What Makes a Good Path? Factoring Manifold Support and Path Geometry

Zhixuan Zhou Tingting Dan Guorong Wu*
Departments of Computer Science and Psychiatry
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
{Tingting_Dan,grwu}@med.unc.edu; zzhixuan@cs.unc.edu

Abstract

The manifold hypothesis suggests that real-world data concentrates on low-dimensional structures, yet navigating meaningful trajectories on latent manifolds remains an open challenge. A common approach is to construct density-aware metrics, but such constructions bundle two distinct questions: (i) where valid data lies and (ii) how we choose to traverse on the manifold. Since different tasks may demand different traversal behaviors even on the same data manifold, specifying these two aspects independently offers additional modeling flexibility. We propose to factor the path objective into a kinetic term governed by a freely chosen geometric prior and a score-based potential, derived from the negative log-density of a pretrained diffusion model, that acts as a soft constraint for manifold adherence. The potential encourages dynamics to evolve along the data support, while the geometric prior remains free to encode task-specific notions of path optimality. To find stationary paths for this objective, we introduce *geodesic force matching* (GFM), a direct-collocation algorithm that discretizes the trajectory and optimizes all waypoints jointly to satisfy the Euler–Lagrange force balance in a least-squares sense. On synthetic manifolds and 3D shape interpolation, our method produces smooth, on-manifold paths, matching strong density-aware baselines at moderate noise and showing clearer advantages with informative priors in data-sparse regions and at low noise where those baselines become numerically unstable. Code is available at `https://github.com/taka2223/GFM-release`.

1 Introduction

The manifold hypothesis posits that real-world data such as images or 3D shapes lie on low-dimensional submanifolds within the ambient space [8, 18, 33, 42]. Modern deep generative models build on this idea, compressing complex distributions into tractable latent spaces [25, 27, 34, 40]. While these models generate realistic samples, they offer no guarantees about paths connecting samples in latent space. Standard linear interpolation implicitly assumes a flat Euclidean geometry and can traverse low-density regions, yielding unrealistic intermediate states [22, 30, 46]. Given strong tools to *represent* the data manifold, how should we *navigate* paths on it in a geometry-aware manner?

We view geometry-aware interpolation as involving two conceptually distinct components: the *data support* $\mathcal{Z}$, defining the locus of valid samples (the "where"), and the *geometric prior* $(\mathcal{M}, g)$, determining the ambient topology and how path length is measured (the "how"). The data support governs semantic validity by restricting paths to high-density regions, while the geometric prior

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

specifies the rules of movement, e.g., whether to follow a straight line in $\mathbb{R}^l$ or a great circle on $\mathbb{S}^{l-1}$. Specifying these components independently offers additional modeling flexibility.

Existing approaches to geometry-aware interpolation share a common design choice: the manifold constraint and the metric are not separately specified. Most existing methods induce a Riemannian metric directly from the data density. The metric is constructed via decoder or encoder Jacobians [3, 4, 12, 38], energy-based models [9], or diffusion scores [7, 36], all assigning high cost to low-density regions to steer paths toward the data support. However, tying the metric to data density makes the metric tensor ill-conditioned in data-sparse regions, producing erratic trajectories—a failure mode we document in Section 5.1. We thus ask whether decoupling $\mathcal{Z}$ from $(\mathcal{M}, g)$ can mitigate these pitfalls while preserving flexibility.

We propose a framework that specifies manifold support and the geometric prior as separate additive terms by treating the latent space as a physical system. The geometric prior determines the system's *kinetic energy* (how easy it is to move), while the data distribution induces a *potential energy* field that pulls trajectories toward the manifold. Balancing these forces recovers paths that traverse the manifold smoothly. We realize the potential via a pretrained diffusion model's score function: since the noisy negative log-density $-\log p_\sigma(z)$ approximates the squared distance to the data support [8, 44], it provides a natural soft potential additive to any path energy.

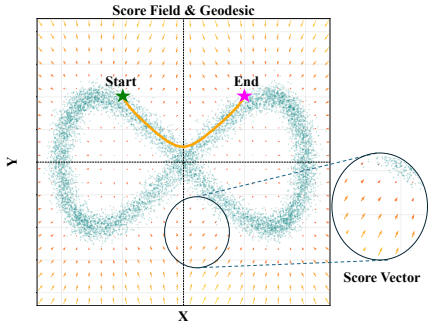


Figure 1: Geodesic dynamics in a score-induced potential. The vector field depicts the score $\nabla \log p_\sigma(x)$, which drives the state toward the data support. The orange path solves the Euler–Lagrange equations balancing metric-induced kinetic energy against the score-induced potential, yielding smooth interpolation along the manifold.

This turns constrained geodesic computation into an unconstrained variational problem. Its Euler–Lagrange equations describe second-order dynamics that balance geometric acceleration against the score-induced potential. To solve them, we introduce **Geodesic Force Matching (GFM)**, a direct collocation algorithm [10, 24] that discretizes the trajectory into waypoints and jointly optimizes them to satisfy the force balance in a least-squares sense. Treating the score as a fixed target removes the need to backpropagate through the score Jacobian, and the resulting solver remains well-behaved at low noise where Jacobian-based metrics fail (Section 5.2).

We summarize our contributions as follows:

- **Additive formulation.** A framework where manifold support and the geometric prior enter the path objective as separate additive terms: a score-based potential encourages adherence to the data support, while the metric encodes task-specific path optimality.
- **Direct-collocation solver.** GFM optimizes all waypoints jointly to match the Euler–Lagrange force balance. It is stable on high-dimensional latents (up to 512×8) and substantially cheaper than Jacobian-based score-geometry baselines.
- **Empirical study.** On synthetic manifolds and 3D shape interpolation, GFM produces on-manifold, smooth trajectories. It is competitive at moderate noise (σ=0.3) and shows a clearer advantage in data-sparse regions with informative priors and at low noise (σ=0.1), where baseline metrics become ill-conditioned.

2 Related Works

Geometry-aware interpolation in latent spaces has been approached through various choices of Riemannian metrics. While these methods differ in how they construct the metric tensor, they share a common design: All geometric structure, including the implicit characterization of the data manifold, is exclusively encoded through the metric.

Kernel-based Metrics Early attempts to equip the latent space with a Riemannian geometry relied on non-parametric estimation of the data support. Pioneered by Tosi et al. [43] and later adapted for deep generative models by Arvanitidis et al. [5], Kapusniak et al. [21], these methods construct a metric directly from the training data embeddings using different kernels like Radial Basis Functions (RBF) or locally adaptive normal distribution (LAND) [2]. While elegant, kernel-based approaches face the curse of dimensionality where distance concentration renders the metric nearly Euclidean in high-dimensional latent spaces.

Pullback Metrics A natural approach induces geometry via the Jacobian of the decoder (or encoder) mapping. These approaches [3, 4, 6, 12, 15, 16, 38] define a Riemannian metric by pulling back the Euclidean metric from the observation space. In modern two-stage models (e.g., Latent Diffusion [34] as well as our settings), the latent space often serves merely as a compression domain, where the valid data manifold forms a complex, implicit substructure within $\mathcal{Z}$ (Details in Section3.1)

Score/Density-Aware Metrics Another line of work leverages pre-trained diffusion models to define density-aware metrics. Some of them [7, 9, 48] prioritized the high density regions, guiding the path to locate on those regions. Several works also try to utilize the Hessian of the log-density (i.e., the Jacobian of score function) [36, 37] to define a metric that penalizes motion along directions of rapid score change. However, like other density-aware approaches, these methods still encode both manifold structure and path geometry through a single metric tensor, precluding independent specification of traversal behavior.

Manifolds with Known Geometry. A significant body of work has integrated Riemannian geometry into generative modeling by assuming the data resides on a manifold with a known metric tensor [28, 29, 31]. In the context of diffusion models, De Bortoli et al. [14] and Huang et al. [19] formulated the stochastic differential equations (SDEs) directly on compact manifolds using the Laplace-Beltrami operator. While theoretically rigorous, these approaches are limited to domains with analytically tractable geometry and cannot easily adapt to the complex, learned submanifolds typical of high-dimensional data [13, 20].

3 Problem Formulation

3.1 Setting

Let $\mathcal{D} = \{x_i\}_{i=1}^N \subset \mathbb{R}^n$ be a dataset encoded into latent space via a pretrained autoencoder (ϕ, ψ) with $\phi : \mathbb{R}^n \to \mathbb{R}^l$, $\psi : \mathbb{R}^l \to \mathbb{R}^n$, and $l \ll n$. Two geometric structures coexist in this latent space.

The base manifold $\mathcal{M} \subseteq \mathbb{R}^l$ (e.g., $\mathbb{R}^l$ or $\mathbb{S}^{l-1}$) is the domain on which paths are defined, equipped with a Riemannian metric g. In ambient coordinates, g is a positive-definite matrix field $G(z)$ with $\langle u, v \rangle_g = u^\top G(z) v$ for tangent vectors $u, v \in T_z\mathcal{M}$. The metric defines the *cost* of motion through latent space.

The data support $\mathcal{Z} := \operatorname{supp}(p_0) \subset \mathcal{M}$ is the implicit set on which the latent data distribution concentrates. Our analysis treats $\mathcal{Z}$ as a smooth, lower-dimensional submanifold; its density may be non-uniform.

A central design choice in our framework is to specify $(\mathcal{M}, g)$ and $\mathcal{Z}$ independently: $(\mathcal{M}, g)$ controls how to move, while $\mathcal{Z}$ controls where motion is valid.

3.2 Geometry-Aware Interpolation

Given endpoints $z_{\text{start}}, z_{\text{end}} \in \mathcal{Z}$, we seek a path $\gamma : [0, 1] \to \mathcal{M}$ that connects them while remaining on the data manifold:

$$\min_{\gamma} \mathcal{E}[\gamma] = \int_0^1 \tfrac{1}{2} \|\dot{\gamma}_t\|_g^2 \, dt \quad \text{s.t.} \quad \gamma_0 = z_{\text{start}},\ \gamma_1 = z_{\text{end}},\ \gamma_t \in \mathcal{Z}. \tag{1}$$

This problem is intractable: $\mathcal{Z}$ is implicit and high-dimensional, precluding both explicit constraint enforcement and projection-based methods. Solving Eq. (1) thus requires three ingredients:

(i) *Data support.* A tractable representation of $\mathcal{Z}$ identifying valid regions without explicit parametrization or hard constraints.

(ii) *Geometric prior.* A base manifold $\mathcal{M}$ and metric g defining global topology and path length, independent of the data distribution.

(iii) *Numerical solver.* An algorithm that finds stationary paths balancing metric-induced kinetic energy and adherence to the data support.

4 Methodology

Key insight. A pretrained diffusion model already exposes what we need to characterize $\mathcal{Z}$: a tractable proxy for manifold distance.

Diffusion models learn to denoise by estimating the gradient of a noise-smoothed data density. Near the data manifold, this gradient (the "score") points toward $\mathcal{Z}$, and under mild regularity assumptions the negative log-density is approximately proportional to the squared distance $d(z, \mathcal{Z})^2$ (Theorem 4.1).

This connection suggests a natural relaxation strategy. The hard constraint $\gamma_t \in \mathcal{Z}$ is equivalent to requiring $d(\gamma_t, \mathcal{Z}) = 0$. While the squared distance d^2 is intractable to compute directly, we can substitute its proxy, the negative log-density, as a soft penalty. The resulting objective is:

$$\min_{\gamma} \ \mathcal{J}[\gamma] = \int_0^1 \ \tfrac{1}{2}\|\dot{\gamma}_t\|_g^2 \ + \ \lambda \mathcal{U}(\gamma_t)\, dt, \tag{2}$$

where the first term is the kinetic energy induced by the geometric prior g, and $\mathcal{U}$ is the score-based potential defined in Eq. (7), whose leading small-noise term penalizes squared distance to the data support. The weight $\lambda > 0$ controls the strength of manifold adherence.

Crucially, the metric g and the potential $\mathcal{U}$ enter as independent additive terms: we can specify the geometric prior without affecting how the data support is characterized, and vice versa.

The remainder of this section makes this precise: we first establish the theoretical justification (Section 4.1), then discuss choices of geometric prior (Section 4.2), and finally present our solver (Section 4.3).

