Neural Manifold Learning Experiments: Comparative Analysis of Five Training Modes

Chaoyi Pan[a,$] Giri Anantharaman[a] Nai-Chieh Huang[a] Claire Jin[a] Daniel Pfrommer[b]
Chenyang Yuan[c] Frank Permenter[c] Guannan Qu[a,†] Nicholas Boffi[a,†] Guanya Shi[a,†]
Max Simchowitz[a,†]

[$]Project lead. [†]Equal advising.

[a]Carnegie Mellon University [b]Massachusetts Institute of Technology [c]Toyota Research Institute

December 20, 2025

Executive Summary

This report presents a comprehensive comparison of five training modes (regression, flow matching, MIP, MIP one-step, and straight flow) across three manifold learning tasks (reconstruction, projection, and Lie algebra). Each mode was evaluated with two loss functions (L1 and L2) and two neural architectures (concatenation and FiLM), for a total of 20 configurations per task. All experiments were conducted across 3 random seeds.

Key Findings:

Flow-based methods (flow matching, MIP, MIP one-step, and straight flow) consistently outperform regression on projection-based geometric metrics, demonstrating superior manifold adherence and boundary enforcement. In contrast, regression achieves the best L2 reconstruction error, excelling at direct supervised learning of point-wise mappings. This reveals a fundamental trade-off: flow-based approaches learn better geometric structure while regression optimizes reconstruction accuracy.

- **Reconstruction Task:** MIP one-step trained with L2 loss using FiLM architecture achieves the best L2 test error (0.003197), demonstrating superior reconstruction accuracy with a single denoising step at evaluation time.
- **Projection Task:** Straight flow trained with L1 loss using FiLM architecture achieves the best boundary metric (0.009769), indicating excellent performance in projecting points onto the target manifold with minimal boundary violations.
- **Lie Algebra Task:** Straight flow trained with L1 loss using FiLM architecture achieves the best average projection metric (0.063612), demonstrating strong performance in maintaining orthogonality to the Lie algebra.

Note on Training Loss: While we evaluate all modes with both L1 and L2 losses for completeness, training flow-based methods (flow matching, straight flow) with L1 loss lacks mathematical grounding, as the conditional flow matching objective is naturally derived for L2 loss. Results with L1-trained flow models should be interpreted with this caveat in mind.

Contents

1 Experimental Setup

1.1 Network Architecture

All experiments compare two conditioning architectures with identical capacity:

- **Concatenation (Concat)**: Concatenates condition c and input x before processing: $[\mathbf{x}; \mathbf{c}]$
- **Feature-wise Linear Modulation (FiLM)**: Applies affine transformations to hidden layer activations conditioned on c: $\text{FiLM}(\mathbf{h}, \mathbf{c}) = \gamma(\mathbf{c}) \odot \mathbf{h} + \beta(\mathbf{c})$
- **Hidden dimension**: 256
- **Hidden layers**: 3
- **Activation**: ReLU
- **Batch size**: 32
- **Training epochs**: 50,000
- **Optimizer**: Adam with learning rate 0.001

1.2 Training Paradigms

Regression: Directly minimizes reconstruction loss between network predictions $\hat{f}(\mathbf{x}, \mathbf{c})$ and target values $f(\mathbf{c})$.

Flow Matching: Learns a conditional vector field $v_\theta(\mathbf{x}, \mathbf{c}, t)$ that transports Gaussian noise to target values via ODE integration:

$$\frac{d\mathbf{x}_t}{dt} = v_\theta(\mathbf{x}_t, \mathbf{c}, t), \quad \mathbf{x}_0 \sim \mathcal{N}(0, I) \tag{1}$$

Evaluated using Euler integration with time conditioning.

Straight Flow: An ablation of flow matching that tests whether time conditioning improves performance. During training, samples interpolated points $\mathbf{x}_t = (1-t)\mathbf{x}_0 + t\mathbf{x}_1$ at random times $t \in [0, 1]$, but always queries the model at $t = 0$ (no time information provided). The model learns to predict $\mathbf{x}_1$ directly from any point along the interpolation path without knowing its position. Evaluated using the same ODE integration as flow matching, but with t always set to 0.

MIP (Manifold Interpolation): Combines flow matching with a denoising term at fixed time $t^* = 0.9$:

$$\mathcal{L}_{\text{MIP}} = \mathcal{L}_{\text{flow}}(t) + \lambda \|\mathbf{x}_{t^*} - f(\mathbf{c})\|^2 \tag{2}$$

The denoising term encourages the learned trajectory to pass through the target manifold at t^*.

MIP One-Step: Uses the same training as MIP but evaluates using only the initial denoising step:

$$\hat{f}(\mathbf{x}_0, \mathbf{c}) = \text{model}(\mathbf{x}_0, \mathbf{c}, t = 0) \tag{3}$$

instead of the full two-step MIP evaluation.

1.3 Training Details

- **Training samples**: $n = 50$ per condition
- **Test samples**: $n = 100,000$
- **Random seeds**: 3 independent runs per configuration
- **Loss functions**: Both L1 and L2 losses evaluated

2 Tasks and Metric Definitions

2.1 Reconstruction Task

Objective: Learn a scalar target function composed of trigonometric terms:

$$f(c) = \sum_{i=1}^{K} w_i \cdot \text{trig}_i(\omega_i c + \phi_i) \tag{4}$$

where $\text{trig}_i \in \{\sin, \cos\}$ alternates, $K = 3$ components, $c \in [0, 1]$. Frequencies ω_i are prime-based to avoid overlaps, and weights $w_i = 1$ (uniform).

