Horizontal Diffusion Models: Score-based Generative Modeling on Frame-Connection Geometry

Sungwoo Park
Korea University
sungwoo_park@korea.ac.kr

Abstract

Geometric modeling has become an essential paradigm in AI for Science, where data often inhabit intrinsically curved or structured spaces. However, existing manifold generative models often rely on closed-form or tractable access to global geometric primitives, such as geodesic distances, parallel transport, or spectral decompositions, which are rarely available beyond canonical geometries. To overcome this limitation, we propose Horizontal Diffusion Models (HDMs), a score-based generative framework defined on the orthonormal frame bundle of a Riemannian manifold. Instead of requiring global geometric operations, HDM lifts Euclidean diffusion processes through the Levi-Civita horizontal distribution, using only local metric and connection information while preserving intrinsic manifold geometry. This construction allows standard Euclidean score networks to be lifted into geometry-consistent and gauge-equivariant horizontal vector fields, bypassing the need for manifold-specific neural architectures. We further derive a horizontal KL objective that reduces to a Euclidean score-matching loss and analyze the curvature-dependent cost of the lift through a generalization bound. Experiments on parametric surfaces and scientific datasets demonstrate high-fidelity generation across diverse manifolds with nontrivial geometry and topology.

1 Introduction

Geometric modeling has become an essential paradigm in AI for Science, where data often inhabit intrinsically curved or structured spaces such as molecules, astrophysical objects, quantum systems, and materials with nontrivial topology or symmetry [Bronstein et al., 2021, Mathieu and Nickel, 2020]. In these scientific domains, capturing and generating data requires models that faithfully respect underlying geometric structures, whether they arise from molecular conformation spaces, orientation groups, biological manifolds, or celestial mechanics. However, the range of geometric structures that can be explored or modeled in practice has been limited by the structural assumptions and tractability of existing generative approaches. Most of existing methods remain fundamentally constrained to spaces that admit global coordinates, explicit embeddings, or closed-form geometric primitives, which excludes a wide variety of naturally occurring and scientifically relevant manifolds. As a result, scientific generative modeling often falls short of harnessing the full richness of geometry observed in real-world datasets, especially as complexity, dimensionality, or curvature increases.

A broad array of geometric generative methods including Riemannian flows, score-based diffusion models, and manifold-aware normalizing flows have been proposed to address these challenges [Brehmer and Cranmer, 2020, Mathieu and Nickel, 2020, De Bortoli et al., 2022, Huang et al., 2022, Thornton et al., 2022, Jo and Hwang, 2024, Bertolini et al., 2025, Courts and Kvinge, 2022]. These approaches typically rely on explicit Riemannian geometric constructions such as exponential and logarithmic maps, geodesic distances, or volume forms, and often require closed-form geodesic

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

computation, parallel transport, or spectral analysis of the Laplace-Beltrami operator. While powerful on highly symmetric or flat (*e.g.*, spheres, flat-torus), such methods encounter fundamental obstacles: explicit geodesics or spectral decompositions are unavailable and parallel transport is computationally intensive and not intrinsic to learned geometry. Consequently, the scientific impact of geometric generative models is limited not only by technical but also by foundational geometric barriers, restricting their reach in scientific applications.

To move beyond these fundamental geometric barriers, we seek a truly intrinsic approach that does not rely on explicit geodesic computations. An elegant solution is found in the geometric and probabilistic machinery of the *orthonormal frame bundle* $\mathcal{O}(\mathcal{M})$ over a Riemannian manifold $(\mathcal{M}, g)$, where Euclidean stochastic processes are lifted into tangent spaces via moving frames. This perspective has deep roots in stochastic differential geometry, originating from the theory of stochastic development and Brownian motion on manifolds [Eells and Sampson, 1964, Elworthy, 1982, Hsu, 2002]. In this framework, the Levi-Civita connection induces a horizontal distribution on $\mathcal{O}(\mathcal{M})$, enabling coordinate-free stochastic flows that preserve intrinsic geometry without relying on geodesics or spectral decompositions [Kobayashi and Nomizu, 1963, Elworthy, 1988, Malliavin, 1978]. Our work builds on this classical foundation and extends it to the generative modeling setting, providing a geometrically consistent framework that bridges differential geometry with modern learning paradigms. The main contribution of this work is twofold:

- **A New Framework of Riemannian Diffusion Models**: We propose a new geometric principle based on horizontal bundle dynamics that generalizes the standard Euclidean diffusion paradigm to Riemannian manifolds. Unlike previous methods constrained by global operation or specific symmetries, our connection-aware formulation serves as a universal foundation for geometric generative modeling.
- **Frame-Bundle Generality Across Riemannian Geometries:** While many existing methods are most mature on canonical geometries such as spheres or flat tori, where certain global operations are readily available, we provide empirical evidence that HDM remains effective across a wider range of geometric settings and scientific datasets considered in this work. In our experiments, HDM attains competitive sample quality and stable training behavior on several manifolds with nontrivial topology, suggesting it can serve as a practical approach for geometric generative modeling.

2 Horizontal Diffusion Models

This section introduces the basic structure of horizontal diffusion model, together with its associated training objective and geometric consistency induced by the frame bundle formulation. For notations not explicitly defined in this section, we refer the reader to Section 6.

Riemannian SGMs. To motivate our construction, we begin by recalling the conventional formulation of score-based generative modeling, which serves as the computational backbone for many diffusion models. Given arbitrary potential $\phi_\bullet : [0,T]\times\mathbb{R}^d \to \mathbb{R}$ and Euclidean Brownian motion W_t, consider the following pair of linear stochastic differential equations (SDEs), which defines a time-symmetric diffusion system in $\mathbb{R}^d$:

$$d\mathbf{E}_t = -\frac{1}{2}\nabla\phi_t(\mathbf{E}_t)dt + dW_t, \quad d\mathbf{E}_s = \left[-\frac{1}{2}\nabla\phi_s(\mathbf{E}_s) + \nabla\log\varrho_s(\mathbf{E}_s)\right]ds + dW_s.$$

These Euclidean forward–reverse SDEs have been widely employed in modern diffusion-based generative models, where the reverse drift involves the score function $\nabla\log\rho_t$ and is approximated by a neural network s_θ. In recent works De Bortoli et al. [2022], this framework has been generalized to *Riemannian manifolds* by replacing the Euclidean gradient ∇ with the Riemannian gradient $\nabla^{\mathcal{M}}$, yielding the Riemannian forward-reverse dynamics with Riemannian potential $\phi_\bullet : [0,T]\times\mathcal{M}\to\mathbb{R}$:

$$d\mathbf{X}_t = -\frac{1}{2}\nabla^{\mathcal{M}}\phi_t(\mathbf{X}_t)dt + dW_t^{\mathcal{M}},\ d\mathbf{X}_s = \left[-\frac{1}{2}\nabla^{\mathcal{M}}\phi_s(\mathbf{X}_s) + \nabla^{\mathcal{M}}\log\rho_s^{\mathcal{M}}(\mathbf{X}_s)\right]ds + dW_s^{\mathcal{M}},$$

where the diffusion $\mathbf{X}_t \in \mathcal{M}$ is governed by Riemannian Brownian motion $W^{\mathcal{M}}$, and all vector fields respect the intrinsic geometry. Despite providing a principled extension of diffusion modeling to curved spaces, these models depend critically on closed-form analytic expressions for geometric operators such as geodesic distance computations $(d_{\mathcal{M}})$, exponential and logarithmic maps, manifold-specific distribution estimations $(\rho_t^{\mathcal{M}})$, and tailored Riemannian score networks $(s_\theta^{\mathcal{M}})$ as highlighted in Table 1. These requirements introduce computational and analytical challenges

Method	$d_{\mathcal{M}}$	exp	log	$\rho_t^{\mathcal{M}}$	$s^{\mathcal{M}}$	g	Γ_{ij}^k	$\mathcal{G}$	s^{Euc}
Closed-form?	✗	✗	✗	✗	✗	✓	✓	✓	✓
De Bortoli et al. [2022]	✓	✓	✓	✓	✓	✗	✗	✗	✗
Thornton et al. [2022]	✓	✓	✓	✓	✓	✗	✗	✗	✗
Chen and Lipman [2024]	✓	✓	✓	✓	△	✗	✗	✗	✗
Diepeveen et al. [2025]	△	△	△	✓	△	✗	✗	✗	✗
Ours	✗	✗	△	✗	✗	✓	✓	✓	✓

Table 1: Comparison of geometric operators used in existing Riemannian generative models.

on curved manifolds, since closed-form expressions are often unavailable. Moreover, the absence of gauge equivariance ($\mathcal{G}$) makes it difficult to move beyond highly symmetric settings, where connection structure and fiber-wise symmetries become indispensable for accurate modeling.

Our framework uniquely addresses these limitations by avoiding manifold-specific closed-form operations while instead leveraging Euclidean score networks (s_θ^{Euc}). In particular, by operating directly on the orthonormal frame bundle, we make use of horizontal lifts defined through connection induced vector fields to ensure that the intrinsic structures such as Riemannian metric (g), Christoffel symbols (Γ_{ij}^k), and gauge equivariance ($\mathcal{G}$) are naturally incorporated as part of the overall formulation. As a result, the method avoids relying on global closed-form operations while retaining structure-preserving dynamics, and it can be implemented efficiently using local geometric information, making it practical for high-quality generative modeling on a broad range of complex manifolds. Please refer to Section 7 for a detailed comparison.

Frame-Connection Geometry. For rigor, we review the geometric structure of the orthonormal frame bundle. Additional notations not included in the main paper are provided in the Section 6.

Let $(\mathcal{M}, g)$ be a connected, oriented, d-dimensional Riemannian manifold. The *orthonormal frame bundle* $\mathcal{O}(\mathcal{M})$ is a principal $O(d)$-bundle over $\mathcal{M}$, with projection map $\pi : \mathcal{O}(\mathcal{M}) \to \mathcal{M}$. Each point $\mathbf{U} = (x, e) \in \mathcal{O}(\mathcal{M})$ consists of a base point $x \in \mathcal{M}$ and a linear isometry $e : \mathbb{R}^d \to T_x\mathcal{M}$, called a *frame*, which maps the standard Euclidean basis to a basis of the tangent space at x satisfying $e^\top g(x)e = I$. The bundle admits a natural right action by $O(d)$: for $h \in O(d)$, the action is given by $\mathbf{U} \cdot h = (x, eh)$. This group action preserves the orthonormality of frames and defines the structure of the principal bundle. The Levi–Civita connection $\nabla_{\mathcal{M}}$ on a Riemannian manifold $(\mathcal{M}, g)$ induces a *principal connection* on the orthonormal frame bundle $\mathcal{O}(\mathcal{M})$ via a $\mathfrak{so}(d)$-valued 1-form $\omega_{\mathbf{U}}(V) := e^{-1}\left(\nabla_{\pi_* V} e\right)$ for any $V \in T_{\mathbf{U}}\mathcal{O}(\mathcal{M})$. With the connection, a tangent vector V is said to be *horizontal* if $\omega_{\mathbf{U}}(V) = 0$, and the corresponding horizontal distribution is given by $\mathrm{H}_{\mathbf{U}} := \ker \omega_{\mathbf{U}} \subset T_{\mathbf{U}}\mathcal{O}(\mathcal{M})$. The canonical basis for this distribution consists of horizontal vector fields $\{H_a := H_e[\varepsilon^a]\}_{a=1}^d$, where each H_a satisfies $\omega(H_a) = 0$ and $\pi_* H_a = e_a := e(\varepsilon_a)$, with $\{\varepsilon_a\}_{a=1}^d$ the standard basis of $\mathbb{R}^d$.

To lift Euclidean forward–reverse SDEs to the frame bundle in a geometry-consistent way, we define a family of horizontal vector fields $H_e(\mathbf{U})[w]$ that map directions $w \in \mathbb{R}^d$ to elements of the horizontal distribution $H_{\mathbf{U}}$ at $\mathbf{U} = (x, e)$. Each $H[w]$ is characterized by $\pi_* H_e(\mathbf{U})[w] = e(w) \in T_x\mathcal{M}$ and $\omega_{\mathbf{U}}(H_e(\mathbf{U})) = 0$. Hence, it represents the horizontal lift of the tangent vector $e(w) \in T_x\mathcal{M}$. In local coordinates, using Christoffel symbols $\Gamma_{ij}^k(x)$, the horizontal lift conditioned on Euclidean vector w can be written in coordinate-free form as a differential operator acting on smooth functions $f \in C^\infty(\mathcal{O}(\mathcal{M}))$:

$$T_{\mathbf{U}}\mathcal{O}(\mathcal{M}) \ni H_e(\mathbf{U})[w]\, f = w^a e_a^i \frac{\partial f}{\partial x^i} - \left(w^a \Gamma_{ij}^k(x)\, e_a^j \frac{\partial f}{\partial e_i^k} \right) \xleftrightarrow{\text{Horizontal Lift}} w \in T_w\mathbb{R}^d. \quad (1)$$

The operator H is central to our construction, as it describes the *horizontal lift* of a Euclidean vector field to the frame-valued vector fields, preserving the intrinsic geometry defined by the Levi–Civita connection. This lift ensures that stochastic dynamics remain consistent with the manifold geometry without the need for global geodesic or exponential maps. This construction will serve as a computationally feasible mechanism for importing the Euclidean score-based framework into manifold settings, without requiring well-known special geometric structure.

Horizontal Diffusion Models. To overcome the computational limitations of conventional Riemannian score-based modeling, we formulate the diffusion dynamics on the orthonormal frame bundle. In this lifted setting, the Riemannian gradient $\nabla^{\mathcal{M}}$ is replaced by the horizontal gradient ∇^{H}. This

formulation defines the Stratonovich forward and reverse-time SDEs entirely through horizontal flows on frame-connection geometry:

$$d\mathbf{U}_t = -\frac{1}{2}\nabla^{\mathrm{H}}\Phi_t(\mathbf{U}_t)\,dt + H_a \circ dW_t^a,\; d\mathbf{U}_s = \left[-\frac{1}{2}\nabla^{\mathrm{H}}\Phi_s(\mathbf{U}_s) + s^{\mathrm{H}}(s,\mathbf{U}_s)\right]ds + H_a \circ dW_s^a, \quad (2)$$

where the reverse drift term is governed by the *horizontal score function*, defined as $s^{\mathrm{H}} := \nabla^{\mathrm{H}} \log \rho^{\mathrm{H}}$. We designate this framework as *Horizontal Diffusion Models* (HDMs). By elevating the stochastic dynamics to the orthonormal frame bundle, HDMs provide a generalization of the standard Riemannian formulation De Bortoli et al. [2022], thereby circumventing the need for global operation. Figure 1 visually elucidates the structural correspondence induced by the canonical projection $\pi : \mathcal{O}(\mathcal{M}) \to \mathcal{M}$, namely $\mathbf{X}_t = \pi(\mathbf{U}_t)$ and $\mathbf{X}_s = \pi(\mathbf{U}_s)$, linking the projected Riemannian diffusion processes $(\mathbf{X}_t, \mathbf{X}_s)$ on $\mathcal{M}$ to the intrinsic horizontal diffusion $(\mathbf{U}_t, \mathbf{U}_s)$ on $\mathcal{O}(\mathcal{M})$.

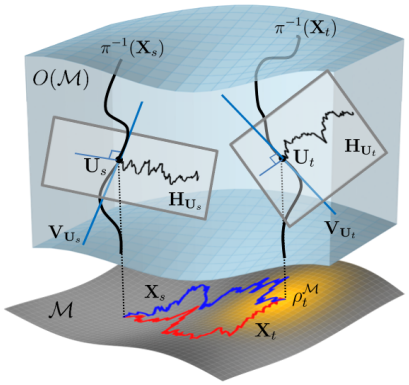


Figure 1: Structural correspondence between Riemannian and Horizontal Diffusion. Our horizontal diffusion evolves as a lifted trajectory on the orthonormal frame bundle **(black)** along horizontal directions induced by the Levi–Civita connection, with vertical directions describing frame rotations. Riemannian diffusion is the projected process on $\mathcal{M}$, shown for the forward process **(red)** and the reverse process **(blue)**.

In the forward process, HDMs inject *horizontal noise* on $\mathbf{H}_{\mathbf{U}_t}$ until the system reaches a stationary distribution $\mathbf{U}_\infty \sim \rho_\infty^{\mathrm{H}}$. Meanwhile, the reverse process progressively removes noise to recover the initial distribution $\mathbf{U}_0 \sim \rho_0^{\mathrm{H}}$. The canonical projection $\pi : \mathcal{O}(\mathcal{M}) \to \mathcal{M}$, defined by $\pi(x, e) = x$, links the bundle geometry to the data manifold. Specifically, the target data distribution is recovered via the pushforward of the reversed trajectory, denoted as $\rho_0^{\mathcal{M}} = \pi_{\#}\rho_0^{\mathrm{H}}$. Consequently, the ultimate goal of score-based modeling here is to approximate the true score function s^{H} with a neural network s_θ^{H}, thereby reconstructing the data distribution $\rho_0^{\mathcal{M}}$.

To formalize the connection between the bundle and the manifold, Proposition 9.4 and Corollary 9.6 establish the existence of the time-dependent marginal distribution ρ_t^{H} on the frame bundle. This is defined via the joint evolution of the initial frame $\mathbf{U}_0$ and the lifted process $\mathbf{U}_t$:

$$\rho_t^{\mathrm{Hor}}(\mathbf{U}) = \int_{\mathcal{O}(\mathcal{M})} p_t^{\mathrm{Hor}}(\mathbf{U} \mid \mathbf{U}_0)\rho_0(\mathbf{U}_0)d\lambda_0(\mathbf{U}_0) \xrightarrow{t\to\infty} \exp\left(-\Phi(\mathbf{U})\right), \quad (3)$$

where λ_0 is the volume measure on $\mathcal{O}(\mathcal{M})$ and $p_t^{\mathrm{H}}(\cdot|\cdot)$ denotes the conditional transition kernel. Finally, while HDMs operate on the orthonormal frame bundle $\mathcal{O}(\mathcal{M})$, our primary objective is to model distributions on the base manifold $\mathcal{M}$. The lifted process allows us to define the marginal manifold distribution by integrating out the fiber components using the Haar measure: $\mathbf{X}_t \sim \rho_t^{\mathcal{M}}(x) := \int_{\pi^{-1}(x)} \rho_t^{\mathrm{H}}(\mathbf{U})\, d\mathrm{Haar}(\mathbf{U})$, where the fiber $\pi^{-1}(x)$ is isomorphic to $O(d)$. The shared region of Figure 1 colored in yellow depicts the Riemannian stationary distribution $\rho_\infty^{\mathcal{M}}$.

Horizontal Score-Matching. With the horizontal diffusion processes in place, we next introduce the *score-matching principle* as the foundation of our learning objective. We aim to develop a score-matching framework formulated directly on the frame bundle as an extended counterpart of conventional approaches. In this setting, the learnable score field s^{H} is constrained to lie in the horizontal distribution, and the discrepancy between forward and reverse drifts is evaluated within this structure. To formalize, we introduce a geometry-aware divergence functional called a *horizontal Kullback–Leibler divergence* as follows:

Proposition 2.1 (Horizontal KL Divergence). *Let $b_t, \tilde{b}_t$ be horizontal drifts on $\mathcal{O}(\mathcal{M})$ inducing path laws $\boldsymbol{\nu}_b$ and $\boldsymbol{\nu}_{\tilde{b}}$. With Sasaki norm $\|\cdot\|_H$, parallel transport $\Gamma^{\mathcal{O}(\mathcal{M})}$, horizontal projection $\mathbf{P}_{\mathrm{H}}(Z) := Z - \omega(Z)^\sharp \cdot U$ [Kobayashi and Nomizu, 1963],*

$$\mathbf{KL}_{\mathrm{H}}[\boldsymbol{\nu}_{\tilde{b}} \| \boldsymbol{\nu}_b] = \frac{1}{2}\,\mathbb{E}_{\boldsymbol{\nu}_{\tilde{b}}}\left[\int_0^T \left\|\mathbf{P}_{\mathrm{H}}(\Delta_t)\right\|_H^2 dt\right], \quad \Delta_t := \Gamma^{\mathcal{O}(\mathcal{M})}_{\mathbf{U}_t^\theta \to \mathbf{U}_t}\left(\tilde{b}_t(\mathbf{U}_t^\theta)\right) - b_t(\mathbf{U}_t). \quad (4)$$

The projection operator $\mathbf{P}_{\mathrm{H}}$ serves to constrain all vector fields within the horizontal distribution, thereby guaranteeing both geometric consistency and invariance under gauge transformations. However, despite the rigorous theoretical basis provided by Proposition 2.1 for defining the horizontal KL divergence, practical numerical implementation remains challenging, primarily due to the necessity of path-wise parallel transport computations. To address the aforementioned issue, we introduce a computationally feasible alternative leveraging gauge-equivariance, thereby enabling efficient adaptation of standard Euclidean neural network architectures.

Gauge-Equivariant Horizontal Lifting. To efficiently evaluate the Horizontal KL divergence, we employ the horizontal lifting map Ψ defined in (5). This mapping projects Euclidean paths $E_\bullet \in C([0,T],\mathbb{R}^d)$ into the *gauge-equivariant quotient path space*, thereby avoiding direct vector field comparisons on the frame bundle:

$$\Psi : C([0,T],\mathbb{R}^d) \xrightarrow{H_e} C([0,T],\mathcal{O}(\mathcal{M}))/\mathcal{G}, \quad \Psi(E_\bullet) := [U_\bullet] = \{U_\bullet \cdot g_\bullet \mid g_\bullet \in \mathcal{G}\}. \tag{5}$$

where $\mathcal{G} := C([0,T],O(d))$ denotes the group of time-dependent gauge transformations. This embedding ensures that geometric evaluations are performed at canonical representatives in the quotient space, thereby eliminating the need for explicit parallel transport. We refer to this formulation as *lifted Euclidean parameterization*, where the gradient of potential and the horizontal score function are consistently parameterized by their Euclidean counterparts:

Proposition 2.2. *Let $\mathbf{E}_t \in \mathbb{R}^d$ be the solution to Euclidean Stratonovich SDE, and let $\mathbf{U}_t = (x_t, e_t) \in \mathcal{O}(\mathcal{M})$ be its horizontal lift defined by $d\mathbf{U}_t = H_{e_t}(\mathbf{U}_t)\big[\circ\, d\mathbf{E}_t\big]$. Then, for every $t \geq 0$,*

$$\nabla^{\mathrm{Hor}}\Phi(\mathbf{U}_t) = H_{e_t}(\mathbf{U}_t)\big[\nabla\phi(\mathbf{E}_t)\big], \quad \nabla^{\mathrm{Hor}}\log\rho_t^{\mathrm{Hor}}(\mathbf{U}_t) = H_{e_t}(\mathbf{U}_t)\big[\nabla\log\rho_t(\mathbf{E}_t)\big], \tag{6}$$

$$s^{\mathrm{H}}(t,\mathbf{U}_t) := \nabla^{\mathrm{Hor}}\log\rho_t^{\mathrm{Hor}}(\mathbf{U}_t) = H_{e_t}(\mathbf{U}_t)\big[s(t,\mathbf{E}_t)\big] = \sum_{a=1}^{d} s^a(t,\mathbf{E}_t)H_a(\mathbf{U}_t). \tag{7}$$

The result follows directly from the definition of the quotient mapping Ψ. This geometric parameterization enables a significant simplification of the theoretical framework. By expressing all horizontal quantities in terms of the latent Euclidean process, our formulation bypasses the need for explicit parallel transport or projection operations, which are otherwise computationally expensive and numerically sensitive. Consequently, the horizontal KL divergence introduced in (4) admits the following Euclidean-form objective, which remains faithful to the geometry $\mathcal{O}(\mathcal{M})$:

$$\mathbf{KL}_{\mathrm{H}}(\theta) := \mathbf{KL}_{\mathrm{H}}[\boldsymbol{\nu}_b \parallel \boldsymbol{\nu}_{\tilde{b}^\theta}] = \frac{1}{2}\int_0^T \sigma_t^2 \mathbb{E}_{\mathbf{E}_t\sim\rho_t}\left[\|s(t,\mathbf{E}_t) - s_\theta(t,\mathbf{E}_t)\|_E^2\right] dt. \tag{8}$$

Here, $b = \frac{1}{2}\nabla^{\mathrm{H}}\Phi + s^{\mathrm{H}}$ and $\tilde{b}^\theta = \frac{1}{2}\nabla^{\mathrm{H}}\Phi + s_\theta^{\mathrm{H}}$ denote the true (*i.e.*, s^{H}) and model-induced (*i.e.*, s_θ^{H}) reverse vector fields along horizontal trajectories, respectively. This result marks a key advancement in manifold-based score modeling. It enables direct computation of the distribution in (3) using the lifted Euclidean score, avoiding the intractability of explicit transition distributions ρ_t^{H} typically estimated via spectral decomposition. Most importantly, the path discrepancy in (8) simplifies to a Euclidean score-matching loss, allowing the use of *standard Euclidean score networks s_θ without structural modification*, while maintaining rigorous geometric consistency.

Horizontal Bridge Process. Extending the previously developed horizontal diffusion framework and associated score-matching methodology, we now provide a rigorous construction of the horizontal diffusion model by explicitly specifying both the potential function and the parameterization of the lifted dynamics. We construct the Euclidean potential function ϕ in (2), which characterizes latent Euclidean Ornstein-Uhlenbeck (OU) processes as follows:

$$\phi(t,\mathbf{E}) = -\gamma_t\sigma_t^{-2}\|\mathbf{E}\|^2 + 2\lambda_t\sigma_t^{-2}\langle v_t,\mathbf{E}\rangle, \quad d\mathbf{E}_t = \big[\lambda_t e_t^{-1}v_t - \gamma_t\mathbf{E}_t\big]dt + \sigma_t \circ dW_t, \tag{9}$$

where the coefficients $(\gamma_t,\lambda_t,\sigma_t) > 0$ and the time-dependent vector $v_t \in \mathbb{R}^d$ are chosen to satisfy model requirements. This OU process admits a Gaussian transition density for the forward process, i.e., $\mathbf{E}_t \sim \mathcal{N}(0,\Sigma_t)$. Next, we lift this construction by harnessing the result in Proposition 2.2 to define the *horizontal Ornstein-Uhlenbeck bridge process* on the frame bundle:

$$d\mathbf{U}_t = \sum_a \left(2\gamma_t\varepsilon^a + \sigma_t^2 v_t^a\right) H_a(\mathbf{U}_t)dt + \sum_a \sigma_t H_a(\mathbf{U}_t)\circ dW_t^a, \quad \Sigma_t = \sigma_t^\top\sigma_t, \tag{10}$$

where $H_a(\mathbf{U}_t) \coloneqq H_{e_t}(\mathbf{U}_t)[\varepsilon^a]$ denote fundamental horizontal vector fields. We specifically design the coefficients such that $\gamma_t, \lambda_t \to 0$ with $\gamma_t \ll \lambda_t$. In this configuration, the drift components vanish, and the horizontal diffusion asymptotically reduces to pure horizontal Brownian motion. Under this design choice, the stationary distribution asymptotically converges to the uniform measure with respect to the Sasaki volume:

$$d\rho_\infty^{\mathrm{H}}(x,e) = d\mathrm{Vol}_{\mathrm{Sas}}(x,e) = d\mathrm{Vol}_g(x) \otimes d\mathrm{Haar}_{O(d)}(e).$$

The samples from stationary distribution can be efficiently generated by independently drawing base points from $d\mathrm{Vol}_g(x)$ and frames from $d\mathrm{Haar}_{O(d)}(e)$, as summarized in Alg. 1. To ensure that the reverse process drives samples toward the data distribution, we introduce the bridge vectors $v_t = -\Delta_x e_t^{-1} + \frac{1}{2}\Gamma(x_t)[\Delta_x, \Delta_x]e_t^{-1}$, which act as time-dependent terms directing the dynamics back to the initial state x_0. These vectors satisfy the zero-mean property $\mathbb{E}_{x_\bullet \sim \rho_\bullet^{\mathrm{H}}}[v_\bullet] = 0$, as established in Lemma 8.3. As a result, the Euclidean score simplifies to $s(t, \mathbf{E}) = -\Sigma_t \mathbf{E}$ for some time-variant Σ_t, and the corresponding reverse horizontal dynamics are given by

$$d\mathbf{U}_s = -\sum_a \left(2\gamma_s \varepsilon^a + \sigma_s^2 v_s^a\right) H_a(\mathbf{U}_s) ds + s^{\mathrm{H}}(s, \mathbf{U}_s) ds + \sum_a \sigma_s H_a(\mathbf{U}_s) \circ dW_s^a, \tag{11}$$

where the horizontal score function takes the form $s^{\mathrm{H}}(s, \mathbf{U}_s) \coloneqq -\sum_{a,b}[\Sigma_s]_{ab}\varepsilon_s^b H_a(\mathbf{U}_s)$. In practice, the bridge vector v_t is computed using the explicit Christoffel symbols Γ as follows:

$$\left[\Gamma(x_t)[\Delta_x, \Delta_x]\right]^i = \sum_{j,k} \Gamma_{jk}^i(x_t)(\Delta_x)^j(\Delta_x)^k. \tag{12}$$

where $\Delta_x \coloneqq x_t - x_0$ denotes the discrepancy between coordinates. In sampling the reverse horizontal dynamics, we proceeds as follows. At each reverse step, we compute the bridge vector b_s using the Christoffel symbol to evaluate score functions s^{H}, and update the reverse SDE. Proposition 9.2 shows that the reverse dynamics in (11) reconstruct the original data distribution $\rho_0^{\mathcal{M}}$. Please refer to Alg. 5 for a detailed algorithm.