4.1 The Score-Based Potential

Let p_0 be the latent distribution supported on $\mathcal{Z}$, and let $p_\sigma := p_0 * \mathcal{N}(0, \sigma^2 I)$ denote the density smoothed by Gaussian noise at scale σ. We work within the tubular neighborhood $U_{\mathcal{Z}}$ of the manifold, where every point z admits a unique orthogonal projection $\Pi(z) \in \mathcal{Z}$.[2]

Characterizing the manifold via scores. The score function provides a principled way to identify the data manifold. As we establish below (Theorem 4.1), the score $\nabla \log p_\sigma(z)$ is approximately proportional to the displacement $\Pi(z) - z$ toward the manifold. This implies that the scaled score vanishes precisely on $\mathcal{Z}$:

$$\lim_{\sigma \to 0} \sigma^2 \nabla \log p_\sigma(z) = 0 \quad \text{for all } z \in \mathcal{Z}, \tag{3}$$

which suggests defining the manifold as the zero-set of this quantity:

$$\mathcal{Z} = \big\{ z \in U_{\mathcal{Z}} \ \big| \ \lim_{\sigma \to 0} \|\sigma^2 \nabla \log p_\sigma(z)\| = 0 \big\}. \tag{4}$$

The dilemma of hard constraints. While Eq. (4) provides a rigorous characterization, directly enforcing it encounters two fundamental obstacles. First, there is a *structural circularity*: identifying the manifold via the vanishing-score condition requires restricting to $U_{\mathcal{Z}}$, otherwise the condition is also satisfied on the medial axis (e.g., the center of a sphere). However, defining $U_{\mathcal{Z}}$ requires prior knowledge of $\mathcal{Z}$, which is precisely what we seek. Second, *hard constraint enforcement is computationally prohibitive*: standard methods such as projected gradient descent require evaluating the expensive Jacobian of the score network $\nabla_z s_\theta(z, \sigma)$ to project trajectories back onto the manifold.

[2]The tubular neighborhood has radius equal to the reach τ of $\mathcal{Z}$; see Lee [26].

From hard constraints to soft potentials. Rather than enforcing the manifold constraint directly, we exploit a deeper connection: the score is not merely an *indicator* of the manifold, but the gradient of a *potential* that measures distance to it. This allows us to convert the hard constraint into a soft penalty that is both tractable and well-defined throughout the ambient space.

Theorem 4.1 (Score-Distance Correspondence)**.** *Let $\mathcal{Z} \subset \mathbb{R}^l$ be a compact, smooth k-dimensional submanifold with reach $\tau > 0$, and assume p_0 has a smooth density on $\mathcal{Z}$ that is bounded away from zero.*[3] *For any $z \in U_{\mathcal{Z}}$ and sufficiently small σ:*

(i) The score acts as a restoring force toward the manifold:

$$\nabla \log p_\sigma(z) = -\frac{z - \Pi(z)}{\sigma^2} + O(1). \tag{5}$$

(ii) The negative log-density approximates the squared distance:

$$-\log p_\sigma(z) = \frac{d(z, \mathcal{Z})^2}{2\sigma^2} + C_\sigma + O(1), \tag{6}$$

where $d(z, \mathcal{Z}) = \|z - \Pi(z)\|$ and C_σ depends only on σ and the codimension $l - k$.

Proof sketch. Part (i) follows from Tweedie's formula: the posterior mean $\mathbb{E}[z_0|z]$ concentrates at $\Pi(z)$ as $\sigma \to 0$. Part (ii) is obtained by applying Laplace's method to the convolution integral defining p_σ. See Appendix F.1 for details. □

Part (i) explains Eq. (4): the scaled score vanishes in the limit on $\mathcal{Z}$, where $z = \Pi(z)$. Part (ii) motivates the operational potential

$$\mathcal{U}_\sigma(z) := -\log p_\sigma(z), \qquad \nabla \mathcal{U}_\sigma(z) = -\nabla \log p_\sigma(z) \approx -s_\theta(z, \sigma). \tag{7}$$

We write $\mathcal{U} = \mathcal{U}_\sigma$ at a fixed noise level. At finite σ, it also reflects on-manifold density. In practice, we use its gradient through $s_\theta(z, \sigma) = (D_\theta(z, \sigma) - z)/\sigma^2$, computed from a pretrained EDM-style denoiser [23] via Tweedie's formula.

Remark 4.2 (Physical interpretation)*.* Our formulation admits a mechanical interpretation: a particle moves under kinetic energy $\frac{1}{2}\|\dot{\gamma}\|_g^2$ (governed by the geometric prior) while confined to a potential well $\mathcal{U}(z)$ centered on $\mathcal{Z}$ (governed by the data distribution). The score acts as the restoring force pulling the particle toward the manifold.

4.2 Geometric Priors: The Choice of Base Space and Metric

Once the data manifold support is characterized by $\mathcal{U}$ (Section 4.1), the geometric prior $(\mathcal{M}, g)$ can be specified independently. The base manifold $\mathcal{M}$ defines the domain on which paths live, while the metric g measures their length. We highlight practical configurations.

Flat Euclidean space **($\mathcal{M} = \mathbb{R}^l, g = \delta$).** The default when no prior topological structure is known: $\mathcal{M} = \mathbb{R}^l$ with the canonical metric $G(z) = \mathbf{I}$.

Hyperspherical space **($\mathcal{M} = \mathbb{S}^{l-1}, g =$ induced).** For latent spaces with high-dimensional Gaussian priors (VAEs, diffusion models), probability mass concentrates on a shell of radius $\sqrt{l}$. We respect this topology by setting $\mathcal{M} = \mathbb{S}^{l-1} \subset \mathbb{R}^l$, with g induced from the ambient inner product.

Pullback geometry **($\mathcal{M} = \mathbb{R}^l, g =$ pullback).** To measure distances in data space rather than latent space, we equip the flat base manifold with the pullback metric $G(z) = J_\psi(z)^\top J_\psi(z)$, where $J_\psi(z)$ is the decoder Jacobian. The recently proposed ProbGEORCE [35] can be viewed as our framework instantiated in this configuration.

Task-specific geometry. A metric can assign unequal costs to directions or motions specified by domain knowledge, as in our torus experiments.

[3] When this density assumption is relaxed (e.g., near-zero density along parts of $\mathcal{Z}$), the $O(1)$ residual in Eq. (6) acquires an additional point-dependent term that depends on local intrinsic density; see Appendix F.1 for a quantitative discussion.

We fix geometry from structural or task knowledge. Among plausible candidates, held-out endpoint pairs can be used to assess decoded-sample validity, task-specific path cost, and numerical stability.

In all cases, the path energy takes the form $\mathcal{E}[\gamma] = \int \frac{1}{2}\|\dot\gamma\|_g^2 + \lambda\mathcal{U}(\gamma)\,dt$: the choice of $(\mathcal{M}, g)$ dictates the kinetic term, while $\mathcal{U}$ dictates the potential term.

4.3 Geodesic Force Matching: Solving via Direct Collocation

We seek approximate stationary paths of the relaxed energy by matching its Euler–Lagrange equation, which balances kinetic energy against manifold adherence.

Proposition 4.3 (Stationarity in ambient space)**.** *A smooth curve $\gamma : [0,1] \to \mathbb{R}^l$ is a critical point of $\mathcal{E}[\gamma] = \int_0^1 \frac{1}{2}\|\dot\gamma(t)\|^2 + \lambda\mathcal{U}(\gamma(t))\,dt$ (under a flat geometric prior) if and only if*

$$\ddot\gamma(t) = \lambda\nabla\mathcal{U}(\gamma(t)). \tag{8}$$

Eq. (8) describes a force balance: the geometric curvature (acceleration) is determined by the gradient of the potential field.[4]

Non-Euclidean priors. When a non-Euclidean prior is imposed (Section 4.2), Eq. (8) adapts naturally. For embedded submanifolds $\mathcal{M} \hookrightarrow \mathbb{R}^l$ with the induced metric (e.g., $\mathbb{S}^{l-1}$), the stationarity condition becomes $P_\gamma(\ddot\gamma) = \lambda P_\gamma(\nabla\mathcal{U})$, where P_z projects onto $T_z\mathcal{M}$. For a general metric $G(z)$, $\ddot\gamma$ is replaced by the covariant derivative $\nabla_{\dot\gamma}\dot\gamma$, and the Euclidean gradient becomes the Riemannian gradient $G^{-1}\nabla\mathcal{U}$ (Appendix C).

Direct collocation. We propose *Geodesic Force Matching* (GFM), a direct-collocation method that solves the boundary-value problem of Eq. (8) without sequential ODE integration. The name reflects the mechanical analogy: GFM matches the Euler–Lagrange force balance globally along the path. We discretize the path into N waypoints $\boldsymbol{z} = \{z_0, \dots, z_{N-1}\}$, fix both endpoints, and optimize the interior residual jointly:

$$\min_{\boldsymbol{z}} \mathcal{L}_{\text{GFM}} = \frac{1}{N}\sum_{i=1}^{N-2} \left\| P_{z_i}\big(\ddot\gamma_i + \lambda s_\theta(z_i, \sigma)\big)\right\|^2, \tag{9}$$

where $\ddot\gamma_i = (z_{i+1} - 2z_i + z_{i-1})/\Delta t^2$ is the finite-difference acceleration. By Eq. (7), a small residual indicates approximate Euler–Lagrange stationarity; it does not establish local or global minimality of the action.

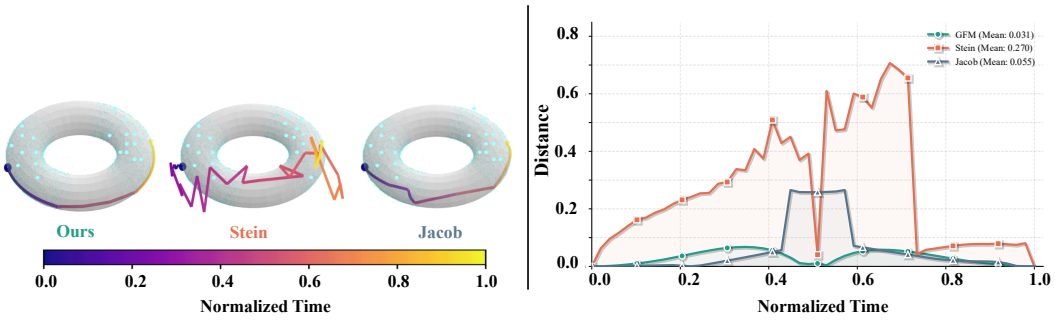
Algorithm 1: Geodesic Force Matching (GFM)

Input : z_0, z_{N-1}; score s_θ; weight λ; waypoints N, $\Delta t = 1/(N-1)$
Output : Trajectory $\boldsymbol{z}^*$
Init $\boldsymbol{z}^{(0)}$ via linear/SLERP interp.
while *not converged* **do**
  for $i = 1, \dots, N-2$ **do**
    $a_i \leftarrow (z_{i+1} - 2z_i + z_{i-1})/\Delta t^2$
    $r_i \leftarrow P_{z_i}\big(a_i + \lambda\,\mathrm{sg}[s_\theta(z_i, \sigma)]\big)$
  end
  $\boldsymbol{z}_{1:N-2} \leftarrow$
   $\boldsymbol{z}_{1:N-2} - \eta\,\nabla_{\boldsymbol{z}_{1:N-2}} \frac{1}{N}\sum_i \|r_i\|^2$
   `// step`
  $z_i \leftarrow R_{\mathcal{M}}(z_i), i = 1, \dots, N-2$
   `// retract`
end
return $\boldsymbol{z}$

Projector and retraction. The projector P_{z_i} encodes the geometric prior: $P_{z_i} = \mathbf{I}$ for the Euclidean prior, and $P_{z_i} = \mathbf{I} - z_i z_i^\top/\|z_i\|^2$ for the hyperspherical prior. For non-Euclidean $\mathcal{M}$, the tangent projector alone does not constrain waypoints to $\mathcal{M}$; we therefore apply a retraction $R_{\mathcal{M}}$ after each gradient step (Algorithm 1, retract step). For $\mathcal{M} = \mathbb{S}^{l-1}$ this is the closed-form $z \mapsto \sqrt{l}\,z/\|z\|$; in our experiments, the radius deviation remains numerically negligible throughout optimization.

