Diffusion Models for Inverse Problems on Riemannian Manifolds

Ivan Lee
ivan.lee@yale.edu
Yale

Xingyu Xu
xingyuxu@andrew.cmu.edu
CMU

Yuejie Chi
yuejie.chi@yale.edu
Yale

Abstract

Score-based diffusion models have emerged as powerful learned priors for Bayesian inverse problems in Euclidean spaces. However, many scientifically important datasets are intrinsically manifold-valued, and existing diffusion-based posterior samplers do not directly apply to these settings. We develop intrinsic posterior samplers for Riemannian manifolds, capturing two complementary diffusion-based paradigms: one separates prior sampling from likelihood correction, while the other combines prior and likelihood guidance in each reverse diffusion step. First, we introduce Manifold DPnP, a manifold generalization of diffusion plug-and-play (DPnP), in which each sampling iteration is reformulated as a manifold-valued SDE driven by Brownian heat flow. We then derive a Riemannian Tweedie's formula, enabling an extension of diffusion posterior sampling (DPS) to manifolds, which we call Manifold DPS. Experiments on protein backbone dihedral angles, real earthquake locations on the sphere, and random-walk trajectory reconstruction illustrate the promising performance of the two samplers for inverse problems on Riemannian manifolds.

1. Introduction

Inverse problems arise throughout science and engineering when one seeks to recover an unknown signal x from indirect, noisy, or incomplete observations y. A typical model takes the form

$$y = \mathcal{A}(x) + \xi,$$

where $\mathcal{A}$ is a known forward operator describing the sensing mechanism and ξ is measurement noise. Such problems are often ill-posed: many possible signals may be consistent with the same observation, and small perturbations in the data may lead to large changes in the reconstruction. A Bayesian formulation addresses this ambiguity by combining the likelihood $p(y \mid x)$ with a prior distribution $\rho^\star$ on the unknown signal, leading to the posterior

$$\pi_y(x) \propto p(y \mid x)\rho^\star(x).$$

Recent work has shown that score-based diffusion models provide a powerful class of learned priors in Euclidean spaces (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Ho et al., 2020; Song et al., 2021). These models learn the score of a family of noised data distributions and use the corresponding reverse diffusion process to generate samples. In inverse problems, the learned unconditional diffusion prior can be combined with the measurement likelihood to sample from a posterior distribution. This idea has led to several diffusion-based inverse problem solvers, including diffusion posterior sampling (DPS) (Chung et al., 2023), denoising diffusion restoration models (Kawar et al., 2022), and diffusion plug-and-play (DPnP) (Xu & Chi, 2024).

Despite this progress, existing diffusion-based inverse problem methods are formulated in Euclidean ambient spaces. Methods that use the word "manifold" in Euclidean inverse problems typically refer to a low-dimensional data manifold embedded in $\mathbb{R}^d$, rather than an intrinsic Riemannian state space (Bora et al., 2017; Chung et al., 2022). This is limiting in applications where many scientific data sets are intrinsically manifold-valued. Examples include protein structure

Accepted to FoGen 2026: Foundations of Deep Generative Models: Understanding Memorization, Generalization, and Reasoning, an ICML 2026 workshop (non-archival).

modelling (Shapovalov & Dunbrack, 2011), cell development and single-cell hierarchies (Klimovskaia et al., 2020), image recognition on non-Euclidean feature spaces (Lui, 2012), geological and earth science data (Karpatne et al., 2018; Peel et al., 2001), and high-energy physics (Brehmer & Cranmer, 2020), to name a few. Thus, there remains a gap between diffusion-based posterior sampling and intrinsic manifold-valued inverse problems.

In this paper, we study Bayesian inverse problems in which the unknown satisfies $x \in \mathcal{M}$ for a Riemannian manifold $\mathcal{M}$. Our goal is to sample from

$$\pi_y(x) \propto \exp(\mathcal{L}(x; y))\rho^\star(x), \qquad x \in \mathcal{M},$$

where $\mathcal{L}(x; y) = \log p(y \mid x)$ is a known log-likelihood and $\rho^\star$ is a learned prior on $\mathcal{M}$. We extend DPnP to manifolds by reformulating the algorithm as a manifold-valued SDE. A crucial ingredient in the SDE is the score of the likelihood function under Brownian heat flow, which we compute using the Bismut–Elworthy–Li formula (Bismut, 1984; Hsu, 2002). We also extend the DPS sampler to manifolds by introducing a Riemannian version of Tweedie's formula, from which we derive the manifold analogues of the key quantities used in the Euclidean DPS sampler. We demonstrate that both samplers achieve superior performance compared to a prior-free baseline on a variety of manifold inverse problems, including a data set of real earthquake locations and a protein-backbone-angle benchmark.

2. Background

2.1. Euclidean diffusion models

We first recall the Euclidean score-based diffusion framework. Let ρ_0 be a data distribution on $\mathbb{R}^d$. A forward diffusion gradually adds noise to the data, producing a family of smoothed densities $(\rho_t)_{0 \le t \le T}$. Score-based diffusion models use a forward SDE

$$\mathrm{d}X_t = f(t, X_t)\,\mathrm{d}t + g(t)\,\mathrm{d}W_t,$$

chosen so that ρ_T is close to a simple reference distribution, such as a standard Gaussian. The corresponding reverse-time SDE is

$$\begin{aligned}\mathrm{d}Y_t = &\left[-f(T-t, Y_t) + g(T-t)^2 \nabla \log \rho_{T-t}(Y_t)\right]\mathrm{d}t \\ &+ g(T-t)\,\mathrm{d}W_t.\end{aligned}$$

Thus sampling requires access to the score field

$$s_t(x) := \nabla \log \rho_t(x).$$

In practice this score is learned by a neural network $s_\theta(t, x)$ using score-matching objectives, and samples are generated by discretizing the reverse-time dynamics.

2.2. Riemannian notation

Let $(\mathcal{M}, g)$ be a connected, complete, d-dimensional Riemannian manifold without boundary. The exponential map $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ maps an initial tangent velocity $v \in T_x\mathcal{M}$ to the endpoint at time one of the geodesic starting from x with velocity v. When y lies away from the cut locus of x, we write $\log_x(y)$ for the inverse of $\exp_x$.

For a smooth function $f : \mathcal{M} \to \mathbb{R}$, the Riemannian gradient $\nabla_\mathcal{M} f$ is the vector field characterized by $\mathrm{d}f_x(v) = \langle \nabla_\mathcal{M} f(x), v\rangle_x$ for $v \in T_x\mathcal{M}$. The divergence of a vector field V is denoted by $\mathrm{div}_\mathcal{M} V$, and the Laplace–Beltrami operator is $\Delta_\mathcal{M} f = \mathrm{div}_\mathcal{M}(\nabla_\mathcal{M} f)$. All densities below are taken with respect to the Riemannian volume measure $\mathrm{vol}_\mathcal{M}$ and all Riemannian manifolds are assumed to be complete and compact.