Training Objective. To instantiate the proposed horizontal diffusion framework, we parameterize the learnable score field via neural networks lifted to the frame bundle. Specifically, we define $\mathrm{NN}^{\mathrm{H}}(t, \mathbf{U}; \theta) = \sum_a \mathrm{NN}^a(t, \mathbf{E}; \theta)\, H_a$, where $\mathrm{NN}(t, \mathbf{E})$ is a Euclidean neural network. We decompose the output as $\mathrm{NN}(t, \mathbf{E}) = s^\theta(t, \mathbf{E}) + v(t, \theta)$, where $v(t, \theta)$ is an auxiliary drift aligned with the analytic bridge vector. Overall, for the horizontal KL divergence from Proposition 2.1, the training objective is given as follows:

$$\mathcal{J}(\theta) = \mathbf{KL}_{\mathrm{H}}(\theta) + \mathbb{E}_{x_0, x_t} \left\| v(t,\theta) - v_t(x_t, x_0) \right\|^2, \tag{13}$$

Gauge Equivariance $\mathcal{G}$. In the geometry of frame bundles, a choice of frame is analogous to a choice of an observer's perspective. A fundamental requirement for a physically consistent generative model is that the generated distribution should not depend on this arbitrary choice. This property is formally termed *gauge equivariance*. While Euclidean diffusion models often leverage equivariance for robustness to symmetries Köhler et al. [2020], Hoogeboom et al. [2022], extending equivariant behavior to geometric setup is non-trivial. As a result, many Riemannian diffusion methods rely on fixed charts or coordinate-dependent approximations and can exhibit *coordinate sensitivity*, where changing the reference frame may degrade generalization on complex geometries. In contrast, HDM is gauge-equivariant by construction:

Proposition 2.3 (Gauge Equivariance)**.** *The proposed horizontal score function is gauge-equivariant, i.e.,* $s^{\mathrm{H}}(t, U \cdot h) = (\mathcal{R}_h)_* s^{\mathrm{H}}(t, U)$ *for all* $h \in O(d)$.

This intrinsic gauge equivariance enables HDM to model the true geometry of complex manifolds without coordinate bias, offering a distinct advantage over existing methods hindered by dependencies on global geometric operations.

The Price of Horizontal Lift. Despite its effectiveness in facilitating generative modeling on complex manifolds, our methodology is subject to inherent geometric distortions. The subsequent theorem rigorously characterizes the *worst-case horizontal KL divergence* through a uniform generalization bound and elucidates how both the expressivity of the neural network and the geometry fundamentally constrain the statistical performance of our approach.

Theorem 2.4 (informal)**.** *Let Θ be the class of L-layer, width-W ReLU score networks with spectral-norm control. Assume $\mathcal{M}$ has sectional curvature bounded by $\kappa_{\max}$ and diameter $D_{\mathcal{M}}$. For any $\delta \in (0,1)$, with probability at least $1-\delta$,*

$$\sup_{\theta\in\Theta} \mathbf{KL}_{\mathrm{H}}(\theta) \leq \frac{6C_{\mathcal{M}}(\kappa_{\max}, D_{\mathcal{M}})\, C_{\mathrm{NN}}(L,W)}{n^{1/4}} + \sqrt{\frac{C^2_{\mathrm{SubG}}(\sigma_{t_m}, \Sigma)\log(1/\delta)}{4nM}}. \tag{14}$$

Here, $C_{\mathcal{M}}$[1] is proportional to the sectional curvature $\kappa_{\max}$, indicating that the generalization error bound increases as the manifold becomes non-flat. To achieve accurate generalization, a sufficiently large n is required to guarantee statistical convergence of the KL bound. This highlights the increased sample complexity and learning difficulty associated with nontrivial geometric structure.

3 Related Work

A broad range of methods has emerged to address the challenge of generative modeling on curved manifolds, notably through the use of Riemannian flows, score-based diffusion models, and manifold-aware normalizing flows. Brehmer and Cranmer [2020] developed flows designed for simultaneous manifold learning and density estimation, significantly advancing manifold-aware generative modeling. Further contributions Mathieu and Nickel [2020] introduced Riemannian continuous normalizing flows. De Bortoli et al. [2022] expanded this direction with Riemannian score-based generative modeling, combining diffusion models and Riemannian geometry. Chen and Lipman [2024] generalized flow-matching on Riemannian manifolds. Huang et al. [2022] proposed Riemannian diffusion models tailored for geometric data, and Thornton et al. [2022] explored Riemannian diffusion Schrödinger bridges. Jo and Hwang [2024] developed generative models via mixtures of Riemannian bridge processes, introducing a flexible approach to manifold-constrained generative modeling. Recently, Lou et al. [2023] expanded these concepts to Riemannian manifolds, demonstrating how diffusion models can facilitate generative modeling on manifolds by leveraging Riemannian score matching. Courts and Kvinge [2022] presented generative models on bundle networks leveraging fiber bundles.

4 Experiments

In this section, we empirically evaluate our framework on a diverse collection of manifolds, encompassing both analytically defined surfaces and real-world scientific domains. Specifically, our experiments span (i) a family of model manifolds and (ii) complex real-world scientific datasets. In all experiments, the quality of the generated samples is evaluated using the (maximal) sliced 2-Wasserstein distance ($i.e., \mathcal{SW}_2, \mathcal{MSW}_2$), which measures the discrepancy between the empirical and generated distributions. Section 10 contains the implementation details on the experimental setup and *additional experiments*, *ablation study* in Appendix 10.

Synthetic Dataset: Gaussian Mixture Model. As our first experiment, we construct synthetic datasets on versatile model manifolds including *Projective*, *Sphere*, *Catenoid*, *Enner* and *Dupin*. For the more intricate cases, the embeddings admit compact forms such as Catenoid $(u,v) \mapsto (\cosh v \cos u, \cosh v \sin u, v)$, Enner $(u,v) \mapsto (\cos u \cos v, \cos u \sin v, \sin u)$, and Dupin $(u,v) \mapsto \left(\frac{\cos u}{1+\sin v}, \frac{\sin u}{1+\sin v}, \frac{\cos v}{1+\sin v}\right)$. Notably, with the exception of the sphere or flat torus, explicit analytic forms for the probability density and geodesic equations are generally infeasible for these parametric surfaces. For each surface, samples are drawn from a Riemannian GMM $p_{\mathcal{M}}(x) = \sum_{k=1}^{K} \mathfrak{a}_k \mathcal{N}_{\mathcal{M}}(x; \mu_k, \Sigma_k)$ where $\sum_k \mathfrak{a}_k = 1$ and $\mathcal{N}_{\mathcal{M}}$ denotes the Gaussian distribution defined on the manifold with Gaussian parameters (μ_k, Σ_k).

Fig. 2 and Table 2 summarize both the qualitative and quantitative generative modeling results on these benchmark surfaces. $\bar{\rho}^{\mathcal{M}}$ is the oracle reference, which measures the sliced Wasserstein distance between the ground-truth and reconstructed results using the theoretically exact reverse dynamics. showing degraded performances. By contrast, our framework operates on the frame bundle, achieving consistently superior results. To support scalability, we additionally report high-dimensional experiments in Appendix 10 (Fig. 5), where HDM maintains stable optimization and sample quality as intrinsic dimension increases, suggesting the frame-bundle computations remain tractable in practice.

Scientific Dataset: Quantum Density. To validate our model on experimentally realistic quantum systems, we utilize the QDataSet Perrier et al. [2022], specifically focusing on the two-qubit control

[1]Please refer to (35) in the appendix for the explicit expressions of all constants ($i.e., C_{\mathcal{M}}, C_{\mathrm{NN}}, C_{\mathrm{SubG}}$).

Table 2: Comparison of Density Estimation on Model Manifolds with sliced 2-Wasserstein distance. Lower values indicate better performance. Prior results on Catenoid, Enner, and Dupin marked with a dagger symbol† are highlighted in light red. The best and second-best results in each column are highlighted in **bold** and blue, respectively.

Model	Density	Projective ($\mathbb{CP}^d$)	Sphere ($\mathbb{S}^d$)	Torus ($\mathbb{T}^d$)	Catenoid†	Enner†	Dupin†
Oracle	$\bar{\rho}^{\mathcal{M}}$	0.032 / 0.136	0.045 / 0.158	0.097 / 0.345	**0.055** / **0.385**	**0.101** / **0.323**	0.112 / 0.710
Riem SGM	$\rho^{\mathcal{M}}$	0.101 / 0.355	0.159 / 0.569	0.213 / 0.711	0.421 / 1.552	0.534 / 1.764	0.811 / 2.022
Riem CFM	$\rho^{\mathcal{M}}$	**0.083** / **0.303**	0.141 / 0.488	0.205 / 0.667	0.379 / 1.426	**0.497** / **1.621**	0.768 / 1.884
Bundle NF	$\rho^{P(\mathcal{M})}$	–	–	**0.196** / **0.635**	–	–	**0.746** / **1.842**
Scaling NF	$\rho^{\mathcal{M}}$	0.085 / 0.244	**0.125** / **0.525**	0.211 / 0.606	**0.317** / **1.356**	0.530 / 1.466	0.803 / 1.762
Pull NF	$\rho^{\mathcal{M}}$	0.138 / 0.431	0.192 / 0.714	0.284 / 0.918	0.531 / 2.051	0.677 / 2.264	1.082 / 2.699
HDM (Ours)	ρ^{H}	**0.075** / **0.235**	**0.079** / **0.268**	**0.102** / **0.351**	**0.161** / **0.721**	**0.211** / **0.654**	**0.327** / **0.959**

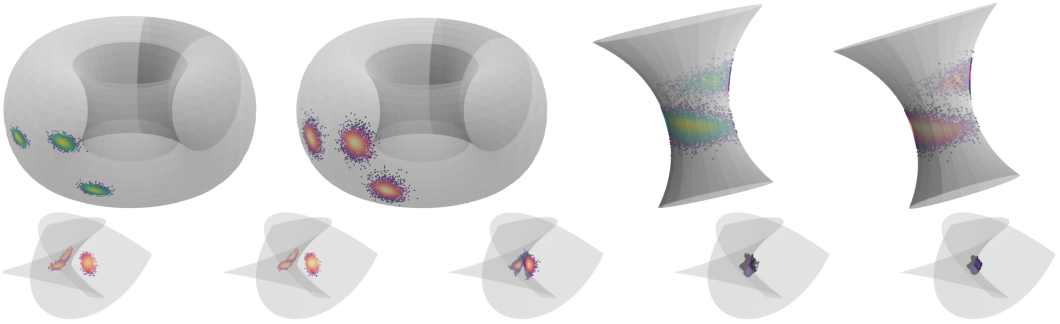


Figure 2: Visualization of Density Estimation on Model Manifolds. *Top:* Density estimation on 2-dimensional surfaces. For each surface, the distribution of the ground truth (left) and generated by ours (right) are shown. Both surfaces are equipped with a Gaussian mixture, and generated samples closely match the true multi-modal density. *Bottom:* The reverse process on the Enneper sampled from proposed method, evolves from right to left showing the density transformation as the system proceeds backward in time.

simulations. Our goal is to model distributions on the *pure-state manifold* $\mathbb{CP}^3$. Since the raw QDataSet provides 52-dimensional Pauli expectation features rather than state vectors directly, we employ a state reconstruction to recover the underlying geometry. First, we estimate the density matrix $\hat{\rho}$ via linear-inversion tomography, solving an overdetermined least-squares problem based on the observed expectations $E_k = \mathrm{Tr}(\rho V_k)$ of an informationally complete operator set. Subsequently, to ensure the data lies strictly on the target manifold, we enforce purity by projecting $\hat{\rho}$ onto its closest rank-one approximation on the direction of principal eigenvector, $\hat{\rho}_{\mathrm{pure}} = |\hat{\psi}\rangle\langle\hat{\psi}|$, thus the dataset consists of these rays $[\hat{\psi}] \in \mathbb{CP}^3$.

Scientific Dataset: Celestial Body. A widely used technique for representing the geometry of closed surfaces is the expansion of the radial function in spherical harmonics. In this approach, parametrized surfaces are given as $r(\theta, \varphi) = r_0(\theta, \varphi) + \sum_{l,m} c_{l,m} Y_{l,m}(\theta, \varphi)$, where (θ, φ) are spherical angular coordinates, $Y_{l,m}$ denote spherical harmonics of degree l and order m, and $c_{l,m}$ are scalar coefficients, naturally inducing a finite-dimensional spherical harmonic coefficient manifold denoted as $\mathcal{M}^{\mathrm{SH}}$. To study density modeling on real observational celestial surface data, we consider datasets provided through NASA planetary science archives, which distribute surface topography as standard spherical harmonic data records and provide surface geometry representations widely used in scientific analysis and downstream modeling tasks. Based on this structure, the experiment considers modeling density on the surfaces of various celestial bodies, exemplified by the *Lunar Spherical Harmonic Dataset (LRO/LOLA SHADR)* Smith et al. [2010], Neumann [2009] and the *Asteroid Spherical Harmonic Dataset (NEAR/Eros)* Miller et al. [2002], Neumann and Smith [2001].

Scientific Dataset: 3D Human Anatomy. We consider modeling density distributions on human anatomical surfaces (*i.e.*, *knee*) that pose nontrivial challenges due to their complex geometry and anatomical variability, using anatomical surface datasets acquired from clinical imaging and provided in spherical harmonic parameterized form through the SPHARM-PDM benchmark suite Styner et al. [2006], Shen et al. [2009]. These datasets consist of three-dimensional anatomical surface geometries represented in the same spherical harmonic coefficient space $\mathcal{M}^{\mathrm{SH}}$ as the celestial body datasets, and have been widely used as standard benchmarks in biomedical shape analysis.

Scientific Dataset: Peptide Conformations. To evaluate HDM on high-dimensional manifolds with non-trivial topology, we consider the dynamics of peptides, specifically *Methionine-Enkephalin*. The geometric structure of a protein peptide backbone is fundamentally determined by its torsion

Table 3: Density Estimation on Scientific Dataset. Each entry reports sliced 2-Wasserstein distance (left) and maximal sliced 2-Wasserstein distance (right). Lower values indicate better performance, with the best and second-best results in **bold** and blue, respectively.

	Celestial Body			Human Anatomy	Molecular Dynamics	
Model	**Eros** ($\mathcal{M}^{\text{SH}}$)	**Moon (IM)** ($\mathcal{M}^{\text{SH}}$)	**Moon (NC)** ($\mathcal{M}^{\text{SH}}$)	**Knee** ($\mathcal{M}^{\text{SH}}$)	**Quantum** ($\mathbb{CP}^3$)	**Peptide** ($\mathbb{R}_+ \times \mathbb{T}^d$)
Riem SGM	0.691 / 1.104	**1.110** / **1.754**	1.201 / 1.798	0.866 / 1.389	0.355 / 1.273	0.684 / 2.168
Riem CFM	0.707 / 1.198	1.200 / 1.896	**1.116** / **1.675**	0.920 / 1.472	**0.276** / **0.968**	**0.517** / **1.603**
Scaling NF	**0.672** / **1.078**	1.258 / 2.015	1.285 / 2.197	**0.825** / **1.405**	0.319 / 1.112	0.796 / 2.401
Pull NF	0.689 / 1.170	-	-	-	0.402 / 1.431	0.612 / 1.984
HDM (Ours)	**0.312** / **0.470**	**0.512** / **0.754**	**0.489** / **0.726**	**0.414** / **0.606**	**0.118** / **0.423**	**0.237** / **0.721**

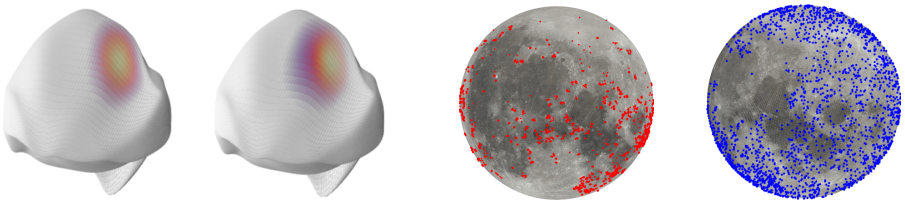


Figure 3: Density Estimation on Human and Celestial Body. *(Left)* True (left) and generated (right) tumor densities on human anatomy (*i.e.*, Knee). *(Right)* Density estimation results for IM and NC lunar regions, demonstrating label-conditional modeling performance.

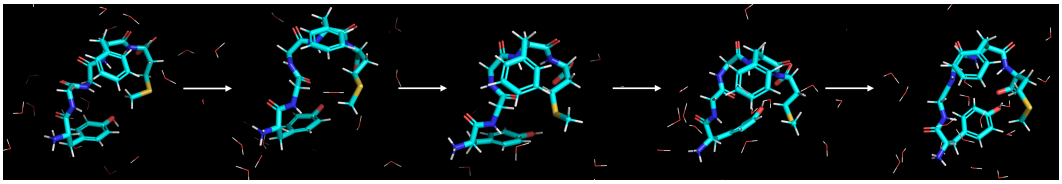


Figure 4: Peptide Conformations. We visualize generated conformational trajectories of Methionine-Enkephalin, where the resulting transitions between backbone configurations are smooth and physically consistent with the toroidal structure.

angles (ϕ, ψ). Consequently, a peptide with d torsion angles can be naturally modeled as a point on an N-dimensional flat torus, $\mathbb{T}^d := \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$. Unlike Euclidean approaches that struggle with the periodic boundary conditions of angular data, our framework respects the intrinsic topology of the torus. We process the molecular dynamics (MD) simulation data into *spatio-temporal trajectories* resulting model space as $\mathbb{R}_+ \times \mathbb{T}^d$, where each sample consists of a sequence of $T = 10$ consecutive time steps. This formulation allows us to capture not only the static equilibrium Boltzmann distribution but also the temporal dynamics on the toroidal manifold, effectively modeling the transition paths between metastable states.

As summarized in Table 3 and Figure 3, our experimental results demonstrate that the proposed method consistently achieves high-fidelity density estimation on a diverse set of surfaces, ranging from celestial bodies (*i.e.*, lunar and Eros craters) to human anatomical surfaces (*i.e.*, knee tumors). While existing methods typically rely on embedding data into the ambient product space $\mathbb{S}^2 \times \mathbb{R}$ and subsequently projecting it back onto the target manifold, induced projection distortions limit reconstruction quality. In contrast, our framework operates on the frame bundle, preserving geometric integrity without projection-induced distortions. The consistently lower sliced Wasserstein distances achieved by our approach underscore its superior capability to model data distributions across geometry. Similarly, our framework learns distributions over torsion-angle representations for peptide conformations respecting their intrinsic periodic structure, thus mitigating artifacts associated with Euclidean embeddings and leading to consistently lower sliced Wasserstein distances compared to existing methods, with Figure 4 providing qualitative visualizations of representative generated trajectories. Please refer to **Ablation Study** in Sec 10.6 for scalability and sensitivity analysis.

5 Conclusion

In this work, we introduced Horizontal Diffusion Models, a novel framework for generative modeling on Riemannian manifolds leveraging orthonormal frame bundles. By horizontally lifting standard Euclidean diffusion processes to the frame bundle, we avoid global Riemannian primitives, assuming access to local differential geometric quantities such as the metric and Levi–Civita connection. Experiments across synthetic and real-world manifold datasets demonstrate superior performance, validating our model's scalability and geometric fidelity. This establishes a robust, geometry-aware pathway for future generative modeling research on complex, curved manifolds.

Acknowledgments and Disclosure of Funding

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-ICT Creative Consilience Program grant funded by the Korea government (MSIT) (IITP-2026-RS-2020-II201819). This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2026-25495269). This research was supported by the Basic Science Research Program through the NRF funded by the Korea government (MSIT) (RS-2026-25522655).

References

Peter L Bartlett, Dylan J Foster, and Matus J Telgarsky. Spectrally-normalized margin bounds for neural networks. *Advances in neural information processing systems*, 30, 2017.

Marco Bertolini, Tuan Le, and Djork-Arné Clevert. Lie group-induced dynamics in score-based generative modeling, 2025. URL `https://openreview.net/forum?id=IARgA4HqjJ`.

Johann Brehmer and Kyle Cranmer. Flows for simultaneous manifold learning and density estimation. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 442–453. Curran Associates, Inc., 2020.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *The Twelfth International Conference on Learning Representations*, 2024. URL `https://openreview.net/forum?id=g7ohDlTITL`.

Nico Courts and Henry Kvinge. Bundle networks: Fiber bundles, local trivializations, and a generative approach to exploring many-to-one maps. In *International Conference on Learning Representations*, 2022.

Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 2406–2422. Curran Associates, Inc., 2022.

Willem Diepeveen, Georgios Batzolis, Zakhar Shumaylov, and Carola-Bibiane Schönlieb. Score-based pullback riemannian geometry: Extracting the data manifold geometry using anisotropic flows. In *Forty-second International Conference on Machine Learning*, 2025. URL `https://openreview.net/forum?id=AJN5btaqNk`.

James Eells and JH Sampson. Harmonic mappings of riemannian manifolds. *American Journal of Mathematics*, 86(1):109–160, 1964.

David Elworthy. *Geometric aspects of diffusions on manifolds*. Springer Berlin Heidelberg, 1988.

Kenneth David Elworthy. *Stochastic Differential Equations on Manifolds*, volume 70 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 1982.

Ulrich G Haussmann and Etienne Pardoux. Time reversal of diffusions. *The Annals of Probability*, pages 1188–1205, 1986.

Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pages 8867–8887. PMLR, 2022.

Elton P Hsu. *Stochastic Analysis on Manifolds*. Number 38. American Mathematical Society, 2002.

Chin-Wei Huang, Milad Aghajohari, Joey Bose, Prakash Panangaden, and Aaron Courville. Riemannian diffusion models. In *Advances in Neural Information Processing Systems*, 2022.

Jaehyeong Jo and Sung Ju Hwang. Generative modeling on manifolds through mixture of riemannian diffusion processes. *arXiv preprint arXiv:2310.07216*, 2024.

Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of Differential Geometry, Vol. I*. Interscience Publishers, 1963.

Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of differential geometry, volume 2*, volume 2. John Wiley & Sons, 1996.

Jonas Köhler, Leon Klein, and Frank Noé. Equivariant flows: exact likelihood generative learning for symmetric densities. In *International conference on machine learning*, pages 5361–5370. PMLR, 2020.

Aaron Lou, Minkai Xu, Adam Farris, and Stefano Ermon. Scaling riemannian diffusion models. *Advances in Neural Information Processing Systems*, 36:80291–80305, 2023.

Paul Malliavin. Stochastic calculus of variation and hypoelliptic operators. *Proceedings of the International Symposium on Stochastic Differential Equations, Kyoto*, pages 195–263, 1978.

Emile Mathieu and Maximilian Nickel. Riemannian continuous normalizing flows. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 2503–2515. Curran Associates, Inc., 2020.

J.K. Miller, A.S. Konopliv, P.G. Antreasian, J.J. Bordi, S. Chesley, C.E. Helfrich, W.M. Owen, T.C. Wang, B.G. Williams, D.K. Yeomans, and D.J. Scheeres. Determination of shape, gravity, and rotational state of asteroid 433 eros. *Icarus*, 155(1):3–17, 2002. doi: 10.1006/icar.2001.6753.

G. Neumann and D. E. Smith. NEAR NLR LEVEL 2 DATA PRODUCTS V1.0. 2001. doi: 10.17189/1520092. NEAR-A-NLR-5-CDR-EROS/ORBIT-V1.0.

G. A. Neumann. Lunar Orbiter Laser Altimeter Spherical Harmonic Model. 2009. doi: 10.17189/1520654. NASA Planetary Data System, LRO-L-LOLA-5-SHADR-V1.0.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Elija Perrier, Akram Youssry, and Chris Ferrie. Qdataset, quantum datasets for machine learning. *Scientific data*, 9(1):582, 2022.

Li Shen, Hany Farid, and Mark A. McPeek. Modeling three-dimensional morphological structures using spherical harmonics. *Evolution*, 63(4):1003–1016, 2009. doi: 10.1111/j.1558-5646.2008.00557.x.

David E Smith, Maria T Zuber, Glenn B Jackson, John F Cavanaugh, Gregory A Neumann, Haris Riris, Xiaoli Sun, Ronald S Zellar, Craig Coltharp, Joseph Connelly, et al. The lunar orbiter laser altimeter investigation on the lunar reconnaissance orbiter mission. *Space Science Reviews*, 150 (1-4):209–241, 2010.

Martin Styner, Ipek Oguz, Shun Xu, Christian Brechbühler, Dimitrios Pantazis, James J. Levitt, Martha E. Shenton, and Guido Gerig. Framework for the statistical shape analysis of brain structures using spharm-pdm. *Insight Journal*, (1071):242, 2006.

James Thornton, Michael Hutchinson, Emile Mathieu, Valentin De Bortoli, Yee Whye Teh, and Arnaud Doucet. Riemannian diffusion schrödinger bridge. *arXiv preprint arXiv:2207.03024*, 2022.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Appendix

6 Notation Table

For the reader's convenience, we provide a summary of the primary symbols and notation employed in this work. The table below serves as a quick reference for all key terms and variables.

Table 4: Summary of notation used throughout the paper.