Detached score. We treat $s_\theta(z_i, \sigma)$ as a fixed target at each iteration (detached from the computational graph), which gives a lagged-target gradient update with the score reevaluated at each iteration. At any fixed point where the residual vanishes, the stationarity condition holds regardless of whether the score Jacobian is included (Appendix F.2). This avoids backpropagation through the

[4] Note that unlike Newtonian mechanics where $\vec a = -\nabla V$ (which conserves $T + V$), the solution to the boundary value problem minimizing $\int (T + V)dt$ satisfies $\vec a = \nabla V$. The path bends toward low-potential regions because spending less time in high-potential regions reduces the integral.



(A) Interpolation with different methods

(B) Distance to Manifold over Time

Figure 2: **Interpolation across a data void on the torus.** Training samples are concentrated on the left and right. **(A)** Trajectories, color-coded by normalized time. **(B)** Distance to the complete analytic torus $\mathcal{T}$. R-GFM uses an analytic geometric prior to remain near $\mathcal{T}$ across the void; Jacob exhibits spikes and Stein drifts away.

score Jacobian and yields roughly an order of magnitude reduction in runtime and memory over the Jacobian-based baseline (Table 4). Across 600 tested ShapeNet paths with SLERP initialization, median residual reductions are 99.51–99.71% (Appendix E.8).

GFM extends to general Riemannian metrics using the contracted Christoffel term (Appendix C); for dense metrics, inversion costs $\mathcal{O}(l^3)$.

5 Experiments

We evaluate Geodesic Force Matching (GFM) on two settings of increasing complexity: a synthetic torus with controlled data voids (Section 5.1) and 3D shape interpolation on ShapeNet (Section 5.2). Implementation details are provided in Appendices D and E.

Baselines. We compare against four representative Riemannian metric learning approaches: **(1). RBF** [21, 43], which defines a conformal metric from radial basis functions centered at training points; **(2). LAND** [5, 9], a nonparametric metric using the regularized inverse of local data variance; **(3). Jacob** [37], the pullback metric $G(z) = J_s(z)^\top J_s(z)$ via the score map; and **(4). Stein** [7], which modifies the ambient geometry by $G = \mathbf{I} + \lambda ss^\top$. The first two derive geometry from training samples, while the latter two derive it from the diffusion score. Full method descriptions are in Appendix A.1.

5.1 Synthetic Experiment: Torus with Data Voids

We first stress-test all methods in a regime where the score signal is unreliable: a 2-torus in $\mathbb{R}^3$ with training samples concentrated in two disjoint clusters, leaving a void in between (Fig. 2A). The task is to interpolate across the void, a setting that falls outside the bounded-density assumption of Theorem 4.1 by design. Writing the complete analytic torus as $\mathcal{T}$, the observed support is $\mathcal{Z} = \mathrm{supp}(p_0) \subsetneq \mathcal{T}$; the void lies outside $\mathcal{Z}$. All methods share an intrinsic-torus initialization. R-GFM additionally uses the analytic torus frame to specify traversal costs across the void.

Results. Figure 2 shows that R-GFM remains near the complete torus across the void, while Stein drifts and Jacob exhibits sharp excursions. Here the finite-noise score supplies density-based attraction and the analytic frame supplies the geometry of continuation. The result demonstrates effective use of an informative prior under sparse coverage.

5.2 3D Shape Interpolation

Setup. We evaluate on the ShapeNet dataset [11] using the 3DShape2VecSet framework [49], which encodes shapes into a 512×8 latent space. The score is provided by a pretrained EDM denoiser [23]. We compare GFM and all baselines at two noise levels, $\sigma \in \{0.1, 0.3\}$. Within each setting, all

score-based methods use the same σ. The levels 0.3 and 0.1 probe smoother and sharper score landscapes, respectively; they and GFM's $\lambda = 1$ were selected empirically. Appendix E.5 shows matched visualizations at additional noise levels, and Appendix E.7 reports a paired sensitivity study.

Metrics. Across $N = 200$ random endpoint pairs and $T = 10$ interpolation steps, we measure Chamfer Distance (CD), Minimum Matching Distance (MMD), 99th-percentile CD (CD_{99}, capturing worst-case generation quality), and F-Score at threshold $\tau = 0.1$ against 5,000 reference samples (definitions in Appendix E.4).

5.3 Results Analysis

Quantitative results. Table 1 shows that at $\sigma = 0.3$, GFM is competitive with the strongest baselines. Stein leads on Sofa MMD/CD_{99} and Bed F@.1, while GFM leads on Bed CD_{99} and Sofa F@.1; Bed MMD ties at the reported precision. The gap widens at $\sigma = 0.1$ (Table 1), where Stein's MMD and CD_{99} roughly double on Sofa while GFM remains on the same order and attains the best MMD and F@.1 across both categories. This is consistent with our motivation: decoupling the metric from the data density allows GFM to operate at lower noise where geometric detail is preserved. RBF achieves competitive CD_{99} but low F@.1 across settings, suggesting it covers only sparse fragments of the manifold.

Table 1: **Quantitative comparison on ShapeNet at $\sigma \in \{0.1, 0.3\}$.** Mean $\pm$ standard deviation across 3 runs. **Bold**: best; underline: second-best. Metric definitions in Appendix E.4.

		Bed			Sofa		
σ	Method	MMD↓	CD_{99} ↓	F@.1↑	MMD↓	CD_{99} ↓	F@.1↑
0.3	RBF	$0.158_{\pm 0.019}$	$0.229_{\pm 0.014}$	$0.015_{\pm 0.014}$	$0.096_{\pm 0.014}$	$0.155_{\pm 0.018}$	$0.625_{\pm 0.017}$
	LAND	$0.200_{\pm 0.015}$	$0.274_{\pm 0.012}$	$0.006_{\pm 0.013}$	$0.158_{\pm 0.019}$	$0.249_{\pm 0.020}$	$0.060_{\pm 0.016}$
	Stein	$0.115_{\pm \mathbf{0.014}}$	$0.205_{\pm 0.019}$	$0.378_{\pm \mathbf{0.014}}$	$0.049_{\pm \mathbf{0.017}}$	$0.104_{\pm \mathbf{0.016}}$	$0.953_{\pm 0.017}$
	Jacob	$0.120_{\pm 0.017}$	$0.175_{\pm 0.013}$	$0.215_{\pm 0.012}$	$0.078_{\pm 0.019}$	$0.131_{\pm 0.019}$	$0.837_{\pm 0.019}$
	GFM (Ours)	$0.115_{\pm 0.019}$	$0.168_{\pm \mathbf{0.021}}$	$0.356_{\pm 0.019}$	$0.058_{\pm 0.019}$	$0.128_{\pm 0.017}$	$0.973_{\pm \mathbf{0.012}}$
0.1	RBF	$0.143_{\pm 0.016}$	$0.220_{\pm 0.020}$	$0.083_{\pm 0.020}$	$0.111_{\pm 0.013}$	$0.188_{\pm \mathbf{0.015}}$	$0.456_{\pm 0.020}$
	LAND	$0.259_{\pm 0.015}$	$0.400_{\pm 0.017}$	$0.000_{\pm 0.014}$	$0.238_{\pm 0.019}$	$0.394_{\pm 0.017}$	$0.008_{\pm 0.016}$
	Stein	$0.123_{\pm 0.017}$	$0.256_{\pm 0.017}$	$0.402_{\pm 0.021}$	$0.103_{\pm 0.018}$	$0.210_{\pm 0.019}$	$0.632_{\pm 0.015}$
	Jacob	$0.143_{\pm 0.016}$	$0.231_{\pm 0.020}$	$0.110_{\pm 0.019}$	$0.111_{\pm 0.021}$	$0.198_{\pm 0.021}$	$0.513_{\pm 0.019}$
	GFM (Ours)	$0.110_{\pm \mathbf{0.015}}$	$0.183_{\pm \mathbf{0.019}}$	$0.424_{\pm \mathbf{0.017}}$	$0.088_{\pm \mathbf{0.013}}$	$0.189_{\pm 0.022}$	$0.754_{\pm \mathbf{0.013}}$

Qualitative analysis. Figure 3A shows representative bed interpolations at $\sigma = 0.1$. Methods with entangled or local metrics (Land, RBF, Jacob) produce intermediates in which the source shape dissolves into scattered fragments before reassembling—consistent with their elevated CD_{99}. Stein retains structural cohesion but exhibits volume loss and geometric distortion. GFM produces a continuous deformation in which the frame expands and the mattress smoothly transforms.

5.4 Decoupling in Practice

A central claim of our framework is that manifold adherence (*where*) and path geometry (*how*) can be specified independently. We support this with two complementary experiments: a 4-way ablation on 3D shapes showing the distinct roles of the two components (Fig. 3B), and a torus study showing the metric can be varied while the score potential is held fixed (Fig. 4).

Complementary roles of potential and geometry. We compare four configurations on the ShapeNet setting of Section 5.2: *Linear* (no potential, Euclidean geometry), *SLERP* (no potential, spherical geometry), *Euclidean + Potential* (score potential with flat metric), and *Full* (score potential with spherical metric). Without the potential, *Linear* and *SLERP* pass through low-density regions and the decoder produces broken artifacts, showing that geometric structure alone cannot locate the data manifold. *SLERP* respects the spherical topology of the latent space but is "geometrically correct" on the hypersphere yet semantically invalid. Both *Euclidean + Potential* and *Full*

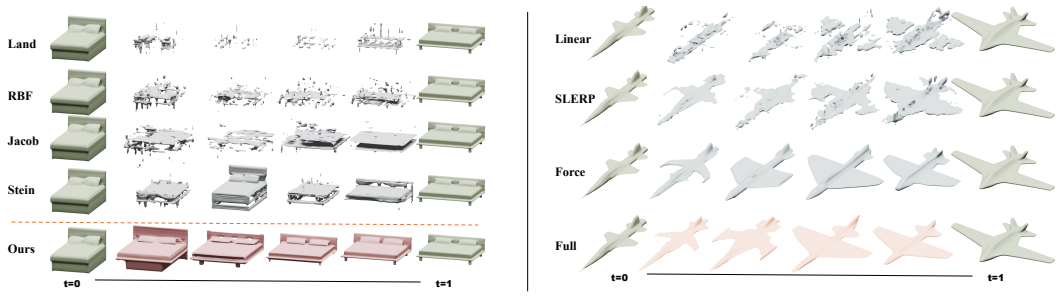


Figure 3: **(A)** GFM vs. baselines (bed category): baselines pass through off-manifold regions producing fragmented intermediates, while GFM produces a continuous deformation. **(B)** Ablation over the four configurations (airplane category): the score potential supports valid intermediates, while the geometric prior changes the traversal behavior.

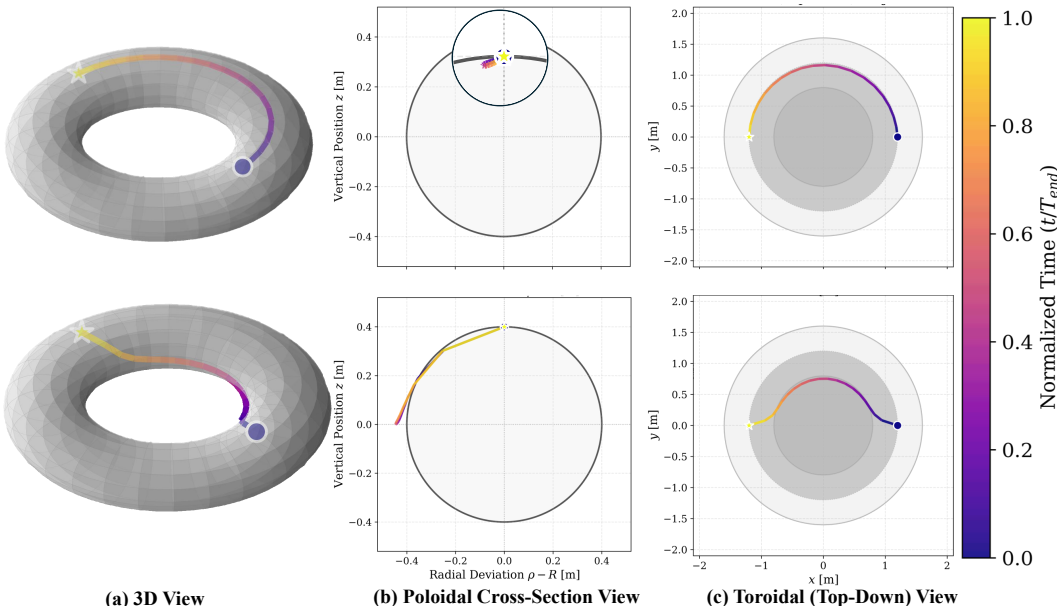


Figure 4: **Metric decoupling on the torus.** Two paths under different anisotropic metrics with the score-based potential held fixed. *Top:* toroidal-favoring metric produces a direct arc along the tube top. *Bottom:* poloidal-favoring metric detours via the inner rim. Both paths remain on-manifold despite different trajectories. Three views: (a) 3D rendering, (b) poloidal cross-section, (c) toroidal top-down view. Color encodes normalized time.

produce valid shapes; the difference between them reflects the choice of path geometry, not validity. Once the potential handles *where*, the geometric prior is free to specify *how*.