2.3. Brownian heat flow and geodesic random walks

For data supported on a Riemannian manifold $(\mathcal{M}, g)$, the natural analogue of Euclidean Brownian noising is Brownian motion on $\mathcal{M}$ (Bortoli et al., 2022; Huang et al., 2022). Let $(B_t^\mathcal{M})_{t \ge 0}$ denote Brownian motion on $\mathcal{M}$. The manifold heat kernel $h_t^\mathcal{M}$ is the fundamental solution of the manifold heat equation. The convolution of a function $f : \mathcal{M} \to \mathbb{R}$ with $h_t^\mathcal{M}$ admits a representation as the action of the heat semigroup on f:

$$f * h_t^\mathcal{M}(x) = \mathbb{E}_x[f(X_t)] := \mathbb{E}[f(X_t) \mid X_0 = x],$$
$$\text{where} \quad \mathrm{d}X_t = \mathrm{d}B_t^\mathcal{M}.$$

Algorithm 1 Geodesic random walk (GRW)

Input: initial point $x_0 \in \mathcal{M}$; drift $b(t,x)$; diffusion $\sigma(t,x)$; horizon T; steps N
Set $\gamma = T/N$
for $k = 0$ **to** $N-1$ **do**
  $t_k = k\gamma$
  Sample $Z_{k+1} \sim N(0, I)$ in $T_{x_k}\mathcal{M}$
  $W_{k+1} = \gamma b(t_k, x_k) + \sqrt{\gamma}\,\sigma(t_k, x_k) Z_{k+1}$
  $x_{k+1} = \exp_{x_k}(W_{k+1})$
end for
Return: x_N

On a compact manifold, Brownian motion has the normalized Riemannian volume measure as its invariant distribution, making Brownian heat flow a natural forward noising process. For a Brownian noising process $\mathrm{d}X_t = \mathrm{d}B_t^{\mathcal{M}}$ with $X_0 \sim \rho_0$, the reverse process over $[0,T]$ is

$$\mathrm{d}Y_t = \nabla_{\mathcal{M}} \log \rho_{T-t}(Y_t)\,\mathrm{d}t + \mathrm{d}B_t^{\mathcal{M}},$$

where $X_t \sim \rho_t$. We write

$$s_t(x) := \nabla_{\mathcal{M}} \log \rho_t(x)$$

for the Riemannian score. Thus Riemannian diffusion models replace Euclidean noising by manifold diffusions and Euclidean scores by Riemannian scores. In practice, the score s_t is unknown and is approximated by a neural network $s_\theta(t,x) \in T_x\mathcal{M}$.

To simulate these dynamics while remaining on the manifold, we use geodesic random walks (GRWs). A GRW performs an Euler-type update in the tangent space and then maps back to the manifold using the exponential map (Jørgensen, 1975; Bortoli et al., 2022). For a diffusion of the form

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t^{\mathcal{M}},$$

where $b(t,x) \in T_x\mathcal{M}$ and $\sigma(t,x) : \mathbb{R}^m \to T_x\mathcal{M}$, the GRW sampler is given in Algorithm 1.

3. Manifold DPnP

Xu & Chi (2024) introduced the diffusion plug-and-play (DPnP) posterior sampler. Each step of the algorithm is a split Gibbs update, which alternates between a proximal-consistency step and a diffusion-denoising step, and is equivalent to solving an SDE.

Theorem 3.1 (DPnP as a heat flow)**.** *One step of DPnP with annealing parameter $\eta > 0$ is equivalent to solving the SDE*

$$\mathrm{d}X_t = \begin{cases} \nabla \log q_{\eta^2 - t}(X_t)\,\mathrm{d}t + \mathrm{d}B_t, & t \in [0, \eta^2], \\ \nabla \log p_{2\eta^2 - t}(X_t)\,\mathrm{d}t + \mathrm{d}B_t, & t \in (\eta^2, 2\eta^2], \end{cases}$$

where p_t and q_t are the probability densities of Brownian motion initialized from p_0 and $\exp(\mathcal{L}(\cdot; y))$, respectively.

Although Theorem 3.1 is stated above using Euclidean notation, the same split-Gibbs structure carries over to a complete Riemannian manifold $\mathcal{M}$ by replacing Euclidean Brownian motion with manifold Brownian motion $B_t^{\mathcal{M}}$, the Euclidean heat semigroup with the manifold heat semigroup, and ∇ with the Riemannian gradient $\nabla_{\mathcal{M}}$. Equivalently, we may define $q_t = \mathrm{e}^{\mathcal{L}(\cdot;y)} * h_t^{\mathcal{M}}$ and $p_t = p_0 * h_t^{\mathcal{M}}$, where $h_t^{\mathcal{M}}$ is the manifold heat kernel. The proof of the theorem is given in Appendix G.

Given estimates $\hat{s}_p$ and $\hat{s}_q$ of the scores $\nabla \log p_t$ and $\nabla \log q_t$, respectively, Algorithm 2 implements the DPnP sampler on manifolds using GRW.

3.1. BEL estimator for score of likelihood

A key ingredient in Algorithm 2 is $\nabla \log q_t$, which, in general, does not admit an explicit representation. As a solution to the manifold heat equation with initial condition given by the likelihood function, q_t admits the representation

Algorithm 2 Manifold DPnP

Input: manifold $\mathcal{M}$, steps N, schedule $\{\eta_k\}$, likelihood $\mathcal{L}$, observation y, scores $\hat{s}_p$, $\hat{s}_q$
Initialize $x_0 \sim \mathrm{Unif}(\mathcal{M})$
for $k = 0$ **to** $N-1$ **do**
  $x_{k+0.5} \leftarrow \mathrm{GRW}(x_k, \hat{s}_q(\eta_k^2 - \cdot, \cdot), \eta_k^2)$ // proximal
  $x_{k+1} \leftarrow \mathrm{GRW}(x_{k+0.5}, \hat{s}_p(\eta_k^2 - \cdot, \cdot), \eta_k^2)$ // denoise
end for
Return: x_N

$q_t(x) = \mathbb{E}_x[\exp(\mathcal{L}(X_t^{\mathcal{M}}; y))]$, where $X_t^{\mathcal{M}}$ is Brownian motion on $\mathcal{M}$. Computing the score of q_t requires taking the gradient of the above expression with respect to x. Given $f : \mathcal{M} \to \mathbb{R}$, the Bismut–Elworthy–Li (BEL) formula (Bismut, 1984; Hsu, 2002) provides the following representation of the gradient of the heat equation solution:

$$\nabla \mathbb{E}_x \left[f(X_T^{\mathcal{M}}) \right] = \frac{1}{T} \mathbb{E}_x \left[f(X_T^{\mathcal{M}}) \int_0^T M_s \, dW_s \right]. \tag{1}$$

Here, $W_s \in \mathbb{R}^n$ is the stochastic anti-development of $X_t^{\mathcal{M}}$, U_t is the horizontal lift of $X_t^{\mathcal{M}}$, and $M_t \in \mathbb{R}^{n \times n}$ solves the ODE

$$\frac{d}{dt} M_t + \frac{1}{2} \widetilde{\mathrm{Ric}}_{U_t} M_t = 0, \qquad M_0 = I. \tag{2}$$

For each $u \in O(\mathcal{M})$, $\widetilde{\mathrm{Ric}}_u$ represents the linear transformation $u^{-1} \circ \mathrm{Ric}_{\pi u} \circ u$. An in-depth treatment of horizontal lifts, stochastic development, and manifold Brownian motion may be found in Hsu (2002).

Note that (1) represents the gradient in Euclidean coordinates, so we left-multiply by U_0 to obtain the tangent vector. With BEL, the score of q_t is given by

$$\begin{aligned} \nabla \log q_t(x) &= \frac{\nabla_x \mathbb{E}_x[\exp(\mathcal{L}(X_t; y))]}{\mathbb{E}_x[\exp(\mathcal{L}(X_t; y))]} \\ &= \frac{\mathbb{E}_x \left[\exp(\mathcal{L}(X_t; y)) \int_0^t M_s \, dW_s \right]}{t \, \mathbb{E}_x[\exp(\mathcal{L}(X_t; y))]}. \end{aligned} \tag{3}$$

For many common manifolds, $\mathrm{Ric} = \lambda I_n$ for $\lambda \in \mathbb{R}$, in which case (2) is solved explicitly by $M_t = e^{-\lambda t/2} I_n$. We compile a list of common manifolds and their Ric operators in Table 1 (Appendix B). When M_t is available explicitly, $f(X_T^{\mathcal{M}}) \int_0^T M_s \, dW_s$ can be simulated without propagating a horizontal frame or solving the Ricci ODE (2); Algorithm 3 gives the resulting estimator. For general manifolds, we instead propagate the orthonormal frame and M_t along the path; see Algorithm 5 in Appendix A. To compute (3), take $f = \exp \mathcal{L}(\cdot\,; y)$.