Symbol	Meaning
$\mathcal{M}, g$	d-dimensional Riemannian manifold and its metric
$\pi : \mathcal{O}(\mathcal{M}) \to \mathcal{M}$	Orthonormal frame bundle (principal $\mathrm{O}(d)$-bundle)
(x^i, e_a^i)	Local coordinates on $\mathcal{O}(\mathcal{M})$; e_a is the a-th frame vector at x
ω	Levi–Civita connection 1-form on $\mathcal{O}(\mathcal{M})$
$\Omega = d\omega + \frac{1}{2}[\omega, \omega]$	Curvature 2-form of ω
$H_e[w]$	Horizontal lift of $w \in \mathbb{R}^d$ at e, giving a vector in $T_U\mathcal{O}(\mathcal{M})$
$\mathbf{U}_t, \mathbf{E}_t$	Horizontal diffusion on $\mathcal{O}(\mathcal{M})$ and latent Euclidean process in $\mathbb{R}^d$
$\nu_t = \mathcal{L}(\mathbf{U}_t, \mathbf{E}_t)$	Joint law of $(\mathbf{U}_t, \mathbf{E}_t)$ on $\mathcal{O}(\mathcal{M}) \times \mathbb{R}^d$
ρ_t	Density of $\mathbf{E}_t$ on $\mathbb{R}^d$
$\rho_t^{\mathrm{H}}(U)$	Density of $\mathbf{U}_t$ on $\mathcal{O}(\mathcal{M})$ (Sasaki volume)
$\rho_t^M(x) = \pi_{\#}\rho_t^{\mathrm{H}}$	Induced marginal on M
$s_\theta^{\mathrm{H}}(t, U)$	Learnable horizontal score field on $\mathcal{O}(\mathcal{M})$
s_*, s	True Euclidean score functions$\nabla \log \rho_t$
$s_*^{\mathcal{M}}, s^{\mathcal{M}}$	True manifold score functions$\nabla^{\mathcal{M}} \log \rho_t^{\mathcal{M}}$
$s_*^{\mathrm{H}}, s^{\mathrm{H}}$	True horizontal score functions$\nabla^{\mathrm{H}} \log \rho_t^{\mathrm{H}}$
$\nu_{[0,T]}, \tilde{\nu}_{[0,T]}$	Path laws of forward / reverse horizontal SDEs on $[0, T]$
$\mathrm{H}_U, \mathrm{Ver}_U$	Horizontal and vertical subspaces of $T_U\mathcal{O}(\mathcal{M})$
$L_t^{\mathbf{E}}$	Generator of horizontal diffusion conditioned on latent path $\mathbf{E}_{[0,t]}$
$p_t^{\mathbf{E}}(U \mid U_0)$	Conditional transition density on $\mathcal{O}(\mathcal{M})$ given $\mathbf{E}_{[0,t]}$
∇^{H}	Horizontal gradient; $\nabla^{\mathrm{H}} f = P_{\mathrm{H}}(\nabla f)$
P_{H}	Projection $T\mathcal{O}(\mathcal{M}) \to H$ onto horizontal subbundle
KL^{H}	Geometry-aware horizontal Kullback–Leibler divergence
$W_t,\ \tilde{W}_s \coloneqq W_s$	Forward and reverse $\mathbb{R}^d$ Brownian motions
$\sigma_t^{(a)} H_a$	Horizontal noise field in lifted SDE
Γ_{ij}^k, R_{irs}^k	Christoffel symbols and Riemann curvature components of (M, g)
$\|\cdot\|_g, \langle \cdot, \cdot \rangle_g$	Norm / inner product induced by g (Sasaki metric on H)
δ_U	Dirac measure at $U \in \mathcal{O}(\mathcal{M})$
$\mathrm{Law}_{\mathbb{P}}(X)$	Distribution of X under probability measure $\mathbb{P}$
$T_U\mathcal{O}(\mathcal{M})$	Tangent space of $\mathcal{O}(\mathcal{M})$ at U

7 Technical Comparison with Existing Works

Numerical Computation of $\rho_t^{\mathcal{M}}$. A principal technique in classical score-based generative modeling on manifolds is the spectral decomposition of the Laplace–Beltrami operator. Given a compact Riemannian manifold $\mathcal{M}$, the Laplace–Beltrami operator $\Delta_{\mathcal{M}}$ admits a discrete set of eigenpairs $\{(\lambda_k, \rho_k)\}_{k=0}^{\infty}$ such that

$$\Delta_{\mathcal{M}}\rho_k = \lambda_k \rho_k, \qquad \rho^{\mathcal{M}}(x) = \sum_{k=0}^{\infty} c_k \rho_k(x), \qquad k = 0, 1, 2, \ldots,$$

with the eigenfunctions $\{\rho_k\}$ forming an orthonormal basis of $L^2(\mathcal{M})$. Existing methods approximate probability densities or construct kernels on the manifold by expanding density $\rho^{\mathcal{M}}$ in spectral basis.

This facilitates analytic calculation of heat kernels and allows the construction of diffusion models via explicit spectral representations. However, the effectiveness of this approach is fundamentally limited to manifolds where the spectral data can be computed in closed form. For general, non-symmetric, or high-curvature manifolds, such explicit decomposition is unavailable or infeasible. Therefore, these methods struggle to scale to arbitrary geometric domains. In contrast, our framework obviates the need for direct spectral decomposition by employing horizontal diffusion processes on the frame bundle. In particular, as established in Proposition 2.2, the horizontal lifting of the Euclidean score function enables an indirect yet generalized mechanism for approximating the data score function on the manifold, irrespective of the analytic tractability or the spectral properties of the underlying space.

Neural Network Architecture. Conventional manifold-based diffusion models often require geometry-specific neural network architectures, with model design and parameterization tailored individually to the underlying manifold for spheres, tori, or other curved spaces. In contrast, our framework employs a single, unified neural network architecture across all experiments, irrespective of the specific manifold geometry. This property is made possible by the use of the horizontal lift, which intrinsically adapts the representation of the score function to the manifold structure at hand.

Geometric Operations. A central feature of our approach is the definition of data dynamics through frame rotations on the orthonormal frame bundle, thereby intrinsically modeling geodesic flows of the underlying manifold. This construction is rooted in Elworthy's stochastic development Elworthy [1988], wherein the evolution of a point on the manifold is governed jointly with its moving frame.Mathematically, the evolution of the frame e_t along a trajectory x_t is described by the stochastic differential equation:

$$de_t = -\sum_{i,j,k} \Gamma_{ij}^k(x_t) e_t^j \circ dx_t^i E_k,$$

where $\Gamma_{ij}^k(x_t)$ are the Christoffel symbols of the Levi-Civita connection, and E_k denotes the standard basis in $\mathbb{R}^d$. This formulation ensures that the frame is parallel transported along the curve x_t, faithfully encoding the connection-induced geometry.

Through this coupled evolution of position and frame, our method is able to realize geodesic and stochastic flows on the manifold in an intrinsic, coordinate-free manner. As a consequence, we are able to model data dynamics without requiring access to global distance functions, explicit geodesic computations. Instead, all essential geometric operations are performed intrinsically via the connection and horizontal lift, emphasizing the mathematical and computational scalability of our framework for generative modeling on manifolds. In contrast, the heavy reliance of most existing methods on global geometric operations critically limits their practical scalability and application to broader classes of manifolds.

8 Lemmas

Lemma 8.1. *Let H_e be the horizontal lift operator defined for $w \in \mathbb{R}^d$, $e \in \mathcal{O}_x(\mathcal{M})$. Then, for the differentiable test function $f \in C^\infty(\mathcal{O}(\mathcal{M}))$, this operator can be written in local coordinate as follows:*

$$H_e(U)[w]f = w^a e_a^i \frac{\partial f}{\partial x^i} - w^a \Gamma_{ij}^k(x)\, e_a^i e_b^j \frac{\partial f}{\partial e_b^k}, \tag{15}$$

where $\mathbf{U} = (x, e) \in \mathcal{O}(\mathcal{M})$, and $\Gamma_{ij}^k(x)$ are the Christoffel symbols of the Levi-Civita connection at $x \in \mathcal{M}$. Then, the map $H_e : T_w\mathbb{R}^d \to T_U\mathcal{O}(\mathcal{M})$ is linear for all $w \in \mathbb{R}^d$.

Proof. Let $\tau \mapsto U(\tau) = (x(\tau), e(\tau))$ be the horizontal curve with $U(0) = U$ and $\pi_*\dot{U}(0) = e(w)$. Horizontality means that every frame vector $e_b(\tau)$ is parallel along $x(\tau)$, *i.e.*, $\dot{e}_b^k + \Gamma_{ij}^k(x)\dot{x}^i e_b^j = 0$, and the projection condition reads $\dot{x}^i(0) = e_a^i w^a$. The chain rule then gives the coordinate form (15):

$$\begin{aligned} H_e(U)[w]f &= \frac{d}{d\tau} f\big(x(\tau), e(\tau)\big)\Big|_{\tau=0} = \dot{x}^i(0)\, \frac{\partial f}{\partial x^i} + \dot{e}_b^k(0)\, \frac{\partial f}{\partial e_b^k} \\ &= e_a^i w^a\, \frac{\partial f}{\partial x^i} - \Gamma_{ij}^k(x)\, \dot{x}^i(0)\, e_b^j\, \frac{\partial f}{\partial e_b^k} = w^a e_a^i\, \frac{\partial f}{\partial x^i} - w^a\, \Gamma_{ij}^k(x)\, e_a^i e_b^j\, \frac{\partial f}{\partial e_b^k}. \end{aligned}$$

For linearity, let $w_1, w_2 \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$, and $f \in C^\infty(\mathcal{O}(\mathcal{M}))$. Substituting $w = w_1 + \lambda w_2$ into (15),

$$\begin{aligned} H_e(U)[w_1 + \lambda w_2]f &= (w_1^a + \lambda w_2^a)\, e_a^i \frac{\partial f}{\partial x^i} - (w_1^a + \lambda w_2^a)\, \Gamma_{ij}^k(x)\, e_a^i e_b^j \frac{\partial f}{\partial e_b^k} \\ &= w_1^a e_a^i \frac{\partial f}{\partial x^i} + \lambda w_2^a e_a^i \frac{\partial f}{\partial x^i} - w_1^a \Gamma_{ij}^k(x)\, e_a^i e_b^j \frac{\partial f}{\partial e_b^k} - \lambda w_2^a \Gamma_{ij}^k(x)\, e_a^i e_b^j \frac{\partial f}{\partial e_b^k} \\ &= \Big(w_1^a e_a^i \frac{\partial f}{\partial x^i} - w_1^a \Gamma_{ij}^k(x)\, e_a^i e_b^j \frac{\partial f}{\partial e_b^k}\Big) + \lambda\Big(w_2^a e_a^i \frac{\partial f}{\partial x^i} - w_2^a \Gamma_{ij}^k(x)\, e_a^i e_b^j \frac{\partial f}{\partial e_b^k}\Big) \\ &= H_e(U)[w_1]f + \lambda\, H_e(U)[w_2]f. \end{aligned}$$

The identity holds for every f, and a tangent vector is determined by its action on $C^\infty(\mathcal{O}(\mathcal{M}))$; hence $H_e(U)[w_1 + \lambda w_2] = H_e(U)[w_1] + \lambda H_e(U)[w_2]$ in $T_U\mathcal{O}(\mathcal{M})$. Applying the same identity to $w = \sum_a w^a \varepsilon_a$ yields the expansion in the fundamental horizontal vector fields used throughout the paper:

$$\begin{aligned} H_e(U)[w] &= H_e(U)\Big[\sum_{a=1}^d w^a \varepsilon_a\Big] = \sum_{a=1}^d w^a\, H_e(U)[\varepsilon_a] = \sum_{a=1}^d w^a\, H_a(U), \\ H_a(U) &= e_a^i \frac{\partial}{\partial x^i} - \Gamma_{ij}^k(x)\, e_a^i e_b^j \frac{\partial}{\partial e_b^k}. \end{aligned} \tag{16}$$

□

Lemma 8.2 (Itô–Stratonovich equivalence for linear SDEs)**.** *Let $\mathbf{E}_t \in \mathbb{R}^d$ be a stochastic process satisfying the following linear Itô SDE:*

$$d\mathbf{E}_t = \alpha_t \mathbf{E}_t dt + \sigma_t^2 dW_t, \tag{17}$$

where $\alpha_t \in \mathbb{R}$ and $\sigma_t^2 \in \mathbb{R}$ are deterministic scalar-valued, time-dependent coefficients, and $W_t \in \mathbb{R}^d$ is a standard d-dimensional Brownian motion with independent components. Then, the equivalent Stratonovich SDE takes the same differential form:

$$\circ d\mathbf{E}_t = \alpha_t \mathbf{E}_t dt + \sigma_t^2 \circ dW_t. \tag{18}$$

That is, the Itô and Stratonovich formulations are identical for this system.

Proof. For an Itô SDE $d\mathbf{E}_t = b_t(\mathbf{E}_t)dt + \Sigma_t(\mathbf{E}_t)dW_t$ with $b_t : \mathbb{R}^d \to \mathbb{R}^d$ and $\Sigma_t : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, the equivalent Stratonovich SDE is

$$d\mathbf{E}_t^i = \Big[b_t^i(\mathbf{E}_t) - \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d}\Sigma_t^{(k,j)}(\mathbf{E}_t)\,\frac{\partial\Sigma_t^{(i,j)}}{\partial\mathbf{E}^k}(\mathbf{E}_t)\Big]dt + \sum_{j=1}^{d}\Sigma_t^{(i,j)}(\mathbf{E}_t)\circ dW_t^j, \qquad i = 1,\dots,d.$$

In (17), $b_t^i(\mathbf{E}) = \alpha_t\mathbf{E}^i$ and $\Sigma_t^{(i,j)}(\mathbf{E}) = \sigma_t^2\delta_{ij}$, which is independent of the state. The correction term therefore vanishes identically:

$$\frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d}\Sigma_t^{(k,j)}(\mathbf{E}_t)\,\frac{\partial\Sigma_t^{(i,j)}}{\partial\mathbf{E}^k}(\mathbf{E}_t) = \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d}\sigma_t^2\delta_{kj}\,\frac{\partial\big(\sigma_t^2\delta_{ij}\big)}{\partial\mathbf{E}^k} = \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d}\sigma_t^2\delta_{kj}\cdot 0 = 0.$$

The same conclusion follows at the level of stochastic integrals. For a continuously differentiable schedule $t \mapsto \sigma_t^2$ and any partition $\{s_k\}$ of $[0,t]$ with mesh $|\pi|$,

$$\int_0^t \sigma_s^2\circ dW_s^i - \int_0^t \sigma_s^2\,dW_s^i = \frac{1}{2}\big\langle\sigma_\bullet^2, W^i\big\rangle_t = \frac{1}{2}\lim_{|\pi|\to 0}\sum_k\big(\sigma^2_{s_{k+1}} - \sigma^2_{s_k}\big)\big(W^i_{s_{k+1}} - W^i_{s_k}\big),$$

$$\Big|\sum_k\big(\sigma^2_{s_{k+1}} - \sigma^2_{s_k}\big)\big(W^i_{s_{k+1}} - W^i_{s_k}\big)\Big| \le \max_k\big|W^i_{s_{k+1}} - W^i_{s_k}\big|\int_0^t\Big|\frac{d\sigma_s^2}{ds}\Big|ds \xrightarrow[|\pi|\to 0]{} 0,$$

where the limit uses the uniform continuity of W^i on $[0,t]$. Substituting into (17) componentwise,

$$\mathbf{E}_t^i - \mathbf{E}_0^i = \int_0^t \alpha_s\mathbf{E}_s^i\,ds + \int_0^t \sigma_s^2\,dW_s^i = \int_0^t \alpha_s\mathbf{E}_s^i\,ds + \int_0^t \sigma_s^2\circ dW_s^i, \qquad i = 1,\dots,d,$$

which is the integral form of (18). □

Lemma 8.3 (Zero mean of the bridge vector)**.** *Let $(\mathbf{U}_t, \mathbf{E}_t) = (x_t, e_t, \mathbf{E}_t)$ be a weak solution to the coupled system with deterministic initial condition $x_0 \in \mathcal{M}$. Define the bridge vector by $v_t := e_t^{-1}\log_{x_t}(x_0) \in \mathbb{R}^d$. Then, the expectation of the bridge vector with respect to the data distribution at time t vanishes, i.e., $\mathbb{E}_{x_t\sim\rho_t^{\mathcal{M}}}[v_t] = 0$.*

Proof. **Step 1 (Christoffel expansion of the bridge vector).** Fix a chart around x_0 and write $\Delta_x := x_t - x_0$. Let $\tau \mapsto y(\tau)$, $\tau \in [0,1]$, be the geodesic with $y(0) = x_t$ and $y(1) = x_0$; then $\log_{x_t}(x_0) = \dot y(0)$ and $|\dot y(0)| = O(|\Delta_x|)$. The Taylor expansion of y combined with the geodesic equation $\ddot y^k + \Gamma_{ij}^k(y)\dot y^i\dot y^j = 0$ gives

$$\begin{aligned} x_0^k = y^k(1) &= y^k(0) + \dot y^k(0) + \tfrac{1}{2}\ddot y^k(0) + O\big(|\Delta_x|^3\big) \\ &= x_t^k + \dot y^k(0) - \tfrac{1}{2}\Gamma_{ij}^k(x_t)\,\dot y^i(0)\dot y^j(0) + O\big(|\Delta_x|^3\big). \end{aligned}$$

Rearranging and substituting the first-order relation $\dot y(0) = -\Delta_x + O(|\Delta_x|^2)$ into the quadratic term,

$$\begin{aligned} \big[\log_{x_t}(x_0)\big]^k = \dot y^k(0) &= -(\Delta_x)^k + \tfrac{1}{2}\Gamma_{ij}^k(x_t)\,\dot y^i(0)\dot y^j(0) + O\big(|\Delta_x|^3\big) \\ &= -(\Delta_x)^k + \tfrac{1}{2}\Gamma_{ij}^k(x_t)\,(\Delta_x)^i(\Delta_x)^j + O\big(|\Delta_x|^3\big) \\ &= -(\Delta_x)^k + \tfrac{1}{2}\Gamma_{ij}^k(x_0)\,(\Delta_x)^i(\Delta_x)^j + O\big(|\Delta_x|^3\big), \end{aligned}$$

where the last line uses $\Gamma_{ij}^k(x_t) = \Gamma_{ij}^k(x_0) + O(|\Delta_x|)$. Hence $\log_{x_t}(x_0) = -\Xi(\Delta_x)$ with

$$\begin{aligned} \Xi^k(\Delta_x) &= (\Delta_x)^k - \tfrac{1}{2}\Gamma_{ij}^k(x_0)\,(\Delta_x)^i(\Delta_x)^j + O\big(|\Delta_x|^3\big), \\ v_t &= -e_t^{-1}\Delta_x + \tfrac{1}{2}\,e_t^{-1}\Gamma(x_t)[\Delta_x,\Delta_x] + O\big(|\Delta_x|^3\big), \end{aligned}$$

which is the computable bridge vector used in the main text. Every term of this expansion carries the left factor e_t^{-1}, and the argument below applies to each term separately.

Step 2 (structure of the transported frame). As in the construction of the forward process, the initial frame is drawn from the normalized Haar measure on the fiber over x_0: $e_0 = \hat{e}_0 h$ with a fixed frame $\hat{e}_0$ at x_0 and $h \sim \mathrm{Haar}_{O(d)}$ independent of the latent driving path $\mathbf{E}_\bullet$, whose law is isotropic, $\mathrm{Law}(h'\mathbf{E}_\bullet) = \mathrm{Law}(\mathbf{E}_\bullet)$ for all $h' \in O(d)$. Let $\Gamma^{\mathcal{M}}_{0\to t} : T_{x_0}\mathcal{M} \to T_{x_t}\mathcal{M}$ denote parallel transport along $x_{[0,t]}$. Horizontality transports each frame vector in parallel, and parallel transport is linear; hence

$$e_t = \Gamma^{\mathcal{M}}_{0\to t} \circ e_0 = \Gamma^{\mathcal{M}}_{0\to t} \circ \hat{e}_0\, h, \qquad e_t^{-1} = h^{-1}\big(\Gamma^{\mathcal{M}}_{0\to t} \circ \hat{e}_0\big)^{-1}.$$

To see that h is independent of the base path, let $\mathbf{U}'_t = (x'_t, e'_t)$ be the horizontal lift of the rotated path $h\mathbf{E}_\bullet$ from $(x_0, \hat{e}_0)$ and set $\mathbf{V}_t := \mathbf{U}'_t \cdot h$. Using $\pi \circ R_h = \pi$ and the right-invariance identity (22),

$$\begin{aligned} d\mathbf{V}_t &= dR_h\big(d\mathbf{U}'_t\big) = dR_h\Big(H_{e'_t}(\mathbf{U}'_t)\big[h \circ d\mathbf{E}_t\big]\Big) \\ &= H_{e'_t h}\big(\mathbf{U}'_t \cdot h\big)\big[\circ d\mathbf{E}_t\big] = H\big(\mathbf{V}_t\big)\big[\circ d\mathbf{E}_t\big], \qquad \mathbf{V}_0 = (x_0, \hat{e}_0 h). \end{aligned}$$

Hence $\mathbf{V}_\bullet = \mathbf{U}_\bullet$ and $x_\bullet = \pi(\mathbf{U}_\bullet) = \pi(\mathbf{U}'_\bullet)$ is the base path of the lift of $h\mathbf{E}_\bullet$ from $(x_0, \hat{e}_0)$. Conditionally on h, the path $h\mathbf{E}_\bullet$ has the law of $\mathbf{E}_\bullet$, which leaves the conditional law of $x_\bullet$ independent of h; therefore h and $x_{[0,t]}$ are independent.

Step 3 (vanishing expectation). Left invariance of the Haar measure under $h \mapsto (-I)h$ gives

$$\begin{aligned} \mathbb{E}\big[h^{-1}\big] &= \int_{O(d)} h^{-1}\, d\mathrm{Haar}_{O(d)}(h) = \int_{O(d)} \big((-I)h\big)^{-1}\, d\mathrm{Haar}_{O(d)}(h) \\ &= -\int_{O(d)} h^{-1}\, d\mathrm{Haar}_{O(d)}(h) = -\mathbb{E}\big[h^{-1}\big], \end{aligned}$$

hence $\mathbb{E}[h^{-1}] = 0$. Conditioning on the base path, which determines $\log_{x_t}(x_0)$ and $\Gamma^{\mathcal{M}}_{0\to t}$, and using Steps 1–2,

$$\begin{aligned} \mathbb{E}\big[v_t\big] &= \mathbb{E}\Big[e_t^{-1} \log_{x_t}(x_0)\Big] = \mathbb{E}\Big[\mathbb{E}\big[e_t^{-1} \,\big|\, x_{[0,t]}\big] \log_{x_t}(x_0)\Big] \\ &= \mathbb{E}\Big[\mathbb{E}\big[h^{-1} \,\big|\, x_{[0,t]}\big]\big(\Gamma^{\mathcal{M}}_{0\to t} \circ \hat{e}_0\big)^{-1} \log_{x_t}(x_0)\Big] \\ &= \mathbb{E}\Big[\mathbb{E}\big[h^{-1}\big]\big(\Gamma^{\mathcal{M}}_{0\to t} \circ \hat{e}_0\big)^{-1} \log_{x_t}(x_0)\Big] = 0. \end{aligned}$$

Since $x_t \sim \rho^{\mathcal{M}}_t$ is the marginal of the base path, this is $\mathbb{E}_{x_t \sim \rho^{\mathcal{M}}_t}[v_t] = 0$. Replacing $\log_{x_t}(x_0)$ by any of the terms $-\Delta_x$ or $\frac{1}{2}\Gamma(x_t)[\Delta_x, \Delta_x]$ of Step 1 leaves the computation unchanged; the truncated bridge vector of the main text therefore has zero mean as well. □

9 Proofs

The proof is organized into two main parts. First, Section 9.1 characterizes the mathematical properties of the horizontal diffusion model, including time-reversal, gauge equivariance, and the structure of the horizontal lift. Second, Section 9.2 presents the learning methodology for this model and establish generalization bounds that hold for arbitrary underlying geometries.

9.1 Theoretical Grounds of Horizontal Diffusion Models

Proposition 9.1 (Horizontal lift of gradients and scores)**.** *Let $\mathbf{E}_t \in \mathbb{R}^d$ be the solution to Euclidean Stratonovich SDE, and let $\mathbf{U}_t = (x_t, e_t) \in \mathcal{O}(\mathcal{M})$ be its horizontal lift defined by $d\mathbf{U}_t = H_{e_t}(\mathbf{U}_t)\big[\circ\, d\mathbf{E}_t\big]$. Then, for every $t \geq 0$,*

$$\nabla^{\mathrm{H}}\Phi(\mathbf{U}_t) = H_{e_t}(\mathbf{U}_t)\big[\nabla\phi(\mathbf{E}_t)\big], \quad \nabla^{\mathrm{H}}\log\rho_t^{\mathrm{H}}(\mathbf{U}_t) = H_{e_t}(\mathbf{U}_t)\big[\nabla\log\rho_t(\mathbf{E}_t)\big], \tag{19}$$

$$s^{\mathrm{H}}(t,\mathbf{U}_t) := \nabla^{\mathrm{H}}\log\rho_t^{\mathrm{H}}(\mathbf{U}_t) = H_{e_t}(\mathbf{U}_t)\big[s(t,\mathbf{E}_t)\big] = \sum_{a=1}^{d} s^a(t,\mathbf{E}_t)H_a(\mathbf{U}_t). \tag{20}$$

Proof. **Setting.** Let the latent path solve $d\mathbf{E}_t = -\frac{1}{2}\nabla\phi_t(\mathbf{E}_t)\,dt + \sigma_t \circ dW_t$ with a deterministic schedule $\sigma_t > 0$, and let $(x_0, e_0) \in \mathcal{O}(\mathcal{M})$ be the initial frame. The horizontal lift path $U_\bullet = \mathrm{HorLift}_{(x_0,e_0)}(E_\bullet)$ is the solution of

$$U_0 = (x_0, e_0), \qquad dU_t = H_{e_t}(U_t)\big[\circ dE_t\big] = \sum_{a=1}^{d} H_a(U_t) \circ dE_t^a,$$

where the second form follows from (16) in Lemma 8.1. The group of time-dependent gauge transformations $\mathcal{G} := C([0,T], O(d))$ acts by $(x_t, e_t)\cdot g_\bullet := (x_t, e_t g_t)$, and the horizontal lifting quotient map is

$$\Psi : C([0,T],\mathbb{R}^d) \to C([0,T],\mathcal{O}(\mathcal{M}))/\mathcal{G},$$
$$\Psi(E_\bullet) := \big[\mathrm{HorLift}_{(x_0,e_0)}(E_\bullet)\big] = \big\{\mathrm{HorLift}_{(x_0,e_0)}(E_\bullet)\cdot g_\bullet \,\big|\, g_\bullet \in \mathcal{G}\big\}.$$

A smooth potential $\phi : \mathbb{R}^d \to \mathbb{R}$ is extended to the frame bundle through Ψ, with $\pi_E(x,e) := e^{-1}(\xi)$ the Euclidean coordinate of the tangent vector $\xi \in T_x\mathcal{M}$ in the frame e, for all $t \in [0,T]$:

$$(\Phi\circ\Psi)(E_\bullet)_t = \Phi(U_t) = \phi\big(\pi_E(U_t)\big) = \phi(E_t).$$

The Sasaki metric on $T_U\mathcal{O}(\mathcal{M})$ is $\langle X, Y\rangle_{\mathrm{Sas}} := g\big(\pi_* X, \pi_* Y\big) + \big\langle \omega(X), \omega(Y)\big\rangle_{\mathfrak{o}(d)}$. The horizontal and vertical subspaces are orthogonal for this metric, and $\{H_a(U)\}_{a=1}^d$ is an orthonormal basis of H_U because $\omega(H_a) = 0$ and $\langle H_a, H_b\rangle_{\mathrm{Sas}} = g(e_a, e_b) = \delta_{ab}$.