Metrics:

- **L1 test error**: $\|\hat{f}(\mathbf{x}, c) - f(c)\|_1$
- **L2 test error**: $\|\hat{f}(\mathbf{x}, c) - f(c)\|_2$

Lower values indicate better reconstruction accuracy.

2.2 Projection Task

Objective: Project points onto a learned low-dimensional subspace embedded in high-dimensional space. The target manifold is a piecewise-linear function that changes slope at discrete boundaries.

Subspace Metric Definitions: Let $c_j \in \{0.1, 0.2, \ldots, 0.9\}$ denote boundary points, and define $\bar{I}_j = [c_j - 0.03, c_j + 0.03]$ as the evaluation interval around each boundary. Let P_j denote the projection matrix onto the subspace spanning interval $[c_j - 0.1, c_j]$ and P_{j+1} the projection onto $[c_j, c_{j+1}]$. We define:

$$P_{j,j+1} = [P_j, P_{j+1}] \tag{5}$$

as the concatenation of these two projection matrices. Let $\tilde{P}_{j,j+1}$ denote the projection onto the span of $P_{j,j+1}$. This is computed via SVD on $P_{j,j+1}$, taking the top $2k$ singular vectors.

We evaluate the following normalized metrics on test points $\hat{f}$ for $c \in \bar{I}_j$:

- **Subspace Diagonal**: $\frac{\|(I - \tilde{P}_{j,j+1})(\hat{f} - f^*)\|}{\|\hat{f} - f^*\|}$ when $c \in \bar{I}_j$ (measures normalized deviation from the correct local subspace at boundaries)
- **Subspace Off-Diagonal**: $\frac{\|(I - \tilde{P}_{j,j+1})(\hat{f} - f^*)\|}{\|\hat{f} - f^*\|}$ when $c \notin \bar{I}_j$ (measures normalized global subspace adherence)
- **Boundary**: $\frac{\|(I - \tilde{P}_{j,j+1})\hat{f}\|}{\|\hat{f}\|}$ (measures normalized discontinuities at the boundary points c_j)

Additionally, we report:

- **L1 test error**: $\|\hat{f}(\mathbf{x}, c) - f(c)\|_1$
- **L2 test error**: $\|\hat{f}(\mathbf{x}, c) - f(c)\|_2$

Lower values indicate better projection accuracy and manifold adherence.

2.3 Lie Algebra Task

Objective: Learn a vector field $f(c) : \mathbb{R} \to \mathbb{R}^d$ that satisfies Lie algebra constraints. The target is constructed using rotation matrices to encode tangent space structure.

Target Function: We define a family of rotation-based target functions with K components:

$$f_i(\alpha, c) = (w_i(c) \cdot \exp(\alpha A_i) \cdot \boldsymbol{e}_1)_{1 \le i \le K} \in \mathbb{R}^{d_{\mathrm{rot}}} \tag{6}$$

where A_i are skew-symmetric matrices (generators of rotations), $\boldsymbol{e}_1$ is a fixed reference vector, and $w_i(c)$ are scalar weighting functions. The full target function is:

$$f(c) = \mathrm{concat}_{1 \le i \le K}(f_i(\alpha_i c, c)) \tag{7}$$

This construction tests the model's ability to learn components rotating at different velocities α_i.

Evaluation Metrics: For each component i at condition c, define the projection onto the tangent space:

$$P_i(c) = \mathrm{Projection}(\mathrm{span}(\exp(c\alpha_i A_i)\boldsymbol{e}_1)) \tag{8}$$

The geometric error for component i is measured as:

$$\text{Projection Error}_i(c) = \frac{\|(I - P_i(c)) f_i(\alpha_i c, c)\|}{\|f_i(\alpha_i c, c)\|} \tag{9}$$

Equivalently, we can measure alignment via cosine similarity:

$$\text{cos-err}_i(c) = \cos\text{-similarity}(\exp(c\alpha_i A_i)\boldsymbol{e}_1, f_i(\alpha_i c, c)) \tag{10}$$

Reported Metrics:

- **Average Cosine Similarity**: Mean cos-err$_i(c)$ across all components and test conditions (higher is better, indicates better tangent space alignment)
- **Minimum Cosine Similarity**: Worst-case cos-err$_i(c)$ across all test points (higher is better)
- **Average Projection**: Mean projection error across all components (lower is better, measures orthogonality to Lie algebra)
- **Maximum Projection**: Maximum projection error (lower is better, measures worst-case deviation)
- **L1 test error**: $\|\hat{f}(\mathbf{x}, c) - f(c)\|_1$
- **L2 test error**: $\|\hat{f}(\mathbf{x}, c) - f(c)\|_2$

Key Insight: When rotation velocities α_i increase with i, faster-rotating components present greater challenges for maintaining manifold constraints. Methods that explicitly model the geometric structure (MIP, flow matching, straight flow) are expected to handle high-velocity components better than direct regression.

3 Reconstruction Task

The reconstruction task evaluates each method's ability to reconstruct manifold points from noisy observations. Performance is measured using L1 and L2 test errors on held-out data.

