Local Manifold Identification with Latent Linear Models and OT Flow

Sherman Khoo [*1] **Song Liu** [*1] **Mark Beaumont** [2]

Abstract

The manifold hypothesis states that high-dimensional data often concentrate near low-dimensional structures. Identifying these local manifold structures allows us to separate signals from ambient noise. In this paper, we study the problem of identifying local manifold structure from data. Given a query point y_0 from a dataset, our goal is to recover the geometry of the data manifold in a neighborhood of y_0 using a *pre-trained* optimal transport flow from the reference to the data distribution. First, we prove that the Brenier optimal transport map preserves manifold structure: the preimage of an m-dimensional data manifold is itself an m-dimensional manifold in the reference space. Second, motivated by this result, we propose a latent variable model that maps a linear model through the transport flow. We prove that the linear approximation error is significantly reduced by the optimal transport map, leading to a tight fit of the non-linear data manifold. Third, noting the intractability of the resulting likelihood, we deploy denoising Fisher score estimation — a recent development from simulation-based inference that learns the Fisher score over parameter–observation pairs — to perform likelihood-based inference effectively. Experiments on both synthetic and real-world datasets demonstrate the effectiveness of the proposed method.

1. Introduction

The manifold hypothesis—that high-dimensional data such as images inherently lie on a low-dimensional manifold—is the foundation of many machine learning algorithms (Fefferman et al., 2016; Pope et al., 2021; Loaiza-Ganem et al., 2024). Recovering this geometric structure allows us to extract signals from the data distribution while removing noise, thereby enabling downstream applications. While directly estimating the global manifold is difficult due to the curse of dimensionality (Narayanan and Mitter, 2010; Ozakin and Gray, 2009), accurately identifying its *local structure*, i.e., the tangent space at a query point, remains highly tractable and is sufficient for many applications. The challenge is to estimate this local structure accurately and efficiently(Aamari and Levrard, 2019; Kamkari et al., 2024).

The study of local manifold recovery has a long history, with the classic tool being local PCA (Kambhatla and Leen, 1997). This method collects the k nearest neighbors of a query point y_0, centers them, and takes the top m eigenvectors of the resulting covariance matrix as an estimate of the tangent space of an m-dimensional data manifold. While local PCA is simple, fast, and comes with finite-sample tangent-space recovery guarantees (Singer and Wu, 2012; Tyagi et al., 2013; Aamari and Levrard, 2019), it is also limited by a fundamental bias-variance trade-off. A linear approximation is only accurate when the neighborhood radius is small relative to the manifold's curvature (Tyagi et al., 2013); however, a small radius captures few samples, resulting in a highly volatile covariance estimate, especially in high dimensions. Consequently, practitioners are forced to trade off linearization bias against estimator variance.

Recent advances in generative modeling shed new light on this problem. Diffusion models (Ho et al., 2020; Song et al., 2021) and flow matching (Lipman et al., 2023; Tong et al., 2024) learn smooth, invertible maps from a simple isotropic reference distribution to the data distribution. Crucially, a growing body of work now shows that these maps implicitly *encode the manifold geometry of the data*: as the flow compresses the reference distribution onto a thin neighborhood of the data manifold $\mathcal{M}$, its Jacobian contracts along the normal directions of $\mathcal{M}$ much more strongly than its tangent directions (Stanczuk et al., 2024; Loaiza-Ganem et al., 2024; Fukumizu et al., 2026). These discoveries naturally raise the question: can we recover the local manifold structure from a *pretrained generative flow*? Among the broader class of generative models, we focus on Brenier Optimal Transport (OT) flow and its practical approximation, OT-Conditional Flow Matching (OT-CFM).

[*]Equal contribution [1]School of Mathematics, University of Bristol, UK [2]School of Biological Sciences, University of Bristol, UK. Correspondence to: Song Liu <song.liu@bristol.ac.uk>.

Published as a paper at the 1st FoGen workshop, ICML 2026, Seoul, South Korea, 2026. Copyright 2026 by the author(s).

In this work, we demonstrate that the *reference space* of OT flow contains a latent preimage that perfectly mirrors the geometry of the data manifold $\mathcal{M}$. By fitting a local linear model on this latent structure and pushing it forward through the pretrained OT map, we generate a curved, high-fidelity local approximation of $\mathcal{M}$ in the data space. Crucially, since the OT flow strongly contracts along normal directions (Fukumizu et al., 2026), this pushforward actively suppresses linearization errors in the reference space. As a result, our model tightly tracks $\mathcal{M}$ across extended neighborhoods where standard local PCA breaks down. To accurately model real-world observations, we further incorporate structured ambient noise orthogonal to the local tangent space, capturing the measurement errors that push data points off the underlying signal manifold.

While pushing this model through a non-linear OT map yields a highly accurate geometric approximation, it also makes evaluating the exact likelihood intractable. However, the model remains *trivial to simulate from*: given the parameters, we sample latent variables and push them through the flow to generate observations. This easy-simulation, hard-likelihood asymmetry is the hallmark of Simulation-Based Inference (SBI) (Cranmer et al., 2020). We adopt denoising score matching (Vincent, 2011), which has been applied to estimate the gradient of the likelihood function (Fisher score) from joint samples in SBI (Sharrock et al., 2024; Khoo et al., 2025). With the score in hand, likelihood-based inference reduces to highly efficient gradient ascent, bypassing the need to ever evaluate the likelihood itself.

To summarize, our core contributions are as follows:

- **Formalization of the OT latent space.** We show that the preimage of an m-dimensional data manifold under the OT map is itself an m-dimensional manifold. This formally defines "the latent twin" of a data manifold, with tangent spaces explicitly linked by the inverse Hessian of the Brenier potential (Proposition 1).
- **OT-based local manifold model.** We propose a latent linear model that approximates the preimage of the data manifold in the reference space. Pushing this subspace forward through a pretrained OT-CFM flow produces a naturally curved local approximation of the data manifold (Definition 1).
- **Rigorous approximation error bounds.** We prove our latent approximation achieves an asymptotically tighter leading-order error bound than the standard tangent-plane approximation used by local PCA. We demonstrate that this error reduction is directly proportional to the spectral anisotropy of the OT Jacobian (Corollary 1).
- **Scalable model inference algorithm.** To bypass the model's intractable likelihood, we adopt a simulation-based inference approach. By estimating the Fisher score from simulated observation-parameter pairs, we perform likelihood-based inference efficiently via gradient ascent on the learned score.

2. Background

2.1. Optimal Transport

Given two probability measures P_0 and P_1 on $\mathbb{R}^d$, the *Monge optimal transport problem* with quadratic cost seeks a map $G : \mathbb{R}^d \to \mathbb{R}^d$ that pushes P_0 forward to P_1 (denoted $G_\# P_0 = P_1$) while minimizing the transport cost:

$$\inf_{G:\, G_\# P_0 = P_1} \int \|z - G(z)\|^2 \, \mathrm{d}P_0(z).$$

Brenier's theorem (Brenier, 1991) establishes that when P_0 has a density with respect to Lebesgue measure, the unique minimizer takes the form $G = \nabla\varphi$ for a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$. The map $G = \nabla\varphi$ is called the *Brenier map*, and φ is the *Brenier potential*. The convexity of φ implies that the Jacobian $J_G = \nabla^2\varphi$ is positive semidefinite.

Computing the exact Brenier map over large datasets is intractable. Earlier approaches approximate the transport map through entropy-regularized linear programming via the Sinkhorn algorithm (Cuturi, 2013), parameterization of the Brenier potential φ with input convex neural networks (Makkuva et al., 2020; Amos et al., 2017), or numerical solution of the Monge–Ampère PDE (Gu et al., 2016). More recent works focus on training straight and continuous flows that closely approximate the OT map (Liu et al., 2022; Tong et al., 2024).

2.2. OT Flow Matching

Flow Matching (FM) (Lipman et al., 2023) learns a continuous-time flow that transports a reference distribution $P_0 = \mathcal{N}(0, I_d)$ to the data distribution P_1. The flow is defined by the ODE

$$\dot{z} = v_\theta(z, t), \quad z(0) = z_0 \sim P_0, \tag{1}$$

and the transport map is the time-1 map $G_\theta(z_0) := z(1)$. In general, the transport map of FM does not approximate the OT flow as the trajectory of the ODE may be strongly curved. Liu et al. (2022) proposes to train the flow repeatedly, to straighten flow, approximating the OT map.

Optimal Transport Conditional Flow Matching (OT-CFM) (Tong et al., 2024) improves upon standard FM by coupling the source and target samples via mini-batch optimal transport. The velocity field v_θ is learned by minimizing the conditional flow matching objective:

$$\mathcal{L}_{\text{OT-CFM}}(\theta) = \mathbb{E}_{t,(z_0,x_1)\sim\pi^*} \left\| v_\theta(z_t, t) - (x_1 - z_0) \right\|^2, \tag{2}$$

where $t \sim \mathrm{Uniform}(0,1)$, $z_t = (1-t)z_0 + t\,x_1$ is the linear interpolant, and π^* is the optimal transport coupling within each mini-batch, obtained by solving the assignment problem $\pi^* = \arg\min_\pi \sum_{ij} \pi_{ij} \|z_{0,i} - x_{1,j}\|^2$.

The OT coupling reduces the variance of the velocity field targets, encouraging straighter transport paths and closely approximating the true OT map. In our paper, we train G_θ using OT-CFM to approximate the true OT map in experiments.

2.3. Score Estimation and Likelihood-Free Inference

Classical latent-variable models such as PPCA (Tipping and Bishop, 1999) and GPLVM (Lawrence, 2004) admit tractable marginal likelihoods because of their Gaussian structure. In contrast, many modern implicit, latent variable models — including the model proposed in this paper — are easy to sample from, but have intractable likelihoods, making direct parameter inference difficult. Specifically, for a parameter ψ, one can simulate observations $Y \sim p(\cdot \mid \psi)$, but cannot evaluate the likelihood $p(Y \mid \psi)$ in closed form. This is the setting of simulation-based inference (SBI; Cranmer et al. 2020).

Recent score-based SBI methods avoid explicit conditional density estimation over high-dimensional observations. Instead, they leverage the technique of denoising score matching (DSM) (Vincent, 2011) to learn the parameter-space score function, from simulated parameter-observation pairs.

Neural posterior score estimation (Sharrock et al., 2024; Geffner et al., 2023) learns the posterior score across a diffusion noise schedule, and generates posterior samples via the reverse-time dynamics. Similarly, the *likelihood score* $s(\psi; Y) = \nabla_\psi \log p(Y \mid \psi)$ can also be directly targeted (Khoo et al., 2025; Jiang et al., 2026). The posterior and likelihood scores are connected by Bayes' rule, and if the prior score is known, the likelihood score can be recovered from the posterior score. Once either the posterior score or likelihood score is estimated, the MLE/MAP can be obtained through stochastic gradient ascent (Robbins and Monro, 1951).

3. Manifold Latent Variable Models

Problem Setting: Let $\{Y_i\}_{i=1}^N \subset \mathbb{R}^d$ be observations sampled from a distribution P_1 supported on or near an m-dimensional manifold $\mathcal{M} \subset \mathbb{R}^d$, where $m \ll d$. Given a query point $y_0 \in \mathbb{R}^d$ near $\mathcal{M}$, the goal is to identify the local tangent space $T_{y_0}\mathcal{M}$ and normal space $N_{y_0}\mathcal{M}$ at y_0.

3.1. Optimal Transport Pullback Preserves Manifold

Suppose G is a Brenier OT map that transports a d-dimensional reference distribution to this data distribution. We show that the preimage of $\mathcal{M}$ under G is also an m-dimensional manifold $\mathcal{M}_0$. In other words, the pullback of the data manifold $\mathcal{M}$ is also an m-dimensional manifold.