Step 1 (directional derivatives). The Stratonovich chain rule applied to both sides of $\Phi(U_t) = \phi(E_t)$ gives

$$d\,\Phi(U_t) = \sum_{a=1}^{d}\big(H_a\Phi\big)(U_t)\circ dE_t^a,$$
$$d\,\phi(E_t) = \sum_{a=1}^{d}\partial_a\phi(E_t)\circ dE_t^a,$$
$$0 = d\big[\Phi(U_t) - \phi(E_t)\big] = \sum_{a=1}^{d}\Big[\big(H_a\Phi\big)(U_t) - \partial_a\phi(E_t)\Big]\circ dE_t^a,$$
$$d\big\langle E^a, W^b\big\rangle_t = \sigma_t\,\delta_{ab}\,dt.$$

Taking the quadratic covariation of the vanishing process with W^b isolates the b-th coefficient:

$$0 = d\big\langle \Phi(U_\bullet) - \phi(E_\bullet), W^b \big\rangle_t = \sum_{a=1}^{d} \Big[\big(H_a\Phi\big)(U_t) - \partial_a\phi(E_t) \Big] d\big\langle E^a, W^b \big\rangle_t$$
$$= \sigma_t \Big[\big(H_b\Phi\big)(U_t) - \partial_b\phi(E_t) \Big] dt.$$

Since $\sigma_t > 0$ and both coefficients are continuous in t, $(H_b\Phi)(U_t) = \partial_b\phi(E_t)$ for all t and $b = 1, \dots, d$, almost surely.

Step 2 (horizontal gradient). The horizontal gradient is the orthogonal projection $\nabla^{\mathrm{H}}\Phi = \mathbf{P}_{\mathrm{H}}(\nabla\Phi)$ of the Sasaki gradient onto H_U. Expanding in the orthonormal basis $\{H_a\}$ and using Step 1 and (16),

$$\nabla^{\mathrm{H}}\Phi(U_t) = \sum_{a=1}^{d} \big\langle \nabla\Phi(U_t), H_a(U_t) \big\rangle_{\mathrm{Sas}} \, H_a(U_t) = \sum_{a=1}^{d} d\Phi\big(H_a(U_t)\big) \, H_a(U_t)$$
$$= \sum_{a=1}^{d} \big(H_a\Phi\big)(U_t) \, H_a(U_t) = \sum_{a=1}^{d} \partial_a\phi(E_t) \, H_a(U_t)$$
$$= H_{e_t}(U_t) \Big[\sum_{a=1}^{d} \partial_a\phi(E_t) \, \varepsilon_a \Big] = H_{e_t}(U_t) \big[\nabla\phi(E_t) \big],$$

which is the first identity in (19).

Step 3 (densities). Let ρ_t be the density of E_t on $\mathbb{R}^d$ and ρ_t^{H} the density of U_t with respect to $d\mu := d\mathrm{Vol}_g \otimes d\mathrm{Haar}_{O(d)}$, where each fiber carries unit Haar mass. The right action is an isometry of the Sasaki metric, which gives $d\mu(u \cdot h) = d\mu(u)$ for $h \in O(d)$, and the push-forward of μ along π_E is the Lebesgue measure on $\mathbb{R}^d$. For every bounded measurable $F : \mathbb{R}^d \to \mathbb{R}$,

$$\int_{\mathbb{R}^d} F(E) \, \rho_t(E) \, dE = \mathbb{E}\big[F(E_t)\big] = \mathbb{E}\big[F\big(\pi_E(U_t)\big)\big] = \int_{\mathcal{O}(\mathcal{M})} F\big(\pi_E(U)\big) \, \rho_t^{\mathrm{H}}(U) \, d\mu(U)$$
$$= \int_{\mathcal{M}} \Big(\int_{O(d)} F\big(\pi_E(u \cdot h)\big) \, \rho_t^{\mathrm{H}}(u \cdot h) \, d\mathrm{Haar}_{O(d)}(h) \Big) d\mathrm{Vol}_g(x)$$
$$= \int_{\mathbb{R}^d} F(E) \, \rho_t^{\mathrm{H}}\big(\Psi(E_\bullet)_t\big) \, dE,$$

where the last equality combines $(\pi_E)_{\#}\mu = \mathrm{Leb}_{\mathbb{R}^d}$ with the unit mass of each $O(d)$-orbit, and ρ_t^{H} is evaluated on the gauge class $\Psi(E_\bullet)_t = [U_t]$. As F is arbitrary,

$$\rho_t^{\mathrm{H}}(U_t) = \rho_t(E_t), \qquad \log \rho_t^{\mathrm{H}}(U_t) = \log \rho_t(E_t), \qquad U_\bullet \in \Psi(E_\bullet).$$

Step 4 (scores). The pair $(\Phi, \phi) = (\log \rho_t^{\mathrm{H}}, \log \rho_t)$ satisfies the relation of the setting by Step 3. Steps 1–2 with this pair and $s(t, E) := \nabla \log \rho_t(E)$ give

$$s^{\mathrm{H}}(t, U_t) := \nabla^{\mathrm{H}} \log \rho_t^{\mathrm{H}}(U_t) = \sum_{a=1}^{d} \big(H_a \log \rho_t^{\mathrm{H}}\big)(U_t) \, H_a(U_t)$$
$$= \sum_{a=1}^{d} \partial_a \log \rho_t(E_t) \, H_a(U_t) = H_{e_t}(U_t) \big[\nabla \log \rho_t(E_t) \big]$$
$$= H_{e_t}(U_t) \big[s(t, E_t) \big] = \sum_{a=1}^{d} s^a(t, E_t) \, H_a(U_t),$$

which proves the second identity in (19) and (20). □

Proposition 9.2 (Gauge Equivariance of horizontal score function)**.** *The proposed horizontal score function is gauge-equivariant in the sense that $s^{\mathrm{H}}(t, U \cdot h) = (\mathcal{R}_h)_* s^{\mathrm{H}}(t, U)$ for all $h \in O(d)$.*

Proof. **Setting.** The frame bundle $\mathcal{O}(\mathcal{M})$ is a principal $O(d)$-bundle with projection π and right action $R_h(U) = U \cdot h = (x, eh)$, which satisfies $\pi \circ R_h = \pi$. The Levi–Civita connection one-form $\omega \in \Omega^1(\mathcal{O}(\mathcal{M}); \mathfrak{o}(d))$ satisfies

$$\omega|_{\mathrm{H}_U} = 0, \qquad R_h^*\omega = \mathrm{Ad}_{h^{-1}}\omega, \qquad dR_h(\mathrm{H}_U) = \mathrm{H}_{U\cdot h}, \qquad \omega(A^\sharp) = A \quad (A \in \mathfrak{o}(d)),$$

with $A^\sharp$ the fundamental vertical vector field of A. For $U = (x, e)$ and $w \in \mathbb{R}^d$, the horizontal lift is the unique vector

$$H(U)[w] \in \mathrm{H}_U \quad \text{such that} \quad d\pi_U\big(H(U)[w]\big) = ew. \tag{21}$$

Step 1 (right-invariance of the lift). The vector $dR_h\big(H(U)[hw]\big)$ lies in $dR_h(\mathrm{H}_U) = \mathrm{H}_{U\cdot h}$, and its projection is

$$d\pi_{U\cdot h}\Big(dR_h\big(H(U)[hw]\big)\Big) = d\big(\pi \circ R_h\big)_U\big(H(U)[hw]\big) = d\pi_U\big(H(U)[hw]\big) = e\,(hw) = (eh)\,w.$$

By uniqueness in (21) at $U \cdot h = (x, eh)$,

$$H\big(U \cdot h\big)[w] = dR_h\big(H(U)[hw]\big), \qquad \forall h \in O(d),\ w \in \mathbb{R}^d. \tag{22}$$

Step 2 (equivariance of the latent score). The latent density depends on E only through $\|E\|$, which gives $\rho_t(hE) = \rho_t(E)$ for $h \in O(d)$. Differentiating in E and using $h^\top = h^{-1}$,

$$\begin{aligned}\log \rho_t(hE) = \log \rho_t(E) \ &\Longrightarrow\ h^\top \nabla \log \rho_t(hE) = \nabla \log \rho_t(E) \\ &\Longrightarrow\ s(t, hE) = h\, s(t, E), \qquad s(t, h^{-1}E) = h^{-1} s(t, E).\end{aligned}$$

The latent coordinate attached to a frame transforms under the right action as

$$\pi_E(U \cdot h) = (eh)^{-1}\xi = h^{-1}e^{-1}\xi = h^{-1}\pi_E(U).$$

Step 3 (equivariance of the horizontal score). With $s^{\mathrm{H}}(t, U) := H(U)\big[s(t, \pi_E(U))\big] \in \mathrm{H}_U$ from Proposition 9.1 and $\mathbf{E}_t = \pi_E(U)$, Steps 1–2 give

$$\begin{aligned}s^{\mathrm{H}}\big(t, U \cdot h\big) &= H\big(U \cdot h\big)\big[s\big(t, \pi_E(U \cdot h)\big)\big] = H\big(U \cdot h\big)\big[s\big(t, h^{-1}\mathbf{E}_t\big)\big] = H\big(U \cdot h\big)\big[h^{-1} s(t, \mathbf{E}_t)\big] \\ &= dR_h\Big(H(U)\big[h\, h^{-1} s(t, \mathbf{E}_t)\big]\Big) = dR_h\Big(H(U)\big[s(t, \mathbf{E}_t)\big]\Big) = (R_h)_*\, s^{\mathrm{H}}(t, U),\end{aligned}$$

which is the statement. The transformed field remains horizontal:

$$\omega_{U\cdot h}\big(dR_h\, s^{\mathrm{H}}(t, U)\big) = \big(R_h^*\omega\big)_U\big(s^{\mathrm{H}}(t, U)\big) = \mathrm{Ad}_{h^{-1}}\,\omega_U\big(s^{\mathrm{H}}(t, U)\big) = \mathrm{Ad}_{h^{-1}}\,0 = 0.$$

Step 4 (compatibility with gauge transformations). Let $g : \mathcal{O}(\mathcal{M}) \to O(d)$ be a smooth gauge transformation, $R_g(U) := U \cdot g(U)$, and let the transformed connection one-form be [Kobayashi and Nomizu, 1996]

$$\omega^g := \mathrm{Ad}_{g^{-1}}\omega + g^{-1}dg.$$

For $X \in T_U\mathcal{O}(\mathcal{M})$ represented by a curve $U(\tau)$ with $U(0) = U$, the product rule applied to $U(\tau) \cdot g(U(\tau)) = \big(U(\tau) \cdot g(U)\big) \cdot \big(g(U)^{-1} g(U(\tau))\big)$ gives

$$\begin{aligned}dR_g(X) &= \frac{d}{d\tau}\Big(U(\tau) \cdot g\big(U(\tau)\big)\Big)\Big|_{\tau=0} = dR_{g(U)}(X) + \big(g(U)^{-1} dg_U(X)\big)^\sharp_{R_g(U)}, \\ \big(R_g^*\omega\big)_U(X) &= \omega_{R_g(U)}\big(dR_{g(U)}X\big) + \omega\Big(\big(g(U)^{-1}dg_U(X)\big)^\sharp\Big) \\ &= \mathrm{Ad}_{g(U)^{-1}}\,\omega_U(X) + g(U)^{-1}dg_U(X) = \omega_U^g(X).\end{aligned} \tag{23}$$

The score expressed in the transformed gauge is $s_g^{\mathrm{H}}(t, U) := \big(dR_g\big)_U^{-1} s^{\mathrm{H}}\big(t, R_g(U)\big)$. By (23) and the horizontality of s^{H},

$$\begin{aligned}\omega_U^g\big(s_g^{\mathrm{H}}(t, U)\big) &= \big(R_g^*\omega\big)_U\Big(\big(dR_g\big)_U^{-1} s^{\mathrm{H}}\big(t, R_g(U)\big)\Big) = \omega_{R_g(U)}\Big(dR_g\big(dR_g\big)^{-1} s^{\mathrm{H}}\big(t, R_g(U)\big)\Big) \\ &= \omega_{R_g(U)}\big(s^{\mathrm{H}}\big(t, R_g(U)\big)\big) = 0.\end{aligned}$$

Hence the transformed score is horizontal with respect to ω^g. For a constant gauge $g \equiv h$, $dg = 0$ and Step 3 give $s_h^{\mathrm{H}}(t, U) = (dR_h)^{-1} dR_h\, s^{\mathrm{H}}(t, U) = s^{\mathrm{H}}(t, U)$, in agreement with the equivariance identity. □

Corollary 9.3 (Isometry–equivariance of the base score)**.** *Let* $\mathrm{Isom} : \mathcal{M} \to \mathcal{M}$ *be any isometry on* $\mathcal{M}$ *whose lift to* $\mathcal{O}(\mathcal{M})$ *preserves the forward marginal* ρ_t^{H}*. Then, the manifold score function satisfies the equivariance:* $s^{\mathcal{M}}\big(t, \mathrm{Isom}(x)\big) = \mathrm{Isom}_* s^{\mathcal{M}}(t, x)$*, where* $\mathrm{Isom}_{*x} : T_x\mathcal{M} \to T_{\mathrm{Isom}(x)}\mathcal{M}$ *denotes the differential (push-forward) of* Isom *at the point* x*.*

Proof. **Step 1 (lift of the isometry).** For $U = (x, e) \in \mathcal{O}(\mathcal{M})$, define

$$\widetilde{\mathrm{Isom}}(U) := \big(\mathrm{Isom}(x),\, d\mathrm{Isom}_x \circ e\big), \qquad \pi \circ \widetilde{\mathrm{Isom}} = \mathrm{Isom} \circ \pi, \qquad \widetilde{\mathrm{Isom}}(U \cdot h) = \widetilde{\mathrm{Isom}}(U) \cdot h,$$

where $d\mathrm{Isom}_x \circ e$ is orthonormal because $d\mathrm{Isom}_x$ is a linear isometry. An isometry preserves the Levi–Civita connection; a frame transported in parallel along $x(\tau)$ is therefore mapped to a frame transported in parallel along $\mathrm{Isom}(x(\tau))$, and $d\widetilde{\mathrm{Isom}}_U(\mathrm{H}_U) = \mathrm{H}_{\widetilde{\mathrm{Isom}}(U)}$. The projection of the image of $H_a(U)$ is

$$\begin{aligned} d\pi\big(d\widetilde{\mathrm{Isom}}_U\, H_a(U)\big) &= d\big(\pi \circ \widetilde{\mathrm{Isom}}\big)_U\, H_a(U) = d\big(\mathrm{Isom} \circ \pi\big)_U\, H_a(U) \\ &= d\mathrm{Isom}_x\big(e(\varepsilon_a)\big) = \big(d\mathrm{Isom}_x \circ e\big)(\varepsilon_a), \end{aligned}$$

and uniqueness of the horizontal lift gives $d\widetilde{\mathrm{Isom}}_U\big(H_a(U)\big) = H_a\big(\widetilde{\mathrm{Isom}}(U)\big)$ for $a = 1, \dots, d$.

Step 2 (equivariance of the horizontal score). The invariance $\rho_t^{\mathrm{H}} \circ \widetilde{\mathrm{Isom}} = \rho_t^{\mathrm{H}}$ and Step 1 give, for each a,

$$\big(H_a \log \rho_t^{\mathrm{H}}\big)\big(\widetilde{\mathrm{Isom}}(U)\big) = \Big(d\widetilde{\mathrm{Isom}}_U\, H_a(U)\Big) \log \rho_t^{\mathrm{H}} = H_a(U)\big(\log \rho_t^{\mathrm{H}} \circ \widetilde{\mathrm{Isom}}\big) = \big(H_a \log \rho_t^{\mathrm{H}}\big)(U).$$

Expanding the score in the orthonormal horizontal basis,

$$\begin{aligned} s^{\mathrm{H}}\big(t, \widetilde{\mathrm{Isom}}(U)\big) &= \sum_{a=1}^{d} \big(H_a \log \rho_t^{\mathrm{H}}\big)\big(\widetilde{\mathrm{Isom}}(U)\big)\, H_a\big(\widetilde{\mathrm{Isom}}(U)\big) \\ &= \sum_{a=1}^{d} \big(H_a \log \rho_t^{\mathrm{H}}\big)(U)\, d\widetilde{\mathrm{Isom}}_U\big(H_a(U)\big) = d\widetilde{\mathrm{Isom}}_U\big[s^{\mathrm{H}}(t, U)\big]. \end{aligned}$$

Step 3 (projection to the base). The base score $s^{\mathcal{M}}(t, x) := d\pi_U\big[s^{\mathrm{H}}(t, U)\big]$, $U \in \pi^{-1}(x)$, is independent of the frame by Proposition 9.2:

$$d\pi_{U \cdot h}\big[s^{\mathrm{H}}(t, U \cdot h)\big] = d\pi_{U \cdot h}\, dR_h\big[s^{\mathrm{H}}(t, U)\big] = d\big(\pi \circ R_h\big)_U\big[s^{\mathrm{H}}(t, U)\big] = d\pi_U\big[s^{\mathrm{H}}(t, U)\big].$$

Combining Steps 1–2 with $\pi \circ \widetilde{\mathrm{Isom}} = \mathrm{Isom} \circ \pi$,

$$\begin{aligned} s^{\mathcal{M}}\big(t, \mathrm{Isom}(x)\big) &= d\pi_{\widetilde{\mathrm{Isom}}(U)}\big[s^{\mathrm{H}}\big(t, \widetilde{\mathrm{Isom}}(U)\big)\big] = d\pi_{\widetilde{\mathrm{Isom}}(U)} \circ d\widetilde{\mathrm{Isom}}_U\big[s^{\mathrm{H}}(t, U)\big] \\ &= d\mathrm{Isom}_x \circ d\pi_U\big[s^{\mathrm{H}}(t, U)\big] = d\mathrm{Isom}_x\big[s^{\mathcal{M}}(t, x)\big] = \mathrm{Isom}_{*x}\, s^{\mathcal{M}}(t, x). \end{aligned}$$

□

Proposition 9.4 (Forward Markovian Density of Horizontal Diffusions on Frame-Bundle)**.** *Let* $\nu_t = \mathcal{L}(\mathbf{U}_t, \mathbf{E}_t)$ *be law of joint dynamics* $\mathbf{U}_t$ *and* $\mathbf{E}_t$ *and* ρ_t^{H} *be its marginal distribution on* $\mathcal{M}$*. Then the horizontal transition distribution* $p_t^E(U \mid \mathbf{U}_0)$ *for every* $x \in \mathcal{M}$ *is well-defined and given as follows:*

$$d\nu_t(U, E) = \Big[\int_{\mathcal{O}(\mathcal{M})} p_t^E(U \mid \mathbf{U}_0)\rho_0(\mathbf{U}_0)d\lambda(\mathbf{U}_0)\Big]\mathcal{N}(\mathbf{m}_t, \Sigma_t; E)dEd\lambda(U). \tag{24}$$

where $p_t^{\mathbf{E}}$ *is* ***Markovian transition*** *between frame-valued states with respect to horizontal lift of* $d\mathbf{E}_t$*. Furthermore, the reverse dynamics* $\mathbf{U}_s$ *is a time-reversal process of* $\mathbf{U}_t$ *in the sense of Haussmann and Pardoux [1986].*

Remark 9.5. The expression of $d\nu_t$ can be regarded as an extension of the Euclidean lifted distribution representation. In particular, the conditional distribution $p_t^{\mathbf{E}}(U \mid \mathbf{U}_0)$ on the frame bundle generalizes the Euclidean transition by incorporating the horizontal lift structure associated with $\circ d\mathbf{E}_t$, while the marginal distribution $\rho_0(\mathbf{U}_0)$ corresponds to the lifted initial distribution. This decomposition preserves a structurally analogous form to the Euclidean lifted dynamics, now formulated on the orthonormal frame bundle $\mathcal{O}(\mathcal{M})$ under the horizontal lifting framework.

Proof. Throughout, the forward dynamics is $d\mathbf{U}_t = -\frac{1}{2}\nabla^{\mathrm{H}}\Phi_t(\mathbf{U}_t)\,dt + \sigma_t \sum_a H_a(\mathbf{U}_t) \circ dW_t^a$, which reduces to Definition 9.7 for $\sigma_t \equiv 1$, and λ denotes the Sasaki volume on $\mathcal{O}(\mathcal{M})$. We use the identity $\sum_a e_a^i e_a^j = g^{ij}(x)$, which follows from $e^\top g(x) e = I$, and the coordinate form $H_a = e_a^i \partial_{x^i} - \Gamma_{ij}^k e_a^i e_b^j \partial_{e_b^k}$ from (16).

Step 1 (generator). By Proposition 9.1, the drift expands in the horizontal basis as

$$-\frac{1}{2}\nabla^{\mathrm{H}}\Phi_t(\mathbf{U}_t) = -\frac{1}{2}\sum_{a=1}^{d}\big(H_a\Phi_t\big)(\mathbf{U}_t)\,H_a(\mathbf{U}_t) = -\frac{1}{2}\sum_{a=1}^{d}\partial_a\phi_t(\mathbf{E}_t)\,H_a(\mathbf{U}_t).$$

For $f \in C^\infty(\mathcal{O}(\mathcal{M}))$, the Stratonovich chain rule followed by the conversion $Y \circ dW^a = Y\,dW^a + \frac{1}{2}d\langle Y, W^a\rangle$ with $Y = \sigma_t(H_a f)(\mathbf{U}_t)$ gives

$$\begin{aligned} df(\mathbf{U}_t) &= -\frac{1}{2}\sum_{a=1}^{d}\big(H_a\Phi_t\big)\big(H_a f\big)(\mathbf{U}_t)\,dt + \sigma_t\sum_{a=1}^{d}\big(H_a f\big)(\mathbf{U}_t)\circ dW_t^a \\ &= \Big[-\frac{1}{2}\sum_{a=1}^{d}\big(H_a\Phi_t\big)\big(H_a f\big) + \frac{\sigma_t^2}{2}\sum_{a=1}^{d}H_a\big(H_a f\big)\Big](\mathbf{U}_t)\,dt + \sigma_t\sum_{a=1}^{d}\big(H_a f\big)(\mathbf{U}_t)\,dW_t^a. \end{aligned}$$

The time-inhomogeneous generator in Hörmander form is therefore

$$\mathcal{L}_t^{\mathbf{E}} = -\frac{1}{2}\sum_{a=1}^{d}\partial_a\phi_t(\mathbf{E}_t)\,H_a + \frac{\sigma_t^2}{2}\sum_{a=1}^{d}H_a H_a, \tag{25}$$

where the superscript records that the drift coefficients are the latent gradients $\partial_a\phi_t(\mathbf{E}_t) = (H_a\Phi_t)(\mathbf{U}_t)$. In the chart (x^i, e_b^k), the components $H_a^{x^i} = e_a^i$ and $H_a^{e_b^k} = -\Gamma_{ij}^k e_a^i e_b^j$ give, for the second-order part,

$$\sum_a H_a^{x^i}H_a^{x^j} = g^{ij}, \qquad \sum_a H_a^{x^i}H_a^{e_b^k} = -g^{ip}\Gamma_{pq}^k e_b^q, \qquad \sum_a H_a^{e_b^k}H_a^{e_c^l} = g^{pr}\Gamma_{pq}^k\Gamma_{rs}^l e_b^q e_c^s,$$

and, with $H_a(e_a^i) = -\Gamma_{mn}^i e_a^m e_a^n$ and $H_a(e_b^j) = -\Gamma_{mn}^j e_a^m e_b^n$, for the first-order part $\mathcal{V} := \sum_a (H_a H_a^\alpha)\partial_\alpha$,

$$\begin{aligned} \sum_a H_a\big(H_a^{x^i}\big) &= \sum_a H_a\big(e_a^i\big) = -\sum_a \Gamma_{mn}^i e_a^m e_a^n = -g^{mn}\Gamma_{mn}^i, \\ \sum_a H_a\big(H_a^{e_b^k}\big) &= -\sum_a\Big[e_a^l\partial_l\Gamma_{ij}^k\,e_a^i e_b^j + \Gamma_{ij}^k H_a(e_a^i)\,e_b^j + \Gamma_{ij}^k e_a^i\,H_a(e_b^j)\Big] \\ &= -g^{li}\partial_l\Gamma_{ij}^k\,e_b^j + g^{mn}\Gamma_{ij}^k\Gamma_{mn}^i\,e_b^j + g^{im}\Gamma_{ij}^k\Gamma_{mn}^j\,e_b^n. \end{aligned}$$

Collecting both parts,

$$\begin{aligned} \sum_{a=1}^{d}H_a H_a f &= \sum_a H_a^\alpha H_a^\beta\,\partial_\alpha\partial_\beta f + \mathcal{V}f \\ &= g^{ij}\frac{\partial^2 f}{\partial x^i\partial x^j} - 2g^{ip}\Gamma_{pq}^k e_b^q\frac{\partial^2 f}{\partial x^i\partial e_b^k} + g^{pr}\Gamma_{pq}^k\Gamma_{rs}^l e_b^q e_c^s\frac{\partial^2 f}{\partial e_b^k\partial e_c^l} - g^{mn}\Gamma_{mn}^i\frac{\partial f}{\partial x^i} \\ &\quad + \Big(-g^{li}\partial_l\Gamma_{ij}^k\,e_b^j + g^{mn}\Gamma_{ij}^k\Gamma_{mn}^i\,e_b^j + g^{im}\Gamma_{ij}^k\Gamma_{mn}^j\,e_b^n\Big)\frac{\partial f}{\partial e_b^k}. \end{aligned}$$

Step 2 (brackets and hypoellipticity). The base components of the bracket vanish by the symmetry $\Gamma^k_{ij} = \Gamma^k_{ji}$ of the torsion-free connection:

$$[H_a, H_b]\, x^k = H_a\big(e^k_b\big) - H_b\big(e^k_a\big) = -\Gamma^k_{ij} e^i_a e^j_b + \Gamma^k_{ij} e^i_b e^j_a = -\big(\Gamma^k_{ij} - \Gamma^k_{ji}\big) e^i_a e^j_b = 0.$$