3.1 Results: Models Trained with L1 Loss

3.1.1 Averaged Results (L1 Loss Training)

Mode	Loss	Arch	L1	L2
straight_flow	l1	concat	0.007651 ± 0.000615	0.009371 ± 0.001167
straight_flow	l1	film	0.004299 ± 0.001390	0.005296 ± 0.001542
mip_one_step_integrate	l1	concat	0.005705 ± 0.002182	0.007312 ± 0.002560
mip_one_step_integrate	l1	film	0.004656 ± 0.001546	0.005832 ± 0.001721
regression	l1	concat	0.006367 ± 0.001027	0.007884 ± 0.000692
regression	l1	film	0.005961 ± 0.001094	0.007284 ± 0.001672
flow	l1	concat	0.021114 ± 0.002823	0.026871 ± 0.004521
flow	l1	film	0.042884 ± 0.005564	0.052019 ± 0.003222
mip	l1	concat	0.004185 ± 0.000247	0.005811 ± 0.000500
mip	l1	film	0.007836 ± 0.003189	0.010120 ± 0.004307

Table 1: Averaged results (trained with L1 loss) for recon task across modes and architectures

3.1.2 Seed-wise Results (L1 Loss Training)

Mode	Loss	Arch	Seed	L1	L2
straight_flow	l1	concat	0	0.007906	0.010877
straight_flow	l1	concat	1	0.006804	0.008032
straight_flow	l1	concat	2	0.008244	0.009206
straight_flow	l1	film	0	0.006165	0.007405
straight_flow	l1	film	1	0.003904	0.004719
straight_flow	l1	film	2	0.002829	0.003763
mip_one_step_integrate	l1	concat	0	0.003881	0.004825
mip_one_step_integrate	l1	concat	1	0.004463	0.006279
mip_one_step_integrate	l1	concat	2	0.008772	0.010834
mip_one_step_integrate	l1	film	0	0.003674	0.004565
mip_one_step_integrate	l1	film	1	0.006839	0.008266
mip_one_step_integrate	l1	film	2	0.003456	0.004665
regression	l1	concat	0	0.004960	0.006917
regression	l1	concat	1	0.007386	0.008493
regression	l1	concat	2	0.006754	0.008242
regression	l1	film	0	0.005248	0.006303
regression	l1	film	1	0.005129	0.005912
regression	l1	film	2	0.007507	0.009638
flow	l1	concat	0	0.023371	0.032170
flow	l1	concat	1	0.017134	0.021124
flow	l1	concat	2	0.022839	0.027319
flow	l1	film	0	0.050699	0.056197
flow	l1	film	1	0.038183	0.048355
flow	l1	film	2	0.039771	0.051505
mip	l1	concat	0	0.003863	0.005119
mip	l1	concat	1	0.004465	0.006034
mip	l1	concat	2	0.004226	0.006282
mip	l1	film	0	0.007724	0.009555
mip	l1	film	1	0.011796	0.015655
mip	l1	film	2	0.003987	0.005152

Table 2: Seed-wise results (trained with L1 loss) for recon task across modes and architectures

3.2 Results: Models Trained with L2 Loss

3.2.1 Averaged Results (L2 Loss Training)

Mode	Loss	Arch	L1	L2
straight_flow	l2	concat	0.003851 ± 0.001085	0.005065 ± 0.001444
straight_flow	l2	film	0.005315 ± 0.003148	0.006799 ± 0.003815
mip_one_step_integrate	l2	concat	0.003159 ± 0.000561	0.004117 ± 0.000519
mip_one_step_integrate	l2	film	0.002381 ± 0.000479	0.003197 ± 0.000525
regression	l2	concat	0.002288 ± 0.000040	0.003314 ± 0.000091
regression	l2	film	0.006915 ± 0.006942	0.008062 ± 0.007377
flow	l2	concat	0.060241 ± 0.009467	0.071551 ± 0.007469
flow	l2	film	0.054293 ± 0.009452	0.066804 ± 0.009595
mip	l2	concat	0.003807 ± 0.000103	0.004862 ± 0.000195
mip	l2	film	0.003112 ± 0.000411	0.004021 ± 0.000316

Table 3: Averaged results (trained with L2 loss) for recon task across modes and architectures

3.2.2 Seed-wise Results (L2 Loss Training)

Mode	Loss	Arch	Seed	L1	L2
straight_flow	l2	concat	0	0.004236	0.005630
straight_flow	l2	concat	1	0.004944	0.006482
straight_flow	l2	concat	2	0.002371	0.003083
straight_flow	l2	film	0	0.003892	0.004798
straight_flow	l2	film	1	0.009681	0.012139
straight_flow	l2	film	2	0.002373	0.003460
mip_one_step_integrate	l2	concat	0	0.002606	0.003622
mip_one_step_integrate	l2	concat	1	0.002945	0.003895
mip_one_step_integrate	l2	concat	2	0.003928	0.004834
mip_one_step_integrate	l2	film	0	0.003050	0.003936
mip_one_step_integrate	l2	film	1	0.002137	0.002885
mip_one_step_integrate	l2	film	2	0.001956	0.002769
regression	l2	concat	0	0.002233	0.003196
regression	l2	concat	1	0.002325	0.003416
regression	l2	concat	2	0.002307	0.003332
regression	l2	film	0	0.016732	0.018494
regression	l2	film	1	0.001995	0.002805
regression	l2	film	2	0.002018	0.002886
flow	l2	concat	0	0.055125	0.068045
flow	l2	concat	1	0.073514	0.081934
flow	l2	concat	2	0.052084	0.064675
flow	l2	film	0	0.047475	0.064479
flow	l2	film	1	0.067658	0.079544
flow	l2	film	2	0.047745	0.056390
mip	l2	concat	0	0.003865	0.005120
mip	l2	concat	1	0.003663	0.004651
mip	l2	concat	2	0.003894	0.004814
mip	l2	film	0	0.003648	0.004468
mip	l2	film	1	0.003037	0.003785
mip	l2	film	2	0.002651	0.003811