The metric can be varied independently. The 4-way ablation establishes that both components contribute, but does not directly show that the metric is a free axis once the potential is fixed. To probe this, we return to the torus setting (Appendix D.3) with *uniform* data coverage to isolate metric effects from data sparsity. We fix the score-based potential $\mathcal{U}$ and vary the anisotropic metric $G(x) = \alpha\,\mathbf{e}_\phi \otimes \mathbf{e}_\phi + \beta\,\mathbf{e}_\theta \otimes \mathbf{e}_\theta + \gamma\,\mathbf{n} \otimes \mathbf{n}$ between two configurations: a toroidal-favoring (α=5.0, β=0.1) and a poloidal-favoring (α=0.05, β=0.5), with γ=1.0 fixed. Both connect the same antipodal endpoints. Figure 4 shows that the two configurations produce qualitatively distinct paths: one a direct arc along the tube top, the other detouring through the inner rim. Both paths remain on the torus surface. This is the empirical counterpart of the additive structure in Eq. (2): the metric changes traversal behavior while both configurations retain adherence to the torus.

6 Conclusion

We presented a framework that disentangles manifold support from the choice of Riemannian metric by using the diffusion log-density as a soft potential, and introduced Geodesic Force Matching (GFM) to compute approximate stationary paths for the resulting variational problem. On synthetic manifolds and ShapeNet, GFM matches strong density-aware baselines at moderate noise and shows clearer advantages with informative priors in data-sparse regions and at low noise.

Limitations. The score-distance correspondence is asymptotic and assumes p_0 bounded away from zero; at finite σ and non-uniform p_0, the potential blends a distance term with an on-manifold density term. The noise scale σ and adherence weight λ are set empirically. The framework inherits the quality of the pretrained score, and extending beyond projection-based constraints to general dense metrics incurs $O(l^3)$ overhead from metric inversion. Promising directions include adaptive (σ, λ) scheduling, density-normalized variants of $\mathcal{U}$, and jointly learning task-specific metrics with the potential.

Acknowledgments and Disclosure of Funding

This work was supported by the National Institutes of Health (AG091653, AG068399, AG084375) and the Foundation of Hope.

Competing Interests

The authors declare no competing interests.

References

[1] Guillaume Alain and Yoshua Bengio. What regularized auto-encoders learn from the data-generating distribution. *The Journal of Machine Learning Research*, 15(1):3563–3593, 2014.

[2] Georgios Arvanitidis, Lars K Hansen, and Søren Hauberg. A locally adaptive normal distribution. *Advances in Neural Information Processing Systems*, 29, 2016.

[3] Georgios Arvanitidis, Lars Kai Hansen, and Søren Hauberg. Latent space oddity: on the curvature of deep generative models. *arXiv preprint arXiv:1710.11379*, 2017.

[4] Georgios Arvanitidis, Soren Hauberg, Philipp Hennig, and Michael Schober. Fast and robust shortest paths on manifolds learned from data. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1506–1515. PMLR, 2019.

[5] Georgios Arvanitidis, Søren Hauberg, and Bernhard Schölkopf. Geometrically enriched latent spaces. *arXiv preprint arXiv:2008.00565*, 2020.

[6] Georgios Arvanitidis, Miguel González-Duque, Alison Pouplin, Dimitris Kalatzis, and Søren Hauberg. Pulling back information geometry. *arXiv preprint arXiv:2106.05367*, 2021.

[7] Simone Azeglio and Arianna Di Bernardo. What's inside your diffusion model? a score-based riemannian metric to explore the data manifold. *arXiv preprint arXiv:2505.11128*, 2025.

[8] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8): 1798–1828, 2013.

[9] Louis Béthune, David Vigouroux, Yilun Du, Rufin VanRullen, Thomas Serre, and Victor Boutin. Follow the energy, find the path: Riemannian metrics from energy-based models. *arXiv preprint arXiv:2505.18230*, 2025.

[10] John T Betts. *Practical methods for optimal control and estimation using nonlinear programming*. SIAM, 2010.

[11] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[12] Nutan Chen, Alexej Klushyn, Richard Kurle, Xueyan Jiang, Justin Bayer, and Patrick Smagt. Metrics for deep generative models. In *International Conference on Artificial Intelligence and Statistics*, pages 1540–1550. PMLR, 2018.

[13] Ricky TQ Chen and Yaron Lipman. Flow matching on general geometries. *arXiv preprint arXiv:2302.03660*, 2023.

[14] Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. *Advances in neural information processing systems*, 35:2406–2422, 2022.

[15] Friso de Kruiff, Erik Bekkers, Ozan Öktem, Carola-Bibiane Schönlieb, and Willem Diepeveen. Pullback flow matching on data manifolds. *arXiv preprint arXiv:2410.04543*, 2024.

[16] Willem Diepeveen, Georgios Batzolis, Zakhar Shumaylov, and Carola-Bibiane Schönlieb. Score-based pullback riemannian geometry: Extracting the data manifold geometry using anisotropic flows. *arXiv preprint arXiv:2410.01950*, 2024.

[17] Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

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

[19] Chin-Wei Huang, Milad Aghajohari, Joey Bose, Prakash Panangaden, and Aaron C Courville. Riemannian diffusion models. *Advances in Neural Information Processing Systems*, 35:2750–2761, 2022.

[20] Jaehyeong Jo and Sung Ju Hwang. Generative modeling on manifolds through mixture of riemannian diffusion processes. *arXiv preprint arXiv:2310.07216*, 2023.

[21] Kacper Kapusniak, Peter Potaptchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael Bronstein, Joey Bose, and Francesco Di Giovanni. Metric flow matching for smooth interpolations on the data manifold. *Advances in Neural Information Processing Systems*, 37:135011–135042, 2024.

[22] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[23] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35: 26565–26577, 2022.

[24] Matthew Kelly. An introduction to trajectory optimization: How to do your own direct collocation. *SIAM review*, 59(4):849–904, 2017.

[25] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[26] John M Lee. *Introduction to Riemannian manifolds*, volume 2. Springer, 2018.

[27] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

[28] Aaron Lou, Derek Lim, Isay Katsman, Leo Huang, Qingxuan Jiang, Ser Nam Lim, and Christopher M De Sa. Neural manifold ordinary differential equations. *Advances in Neural Information Processing Systems*, 33:17548–17558, 2020.

[29] Emile Mathieu and Maximilian Nickel. Riemannian continuous normalizing flows. *Advances in neural information processing systems*, 33:2503–2515, 2020.

[30] Mike Yan Michelis and Quentin Becker. On linear interpolation in the latent space of deep generative models. *arXiv preprint arXiv:2105.03663*, 2021.

[31] Benjamin Kurt Miller, Ricky TQ Chen, Anuroop Sriram, and Brandon M Wood. Flowmm: Generating materials with riemannian flow matching. *arXiv preprint arXiv:2406.04713*, 2024.

[32] Liviu Nicolaescu. *An invitation to Morse theory*. Springer, 2007.

[33] Phillip Pope, Chen Zhu, Ahmed Abdelkader, Micah Goldblum, and Tom Goldstein. The intrinsic dimension of images and its impact on learning. *arXiv preprint arXiv:2104.08894*, 2021.

[34] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[35] Frederik Möbius Rygaard, Shen Zhu, Yinzhu Jin, Søren Hauberg, and Tom Fletcher. Likely interpolants of generative models. *arXiv preprint arXiv:2510.26266*, 2025.

[36] Shinnosuke Saito and Takashi Matsubara. Image interpolation with score-based riemannian metrics of diffusion models. *arXiv preprint arXiv:2504.20288*, 2025.

[37] Shinnosuke Saito and Takashi Matsubara. Be tangential to manifold: Discovering riemannian metric for diffusion models. *arXiv preprint arXiv:2510.05509*, 2025.

[38] Hang Shao, Abhishek Kumar, and P Thomas Fletcher. The riemannian geometry of deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 315–323, 2018.

[39] Noam Shazeer. Glu variants improve transformer. *arXiv preprint arXiv:2002.05202*, 2020.

[40] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

[41] Jan Pawel Stanczuk, Georgios Batzolis, Teo Deveney, and Carola-Bibiane Schönlieb. Diffusion models encode the intrinsic dimension of data manifolds. In *Forty-first International Conference on Machine Learning*, 2024.

[42] Joshua B Tenenbaum, Vin de Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

[43] Alessandra Tosi, Søren Hauberg, Alfredo Vellido, and Neil D Lawrence. Metrics for probabilistic geometries. *arXiv preprint arXiv:1411.7432*, 2014.

[44] S. R. S. Varadhan. Diffusion processes in a small time interval. *Communications on Pure and Applied Mathematics*, 20(4):659–685, 1967. doi: 10.1002/cpa.3160200404.

[45] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[46] Tom White. Sampling generative networks. *arXiv preprint arXiv:1609.04468*, 2016.

[47] Roderick Wong. *Asymptotic approximations of integrals*. SIAM, 2001.

[48] Qingtao Yu, Jaskirat Singh, Zhaoyuan Yang, Peter Henry Tu, Jing Zhang, Hongdong Li, Richard Hartley, and Dylan Campbell. Probability density geodesics in image diffusion latent space. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 27989–27998, 2025.

[49] Biao Zhang, Jiapeng Tang, Matthias Niessner, and Peter Wonka. 3dshape2vecset: A 3d shape representation for neural fields and generative diffusion models. *ACM Transactions On Graphics (TOG)*, 42(4):1–16, 2023.

A Additional Experiments Details

A.1 Baseline Method Descriptions

We provide full descriptions of the four baselines compared in Section 5.1 and 5.2.

RBF-based metrics. [21, 43] define a conformal metric $G(z) = \Omega(z)\mathbf{I}$, where the scaling factor $\Omega(z)$ is parameterized by a weighted sum of Radial Basis Functions centered at training points. This approach adapts the local volume but remains isotropic.

LAND metric. [5, 9] introduce a nonparametric, diagonal metric defined as the regularized inverse of local data variance. Unlike conformal methods, LAND explicitly captures local anisotropy by adapting to the directional spread of data without requiring a training phase.

Score-Jacobian metric. [37] defines a metric $G(z) = J_s(z)^\top J_s(z)$ as the pullback of the Euclidean metric via the score map $z \mapsto s_\theta(z)$. It penalizes directions where the score changes rapidly (the manifold normal), optimizing trajectories via the energy functional $\mathcal{E} \approx \sum \|s_\theta(z_{i+1}) - s_\theta(z_i)\|^2$.

Stein score metric. [7] propose a regularization-based approach that modifies the ambient geometry using the score outer product: $G = \mathbf{I} + \lambda ss^\top$. This penalizes motion along the score direction, effectively constraining the dynamics to the manifold's tangent space.

B Preliminaries on Diffusion Models

We provide the necessary background on diffusion models, specifically focusing on the *Elucidating the Design Space* (EDM) framework [23] employed in our experiments.