For the frame components, $H_b(e^k_c) = -\Gamma^k_{ij} e^i_b e^j_c$ and the product rule give

$$\begin{aligned}
H_a\big(H_b\, e^k_c\big) &= -e^p_a e^i_b e^q_c\, \partial_p \Gamma^k_{iq} + \Gamma^k_{jq}\Gamma^j_{pi}\, e^p_a e^i_b e^q_c + \Gamma^k_{ij}\Gamma^j_{pq}\, e^p_a e^i_b e^q_c, \\
[H_a, H_b]\, e^k_c &= H_a\big(H_b\, e^k_c\big) - H_b\big(H_a\, e^k_c\big) \\
&= -e^p_a e^i_b e^q_c \Big[\partial_p \Gamma^k_{iq} - \partial_i \Gamma^k_{pq} + \Gamma^k_{pj}\Gamma^j_{iq} - \Gamma^k_{ij}\Gamma^j_{pq}\Big] = -R^k_{qpi}(x)\, e^p_a e^i_b e^q_c,
\end{aligned}$$

where the term $\Gamma^k_{jq}\Gamma^j_{pi} e^p_a e^i_b$ is symmetric in (a,b) and cancels in the bracket, and $R^k_{qpi} = \partial_p \Gamma^k_{iq} - \partial_i \Gamma^k_{pq} + \Gamma^k_{pj}\Gamma^j_{iq} - \Gamma^k_{ij}\Gamma^j_{pq}$ are the components of $R(\partial_p, \partial_i)\partial_q$. Hence

$$\begin{aligned}
[H_a, H_b] &= -R^k_{lij}(x)\, e^i_a e^j_b e^l_c\, \frac{\partial}{\partial e^k_c} = -\Omega(H_a, H_b)^\sharp, \\
\omega\big([H_a, H_b]\big) &= -\Omega(H_a, H_b) = -e^{-1} R(e_a, e_b)\, e,
\end{aligned} \tag{26}$$

with $\Omega = d\omega + \frac{1}{2}[\omega, \omega]$ the curvature two-form, in agreement with Cartan's structure equation $\omega([Y,Z]) = -d\omega(Y,Z) = -\Omega(Y,Z)$ for horizontal Y, Z. The bracket is purely vertical. Whenever $\{R_x(e_a, e_b)\}_{a<b}$ spans $\mathfrak{so}(T_x\mathcal{M})$, which holds for non-degenerate curvature,

$$\begin{aligned}
\mathrm{Lie}\{H_1, \dots, H_d\}_U &\supseteq \mathrm{span}\{H_a(U)\}_a \oplus \mathrm{span}\{[H_a, H_b](U)\}_{a<b} \\
&= \mathrm{H}_U \oplus \mathrm{Ver}_U = T_U\mathcal{O}(\mathcal{M}), \qquad \forall\, U \in \mathcal{O}(\mathcal{M}).
\end{aligned}$$

Hörmander's bracket condition holds for (25), and $\partial_t - \mathcal{L}^{\mathbf{E}}_t$ is hypoelliptic. The law of $\mathbf{U}_t$ started at $\mathbf{U}_0$ therefore admits a smooth transition density $\bar{p}_t(U \mid \mathbf{U}_0)$ with respect to λ. The fundamental horizontal vector fields are divergence-free with respect to λ [Hsu, 2002], which gives $\int (H_a f)\, p\, d\lambda = -\int f\, (H_a p)\, d\lambda$ and the formal adjoint

$$\begin{aligned}
\int_{\mathcal{O}(\mathcal{M})} \big(\mathcal{L}^{\mathbf{E}}_t f\big)\, p\, d\lambda &= \int_{\mathcal{O}(\mathcal{M})} \Big[-\frac{1}{2}\sum_a \big(H_a\Phi_t\big)\big(H_a f\big) + \frac{\sigma_t^2}{2}\sum_a H_a H_a f\Big] p\, d\lambda \\
&= \int_{\mathcal{O}(\mathcal{M})} f\, \Big[\frac{1}{2}\sum_a H_a\big((H_a\Phi_t)\, p\big) + \frac{\sigma_t^2}{2}\sum_a H_a H_a p\Big] d\lambda \\
&= \int_{\mathcal{O}(\mathcal{M})} f\, \big(\mathcal{L}^{\mathbf{E}}_t\big)^* p\, d\lambda.
\end{aligned}$$

The density solves the forward Kolmogorov equation with Dirac initial condition,

$$\begin{aligned}
\partial_t \bar{p}_t(\cdot \mid \mathbf{U}_0) &= \big(\mathcal{L}^{\mathbf{E}}_t\big)^* \bar{p}_t(\cdot \mid \mathbf{U}_0) = \frac{1}{2}\sum_a H_a\big((H_a\Phi_t)\, \bar{p}_t\big) + \frac{\sigma_t^2}{2}\sum_a H_a H_a \bar{p}_t, \\
\bar{p}_0(\cdot \mid \mathbf{U}_0) &= \delta_{\mathbf{U}_0}.
\end{aligned}$$

Step 3 (Markov property and marginal law). Writing $\bar{p}_{s,t}(U \mid U')$ for the transition density from time s to time t ($\bar{p}_{0,t} = \bar{p}_t$), the Markov property of the strong solution yields the Chapman–Kolmogorov identity, and averaging over the initial law yields the marginal density:

$$\begin{aligned}
\int_{\mathcal{O}(\mathcal{M})} \bar{p}_{0,s}\big(U' \mid \mathbf{U}_0\big)\, \bar{p}_{s,t}\big(U \mid U'\big)\, d\lambda(U') &= \bar{p}_{0,t}\big(U \mid \mathbf{U}_0\big), \qquad 0 \le s \le t, \\
\rho^{\mathrm{H}}_t(U) &= \int_{\mathcal{O}(\mathcal{M})} \bar{p}_t\big(U \mid \mathbf{U}_0\big)\rho_0(\mathbf{U}_0)\, d\lambda(\mathbf{U}_0), \\
\partial_t \rho^{\mathrm{H}}_t &= \int_{\mathcal{O}(\mathcal{M})} \big(\mathcal{L}^{\mathbf{E}}_t\big)^* \bar{p}_t\big(\cdot \mid \mathbf{U}_0\big)\rho_0(\mathbf{U}_0)\, d\lambda(\mathbf{U}_0) = \big(\mathcal{L}^{\mathbf{E}}_t\big)^* \rho^{\mathrm{H}}_t.
\end{aligned}$$

Step 4 (joint law). Let $p_t^{\mathbf{E}}(U \mid \mathbf{U}_0)$ denote the density of the conditional law of $\mathbf{U}_t$ given $\mathbf{U}_0$ and $\mathbf{E}_t = E$. Since $\mathbf{E}_t \sim \mathcal{N}(\mathbf{m}_t, \Sigma_t)$ independently of $\mathbf{U}_0$, the disintegration of ν_t reads

$$\begin{aligned} d\nu_t(U,E) &= \int_{\mathcal{O}(\mathcal{M})} \mathbb{P}\big(\mathbf{U}_t \in dU \mid \mathbf{U}_0, \mathbf{E}_t = E\big)\, \mathbb{P}\big(\mathbf{E}_t \in dE \mid \mathbf{U}_0\big)\, \rho_0(\mathbf{U}_0)\, d\lambda(\mathbf{U}_0) \\ &= \int_{\mathcal{O}(\mathcal{M})} p_t^{\mathbf{E}}\big(U \mid \mathbf{U}_0\big) d\lambda(U)\, \mathcal{N}(\mathbf{m}_t, \Sigma_t; E)\, dE\, \rho_0(\mathbf{U}_0)\, d\lambda(\mathbf{U}_0) \\ &= \Big[\int_{\mathcal{O}(\mathcal{M})} p_t^{\mathbf{E}}\big(U \mid \mathbf{U}_0\big) \rho_0(\mathbf{U}_0)\, d\lambda(\mathbf{U}_0)\Big] \mathcal{N}(\mathbf{m}_t, \Sigma_t; E)\, dE\, d\lambda(U), \end{aligned}$$

which is the stated representation. Integrating out E recovers the hypoelliptic transition density of Step 2 by the law of total probability,

$$\begin{aligned} \int_{\mathbb{R}^d} d\nu_t(U,E) &= \Big[\int_{\mathcal{O}(\mathcal{M})} \bar{p}_t\big(U \mid \mathbf{U}_0\big) \rho_0(\mathbf{U}_0)\, d\lambda(\mathbf{U}_0)\Big] d\lambda(U), \\ \bar{p}_t\big(U \mid \mathbf{U}_0\big) &= \int_{\mathbb{R}^d} p_t^{\mathbf{E}}\big(U \mid \mathbf{U}_0\big)\, \mathcal{N}(\mathbf{m}_t, \Sigma_t; E)\, dE. \end{aligned}$$

Step 5 (stationary law). At constant temperature $\sigma_t \equiv 1$ and for a time-independent confining potential Φ, the identity $H_a e^{-\Phi} = -(H_a\Phi)e^{-\Phi}$ gives

$$\begin{aligned} \big(\mathcal{L}^{\mathbf{E}}\big)^* e^{-\Phi} &= \frac{1}{2}\sum_a H_a\big((H_a\Phi)\, e^{-\Phi}\big) + \frac{1}{2}\sum_a H_a\big(H_a\, e^{-\Phi}\big) \\ &= \frac{1}{2}\sum_a H_a\big((H_a\Phi)\, e^{-\Phi}\big) - \frac{1}{2}\sum_a H_a\big((H_a\Phi)\, e^{-\Phi}\big) = 0. \end{aligned}$$

The Gibbs measure $\nu_\infty(dU) = Z^{-1} e^{-\Phi(U)} d\mathrm{Vol}_{\mathrm{Sas}}(U)$ with $Z = \int_{\mathcal{O}(\mathcal{M})} e^{-\Phi} d\mathrm{Vol}_{\mathrm{Sas}} < \infty$ is therefore invariant. Hypoellipticity (Step 2) and confinement make the lifted diffusion ergodic with ν_∞ as its unique invariant measure, and $\rho_t^{\mathrm{H}}\, d\lambda \to \nu_\infty$ as $t \to \infty$.

Step 6 (time reversal). Let $\check{\mathbf{U}}_s := \mathbf{U}_{T-s}$, $s \in [0,T]$, whose density is ρ_{T-s}^{H}. Using Step 3, $H_a\rho^{\mathrm{H}} = (H_a \log \rho^{\mathrm{H}})\rho^{\mathrm{H}}$, and splitting $-\frac{\sigma^2}{2} = \frac{\sigma^2}{2} - \sigma^2$,

$$\begin{aligned} \partial_s \rho_{T-s}^{\mathrm{H}} &= -\big(\mathcal{L}_{T-s}^{\mathbf{E}}\big)^* \rho_{T-s}^{\mathrm{H}} \\ &= -\frac{1}{2}\sum_a H_a\big((H_a\Phi_{T-s})\, \rho_{T-s}^{\mathrm{H}}\big) - \frac{\sigma_{T-s}^2}{2}\sum_a H_a H_a \rho_{T-s}^{\mathrm{H}} \\ &= -\frac{1}{2}\sum_a H_a\big((H_a\Phi_{T-s})\, \rho_{T-s}^{\mathrm{H}}\big) - \sigma_{T-s}^2 \sum_a H_a\big((H_a \log \rho_{T-s}^{\mathrm{H}})\, \rho_{T-s}^{\mathrm{H}}\big) \\ &\quad + \frac{\sigma_{T-s}^2}{2}\sum_a H_a H_a \rho_{T-s}^{\mathrm{H}} \\ &= -\sum_a H_a\Big(\Big[\frac{1}{2} H_a\Phi_{T-s} + \sigma_{T-s}^2\, H_a \log \rho_{T-s}^{\mathrm{H}}\Big]\rho_{T-s}^{\mathrm{H}}\Big) + \frac{\sigma_{T-s}^2}{2}\sum_a H_a H_a \rho_{T-s}^{\mathrm{H}}. \end{aligned}$$

By the adjoint formula of Step 2, the last line is the forward Kolmogorov equation of the diffusion whose drift has coefficient $\frac{1}{2}H_a\Phi_{T-s} + \sigma_{T-s}^2 H_a \log \rho_{T-s}^{\mathrm{H}}$ along H_a. With $\nabla^{\mathrm{H}}\Phi = \sum_a (H_a\Phi) H_a$ and $s^{\mathrm{H}} = \sum_a (H_a \log \rho^{\mathrm{H}}) H_a$, the Haussmann–Pardoux theorem Haussmann and Pardoux [1986] identifies $\check{\mathbf{U}}$ as the solution of

$$\begin{aligned} d\check{\mathbf{U}}_s &= \Big(\frac{1}{2}\nabla^{\mathrm{H}}\Phi_{T-s}(\check{\mathbf{U}}_s) + \sigma_{T-s}^2\, s^{\mathrm{H}}\big(T-s, \check{\mathbf{U}}_s\big)\Big) ds + \sigma_{T-s}\sum_a H_a(\check{\mathbf{U}}_s) \circ d\bar{W}_s^a, \\ d\bar{W}_s^a &= dW_{T-s}^a, \\ \check{\mathcal{L}}_s &= \sum_a \Big(\frac{1}{2} H_a\Phi_{T-s} + \sigma_{T-s}^2\, H_a \log \rho_{T-s}^{\mathrm{H}}\Big) H_a + \frac{\sigma_{T-s}^2}{2}\sum_a H_a H_a, \\ \partial_s \rho_{T-s}^{\mathrm{H}} &= \check{\mathcal{L}}_s^*\, \rho_{T-s}^{\mathrm{H}}. \end{aligned}$$

For $\sigma_t \equiv 1$ and with s indexing reverse time, this is the reverse SDE of Definition 9.7. The forward–reverse pair $(\mathcal{L}_t^{\mathbf{E}}, \check{\mathcal{L}}_{T-t})$ shares the marginals ρ_t^{H}, and the kernel $p_t^{\mathbf{E}}$ retains the lifted Gaussian representation of Step 4. □

Corollary 9.6 (Marginal Density on the Base Manifold)**.** *Let $\nu_t = \mathcal{L}(\mathbf{U}_t, \mathbf{E}_t)$ denote the joint distribution of the lifted frame-valued state $\mathbf{U}_t \in \mathcal{O}(\mathcal{M})$ and the latent Euclidean path terminal value $\mathbf{E}_t \in \mathbb{R}^d$ defined earlier, and let ρ_t^{H} be the induced marginal distribution on the base manifold $\mathcal{M}$. Then for every $x \in \mathcal{M}$, the marginal distribution $\rho_t^{\mathrm{H}}(x)$ is given by the fiber-wise integral*

$$\rho_t^{\mathrm{H}}(x) = \int_{O_x(M)} \int_{\mathbb{R}^d} \int_{\mathcal{O}(\mathcal{M})} p_t^{\mathbf{E}}(U \mid \mathbf{U}_0) \rho_0(\mathbf{U}_0) \mathcal{N}(\mathbf{m}_t, \Sigma_t; E) d\Xi(U, \mathbf{U}_0, E),$$

where $O_x(\mathcal{M}) := \pi^{-1}(\{x\})$ is the orthonormal frame fiber over $x \in M$, and the joint measure Ξ is defined by

$$\Xi := \lambda_0 \otimes \mathit{Leb}_{\mathbb{R}^d} \otimes \mathit{Haar}_{\mathrm{O}(d)},$$

with λ_0 denoting the reference volume measure on $\mathcal{O}(\mathcal{M})$, $\mathit{Leb}_{\mathbb{R}^d}$ the Lebesgue measure on the latent space, and $\mathit{Haar}_{\mathrm{O}(d)}$ the normalized Haar measure on the orthonormal group fibers.

Proof. The reference volume disintegrates along π into the Riemannian volume on the base and the normalized Haar measure on each fiber, $d\lambda_0(U) = d\mathrm{Haar}_{O(d)}(U)\, d\mathrm{Vol}_g(x)$ for $U \in O_x(\mathcal{M})$. For every bounded measurable $F : \mathcal{M} \to \mathbb{R}$, Proposition 9.4 and Fubini's theorem give

$$\begin{aligned}
\mathbb{E}\big[F(\pi(\mathbf{U}_t))\big] &= \int_{\mathcal{O}(\mathcal{M}) \times \mathbb{R}^d} F\big(\pi(U)\big)\, d\nu_t(U, E) \\
&= \int_{\mathcal{O}(\mathcal{M})} \int_{\mathbb{R}^d} F\big(\pi(U)\big) \Big[\int_{\mathcal{O}(\mathcal{M})} p_t^{\mathbf{E}}\big(U \mid \mathbf{U}_0\big) \rho_0(\mathbf{U}_0)\, d\lambda_0(\mathbf{U}_0) \Big] \\
&\qquad \times \mathcal{N}(\mathbf{m}_t, \Sigma_t; E)\, dE\, d\lambda_0(U) \\
&= \int_{\mathcal{M}} F(x) \Big[\int_{O_x(\mathcal{M})} \int_{\mathbb{R}^d} \int_{\mathcal{O}(\mathcal{M})} p_t^{\mathbf{E}}\big(U \mid \mathbf{U}_0\big) \rho_0(\mathbf{U}_0)\, \mathcal{N}(\mathbf{m}_t, \Sigma_t; E) \\
&\qquad \times d\lambda_0(\mathbf{U}_0)\, dE\, d\mathrm{Haar}_{O(d)}(U) \Big] d\mathrm{Vol}_g(x) \\
&= \int_{\mathcal{M}} F(x) \Big[\int_{O_x(\mathcal{M})} \int_{\mathbb{R}^d} \int_{\mathcal{O}(\mathcal{M})} p_t^{\mathbf{E}}\big(U \mid \mathbf{U}_0\big) \rho_0(\mathbf{U}_0)\, \mathcal{N}(\mathbf{m}_t, \Sigma_t; E) \\
&\qquad \times d\Xi(U, \mathbf{U}_0, E) \Big] d\mathrm{Vol}_g(x),
\end{aligned}$$

where $d\Xi(U, \mathbf{U}_0, E) = d\lambda_0(\mathbf{U}_0)\, dE\, d\mathrm{Haar}_{O(d)}(U)$ collects the three integrations. Since F is arbitrary, the bracket is the density of $\pi_{\#}\mathcal{L}(\mathbf{U}_t)$ with respect to $d\mathrm{Vol}_g$, which is the stated fiber-wise integral for $\rho_t^{\mathrm{H}}(x)$. □

Definition 9.7 (Horizontal Diffusion Models)**.** Let $\boldsymbol{\varepsilon} = \{\varepsilon_1, \ldots, \varepsilon_d\}$ denote the standard basis of $\mathbb{R}^d$. The horizontal diffusion models are the system of forward-reverse Stratonovich SDEs on the orthonormal frame bundle $\mathcal{O}(\mathcal{M})$, with local coordinate $\mathbf{U}_t = (x_t, e_t) \in \mathcal{O}(\mathcal{M})$, evolves according to

$$\begin{bmatrix} d\mathbf{U}_t \\ d\mathbf{U}_s \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\nabla^{\mathrm{H}}\Phi_t(\mathbf{U}_t) \\ \frac{1}{2}\nabla^{\mathrm{H}}\Phi_s(\mathbf{U}_s) + s^{\mathrm{H}}(s, \mathbf{U}_s) \end{bmatrix} \begin{bmatrix} dt \\ ds \end{bmatrix}^{\intercal} + \begin{bmatrix} H_a(\mathbf{U}_t) \\ H_a(\mathbf{U}_s) \end{bmatrix} \begin{bmatrix} \circ\, dW_t^a \\ \circ\, dW_s^a \end{bmatrix}^{\intercal}, \tag{27}$$

Then, the **horizontal score function** s^{H} is defined as time-indexed horizontal vector fields

$$s^{\mathrm{H}}(t, \mathbf{U}_t) \coloneqq \nabla^{\mathrm{H}} \log \rho_t^{\mathrm{H}}(\mathbf{U}_t), \tag{28}$$

which lies in the horizontal distribution $\mathrm{H}_{\mathbf{U}_t} \subset T_{\mathbf{U}_t}\mathcal{O}(\mathcal{M})$ at each point $\mathbf{U}_t \in \mathcal{O}(\mathcal{M})$. Here, ρ_t^{H} denotes a Markovian probability distribution with respect to forward frame dynamics, *i.e.*, $\mathbf{U}_t \sim \rho_t^{\mathrm{H}}$.

We end this section by integrating the preceding theoretical results to elucidate the rationale of horizontal diffusion models. For a frame $e : \mathbb{R}^d \simeq T_x\mathcal{M}$ and a vector $w \in \mathbb{R}^d$, the horizontal lift $H_e(\mathbf{U})[w] \in \mathrm{H}_{\mathbf{U}}$ at $\mathbf{U} := (x, e)$ is characterized by $\omega\big(H_e(\mathbf{U})[w]\big) = 0$ and $\pi_* H_e(\mathbf{U})[w] = e(w)$. The latent Euclidean diffusion and its stochastic development $\mathbf{U}_t = (x_t, e_t)$ are related through Lemma 8.1 and Proposition 9.1 as

$$
\begin{aligned}
d\mathbf{E}_t &= -\frac{1}{2}\nabla\phi_t(\mathbf{E}_t)\,dt + \sigma_t \circ dW_t, \qquad \mathbf{E}_0 \sim \rho_0,\\
d\mathbf{U}_t &= H_{e_t}(\mathbf{U}_t)\big[\circ d\mathbf{E}_t\big]\\
&= H_{e_t}(\mathbf{U}_t)\Big[-\frac{1}{2}\nabla\phi_t(\mathbf{E}_t)\Big]dt + \sigma_t \sum_{a=1}^{d} H_{e_t}(\mathbf{U}_t)[\varepsilon_a] \circ dW_t^a\\
&= -\frac{1}{2}\nabla^{\mathrm{H}}\Phi_t(\mathbf{U}_t)\,dt + \sigma_t \sum_{a=1}^{d} H_a(\mathbf{U}_t) \circ dW_t^a,
\end{aligned}
$$

which coincides with the forward equation of Definition 9.7 for $\sigma_t \equiv 1$. Expanding the horizontal gradient in the Sasaki-orthonormal basis $\{H_a\}$, using $(H_a\Phi_t)(\mathbf{U}_t) = \partial_{E^a}\phi_t(\mathbf{E}_t)$ from Step 1 of the proof of Proposition 9.1, and inserting the coordinate form (16),

$$
\begin{aligned}
\nabla^{\mathrm{H}}\Phi_t(\mathbf{U}_t) &= \sum_{a=1}^{d}\big(H_a\Phi_t\big)(\mathbf{U}_t)\,H_a(\mathbf{U}_t) = \sum_{a=1}^{d}\Big(e_a^i\frac{\partial\Phi_t}{\partial x^i} - \Gamma_{ij}^k(x_t)\,e_a^i e_b^j\frac{\partial\Phi_t}{\partial e_b^k}\Big)H_a(\mathbf{U}_t)\\
&= \sum_{a=1}^{d}\partial_{E^a}\phi_t(\mathbf{E}_t)\,H_a(\mathbf{U}_t) = \sum_{a=1}^{d}\partial_{E^a}\phi_t(\mathbf{E}_t)\Big(e_a^i\frac{\partial}{\partial x^i} - \Gamma_{ij}^k(x_t)\,e_a^i e_b^j\frac{\partial}{\partial e_b^k}\Big). \qquad (29)
\end{aligned}
$$

Reading off the x^i- and e_b^k-components of the forward equation with (29) gives the lifted forward process in local coordinates:

$$
\begin{aligned}
dx_t^i &= -\frac{1}{2}\sum_{a=1}^{d}\partial_{E^a}\phi_t(\mathbf{E}_t)\,e_a^i\,dt + \sigma_t\sum_{a=1}^{d}e_a^i \circ dW_t^a = \sum_{a=1}^{d}e_a^i \circ d\mathbf{E}_t^a,\\
de_{b,t}^k &= \frac{1}{2}\sum_{a=1}^{d}\partial_{E^a}\phi_t(\mathbf{E}_t)\,\Gamma_{ij}^k(x_t)\,e_a^i e_b^j\,dt - \sigma_t\sum_{a=1}^{d}\Gamma_{ij}^k(x_t)\,e_a^i e_b^j \circ dW_t^a = -\Gamma_{ij}^k(x_t)\,e_b^j \circ dx_t^i,
\end{aligned}
$$

where the last identity expresses that each frame vector is transported in parallel along $x_\bullet$. By linearity of the lift and Proposition 9.1, the Euclidean reverse drift is lifted term by term,

$$
\begin{aligned}
H_{e_t}(\mathbf{U}_t)\Big[\frac{1}{2}\nabla\phi_t(\mathbf{E}_t) + \sigma_t^2\nabla\log\rho_t(\mathbf{E}_t)\Big] &= \frac{1}{2}\nabla^{\mathrm{H}}\Phi_t(\mathbf{U}_t) + \sigma_t^2\,\nabla^{\mathrm{H}}\log\rho_t^{\mathrm{H}}(\mathbf{U}_t)\\
&= \frac{1}{2}\nabla^{\mathrm{H}}\Phi_t(\mathbf{U}_t) + \sigma_t^2\,s^{\mathrm{H}}(t,\mathbf{U}_t),
\end{aligned}
$$

with ρ_t^{H} the marginal of $\mathbf{U}_t$ under the forward dynamics. Step 6 of the proof of Proposition 9.4 then gives the lifted reverse Stratonovich SDE, with s indexing reverse time:

$$
d\mathbf{U}_s = \Big[\frac{1}{2}\nabla^{\mathrm{H}}\Phi_s(\mathbf{U}_s) + \sigma_s^2\nabla^{\mathrm{H}}\log\rho_s^{\mathrm{H}}(\mathbf{U}_s)\Big]\,ds + \sigma_s\sum_{a=1}^{d}H_{e_s}(\mathbf{U}_s)[\varepsilon_a] \circ dW_s^a, \qquad (30)
$$

where H_{e_s} denotes the horizontal lift with respect to the moving frame e_s at $x_s = \pi(\mathbf{U}_s)$. Proposition 9.4 establishes that $\mathbf{U}_s$ constitutes the genuine time-reversal process of $\mathbf{U}_t$, and further demonstrates that the associated probability density admits a joint representation in terms of both the latent process $\mathbf{E}$ and the data process $\mathbf{U}$. Corollary 9.6 subsequently characterizes the resulting data density on the manifold, showing how it emerges under canonical projection from the frame bundle. Finally, the horizontal score field s^{H} is gauge-equivariant by Proposition 9.2: for every $h \in O(d)$,

$$
s^{\mathrm{H}}(t, R_h(\mathbf{U})) = dR_h\big(s^{\mathrm{H}}(t,\mathbf{U})\big), \qquad (31)
$$

and for a smooth gauge transformation $g : \mathcal{O}(\mathcal{M}) \to O(d)$ the score expressed in the transformed gauge remains horizontal with respect to ω^g (Step 4 of the same proof). This ensures that our construction is consistent with the natural $O(d)$-equivariance structure of the frame bundle geometry.