Table 4: Seed-wise results (trained with L2 loss) for recon task across modes and architectures

3.3 Reconstruction Task Summary

Across all configurations, models trained with L2 loss generally outperform L1-trained models on both L1 and L2 test metrics. The regression baseline achieves the lowest L1 test error (0.002288), while MIP one-step achieves the best L2 test error (0.003197). The MIP one-step approach demonstrates that a single denoising step can achieve competitive or superior reconstruction performance compared to the traditional two-step MIP approach. Straight flow shows moderate performance, suggesting that time conditioning provides valuable information for reconstruction tasks.

4 Projection Task

The projection task evaluates each method's ability to project arbitrary points onto the learned manifold. Performance is assessed using geometric metrics measuring subspace adherence and boundary discontinuities.

4.1 Results: Models Trained with L1 Loss

4.1.1 Averaged Results (L1 Loss Training)

Mode	Loss	Arch	L1	L2	Subspace Diag	Subspace
straight_flow	l1	concat	0.961537 ± 0.008393	1.388102 ± 0.009159	0.088869 ± 0.001517	0.782543 ±
straight_flow	l1	film	0.966100 ± 0.002405	1.402517 ± 0.003180	0.088979 ± 0.002925	0.782738 ±
mip_one_step_integrate	l1	concat	0.952631 ± 0.005196	1.375615 ± 0.007067	0.096746 ± 0.002745	0.782950 ±
mip_one_step_integrate	l1	film	0.954570 ± 0.004069	1.381055 ± 0.011186	0.094953 ± 0.003293	0.782912 ±
regression	l1	concat	0.934109 ± 0.000258	1.348542 ± 0.001355	0.645471 ± 0.050949	0.758075 ±
regression	l1	film	0.938874 ± 0.008625	1.359889 ± 0.021979	0.114321 ± 0.007385	0.783407 ±
flow	l1	concat	1.008413 ± 0.013867	1.430288 ± 0.015294	0.195982 ± 0.006026	0.782279 ±
flow	l1	film	0.984554 ± 0.004379	1.404958 ± 0.012633	0.190697 ± 0.018725	0.781708 ±
mip	l1	concat	0.982445 ± 0.005516	1.406487 ± 0.008115	0.086279 ± 0.002817	0.782500 ±
mip	l1	film	0.965555 ± 0.001022	1.382665 ± 0.001200	0.107176 ± 0.000703	0.782379 ±

Table 5: Averaged results (trained with L1 loss) for proj task across modes and architectures

4.1.2 Seed-wise Results (L1 Loss Training)

Mode	Loss	Arch	Seed	L1	L2	Subspace Diag	Subspace Off-Diag	Boundary
straight_flow	l1	concat	0	0.950777	1.378141	0.089966	0.782858	0.011763
straight_flow	l1	concat	1	0.962579	1.385911	0.086725	0.783008	0.012132
straight_flow	l1	concat	2	0.971257	1.400254	0.089917	0.781764	0.010325
straight_flow	l1	film	0	0.963650	1.402508	0.090370	0.782811	0.008506
straight_flow	l1	film	1	0.969367	1.406416	0.084910	0.783439	0.012071
straight_flow	l1	film	2	0.965284	1.398627	0.091657	0.781964	0.008731
mip_one_step_integrate	l1	concat	0	0.948438	1.371865	0.095191	0.783143	0.017119
mip_one_step_integrate	l1	concat	1	0.949500	1.369466	0.094443	0.783822	0.019266
mip_one_step_integrate	l1	concat	2	0.959954	1.385512	0.100604	0.781885	0.019079
mip_one_step_integrate	l1	film	0	0.954311	1.370723	0.093130	0.783806	0.015496
mip_one_step_integrate	l1	film	1	0.949720	1.375848	0.092153	0.783068	0.013733
mip_one_step_integrate	l1	film	2	0.959678	1.396595	0.099577	0.781861	0.015733
regression	l1	concat	0	0.933858	1.349405	0.607174	0.762349	0.385087
regression	l1	concat	1	0.934464	1.349592	0.717474	0.751762	0.392547
regression	l1	concat	2	0.934005	1.346629	0.611765	0.760113	0.398936
regression	l1	film	0	0.950696	1.390935	0.124129	0.782950	0.035222
regression	l1	film	1	0.935568	1.343060	0.106307	0.783671	0.020470
regression	l1	film	2	0.930360	1.345672	0.112525	0.783601	0.033193
flow	l1	concat	0	1.028023	1.451764	0.196102	0.782906	0.101709
flow	l1	concat	1	0.998805	1.421769	0.203300	0.780756	0.103574
flow	l1	concat	2	0.998410	1.417330	0.188542	0.783174	0.097786
flow	l1	film	0	0.978532	1.388090	0.212672	0.779576	0.093783
flow	l1	film	1	0.988816	1.418493	0.166913	0.783566	0.089496
flow	l1	film	2	0.986315	1.408291	0.192504	0.781982	0.096300
mip	l1	concat	0	0.978350	1.399293	0.089940	0.781320	0.032891
mip	l1	concat	1	0.990242	1.417827	0.085810	0.783024	0.031739
mip	l1	concat	2	0.978743	1.402341	0.083088	0.783157	0.019507
mip	l1	film	0	0.964850	1.381951	0.106794	0.781406	0.033723
mip	l1	film	1	0.967001	1.384356	0.108162	0.783058	0.037610
mip	l1	film	2	0.964815	1.381688	0.106572	0.782674	0.041882