Proposition 1 (Manifold preservation under OT pullback)**.** *Let $P_0 = \mathcal{N}(0, I_d)$ and let P_1 be a probability measure on $\mathbb{R}^d$ with density f_1 satisfying $0 < \lambda \le f_1 \le \Lambda < \infty$ on compact sets (e.g., P_1 is a Gaussian-smoothed distribution supported near a smooth m-dimensional submanifold $\mathcal{M}$). Let $G = \nabla\varphi$ be the Brenier optimal transport map from P_0 to P_1, Then:*

1. *$\mathcal{M}_0 := G^{-1}(\mathcal{M})$ is a smooth m-dimensional submanifold of $\mathbb{R}^d$.*
2. *For each $y \in \mathcal{M}$ with $z = G^{-1}(y)$,*

$$T_z\mathcal{M}_0 = [\nabla^2\varphi(z)]^{-1} \cdot T_y\mathcal{M}. \tag{3}$$

The proof of this result can be found in Appendix A. Proposition 1 shows that a manifold in the data space will also have a "twin" manifold structure in the reference space, with the exact same dimension, whose tangent space is precisely defined under (3). See Figure 1 for an illustration.

3.2. OT Latent Variable Model

In this paper, instead of directly modeling the manifold $\mathcal{M}$ in the data space, we propose to model its latent twin in the reference space. We specify the latent manifold to be an affine subspace, which is then pushed through the OT map G, to produce a curved approximation of the data manifold $\mathcal{M}$. Now we present our generative model for an observation Y near a query point y_0:

Definition 1 (Manifold Latent Variable Model)**.** *Given a query point $y_0 \in \mathbb{R}^d$ and a pretrained OT flow $G_\theta : \mathbb{R}^d \to \mathbb{R}^d$, define $w_0 := G_\theta^{-1}(y_0)$. The generative model for observations $Y \in \mathbb{R}^d$ near y_0 is:*

$$\begin{aligned} Y &= G_\theta(w_0 + Wx) + \Pi_\perp(x, W)\,\varepsilon, \\ x &\sim \mathcal{N}(0, I_m), \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_d), \end{aligned} \tag{4}$$

where the components are defined as follows:

1. ***Signal direction.*** *$W \in \mathbb{R}^{d\times m}$ is the parameter to be estimated. Its columns span an m-dimensional affine subspace $\{w_0 + Wx : x \in \mathbb{R}^m\}$ in the reference space.*
2. ***Signal manifold.*** *The image $\mathcal{M}_W := \{G_\theta(w_0 + Wx) : x \in \mathbb{R}^m\}$ is a curved m-dimensional submanifold in data space that locally approximates $\mathcal{M}$ near y_0.*
3. ***Local tangent space.*** *At latent coordinate x, the tangent space of $\mathcal{M}_W$ at the signal point $G_\theta(w_0 + Wx)$ is spanned by the columns of*

$$T(x, W) = J_{G_\theta}(w_0 + Wx) \cdot W \in \mathbb{R}^{d\times m}, \tag{5}$$

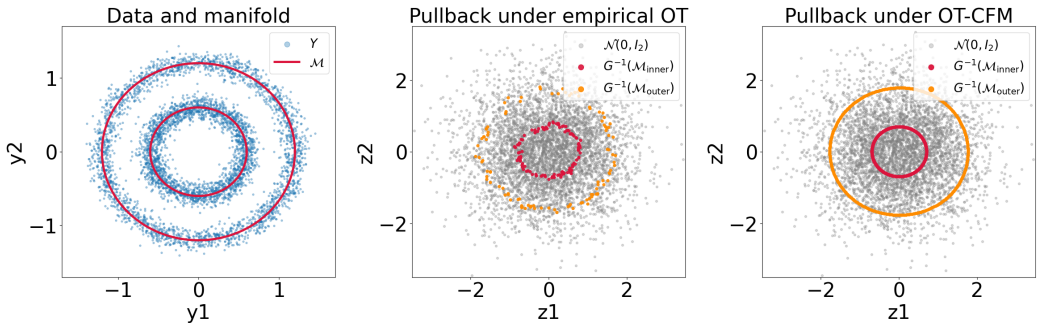

Figure 1. Illustration of $G^{-1}(\mathcal{M})$ for a noisy double-ring target $\mathcal{M}$ with reference $\mathcal{N}(0, I_2)$. Both empirical OT (center) and OT-CFM (right) reveal clear manifold structure in the preimage of $\mathcal{M}$ — specifically, two disjoint rings of the same dimension as $\mathcal{M}$, consistent with Proposition 1. Empirical OT (Hungarian on 5000×5000 samples) yields only a discrete coupling rather than a continuous map, so we plot the preimage using data points lying close to $\mathcal{M}$, producing a noisy point cloud. The continuous OT-CFM map, in contrast, can be inverted at any point of $\mathcal{M}$ via reverse-time ODE integration, yielding clean preimage circles. Gray points are reference samples.

where $J_{G_\theta}(z) = \partial G_\theta / \partial z$ is the Jacobian of the OT map.

4. ***Normal noise.*** *Let $Q(x, W) \in \mathbb{R}^{d \times m}$ be the orthonormal basis of $\mathrm{col}(T(x, W))$ obtained via QR decomposition. The noise is projected onto the $(d - m)$-dimensional normal subspace via*

$$\Pi_\perp(x, W) = I_d - Q(x, W)\, Q(x, W)^\top. \qquad (6)$$

Note that another tempting choice of latent model is

$$Y = G_\theta(w_0 + Wx + \epsilon), \qquad \epsilon \sim \mathcal{N}(0, \sigma^2 I_d),$$

where isotropic noise is added to the linear model and then pushed forward through the OT map. This model has a tractable likelihood: applying the change of variables rule gives

$$p(Y \mid W) = p_Z\big(G_\theta^{-1}(Y); W\big)\, \big|\det J_{G_\theta^{-1}}(Y)\big|,$$

where $Z := w_0 + WX + \epsilon \sim \mathcal{N}(w_0, WW^\top + \sigma^2 I)$ is Gaussian. Since the Jacobian determinant does not depend on W, MLE reduces to fitting PCA to the pullback local samples $\{G_\theta^{-1}(Y_{i,\mathrm{local}})\}$. We refer to this method as *Pullback PCA*.

Although computationally attractive, this latter model does not reflect the manifold data-generation assumption. In many applications, data points are sampled from a manifold $\mathcal{M}$ (via the latent coordinate x) and then perturbed by measurement errors or environmental fluctuations that displace them *away from* $\mathcal{M}$. Pushing reference-space error through the nonlinear G_θ produces a different noise structure that does not match how noise enters observations in practice.

This also explains our structured noise over the isotropic Gaussian noise $\mathcal{N}(0, \sigma^2 I_d)$ used in classical latent variable models such as PPCA (Tipping and Bishop, 1999). The projection $\Pi_\perp$ ensures that all on-manifold variation is attributed to the signal x and all off-manifold variation is attributed to the noise ε. This separation makes the model more interpretable — W controls the signal subspace and σ controls the off-manifold noise level — and ensures that the likelihood faithfully represents the assumed data-generating mechanism while remaining identifiable. We will benchmark our method against Pullback PCA.

Although Proposition 1 shows the preimage of $\mathcal{M}$ is also a manifold, it does not state the local manifold structure in the reference space can be approximated by a linear model. The advantage of pushing the linear model through an OT map remains unclear. In the next section, we show why fitting a local linear model in the reference space is much less restrictive than fitting a local linear model directly in the data space.

3.3. Local Linear Approximation in Reference or Data Space?

A classic approach to identifying the local manifold structure around y_0 is local PCA, which fits a linear model $y_0 + Wx$ in the data space using samples in the neighborhood of y_0; the tangent space at y_0 is then estimated as the column span of W. Our model (4), by contrast, fits a linear model $w_0 + Wx$ in the *reference space* to the preimage $G^{-1}(\mathcal{M})$ and pushes it through G_θ to obtain a curved approximation of the data manifold. The question is: which is better?

Intuitively, fitting a linear model in the reference space yields a better approximation. In generative modelling, the reference distribution is typically full-dimensional and isotropic (e.g., Gaussian), while the target distribution concentrates on a thin neighborhood of the data manifold $\mathcal{M}$. The OT map G_θ achieves this concentration by strongly compressing distances in directions perpendicular to $\mathcal{M}$ — the spectral gap of $\nabla^2 \varphi$ formalizes this (Fukumizu et al., 2026). Consequently, when we approximate the manifold by a linear model in the reference space, any approximation

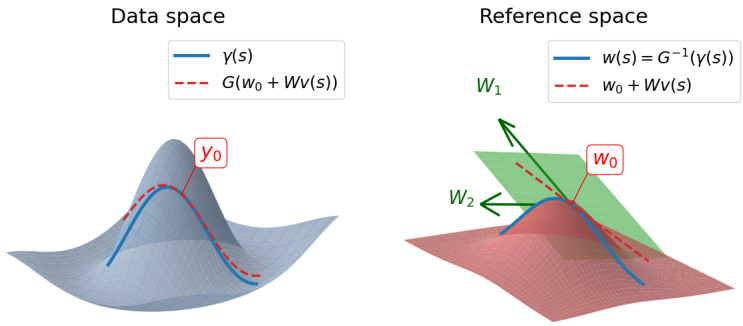

Figure 2. Geometric objects in Definition 2. $\gamma(s)$ is a curve on $\mathcal{M}$. $w(s)$ is its preimage on $\mathcal{M}_0$. The red dashed line on the right panel is $w_0 + Wv(s)$, which is the orthogonal projection of $w(s)$ onto the affine space $w_0 + T_{w_0}\mathcal{M}_0$. Its pushforward through G is the red dashed line on the left. $v(s)$ represents the coordinates on the green plane.

error in the normal direction is crushed by OT.

The following theory makes this intuition precise.

Definition 2 (Local geometry at y_0)**.** *Let $y_0 \in \mathcal{M}$ be the query point in the data space, and let $w_0 = G^{-1}(y_0) \in \mathcal{M}_0$ be its latent preimage. We denote the maximum extrinsic curvatures of $\mathcal{M}$ at y_0 and $\mathcal{M}_0$ at w_0 by κ and κ_0, respectively.*

Let $T \in \mathbb{R}^{d\times m}$ and $W \in \mathbb{R}^{d\times m}$ be matrices whose orthonormal columns span $T_{y_0}\mathcal{M}$ and $T_{w_0}\mathcal{M}_0$, respectively. For any smooth curve $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ with $\gamma(0) = y_0$, parameterized by arc length, define the following latent and ambient coordinate trajectories:

$$\begin{aligned} w(s) &:= G^{-1}(\gamma(s)) \in \mathcal{M}_0, \\ v(s) &:= W^\top\big(w(s) - w_0\big) \in \mathbb{R}^m, \\ u(s) &:= T^\top\big(\gamma(s) - y_0\big) \in \mathbb{R}^m. \end{aligned}$$

In plain words, $\gamma(s)$ is a smooth curve on $\mathcal{M}$, while $w(s)$ is its preimage on $\mathcal{M}_0$. $w_0 + Wv(s)$ is the orthogonal projection of $w(s)$ on the tangent plane $T_{w_0}M_0$, thus $v(s)$ can be seen as the coordinates on this tangent plane. Similarly, $u(s)$ acts as coordinates on the ambient tangent plane. We plot these objects in Figure 2.

Now, let us make an assumption about the pretrained OT map G_θ.

Assumption 1 (Manifold-aligned eigenvalue structure)**.** *The Hessian $\nabla^2\varphi(w_0)$ admits an orthonormal eigenbasis $\{e_k\}_{k=1}^d$ such that $\{e_k\}_{k=1}^m$ spans $T_{w_0}\mathcal{M}_0$ and $\{e_k\}_{k=m+1}^d$ spans $N_{w_0}\mathcal{M}_0$, with corresponding eigenvalues*

$$\lambda_k\big(\nabla^2\varphi(w_0)\big) = \begin{cases} \Theta(1) & k = 1, \dots, m, \\ O(\rho) & k = m+1, \dots, d, \end{cases} \tag{7}$$

for a small spectral parameter $\rho > 0$.

