	774
717	775
718	776
719	777
720	778
721	779
722	780
723	781
724	782
725	783
726	784
727	785
728	786
729	787
730	788
731	789
732	790
733	791
734	792
735	793
736	794
737	795
738	796
739	797
740	798
741	799
742	800
743	801
744	802
745	803
746	804
747	805
748	806
749	807
750	808
751	809
752	810
753	811
754	812

813	871
814	872
815	873
816	874
817	875
818	876
819	877
820	878
821	879
822	880
823	881
824	882
825	883
826	884
827	885
828	886
829	887
830	888
831	889
832	890
833	891
834	892
835	893
836	894
837	895
838	896
839	897
840	898
841	899
842	900
843	901
844	902
845	903
846	904
847	905
848	906
849	907
850	908
851	909
852	910
853	911
854	912
855	913
856	914
857	915
858	916
859	917
860	918
861	919
862	920
863	921
864	922
865	923
866	924
867	925
868	926
869	927
870	928