B.1 Diffusion and the EDM Framework

Diffusion models generate data by reversing a gradual noise corruption process. We consider a data distribution $p_0(z_0)$ supported on a manifold $\mathcal{Z} \subset \mathbb{R}^l$. The forward process defines a family of smoothed distributions $p_\sigma(z)$ obtained by convolving the data with isotropic Gaussian noise at continuous scales $\sigma > 0$:

$$p_\sigma(z) = \int p_0(z_0)\mathcal{N}(z; z_0, \sigma^2\mathbf{I})d\mathrm{vol}_{\mathcal{Z}}(z_0). \tag{10}$$

Unlike earlier formulations that rely on discrete timesteps, EDM treats the noise level σ as a continuous variable. The framework trains a *denoiser* network $D_\theta(z, \sigma)$ parameterized to estimate the clean data z_0 from a noisy observation $z = z_0 + n$, where $n \sim \mathcal{N}(0, \sigma^2\mathbf{I})$. This is known as the x-prediction parameterization.The model is trained by minimizing the weighted L_2 denoising error over a distribution of noise levels $p(\sigma)$:

$$\mathcal{L}_{\text{EDM}}(\theta) = \mathbb{E}_{\sigma,z,\epsilon}\left[\lambda(\sigma)\left\|D_\theta(z + \sigma\epsilon, \sigma) - z\right\|^2\right], \tag{11}$$

where $\lambda(\sigma)$ is a weighting function designed to balance the loss contribution across noise scales.

B.2 Score Estimation via Tweedie's Formula

Our Geodesic Force Matching framework requires the score function $\nabla_z \log p_\sigma(z)$ to define the manifold potential $\mathcal{U}(z)$. Although EDM is trained to predict the clean data z_0 rather than the score directly, the two are strictly equivalent via Tweedie's formula [17]. For a Gaussian noise kernel, the score of the smoothed density is proportional to the denoising residual:

$$\nabla_z \log p_\sigma(z) = \frac{\mathbb{E}[z_0 \mid z] - z}{\sigma^2}. \tag{12}$$

Using the optimal denoiser $D_\theta(z, \sigma) \approx \mathbb{E}[z_0 \mid z]$, we recover the score function required for our optimization:

$$s_\theta(z, \sigma) := \frac{D_\theta(z, \sigma) - z}{\sigma^2} \approx \nabla_z \log p_\sigma(z). \tag{13}$$

Equation (13) serves as the computational bridge in our framework:

- It provides the manifold restoring force pointing toward high-density regions (as utilized in Theorem 4.1).
- It defines the soft potential $\mathcal{U}(z) = -\log p_\sigma(z)$ used to penalize manifold violations during geodesic optimization.

B.3 Noise Level as a Geometric Scale

The noise level σ plays a dual role in our geometric setting. Theoretically, the manifold constraint is defined as $\sigma \to 0$. However, optimizing directly with a vanishing σ is numerically unstable because the score magnitude scales with $1/\sigma^2$, resulting in exploding gradients. Conversely, a large σ washes out geometric details. As demonstrated in Section 5.2, our framework leverages the robustness of the EDM denoiser to operate in a low-noise regime (e.g., $\sigma \approx 0.1$), which balances the fidelity of the manifold approximation with the smoothness required for stable trajectory optimization.

C Extension to General Riemannian Manifolds

In the main text, we focused on efficient solvers for embedded geometric priors where constraints are handled via projections. Here, we derive the Geodesic Force Matching (GFM) formulation for general Riemannian manifolds and detail its efficient implementation.

C.1 Riemannian Stationarity Conditions

Consider the latent space $\mathcal{M} = \mathbb{R}^l$ equipped with a position-dependent Riemannian metric $G(z)$. The variational energy functional is:

$$\mathcal{E}[\gamma] = \int_0^1 \frac{1}{2}\langle\dot\gamma(t), \dot\gamma(t)\rangle_{G(\gamma(t))} + \lambda\mathcal{U}(\gamma(t))dt. \tag{14}$$

The Euler–Lagrange stationarity equation balances the geometric covariant acceleration against the potential force:

$$\nabla_{\dot\gamma}\dot\gamma = \lambda \operatorname{grad}_G \mathcal{U}(\gamma). \tag{15}$$

In local coordinates $\gamma(t) = z(t)$, this expands to a system of ODEs. We isolate the geometric curvature term, denoted as the quadratic Christoffel vector $\mathbf{\Gamma}(\dot z, \dot z)$:

$$\ddot z(t) + \mathbf{\Gamma}(\dot z, \dot z) = \lambda G^{-1}(z)\nabla\mathcal{U}(z), \tag{16}$$

where the k-th component of the curvature vector is given by the contraction $\mathbf{\Gamma}^k(\dot z, \dot z) = \Gamma^k_{ij}(z)\dot z^i \dot z^j$.

C.2 Riemannian GFM Algorithm

We solve Eq. (16) by discretizing the trajectory into N waypoints $\boldsymbol{z} = \{z_0, \dots, z_{N-1}\}$ and fixed endpoints. Using the relation $\nabla\mathcal{U}(z) \approx -s_\theta(z)$ (since the score is the negative gradient of the potential), the force balance condition at each step becomes:

$$a_{\text{cov}}(z_i) - F_{\text{riem}}(z_i) \approx 0, \tag{17}$$

where a_{cov} approximates the LHS of Eq. (16) and F_{riem} approximates the RHS. The optimization objective is:

$$\mathcal{L}_{\text{R-GFM}}(\boldsymbol{z}) = \frac{1}{N}\sum_{i=1}^{N-2} \left\| (a_i + \mathbf{\Gamma}(v_i, v_i)) - (-\lambda G_i^{-1} s_\theta(z_i)) \right\|^2_{G(z_i)}. \tag{18}$$

Algorithm 2 details the implementation.

Algorithm 2: Riemannian Geodesic Force Matching

Input : Start z_0, End z_{N-1}, Score s_θ, Waypoints N, $\Delta t = 1/(N-1)$, Weight λ
Output : Optimized Trajectory $\boldsymbol{z}^*$

`/* 1. Initialization */`
Initialize $\boldsymbol{z}^{(0)}$ via Linear Interpolation;
`/* 2. Optimization Loop */`
while *not converged* **do**
  for $i \leftarrow 1$ **to** $N-2$ **do**
    `/* 1. Compute Metric and Inverse */`
    $G_i \leftarrow G(z_i)$;
    $G_i^{-1} \leftarrow \text{inverse}(G_i)$;
    `/* 2. Compute Kinematics */`
    $v_i \leftarrow (z_{i+1} - z_{i-1})/(2\Delta t)$ `// Velocity` $\dot{z}$
    $a_i \leftarrow (z_{i+1} - 2z_i + z_{i-1})/\Delta t^2$ `// Euclidean Accel` $\ddot{z}$
    `/* 3. Compute Geometric Curvature (Sec. C.3) */`
    $\boldsymbol{\Gamma}_{\text{quad}} \leftarrow \text{compute_gamma_vector}(G_i^{-1}, \partial G_i, v_i)$ `//` $\boldsymbol{\Gamma}(v_i, v_i)$
    `/* 4. Force Matching Residual */`
    $a_{\text{cov}} \leftarrow a_i + \boldsymbol{\Gamma}_{\text{quad}}$ `// Covariant Accel` $\nabla_{\dot{\gamma}}\dot{\gamma}$
    $F_{\text{riem}} \leftarrow -\lambda G_i^{-1} \text{sg}[s_\theta(z_i)]$ `// Data Force` $G^{-1}\nabla\mathcal{U}$
    $r_i \leftarrow a_{\text{cov}} - F_{\text{riem}}$
    `/* 5. Riemannian Loss */`
    $\ell_i \leftarrow r_i^\top G_i r_i$
  end
  $\mathcal{L}_{\text{Total}} \leftarrow \frac{1}{N}\sum \ell_i$
  Update $\boldsymbol{z}_{1:N-2}$ via Adam using $\nabla\mathcal{L}_{\text{Total}}$
end
return $\boldsymbol{z}$

C.3 Efficient Computation via Automatic Differentiation

Direct evaluation of the Riemannian acceleration $\ddot{z} = -\boldsymbol{\Gamma}(\dot{z}, \dot{z})$ requires instantiating the rank-3 Christoffel tensor, incurring $\mathcal{O}(l^3)$ costs. To implement the curvature computation step (compute Γ vector) in Algorithm 2 efficiently, we decompose this term using Lagrangian mechanics. Specifically, the term $\boldsymbol{\Gamma}(v, v)$ can be computed exactly using only Jacobian-Vector Products (JVP) and gradients, without forming the Christoffel symbols:

$$\boldsymbol{\Gamma}(v, v) = G(z)^{-1}\left[\mathcal{J}_{f(z)} \cdot v - \nabla_z \mathcal{K}(z, v)\right], \tag{19}$$

where:

- $v = \dot{z}$ is the velocity vector.
- $f(z) = G(z)v$ is the metric-velocity product map.
- $\mathcal{J}_f(z) \cdot v$ denotes the JVP of f along v, computing $(\partial_v G)v$ via forward-mode AD.
- $\mathcal{K}(z, v) = \frac{1}{2}v^\top G(z)v$ is the kinetic energy.

This decomposition reduces the cost of the curvature term to $\mathcal{O}(l^2)$, leaving the metric inversion as the sole $\mathcal{O}(l^3)$ bottleneck.

D More Details and Results about Synthetic Torus

D.1 Common Setup

Both experiments use a torus embedded in $\mathbb{R}^3$ with major radius $R = 1.2$ and minor radius $r = 0.4$, parametrized by toroidal angle $\theta \in [0, 2\pi)$ (around the central axis) and poloidal angle $\phi \in [0, 2\pi)$ (around the tube cross-section):

$$x = (R + r\cos\phi)\cos\theta, \quad y = (R + r\cos\phi)\sin\theta, \quad z = r\sin\phi. \tag{20}$$

We train an EDM-style denoiser [23] consisting of a 4-layer residual MLP (hidden dimension 128, sinusoidal noise embedding, $\sim$132K parameters) using the EDM loss with log-normal noise sampling ($P_{\text{mean}} = -1.2$, $P_{\text{std}} = 1.2$), Adam optimizer (learning rate 10^{-3}, batch size 2048, 1000 epochs).

Anisotropic metric. Both experiments use R-GFM (Alg. 2) with an anisotropic metric defined using the torus's natural frame—toroidal tangent $\mathbf{e}_\theta$, poloidal tangent $\mathbf{e}_\phi$, and surface normal $\mathbf{n}$:

$$G(x) = \alpha\, \mathbf{e}_\phi \otimes \mathbf{e}_\phi + \beta\, \mathbf{e}_\theta \otimes \mathbf{e}_\theta + \gamma\, \mathbf{n} \otimes \mathbf{n}, \tag{21}$$

where $\alpha, \beta, \gamma > 0$ control movement costs: small α favors poloidal motion (wrapping around the tube), small β favors toroidal motion (following the central axis), and large γ penalizes leaving the surface.

D.2 Out-of-Distribution Interpolation

This experiment tests interpolation across regions with no training data. Training data is concentrated in two disjoint clusters: Cluster A at $\theta \in (-\pi/3, \pi/3)$ and Cluster B at $\theta \in (\pi - \pi/3, \pi + \pi/3)$, leaving approximately one-third of the surface as a void. We sample $N = 20{,}000$ points with Gaussian noise ($\sigma_{\text{noise}} = 0.005$).

Let $\mathcal{T}$ be the complete torus and $\mathcal{Z}$ the support of the clean patch distribution before observation noise. The void $\mathcal{T} \setminus \mathcal{Z}$ is outside the data support. Gaussian smoothing makes p_σ positive, but its small-noise distance term is governed by $\mathcal{Z}$, not by the missing surface. R-GFM uses the hand-specified analytic frame to define traversal costs there.

All methods optimize 50 waypoints using Adam, initialized by linear interpolation in intrinsic torus coordinates.

Our R-GFM uses metric parameters $\alpha = 1.0$, $\beta = 0.05$, $\gamma = 1.0$, learning rate 5×10^{-3}, $\sigma = 0.01$, $\lambda = 1.0$, and 1500 iterations.

The Score-Jacobian baseline [37] uses learning rate 10^{-2}, 2000 iterations.

The Stein Score baseline [7] uses learning rate 10^{-4}, 2000 iterations, with score evaluated at segment midpoints.