In plain words, this assumption says that the Jacobian of the OT map at w_0 behaves anisotropically: along directions normal to $\mathcal{M}_0$, the eigenvalues are of order ρ, so the map locally compresses the normal subspace by a factor proportional to ρ. The smaller ρ is, the stronger this normal contraction. Along tangent directions, by contrast, the eigenvalues are $\Theta(1)$: the map neither contracts nor expands appreciably, acting on the tangent subspace by a constant that does not depend on ρ. The ratio of normal to tangent eigenvalues is therefore of order ρ, formalizing the qualitative statement that the OT map crushes the reference samples onto the data manifold in directions transverse to it.

In the manifold-concentration regime studied by Fukumizu et al. (2026), ρ is connected to the off-manifold noise scale of P_1 and the assumption holds when the data is concentrated near $\mathcal{M}$. We have also validated this spectral gap on a synthetic dataset in Figure 7.

Corollary 1 (Local linear approximation through the OT flow)**.** *Suppose Proposition 1 and Assumption 1 hold. With the notation of Definition 2, the **latent tangent-plane approximation** incurs the approximation error*

$$\big\|\gamma(s) - G\big(w_0 + Wv(s)\big)\big\| = O\big(\rho\,\kappa_0\,s^2\big) + O(s^3), \tag{8}$$

*while the standard **ambient tangent-plane approximation** incurs the approximation error*

$$\big\|\gamma(s) - \big(y_0 + Tu(s)\big)\big\| = O(\kappa\,s^2) + O(s^3), \tag{9}$$

as $s \to 0$. Hence, whenever $\rho\kappa_0 \ll \kappa$, our model yields an asymptotically smaller linearization error.

The proof of the corollary can be found in Appendix B.

The corollary formalizes the intuition that the OT flow "does the curving for us". Initially, the pullback manifold $\mathcal{M}_0$ in the reference space is approximated linearly with curvature error $O(\kappa_0 s^2)$. When the pullback manifold is mapped back

to data space via J_{G_θ}, the OT map's normal contraction shrinks it by a further factor of ρ, where ρ is the spectral gap of $\nabla^2\varphi$ at w_0. The tangent-plane approximation in data space enjoys no such contraction, so its linear approximation error remains at $O(\kappa s^2)$. This is the theoretical justification for our latent variable model (4): by parameterising the signal as a linear subspace in reference space and pushing it through G_θ, we obtain a local approximation that is tighter than data-space tangent approximation such as local PCA whenever $\rho\kappa_0 \ll \kappa$.

The corollary predicts that the reference space linear approximation is valid over a neighborhood of radius $O(\rho^{-1/2})$ times larger than the tangent-plane approximation. By Fukumizu et al., ρ scales with the off-manifold noise scale, and is small when the data is distributed tightly around the manifold. The flow-based approximation therefore enjoys a substantially larger usable neighborhood, allowing the method to exploit more local data along the curved manifold for estimating W. We validate this prediction in Figure 3.

4. Likelihood Free Inference

To estimate the parameter W in our latent variable model (4), one might resort to likelihood-based inference. However, the marginal likelihood, i.e.,

$$p(Y \mid W) = \int p(Y \mid x, W)\, p(x) \mathrm{d}x$$

is analytically intractable because both the signal $G_\theta(w_0 + Wx)$ and the noise projection $\Pi_\perp(x, W)$ depends on the latent variable x through the OT map G_θ. This prevents the application of Gaussian identities. However, the model is *trivial to simulate from*: sample x and ε, compute $G_\theta(w_0+Wx)$ by integrating the flow, obtain $T(x, W)$ via Jacobian-vector Product (JVP), then Q via QR decomposition, and finally add the projected noise $\Pi_\perp(x, W)\varepsilon$. The intractability of likelihood and the simplicity of simulation motivate a *simulation-based likelihood inference*.

We adopt DSM (Vincent, 2011) to learn the score of the posterior distribution $\nabla_W \log p(W \mid Y)$ from simulated parameter–observation pairs. Place a Gaussian prior on the parameters: $W_{i,j} \sim \mathcal{N}(0, \sigma^2_{\text{prior}})$. We generate training data by sampling W from this prior and simulating Y from model (4), producing pairs (Y, W) from the joint distribution $p(Y, W) = p(Y \mid W)p(W)$. We then corrupt the parameters by adding Gaussian noise:

$$\widetilde{W}_{i,j} = W_{i,j} + \sigma_w \eta_{i,j}, \qquad \eta_{i,j} \sim \mathcal{N}(0, 1) \tag{10}$$

where $\sigma_w > 0$ controls the corruption level. A score model $s_\phi(Y, \widetilde{W})$ is trained to predict the denoising direction by minimizing

$$\mathcal{L}_{\text{DSM}}(\phi) = \mathbb{E}_{W,Y,\eta} \left\| s_\phi(Y, \widetilde{W}) + \frac{\eta}{\sigma_w} \right\|^2. \tag{11}$$

By the classical result of Vincent (2011), the minimizer of (11) satisfies

$$s_{\phi^*}(Y, \widetilde{W}) = \nabla_{\widetilde{W}} \log p_{\sigma_w}(\widetilde{W} \mid Y), \tag{12}$$

the score of the noise-smoothed posterior. As $\sigma_w \to 0$, this converges to the true posterior score $\nabla_W \log p(W \mid Y)$. The likelihood score can then be approximated by subtracting the known prior score via Bayes' rule:

$$\nabla_W \log p(Y \mid W) = \underbrace{s_{\phi^*}(Y, W)}_{\approx \nabla_W \log p(W|Y)} - \underbrace{\nabla_W \log p(W)}_{= -W/\sigma^2_{\text{prior}}}. \tag{13}$$

Note, instead of MLE, one may also perform Maximum a Posteriori (MAP) on W, then the subtraction of the prior score in (13) is optional. Once $\nabla_W \log p(Y \mid W)$ is estimated, a simple gradient ascent algorithm can be used to find the MLE $\widehat{W}$. We provide more details about the likelihood free inference procedure in Section C as well as the full inference algorithm in Algorithm 1.

5. Experiments

We evaluate our approach against two baselines: *local PCA* (Kambhatla and Leen, 1997) and *pullback PCA* (Section 3.2). Comparing against local PCA validates the asymptotic error reduction established in Corollary 1. Meanwhile, the comparison with pullback PCA isolates the importance of properly modeling structured noise in the ambient space. All methods estimate the linear basis $\widehat{W}$ from the same local neighborhood centered at the query y_0. The estimated manifolds are defined as follows:

- **Local PCA:** $\widehat{\mathcal{M}} = \{y_0 + \widehat{W}x \mid x \in \mathbb{R}^m\}$.
- **Pullback PCA and Ours:** $\widehat{\mathcal{M}} = \{G_\theta(w_0 + \widehat{W}x) \mid x \in \mathbb{R}^m\}$.

For our synthetic data experiments and ablation studies, we define the true data manifold $\mathcal{M}$ as a 1D ring of radius 1 for simplicity. We quantify the performance of each method by computing the mean and maximum orthogonal distances from the estimated manifold $\widehat{\mathcal{M}}$ to the true manifold $\mathcal{M}$ in the local neighborhood. In the real-world data experiments, we evaluate the quality of $\widehat{\mathcal{M}}$ through a held-out reconstruction error, which measures how closely $\widehat{\mathcal{M}}$ reproduces the neighbors of y_0 in a held-out dataset. Further details on experimental settings can be found at Section E and G.

5.1. Manifold Estimation Error with Growing Neighborhood

Corollary 1 states that the estimated local manifold of all methods would incur larger linearization error as the neighborhood increases. We test this claim quantitatively on the

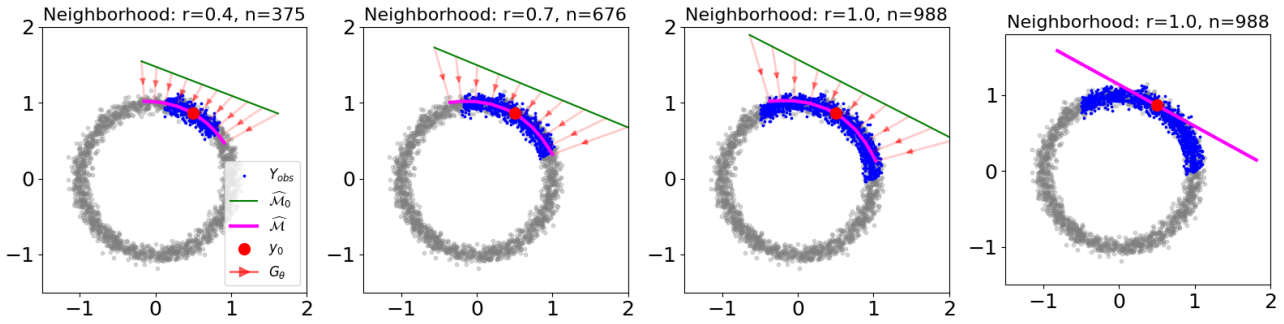

Figure 3. **Effective neighborhood of linear approximation in reference and data space.** Across panels, the local neighborhood used for fitting (blue, $\|y - y_0\| < r$) is enlarged from $r = 0.4$ to $r = 1.0$. The first three panels show a linear latent model $\widehat{\mathcal{M}}_0 = w_0 + \widehat{W}x$ (green) is fit by Algorithm 1 and pushed forward through the OT map G_θ (red arrows) to give the estimated manifold $\widehat{\mathcal{M}}$ (magenta). The estimated manifold tracks the true circular manifold across the full neighborhood at every radius, *including the largest one*, despite the latent model being strictly linear. This robustness is due to OT map's normal contraction: linearization error along the direction perpendicular to the manifold is strongly suppressed by G_θ, so a flat latent approximation produces an accurate curved data-space estimate even over wide neighborhoods. **Right:** local PCA fit in the data space at $r = 1.0$ (same neighborhood as the third panel), shown for comparison. Without an OT map to absorb normal error, the linear fit is rigid and fail to track the true manifold on both ends of the manifold. See Figure 6 for a more detailed comparison and Figure 4 for a quantitative comparison.

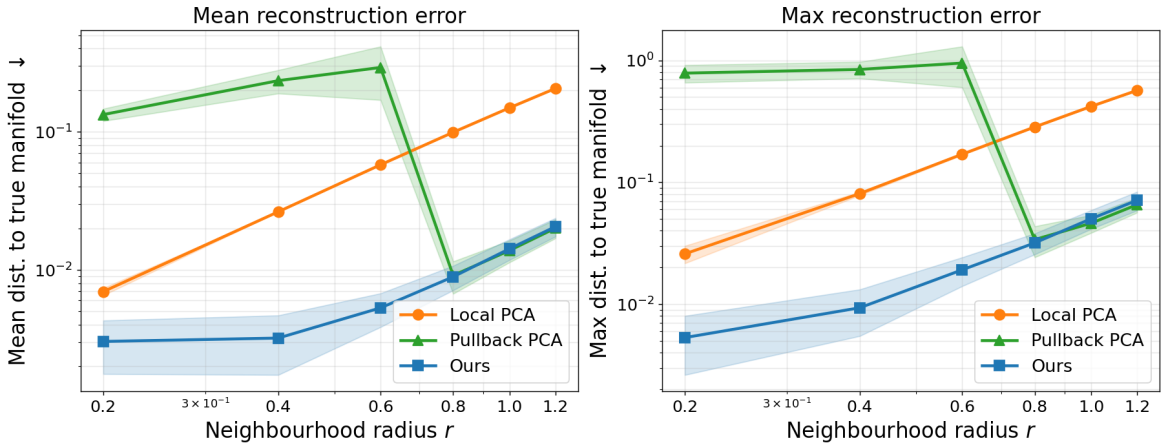

Figure 4. Reconstruction error vs. neighborhood radius r on a noisy unit circle embedded in a 2D space. **Left**: mean error; **right**: max error. Bands show std. over 12 query points $\times$ 10 seeds.

same synthetic dataset used in the visualization in Figure 3. It can be seen in Figure 4, for all methods, as the neighborhood radius (r) increases, the manifold reconstruction error increases. Moreover, Corollary 1 predicts that our method will have a linearization error that is lower than the local PCA, up to a constant that is determined by the spectral gap defined in Assumption 1. Indeed, we can see the reconstruction error of our method also increases, but remains lower than the local PCA up to a constant. We can also observe that when the neighborhood radius is small, pullback PCA does not perform well. However, its error drops suddenly as the neighborhood increases. This is because when the neighborhood is small, the off-manifold noise overwhelms the local manifold structure. After the inverse OT, the signal, i.e., the preimage manifold structure is unidentifiable from the pullback samples. However, when the neighborhood is large, the local structure survives the inverse OT map, thus the pullback PCA can recover the correct tangent space of the preimage. In comparison, our model works robustly in both small and large neighborhood radius settings. In Appendix F, we further demonstrate our model's robustness to varying off-manifold noise levels (Figure 8) and flow training quality 7, consistently outperforming local PCA and pullback PCA.