Table 6: Seed-wise results (trained with L1 loss) for proj task across modes and architectures

4.2 Results: Models Trained with L2 Loss

4.2.1 Averaged Results (L2 Loss Training)

Mode	Loss	Arch	L1	L2	Subspace Diag	Subspace
straight_flow	l2	concat	0.953148 ± 0.007896	1.368337 ± 0.008432	0.100474 ± 0.010328	0.782947 ±
straight_flow	l2	film	0.953106 ± 0.001934	1.382213 ± 0.003216	0.092751 ± 0.001830	0.783178 ±
mip_one_step_integrate	l2	concat	0.943220 ± 0.002871	1.359088 ± 0.003868	0.096157 ± 0.003722	0.782377 ±
mip_one_step_integrate	l2	film	0.940340 ± 0.006580	1.359007 ± 0.011974	0.098788 ± 0.001163	0.783082 ±
regression	l2	concat	0.941179 ± 0.006019	1.356883 ± 0.005875	0.721983 ± 0.024342	0.782519 ±
regression	l2	film	0.941871 ± 0.005438	1.353876 ± 0.005981	0.109260 ± 0.006132	0.783310 ±
flow	l2	concat	0.998419 ± 0.010089	1.426732 ± 0.015232	0.194247 ± 0.012624	0.781902 ±
flow	l2	film	0.986891 ± 0.011952	1.404517 ± 0.010156	0.191611 ± 0.006451	0.782208 ±
mip	l2	concat	0.978285 ± 0.002389	1.401722 ± 0.002520	0.093925 ± 0.004407	0.782092 ±
mip	l2	film	0.971166 ± 0.004485	1.395023 ± 0.004815	0.097735 ± 0.002881	0.782565 ±

Table 7: Averaged results (trained with L2 loss) for proj task across modes and architectures

4.2.2 Seed-wise Results (L2 Loss Training)

Mode	Loss	Arch	Seed	L1	L2	Subspace Diag	Subspace Off-Diag	Boundary
straight_flow	l2	concat	0	0.964274	1.379420	0.115070	0.782642	0.029684
straight_flow	l2	concat	1	0.948415	1.366605	0.092690	0.782657	0.014269
straight_flow	l2	concat	2	0.946756	1.358986	0.093664	0.783543	0.015598
straight_flow	l2	film	0	0.955281	1.384740	0.090293	0.783236	0.014721
straight_flow	l2	film	1	0.950582	1.377675	0.093279	0.782857	0.010084
straight_flow	l2	film	2	0.953455	1.384223	0.094681	0.783440	0.011793
mip_one_step_integrate	l2	concat	0	0.946711	1.363403	0.090894	0.782186	0.014116
mip_one_step_integrate	l2	concat	1	0.939680	1.354019	0.098742	0.782859	0.014537
mip_one_step_integrate	l2	concat	2	0.943268	1.359842	0.098836	0.782085	0.017464
mip_one_step_integrate	l2	film	0	0.936463	1.350875	0.097907	0.783591	0.016643
mip_one_step_integrate	l2	film	1	0.934951	1.350210	0.098027	0.782793	0.018961
mip_one_step_integrate	l2	film	2	0.949605	1.375936	0.100431	0.782863	0.013540
regression	l2	concat	0	0.947243	1.364515	0.697390	0.772229	0.437604
regression	l2	concat	1	0.943320	1.355912	0.755140	0.789948	0.513411
regression	l2	concat	2	0.932974	1.350222	0.713418	0.785379	0.431160
regression	l2	film	0	0.934737	1.345641	0.116227	0.782274	0.030266
regression	l2	film	1	0.942952	1.356323	0.101304	0.783709	0.021264
regression	l2	film	2	0.947925	1.359665	0.110247	0.783946	0.026435
flow	l2	concat	0	0.994820	1.420759	0.194995	0.781539	0.097962
flow	l2	concat	1	0.988260	1.411795	0.178426	0.782324	0.100950
flow	l2	concat	2	1.012175	1.447642	0.209321	0.781843	0.080709
flow	l2	film	0	0.985680	1.400545	0.192682	0.782781	0.099145
flow	l2	film	1	0.972895	1.394550	0.183230	0.782909	0.104069
flow	l2	film	2	1.002097	1.418457	0.198922	0.780935	0.090557
mip	l2	concat	0	0.978513	1.401043	0.100064	0.781178	0.028088
mip	l2	concat	1	0.975253	1.399032	0.089928	0.782937	0.026146
mip	l2	concat	2	0.981090	1.405091	0.091783	0.782162	0.024452
mip	l2	film	0	0.977040	1.401209	0.095745	0.783356	0.033471
mip	l2	film	1	0.970299	1.394394	0.095651	0.782098	0.027525
mip	l2	film	2	0.966158	1.389465	0.101809	0.782240	0.034429