We measure deviation from the complete analytic torus (including the void), rather than distance to the observed support: $d_{\text{torus}}(x, y, z) = \left| \sqrt{(\sqrt{x^2 + y^2} - R)^2 + z^2} - r \right|$.

D.3 Metric Decoupling

This experiment demonstrates that varying the Riemannian metric while fixing the score-based potential yields qualitatively different paths. All of them remain on-manifold but with distinct trajectories. We use uniform coverage ($N = 20{,}000$ points, $\sigma_{\text{noise}} = 0.005$) to isolate metric effects from data sparsity.

Configurations. We test two metric configurations (Table 2) with endpoints at antipodal positions $z_{start} = (1.2, 0, 0.4)$ and $z_{end} = (-1.2, 0, 0.4)$. All use $\sigma = 0.1$, $\lambda = 1.0$, gradient clipping (max norm 10), and linear initialization in intrinsic coordinates; Both Config A and B uses 20 waypoints with learning rate 10^{-3} all for 1500 iterations.

Table 2: Metric configurations for the decoupling experiment.

Config	α	β	γ	Behavior
A	5.0	0.1	1.0	Toroidal-favoring: direct arc along tube top
B	0.05	0.5	1.0	Poloidal-favoring: detours via inner rim

As shown in Figure 4, the two configurations produce qualitatively distinct paths. Config A (top row), with low toroidal cost, takes a direct arc along the tube top, maintaining nearly constant poloidal angle throughout. Config B (bottom row), with low poloidal cost, detours through the inner rim: the path descends from the tube top, traverses a shorter toroidal arc near the hole, then returns to the top at the endpoint, clearly visible in the toroidal view where the trajectory curves inward. Despite their

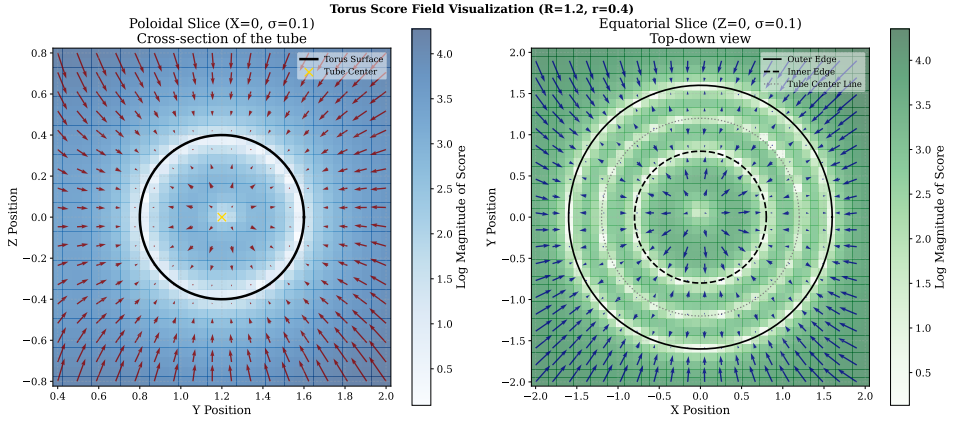


Figure 5: **Learned score field on the torus.** *Left:* Poloidal slice ($x = 0$) showing the tube cross-section. Score vectors point toward the torus surface from both inside and outside, with magnitude (color) decreasing near the surface. *Right:* Equatorial slice ($z = 0$) showing the top-down view. Notably, the region in the Tube Center (right, slim dashed line) exhibits low-magnitude scores with ambiguous directions, this is the *medial axis* where points are equidistant from multiple parts of the manifold.

different trajectories, both paths remain on the torus surface, showing that these two metric choices change traversal behavior while retaining surface adherence.

Score field visualization. We visualize the learned score field on the poloidal slice ($x = 0$) and equatorial slice ($z = 0$). The vectors point toward the torus surface from both outside and inside the hole, with near-zero magnitude on the surface itself, confirming correct geometry capture.

E More Details and Results about 3D Shape Interpolation Experiments

E.1 Dataset and Pretrained Models

We evaluate on the ShapeNet dataset [11], focusing on beds (category 5) and sofas (category 47). Each shape is represented as an occupancy field at 128^3 resolution, with point clouds of 2048 surface points for encoding.

Autoencoder. We use the 3DShape2VecSet framework [49], a Transformer-based VAE [25, 45] that encodes point clouds into latent vector sets. The encoder applies farthest point sampling to select 512 tokens, cross-attention to aggregate information. The decoder uses 24 Transformer blocks with GEGLU [39] activations to reconstruct occupancy values via cross-attention. Each latent has dimension $512 \times 8 = 4096$.

Diffusion model. We use an EDM-style diffusion model [23] operating in the autoencoder's latent space.

Both models we use the pre-trained weight from the official codebase of 3DShape2VecSet.

All of the experiments are conducted with a single NVIDIA RTX 6000 Ada GPU.

E.2 Hyperparameters

All methods optimize 10 waypoints initialized via SLERP. Table 3 summarizes the hyperparameters. For **RBF metric**, K is the number of centroids and for **LAND metric**, N_{ref} is the number of reference samples.

Table 3: Interpolation method hyperparameters. All methods use 10 waypoints with SLERP initialization and Adam optimizer.

Method	σ	λ	LR	Iterations	Other
GFM (Ours)	0.1, 0.3	1.0	0.01	400	Spherical projection
Jacob	0.1, 0.3	–	0.001	1000	Cosine annealing
Stein	0.1, 0.3	1000	0.001	500	Midpoint score eval
RBF	–	–	0.01	2000	K=1000
LAND	–	–	0.01	400	N_{ref}=1000

E.3 Implementation Details

Score computation. Given an EDM-style denoiser $D_\theta(z, \sigma)$ that predicts the clean sample, the score is computed as

$$s_\theta(z, \sigma) = \frac{D_\theta(z, \sigma) - z}{\sigma^2}. \tag{22}$$

SLERP initialization. All methods initialize waypoints via spherical linear interpolation. For endpoints z_{start}, z_{end} with unit directions $\hat{z}_i = z_i/\|z_i\|$ and angle $\theta = \arccos(\hat{z}_{start} \cdot \hat{z}_{end})$, the interpolated direction at parameter t is

$$\hat{z}(t) = \frac{\sin((1-t)\theta)}{\sin\theta}\hat{z}_{start} + \frac{\sin(t\theta)}{\sin\theta}\hat{z}_{end}, \tag{23}$$

with norms interpolated linearly.

Noisy optimization. For score-based methods (GFM, Jacob, Stein), optimization proceeds in noisy latent space: endpoints are perturbed as $\tilde{z}_i = z_i + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$, and the score is evaluated at the same σ. After optimization, each waypoint is denoised using 5 steps of the 2nd-order Heun sampler from $\sigma_{\text{start}} = \sigma$ to $\sigma_{\text{min}} = 0.002$. For all the methods we set $\sigma = 0.1$

E.4 Evaluation Metrics

Here we give the definitions of the evaluation metrics. Let $S_g = \{X\}$ and $S_r = \{Y\}$ denote the set of generated and reference shapes, respectively. We evaluate the quality of the generated distribution using the **Chamfer Distance** as the base metric between two point clouds X and Y:

$$d_{\text{CD}}(X, Y) = \frac{1}{|X|}\sum_{x\in X}\min_{y\in Y}\|x-y\|_2^2 + \frac{1}{|Y|}\sum_{y\in Y}\min_{x\in X}\|y-x\|_2^2 \tag{24}$$

On top of Chamfer Distance, we further have: **Minimum Matching Distance (MMD).** We use MMD to evaluate how well the reference modes are covered by the generated samples (corresponding to *Recall*). It is defined as the average distance from each reference shape to its nearest generated neighbor:

$$\text{MMD}(S_g, S_r) = \frac{1}{|S_r|}\sum_{Y\in S_r}\min_{X\in S_g} d_{\text{CD}}(Y, X) \tag{25}$$

99$^{\text{th}}$ Percentile Chamfer Distance (CD$_{99}$). To robustly assess the geometric integrity and identify "worst-case" generation failures (e.g., outliers or disjoint artifacts), we report the 99$^{\text{th}}$ percentile of the distances from generated shapes to the reference manifold:

$$\text{CD}_{99}(S_g, S_r) = P_{99}\left(\{\min_{Y\in S_r} d_{\text{CD}}(X, Y) \mid X \in S_g\}\right) \tag{26}$$

where $P_{99}(\cdot)$ denotes the 99$^{\text{th}}$ percentile of a set of scalars.

F-Score (F@τ). We define Precision and Recall at a fixed threshold $\tau = 0.1$ to quantify the percentage of "high-quality" samples and "covered" reference modes:

$$\text{Precision}(\tau) = \frac{1}{|S_g|}\sum_{X\in S_g}\mathbb{I}\left[\min_{Y\in S_r} d_{\text{CD}}(X, Y) < \tau\right] \tag{27}$$

$$\text{Recall}(\tau) = \frac{1}{|S_r|}\sum_{Y\in S_r}\mathbb{I}\left[\min_{X\in S_g} d_{\text{CD}}(Y, X) < \tau\right] \tag{28}$$

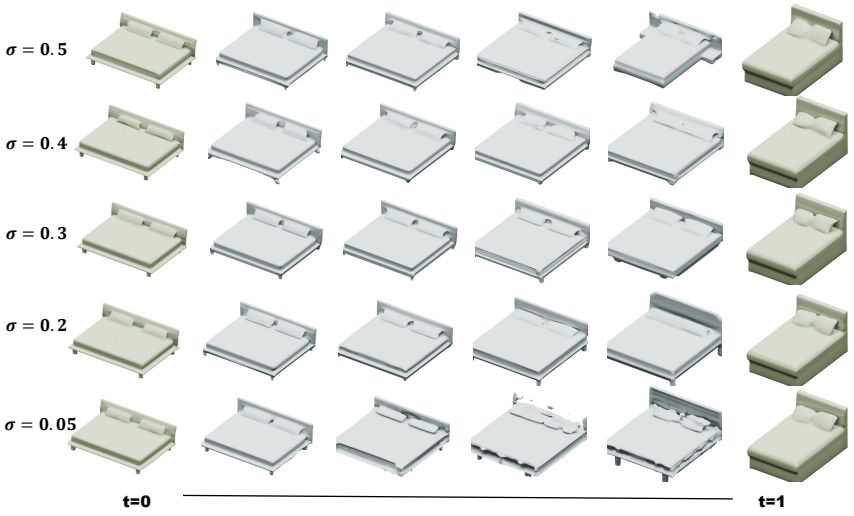


Figure 6: **GFM (Ours) across noise scales.** Interpolation trajectories generated by our method at $\sigma \in \{0.5, 0.4, 0.3, 0.2, 0.05\}$. Our approach maintains geometrically coherent and semantically meaningful interpolations across all noise levels, including the challenging low-noise regime ($\sigma = 0.05$) where the score field becomes highly peaked.

Table 4: The efficiency benchmark of different methods

Method	Time (s / 600 iters)	Memory (GB)
Jacob	423.18 ± 9.09	11.7
Stein	49.07 ± 0.79	1.3
GFM	41.75 ± 0.99	1.3

The F-Score is calculated as the harmonic mean:

$$\text{F-Score}(\tau) = \frac{2 \cdot \text{Precision}(\tau) \cdot \text{Recall}(\tau)}{\text{Precision}(\tau) + \text{Recall}(\tau)} \tag{29}$$

E.5 Additional Visualization Results (different noise scale)

As discussed in Section 5.2, the noise level σ critically affects both optimization stability and geometric fidelity. To provide a comprehensive view of this trade-off, we visualize interpolation trajectories for all score-based methods (Jacob, Stein, and Ours) across a range of noise scales $\sigma \in \{0.5, 0.4, 0.3, 0.2, 0.05\}$. As we can see from Figs. 6, 7, 8, at higher noise levels ($\sigma \geq 0.4$), all methods produce relatively stable trajectories, though fine geometric details are smoothed out. As σ decreases, the score magnitude increases (recall $\|s_\theta\| \propto 1/\sigma^2$), creating a more challenging optimization landscape. The visualizations reveal that baseline methods (Jacob and Stein) begin to exhibit instabilities, manifesting as fragmented geometry or off-manifold drift, and fail catastrophically at low noise ($\sigma = 0.05$). In contrast, our GFM framework maintains coherent interpolations even at $\sigma = 0.05$, demonstrating the robustness advantage of the direct collocation approach over methods that rely on well-behaved score fields throughout the trajectory.