9.2 Training Horizontal Diffusion Models

Proposition 9.8 (Horizontal KL Divergence)**.** *Let $b_t(\mathbf{U}_t)$ and $\tilde{b}_t(\mathbf{U}_t^\theta)$ be horizontal drifts on the frame bundle $\mathcal{O}(\mathcal{M})$ with path laws $\boldsymbol{\nu}_b$ and $\boldsymbol{\nu}_{\tilde{b}}$, and let $\mathbf{P}_{\mathrm{H}}(Z) := Z - \omega(Z)^\sharp \cdot U$ project[2] to the horizontal subspace $\mathrm{H}_U \subset T_U\mathcal{O}(\mathcal{M})$ via the Levi–Civita connection ω, and $\|\cdot\|_H$ be the Sasaki metric norm on H_U, $\Gamma^{\mathcal{O}(\mathcal{M})}$ be parallel transport. Then, the horizontal KL divergence is given by*

$$\mathbf{KL}_{\mathrm{H}}[\boldsymbol{\nu}_{\tilde{b}} \| \boldsymbol{\nu}_b] = \frac{1}{2}\mathbb{E}_{\boldsymbol{\nu}_{\tilde{b}}}\left[\int_0^T \|\mathbf{P}_{\mathrm{H}}(\Delta_t)\|_H^2 \, dt\right], \quad \Delta_t := \Gamma^{\mathcal{O}(\mathcal{M})}_{\mathbf{U}_t^\theta \to \mathbf{U}_t}\left(\tilde{b}_t(\mathbf{U}_t^\theta)\right) - b_t(\mathbf{U}_t),$$

Proof. **Step 1 (forward, reverse, and model dynamics).** Fix the latent dynamics of Proposition 9.1 and let $s_\theta^{\mathrm{H}} : [0,T] \times \mathcal{O}(\mathcal{M}) \to \mathrm{H}$ be a learnable horizontal score. The forward SDE, the true reverse SDE of Proposition 9.4, and the learned reverse SDE are

$$d\mathbf{U}_t = -\frac{1}{2}\nabla^{\mathrm{H}}\Phi_t(\mathbf{U}_t)\,dt + \sigma_t \sum_a H_a(\mathbf{U}_t) \circ dW_t^a, \qquad \mathbf{U}_0 \sim \rho_0^{\mathrm{H}}, \tag{32}$$

$$d\mathbf{U}_s = \Big[\frac{1}{2}\nabla^{\mathrm{H}}\Phi_s(\mathbf{U}_s) + \sigma_s^2\, s^{\mathrm{H}}(s,\mathbf{U}_s)\Big]ds + \sigma_s \sum_a H_a(\mathbf{U}_s) \circ d\widetilde{W}_s^a, \qquad \mathbf{U}_0 \sim \rho_T^{\mathrm{H}}, \tag{33}$$

$$d\mathbf{U}_s = \Big[\frac{1}{2}\nabla^{\mathrm{H}}\Phi_s(\mathbf{U}_s) + \sigma_s^2\, s_\theta^{\mathrm{H}}(s,\mathbf{U}_s)\Big]ds + \sigma_s \sum_a H_a(\mathbf{U}_s) \circ d\widehat{W}_s^a, \qquad \mathbf{U}_0 \sim \rho_T^{\mathrm{H}},$$

with path laws $\nu_{[0,T]} := \mathrm{Law}_{\mathbb{P}}(\mathbf{U}_\bullet)$, $\tilde{\nu}_{[0,T]} := \mathrm{Law}_{\tilde{\mathbb{P}}}(\mathbf{U}_\bullet)$, and $\tilde{\nu}^\theta_{[0,T]} := \mathrm{Law}_{\tilde{\mathbb{P}}^\theta}(\mathbf{U}_\bullet)$. The three SDEs share the diffusion fields $\sigma\, H_a$ and hence the Itô–Stratonovich correction $\frac{1}{2}\sigma^2\mathcal{V}$, with $\mathcal{V} = \sum_a (H_a H_a^\alpha)\partial_\alpha$ computed in local coordinates in Step 1 of the proof of Proposition 9.4:

$$d\mathbf{U}_t = \begin{cases} \Big(-\frac{1}{2}\nabla^{\mathrm{H}}\Phi_t + \frac{1}{2}\sigma_t^2\mathcal{V}\Big)(\mathbf{U}_t)\,dt + \sigma_t \sum_a H_a(\mathbf{U}_t)\, dW_t^a, & \text{(forward)}, \\ \Big(\frac{1}{2}\nabla^{\mathrm{H}}\Phi_t + \sigma_t^2 s^{\mathrm{H}}(t,\cdot) + \frac{1}{2}\sigma_t^2\mathcal{V}\Big)(\mathbf{U}_t)\,dt + \sigma_t \sum_a H_a(\mathbf{U}_t)\, d\widetilde{W}_t^a, & \text{(reverse)}, \\ \Big(\frac{1}{2}\nabla^{\mathrm{H}}\Phi_t + \sigma_t^2 s_\theta^{\mathrm{H}}(t,\cdot) + \frac{1}{2}\sigma_t^2\mathcal{V}\Big)(\mathbf{U}_t)\,dt + \sigma_t \sum_a H_a(\mathbf{U}_t)\, d\widehat{W}_t^a, & \text{(denoiser)}. \end{cases}$$

The drift of the reverse and denoiser SDEs differ by the horizontal field $\sigma_t^2\big(s_\theta^{\mathrm{H}} - s^{\mathrm{H}}\big)$ in both forms. We first treat two horizontal drifts $b_t, \tilde{b}_t$ driven by the common noise $\sum_a H_a \circ dW^a$ of Definition 9.7, and then insert the score parameterization.

Step 2 (horizontal Girsanov formula). Realize both path laws on the canonical space $C([0,T], \mathcal{O}(\mathcal{M}))$ with coordinate process $\mathbf{U}_\bullet$. The two drifts are then compared along a common path, $\mathbf{U}_t^\theta = \mathbf{U}_t$, for which $\Gamma^{\mathcal{O}(\mathcal{M})}_{\mathbf{U}_t^\theta \to \mathbf{U}_t}$ is the identity of $T_{\mathbf{U}_t}\mathcal{O}(\mathcal{M})$. The discrepancy is horizontal, and its components in the basis $\{H_a\}$ follow from $g_H(V,W) := g\big(\pi_* V, \pi_* W\big)$ on H_U and $g_H(H_a, H_b) = g(e_a, e_b) = \delta_{ab}$:

$$\begin{aligned} \Delta_t &= \tilde{b}_t(\mathbf{U}_t) - b_t(\mathbf{U}_t) \in \mathrm{H}_{\mathbf{U}_t}, \\ \mathbf{P}_{\mathrm{H}}(\Delta_t) &= \Delta_t - \omega(\Delta_t)^\sharp \cdot \mathbf{U}_t = \Delta_t, \\ \Delta_t &= \sum_{a=1}^d \Delta_t^a\, H_a(\mathbf{U}_t), \qquad \Delta_t^a := g_H\big(\Delta_t, H_a(\mathbf{U}_t)\big), \\ \big\|\mathbf{P}_{\mathrm{H}}(\Delta_t)\big\|_H^2 &= \sum_{a,b=1}^d \Delta_t^a \Delta_t^b\, g_H\big(H_a, H_b\big) = \sum_{a=1}^d \big(\Delta_t^a\big)^2. \end{aligned}$$

Under $\boldsymbol{\nu}_b$ the path solves $d\mathbf{U}_t = b_t\,dt + \sum_a H_a \circ dW_t^a$ with a Brownian motion W. Shifting the noise by the discrepancy, $\widetilde{W}_t^a := W_t^a - \int_0^t \Delta_r^a\, dr$, and using that the Stratonovich integral against an

absolutely continuous path is a Lebesgue integral,

$$\begin{aligned} d\mathbf{U}_t &= b_t\,dt + \sum_a H_a \circ d\widetilde{W}_t^a + \sum_a \Delta_t^a H_a(\mathbf{U}_t)\,dt \\ &= \big(b_t + \Delta_t\big)dt + \sum_a H_a \circ d\widetilde{W}_t^a = \tilde{b}_t\,dt + \sum_a H_a \circ d\widetilde{W}_t^a. \end{aligned}$$

Under Novikov's condition $\mathbb{E}\exp\big(\frac{1}{2}\int_0^T \|\Delta_t\|_H^2 dt\big) < \infty$, the exponential

$$\widehat{Z}_T^H := \exp\Big(\sum_{a=1}^d \int_0^T \Delta_t^a\,dW_t^a - \frac{1}{2}\int_0^T \sum_{a=1}^d \big(\Delta_t^a\big)^2 dt\Big)$$

is a martingale density; under $\mathbb{Q} := \widehat{Z}_T^H \cdot \mathbb{P}$ the process $\widetilde{W}$ is a Brownian motion, and the path then has law $\boldsymbol{\nu}_{\tilde{b}}$. Hence $\widehat{Z}_T^H = d\boldsymbol{\nu}_{\tilde{b}}/d\boldsymbol{\nu}_b$, and substituting $dW_t^a = d\widetilde{W}_t^a + \Delta_t^a dt$,

$$\begin{aligned} \mathbf{KL}_{\mathrm{H}}\big[\boldsymbol{\nu}_{\tilde{b}}\|\boldsymbol{\nu}_b\big] &= \mathbb{E}_{\boldsymbol{\nu}_{\tilde{b}}}\Big[\log\frac{d\boldsymbol{\nu}_{\tilde{b}}}{d\boldsymbol{\nu}_b}\Big] = \mathbb{E}_{\mathbb{Q}}\Big[\sum_a \int_0^T \Delta_t^a\,dW_t^a - \frac{1}{2}\int_0^T \sum_a \big(\Delta_t^a\big)^2 dt\Big] \\ &= \mathbb{E}_{\mathbb{Q}}\Big[\sum_a \int_0^T \Delta_t^a\,d\widetilde{W}_t^a + \int_0^T \sum_a \big(\Delta_t^a\big)^2 dt - \frac{1}{2}\int_0^T \sum_a \big(\Delta_t^a\big)^2 dt\Big] \\ &= \underbrace{\mathbb{E}_{\mathbb{Q}}\Big[\sum_a \int_0^T \Delta_t^a\,d\widetilde{W}_t^a\Big]}_{=0} + \frac{1}{2}\,\mathbb{E}_{\mathbb{Q}}\Big[\int_0^T \sum_a \big(\Delta_t^a\big)^2 dt\Big] = \frac{1}{2}\,\mathbb{E}_{\boldsymbol{\nu}_{\tilde{b}}}\Big[\int_0^T \big\|\mathbf{P}_{\mathrm{H}}(\Delta_t)\big\|_H^2 dt\Big], \end{aligned}$$

which is the statement. With the diffusion fields $\sigma_t H_a$ of Step 1, the same computation applies to $\sigma_t^{-1}\Delta_t$ in place of Δ_t.

Step 3 (pull-back to the latent path space). The horizontal SDE admits a unique strong solution for each latent path, and the lifting map Ψ is measurable and almost surely injective for a fixed initial frame, with measurable left inverse $\Psi^{-1}(\mathbf{U}_\bullet) = \big(e_t^{-1}\pi_*\dot{\mathbf{U}}_t\big)_{t\in[0,T]}$. Consequently,

$$\begin{aligned} &\mathcal{F}_t^E = \sigma(E_r : 0 \le r \le t) = \sigma(\mathbf{U}_r : 0 \le r \le t) = \mathcal{F}_t^U, \\ &\mu_{[0,T]} := \mathbb{P}\circ\Psi^{-1}, \qquad \tilde{\mu}_{[0,T]} := \tilde{\mathbb{P}}\circ\Psi^{-1}, \qquad \tilde{\mu}^\theta_{[0,T]} := \tilde{\mathbb{P}}^\theta\circ\Psi^{-1}. \end{aligned}$$

For the reverse and denoiser dynamics, Proposition 9.1 gives the components of the normalized discrepancy as functions of the latent state, $\sigma_t^{-1}\Delta_t^a = \sigma_t\big(s_\theta^a - s^a\big)(t, E_t)$, and the noise increments are latent increments, $\sigma_t\,dW_t^a = dE_t^a - \big[\frac{1}{2}\partial_a\phi_t(E_t) + \sigma_t^2 s^a(t,E_t)\big]dt$. The density $\widehat{Z}_T^H$ is $\mathcal{F}_T^U$-measurable and, evaluated at $\mathbf{U}_\bullet = \Psi(E_\bullet)$,

$$\begin{aligned} \widehat{Z}_T^H\big(\Psi(E_\bullet)\big) &= \exp\Big(\sum_a \int_0^T \sigma_t\big(s_\theta^a - s^a\big)(t,E_t)\,dW_t^a - \frac{1}{2}\int_0^T \sigma_t^2\big\|s_\theta(t,E_t) - s(t,E_t)\big\|^2 dt\Big) \\ &= \frac{d\tilde{\mu}^\theta_{[0,T]}}{d\tilde{\mu}_{[0,T]}}(E_\bullet), \end{aligned}$$

which is the Euclidean Girsanov density between the latent SDEs with drifts $\frac{1}{2}\nabla\phi_t + \sigma_t^2 s$ and $\frac{1}{2}\nabla\phi_t + \sigma_t^2 s_\theta$. The horizontal density on $C([0,T],\mathcal{O}(\mathcal{M}))$ and the latent density on $C([0,T],\mathbb{R}^d)$ therefore coincide through Ψ.

Step 4 (Euclidean score-matching form). By Proposition 9.1, the true and learnable horizontal scores expand as

$$s^{\mathrm{H}}(t,U) = \sum_{a=1}^d s^a(t,\mathbf{E}_t)H_a(U), \qquad s_\theta^{\mathrm{H}}(t,U) = \sum_{a=1}^d s_\theta^a(t,\mathbf{E}_t)H_a(U).$$

The discrepancy between the denoiser and reverse drifts and its normalized squared norm are

$$\begin{aligned}\Delta_t &= \Big[\frac{1}{2}\nabla^{\mathrm{H}}\Phi_t + \sigma_t^2 s_\theta^{\mathrm{H}}(t,\cdot)\Big](\mathbf{U}_t) - \Big[\frac{1}{2}\nabla^{\mathrm{H}}\Phi_t + \sigma_t^2 s^{\mathrm{H}}(t,\cdot)\Big](\mathbf{U}_t)\\ &= \sigma_t^2 \sum_{a=1}^{d} \big(s_\theta^a(t,\mathbf{E}_t) - s^a(t,\mathbf{E}_t)\big) H_a(\mathbf{U}_t),\\ \big\|\sigma_t^{-1}\mathbf{P}_{\mathrm{H}}(\Delta_t)\big\|_H^2 &= \sigma_t^{-2}\,\sigma_t^4 \sum_{a,b=1}^{d} \big(s_\theta^a - s^a\big)\big(s_\theta^b - s^b\big)(t,\mathbf{E}_t)\, g_H\big(H_a, H_b\big)\\ &= \sigma_t^2 \sum_{a=1}^{d} \big|s_\theta^a(t,\mathbf{E}_t) - s^a(t,\mathbf{E}_t)\big|^2.\end{aligned}$$

Step 2 with the roles of the two drifts exchanged, the re-indexing $t = T - s$ under which the true reverse path has marginals ρ_t^{H}, and $\mathbf{E}_t \sim \rho_t$ (Step 3 of the proof of Proposition 9.1) give

$$\begin{aligned}\mathbf{KL}_{\mathrm{H}}\big(\tilde\nu_{[0,T]} \| \tilde\nu^\theta_{[0,T]}\big) &= \frac{1}{2}\,\mathbb{E}_{\tilde\nu_{[0,T]}}\Big[\int_0^T \big\|\sigma_t^{-1}\mathbf{P}_{\mathrm{H}}(\Delta_t)\big\|_H^2 dt\Big]\\ &= \frac{1}{2}\int_0^T \sigma_t^2\,\mathbb{E}_{\mathbf{E}_t\sim\rho_t}\Big[\big\|s(t,\mathbf{E}_t) - s_\theta(t,\mathbf{E}_t)\big\|^2\Big] dt,\end{aligned}$$

which is the objective (8); the direction $\mathbf{KL}_{\mathrm{H}}(\tilde\nu^\theta_{[0,T]} \| \tilde\nu_{[0,T]})$ has the same integrand with the expectation taken under $\tilde\nu^\theta_{[0,T]}$. The weight σ_t^2 is the diffusion strength at time t, and the loss is invariant to the choice of local coordinates because it depends on the frame only through the orthonormality $g_H(H_a, H_b) = \delta_{ab}$. □

Theorem 9.9 (Uniform Generalization Bound for Worst-Case Horizontal KL)**.** *Assume the neural score networks are parameterized as L-layer, width W feedforward ReLU networks with spectral norm bounds s_ℓ, (2,1)-norm bounds b_ℓ, activation Lipschitz constants ρ_ℓ. Assume the manifold $\mathcal{M}$ has sectional curvature $|K(\pi)| \le \kappa_{\max}$ and diameter $\mathrm{Diam}(\mathcal{M})$; noise scales σ_{t_m} are bounded below by $\sigma_{\min}$. Let n be the number of samples per time, M the number of time steps, D the maximal function class diameter, and $\widehat{\mathbf{KL}}_{\mathrm{H}}(\theta)$ the empirical counterpart of $\mathbf{KL}_{\mathrm{H}}(\theta)$ in* (8) *computed from the nM training samples. Then for any $\delta \in (0,1)$, with probability at least $1-\delta$, the following holds:*

$$\sup_{\theta\in\Theta}\Big\{\mathbf{KL}_{\mathrm{H}}(\theta) - \widehat{\mathbf{KL}}_{\mathrm{H}}(\theta)\Big\} \;\le\; \frac{6C_{\mathcal{M}}(\sigma_{\max},\kappa_{\max},\mathrm{Diam}(\mathcal{M}),D)\cdot C_{\mathrm{NN}}(L,W,s_\ell,b_\ell,\rho_\ell,\bar D)}{n^{1/4}} + \sqrt{\frac{C_{\mathrm{SubG}}^2(\sigma_{t_m}, L_{t_m}^{\mathrm{tot}}, \lambda_{\max}(\Sigma_{t_m}))\log(1/\delta)}{4nM}}, \quad (34)$$

where Eq (35) *provides full description of constants $C_{\mathcal{M}}$, C_{NN} and C_{SubG}.*

Remark 9.10. The result states for the case where the neural network score models are drawn from the classical family of L-layer, width-W feedforward ReLU networks with uniformly bounded spectral norms and (2,1)-norms, for which explicit covering number bounds are available via classical results Bartlett et al. [2017]. However, the generalization argument and the overall structure of the proof are not tied to these specific architectural assumptions. In particular, the identical analysis applies for any hypothesis class $\mathfrak{F}$ whose empirical covering numbers can be suitably bounded. The key requirement is the ability to upper-bound the entropy number of the hypothesis class with respect to the empirical L_2 metric induced by the data distribution.

Proof. **Step 1 (function classes and risks).** Let $t_m = mT/M$, $m = 1,\dots,M$, be the time grid and $E_{t_m}^{(i)} \sim \rho_{t_m} = \mathcal{N}(0,\Sigma_{t_m})$, $i = 1,\dots,n$, independent samples, with network inputs restricted to the

ball $\{\|E\| \le \bar{D}\}$. The scaled score errors, their horizontal lifts, and the bounds used below are

$$\mathfrak{F} := \Big\{ f_{\theta,m}(E) = \sigma_{t_m}\big[s_\theta(t_m,E) - s(t_m,E)\big] : \theta \in \Theta,\ m = 1,\dots,M \Big\},$$
$$\mathfrak{G} := \Big\{ g_{\theta,m}(E) = \big\|H_{e_{t_m}}\big[f_{\theta,m}(E)\big]\big\|_H^2 : f_{\theta,m} \in \mathfrak{F} \Big\},$$
$$\bar{C}_f := \sup_{\theta,m,\,\|E\|\le\bar{D}} \big\|f_{\theta,m}(E)\big\|, \qquad \sigma_{\max} := \max_m \sigma_{t_m}.$$

The population and empirical risks of the lifted class, and the time-discretized objective (8) with its empirical counterpart, are

$$\mathcal{R}^{\mathrm{H}}(\theta) = \frac{T}{M}\sum_{m=1}^{M} \mathbb{E}_{E_{t_m}\sim\rho_{t_m}}\big[g_{\theta,m}(E_{t_m})\big], \qquad \widehat{\mathcal{R}}_n^{\mathrm{H}}(\theta) = \frac{T}{nM}\sum_{m=1}^{M}\sum_{i=1}^{n} g_{\theta,m}\big(E_{t_m}^{(i)}\big),$$
$$\mathbf{KL}_{\mathrm{H}}(\theta) = \frac{1}{2}\cdot\frac{T}{M}\sum_{m=1}^{M}\sigma_{t_m}^2\,\mathbb{E}\big\|s(t_m,E_{t_m}) - s_\theta(t_m,E_{t_m})\big\|^2 = \frac{1}{2}\,\mathcal{R}^{\mathrm{H}}(\theta),$$
$$\widehat{\mathbf{KL}}_{\mathrm{H}}(\theta) = \frac{1}{2}\,\widehat{\mathcal{R}}_n^{\mathrm{H}}(\theta),$$

where $\mathbf{KL}_{\mathrm{H}}(\theta) = \frac{1}{2}\mathcal{R}^{\mathrm{H}}(\theta)$ uses the Sasaki isometry $\|H_e[w]\|_H = \|w\|$ (Step 4). The quantity to control is therefore

$$\sup_{\theta\in\Theta}\Big\{\mathbf{KL}_{\mathrm{H}}(\theta) - \widehat{\mathbf{KL}}_{\mathrm{H}}(\theta)\Big\} = \frac{1}{2}\sup_{\theta\in\Theta}\Big\{\mathcal{R}^{\mathrm{H}}(\theta) - \widehat{\mathcal{R}}_n^{\mathrm{H}}(\theta)\Big\} =: \frac{1}{2}\,\mathcal{Z}.$$

Step 2 (symmetrization). Let $(E_{t_m}^{(i)\prime})$ be an independent ghost sample and (ε_{mi}) i.i.d. Rademacher variables. Jensen's inequality, the exchangeability of $E_{t_m}^{(i)}$ and $E_{t_m}^{(i)\prime}$, and the symmetry $-\varepsilon_{mi} \overset{d}{=} \varepsilon_{mi}$ give

$$\begin{aligned}
\mathbb{E}\,\mathcal{Z} &= \mathbb{E}\sup_\theta\Big\{\frac{T}{nM}\sum_{m,i}\Big(\mathbb{E}'\big[g_{\theta,m}(E_{t_m}^{(i)\prime})\big] - g_{\theta,m}(E_{t_m}^{(i)})\Big)\Big\} \\
&\le \mathbb{E}\sup_\theta\Big\{\frac{T}{nM}\sum_{m,i}\Big(g_{\theta,m}(E_{t_m}^{(i)\prime}) - g_{\theta,m}(E_{t_m}^{(i)})\Big)\Big\} \\
&= \mathbb{E}\sup_\theta\Big\{\frac{T}{nM}\sum_{m,i}\varepsilon_{mi}\Big(g_{\theta,m}(E_{t_m}^{(i)\prime}) - g_{\theta,m}(E_{t_m}^{(i)})\Big)\Big\} \le 2T\,\mathbb{E}\,\mathfrak{R}_{n,M},
\end{aligned}$$
$$\mathfrak{R}_{n,M} := \mathbb{E}_\varepsilon \sup_\theta \frac{1}{nM}\sum_{m=1}^{M}\sum_{i=1}^{n}\varepsilon_{mi}\,g_{\theta,m}\big(E_{t_m}^{(i)}\big).$$

Step 3 (concentration). Assume $s_\theta(t_m,\cdot)$ is $L_{t_m}(\theta)$-Lipschitz, and let $h_{\theta,m} := s_\theta(t_m,\cdot) - s(t_m,\cdot)$ with the Gaussian score $s(t_m,E) = -\Sigma_{t_m}^{-1}E$ of the forward marginal. For $\|x\|, \|y\| \le \bar{D}$,

$$\begin{aligned}
L_{t_m}^{\mathrm{tot}} &:= \sup_\theta L_{t_m}(\theta) + \|\Sigma_{t_m}^{-1}\|, \\
\big\|h_{\theta,m}(x) - h_{\theta,m}(y)\big\| &\le \big\|s_\theta(t_m,x) - s_\theta(t_m,y)\big\| + \big\|\Sigma_{t_m}^{-1}(x-y)\big\| \\
&\le \big(L_{t_m}(\theta) + \|\Sigma_{t_m}^{-1}\|\big)\|x-y\| \le L_{t_m}^{\mathrm{tot}}\|x-y\|, \\
\big|g_{\theta,m}(x) - g_{\theta,m}(y)\big| &= \Big|\big\|f_{\theta,m}(x)\big\|^2 - \big\|f_{\theta,m}(y)\big\|^2\Big| \\
&= \Big|\big\langle f_{\theta,m}(x) - f_{\theta,m}(y),\ f_{\theta,m}(x) + f_{\theta,m}(y)\big\rangle\Big| \\
&\le \big\|f_{\theta,m}(x) - f_{\theta,m}(y)\big\|\Big(\big\|f_{\theta,m}(x)\big\| + \big\|f_{\theta,m}(y)\big\|\Big) \\
&\le 2\bar{C}_f\,\sigma_{t_m}L_{t_m}^{\mathrm{tot}}\,\|x-y\|,
\end{aligned}$$

and composing with the 1-Lipschitz radial projection onto $\{\|E\| \le \bar{D}\}$ extends the bound to $\mathbb{R}^d$. Write $E_{t_m}^{(i)} = \Sigma_{t_m}^{1/2}\xi_m^{(i)}$ with a standard Gaussian vector $\xi = (\xi_m^{(i)}) \in \mathbb{R}^{dnM}$. For two realizations

ξ, ξ', the Cauchy–Schwarz inequality gives

$$\big|\mathcal{Z}(\xi) - \mathcal{Z}(\xi')\big| \le \sup_\theta \big|\widehat{\mathcal{R}}_n^{\mathrm{H}}(\theta;\xi) - \widehat{\mathcal{R}}_n^{\mathrm{H}}(\theta;\xi')\big| \le \frac{T}{nM} \sum_{m,i} 2\bar{C}_f\, \sigma_{t_m} L_{t_m}^{\mathrm{tot}} \sqrt{\lambda_{\max}(\Sigma_{t_m})}\, \big\|\xi_m^{(i)} - \xi_m^{(i)\prime}\big\|$$

$$\le \frac{T}{\sqrt{nM}} \max_m \Big(2\bar{C}_f\, \sigma_{t_m} L_{t_m}^{\mathrm{tot}} \sqrt{\lambda_{\max}(\Sigma_{t_m})}\Big)\, \|\xi - \xi'\|.$$

Gaussian concentration for Lipschitz functions [Vershynin, 2018] yields, with probability at least $1-\delta$,

$$\mathcal{Z} \le \mathbb{E}\,\mathcal{Z} + T \max_m \Big(2\bar{C}_f\, \sigma_{t_m} L_{t_m}^{\mathrm{tot}} \sqrt{\lambda_{\max}(\Sigma_{t_m})}\Big) \sqrt{\frac{2\log(1/\delta)}{nM}} = \mathbb{E}\,\mathcal{Z} + C_{\mathrm{SubG}} \sqrt{\frac{\log(1/\delta)}{nM}}.$$

Step 4 (geometry of the horizontal lift). For $U = (x, e)$ and $w \in \mathbb{R}^d$, the Sasaki norm of the lift is

$$\big\|H_e[w]\big\|_H^2 = g_H\big(H_e[w], H_e[w]\big) = g_x\big(ew, ew\big) = \sum_{a,c} w^a w^c\, g_x(e_a, e_c) = \|w\|^2.$$

The $(d + d^2)$-dimensional chart representation of the same vector (Algorithm 4) also carries the frame velocity $\dot{e}_b^k = -\Gamma_{ij}^k(x) e_a^i w^a e_b^j$. Measuring both blocks with g and writing $v = ew$,

$$\big\|H_e[w]\big\|_{\mathrm{chart}}^2 := g_{pq}\dot{x}^p\dot{x}^q + \sum_{b=1}^d g_{pq}\dot{e}_b^p\dot{e}_b^q = g_{pq}e_a^p e_c^q w^a w^c + \sum_{b=1}^d g_{pq}\Gamma_{ij}^p\Gamma_{mn}^q\, v^i e_b^j\, v^m e_b^n$$

$$= \|w\|^2 + \sum_{b=1}^d \big\|\Gamma_x(v, e_b)\big\|_g^2.$$

In normal coordinates centered at x_0 with $r = d(x, x_0) = |x|$, the curvature expansion $\Gamma_{ij}^k(x) = -\frac{1}{3}\big(R_{ijl}^k + R_{jil}^k\big)(x_0)\, x^l + \mathcal{O}(r^2)$ and the curvature bound $\|R(X,Y)Z\|_g \le \kappa_{\max}\|X\|\|Y\|\|Z\|$ give

$$\Gamma_x(v, u) = -\frac{1}{3}\Big[R(u, x)v + R(v, x)u\Big] + \mathcal{O}(r^2),$$

$$\big\|\Gamma_x(v, e_b)\big\|_g \le \frac{1}{3}\Big(\big\|R(e_b, x)v\big\|_g + \big\|R(v, x)e_b\big\|_g\Big) + \mathcal{O}(r^2) \le \frac{2}{3}\,\kappa_{\max}\, r\, \|v\|_g + \mathcal{O}(r^2),$$

$$\sum_{b=1}^d \big\|\Gamma_x(v, e_b)\big\|_g^2 \le \frac{4d}{9}\,\kappa_{\max}^2\, r^2\, \|v\|_g^2 + \mathcal{O}(r^3) = \frac{4d}{9}\,\kappa_{\max}^2\, r^2\, \|w\|^2 + \mathcal{O}(r^3).$$