5.2. Local Manifold Recovery on Real-world Images

We now move on to high-dimensional image datasets, which exhibit different notions of locality. For the MNIST digits dataset (mni, 1998), the locality is visual similarity within the same digit class. For the Frey face dataset (Roweis and Saul, 2000), since images are collected from a single video,

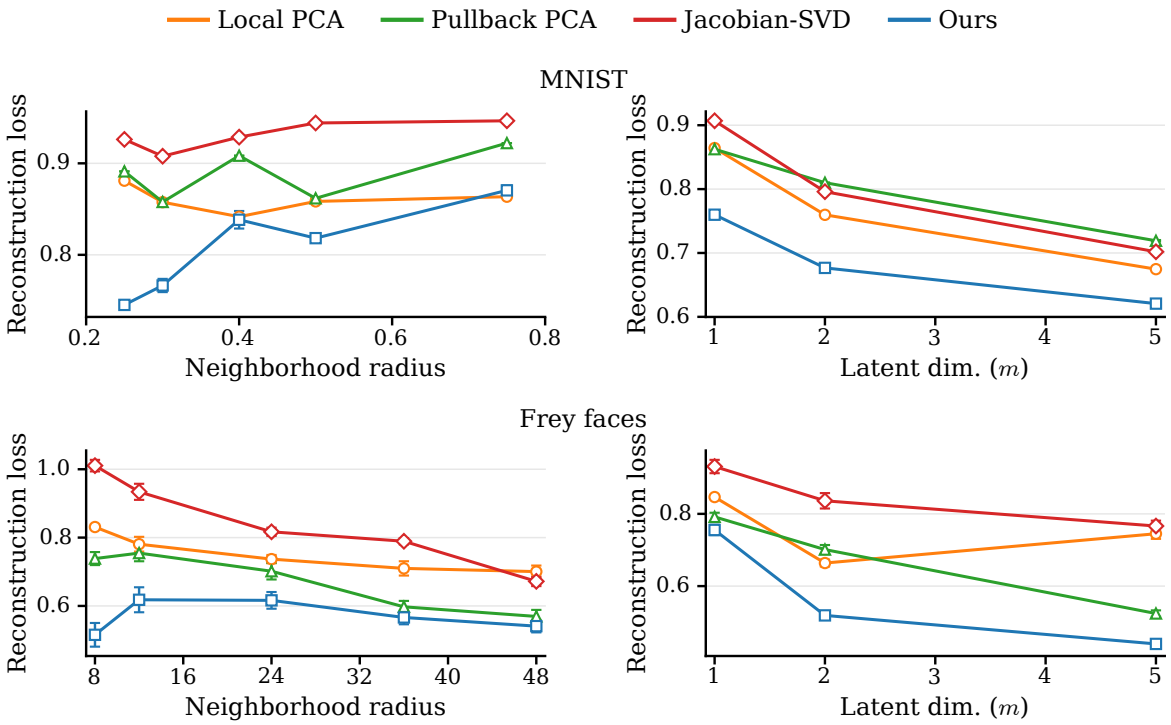

Figure 5. Held-out reconstruction loss with increasing neighborhood size and latent dimension on MNIST and Frey Faces dataset, in top and bottom row respectively. (Lower is better, error bars are the SE across 10 seeds)

locality is represented by temporal contiguity. Finally, for the COIL-20 dataset (Nene et al., 1996), which contains various images captured from different viewpoint, we use this viewpoint angle as a notion of locality. For each dataset, we train an OT-CFM on the entire dataset, and estimate a local manifold around a small neighborhood of a query or reference image y_0. In addition to the baseline methods discussed previously, we also include a local manifold estimated from the singular directions of the Jacobian evaluated at w_0. We provide further details of this performance metric and chosen local observations in Appendix G.

Figure 5 presents the quantitative results for the MNIST and Frey Faces datasets. Across both datasets, our proposed method consistently achieves a lower held-out reconstruction error than the baselines, particularly maintaining its advantage as the latent dimension m increases (right column). Notably, the purely local Jacobian-SVD baseline performs poorly across all settings, demonstrating that simply extracting the singular directions of OT flow's Jacobian at w_0 is insufficient at recovering the local geometry. We also observe a natural geometric degradation in the MNIST radius sweep (top-left): while our method provides a massive reduction in error at small to moderate radii, the error eventually converges toward that of local PCA at the largest radius ($r \approx 0.8$), showing the inherent limits of fitting a linear latent model over a highly nonlinear image neighborhoods. Further qualitative visualizations of these local manifold approximations are provided in Appendices G.1 and G.2, alongside additional experimental results for the COIL-20 dataset in Appendix G.3.

6. Conclusion and Limitations

We propose a novel latent variable model for identifying local manifold structure by modeling the the preimage of the manifold in the reference space, and pushing it through the OT map to get a curved manifold in data space. Although experiments on toy and real-world datasets have been promising, our approach has several limitations that present exciting avenues for future work. First, the validity of Assumption 1 needs further empirical verification. The middle panel of Figure 7 shows that it holds approximately for well-trained OT-CFM flows on low-dimensional toy targets. Whether the assumption holds for more general high-dimensional datasets is an open empirical question. Second, our generative model assumes the intrinsic dimension m of the data manifold is known a priori. Automatically determining the intrinsic dimension within our framework is a direction for future work. Third, despite the effectiveness of our method on real-world image datasets, estimating a high-dimensional W remains a challenging problem. How to leverage inductive bias to effectively compute the Fisher score is an interesting future direction.

Acknowledgements

We thank Prof. Kenji Fukumizu for helpful discussions and two anonymous reviewers for their insightful comments. SK was supported by the EPSRC Center for Doctoral Training in Computational Statistics and Data Science, grant number EP/S023569/1.

References

MNIST Database of Handwritten Digits, 1998.

Eddie Aamari and Clément Levrard. Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, 47(1):177–204, 2019. doi: 10.1214/18-AOS1685.

Brandon Amos, Lei Xu, and J Zico Kolter. Input convex neural networks. In *International Conference on Machine Learning*, pages 146–155. PMLR, 2017.

Silvere Bonnabel. Stochastic gradient descent on riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229, 2013.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.

Kenji Fukumizu, Wei Huang, Han Bao, Shuntuo Xu, and Nisha Chandramoorthy. Flow matching from viewpoint of proximal operators. *arXiv preprint arXiv:2602.12683*, 2026.

Tomas Geffner, George Papamakarios, and Andriy Mnih. Compositional score modeling for simulation-based inference. In *International Conference on Machine Learning*, pages 11098–11116. PMLR, 2023.

David Greenberg, Marcel Nonnenmacher, and Jakob Macke. Automatic posterior transformation for likelihood-free inference. In *International Conference on Machine Learning*, pages 2404–2414. PMLR, 2019.

Xianfeng Gu, Feng Luo, Jian Sun, and Shing-Tung Yau. Variational principles for Minkowski type problems, discrete optimal transport, and discrete Monge–Ampère equations. *Asian Journal of Mathematics*, 20(2):383–398, 2016.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.

Haoyu Jiang, Yuexi Wang, and Yun Yang. Likelihood-free inference via structured score matching. *arXiv preprint arXiv:2603.29054*, 2026.

Nandakishore Kambhatla and Todd K. Leen. Dimension reduction by local principal component analysis. *Neural Computation*, 9(7):1493–1516, 1997. doi: 10.1162/neco.1997.9.7.1493.

Hamidreza Kamkari, Brendan Leigh Ross, Rasa Hosseinzadeh, Jesse C Cresswell, and Gabriel Loaiza-Ganem. A geometric view of data complexity: Efficient local intrinsic dimension estimation with diffusion models. In *Advances in Neural Information Processing Systems*, volume 37, pages 38307–38354, 2024.

Sherman Khoo, Yakun Wang, Song Liu, and Mark Beaumont. Direct Fisher score estimation for likelihood maximization. In *Advances in Neural Information Processing Systems*, 2025.

Neil D Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *Advances in Neural Information Processing Systems*, volume 16, 2004.

John M Lee. *Introduction to Riemannian Manifolds*. Springer, 2nd edition, 2018.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow, September 2022. URL http://arxiv.org/abs/2209.03003. arXiv:2209.03003 [cs].

Gabriel Loaiza-Ganem, Brendan Leigh Ross, Rasa Hosseinzadeh, Anthony L. Caterini, and Jesse C. Cresswell. Deep generative models through the lens of the manifold hypothesis: A survey and new connections. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=a90WpmSi0I.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Ashok Vardhan Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason Lee. Optimal transport mapping via input convex neural networks. In *International Conference on Machine Learning*, pages 6672–6681. PMLR, 2020.

Hariharan Narayanan and Sanjoy Mitter. Sample complexity of testing the manifold hypothesis. In *Advances in Neural Information Processing Systems*, volume 23, 2010.

S.A. Nene, S.K. Nayar, and H. Murase. Columbia Object Image Library (COIL-20). In *Technical Report, Department of Computer Science, Columbia University CUCS-005-96*, Feb 1996.

Arkadas Ozakin and Alexander G Gray. Submanifold density estimation. In *Advances in Neural Information Processing Systems*, volume 22, 2009.

George Papamakarios and Iain Murray. Fast ε-free inference of simulation models with Bayesian conditional density estimation. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

George Papamakarios, David Sterratt, and Iain Murray. Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. In *International Conference on Artificial Intelligence and Statistics*, pages 837–848. PMLR, 2019.

Phil Pope, Chen Zhu, Ahmed Abdelkader, Micah Goldblum, and Tom Goldstein. The intrinsic dimension of images and its impact on learning. In *International Conference on Learning Representations*, 2021.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290 (5500):2323–2326, 2000.

Louis Sharrock, Jack Simons, Song Liu, and Mark Beaumont. Sequential neural score estimation: Likelihood-free inference with conditional score based diffusion models. In *International Conference on Machine Learning*, pages 44565–44602. PMLR, 2024.

Amit Singer and Hau-Tieng Wu. Vector diffusion maps and the connection Laplacian. *Communications on Pure and Applied Mathematics*, 65(8):1067–1144, 2012. doi: 10.1002/cpa.21395.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Jan Pawel Stanczuk, Georgios Batzolis, Teo Deveney, and Carola-Bibiane Schönlieb. Diffusion models encode the intrinsic dimension of data manifolds. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 46412–46440. PMLR, 2024. URL https://proceedings.mlr.press/v235/stanczuk24a.html.

Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B*, 61(3):611–622, 1999.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024.

Hemant Tyagi, Elif Vural, and Pascal Frossard. Tangent space estimation for smooth embeddings of Riemannian manifolds. *Information and Inference: A Journal of the IMA*, 2(1):69–114, 2013. doi: 10.1093/imaiai/iat003.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23 (7):1661–1674, 2011.

A. Proof of Proposition 1

Proof. First, for any Brenier transport map, the vector field is the gradient field of a convex function φ.

By Caffarelli's regularity theorem, with the densities f_0 and f_1 lower-bounded by a constant $\epsilon > 0$, $\nabla\varphi$ is smooth, $\nabla^2\varphi \succ 0$, and its eigenvalues are also lower-bounded by a constant $\lambda_{\min} > 0$. The Inverse Function Theorem states that the inverse transport $G^{-1} = (\nabla\varphi)^{-1}$ exists and is also smooth.

Since $\mathcal{M}$ is a smooth manifold, for an arbitrary point $y \in \mathcal{M}$, there exists an open neighborhood $U \subset \mathbb{R}^d$ and a smooth function $\psi : U \to \mathbb{R}^{d-m}$, such that $U \cap \mathcal{M} = \psi^{-1}(0)$.