E.6 Computational Benchmark

Here we test both memory and time efficiency of the optimization-based methods (**Stein**, **Jacob**, **GFM**). We test the efficiency with batch size 20, iteration 600 and repeat 3 times. One can see in Table 4 that Jacob is slow and memory consuming due to the Jacobian computation

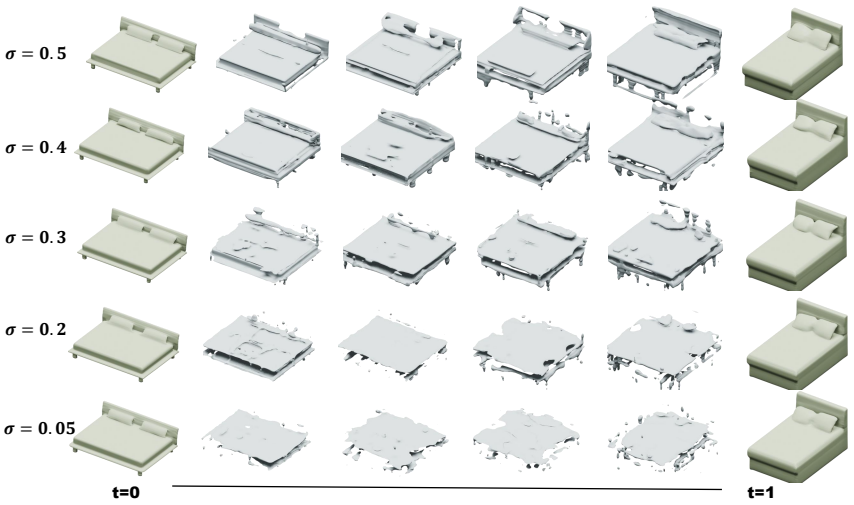


Figure 7: **Score-Jacobian metric (Jacob) across noise scales.** As the noise level decreases, the Jacobian of the score $J_s(z)$ becomes increasingly ill-conditioned in low-density regions, causing the optimization to destabilize. At $\sigma \leq 0.2$, trajectories exhibit fragmentation and off-manifold artifacts, reflecting the method's reliance on well-behaved score gradients throughout the path.

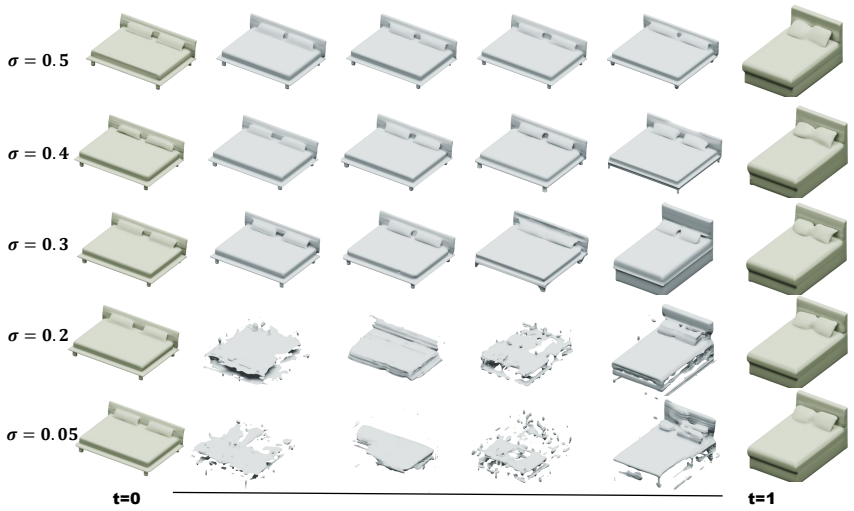


Figure 8: **Stein score metric across noise scales.** The Stein metric $G = I + \lambda ss^\top$ degrades gracefully at moderate noise levels but suffers from geometric distortions as σ decreases. At low noise, the score outer product $ss^\top$ dominates the metric, creating anisotropic warping that leads to volume loss and structural inconsistencies in the interpolated shapes.

E.7 Sensitivity to Noise and Adherence Weight

We evaluate $\lambda \in \{0, 0.1, 0.3, 1, 3, 10\}$ at $\sigma \in \{0.1, 0.3\}$ using 50 fixed endpoint pairs per category and three paired seeds. Endpoints and noise realizations are held fixed across λ; each run uses 10 waypoints, 400 optimization iterations, and five-step denoising. All conditions within a category share a pool of 300 clean reference endpoints. Tables 5 and 6 report mean and sample standard deviation across seeds. This reference pool differs from the 5,000-sample pool in Table 1, so comparisons are within this sweep.

The common stable range is $\lambda \in [0.1, 1]$, with the most consistent plateau at 0.3–1. At $\lambda = 3$, behavior depends on noise level; $\lambda = 10$ increases CD_{99} and reduces F@.1 in both categories. These trends support $\lambda = 1$ as an empirical working point across the two tested noise levels.

Table 5: Paired adherence-weight sweep at $\sigma = 0.1$. Mean $\pm$ sample standard deviation across three seeds.

Category	λ	MMD↓	CD_{99} ↓	F@.1↑
Bed	0	0.0767 ± 0.0011	0.1554 ± 0.0115	0.7160 ± 0.0099
	0.1	0.0764 ± 0.0013	0.1495 ± 0.0067	0.7464 ± 0.0146
	0.3	0.0763 ± 0.0016	0.1416 ± 0.0095	0.7807 ± 0.0079
	1	0.0759 ± 0.0015	0.1463 ± 0.0045	0.7988 ± 0.0120
	3	0.0766 ± 0.0010	0.2029 ± 0.0053	0.7166 ± 0.0034
	10	0.0774 ± 0.0009	0.2988 ± 0.0036	0.6393 ± 0.0122
Sofa	0	0.0637 ± 0.0005	0.1506 ± 0.0074	0.8693 ± 0.0087
	0.1	0.0635 ± 0.0003	0.1468 ± 0.0095	0.8870 ± 0.0035
	0.3	0.0631 ± 0.0005	0.1459 ± 0.0085	0.9139 ± 0.0045
	1	0.0628 ± 0.0004	0.1433 ± 0.0115	0.9204 ± 0.0062
	3	0.0633 ± 0.0004	0.2154 ± 0.0132	0.8524 ± 0.0102
	10	0.0639 ± 0.0004	0.2956 ± 0.0284	0.7036 ± 0.0053

Table 6: Paired adherence-weight sweep at $\sigma = 0.3$. Mean $\pm$ sample standard deviation across three seeds.

Category	λ	MMD↓	CD_{99} ↓	F@.1↑
Bed	0	0.0767 ± 0.0017	0.1421 ± 0.0027	0.7929 ± 0.0127
	0.1	0.0765 ± 0.0018	0.1377 ± 0.0039	0.8045 ± 0.0139
	0.3	0.0761 ± 0.0019	0.1335 ± 0.0034	0.8202 ± 0.0192
	1	0.0758 ± 0.0018	0.1389 ± 0.0044	0.8374 ± 0.0151
	3	0.0759 ± 0.0020	0.1400 ± 0.0106	0.8492 ± 0.0253
	10	0.0790 ± 0.0010	0.3229 ± 0.0097	0.4427 ± 0.0108
Sofa	0	0.0632 ± 0.0005	0.1424 ± 0.0092	0.9126 ± 0.0073
	0.1	0.0631 ± 0.0003	0.1399 ± 0.0082	0.9172 ± 0.0070
	0.3	0.0629 ± 0.0004	0.1374 ± 0.0111	0.9274 ± 0.0063
	1	0.0625 ± 0.0002	0.1317 ± 0.0100	0.9411 ± 0.0043
	3	0.0626 ± 0.0004	0.1309 ± 0.0117	0.9449 ± 0.0062
	10	0.0647 ± 0.0001	0.3752 ± 0.0130	0.5458 ± 0.0040

E.8 Optimization Reliability

We assess the discrete Euler–Lagrange residual under the prescribed SLERP initialization on the Bed and Sofa endpoint pairs. Let R_0 and R_f denote the path-level residual objective in Eq. (9) before and after optimization; the relative decrease is $(R_0 - R_f)/R_0$. Table 7 summarizes 150 paths per category/noise condition (600 in total). All paths complete the prescribed optimization without numerical or runtime failures, with median residual reductions of 99.51–99.71%. These statistics measure attainment of approximate stationarity for the tested endpoints and initialization.

Table 7: Residual reduction under SLERP initialization. Each reported median is computed across paths; relative decrease is the median of per-path ratios. Failures denote numerical or runtime failures.

Category	σ	Median R_0	Median R_f	Median decrease	Failures
Bed	0.1	35.13	0.1666	99.52%	0/150
Bed	0.3	3.760	0.01305	99.65%	0/150
Sofa	0.1	34.37	0.1645	99.51%	0/150
Sofa	0.3	3.932	0.01141	99.71%	0/150

F Proofs

F.1 Proof of Theorem 4.1

We prove the two parts separately, following the framework of [41] with additional details from [1].

F.1.1 Common Setup

Let p_0 be a probability distribution supported on $\mathcal{Z}$ with smooth density (with respect to the volume measure on $\mathcal{Z}$) bounded away from zero. The noise-perturbed distribution is:

$$p_\sigma(z) = \int_{\mathcal{Z}} \mathcal{N}(z; z_0, \sigma^2 \mathbf{I})\, p_0(z_0)\, d\mathrm{vol}_{\mathcal{Z}}(z_0). \tag{30}$$

By the Tubular Neighborhood Theorem [26][Theorem 5.25], every compact embedded submanifold $\mathcal{Z} \subset \mathbb{R}^l$ admits a uniform tubular neighborhood U_τ of radius $\tau > 0$ (the reach of $\mathcal{Z}$). Within U_τ, every point z has a unique orthogonal projection $\Pi(z) \in \mathcal{Z}$, and we can write $z = \Pi(z) + d(z) \cdot n(z)$ where $d(z) = \|z - \Pi(z)\|$ and $n(z)$ is the unit normal vector.

The following lemmas characterize the geometry of the distance function restricted to the manifold:

Lemma F.1 (Critical Point Structure, [41][Theorem D.8])**.** *Let $z \in U_\tau$ and define $f_z(z_0) := \|z - z_0\|^2$. Then:*

1. *$\Pi(z)$ is a critical point of $f_z|_{\mathcal{Z}}$.*
2. *$\Pi(z)$ is non-degenerate if and only if z is a non-focal point.*
3. *The index of f_z at $\Pi(z)$ equals the number of focal points on the segment from $\Pi(z)$ to z.*

Since the endpoint map restricted to U_τ is a diffeomorphism, there are no focal points within the tubular neighborhood. Hence $\Pi(z)$ is a non-degenerate minimum of $f_z|_{\mathcal{Z}}$.