4. Manifold DPS

In the framework of training-free guidance (Ye et al., 2024), the goal of guided diffusion is to sample from the distribution $\rho_0(x \mid y) \propto f_y(x) \rho^\star(x)$ given $f_y : \mathcal{X} \to \mathbb{R}_{\geq 0}$. For posterior sampling, as in our case, we take $f_y(x) = p(y \mid x) = \exp(\mathcal{L}(x; y))$. Define

$$\begin{aligned} f_y(x, t) &\coloneqq \int f_y(x_0) \rho_{0|t}(x_0 \mid x) \, dx_0 \\ &= \mathbb{E}\left[f_y(X_0) \mid X_t = x \right]. \end{aligned}$$

Given a stochastic process with initial distribution $\rho_0(\cdot \mid y)$, the score of its density at time t admits the decomposition $\nabla_x \log \rho_t(x \mid y) = \nabla_x \log \rho_t(x) + \nabla_x \log f_y(x, t)$. The DPS sampler (Chung et al., 2024) approximates the expectation with a point estimate $f_y(x, t) \approx f_y(\mathbb{E}[X_0 \mid X_t = x])$.

Algorithm 3 BEL estimator with explicit M_t

Input: $f : \mathcal{M} \to \mathbb{R}$; $x \in \mathcal{M}$; horizon t; step size h; samples B
Pick orthonormal frame u_0 over x; set $N = \lfloor t/h \rfloor$
for $i = 1$ **to** B **do**
  Set $x_0 = x$
  for $k = 0$ **to** $N - 1$ **do**
    Pick orthonormal frame u_k over x_k
    Sample $\zeta_k \sim N(0, I_n)$
    $x_{k+1} = \exp_{x_k}(u_k \sqrt{h}\, \zeta_k)$
  end for
  $S_i = f(x_N) \sum_{k=0}^{N} M_{kh}(\sqrt{h}\, \zeta_k)$
  $D_i = f(x_N)$
end for
Return: $u_0 \frac{\sum_{i=1}^{B} S_i}{t \sum_{i=1}^{B} D_i}$

4.1. Manifold extension

In the manifold setting, $\mathbb{E}[X_0 \mid X_t]$ is not defined. Instead, we define an analog

$$x_0(x,t) \coloneqq \exp_x \left(\mathbb{E}_{X_0 \sim \rho_{0|t}(\cdot|x)} \left[\log_x(X_0) \right] \right).$$

This choice is motivated by a "manifold Tweedie" formula in the following sense. On compact, complete Riemannian manifolds, Varadhan's asymptotics for the heat kernel (Bortoli et al., 2022, Equation (8)) implies

$$\lim_{t \to 0} t \nabla_x \log \rho_{t|0}(x \mid x_0) = \log_x(x_0) \tag{4}$$

whenever x_0 is away from the cut locus of x. Note that we can write

$$\begin{aligned} \nabla_x \log \rho_t(x) &= \frac{1}{\rho_t(x)} \int \rho_0(x_0) \nabla_x \rho_{t|0}(x \mid x_0)\, dx_0 \\ &= \int \rho_{0|t}(x_0 \mid x) \nabla_x \log \rho_{t|0}(x \mid x_0)\, dx_0 \\ &= \mathbb{E}_{X_0 \sim \rho_{0|t}(\cdot|x)} \left[\nabla_x \log \rho_{t|0}(x \mid X_0) \right]. \end{aligned}$$

Substituting Varadhan's approximation (4), we obtain the Tweedie-like approximation

$$\nabla_x \log \rho_t(x) \approx \frac{1}{t} \mathbb{E}_{X_0 \sim \rho_{0|t}(\cdot|x)} \left[\log_x(X_0) \right],$$

which implies that $x_0(x,t) \approx \exp_x(t \nabla_x \log \rho_t(x))$. We thereby obtain the estimator

$$\hat{x}_0(x,t) \coloneqq \exp_x(t s_t(x)) \tag{5}$$

of $x_0(x,t)$, where $s_t(x)$ is the score function. The *Manifold DPS* (MDPS) Algorithm 4 simulates the reverse SDE of the manifold Brownian noising process using the approximation $f_y(x,t) \approx f_y(\hat{x}_0(x,t))$ and outputs an approximate sample from $p_0(\cdot \mid y)$.

5. Experiments

We evaluate our samplers on three benchmarks: a sensor-based earthquake localization problem on the sphere S^2 using real seismic data; a path reconstruction problem on the product manifold $(S^2)^T$; and a protein-backbone-angle problem on the torus $\mathbb{T}^2$. An additional synthetic benchmark on the flat torus $\mathbb{T}^d$ is provided in Appendix F. On each benchmark we evaluate Manifold DPnP and Manifold DPS against a prior-free Metropolis-adjusted Langevin (MALA) baseline.

Algorithm 4 Manifold DPS (MDPS)

Input: step size $h > 0$; steps N; horizon T; prior score s_t; likelihood f_y; guidance strength ρ_t
Sample $x_0 \sim \mathrm{Unif}(\mathcal{M})$
for $t = 0$ **to** $N - 1$ **do**
  $v_t \leftarrow s_{T-t}(x_t) + \rho_t \nabla_{x_t} \log f_y(\hat{x}_0(x_t, T - t))$
  Sample $\zeta_t \sim N(0, I)$ in $T_{x_t}\mathcal{M}$
  $x_{t+1} \leftarrow \exp_{x_t}(hv_t + \sqrt{h}\,\zeta_t)$
end for
Return: x_N

5.1. The "two-point" problem on earthquake data

We conduct experiments on real earthquake data; see Appendix C for the train/test split and additional details. We consider the following inverse problem on the sphere. An earthquake occurs at an unknown location $x \in \mathbb{S}^2$, and a sensor located at $y \in \mathbb{S}^2$ records a scalar signal whose strength is maximal when $y = x$ and decays with the distance from the source. A natural model is $I(x, y) = \exp(\beta\langle x, y\rangle - \beta)$, where $\beta > 0$ controls the decay rate, so that $I(x, y) = 1$ when $x = y$ and decreases smoothly with distance.

Given a measurement $u \in \mathbb{R}$, we model the observation noise as $U \mid x, y \sim \mathcal{N}(I(x, y), \sigma^2)$, so the likelihood is $p(u \mid x, y) \propto \exp\left(-\frac{(u-I(x,y))^2}{2\sigma^2}\right)$. Suppose we have collected two measurements (y_1, u_1) and (y_2, u_2), and wish to recover the earthquake location x. Even in the noiseless case $\sigma = 0$, elementary geometry shows that this inverse problem has multiple solutions in general.

We therefore apply Manifold DPnP and MDPS to this problem, expecting that the prior learned by the generative model provides the regularization needed to select the correct solution among the geometrically feasible candidates. We compare against a baseline that does not use any prior information on the data distribution. Formally, this baseline samples from the posterior

$$q(x) \propto p(u_1 \mid x, y_1)\, p(u_2 \mid x, y_2). \tag{6}$$

Baseline sampler. We use MALA to sample from q with the Markov chain $x_+ = \exp_x(\tau \nabla_{S^2} \log q(x) + \sqrt{2\tau}\, v)$, $v \sim \mathrm{Proj}_{T_x S^2} N(0, I)$, using $\tau = 1$ and 5000 burn-in steps.

Score network. Following Bortoli et al. (2022), our neural net for estimating $\nabla \log p_t$ uses 5 hidden layers of dimension 512, trained using implicit score matching.

Results. We sample $\{x_i\}_{i=1}^B$ earthquake locations, sample $\{(y_1^i, y_2^i)\}_{i=1}^B$ sensor locations uniformly at random, and sample observations $\{(u_1^i, u_2^i)\}_{i=1}^B$ from (6) using $x = x_i$ and $(y_1, y_2) = (y_1^i, y_2^i)$. For each sampler we track the mean geodesic distance between x_i and the current sample estimate over the course of sampling. Figure 1 plots this distance against fractional progress through each sampler: both DPnP and MDPS converge well below the likelihood-only MALA baseline, with DPnP converging faster and MDPS reaching a comparable distance by the end.

5.2. Random trajectory on sphere

Given a path $X(1), X(2), \ldots, X(T)$ on S^2, we view it as an object on the product manifold $\bigtimes_{i=1}^T S^2$, whose tangent space and Brownian heat flow are products of the corresponding individual structures. Each path consists of $T = 4$ points generated from a momentum-driven random walk on S^2; the full description of the prior distribution and parameter values used is given in Appendix D.