Define $V := G^{-1}(U)$. Then we can rewrite the local preimage $V \cap \mathcal{M}_0$ as

$$\{z \in V : G(z) \in \mathcal{M}\} = \{z \in V : \psi(G(z)) = 0\} = (\psi \circ G)^{-1}(0).$$

The composition $\psi \circ G$ is a map from V to $\mathbb{R}^{d-m}$. We know that J_ψ has rank $d - m$ by definition, and $J_G = \nabla^2\varphi$ is invertible. Therefore, $J_{\psi\circ G} = J_\psi J_G$ has full row rank $d - m$ everywhere in V, which means that 0 is a regular value of $\psi \circ G$.

Applying the Preimage Theorem to $\psi \circ G$, we get that the level set $\{z \in V : (\psi \circ G)(z) = 0\}$ is a manifold with dimension $d - (d - m) = m$. Because this holds locally in a neighborhood V for every $z \in \mathcal{M}_0$, the entire preimage $\mathcal{M}_0$ is an m-dimensional manifold. This proves our first claim.

Furthermore, the Preimage Theorem states that the tangent space is given by the kernel of the Jacobian:

$$T_z\mathcal{M}_0 = \ker(J_{\psi\circ G}) = \ker(J_\psi J_G) = J_G^{-1}\ker(J_\psi) = J_G^{-1}T_y(\mathcal{M}),$$

where the third equality is due to the fact that $J_G = \nabla^2\varphi$ is an invertible linear operator (guaranteed by Caffarelli's theorem), and the fourth equality follows from the definition of ψ, which dictates that $T_y\mathcal{M} = \ker(J_\psi)$. This concludes the proof. □

B. Proof of Corollary 1

Proof. Our proof consists of three parts, derive a linear approximation error in the reference space, analyse the error after pushing through G_θ, derive and compare with linear approximation error in the data space. From now on, we simplify G_θ as G.

Step 1: Linear approximation in the reference space. The pullback curve $w(s) = G^{-1}(\gamma(s))$ lies on the smooth m-dimensional submanifold $\mathcal{M}_0$. By the second-fundamental-form characterisation of submanifold curvature (e.g., Lee 2018, Chapter 8), there exists a C^2 local coordinate map $v : (-\epsilon, \epsilon) \to \mathbb{R}^m$ with $v(0) = 0$ that absorbs the tangent component of $\ddot{w}(0)$, giving

$$w(s) = w_0 + Wv(s) + \tfrac{1}{2}N_0\, s^2 + O(s^3), \tag{14}$$

where $N_0 \in N_{w_0}\mathcal{M}_0$ is the normal component of $\ddot{w}(0)$. Because γ is parameterized by arc length ($\|\gamma'(0)\| = 1$), the initial speed of the pullback curve is $\|w'(0)\| = \|J_G(w_0)^{-1}\gamma'(0)\|$. By Assumption 1, the tangent eigenvalues of $J_G(w_0)$ are $\Theta(1)$, ensuring $\|w'(0)\| = \Theta(1)$. Consequently, the magnitude of the normal acceleration scales strictly with the latent curvature, yielding $\|N_0\| = \Theta(\kappa_0\|w'(0)\|^2) = \Theta(\kappa_0)$.

Step 2: Pushing the approximation through G. Rearranging Step 1 yields $w_0 + Wv(s) = w(s) - \frac{1}{2}N_0 s^2 + O(s^3)$. Taylor-expanding the map G around the true curve point $w(s)$, we obtain:

$$\begin{aligned} G(w_0 + Wv(s)) &= G\big(w(s) - \tfrac{1}{2}N_0 s^2 + O(s^3)\big) \\ &= G(w(s)) + J_G(w(s)) \cdot \big(-\tfrac{1}{2}N_0 s^2\big) + O(s^3) \\ &= \gamma(s) - \tfrac{1}{2}J_G(w(s))\, N_0\, s^2 + O(s^3), \end{aligned}$$

where the $O(s^3)$ term absorbs both the higher-order terms of the Taylor remainder (which are $O(\|N_0 s^2\|^2) = O(s^4)$) and the $O(s^3)$ displacement from Step 1. Assuming standard Caffarelli regularity for the Brenier potential ($\varphi \in C^3$), the Jacobian J_G is locally Lipschitz. Since $w(s) = w_0 + O(s)$, it follows that $J_G(w(s)) = J_G(w_0) + O(s)$. Substituting this yields:

$$\gamma(s) - G(w_0 + Wv(s)) = \tfrac{1}{2}J_G(w_0)\, N_0\, s^2 + O(s^3).$$

By definition, $N_0 \in N_{w_0}\mathcal{M}_0$. Because $\nabla^2\varphi(w_0)$ has eigenvalues of order $O(\rho)$ strictly on this normal subspace (Assumption 1), the matrix-vector product is bounded by:

$$\|J_G(w_0)N_0\| \leq O(\rho)\|N_0\| = O(\rho\kappa_0).$$

Therefore, $\|\gamma(s) - G(w_0 + Wv(s))\| = O(\rho\kappa_0 s^2) + O(s^3)$, which establishes (8).

Step 3: Comparison with the ambient tangent plane. Applying the identical second-fundamental-form expansion directly to γ on the ambient manifold $\mathcal{M}$ gives the standard tangent-plane approximation error: $\|\gamma(s) - (y_0 + Tu(s))\| = O(\kappa s^2) + O(s^3)$, which proves (9). Comparing the two bounds, the flow-based latent approximation achieves an asymptotically smaller leading-order error whenever $\rho\kappa_0 \ll \kappa$. □

C. Likelihood-Free Inference

C.1. Intractability of the Likelihood

To estimate the parameter W, we seek to perform maximum likelihood or maximum a posteriori estimation. However, the likelihood $p(Y \mid W)$ of model (4) is not available in closed form. Marginalizing over the latent variable x yields the integral

$$p(Y \mid W) = \int \underbrace{\mathcal{N}(x \mid 0, I_m)}_{p(x)} \underbrace{\mathcal{N}\big(Y \mid G_\theta(w_0 + Wx),\, \sigma^2\Pi_\perp(x, W)\big)}_{p(Y|x,W)} \,\mathrm{d}x, \tag{15}$$

where the conditional $p(Y \mid x, W)$ is a degenerate Gaussian whose covariance $\sigma^2\Pi_\perp(x, W)$ varies with x through the Jacobian of the nonlinear flow G_θ. Since both the mean $G_\theta(w_0 + Wx)$ and the covariance $\Pi_\perp(x, W)$ depend nonlinearly on x, standard Gaussian integration identities do not apply, and the integral has no obvious closed-form expression.

C.2. DSM Training Procedure

First, we sample pairs of $\{W\}$ and for each W, we sample a paired Y from model (4) using the pretrained flow G_θ, to form a paired set $\mathcal{D}_{\mathrm{DSM}} := \{(Y, W)\}$. We then start the training loop: at each iteration, sample a batch from $\mathcal{D}_{\mathrm{DSM}}$, corrupt the parameters as (10), and update ϕ by gradient descent on (11). Crucially, after simulating the training pairs $\mathcal{D}_{\mathrm{DSM}}$, the training procedure does not have to interact with G_θ nor marginalize over x, making the score gradient update highly efficient and entirely integration-free.

C.3. Gradient Ascent for MLE

Given a set of local observations $\{Y_i\}_{i=1}^n$ in a neighborhood of the query point y_0, we estimate the signal direction W by gradient ascent on the log-likelihood. At each step, the posterior score is obtained by querying the trained score network s_ϕ at the current parameter estimate and averaging over the local observations. The likelihood score is then recovered by (13), and the parameters are updated iteratively:

$$W^{(k+1)} = W^{(k)} + \gamma \left[\frac{1}{n} \sum_{i=1}^{n} s_\phi\Big(Y_i,\, W^{(k)}\Big) + \frac{W^{(k)}}{\sigma^2_{\mathrm{prior}}} \right], \tag{16}$$

where $\gamma > 0$ is the learning rate. Upon convergence, the estimated tangent space at y_0 is read off as $\hat{T} = \mathrm{col}(J_{G_\theta}(w_0) \cdot \widehat{W})$, and the orthonormal tangent basis is obtained via QR decomposition. The full inference procedure is summarized in Algorithm 1.

A key advantage of the amortized approach is the computational cost at inference time. Each gradient step in (16) requires only forward passes through the score network s_ϕ — a small neural network evaluated at n input pairs (Y_i, W). No integration, no root-finding for latent variables, and no Jacobian computation are needed during inference. The expensive operations — simulating (Y, W) pairs and integrating the flow, are performed once before training the score network. The Jacobian–vector product $J_{G_\theta}(w_0) \cdot \widehat{W}$ in the final step is computed only once at convergence to extract the tangent basis in data space.

Algorithm 1 Local Manifold Identification via Likelihood Free Inference

Require: Trained score network s_ϕ, pretrained flow G_θ, query point y_0, local observations $\{Y_i\}_{i=1}^n$, learning rate γ, prior std σ_{prior}

Ensure: Estimated tangent basis $\hat{Q} \in \mathbb{R}^{d\times m}$

$w_0 \leftarrow G_\theta^{-1}(y_0)$
Initialize $W^{(0)} \in \mathbb{R}^{d\times m}$ randomly
for $k = 0, 1, 2, \ldots$ until convergence **do**
  $\hat{g}_{\text{post}} \leftarrow \frac{1}{n}\sum_{i=1}^n s_\phi(Y_i, W^{(k)})$ {Average posterior score}
  $\hat{g}_{\text{lik}} \leftarrow \hat{g}_{\text{post}} + W^{(k)}/\sigma^2_{\text{prior}}$ {Subtract prior score $\nabla \log p(W)$, optional}
  $W^{(k+1)} \leftarrow W^{(k)} + \gamma\, \hat{g}_{\text{lik}}$
end for
$\hat{Q}, _ \leftarrow \text{QR}(J_{G_\theta}(w_0) \cdot \widehat{W})$ {Tangent basis in data space}
return $\hat{Q}$

D. Denoising Score Matching in SBI

Classical neural SBI methods learn the posterior (Papamakarios and Murray, 2016; Greenberg et al., 2019) or the likelihood (Papamakarios et al., 2019) directly via normalizing flows. More recent work targets the *score function* — the gradient of the log-density — rather than the density itself. The score can be learned via a least-squares training procedure known as *denoising score matching* (DSM; Vincent 2011). Given samples (θ, x) from the joint distribution, DSM corrupts the parameters as $\tilde{\theta} = \theta + \sigma\eta$ with $\eta \sim \mathcal{N}(0, I)$ and trains a network $s_\phi(x, \tilde{\theta})$ by minimizing

$$\mathcal{L}_{\text{DSM}}(\phi) = \mathbb{E}_{\theta,x,\eta}\left\| s_\phi(x, \tilde{\theta}) + \frac{\eta}{\sigma}\right\|^2. \tag{17}$$

The minimizer is attained at $s_{\phi^*}(x, \tilde{\theta}) = \nabla_{\tilde{\theta}} \log p_\sigma(\tilde{\theta} \mid x)$, the score of the noise-smoothed posterior. As $\sigma \to 0$, this converges to the score of the true posterior $\nabla_\theta \log p(\theta \mid x)$.

E. Experimental Settings

E.1. Evaluation Metric on Toy Dataset

To quantitatively assess the geometric fidelity of the estimated local manifolds, we measure the distance from a discretized grid of points on the predicted curves to the true underlying manifold (a circle of radius R in our toy experiments). To ensure a fair comparison, the predicted curves for all methods are explicitly rescaled so that they span the exact same arc length in the ambient data space. We evaluate performance using two key metrics:

- **Mean Reconstruction Error:** The average orthogonal distance between the points on the predicted curve and the true manifold. This metric captures the overall geometric alignment and average fidelity of the local approximation across the evaluated neighborhood.
- **Max Reconstruction Error:** The maximum orthogonal distance from any point on the predicted curve to the true manifold. This metric isolates the worst-case deviation, typically highlighting the extreme linearization error where the predicted curve diverges most sharply from the true manifold at the outer edges of the local neighborhood.