Table 8: Seed-wise results (trained with L2 loss) for proj task across modes and architectures

4.3 Projection Task Summary

The projection task reveals interesting trade-offs between different geometric constraints. Flow-based methods (straight flow, MIP, MIP one-step) significantly outperform regression on boundary enforcement and subspace adherence metrics. Straight flow trained with L1 loss achieves the best boundary metric (0.009769), demonstrating excellent manifold constraint satisfaction. MIP trained with L1 loss achieves strong subspace diagonal performance (0.086279). The strong performance of straight flow indicates that the ablation of time conditioning does not significantly impair geometric learning for projection tasks, and may even provide a beneficial inductive bias for boundary enforcement.

5 Lie Algebra Task

The Lie algebra task evaluates each method's ability to learn the tangent space structure of the manifold. Performance is measured using cosine similarity (measuring alignment with the true tangent space) and projection metrics (measuring orthogonality to the Lie algebra).

5.1 Results: Models Trained with L1 Loss

5.1.1 Averaged Results (L1 Loss Training)

Mode	Loss	Arch	L1	L2	Avg Cos Sim	Min Co
straight_flow	l1	concat	0.943249 ± 0.004702	1.611076 ± 0.004268	0.049991 ± 0.004831	-0.219003 ±
straight_flow	l1	film	0.946009 ± 0.012443	1.639597 ± 0.032941	0.046953 ± 0.005853	-0.230633 ±
mip_one_step_integrate	l1	concat	0.928837 ± 0.006404	1.600301 ± 0.007844	0.050264 ± 0.005242	-0.253179 ±
mip_one_step_integrate	l1	film	0.935447 ± 0.010640	1.608270 ± 0.024187	0.057269 ± 0.008716	-0.232594 ±
regression	l1	concat	0.901953 ± 0.005011	1.552268 ± 0.008365	0.047873 ± 0.004473	-0.258200 ±
regression	l1	film	0.924863 ± 0.006419	1.595575 ± 0.017823	0.055398 ± 0.006371	-0.258873 ±
flow	l1	concat	0.978049 ± 0.013907	1.572010 ± 0.018199	0.073508 ± 0.023696	-0.305286 ±
flow	l1	film	0.970517 ± 0.005058	1.547791 ± 0.022525	0.036151 ± 0.010965	-0.304738 ±
mip	l1	concat	0.954738 ± 0.004323	1.604481 ± 0.011878	0.048726 ± 0.007022	-0.224520 ±
mip	l1	film	0.943563 ± 0.000880	1.590165 ± 0.006388	0.055459 ± 0.002709	-0.256903 ±

Table 9: Averaged results (trained with L1 loss) for lie task across modes and architectures

5.1.2 Seed-wise Results (L1 Loss Training)

Mode	Loss	Arch	Seed	L1	L2	Avg Cos Sim	Min Cos Sim	Avg Projection	M
straight_flow	l1	concat	0	0.945057	1.614446	0.048890	-0.242538	0.077089	
straight_flow	l1	concat	1	0.947885	1.613727	0.056380	-0.203480	0.075994	
straight_flow	l1	concat	2	0.936803	1.605055	0.044702	-0.210989	0.071022	
straight_flow	l1	film	0	0.957625	1.664108	0.047827	-0.246626	0.062953	
straight_flow	l1	film	1	0.951650	1.661650	0.039388	-0.216429	0.062925	
straight_flow	l1	film	2	0.928754	1.593034	0.053645	-0.228845	0.064957	
mip_one_step_integrate	l1	concat	0	0.934652	1.607832	0.045056	-0.252297	0.076202	
mip_one_step_integrate	l1	concat	1	0.931943	1.603589	0.048297	-0.252077	0.084534	
mip_one_step_integrate	l1	concat	2	0.919917	1.589482	0.057437	-0.255164	0.077929	
mip_one_step_integrate	l1	film	0	0.948843	1.634568	0.059711	-0.233211	0.069804	
mip_one_step_integrate	l1	film	1	0.922813	1.576177	0.045584	-0.208270	0.090579	
mip_one_step_integrate	l1	film	2	0.934685	1.614064	0.066511	-0.256299	0.065687	
regression	l1	concat	0	0.896970	1.542609	0.044541	-0.265027	0.084472	
regression	l1	concat	1	0.908808	1.563013	0.044881	-0.249526	0.090134	
regression	l1	concat	2	0.900082	1.551181	0.054196	-0.260046	0.100895	
regression	l1	film	0	0.916776	1.576981	0.050868	-0.257508	0.071023	
regression	l1	film	1	0.932477	1.619609	0.050918	-0.275497	0.072887	
regression	l1	film	2	0.925337	1.590136	0.064408	-0.243612	0.074245	
flow	l1	concat	0	0.997657	1.596292	0.078983	-0.337836	0.204257	
flow	l1	concat	1	0.966928	1.552480	0.042139	-0.303568	0.250473	
flow	l1	concat	2	0.969561	1.567259	0.099401	-0.274454	0.197342	
flow	l1	film	0	0.967071	1.519345	0.038508	-0.316223	0.254242	
flow	l1	film	1	0.977669	1.574430	0.048244	-0.336648	0.219382	
flow	l1	film	2	0.966812	1.549599	0.021699	-0.261342	0.214051	
mip	l1	concat	0	0.948668	1.588198	0.052253	-0.209427	0.125433	
mip	l1	concat	1	0.958396	1.609043	0.038923	-0.240682	0.115411	
mip	l1	concat	2	0.957152	1.616201	0.055002	-0.223450	0.109288	
mip	l1	film	0	0.942383	1.594994	0.057495	-0.255973	0.100651	
mip	l1	film	1	0.944497	1.594364	0.051630	-0.266231	0.097163	
mip	l1	film	2	0.943809	1.581139	0.057251	-0.248505	0.105057	