To leading order in $r \le \mathrm{Diam}(\mathcal{M})$, both representations are quadratic forms $w^\top Q_e\, w$ with $Q_e = I$ for the Sasaki norm and

$$\|w\|^2 = \big\|H_e[w]\big\|_H^2 \le \big\|H_e[w]\big\|_{\mathrm{chart}}^2 \le \Big(1 + \frac{4d}{9}\kappa_{\max}^2 \mathrm{Diam}^2(\mathcal{M})\Big)\|w\|^2,$$

$$\|Q_e\| \le 1 + \frac{4d}{9}\kappa_{\max}^2 \mathrm{Diam}^2(\mathcal{M}).$$

For $f, \tilde{f}$ with $\|f\|, \|\tilde{f}\| \le \bar{C}_f$, the map $f \mapsto \|H_e[f]\|^2$ is therefore Lipschitz with a curvature-dependent constant, which we use for both representations:

$$\Big|\big\|H_e[f]\big\|^2 - \big\|H_e[\tilde{f}]\big\|^2\Big| = \big|(f - \tilde{f})^\top Q_e\, (f + \tilde{f})\big| \le \|Q_e\|\, \|f - \tilde{f}\|\big(\|f\| + \|\tilde{f}\|\big)$$

$$\le 2\bar{C}_f\Big(1 + \frac{4d}{9}\kappa_{\max}^2 \mathrm{Diam}^2(\mathcal{M})\Big)\|f - \tilde{f}\|.$$

Step 5 (covering numbers). Let $d_{P_{nM}}$ and $d'_{P_{nM}}$ be the empirical L_2-metrics on $\mathfrak{F}$ and $\mathfrak{G}$ over the joint sample,

$$d_{P_{nM}}(f, \tilde{f}) := \Big(\frac{1}{nM}\sum_{m=1}^M \sum_{i=1}^n \big\|f(t_m, E_{t_m}^{(i)}) - \tilde{f}(t_m, E_{t_m}^{(i)})\big\|^2\Big)^{1/2},$$

$$d'_{P_{nM}}(g, \tilde{g}) := \Big(\frac{1}{nM}\sum_{m=1}^M \sum_{i=1}^n \big|g(t_m, E_{t_m}^{(i)}) - \tilde{g}(t_m, E_{t_m}^{(i)})\big|^2\Big)^{1/2}.$$

The score network is $F_\theta(t,E) := W_L\,\sigma(W_{L-1}\,\sigma(\cdots\sigma(W_1 z)\cdots))$ acting on the input $z=(\tau(t),E)$ with $\|z\|\le\bar D$, ReLU activation, widths at most W, $\|W_\ell\|_{\mathrm{op}}\le s_\ell$ and $\|W_\ell^\top\|_{2,1}\le b_\ell$, and $f_{\theta,m}=\sigma_{t_m}\big[F_\theta(t_m,\cdot)-s(t_m,\cdot)\big]$ with a common θ across time. Pointwise application of Step 4, the fixed shift $s(t_m,\cdot)$, and the spectral covering bound of Bartlett et al. [2017] give the chain

$$\begin{aligned}
d'_{P_{nM}}\big(g_\theta,g_{\theta'}\big) &\le 2\bar C_f\Big(1+\frac{4d}{9}\kappa_{\max}^2\mathrm{Diam}^2(\mathcal{M})\Big)\,d_{P_{nM}}\big(f_\theta,f_{\theta'}\big)\\
&\le 2\bar C_f\,\sigma_{\max}\Big(1+\frac{4d}{9}\kappa_{\max}^2\mathrm{Diam}^2(\mathcal{M})\Big)\,d_{P_{nM}}\big(F_\theta,F_{\theta'}\big),\\
\log\mathcal{N}\big(\mathfrak{G},d'_{P_{nM}},\eta\big) &\le \log\mathcal{N}\Big(\{F_\theta\},d_{P_{nM}},\frac{\eta}{2\bar C_f\,\sigma_{\max}\big(1+\frac{4d}{9}\kappa_{\max}^2\mathrm{Diam}^2(\mathcal{M})\big)}\Big)\\
&\le \frac{\Big[2\bar C_f\,\sigma_{\max}\big(1+\frac{4d}{9}\kappa_{\max}^2\mathrm{Diam}^2(\mathcal{M})\big)C_{\mathrm{NN}}\Big]^2}{\eta^2},
\end{aligned}$$

where $C_{\mathrm{NN}}^2=\bar D^2\ln(2W^2)\big(\prod_\ell s_\ell^2\rho_\ell^2\big)\big(\sum_\ell(b_\ell/s_\ell)^{2/3}\big)^3$ is the constant of Bartlett et al. [2017] in the normalized metric.

Step 6 (Dudley's entropy integral). Let D be the $d'_{P_{nM}}$-diameter of $\mathfrak{G}$. Dudley's chaining bound with truncation at scale $\alpha>0$ [Bartlett et al., 2017, Lemma A.5], Step 5, and the choice $\alpha=2\bar C_f\sigma_{\max}(1+\frac{4d}{9}\kappa_{\max}^2\mathrm{Diam}^2(\mathcal{M}))C_{\mathrm{NN}}/\sqrt{nM}$ give

$$\begin{aligned}
\mathfrak{R}_{n,M} &\le \inf_{\alpha>0}\Big\{4\alpha+\frac{12}{\sqrt{nM}}\int_\alpha^D\sqrt{\log\mathcal{N}\big(\mathfrak{G},d'_{P_{nM}},\eta\big)}\,d\eta\Big\}\\
&\le \inf_{\alpha>0}\Big\{4\alpha+\frac{12}{\sqrt{nM}}\,2\bar C_f\sigma_{\max}\Big(1+\frac{4d}{9}\kappa_{\max}^2\mathrm{Diam}^2(\mathcal{M})\Big)C_{\mathrm{NN}}\,\log^+\frac{D}{\alpha}\Big\}\\
&\le \frac{2\bar C_f\sigma_{\max}\big(1+\frac{4d}{9}\kappa_{\max}^2\mathrm{Diam}^2(\mathcal{M})\big)C_{\mathrm{NN}}}{\sqrt{nM}}\\
&\qquad\times\Big[4+12\log^+\frac{D}{2\bar C_f\sigma_{\max}\big(1+\frac{4d}{9}\kappa_{\max}^2\mathrm{Diam}^2(\mathcal{M})\big)C_{\mathrm{NN}}}+6\log(nM)\Big].
\end{aligned}$$

With $\log y\le y/e$ applied to $y=(nM)^{1/4}$, $\log(nM)/\sqrt{nM}\le(4/e)(nM)^{-1/4}$, and $(nM)^{-1/2}\le(nM)^{-1/4}\le n^{-1/4}$; hence

$$\begin{aligned}
T\,\mathfrak{R}_{n,M} &\le \frac{2T\bar C_f\sigma_{\max}\big(1+\frac{4d}{9}\kappa_{\max}^2\mathrm{Diam}^2(\mathcal{M})\big)C_{\mathrm{NN}}}{n^{1/4}}\\
&\qquad\times\Big[4+\frac{24}{e}+12\log^+\frac{D}{2\bar C_f\sigma_{\max}\big(1+\frac{4d}{9}\kappa_{\max}^2\mathrm{Diam}^2(\mathcal{M})\big)C_{\mathrm{NN}}}\Big]=\frac{6\,C_{\mathcal{M}}\,C_{\mathrm{NN}}}{n^{1/4}}.
\end{aligned}$$

Step 7 (collecting the bounds). Steps 1–3 and 6 give, with probability at least $1-\delta$,

$$\begin{aligned}
\sup_{\theta\in\Theta}\Big\{\mathbf{KL}_{\mathrm{H}}(\theta)-\widehat{\mathbf{KL}}_{\mathrm{H}}(\theta)\Big\}=\frac12\,\mathcal{Z}&\le\frac12\,\mathbb{E}\,\mathcal{Z}+\frac12\,C_{\mathrm{SubG}}\sqrt{\frac{\log(1/\delta)}{nM}}\\
&\le T\,\mathbb{E}\,\mathfrak{R}_{n,M}+\sqrt{\frac{C_{\mathrm{SubG}}^2\log(1/\delta)}{4nM}}\\
&\le\frac{6\,C_{\mathcal{M}}\,C_{\mathrm{NN}}}{n^{1/4}}+\sqrt{\frac{C_{\mathrm{SubG}}^2\log(1/\delta)}{4nM}},
\end{aligned}$$

and consequently $\mathbf{KL}_{\mathrm{H}}(\theta)\le\widehat{\mathbf{KL}}_{\mathrm{H}}(\theta)+6C_{\mathcal{M}}C_{\mathrm{NN}}n^{-1/4}+\sqrt{C_{\mathrm{SubG}}^2\log(1/\delta)/(4nM)}$ uniformly over Θ. The three constants C_{NN}, $C_{\mathcal{M}}$, and C_{SubG} reflect the contribution of network expressivity,

manifold geometry, and stochastic noise, respectively:

$$C_{\mathrm{NN}}(L,W) \coloneqq \left[\bar{D}^2 \ln(2W^2) \left(\prod_{\ell=1}^{L} s_\ell^2 \rho_\ell^2\right) \left(\sum_{\ell=1}^{L} \left(\frac{b_\ell}{s_\ell}\right)^{2/3}\right)^3\right]^{1/2}, \tag{35}$$

$$C_{\mathcal{M}}(\sigma,\kappa,\mathrm{Diam}(\mathcal{M})) \coloneqq \frac{T\bar{C}_f\,\sigma_{\max}}{3}\Big(1+\frac{4d}{9}\kappa_{\max}^2 \mathrm{Diam}^2(\mathcal{M})\Big) \times \Big[4+\frac{24}{e}+12\log^+ \frac{D}{2\bar{C}_f\sigma_{\max}\big(1+\frac{4d}{9}\kappa_{\max}^2\mathrm{Diam}^2(\mathcal{M})\big)C_{\mathrm{NN}}}\Big], \tag{36}$$

$$C_{\mathrm{SubG}}(\sigma_t,\Sigma) \coloneqq 2\sqrt{2}\,T\,\bar{C}_f \max_m \Big(\sigma_{t_m} L_{t_m}^{\mathrm{tot}} \sqrt{\lambda_{\max}(\Sigma_{t_m})}\Big). \tag{37}$$

□

10 Implementation Details

In this section, we provide comprehensive details on the implementation of our proposed horizontal diffusion model. We first specify the choice of hyperparameters used throughout the experiments. Subsequently, we formally define the construction of synthetic benchmark datasets, encompassing a suite of analytically parameterized surfaces that serve as controlled environments for evaluating geometric generative modeling. Finally, we describe the real-world datasets, including celestial surfaces and protein structures, modeled via spherical harmonics parameterizations and sourced from publicly available scientific repositories. All implementation specifics, including the design of neural architectures, training protocols, and evaluation metrics, are reported to ensure reproducibility and facilitate future extensions of our approach.

10.1 Hyperparameters

Recall that the proposed (reverse) horizontal diffusion model for the Ornstein-Uhlenbeck (OU) bridge process on the orthonormal frame bundle is given by

$$dU_s = -\sum_a \left(2\gamma_s \epsilon^a + \sigma_s^2 v_s^a\right) H_a(U_s) ds + s^{\mathrm{H}}(s, U_s) ds + \sum_a \sigma_s H_a(U_s) \circ dW_s^a,$$

It depends on several time-dependent hyperparameters that critically determine both the geometry-adaptivity and empirical behavior of the model. Here, we detail the practical design and rationale for each hyperparameter as implemented in our framework.

Damping Coefficient γ_s. The role of this coefficient is to govern the mean-reverting drift of the latent Euclidean process, corresponding to the dissipative force in the OU process. We use a constant value, $\gamma_s \equiv \gamma$, for all s, as this ensures numerical stability and avoids extreme drift behaviors.

Diffusion Scale Σ_s, σ_s. The diffusion parameter σ_s scales noise intensity of both forward and reverse OU process in the latent space. It is designed as $\sigma_s^2 = \gamma \Sigma_s$ so that we have $\gamma I = \sigma_s^2 \Sigma_s^{-1}$. The other parameter is exponentially controlled by using the formula $\Sigma_s = \Sigma_0 \exp(-\gamma s)$, with σ_s computed at each time step.

Bridge Intensity λ_s. The role of this coefficient is to control the strength of the attraction term v_s, driving the process toward the data manifold during reverse diffusion. In this paper, several scheduling options are considered as listed in following: (1) Exponential decay: $\lambda_s = \lambda_{\max} e^{-\alpha s}$, (2) Polynomial decay: $\lambda_s = \lambda_{\max}(1 - s/T)^k$, (3) Sigmoid decay: $\lambda_s = \lambda_{\max}/(1 + e^{\alpha(s - T/2)})$. In the experiment on sphere and torus, we used polynomial decay intensity function, and used exponential decay for other setups.

10.2 Synthetic Dataset

Enner Surface. This first model space is a complete, orientable *minimal surface* in $\mathbb{R}^3$. It is most conveniently introduced through the polynomial immersion defined as follows:

$$f(u, v) = \left(u - \tfrac{u^3}{3} + uv^2, v - \tfrac{v^3}{3} + vu^2, u^2 - v^2\right), \qquad (u, v) \in \mathbb{R}^2.$$

In order to induce gradient fields, we apply differentiation $\mathbf{X}$ with respect to its parameters and yields the followings:

$$\partial_u f = \left(1 - u^2 + v^2, 2uv, 2u\right), \quad \partial_v f = \left(2uv, 1 - v^2 + u^2, -2v\right).$$

Taking inner products produces the first fundamental form

$$g_{11} = g_{22} = (1 + r^2)^2, \quad g_{12} = 0, \qquad r^2 = u^2 + v^2.$$

In other means, the induced metric $g = (1 + r^2)^2(du^2 + dv^2)$ is *conformal* to the Euclidean metric and we have $\det g = (1 + r^2)^4$. Taking derivatives on both component of Riemannian metric, we have

$$\partial_u g_{11} = 4u(1 + r^2), \quad \partial_v g_{11} = 4v(1 + r^2), \qquad \partial_u g_{22} = 4u(1 + r^2), \quad \partial_v g_{22} = 4v(1 + r^2).$$

Substituting these results into the Levi–Civita formula

$$\Gamma^i_{jk} = \frac{1}{2} g^{i\ell} \big(\partial_j g_{k\ell} + \partial_k g_{j\ell} - \partial_\ell g_{jk} \big)$$

gives the non–zero components of Christoffel symbols:

$$\Gamma^u{}_{uu} = \Gamma^v{}_{uv} = \Gamma^u{}_{uv} = \frac{2u}{1+r^2},$$

$$\Gamma^v{}_{vv} = \Gamma^u{}_{uv} = \Gamma^v{}_{vu} = \frac{2v}{1+r^2},$$

$$\Gamma^u{}_{vv} = -\frac{2u}{1+r^2}, \qquad \Gamma^v{}_{uu} = -\frac{2v}{1+r^2}.$$

Catenoid Surface. This second model space is another complete, orientable *minimal surface* embedded in $\mathbb{R}^3$. A convenient immersion of the catenoid can be explicitly given by the parameterization

$$f(u,v) = (\cosh v \cos u, \cosh v \sin u, v), \quad \text{for} \quad (u,v) \in [0, 2\pi) \times \mathbb{R}. \tag{38}$$

To analyze gradient fields and geometric properties, we compute the partial derivatives of the immersion f with respect to the local parameters u and v, obtaining $\partial_u f = (-\cosh v \sin u, \cosh v \cos u, 0)$ and $\partial_v f = (\sinh v \cos u, \sinh v \sin u, 1)$. Taking the inner products of these tangent vectors yields the components of the induced metric tensor, known as the first fundamental form. Explicitly, we have $g_{11} = \langle \partial_u f, \partial_u f \rangle = \cosh^2 v$, $g_{22} = \langle \partial_v f, \partial_v f \rangle = \cosh^2 v + \sinh^2 v = \cosh^2 v$, and $g_{12} = \langle \partial_u f, \partial_v f \rangle = 0$. Hence, the metric tensor is conformally equivalent to the Euclidean metric, and is concisely expressed as $g = \cosh^2 v, (du^2 + dv^2)$, with determinant $\det(g) = \cosh^4 v$.

The Christoffel symbols associated with the Levi-Civita connection can be computed from the metric derivatives. Since the metric depends solely on v, we have $\partial_u g_{ij} = 0$ for all (i, j). The non-zero metric derivatives are $\partial_v g_{11} = \partial_v g_{22} = 2 \cosh v \cdot 2 \sinh v = \sinh(2v)$. Using the Levi-Civita formula, $\Gamma^i_{jk} = \frac{1}{2} g^{i\ell} (\partial_j g_{k\ell} + \partial_k g_{j\ell} - \partial_\ell g_{jk})$, and the inverse metric components $g^{11} = g^{22} = \cosh^{-2} v$, we find the non-zero Christoffel symbols explicitly as follows: $\Gamma^u_{uv} = \Gamma^u_{vu} = \tanh v$, $\Gamma^v_{uu} = -\tanh v$, and $\Gamma^v_{vv} = \tanh v$. These Christoffel symbols reflect how tangent vectors are parallel transported across the curved geometry of the catenoid surface.

Torus. This space represents a classical example of a compact, orientable surface with genus one, smoothly embedded in $\mathbb{R}^3$. While the flat torus has been extensively studied in various contexts, generative modeling of the embedded, curved (non-flat) torus remains unexplored. In this study, we focus on this embedded torus as our primary model space, employing the following standard parameterization:

$$f(u,v) = ((R + r\cos v)\cos u, (R + r\cos v)\sin u, r \sin v), \quad \text{with} \quad (u,v) \in [0, 2\pi)^2, \tag{39}$$

where the constants $R > r > 0$ represent the major (central) and minor (tube) radii, respectively. To examine the geometry of this surface, we first compute the tangent vectors derived from the parameterization. The partial derivatives of the embedding are thus explicitly given by

$$\partial_u f = (-(R + r\cos v)\sin u, (R + r\cos v)\cos u, 0), \tag{40}$$

$$\partial_v f = (-r \sin v \cos u, -r \sin v \sin u, r \cos v). \tag{41}$$

Next, we determine the induced metric, or the first fundamental form, by taking inner products of these tangent vectors. The resulting metric tensor components become $g_{11} = \langle \partial_u f, \partial_u f \rangle = (R + r\cos v)^2$, $g_{22} = \langle \partial_v f, \partial_v f \rangle = r^2$, and $g_{12} = \langle \partial_u f, \partial_v f \rangle = 0$, yielding a diagonal metric tensor. Hence, the metric can be succinctly expressed as $g = (R + r\cos v)^2 du^2 + r^2 dv^2$, with determinant given by $\det(g) = r^2(R + r\cos v)^2$. To analyze curvature properties and parallel transport, we compute the derivatives of the metric tensor. Noting the metric's dependence solely on the parameter v, we find $\partial_u g_{ij} = 0$ for all (i, j). Non-zero derivatives are explicitly computed as $\partial_v g_{11} = -2r(R + r\cos v)\sin v$ and $\partial_v g_{22} = 0$.

Using the inverse metric tensor components, $g^{11} = (R + r\cos v)^{-2}$ and $g^{22} = r^{-2}$, we apply the Levi-Civita connection formula, $\Gamma^i_{jk} = \frac{1}{2} g^{i\ell} (\partial_j g_{k\ell} + \partial_k g_{j\ell} - \partial_\ell g_{jk})$, to derive the non-zero

Christoffel symbols explicitly as follows:

$$\Gamma^u_uv = \Gamma^u_vu = -\frac{r\sin v}{R + r\cos v}, \tag{42}$$

$$\Gamma^v_uu = \frac{(R + r\cos v)\sin v}{r}, \tag{43}$$

$$\Gamma^v_vv = 0. \tag{44}$$

These Christoffel symbols characterize how vector fields evolve when parallel transported along curves on the torus surface.

Dupin Cyclide. The Dupin Cyclide is a classical example of a toroidal surface distinguished by its constant principal curvatures, and can be represented as a surface of revolution in $\mathbb{R}^3$. As with the standard torus, the Dupin torus provides a canonical model for studying nontrivial extrinsic geometry in generative modeling on curved surfaces. A standard parametrization is defined as follows:

$$f(u, v) = ((a + b\cos v)\cos u, (a + b\cos v)\sin u, b\sin v), \qquad (u, v) \in [0, 2\pi)^2,$$

where $a > b > 0$ represent the major and minor radii, respectively. This surface is a particular case of a Dupin cyclide. The tangent vectors are:

$$\partial_u f = (-(a + b\cos v)\sin u, (a + b\cos v)\cos u, 0),$$
$$\partial_v f = (-b\sin v\cos u, -b\sin v\sin u, b\cos v).$$

The first fundamental form has components:

$$g_{11} = \langle\partial_u f, \partial_u f\rangle = (a + b\cos v)^2, \qquad g_{22} = \langle\partial_v f, \partial_v f\rangle = b^2, \qquad g_{12} = \langle\partial_u f, \partial_v f\rangle = 0.$$

Thus, the induced Riemannian metric is diagonal and can be presented as follows:

$$g = (a + b\cos v)^2 du^2 + b^2 dv^2, \qquad \det g = (a + b\cos v)^2 b^2.$$

The derivatives of the metric components are:

$$\partial_v g_{11} = -2b(a + b\cos v)\sin v, \qquad \partial_v g_{22} = 0, \qquad \partial_u g_{ij} = 0.$$

The inverse metric is:

$$g^{11} = (a + b\cos v)^{-2}, \qquad g^{22} = b^{-2}.$$

Applying the Levi-Civita connection formula, the nonzero Christoffel symbols are:

$$\Gamma^u_{uv} = \Gamma^u_{vu} = -\frac{b\sin v}{a + b\cos v},$$
$$\Gamma^v_{uu} = \frac{(a + b\cos v)\sin v}{b},$$
$$\Gamma^v_{vv} = 0.$$

Gaussian Mixture Model on Embedded Manifolds. Let $\mathcal{M} = \Phi(U) \subset \mathbb{R}^3$ be a smooth two-dimensional embedded manifold parametrized by $\Phi : U \subset \mathbb{R}^2 \to \mathbb{R}^3$, with local coordinates (u, v). A Gaussian mixture distribution defined on the intrinsic parameter domain U has the form $p_U(u, v) = \sum_{k=1}^K \alpha_k \mathcal{N}((u, v); \mu_k, \Sigma_k)$, with $\alpha_k \geq 0$, $\sum_{k=1}^K \alpha_k = 1$, and $\Sigma_k \succ 0$, which induces a density on the manifold given by $p_{\mathcal{M}}(x) = p_U(u, v)/\sqrt{\det g(u, v)}|_{x=\Phi(u,v)}$, where $g(u, v) = J_\Phi(u, v)^\top J_\Phi(u, v)$ is the induced metric from the Jacobian J_Φ of the parametrization. Sampling from this manifold density involves drawing $(u, v) \sim p_U$ in parameter space and mapping to the manifold via $x = \Phi(u, v)$. Likelihood evaluation and inference tasks similarly involve pulling back points on the manifold to the parameter space coordinates $(u, v) = \Phi^{-1}(x)$. As an illustrative example, consider the catenoid parametrization $f(u, v) = (\cosh v\cos u, \cosh v\sin u, v)$ with $(u, v) \in [0, 2\pi) \times \mathbb{R}$. Its tangent vectors $f_u = (-\cosh v\sin u, \cosh v\cos u, 0)$ and $f_v = (\sinh v\cos u, \sinh v\sin u, 1)$ yield metric coefficients $E = \cosh^2 v$, $F = 0$, and $G = \cosh^2 v$, giving $\det g(u, v) = \cosh^4 v$. Thus, the induced density on the catenoid is explicitly $p_{\mathcal{M}}(x) = p_U(u, v)/\cosh^2 v|_{x=f(u,v)}$, clearly demonstrating how geometric distortion due to embedding is corrected by the induced metric.

10.3 Details on Model Riemannian Manifolds

Kelvin inversion on $\mathbb{R}^3$. For a given point $x = (x_1, x_2, x_3)^\top \in \mathbb{R}^3$, center $c \in \mathbb{R}^3$, and radius $R > 0$, the *Kelvin inversion* $I_{c,R}$ is defined as

$$I_{c,R}(x) = c + \frac{R^2}{\|x - c\|^2}(x - c).$$

This mapping is a global involution on $\mathbb{R}^3 \setminus \{c\}$, satisfying $I_{c,R}^{-1} = I_{c,R}$.

Standard torus parameterization. Consider a standard torus in $\mathbb{R}^3$ with major radius $R_0 > 0$ and minor radius $0 < r < R_0$. The torus can be explicitly parameterized by angles $(\Theta, \Phi) \in [0, 2\pi) \times [0, 2\pi)$ as follows:

$$T(\Theta, \Phi) = \big((R_0 + r\cos\Phi)\cos\Theta, (R_0 + r\cos\Phi)\sin\Theta, r\sin\Phi\big).$$

Conversely, given a point $(x, y, z)^\top$ on this torus, its associated angular coordinates (Θ, Φ) can be uniquely recovered through the inverse relations:

$$\Theta = \operatorname{atan2}(y, x), \qquad \Phi = \operatorname{atan2}\left(z, \sqrt{x^2 + y^2} - R_0\right).$$

Stereographic projection and its inverse. Let $N = (0, 0, 1)^\top$ be the north pole of the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$. The stereographic projection from $\mathbb{S}^2 \setminus \{N\}$ onto the complex plane $\mathbb{C}$ is given by

$$s : \mathbb{S}^2 \setminus \{N\} \to \mathbb{C}, \quad s(y) = \frac{y_1 + iy_2}{1 - y_3}, \quad \text{for} \quad y = (y_1, y_2, y_3) \in \mathbb{S}^2 \setminus \{N\}.$$

Its inverse mapping is explicitly defined as

$$s^{-1} : \mathbb{C} \to \mathbb{S}^2 \setminus \{N\}, \quad s^{-1}(z) = \frac{1}{1 + |z|^2}\big(2\operatorname{Re} z, 2\operatorname{Im} z, |z|^2 - 1\big), \quad z \in \mathbb{C}.$$

This stereographic projection naturally extends to a bijection between the Riemann sphere $\mathbb{C} \cup \{\infty\}$ and the sphere $\mathbb{S}^2$ by identifying the north pole N with the point at infinity:

$$s(N) = \infty, \quad s^{-1}(\infty) = N.$$

Conversion between Catenoid and Sphere. Given new notations and functions described above, let us consider the standard parametrization of the catenoid as described in Eq. 38. Then, we introduce an exponential compactification given by the mapping

$$\psi(u, v) = \alpha e^{v+iu} \in \mathbb{C} \cup \{\infty\}, \quad \text{with scale factor } \alpha > 0.$$

Leveraging this compactification, we construct the following explicit bijection from the cylindrical domain $[0, 2\pi) \times \mathbb{R}$ onto the sphere $\mathbb{S}^2$: $\Phi_{\text{cat}\to S^2}(u, v) = s^{-1}\big(\psi(u, v)\big)$, where $s : \mathbb{S}^2 \to \mathbb{C} \cup \{\infty\}$ denotes the stereographic projection. In particular, under this mapping, the cylindrical coordinates map naturally to spherical coordinates. Moreover, this transformation admits an explicit inverse: For any point $y \in \mathbb{S}^2$, one first computes its stereographic projection $z = s(y)$, after which the original catenoid coordinates are recovered as

$$u = \arg(z), \qquad v = \log(|z|/\alpha).$$

Consequently, any sampling procedures, density functions, or vector fields defined originally in the (u, v)-parametrization of the catenoid can be straightforwardly pulled forward onto $\mathbb{S}^2$ via $\Phi_{\text{cat}\to S^2}$, and similarly pushed back onto the catenoid by means of the analytic inverse mapping presented above. This construction thus facilitates geometric transfer between these distinct spaces, enabling standard methods developed for canonical spherical geometries to be effectively adapted and evaluated within the analytically more complex catenoid setting.