F. Additional Experimental Results

In this section, we add some additional figures and tables to supplement our experimental results in Section 5. All experiments are run on the same 1D ring dataset used in Figure 3 with two ambient dimensions.

First, we show an illustration of local PCA vs. our linear model on this dataset in Figure 6. It can be seen that our method traces the true manifold nicely, while the local PCA makes large approximation error at the both ends.

Second, we show how the quality of the pretrained flow affects the recovered flow. In Figure 7, the leftmost plot shows that as we train the OT-CFM model with more iterations, the manifold reconstruction error reduce significantly. However, after 2000 training epochs, the flow model may have over-fitted, leading to an increased reconstruction error.

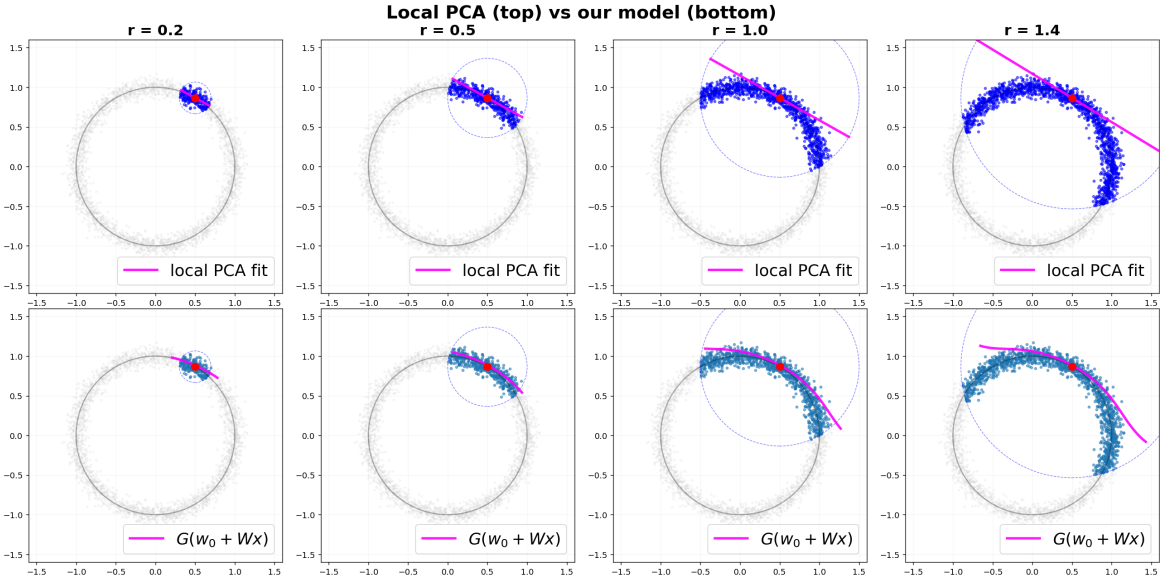

Figure 6. Comparison of Local PCA and our methods on 1D circle dataset. Our method has a significantly lower approximation error comparing to Local PCA (particularly at the two endpoints of the manifold).

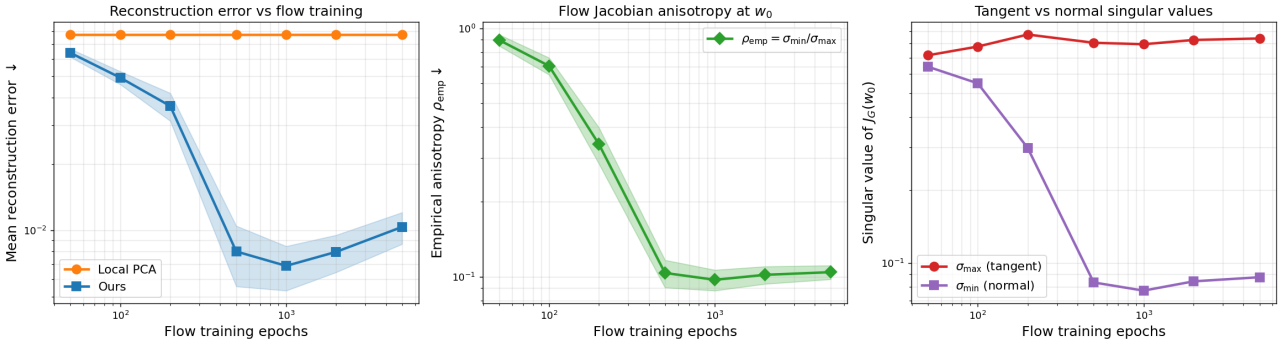

Figure 7. Ablation study, how the quality of the pretrained flow affects the recovered manifold.

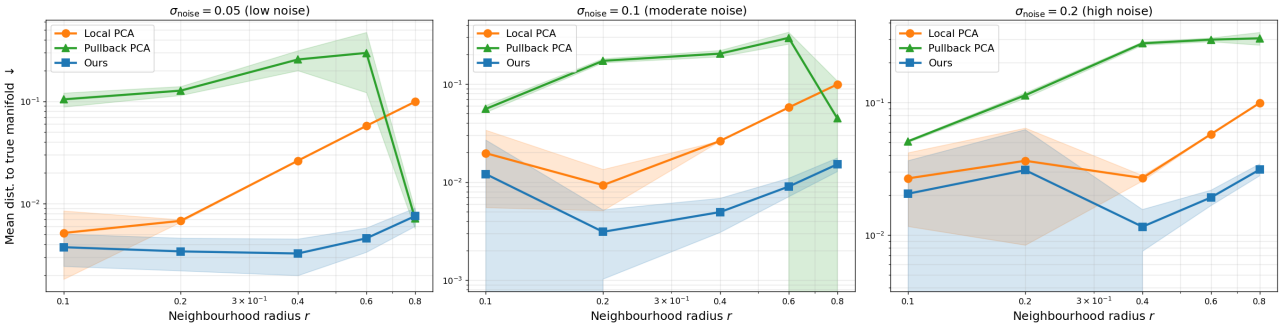

Figure 8. Ablation study, how the the dataset noise affects the recovered manifold.

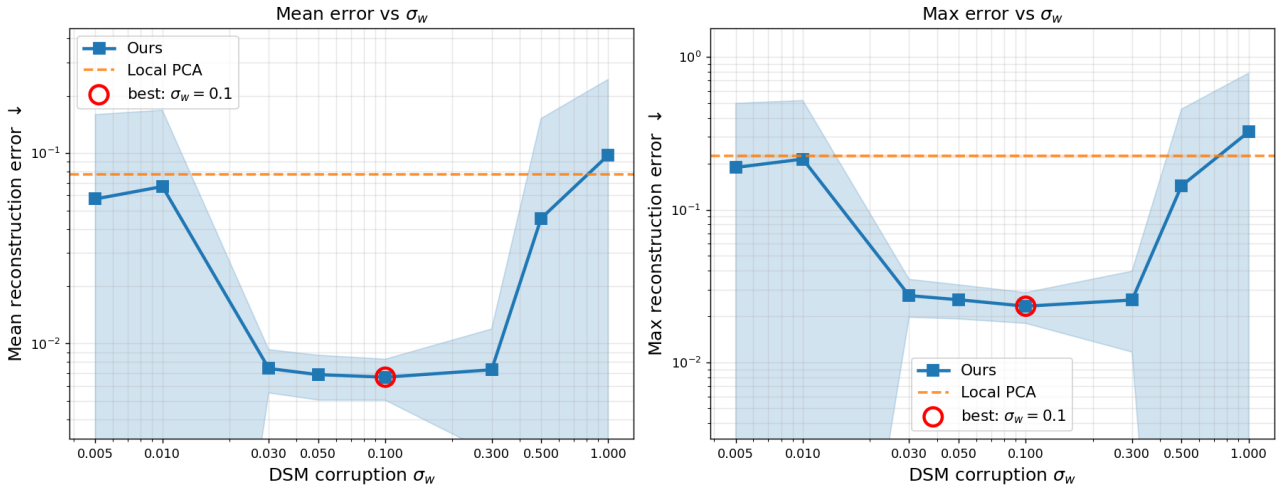

Figure 9. Ablation study, how the prior W affects the recovered manifold.

The center and the right plot show the spectral gap of tangent vs. normal space of G_θ increase as the training goes on. This is a direct validation of our Assumption 1.

Third, we compare the impact of the data noise on our model. We change the off-manifold noise to 0.05, 0.1 and 0.2 and plot the manifold reconstruction error. It can be seen our method remains the most robust in all three settings with the smallest reconstruction error.

Finally, we compare the impact of the σ_W, i.e., noise we add on to W when performing denoising score matching. It can be seen that there exists a global sweet spot, in which we attain the lowest reconstruction error. However, even a suboptimal choice, e.g., $\sigma_W = 0.01$, our method still achieves a lower manifold reconstruction error comparing to local PCA.

G. Real-world dataset experiments

We describe the full experimental setting used in Section 5.2. We first provide the settings that are shared across all three image datasets used (MNIST, Frey Faces, COIL-20), before specifying dataset-specific settings.

OT Flow Pre-training For each dataset we train a OT-CFM model (Tong et al., 2024) to convergence, and use the frozen model for subsequent local manifold learning. The reference distribution is $P_0 = N(0, I_d)$ in pixel space, and the velocity network $v_\theta(z, t)$ is prameterized by either a ConvNet or U-Net. Training was done with the AdamW optimizer (Loshchilov and Hutter, 2017) with a fixed step size of 2×10^{-4}.

Table 1. Configuration of pre-trained OT-CFM flows used in all experiments.

Dataset	Image shape	Architecture	Hidden dim.	Blocks	Batch	Epochs	Params
MNIST	$1 \times 28 \times 28$	ConvNet	192	10	128	250	7.2M
Frey Faces	$1 \times 28 \times 28$	U-Net	64	1	128	300	3.0M
COIL-20	$1 \times 128 \times 128$	U-Net	96	2	16	800	11.4M

Evaluation Protocol For each reference/query image, we generate a local neighborhood around the reference image, representing the local observations used for the local manifold estimation. This local observation dataset is split equally into a train/test split, and the test dataset is chosen such that it is interspersed equally across the local geometry.

Given the local manifold approximation $\mathcal{M}$ for each method, we follow the same evaluation pipeline. Firstly, the manifold approximation is rendered on a single, shared coordinate grid in image space. Then, for each held-out, test image, we locate the nearest rendered image point and calculate a reconstruction error, which is finally pooled across all test images.

Specifically, for the direction matrix W for each local manifold approximation method, we approximate the images by integrating $G_\theta(w_0 + W\alpha)$. We choose a grid of points $\{\alpha_1, \ldots, \alpha_K\}$, based on the pullback PCA score. This coordinate grid is shared across all method baselines, and we render the images along this grid. Then, for each held-out image y_i, we locate the nearest rendered grid image, $\hat{\alpha}(y_i) = \arg\min_{k \in \{1,\ldots,K\}} \|y_i - c(\alpha_k)\|_2$, where $c(\alpha_k)$ is the rendered image at the grid point. $\hat{\alpha}(y_i)$ represents how well the rendered approximation covers the held-out data manifold. Finally, we use reconstruction error $\|y_i - c(\hat{\alpha}(y_i))\|_2^2$, which is pooled and normalized across all test images y_i.

For each method we render the chart on the shared α-grid, producing K pixel-space images $\{c(\alpha_k)\}_{k=1}^K$, and assign each held-out image y to its nearest chart point $\hat{\alpha}(y) = \arg\min_k \|y - c(\alpha_k)\|_2$. We then report the following metrics, all computed on the held-out half of the neighborhood unless otherwise stated:

Baseline Methods For the PCA methods, which is done either on the data-space neighborhood points or on the reference space inverted neighborhood points, we obtain the manifold direction W using the first m principal components. For the Jacobian flow method, we take the leading m singular vector of the Jacobian $J_{G_\theta}(w_0)$.