The path reconstruction problem. We study the problem of reconstructing the entire path given only the noised position of the two endpoints. Specifically, given a path $x \in \mathbb{S}^{2\times T}$, we observe $(y_o, y_T) \in \mathbb{S}^2 \times \mathbb{S}^2$ with likelihood

$$\mathbb{P}(y_o, y_T \mid x) \propto \exp(\kappa(\langle y_o, x(0)\rangle + \langle y_T, x(T)\rangle). \tag{7}$$

In Figure 2, we sampled a path x from the prior, sampled multiple observations (y_o, y_T) from (7), and displayed the average posterior sample under each method.

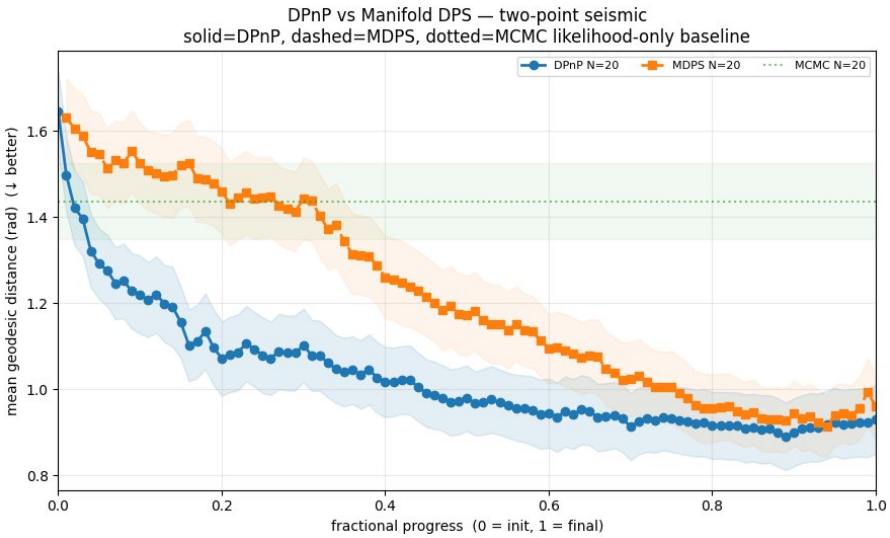


Figure 1. Manifold DPnP and MDPS substantially outperform the prior-free MALA baseline on the two-point problem, with DPnP converging faster.

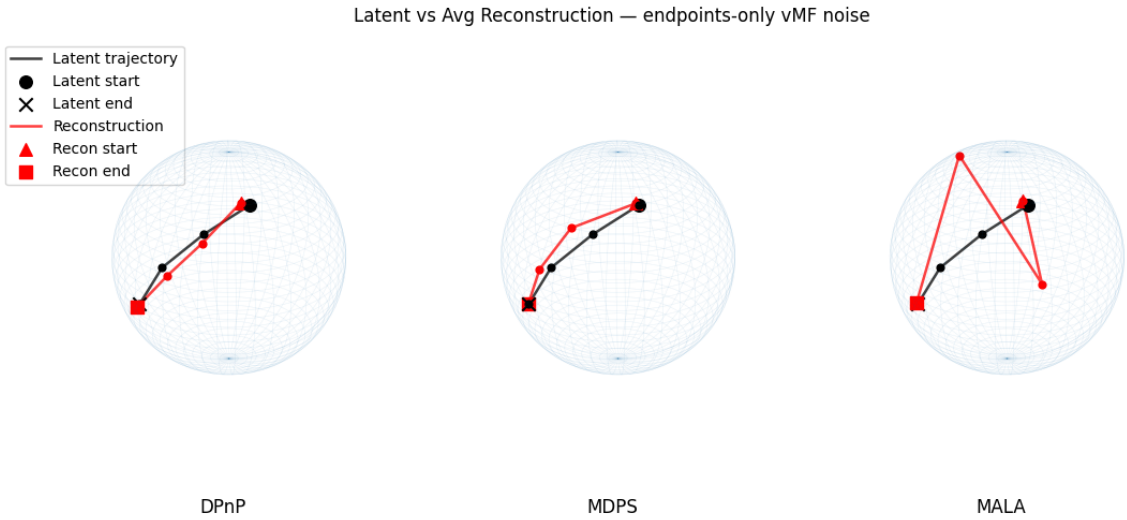


Figure 2. A representative reconstruction using DPnP, MDPS, and the MALA baseline.

Figure 3 compares the performance of DPnP, MDPS, and the MALA baseline in terms of their reconstruction gap. We sample latent paths $x^{(i)}$, $i = 1, \ldots, B$, sample observations $(y_o^{(i)}, y_T^{(i)})$ from (7), and sample $\widehat{x}^{(ij)}$, $j = 1, \ldots, N$ from the posterior. The averaged reconstruction gap is

$$L = \frac{1}{B} \sum_{i=1}^{B} \frac{1}{T} \sum_{t=1}^{T} d_{geo}(x^{(i)}(t), \widehat{x}^{(i)}(t)),$$

where $\widehat{x}^{(i)}(t)$ is the per-observation spherical mean of the N posterior samples.

5.3. Protein backbone angles on $\mathbb{T}^2$

The Ramachandran dataset of Lovell et al. (2003), a standard benchmark in manifold sampling (Davis et al., 2025; Chen & Lipman, 2023), provides protein backbone dihedral pairs

$$x = (\phi, \psi) \in S^1 \times S^1 \cong \mathbb{T}^2.$$

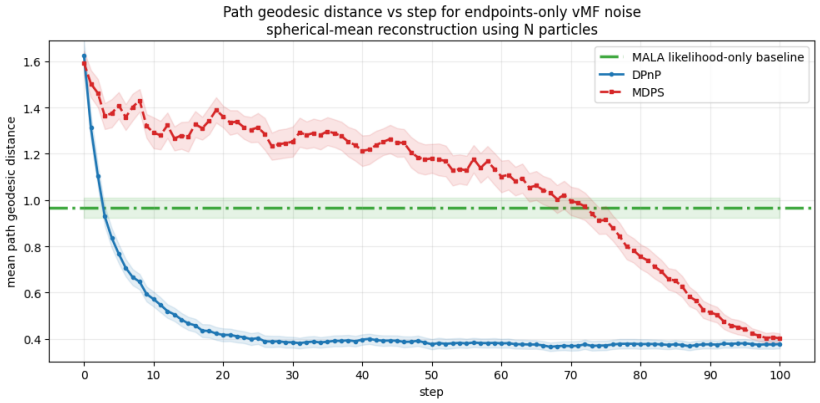


Figure 3. DPnP and MDPS perform better than the MALA likelihood-only baseline on the trajectory reconstruction problem.