Table 10: Seed-wise results (trained with L1 loss) for lie task across modes and architectures

5.2 Results: Models Trained with L2 Loss

5.2.1 Averaged Results (L2 Loss Training)

Mode	Loss	Arch	L1	L2	Avg Cos Sim	Min Co
straight_flow	l2	concat	0.929639 ± 0.006622	1.589545 ± 0.008867	0.041678 ± 0.006557	-0.247432 ±
straight_flow	l2	film	0.929072 ± 0.012654	1.624128 ± 0.031358	0.044650 ± 0.003235	-0.247986 ±
mip_one_step_integrate	l2	concat	0.912248 ± 0.002952	1.571566 ± 0.001903	0.044528 ± 0.003164	-0.253084 ±
mip_one_step_integrate	l2	film	0.915300 ± 0.003585	1.578945 ± 0.003461	0.049461 ± 0.006483	-0.252221 ±
regression	l2	concat	0.906051 ± 0.001520	1.556111 ± 0.007791	0.045938 ± 0.008341	-0.248872 ±
regression	l2	film	0.920272 ± 0.003046	1.580328 ± 0.011117	0.042800 ± 0.009726	-0.247779 ±
flow	l2	concat	1.005586 ± 0.005910	1.646773 ± 0.016019	0.025465 ± 0.023687	-0.236273 ±
flow	l2	film	0.964576 ± 0.004997	1.592203 ± 0.016294	0.052675 ± 0.016302	-0.200561 ±
mip	l2	concat	0.963025 ± 0.001503	1.633731 ± 0.005640	0.048842 ± 0.007701	-0.243064 ±
mip	l2	film	0.946999 ± 0.004336	1.597821 ± 0.011389	0.044041 ± 0.009503	-0.246990 ±

Table 11: Averaged results (trained with L2 loss) for lie task across modes and architectures

5.2.2 Seed-wise Results (L2 Loss Training)

Mode	Loss	Arch	Seed	L1	L2	Avg Cos Sim	Min Cos Sim	Avg Projection	M
straight_flow	l2	concat	0	0.937911	1.600918	0.050345	-0.223275	0.103216	
straight_flow	l2	concat	1	0.921701	1.579285	0.040200	-0.238940	0.080835	
straight_flow	l2	concat	2	0.929306	1.588431	0.034489	-0.280082	0.101421	
straight_flow	l2	film	0	0.911277	1.579829	0.040324	-0.255229	0.064148	
straight_flow	l2	film	1	0.939601	1.648067	0.048103	-0.230049	0.065189	
straight_flow	l2	film	2	0.936339	1.644490	0.045523	-0.258681	0.059915	
mip_one_step_integrate	l2	concat	0	0.908348	1.569144	0.045838	-0.257713	0.080490	
mip_one_step_integrate	l2	concat	1	0.915487	1.573793	0.047579	-0.245846	0.080008	
mip_one_step_integrate	l2	concat	2	0.912908	1.571761	0.040167	-0.255693	0.103283	
mip_one_step_integrate	l2	film	0	0.916353	1.575313	0.042242	-0.267718	0.068949	
mip_one_step_integrate	l2	film	1	0.910479	1.577919	0.057966	-0.236433	0.083541	
mip_one_step_integrate	l2	film	2	0.919068	1.583602	0.048174	-0.252511	0.071080	
regression	l2	concat	0	0.904620	1.552595	0.052269	-0.263586	0.105363	
regression	l2	concat	1	0.905376	1.566912	0.051393	-0.242916	0.118921	
regression	l2	concat	2	0.908156	1.548826	0.034153	-0.240114	0.112513	
regression	l2	film	0	0.916001	1.574363	0.056414	-0.272641	0.067651	
regression	l2	film	1	0.921921	1.595909	0.037693	-0.251239	0.067209	
regression	l2	film	2	0.922895	1.570713	0.034294	-0.219457	0.126687	
flow	l2	concat	0	1.011851	1.659473	0.044556	-0.258069	0.233652	
flow	l2	concat	1	0.997662	1.624177	-0.007919	-0.211200	0.249057	
flow	l2	concat	2	1.007245	1.656669	0.039756	-0.239551	0.241106	
flow	l2	film	0	0.970061	1.603030	0.064958	-0.229548	0.235126	
flow	l2	film	1	0.965692	1.604406	0.029638	-0.200304	0.200663	
flow	l2	film	2	0.957974	1.569174	0.063429	-0.171831	0.249046	
mip	l2	concat	0	0.965122	1.639996	0.045740	-0.258110	0.083920	
mip	l2	concat	1	0.961676	1.626323	0.059434	-0.235963	0.080952	
mip	l2	concat	2	0.962277	1.634873	0.041351	-0.235119	0.078445	
mip	l2	film	0	0.941752	1.593201	0.044801	-0.233324	0.126003	
mip	l2	film	1	0.946876	1.586769	0.032041	-0.271496	0.119464	
mip	l2	film	2	0.952370	1.613493	0.055281	-0.236151	0.104387	