For our proposed method, note that in the likelihood-free estimation of W, as opposed to the approach discussed in Appendix C.2, which uses a pretrained score network, allowing for amortization over the parameters W, we use a non-amortized approach which estimates a new likelihood-score for each parameter iterate $W^{(k)}$ in Algorithm 1. This approach is proposed as Fisher score matching MLE in Khoo et al. (2025). As opposed to a full neural network score parameterization, we use a linear score model with a handcrafted image feature map. Furthermore, because we are interested only in the direction of W, we use Riemannian SGD (Bonnabel, 2013) on a Stiefel manifold.

All experiments was ran with a Nvidia RTX 4090 GPU, with 24 GB of RAM.

G.1. MNIST dataset

We use the MNIST dataset (mni, 1998), with the reference image being the digit "3", as seen in the top left image of Figure 10. We calculate the radius to this image using lower-order image moments as a measure of distance. For our proposed method which leverages Fisher score matching (Khoo et al., 2025) to obtain the MLE, we use a proposal distribution for W which is a standard Gaussian with proposal scale of 0.5, and with 1024 proposals per iteration. We ran the Riemannian SGD with a total of 40 steps with a learning rate of 1×10^{-2}.

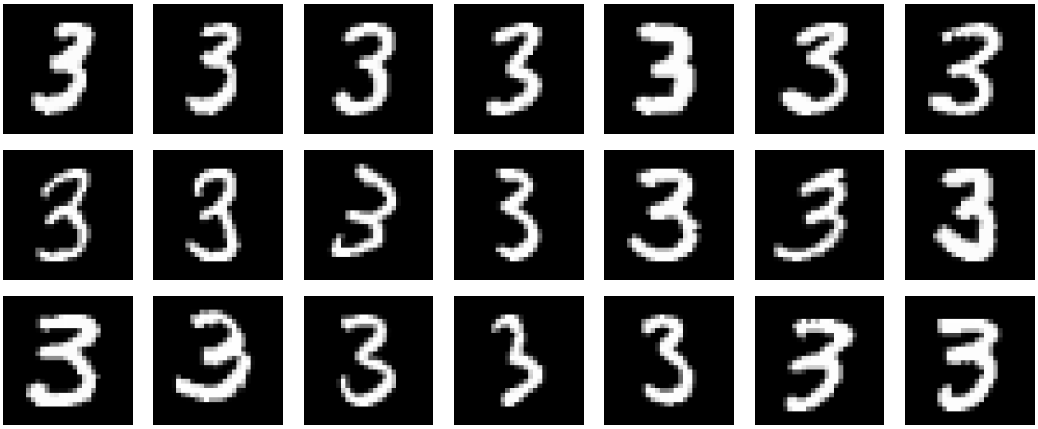

Figure 10. Local observations for MNIST dataset ($r = 0.25$). Reference image is in the top left corner.

As we can see in Figures 11 and 12, our proposed method is able to provide a better approximation to the local manifold, varying smoothly across grid points and staying on-manifold.

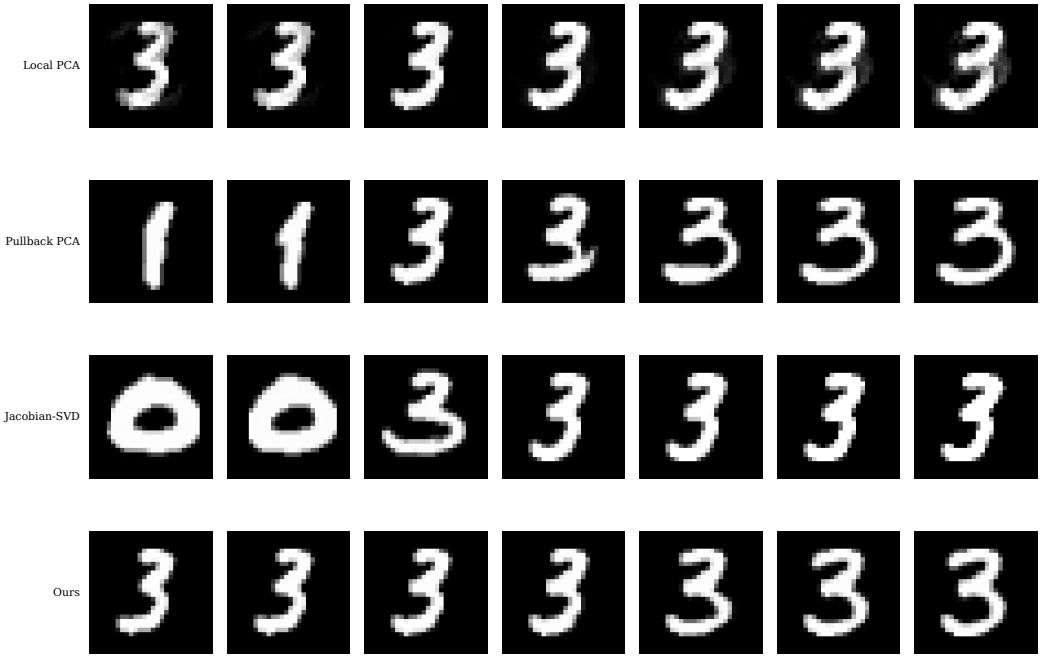

Figure 11. MNIST manifold approximation across grid points.

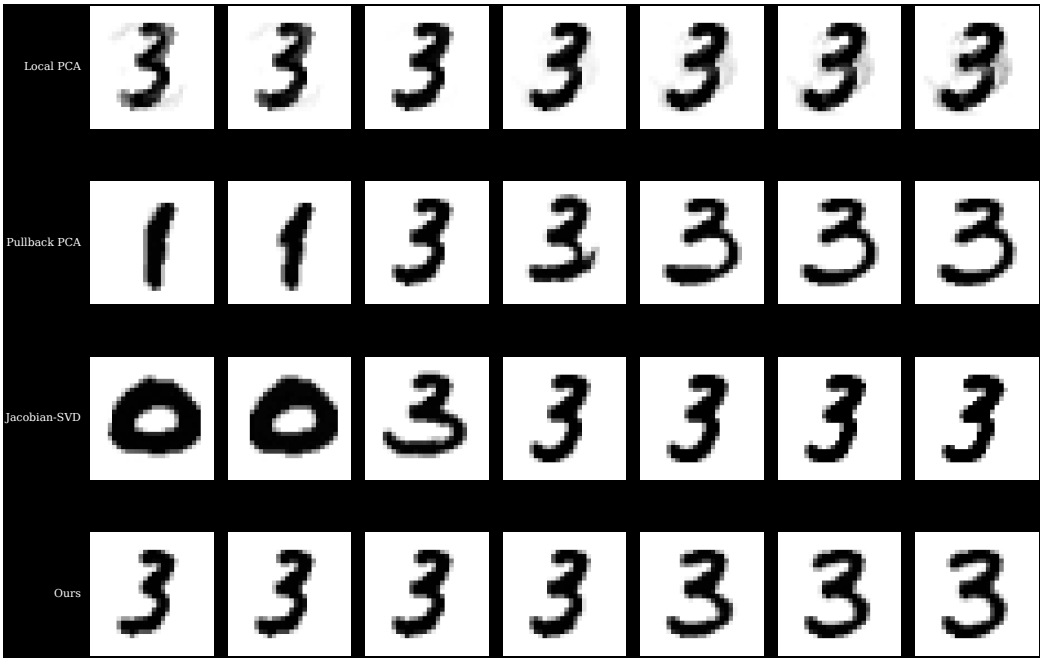

Figure 12. MNIST manifold approximation across grid points. (Inverted colors)

Similarly, Figure 13 reflects that our proposed method is able to provide better reconstruction error over a range of latent dimension and neighborhood radius. This is further reflected in the tangent direction of the various baseline methods in data-space, from Figure 14. As we can see, our proposed method is able to better capture the tangent direction representing the variation in the digits within the local neighborhood.

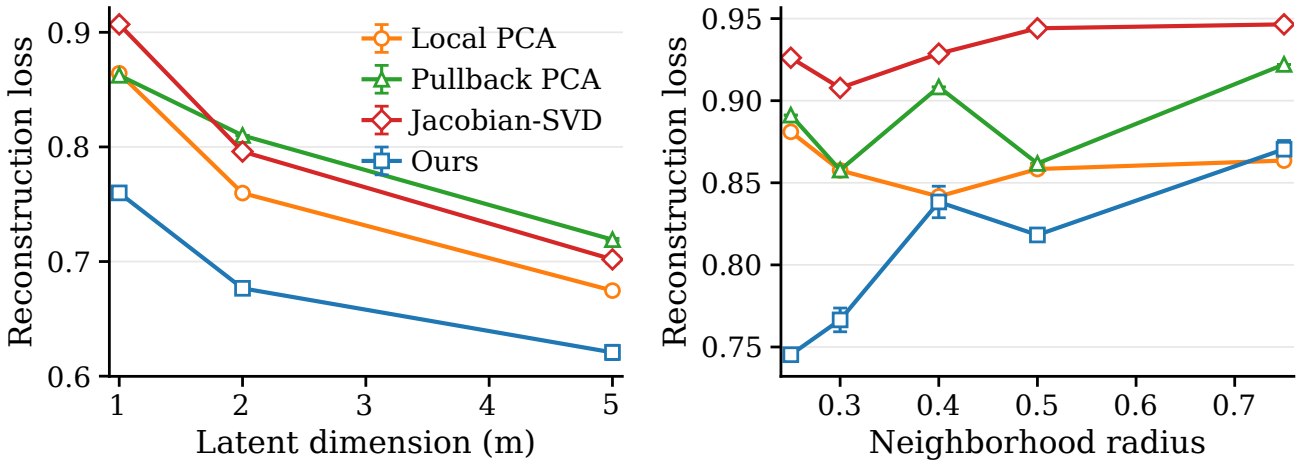

Figure 13. MNIST reconstruction error across latent dimension and neighborhood radius.

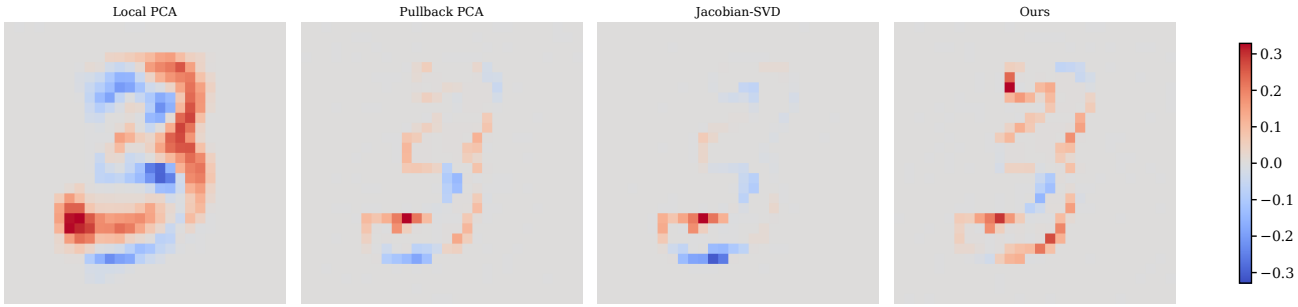

Figure 14. MNIST tangent direction.

In order to investigate the robustness of our evaluation to different choices of grid points, we rescale the grid points used for evaluation using a scaling factor, and provide the ablation in Figure 15. Generally, our proposed method is able to stay competitive and provide better local manifold approximation across a range of chosen grid points.

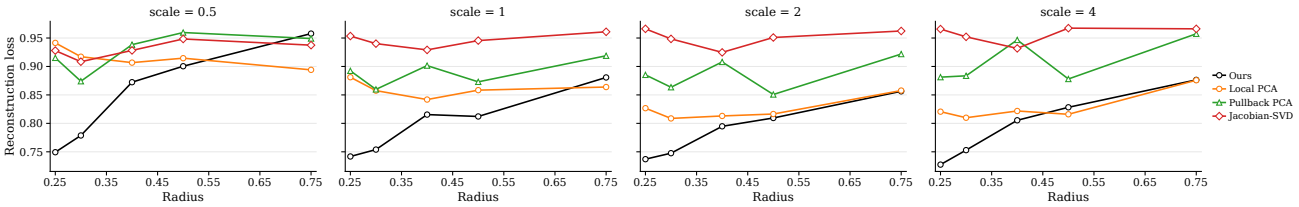

Figure 15. MNIST grid points rescaling.