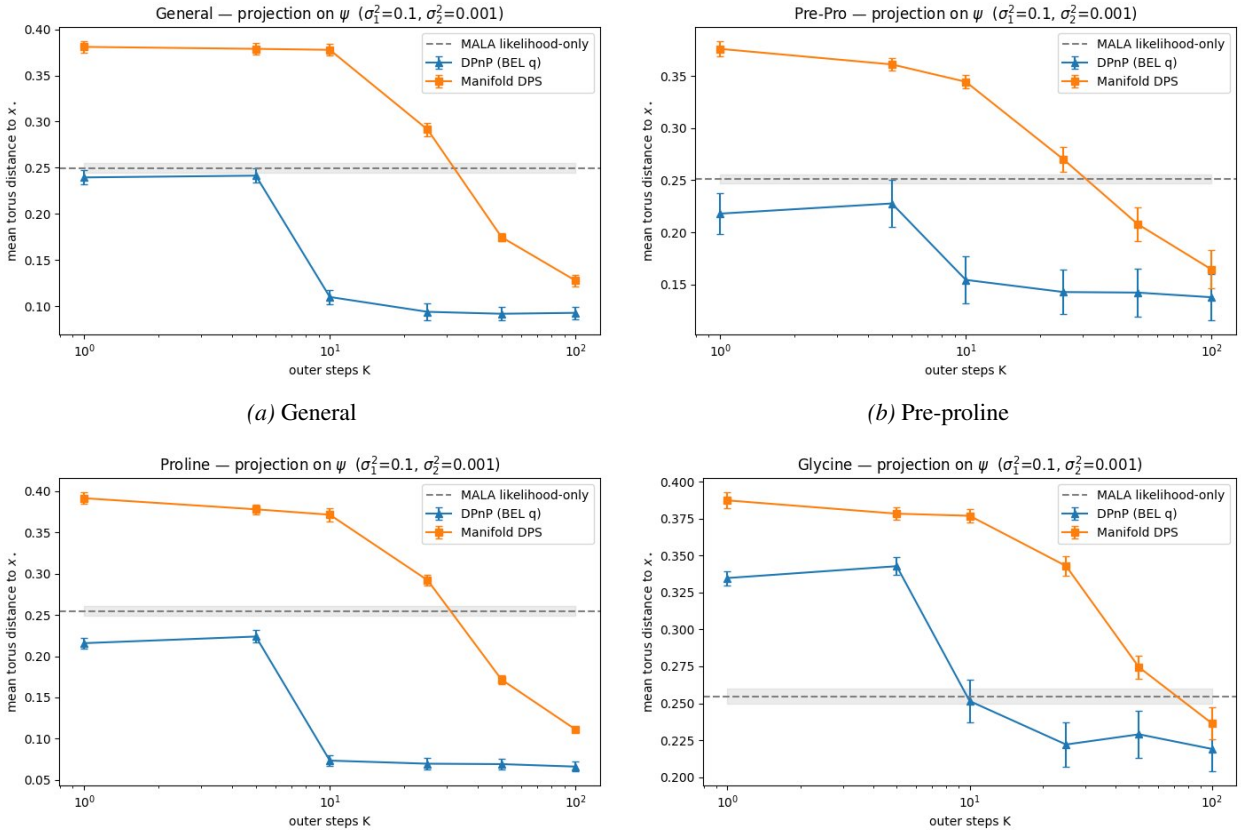
It is a natural test for manifold density modeling: both coordinates are periodic and the empirical distribution is highly multimodal, with modes corresponding to common backbone conformations such as α-helices, β-sheets, and left-handed helices.

We pose an inverse problem with asymmetric observation noise,

$$y = (\phi + \xi_1, \psi + \xi_2), \qquad \xi_1 \sim \mathcal{N}(0, \sigma_1^2), \quad \xi_2 \sim \mathcal{N}(0, \sigma_2^2),$$

with $\sigma_1^2 = 10^{-1}$ and $\sigma_2^2 = 10^{-3}$. The ψ coordinate is observed nearly precisely while ϕ is observed only weakly, forcing the sampler to use the learned prior to fill in the conditional structure of ϕ given ψ.

Figure 4 reports the mean torus distance to the true angle pair as a function of the number of outer steps K, for four standard residue classes (general, pre-proline, proline, and glycine). On the general, pre-proline, and proline classes, both Manifold DPnP and MDPS fall well below the prior-free MALA baseline as K increases, with DPnP converging fastest; representative posterior samples for each class are deferred to Appendix E, where DPnP and MDPS concentrate on the conditional mode selected by the prior while the MALA baseline spreads along the weakly observed ϕ axis. Glycine is a harder case: its empirical distribution is dominated by two modes whose ψ marginals overlap almost entirely, so conditioning on the precisely observed ψ coordinate does little to distinguish them, and the prior provides correspondingly limited information about ϕ. As a result, the generative-prior samplers improve only marginally over the baseline on glycine.



(a) General

(b) Pre-proline

(c) Proline

(d) Glycine

Figure 4. Mean torus distance to x_{true} versus the number of outer steps K on the Ramachandran problem, for four residue classes. Manifold DPnP and MDPS fall well below the prior-free MALA baseline (dashed) on general, pre-proline, and proline; on glycine the improvement is smaller because the two dominant modes overlap in ψ (see text).

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Besse, A. L. *Einstein Manifolds*. Classics in Mathematics. Springer, Berlin, Heidelberg, 1987. doi: 10.1007/978-3-540-74311-8.

Bismut, J.-M. *Large Deviations and the Malliavin Calculus*, volume 45 of *Progress in Mathematics*. Birkhäuser, Boston, 1984. ISBN 9780817632205.

Bora, A., Jalal, A., Price, E., and Dimakis, A. G. Compressed sensing using generative models. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 537–546, 2017.

Bortoli, V. D., Mathieu, E., Hutchinson, M., Thornton, J., Teh, Y. W., and Doucet, A. Riemannian score-based generative modeling, 2022. URL https://arxiv.org/abs/2202.02763.

Brehmer, J. and Cranmer, K. Flows for simultaneous manifold learning and density estimation. In *Advances in Neural Information Processing Systems*, volume 33, pp. 442–453, 2020.

Chen, R. T. Q. and Lipman, Y. Flow matching on general geometries, 2023. URL https://arxiv.org/abs/2302.03660.

Chung, H., Sim, B., Ryu, D., and Ye, J. C. Improving diffusion models for inverse problems using manifold constraints. In *Advances in Neural Information Processing Systems*, volume 35, pp. 25683–25696, 2022.

Chung, H., Kim, J., McCann, M. T., Klasky, M. L., and Ye, J. C. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

Chung, H., Kim, J., McCann, M. T., Klasky, M. L., and Ye, J. C. Diffusion posterior sampling for general noisy inverse problems, 2024. URL https://arxiv.org/abs/2209.14687.

Davis, O., Albergo, M. S., Boffi, N. M., Bronstein, M. M., and Bose, A. J. Generalised flow maps for few-step generative modelling on riemannian manifolds, 2025. URL https://arxiv.org/abs/2510.21608.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Hsu, E. P. *Stochastic Analysis on Manifolds*, volume 38 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002. ISBN 978-0-8218-0802-3. doi: 10.1090/gsm/038.

Huang, C.-W., Aghajohari, M., Bose, J., Panangaden, P., and Courville, A. Riemannian diffusion models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 2750–2761, 2022.

Jørgensen, E. The central limit problem for geodesic random walks. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32:1–64, 1975. doi: 10.1007/BF00533088.

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., and Kumar, V. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 31(8):1544–1554, 2018. doi: 10.1109/TKDE.2018.2861006.

Kawar, B., Elad, M., Ermon, S., and Song, J. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 23593–23606, 2022.

Klimovskaia, A., Lopez-Paz, D., Bottou, L., and Nickel, M. Poincaré maps for analyzing complex hierarchies in single-cell data. *Nature Communications*, 11(1):2966, 2020. doi: 10.1038/s41467-020-16822-4.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, 50(3):437–450, 2003. doi: 10.1002/prot.10286.

Lui, L. M. *Advances in Computational Conformal Geometry for Image Processing*. International Press, 2012.

Milnor, J. Curvatures of left invariant metrics on Lie groups. *Advances in Mathematics*, 21(3):293–329, 1976. doi: 10.1016/S0001-8708(76)80002-3.

Peel, D., Whiten, W. J., and McLachlan, G. J. Fitting mixtures of Kent distributions to aid in joint set identification. *Journal of the American Statistical Association*, 96(453):56–63, 2001. doi: 10.1198/016214501750332992.

Petersen, P. *Riemannian Geometry*, volume 171 of *Graduate Texts in Mathematics*. Springer, Cham, 3 edition, 2016. doi: 10.1007/978-3-319-26654-1.

Shapovalov, M. V. and Dunbrack, Roland L., J. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure*, 19(6):844–858, 2011. doi: 10.1016/j.str.2011.03.019.

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 2256–2265, 2015.

Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Xu, X. and Chi, Y. Provably robust score-based diffusion posterior sampling for plug-and-play image reconstruction, 2024. URL https://arxiv.org/abs/2403.17042.

Ye, H., Lin, H., Han, J., Xu, M., Liu, S., Liang, Y., Ma, J., Zou, J., and Ermon, S. TFG: Unified training-free guidance for diffusion models. 2024. URL https://arxiv.org/abs/2409.15761.