Table 12: Seed-wise results (trained with L2 loss) for lie task across modes and architectures

5.3 Lie Algebra Task Summary

The Lie algebra task shows significant performance differences across training modes. Straight flow trained with L1 loss achieves the best average projection metric (0.063612), indicating superior orthogonality constraints to the Lie algebra. Flow matching achieves the highest average cosine similarity (0.073508), suggesting good alignment with tangent space directions, but suffers from poor projection metrics. MIP trained with L2 loss achieves the best maximum projection metric (0.116828), demonstrating consistent performance across all test samples. The strong performance of straight flow suggests that removing time conditioning may provide a beneficial inductive bias for learning geometric constraints in Lie algebra tasks.

6 Cross-Task Analysis

6.1 Mode Performance Comparison

Regression: The baseline regression approach performs strongly on reconstruction (best L1 error: 0.002288) but struggles significantly with geometric metrics in projection and Lie algebra tasks. This demonstrates that direct supervised learning excels at point-wise accuracy but fails to capture manifold structure.

Flow Matching: Shows mixed results across tasks. Achieves strong tangent space alignment (highest cosine similarity: 0.073508 in Lie algebra) but exhibits weaker performance on geometric constraint enforcement, particularly in projection tasks. Time conditioning appears beneficial for some aspects of geometric learning but not universally superior.

MIP (Two-Step): Achieves strong geometric constraint satisfaction, particularly for subspace diagonal errors in the projection task (0.086279). The two-step evaluation provides robustness across multiple metrics and demonstrates the value of the denoising objective for manifold learning.

MIP One-Step: Achieves the best reconstruction L2 error (0.003197) and competitive performance on geometric metrics. This suggests that stopping at the initial denoising step can provide superior manifold adherence while reducing computational cost at inference time. The single-step approach offers an excellent balance between reconstruction accuracy and geometric constraint satisfaction.

Straight Flow: Emerges as a strong performer on geometric metrics, achieving the best boundary metric (0.009769) in projection and best average projection metric (0.063612) in Lie algebra. The ablation study reveals that removing time conditioning does not impair geometric learning and may even provide beneficial inductive biases for certain manifold learning tasks. Straight flow demonstrates that simpler models without time conditioning can match or exceed the performance of more complex time-conditioned approaches.

6.2 Architecture Comparison

Both concatenation and FiLM architectures show competitive performance across tasks. FiLM architecture tends to achieve slightly better results on the most critical metrics (e.g., reconstruction L2 error, projection boundary), while concatenation shows more consistent performance across secondary metrics. The choice of architecture appears less critical than the choice of training paradigm.

6.3 Loss Function Analysis

Models trained with L2 loss generally perform better on L2 test metrics and achieve smoother manifold reconstructions, while L1-trained models excel at L1 metrics and boundary enforcement. The choice of training loss significantly impacts the learned manifold properties, with L2 loss favoring smooth reconstructions and L1 loss favoring sparse, boundary-respecting solutions. However, it should be noted that training flow-based methods with L1 loss lacks mathematical grounding, as the conditional flow matching objective is naturally derived for L2 loss.

7 Conclusions

This comprehensive evaluation across 60 configurations per task (5 modes $\times$ 2 losses $\times$ 2 architectures $\times$ 3 seeds) reveals several key insights:

1. **Flow-based methods excel at geometric learning:** Methods that model trajectories through the manifold (flow matching, MIP, MIP one-step, straight flow) consistently outperform regression on projection-based geometric metrics, demonstrating superior manifold structure learning.

2. **Regression maintains reconstruction advantage:** Despite weaker geometric performance, regression achieves the best L2 reconstruction error, highlighting a fundamental trade-off between point-wise accuracy and geometric constraint satisfaction.

3. **Straight flow emerges as a strong ablation:** Removing time conditioning from flow matching produces a simpler model that matches or exceeds the performance of time-conditioned approaches on key geometric metrics (projection boundary, Lie algebra projection). This suggests that time conditioning may not be necessary for effective geometric learning and that simpler inductive biases can be equally or more effective.

4. **MIP one-step offers excellent balance:** The single-step evaluation variant achieves top reconstruction performance while maintaining strong geometric metrics, providing an efficient alternative to traditional two-step approaches.

5. **Task-specific optimization matters:** No single configuration dominates all metrics. The optimal choice depends on whether the priority is reconstruction accuracy, geometric constraint satisfaction, or tangent space learning.

6. **Training loss and architecture choices significantly impact performance:** L2 loss favors smooth manifolds, L1 loss favors boundary enforcement, and FiLM architecture shows slight advantages on key metrics. However, training flow-based methods with L1 loss should be interpreted cautiously as it lacks mathematical grounding.

The results demonstrate that manifold learning via denoising diffusion provides a flexible framework where evaluation strategies (one-step vs two-step) and model complexity (time-conditioned vs straight flow) can be tailored to specific application requirements, offering valuable trade-offs between computational efficiency, model simplicity, and different notions of manifold quality.