G.2. Frey Faces dataset

We use the Frey Face dataset (Roweis and Saul, 2000), which consists of 2000 video frames. Each frame is assigned a temporal index t, and since the video is of an individual's face with different expression and pose, the temporal index t encodes a smooth manifold. Using a fixed anchor image at t_0, we define a temporal radius r_t around the anchor image, and select the local neighborhood based on $[t_0 - r_t, t_0 + r_t]$. The local observations are shown in Figure 16.

We use a proposal distribution for W which is a standard Gaussian with proposal scale of 0.1, and with 2048 proposals per iteration, and ran the Riemannian SGD with a total of 80 steps with a learning rate of 6×10^{-3}.

Figure 16. Local observations for Frey face dataset ($r = 12$)

From Figure 18, we can see that across neighborhood radius and latent dimensions, our proposed method provides a better reconstruction loss compared to baseline methods. From Figure 17, we can see that the local PCA fails to provide a good representation of the local manifold, and the Jacobian-SVD approach seems to go off-manifold. On the other hand, our proposed method varies smoothly, and matches the variation of the local observation.

Figure 17. Frey face manifold approximation across grid points

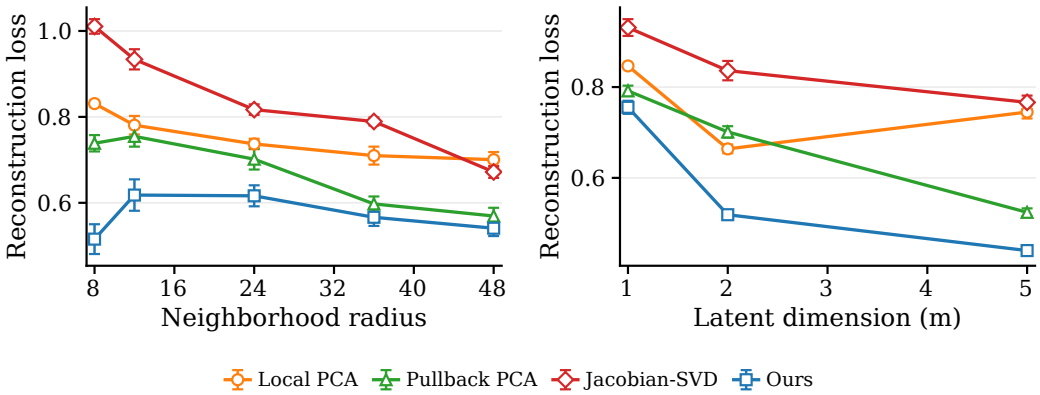

Figure 18. Frey face reconstruction error across latent dimension and neighborhood radius.

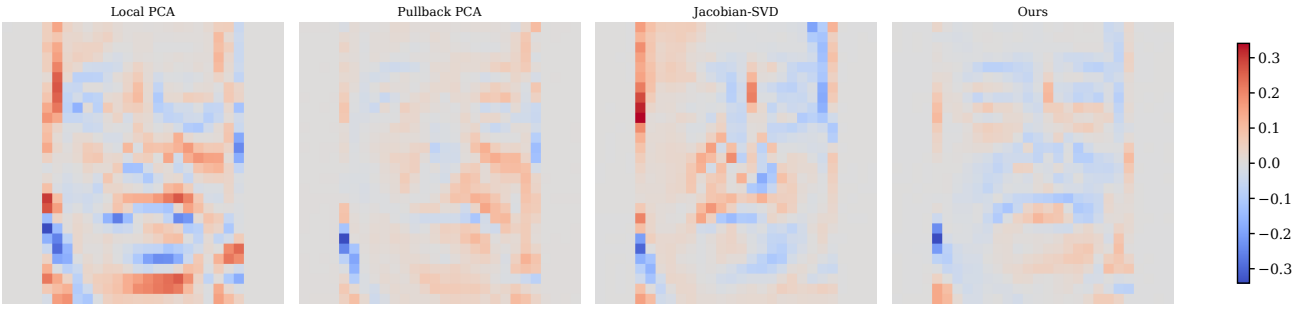

Figure 19. Frey face tangent direction

G.3. COIL-20 dataset

We use the COIL-20 (Nene et al., 1996), which consists of 1440 images which are constructed from 20 objects horizontally rotated across 72 viewpoints. The angular radius thus serves as a natural measure of locality. The local observations are provided in Figure 20.

We use a proposal distribution for W which is a standard Gaussian with proposal scale of 0.05, and with 512 proposals per iteration, and ran the Riemannian SGD with a total of 10 steps with a learning rate of 4×10^{-3}.

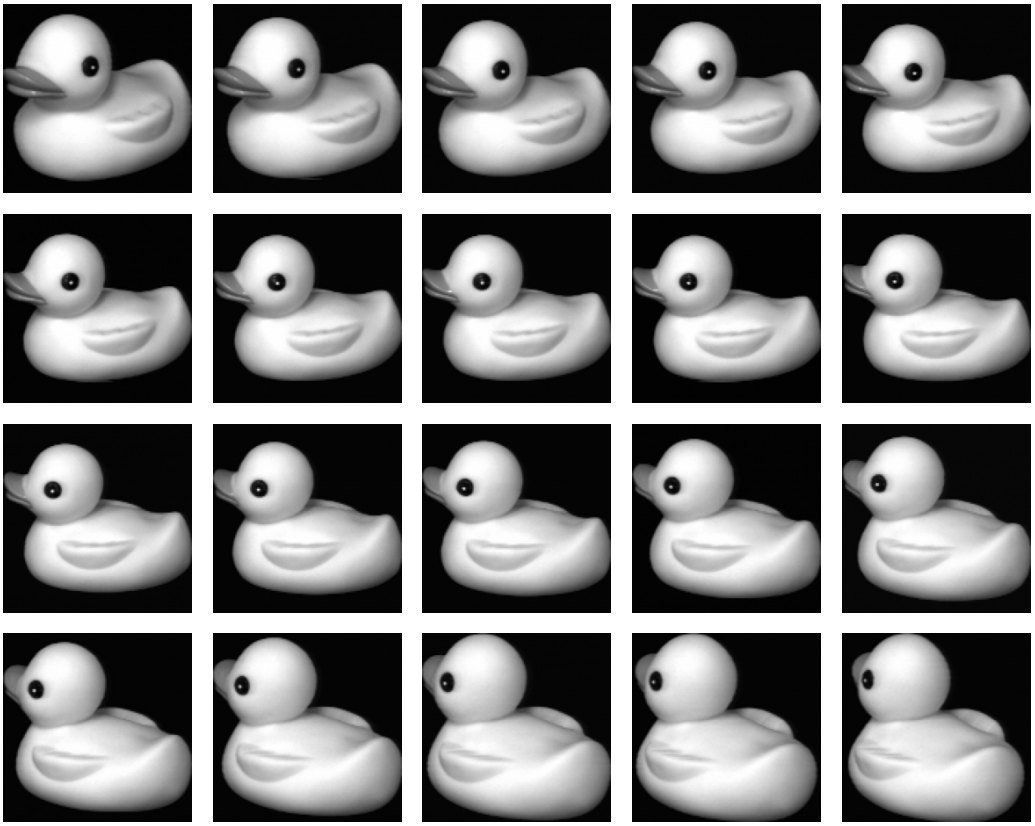

Figure 20. Local observations for COIL-20 dataset

Figures 21 and 23 provide a qualitative assessment of the local manifold approximation of our method, and show that our proposed method is able to obtain a smoother manifold compared to existing baselines. Similarly, this is shown in the quantitative benchmark in Figure 22 as well.

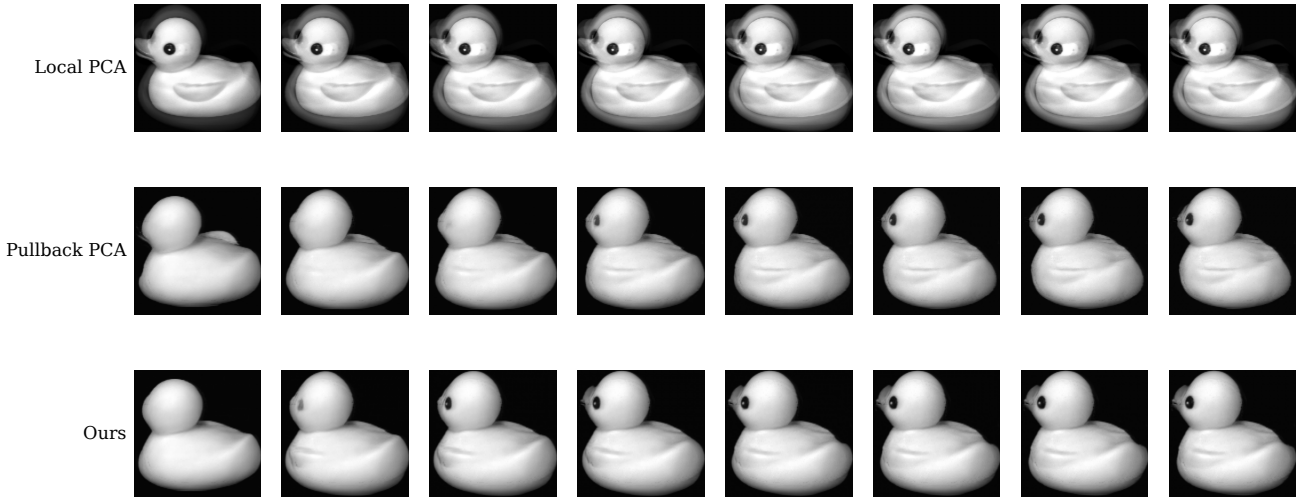

Figure 21. COIL-20 manifold approximation across grid points

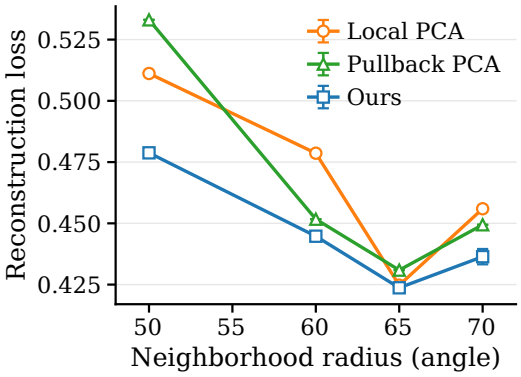

Figure 22. COIL-20 reconstruction error across neighborhood radius

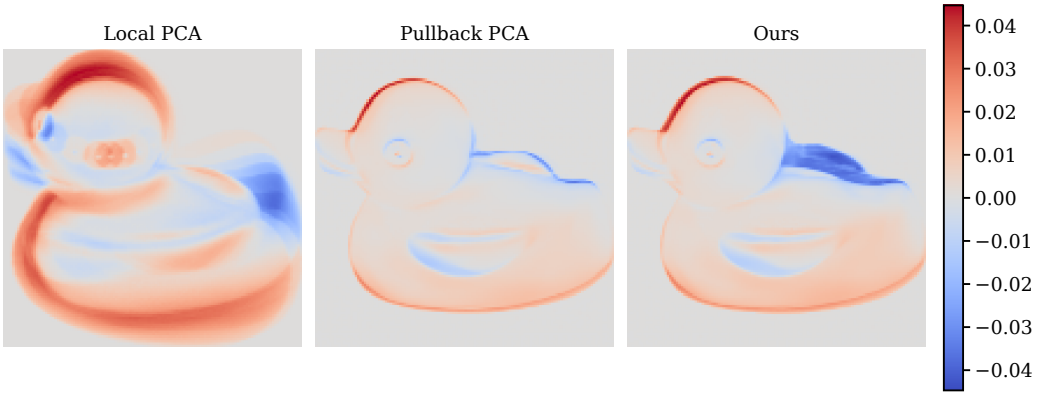

Figure 23. COIL-20 tangent direction