A. General BEL Estimator

For general Riemannian manifolds, the Ricci ODE (2) does not admit a closed-form solution, so we propagate the orthonormal frame and M_t along each Brownian sample path; see Algorithm 5.

Algorithm 5 General BEL estimator

Input: $f : \mathcal{M} \to \mathbb{R}$; $x \in \mathcal{M}$; horizon t; step size h; samples B
Pick orthonormal frame u_0 over x; set $x_0 = x$, $M_0 = I$, $N = \lfloor t/h \rfloor$
for $i = 1$ **to** B **do**
  for $k = 0$ **to** $N - 1$ **do**
    Sample $\zeta_k \sim N(0, I_n)$
    $x_{k+1} = \exp_{x_k}(u_k \sqrt{h}\, \zeta_k)$
    $u_{k+1} = P_{x_k \to x_{k+1}}(u_k)$ // parallel transport
    $M_{k+1} = M_k - \frac{1}{2} h\, \widetilde{\mathrm{Ric}}_{u_k} M_k$
  end for
  $S_i = f(x_N) \sum_{k=0}^{N} M_k(\sqrt{h}\, \zeta_k)$
  $D_i = f(x_N)$
end for
Return: $u_0 \frac{\sum_{i=1}^{B} S_i}{t \sum_{i=1}^{B} D_i}$

B. Ricci Operators for Common Manifolds

Manifold $\mathcal{M}$	Metric / normalization	Ricci operator
$\mathbb{R}^n$	Euclidean metric	0
$\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$	Flat quotient metric	0
$\mathrm{Sym}(n)$	Frobenius metric, $\langle A, B\rangle = \mathrm{tr}(AB)$	0
S^n	Unit round sphere, sectional curvature $+1$	$(n-1)I_n$
$\mathbb{H}^n$	Hyperbolic space, sectional curvature -1	$-(n-1)I_n$
$\mathbb{H}P^m$	Standard symmetric-space metric, sectional curvatures in $[1, 4]$	$4(m+2)I_{4m}$
$\mathrm{Gr}(k, n)$	Real Grassmannian with canonical quotient metric	$(n-2)I_{k(n-k)}$
$\mathrm{SO}(n)$	Bi-invariant metric $\langle X, Y\rangle = -\frac{1}{2}\,\mathrm{tr}(XY)$	$\frac{n-2}{2} I_{\frac{n(n-1)}{2}}$
$\mathrm{SU}(n)$	Bi-invariant metric $\langle X, Y\rangle = -\,\mathrm{Re}\,\mathrm{tr}(XY)$	$\frac{n}{2} I_{n^2-1}$

Table 1. Ricci operators for common manifolds. The constants depend on the metric normalization. The flat and constant-curvature examples are standard (Petersen, 2016). The compact symmetric-space examples, including $\mathbb{H}P^m$ and Grassmannians, may be found in Besse (1987). For compact Lie groups with bi-invariant metrics, see Milnor (1976).

C. Earthquake Data: Train/Test Split

We use earthquake locations from the NOAA significant-earthquakes catalog, available at NOAA NCEI. Figure 5 visualizes the data and our train-test split.

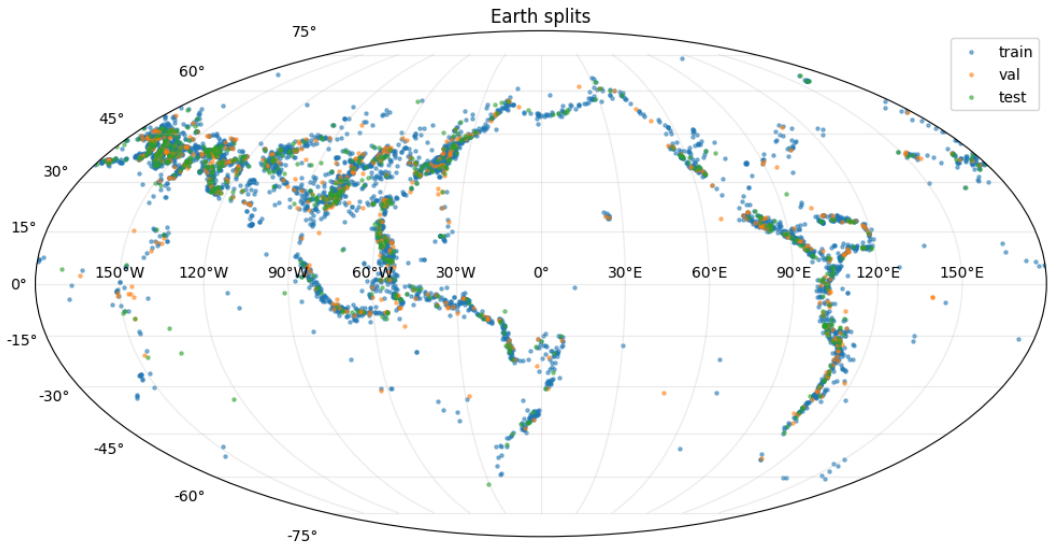


Figure 5. Train-test split for earthquake data.

D. Trajectory Prior Distribution

There are tunable parameters $\mu \in S^2, \kappa, s, \rho, \sigma, \Delta t \in \mathbb{R}$ corresponding to the center of the initial point, the concentration of the initial point, the speed, the momentum, the noise variance, and the time step. On initialization, sample $X(0)$ from the vMF distribution centered at μ and sample a random tangent vector $v(0) \in T_{X(0)}S^2$ by projecting a standard Gaussian onto the tangent space. Given $X(t) \in S^2$ and $v(t) \in T_{X(t)}S^2$, set

$$X(t+1) = \exp_{X(t)}\left(\frac{v(t)}{\|v(t)\|} s\Delta t\right).$$

Let $\tilde{v}(t+1)$ be the parallel transport of $v(t)$ to $X(t+1)$, let $\epsilon(t+1)$ be Gaussian noise in $T_{X(t+1)}S^2$, and set

$$v(t+1) = \mathrm{Proj}_{T_{X(t+1)}S^2}\left(\rho\tilde{v}(t+1) + \sigma\epsilon(t+1)\right).$$

For our experiment, we used $\mu = (0, 1, 0)$, $\kappa = 120$, $s = 10$, $\rho = 0.98$, $\sigma = 0.5$, $\Delta_t = 0.5$. We picked a large ρ and low σ to obtain smooth paths.

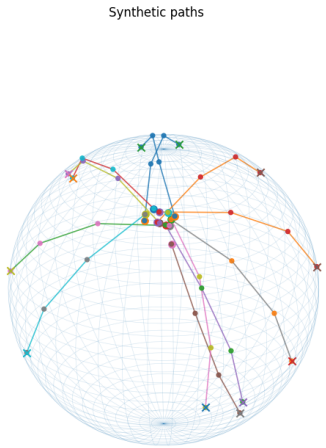


Figure 6. Sample paths from the prior distribution.

E. Ramachandran Posterior Samples

Figures 7 to 10 show representative posterior samples on $\mathbb{T}^2$ for the four residue classes of Section 5.3. In each figure, the first panel shows the empirical prior and the remaining panels overlay samples from Manifold DPnP, MDPS, and the MALA baseline, together with the true angle pair x_{true} (star) and the observation y (cross). Because ψ is observed precisely while ϕ is observed only weakly, the MALA baseline spreads along the ϕ axis at fixed ψ, whereas DPnP and MDPS use the learned prior to concentrate on the appropriate mode. The two dominant modes in the glycine panel (Figure 10) overlap almost entirely in ψ, illustrating why the prior is least informative there.