Lemma F.2 (Gaussian Concentration, [41][Lemma D.9])**.** *For $z \in U_\tau$, there exists an open neighborhood $E \subset \mathcal{Z}$ of $\Pi(z)$ such that:*

$$\frac{\int_{\mathcal{Z} \setminus E} \mathcal{N}(z_0; z, \sigma^2 I)\, d\mathrm{vol}}{\int_E \mathcal{N}(z_0; z, \sigma^2 I)\, d\mathrm{vol}} \xrightarrow{\sigma \to 0} 0. \tag{31}$$

Proof sketch. By Lemma F.1, $\Pi(z)$ is a non-degenerate minimum of $f_z(z_0) := \|z - z_0\|^2$ on $\mathcal{Z}$. The Morse Lemma [32] provides coordinates where $f_z(z_0) = f_z(\Pi(z)) + \|\varphi(z_0)\|^2$ near $\Pi(z)$. Since $\mathcal{Z}$ is compact and $\Pi(z)$ is the unique minimizer, $a := \inf_{\mathcal{Z} \setminus E}(f_z - f_z(\Pi(z))) > 0$. Choose a smaller neighborhood $E' \Subset E$ of positive volume with $f_z(y) - f_z(\Pi(z)) \le a/2$ on E'. Then

$$\frac{\int_{\mathcal{Z} \setminus E} e^{-f_z/(2\sigma^2)}\, d\mathrm{vol}}{\int_E e^{-f_z/(2\sigma^2)}\, d\mathrm{vol}} \le \frac{\mathrm{Vol}(\mathcal{Z} \setminus E)}{\mathrm{Vol}(E')} e^{-a/(4\sigma^2)} \longrightarrow 0. \tag{32}$$

□

F.1.2 Proof of Part (i): Score as Restoring Force

Direct computation of the score yields:

$$\nabla_z \log p_\sigma(z) = \frac{\nabla_z p_\sigma(z)}{p_\sigma(z)} = \frac{\int_{\mathcal{Z}} \nabla_z \mathcal{N}(z; z_0, \sigma^2 I)\, p_0(z_0)\, d\mathrm{vol}}{\int_{\mathcal{Z}} \mathcal{N}(z; z_0, \sigma^2 I)\, p_0(z_0)\, d\mathrm{vol}}. \tag{33}$$

Using $\nabla_z \mathcal{N}(z; z_0, \sigma^2 I) = -\frac{z - z_0}{\sigma^2} \mathcal{N}(z; z_0, \sigma^2 I)$, we obtain Tweedie's formula [17]:

$$\nabla_z \log p_\sigma(z) = -\frac{z - \mathbb{E}_{p_\sigma(z_0|z)}[z_0]}{\sigma^2}, \tag{34}$$

where the posterior is $p_\sigma(z_0|z) \propto \mathcal{N}(z; z_0, \sigma^2 I) p_0(z_0)$.

It remains to show that $\mathbb{E}_{p_\sigma(z_0|z)}[z_0] \to \Pi(z)$ as $\sigma \to 0$. By Lemma F.2, the posterior concentrates near $\Pi(z)$. Applying Laplace's method [47] to

$$\mathbb{E}_{p_\sigma(z_0|z)}[z_0] = \frac{\int_{\mathcal{Z}} z_0\, p_0(z_0)\, e^{-\|z-z_0\|^2/(2\sigma^2)}\, d\mathrm{vol}}{\int_{\mathcal{Z}} p_0(z_0)\, e^{-\|z-z_0\|^2/(2\sigma^2)}\, d\mathrm{vol}}, \tag{35}$$

we introduce Morse coordinates $\varphi = (\varphi_1, \ldots, \varphi_k)$ near $\Pi(z)$ such that $\|z - z_0\|^2 = \|z - \Pi(z)\|^2 + \|\varphi\|^2$. Let $z_0(\varphi)$ denote the inverse coordinate map; then $z_0(0) = \Pi(z)$ and we can expand

$$z_0(\varphi) = \Pi(z) + J\varphi + O(\|\varphi\|^2), \tag{36}$$

where J is the Jacobian of the coordinate map at $\varphi = 0$. Under the Gaussian weight $\exp(-\|\varphi\|^2/(2\sigma^2))$, the leading odd contribution vanishes by symmetry. Expanding the smooth density and volume factor shows that the remaining linear and quadratic contributions are $O(\sigma^2)$. Therefore:

$$\mathbb{E}_{p_\sigma(z_0|z)}[z_0] = \Pi(z) + O(\sigma^2). \tag{37}$$

Substituting into Eq. (34):

$$\nabla_z \log p_\sigma(z) = -\frac{z - \Pi(z) - O(\sigma^2)}{\sigma^2} = -\frac{z - \Pi(z)}{\sigma^2} + O(1), \tag{38}$$

establishing part (i). Note that when $z \in \mathcal{Z}$, we have $\Pi(z) = z$, so the leading term vanishes—justifying Eq. (3) in the main text.

Corollary F.3 (Score Vanishing on Manifold)**.** *For any $z \in \mathcal{Z}$, we have $\Pi(z) = z$, hence:*

$$\sigma^2 \nabla_z \log p_\sigma(z) = -(z - \Pi(z)) + O(\sigma^2) = O(\sigma^2) \xrightarrow{\sigma \to 0} 0. \tag{39}$$

Conversely, for $z \in U_{\mathcal{Z}} \setminus \mathcal{Z}$, we have $z - \Pi(z) \neq 0$, so the limit is non-zero. This justifies the characterization in Eq. (4)*:*

$$\mathcal{Z} = \big\{ z \in U_{\mathcal{Z}} \;\big|\; \lim_{\sigma \to 0} \|\sigma^2 \nabla \log p_\sigma(z)\| = 0 \big\}. \tag{40}$$

Remark F.4. The restriction to $U_{\mathcal{Z}}$ is necessary: outside the tubular neighborhood, the score may vanish at points on the medial axis (e.g., the center of a sphere), which are equidistant from multiple parts of the manifold.

F.1.3 Proof of Part (ii): Log-Density as Distance Surrogate

Taking the negative logarithm of $p_\sigma(z)$:

$$-\log p_\sigma(z) = \frac{l}{2}\log(2\pi\sigma^2) - \log \int_{\mathcal{Z}} \exp\left(-\frac{\|z - z_0\|^2}{2\sigma^2}\right) p_0(z_0)\, d\mathrm{vol}_{\mathcal{Z}}(z_0). \tag{41}$$

By Lemma F.1, the function $f_z(z_0) := \|z - z_0\|^2$ attains its unique minimum on $\mathcal{Z}$ at $z_0 = \Pi(z)$, with minimum value $d(z, \mathcal{Z})^2$. Using Morse coordinates as in part (i):

$$\|z - z_0\|^2 = d(z, \mathcal{Z})^2 + \sum_{i=1}^{k} \varphi_i^2, \tag{42}$$

the integral becomes:

$$\begin{aligned} &\int_{\mathcal{Z}} \exp\left(-\frac{\|z - z_0\|^2}{2\sigma^2}\right) p_0(z_0)\, d\mathrm{vol}_{\mathcal{Z}}(z_0) \\ &\qquad = \exp\left(-\frac{d(z, \mathcal{Z})^2}{2\sigma^2}\right) \int \exp\left(-\frac{\|\varphi\|^2}{2\sigma^2}\right) p_0(\varphi) |J(\varphi)|\, d\varphi, \end{aligned} \tag{43}$$

where $|J(\varphi)|$ is the Jacobian of the coordinate transformation.

Applying the standard Laplace approximation [47]:

$$\int \exp\left(-\frac{\|\varphi\|^2}{2\sigma^2}\right) p_0(\varphi) |J(\varphi)|\, d\varphi = (2\pi\sigma^2)^{k/2} \cdot p_0(\Pi(z)) \cdot |J(0)| \cdot (1 + O(\sigma^2)), \tag{44}$$

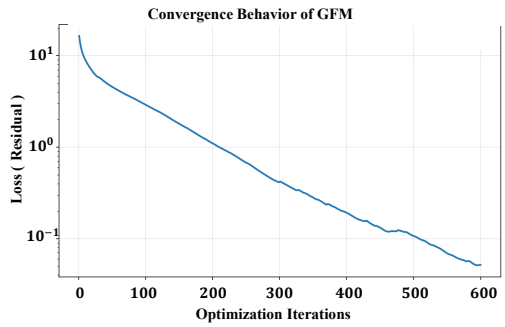


Figure 9: Representative residual trajectory for score-detached optimization on a 10-waypoint shape interpolation.

where $k = \dim(\mathcal{Z})$ and $|J(0)|$ encodes local geometric factors.

Combining terms:

$$-\log p_\sigma(z) = \frac{l}{2}\log(2\pi\sigma^2) + \frac{d(z,\mathcal{Z})^2}{2\sigma^2} - \frac{k}{2}\log(2\pi\sigma^2) - \log p_0(\Pi(z)) - \log|J(0)| + O(\sigma^2) \tag{45}$$

$$= \frac{d(z,\mathcal{Z})^2}{2\sigma^2} + \underbrace{\frac{l-k}{2}\log(2\pi\sigma^2)}_{C_\sigma} + O(1), \tag{46}$$

where the $O(1)$ term absorbs $-\log p_0(\Pi(z))$ and geometric factors, which are bounded on compact subsets of U_τ. This establishes part (ii). □

Behavior under non-uniform p_0. At finite σ, the residual includes $-\log p_0(\Pi(z))$ and local geometric factors. Thus $\mathcal{U}$ combines a leading distance penalty with an on-manifold density preference. For the torus with a void, $\mathcal{Z} \subsetneq \mathcal{T}$: the missing surface is outside the support, and the theorem does not identify its continuation. Section 5.1 instead tests how an analytic geometric prior complements a finite-noise score.

F.2 Stationarity of the Detached Iteration

Define the residual $r(z) := \ddot{\gamma}(z) + \lambda s_\theta(z)$ and the objective $L(z) = \|r(z)\|^2$. The gradient is

$$\nabla_z L = 2\sum_i r_i^\top \left(\frac{\partial \ddot{\gamma}_i}{\partial z} + \lambda \frac{\partial s_\theta(z_i)}{\partial z}\right). \tag{47}$$

At a point z^* where $r(z^*) = 0$, we have $\nabla_z L|_{z^*} = 0$ identically, independent of the score Jacobian term. Thus, every zero-residual point is stationary for both the full residual objective and its detached-score surrogate. This fixed-point statement does not establish convergence of the iteration or minimality of the action. □

Figure 9 shows a representative residual trace. Table 7 reports aggregate residual reductions and numerical/runtime failures across 600 paths under SLERP initialization.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Section 1 accurately reflect the contributions and scope. We claim competitive performance at moderate noise and clearer advantages with an informative torus prior in data-sparse regions and at low noise; this is supported by Section 5.1 (torus with voids) and Section 5.3 (ShapeNet at $\sigma \in \{0.1, 0.3\}$).

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Conclusion part discusses limitations explicitly: (i) Theorem 4.1 is asymptotic and assumes p_0 bounded away from zero; (ii) the noise scale σ and adherence weight λ are set empirically with no automatic heuristic; (iii) the framework inherits the quality of the pretrained score; (iv) dense metric inversion incurs $O(l^3)$ overhead. Section 4.3 distinguishes stationarity from minimality, and Section 5.1 identifies the supplied torus prior.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Theorem 4.1 states all assumptions (compact smooth submanifold with positive reach, bounded-below density), with a footnote on the relaxation when density assumption is violated. The full proof is in Appendix F.1, including Lemmas F.1 and F.2. Proposition 4.3 is stated and used in Section 4.3.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Algorithm 1 specifies the GFM procedure in main text. Full hyperparameters and dataset configurations are in Appendices D (synthetic torus) and E (ShapeNet). Pretrained weights are obtained from the publicly available 3DShape2VecSet codebase

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Code is provided with instructions to reproduce the main experimental results. ShapeNet and 3DShape2VecSet pretrained weights are publicly available via the original sources cited.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Appendix D provides full settings for the synthetic torus experiments (denoiser architecture, optimizer, learning rates, iteration counts, metric parameters). Appendix E provides full settings for ShapeNet experiments, including Table 3 summarizing per-method hyperparameters.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

Justification: Table 1 reports mean ± standard deviation across 3 independent runs for all methods and metrics. Tables 5 and 6 report mean ± sample standard deviation across 3 paired seeds. Table 4 reports mean ± standard deviation across 3 efficiency benchmark runs.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix E specifies all experiments use a single NVIDIA RTX 6000 Ada GPU. Table 4 reports per-method runtime and memory usage at fixed batch size and iteration count.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed the NeurIPS Code of Ethics and confirm that this work conforms to it. The work involves no human subjects, no scraped or proprietary data, and no deployment with direct societal impact.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.

- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Our method enables geometry-aware interpolation in pretrained generative model latent spaces, with positive applications in 3D content creation, scientific design, and shape analysis. Negative impacts are inherited from the underlying pretrained generative models (e.g., potential misuse for synthetic content generation); our method does not introduce additional risks beyond those already present in the underlying models.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: We do not release new pretrained generative models. Our method uses publicly available pretrained models (3DShape2VecSet, EDM denoiser) under their original licenses.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: ShapeNet [11], 3DShape2VecSet [49], and EDM [23] are properly cited. We use these assets according to their original licenses and terms of use.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: We release source code for the GFM solver.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: This work does not involve crowdsourcing or human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This work does not involve human subjects research.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs were used only for writing assistance (proofreading, phrasing) and did not contribute to the core methodology or scientific content of this research.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.