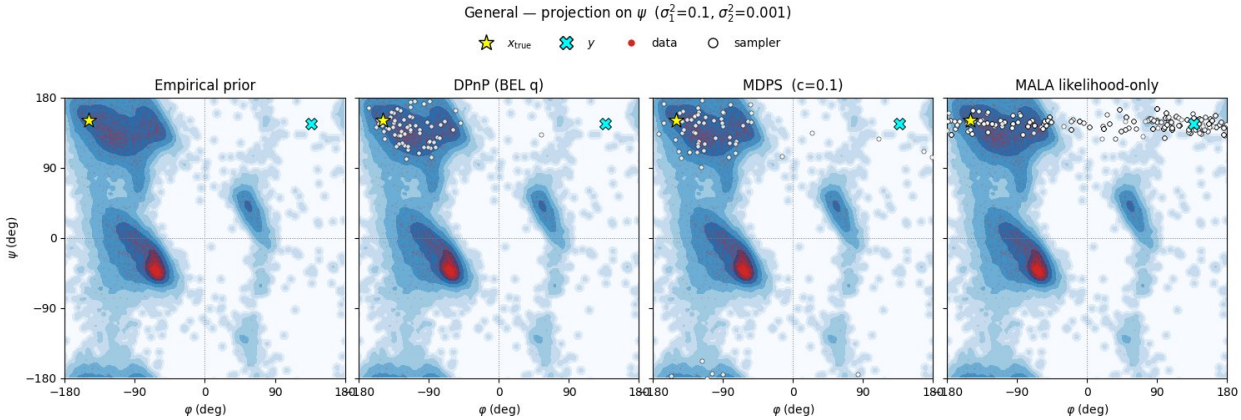


Figure 7. General: representative posterior samples on $\mathbb{T}^2$.

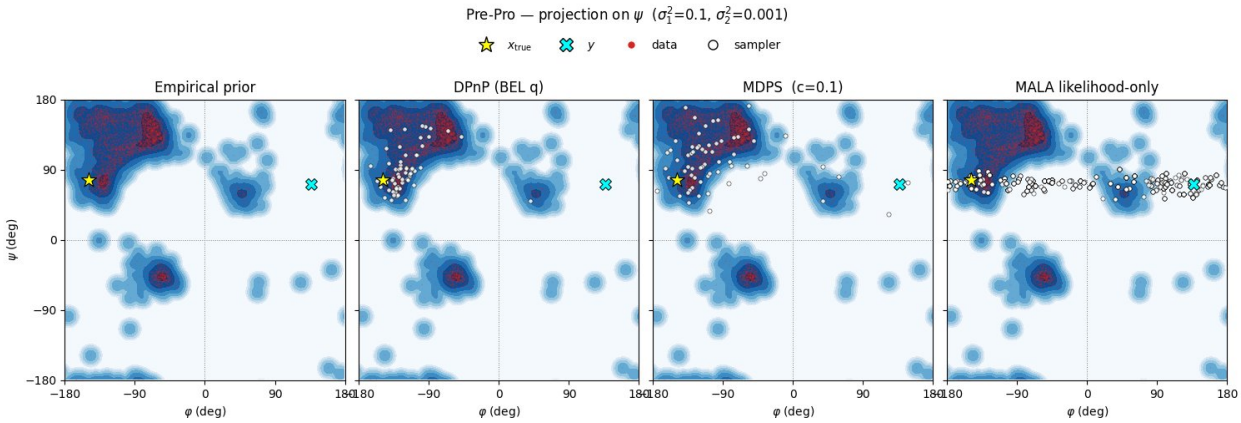


Figure 8. Pre-proline: representative posterior samples on $\mathbb{T}^2$.

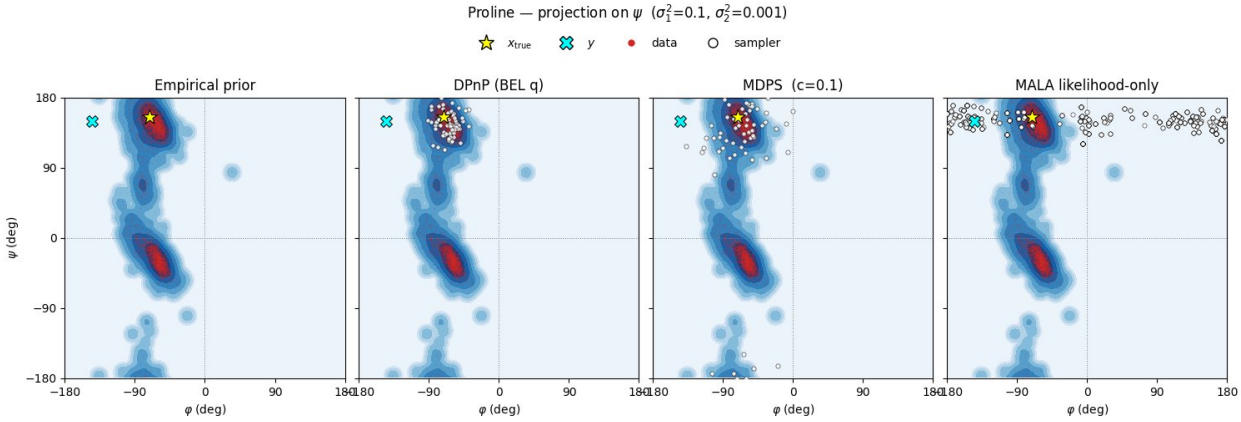


Figure 9. Proline: representative posterior samples on $\mathbb{T}^2$.

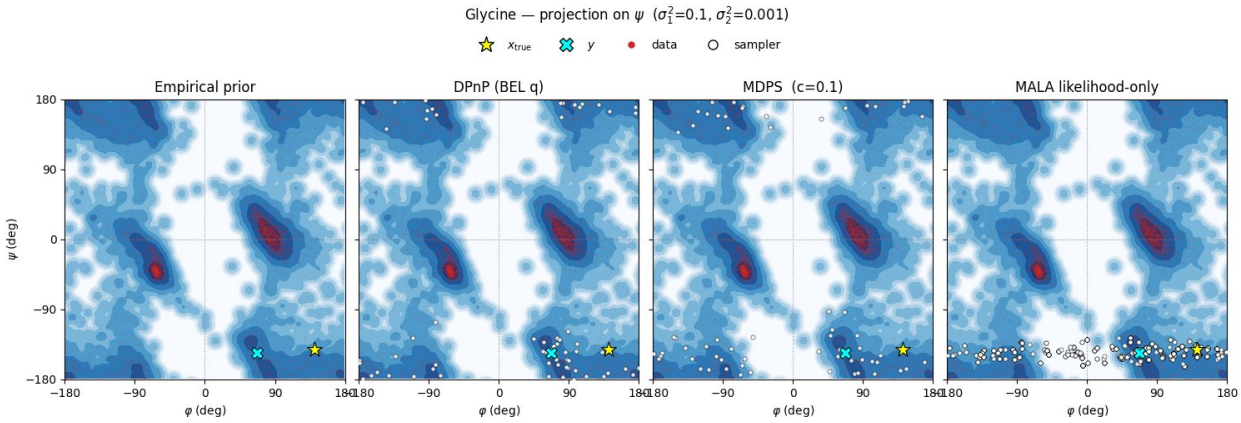


Figure 10. Glycine: representative posterior samples on $\mathbb{T}^2$. The two dominant modes have largely overlapping ψ marginals, so the precisely observed ψ does little to disambiguate them.

F. Experiments on the Flat Torus

As an additional synthetic benchmark, we evaluate the samplers on a wrapped Gaussian mixture on the flat torus. We view $\mathbb{T}^d \cong \mathbb{R}^d/\mathbb{Z}^d$. The universal cover $\pi : \mathbb{R}^d \to \mathbb{T}^d$ of $\mathbb{T}^d$ is given by $\pi(x) = x \bmod 1$. Given a Gaussian $\tilde{Z} \sim N(\mu, \Sigma)$ on $\mathbb{R}^d$ with density $\tilde{\phi}_{\mu,\Sigma}$, its π-pushforward is the wrapped Gaussian $Z := \pi(\tilde{Z}) \sim \mathrm{N}^{\mathbb{T}}(\mu, \Sigma)$ on $\mathbb{T}^d$, with density $\phi_{\mu,\Sigma}(x) = \sum_{n\in\mathbb{Z}^d} \tilde{\phi}_{\mu,\Sigma}(x+n)$, which we approximate by truncating the sum to $n \in [-K, K]^d$. Wrapped Gaussian mixtures are obtained analogously. We choose the prior to be a three-component wrapped Gaussian mixture and the likelihood to be

$$p_{y|x}(\cdot \mid x) \sim \mathrm{N}^{\mathbb{T}}(x, \Sigma_y), \quad \Sigma_y = \mathrm{diag}(0.001, 0.1). \tag{8}$$

Under this likelihood, y_1 is concentrated around x_1 while y_2 is extremely noisy, so the forward operator essentially projects onto the first coordinate of x. The score network has 4 hidden layers of dimension 128, trained using denoising score matching. Both DPnP and MDPS recover the correct conditional structure of the weakly observed coordinate, while the prior-free MALA baseline does not.

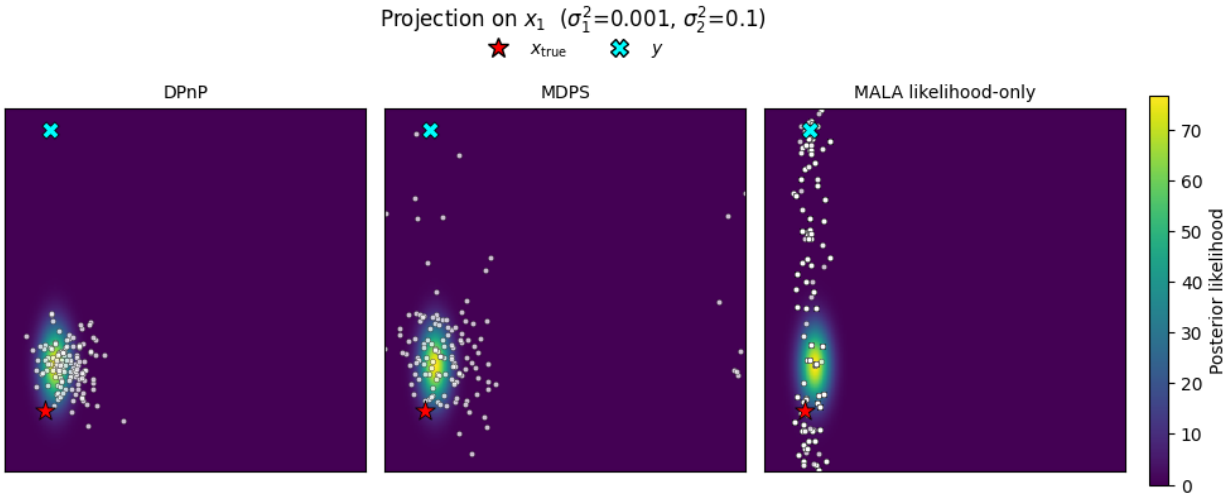


Figure 11. A representative example in which DPnP and MDPS correctly identify the x_2 component while the MALA sampler fails.

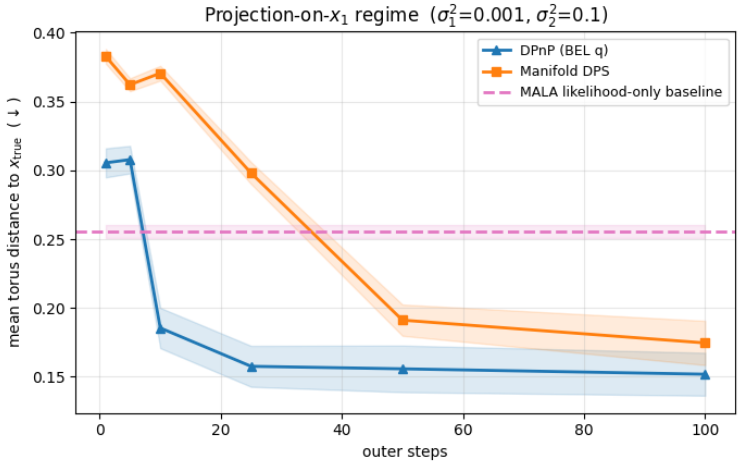


Figure 12. Both DPnP and MDPS outperform the MALA baseline on the torus.

G. Proof of Theorem 3.1

Proof. We prove a general reverse-time representation for a single proximal step, and then apply it twice.

Step 1: a generic proximal step as reverse heat flow. Let r_0 be a probability density on $\mathbb{R}^d$, and let

$$r_t \coloneqq r_0 * \varphi_t, \qquad \varphi_t(z) \coloneqq (2\pi t)^{-d/2} \exp\left(-\frac{\|z\|^2}{2t}\right),$$

be its heat evolution. Fix $\tau > 0$, and consider the forward diffusion

$$\mathrm{d}Z_t = \mathrm{d}W_t, \qquad Z_0 \sim r_0, \qquad 0 \le t \le \tau.$$

Then Z_t has density r_t.

Now define the time-reversed process $\widehat{Z}_s \coloneqq Z_{\tau-s}$, $0 \le s \le \tau$. By the standard time-reversal formula for diffusions with unit diffusion matrix, the reversed process is again a diffusion, with drift given by the score of the time-$(\tau - s)$ marginal:

$$\mathrm{d}\widehat{Z}_s = \nabla \log r_{\tau-s}(\widehat{Z}_s)\,\mathrm{d}s + \mathrm{d}\widehat{W}_s, \qquad 0 \le s \le \tau,$$

for some Brownian motion $(\widehat{W}_s)_{0\le s\le\tau}$.

Condition now on the terminal value of the forward process, equivalently on the initial value of the reversed process: $\widehat{Z}_0 = Z_\tau = x$. Under this conditioning, $\widehat{Z}$ is precisely the reverse-time diffusion started from x,

$$\mathrm{d}X_s = \nabla \log r_{\tau-s}(X_s)\,\mathrm{d}s + \mathrm{d}B_s, \qquad X_0 = x, \qquad 0 \le s \le \tau.$$

Its terminal law is $\mathsf{Law}(X_\tau) = \mathsf{Law}(Z_0 \mid Z_\tau = x)$.

Since $Z_\tau = Z_0 + G$ with $G \sim \mathcal{N}(0, \tau I_d)$ independent of Z_0, Bayes' rule gives

$$\mathsf{Law}(Z_0 \in \mathrm{d}u \mid Z_\tau = x) \propto r_0(u)\,\varphi_\tau(x-u)\,\mathrm{d}u \propto r_0(u) \exp\left(-\frac{\|u-x\|^2}{2\tau}\right)\mathrm{d}u.$$

Therefore the endpoint X_τ of the reverse diffusion has exactly the proximal law with base density r_0 and quadratic parameter τ.

Step II: application to the denoising substep. Take $r_0 = p^\star$ and $\tau = \eta^2$. Then the diffusion denoising sampler from input x produces exactly the terminal value at time η^2 of

$$\mathrm{d}X_t = \nabla \log p_{\eta^2-t}(X_t)\,\mathrm{d}t + \mathrm{d}B_t, \qquad X_0 = x, \qquad 0 \le t \le \eta^2.$$

Step III: application to the consistency substep. Let $q_0(u) \propto \exp(\mathcal{L}(u; y))$, and let q_t be its heat evolution. Applying the same argument with $r_0 = q_0$ and $\tau = \eta^2$, we find that, conditional on the intermediate state X_{η^2}, one proximal consistency step is exactly the terminal value at time $2\eta^2$ of

$$\mathrm{d}X_t = \nabla \log q_{2\eta^2-t}(X_t)\,\mathrm{d}t + \mathrm{d}B_t, \qquad \eta^2 < t \le 2\eta^2.$$

Step IV: concatenation. Composing the two Markov kernels yields the piecewise diffusion

$$\mathrm{d}X_t = \begin{cases} \nabla \log p_{\eta^2-t}(X_t)\,\mathrm{d}t + \mathrm{d}B_t, & 0 \le t \le \eta^2, \\ \nabla \log q_{2\eta^2-t}(X_t)\,\mathrm{d}t + \mathrm{d}B_t, & \eta^2 < t \le 2\eta^2, \end{cases} \qquad X_0 = x.$$

Its endpoint $X_{2\eta^2}$ is exactly the output of one full iteration of DPnP, in the order encoded by the displayed SDE. □