Conversion between Enneper Surface and Torus. The classical Enneper surface admits a polynomial parameterization, which is known to possess intricate self-intersecting geometry, and hence is

generally not globally injective. For practical applicability, it is beneficial to restrict our attention to a suitable, simply connected subdomain

$$D_R = \{(u, v) \in \mathbb{R}^2 : \, u^2 + v^2 < R^2\},$$

with the radius parameter $R > 0$ chosen sufficiently small to moderate, ensuring that the restriction $X_{\text{enn}}|_{D_R}$ is injective over the support of interest. Such a selection of domain radius is essential to avoid self-intersections and ensure well-posedness of the ensuing geometric transformations. Next, we introduce polar coordinates within the domain D_R, defined explicitly by $r = \sqrt{u^2 + v^2}$ and $\theta = \operatorname{atan2}(v, u)$, providing a natural angular and radial parametrization of the Enneper surface domain. To facilitate the mapping onto the torus, we define a strictly increasing bijection $\eta : [0, \infty) \to [0, 1)$ through

$$\eta(r) = \frac{2}{\pi} \arctan(\beta r), \qquad \text{with a scaling parameter } \beta > 0,$$

which possesses the explicit inverse given by $r = \frac{1}{\beta} \tan\big(\frac{\pi}{2}\eta\big)$. This bijection is chosen specifically to achieve a smooth radial rescaling, mapping the infinite radial extent of the Enneper surface onto a finite interval suitable for embedding into the toroidal geometry. Furthermore, angles are mapped onto the standard toroidal coordinates by assigning $\Theta = \theta \bmod 2\pi$, thus ensuring periodic continuity in the angular direction, and $\Phi = 2\pi\eta(r) = 4 \arctan(\beta r)$, thereby embedding the radial direction smoothly onto the torus. Consequently, we define the explicit embedding from the Enneper surface domain to the torus as

$$\Phi_{\text{enn}\to T^2}(u, v) = T(\Theta, \Phi),$$

where T denotes the standard torus parametrization introduced previously. By construction, this mapping is continuous, smoothly varying, and injective when restricted to the chosen domain D_R, thereby establishing a well-defined bijection onto its toroidal image. Explicit recovery of the original coordinates from the toroidal embedding is straightforward: the inverse map explicitly yields $r = \frac{1}{\beta} \tan(\Phi/4)$, directly recovers the angular coordinate via $\theta = \Theta$, and subsequently produces the original planar coordinates by $(u, v) = (r \cos\theta, r \sin\theta)$. If required, the inverse chart of the original polynomial parametrization X_{enn} can then be applied, completing the full cycle of the geometric transformation between the Enneper surface and the torus.

Conversion between Dupin Cyclide and Torus via Inversion. A Dupin cyclide can be analytically represented as an inversion of a standard geometric surface such as a torus, cone, or cylinder, and thus admits a global angle-preserving bijection onto these simpler shapes through Kelvin inversion in $\mathbb{R}^3 \setminus \{c\}$. Given torus parameters (R_0, r) along with an inversion defined by a center $c \in \mathbb{R}^3$ and radius $R > 0$, a Dupin cyclide can be explicitly obtained via the composition $C(\Theta, \Phi) = I_{c,R}\big(T(\Theta, \Phi)\big)$, where $T(\Theta, \Phi)$ is the standard parameterization of the torus. Conversely, given an arbitrary Dupin cyclide C, there exist suitable parameters (c, R) that render the image $Y = I_{c,R}(C)$ into a standard torus configuration. Thus, the explicit transformation $\Phi_{\text{dup}\to T^2}(x)$ defined by $(\Theta, \Phi) = \big(\operatorname{atan2}(Y_2, Y_1), \operatorname{atan2}(Y_3, \sqrt{Y_1^2 + Y_2^2} - R_0)\big)$, with $Y = I_{c,R}(x)$, provides a global bijection from the Dupin cyclide (excluding only the inversion center c) onto the standard torus, subject to identification along the torus seam. Its inverse is explicitly given by $x = I_{c,R}\big(T(\Theta, \Phi)\big)$, allowing full recovery of cyclide coordinates from toroidal parameters.

10.4 Scientific Dataset

Lunar Spherical Harmonic Dataset (LRO/LOLA SHADR). To precisely describe the lunar surface geometry, we employ the spherical harmonic coefficient dataset released by NASA's Lunar Reconnaissance Orbiter (LRO) mission, specifically the Lunar Orbiter Laser Altimeter (LOLA) Spherical Harmonic Data Record (SHADR) Smith et al. [2010], Neumann [2009]. This dataset (*e.g.*, `lro_ltm05_2050_sha.tab`) represents the Moon's reference surface as a truncated expansion in real spherical harmonics:

$$r(\theta, \varphi) = r_0(\theta, \varphi) + \sum_{l=1}^{L_{\max}} \sum_{m=0}^{l} \left[C_{l,m} \cos(m\varphi) + S_{l,m} \sin(m\varphi)\right] P_{l,m}(\cos\theta), \tag{45}$$

where r_0 is a reference radius, $P_{l,m}$ are the associated Legendre polynomials, and $(C_{l,m}, S_{l,m})$ are the tabulated spherical harmonic coefficients up to degree and order $L_{\max} = 2050$. This high-resolution model enables a faithful reconstruction of the lunar topography, capturing both global

structure and fine-scale features. The dataset is distributed via NASA's Planetary Data System (PDS) and constitutes the standard scientific basis for lunar surface analysis and generative modeling. Empirically, we found that retaining only the leading 50 harmonic coefficients suffices to capture the essential topographic features for generative modeling. Accordingly, our experiments utilize coefficients up to degree $L_{\max} = 50$. To train the diffusion model, we sample 2^{10} points from the spherical harmonic parameterization of the lunar surface and employ these as data instances $x_0 \sim \rho_0^{\mathcal{M}}$ to define the score matching loss.

Asteroid Spherical Harmonic Dataset (NEAR/Eros). In second real-world dataset, we reconstruct and analyze the shape of asteroid 433 Eros by utilizing the spherical harmonic coefficient dataset acquired by the NEAR Shoemaker mission Miller et al. [2002]. The dataset provides a tabulated set of real spherical harmonic coefficients $(C_{l,m})$ up to a specified maximum degree $L_{\max}$, as determined from laser altimeter and optical data. The surface is parameterized in the form:

$$r(\theta, \varphi) = r_0(\theta, \varphi) + \sum_{l=0}^{L_{\max}} \sum_{m=-l}^{l} C_{l,m} Y_{l,m}(\theta, \varphi), \tag{46}$$

where $Y_{l,m}$ denotes the (complex or real-valued) spherical harmonics. Each row in the dataset specifies $(l, m, C_{l,m})$, which are filtered by thresholding or truncation in downstream processing. This high-resolution expansion enables both scientific study and generative modeling of the asteroid's three-dimensional surface geometry. The dataset is publicly available as part of the NEAR A Shoemaker mission archive distributed by NASA's Planetary Data System (PDS) Neumann and Smith [2001]. To maintain consistency with our lunar surface experiments, we adopt an analogous experimental setup for the Eros dataset. Specifically, we retain the leading 50 spherical harmonic coefficients to parameterize the asteroid's surface and uniformly sample 2^{10} points for use in training the diffusion model via the score matching loss.

Shape Analysis of Human Anatomy Surfaces. To demonstrate the applicability of our framework to biomedical shape analysis, we consider publicly available anatomical datasets distributed with the SPHARM-PDM software [Styner et al., 2006, Shen et al., 2009]. Specifically, we adopt (i) the *Hippocampi dataset*, which consists of left and right hippocampus surface models from multiple subjects divided into two clinical groups, and (ii) the *Knee dataset*, which provides volumetric MRI data of the knee joint along with preprocessed surface reconstructions. Both datasets are parameterized onto the sphere using spherical harmonics, enabling direct comparison of 3D morphologies across subjects. These anatomical surfaces serve as standard benchmarks in neuroimaging and musculoskeletal research, providing a controlled experimental setting for validating geometric learning algorithms.

Quantum Tomography and Dynamics. To prepare the quantum dataset, we leverage the publicly available QDataSet Perrier et al. [2022], focusing on its two-qubit control simulations. In contrast to classical surface datasets that provide explicit geometric representations, the raw QDataSet does not directly provide state vectors; instead, each sample is given as a 52-dimensional feature vector of Pauli expectation values corresponding to an informationally complete operator set. To recover states on the *pure-state manifold* $\mathbb{CP}^3$, we first reconstruct a two-qubit density matrix via linear-inversion tomography: given measurement operators $\{V_k\}_{k=1}^{52}$ and observed expectations $E_k = \mathrm{Tr}(\rho V_k)$, we solve an overdetermined least-squares problem to estimate $\hat{\rho}$. Since our target distribution is supported strictly on pure states, we then enforce purity by projecting $\hat{\rho}$ onto its closest rank-one approximation along the direction of its principal eigenvector, defining

$$\hat{\rho}_{\mathrm{pure}} = |\hat{\psi}\rangle\langle\hat{\psi}|,$$

where $|\hat{\psi}\rangle$ is the eigenvector of $\hat{\rho}$ associated with the largest eigenvalue. The resulting dataset is thus represented by the corresponding rays $[\hat{\psi}] \in \mathbb{CP}^3$.

10.5 Neural Network Architecture

Our framework employs a UNet-style multilayer perceptron designed for structured vector inputs. The input feature is the concatenation of the latent state $E \in \mathbb{R}^d$, the displacement $\Delta x \in \mathbb{R}^d$, and the flattened inverse frame matrix $\mathrm{vec}(e^{-1}) \in \mathbb{R}^{d^2}$, while the time variable t is encoded by a sinusoidal embedding $\tau(t) \in \mathbb{R}^T$ and injected throughout the network via FiLM conditioning Perez et al. [2018].

Table 5: Layer composition of ScoreUNet. Each encoder stage projects the features to a higher channel dimension, the bottleneck applies two residual transformations, and the decoder fuses skip connections back to lower channels. Both heads output d-dimensional vectors for the score and bridge predictions.

Block	**#Linear Layers**	**Input → Output Dim**	**Notes**
Input concat	–	$[E \in \mathbb{R}^d,\ \Delta x \in \mathbb{R}^d,\ \mathrm{vec}(e^{-1}) \in \mathbb{R}^{d^2}]$	Time embedding $\tau(t) \in \mathbb{R}^T$ injected via FiLM
Encoder stage s	$1 + 3R$	$C_{s-1} \to C_s$	in-projection (1) + R ResBlocks (fc1+fc2+FiLM)
Bottleneck	2×3	$C_L \to C_L$	two ResBlocks, each with (fc1+fc2+FiLM)
Decoder stage s	$1 + 3R$	$(C_{s+1} \oplus C_{s+1}) \to C_s$	merge-projection (1) + R ResBlocks
Score head	1	$C_1 \to d$	LN + Act + Linear
Bridge head	1	$C_1 \to d$	LN + Act + Linear
Per-ResBlock	3	$C \to C$	fc1, fc2, and FiLM($T \to 2C$)

The encoder consists of successive stages, each performing a linear in-projection followed by R FiLM-modulated residual blocks, thereby projecting features to higher channel dimensions and storing intermediate outputs as skip connections.

At the bottleneck, two residual transformations refine the representation with layer normalization, activation, and FiLM-modulated affine conditioning. The decoder mirrors the encoder by concatenating the current hidden representation with the corresponding skip features, applying a merge projection, and passing the result through R residual blocks, progressively reducing the channel dimension back to the base width. Finally, two parallel output heads map the decoded representation to task-specific predictions: the *score head* outputs d-dimensional vectors corresponding to the learned score function $s(t, E)$, and the *bridge head* outputs d-dimensional vectors modeling the bridge function $v(t, \Delta x, e^{-1})$. This architecture preserves the UNet property of combining hierarchical feature extraction with precise skip connections, while FiLM conditioning ensures that temporal information influences all layers of the computation.

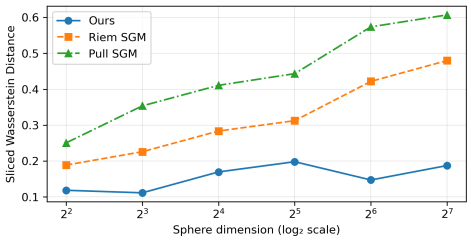


Figure 5: Scalability Analysis on Dimensionality on High-Dimensional Spheres. Performance of different methods as the dimension of the sphere increases. Unlike Riemannian SGM and Pull NF, our HDM maintains stable and accurate density estimation even in high-dimensional settings by leveraging Euclidean-based modeling.

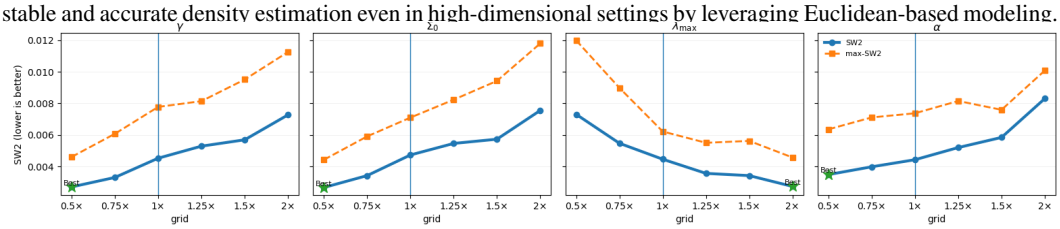


Figure 6: Sensitivity Analysis on High-Dimensional Spheres. Each panel illustrates sensitivity analysis by systematically varying a single parameter (γ, Σ_0, $\lambda_{\max}$, α) through multiplicative perturbations around the baseline (vertical blue line at $1\times$), while keeping all other parameters constant. The performance metrics are reported as the sliced Wasserstein-2 distance and the max-sliced Wasserstein-2 distance between the target and generated samples.

10.6 Ablation Study

Scalability Analysis on Dimensionality. We systematically investigate how increasing the intrinsic dimension of the sphere $\mathbb{S}^n$ affects density estimation performance. Unlike prior methods such as Riemannian SGM and Pull NF, HDM leverages Euclidean neural networks for score modeling, leading to substantially more robust and stable reconstruction as dimensionality grows. Across spheres of increasing dimension, existing methods exhibit pronounced performance degradation, whereas HDM maintains consistently high accuracy by capitalizing on the favorable optimization and expressivity properties of Euclidean-based modeling. While the lifted formulation introduces state variables of size $d + d^2$, these additional degrees of freedom largely arise from fixed, non-trainable geometric components (*e.g.*, frame- and connection-related quantities) rather than learnable parameters of nueral networks, thus the total GPU momery used in this framework is soley focused on Euclidean score function, which is supposed to be lifted by using horizontal lift. Empirically, the peak GPU memory footprint scales comparably to Riemannian SGM as d increases: when moving from $d = 64$ to $d = 128$, Riemannian SGM grows from $1.6\,\text{GB}$ to $2.4\,\text{GB}$, whereas HDM increases from $2.0\,\text{GB}$ to $2.5\,\text{GB}$, demonstrating similar memory expansion despite operating in the lifted representation.

Sensitivity Analysis on Hyperparameters. We perform a comprehensive sensitivity analysis on key hyperparameters to evaluate their impact on model performance in Figure 6. Specifically, we systematically vary each of the parameters $(\gamma, \Sigma_0, \lambda_{\max}, \alpha)$ introduced in Sec 10.1 individually, while holding the remaining parameters fixed at their baseline values. Empirically, the analysis demonstrates that smaller values of γ, Σ_0, and α lead to improved alignment between the generated and target distributions, whereas larger values of $\lambda_{\max}$ enhance performance within the tested range. The experimental setup for this analysis is standardized across all evaluations, using total integration time $T = 1$, $N = 600$ integration steps, 4096 particles, an exponential schedule for the diffusion parameter λ_s, 50 random projections, and Wasserstein exponent $p = 2$.

10.7 Algorithm

This section presents the core algorithms for sampling and simulating horizontal diffusions on the frame bundle. We introduce methods for stationary distribution sampling, orthonormal frame construction, horizontal lift evaluation, and reverse-time diffusion simulation.

Sampling Stationary Distribution on $\mathcal{O}(\mathcal{M})$. we first outline Algorithm 1, which samples independent points $(x, e) \in \mathcal{O}(\mathcal{M})$ from the stationary distribution given by the product measure $d\mathrm{Vol}_g \otimes d\mathrm{Haar}_{O(d)}$. The algorithm first draws a base point x according to the Riemannian volume measure, followed by generating an orthonormal frame e at x via QR factorization of a Gaussian matrix. Thus, it ensures Haar-distributed randomness in each fiber. As the base and fiber components are sampled independently, this procedure yields unbiased samples from the stationary law of the horizontal diffusion on the frame bundle.

Algorithm 1 SAMPLING STATIONARY MEASURE $d\rho_\infty^{\mathrm{H}} = d\mathrm{Vol}_g \otimes d\mathrm{Haar}_{O(d)}$

Require: Riemannian manifold $(\mathcal{M}, g)$. REJECTIONVOLUMESAMPLE() that returns $x \sim d\mathrm{Vol}_g$;
Ensure: $(x, e) \in \mathcal{O}(\mathcal{M})$ sampled from the product measure.

Sampling the base point

$x \leftarrow$ REJECTIONVOLUMESAMPLE()

Sampling the orthonormal frame at x

$Z \leftarrow \mathrm{randn}(d, d)$ ▷ entries i.i.d. $\mathcal{N}(0, 1)$

$(Q, R) \leftarrow \mathrm{QR}(Z)$ ▷ column-orthonormal Q

$Q \leftarrow Q\mathrm{diag}\big(\mathrm{sign}(\mathrm{diag}(R))\big)$ ▷ enforce $Q \in O(d)$ (Haar)

$e \leftarrow$ FRAMEFROMMATRIX(x, Q) ▷ convert Q to orthonormal frame in $T_x\mathcal{M}$

Return

return (x, e)

Given a point $x \in \mathcal{M}$ and a Haar-random orthogonal matrix $Q \in O(d)$, Algorithm 2 constructs an orthonormal frame in $T_x\mathcal{M}$ by first orthonormalising the local coordinate basis via a metric-aware Gram–Schmidt process, and subsequently applying the columns of Q as coefficients to form the final frame. This yields an isometric linear map $e : \mathbb{R}^d \to T_x\mathcal{M}$, whose orientation is distributed according to the Haar measure on the fibre over x.

Algorithm 2 FRAMEFROMMATRIX(x, Q)

Require: Point $x \in \mathcal{M}$, Haar matrix $Q \in O(d)$
Ensure: Orthonormal frame $e = \big(e_1, \ldots, e_d\big)$ at x

$\big\{b_i = \frac{\partial}{\partial x^i}\big|_x\big\}_{i=1}^d \leftarrow$ LOCALBASIS(x) ▷ chart differentials

$\big\{\hat{e}_i\big\}_{i=1}^d \leftarrow$ GRAMSCHMIDT$(\{b_i\}, g_x)$ ▷ produces g-orthonormal vectors

for $a = 1$ **to** d **do**

$\quad e_a = \sum_{i=1}^{d} \hat{e}_i Q_{ia}$ ▷ Embed Haar matrix

end for

return $e = (e_1, \ldots, e_d)$ ▷ linear map $e : \mathbb{R}^d \to T_x\mathcal{M}$

In Algorithm 3, a simple rejection sampling scheme is used to generate points in local coordinates $y \in B$ with respect to the intrinsic Riemannian volume. Proposals are drawn uniformly from the coordinate patch, and accepted with probability proportional to the volume density $\mathrm{VolDens}(y)$, normalized by an upper bound M. This procedure yields samples exactly distributed according to the manifold's volume element, serving as an accessible alternative to more complex MCMC methods when no explicit inverse map is available.

Algorithm 3 REJECTIONVOLUMESAMPLE()

Require: **Local Coordinate domain** $B \subset \mathcal{M}$, **Determinant oracle** $\mathrm{VolDens}(y) \equiv \sqrt{\det g_{ij}(y)}$ for $y \in B$, Constant $M \geq \sup_{y \in B} \mathrm{VolDens}(y)$

Ensure: Local coordinate vector $y \in B$ distributed according to the intrinsic volume measure $\mathrm{VolDens}(y)dy$

repeat
  $y \leftarrow \mathrm{Unif}(B)$ ▷ draw uniformly in coordinate region
  $p \leftarrow \mathrm{VolDens}(y)/M$ ▷ $0 < p \leq 1$
  $u \leftarrow \mathrm{Unif}(0, 1)$
until $u < p$ ▷ accept with probability p
return $y^\star \leftarrow y$

Sampling Reverse Horizontal Diffusion $\mathbf{U}_s$. Algorithm 4 computes the horizontal lift matrix H at a given frame bundle state $U = (x, e)$, where $x \in \mathcal{M}$ and $e \in O(d)$. The resulting $H \in \mathbb{R}^{(d+d^2)\times d}$ encodes the horizontal distribution associated with the Levi-Civita connection at x. Each column H_a concatenates the base-point direction $e_{\cdot a}$ and the frame evolution $-\Gamma(e_{\cdot a})e$, where Γ denotes the Christoffel symbol tensor at x. The upper d rows of H_a yield the directional derivative along $e_{\cdot a}$ in $T_x\mathcal{M}$, while the lower d^2 rows encode the corresponding infinitesimal frame rotation induced by parallel transport. This construction provides a local trivialization of the horizontal distribution, ensuring compatibility with the manifold's Riemannian geometry and connection.

Algorithm 4 HORIZONTALLIFT($U{=}(x, e)$)

Require: State $U = (x, e)$ with *base point* $x \in \mathcal{M}$ and *frame matrix* $e = [e_{ia}] \in \mathrm{O}(d)$

Ensure: Horizontal lift matrix $H = \big[H_1 \mid \cdots \mid H_d\big]$ such that $H_a = \big(e_{\cdot a}, -\Gamma(e_{\cdot a})e\big)$

$\Gamma \leftarrow \textsc{Christoffel}(x)$ ▷ Γ^l_{jk} tensor of size $d \times d \times d$
$H \leftarrow \mathbf{0}_{(d+d^2)\times d}$
for $a = 1$ **to** d **do**
  // base–point component
  for $i = 1$ **to** d **do**
    $H_{i,a} \leftarrow e_{ia}$ ▷ $(H_a)^x = e_{\cdot a}$
  end for
  // vertical (frame) component
  for $l = 1$ **to** d **do**
    for $m = 1$ **to** d **do**
      $s \leftarrow 0$
      for $j = 1$ **to** d **do**
        for $k = 1$ **to** d **do**
          $s \leftarrow s + \Gamma^l_{jk} e_{ja} e_{km}$
        end for
      end for
      $r \leftarrow d + (l-1)d + m$ ▷ row index for (l, m) pair
      $H_{r,a} \leftarrow -s$
    end for
  end for
end for
return H

With the definition of horizontal lift of Euclidean latent vectors, Algorithm 5 describes a discretized scheme for sampling from the reverse-time dynamics of the horizontal diffusion process on the frame bundle $\mathcal{O}(\mathcal{M})$. The procedure reconstructs approximate samples from the initial data distribution ρ_0^{H} by integrating the reverse Stratonovich SDE backward from a given terminal state.

Given a terminal state (U_T, E_T), the algorithm first precomputes the reverse-time grid and associated scalar schedules, including stepwise values for the bridge parameter λ, the drift γ, and the noise strength σ^2. Initialization is performed by setting the latent and geometric states to the prescribed terminal values.

Algorithm 5 REVERSEHORIZONTALDIFFUSION$(U_T, E_T; T, N)$

Require: Terminal state $U_T = (x_T, e_T) \in \mathcal{O}(\mathcal{M})$, Terminal latent $E_T \in \mathbb{R}^d$, Time horizon $T > 0$, Number of steps N with step size $\Delta s = T/N$.
Ensure: Approximate sample U_0 from the data distribution ρ_0^{H}

Pre–compute time grids and scalar schedules
for $i \leftarrow 0$ **to** N **do**
  $s_i \leftarrow i\Delta s, \quad \lambda_i \leftarrow \lambda(s_i), \gamma_i \leftarrow \gamma(s_i), \sigma_i^2 \leftarrow \sigma^2(s_i) = \gamma_i \Sigma(s_i)$
end for

Initialize reverse variables
$U \leftarrow U_T, \quad E \leftarrow E_T$
for $i \leftarrow N-1$ **down to** 0 **do**
  $s \leftarrow s_i, \quad \Delta s \leftarrow s_{i+1} - s_i$ ▷ reverse time increment
  //geometry: bridge vector in frame coordinates
  $x \leftarrow \text{INITIALPOINT}(U)$, $e \leftarrow \text{INITIALFRAME}(U)$
  $v \leftarrow v(t, \theta)$
  //scores and Stratonovich drift
  $b_{\mathrm{rev}} \leftarrow -\lambda_i v + \gamma_i E + \sigma_i^2 s_\theta(E, s)$
  $\xi \leftarrow \sigma_i \sqrt{\Delta s} * \text{randn}(d)$ ▷ $\xi \sim \mathcal{N}(0, \sigma_i^2 \Delta s I_d)$
  $dE \leftarrow b_{\mathrm{rev}} \Delta s + \xi$
  $E \leftarrow E + dE$
  //horizontal Stratonovich update of the frame path
  $H \leftarrow \text{HORIZONTALLIFT}(U)$ ▷ $H \in \mathbb{R}^{(d+d^2)\times d}$
  $U \leftarrow U + H dE$ ▷ Stratonovich step $dU = H(U) \circ dE$
  if $i \equiv 0 \pmod{P}$ **then** ▷ every P steps re–orthonormalise
    $e \leftarrow \text{ORTHONORMALIZEMETRIC}(e)$ ▷ $e^T g(x) e = I$
    $U \leftarrow (x, e)$
  end if
end for
return $U_0 \leftarrow U$

The main loop then iteratively integrates the reverse SDE from time T to 0 in N discrete steps. At each iteration, the current frame and base point are extracted, and the neural networks infer the bridge vector $v = v(t, \theta)$, quantifying displacement relative to the anchor. The reverse drift combines three contributions: a deterministic pull towards x_0, a restoring force in the latent, and a score-driven correction based on the model's learned score $\text{SCORE}(E, s)$. Stochastic increments ξ simulate Gaussian noise compatible with the instantaneous covariance.

The latent vector E is updated via an Euler–Maruyama step, and the corresponding geometric state U is propagated horizontally in the frame bundle according to the Stratonovich rule $dU = H(U) \circ dE$. To ensure numerical stability, we prform periodic re-orthonormalization of the frame using the manifold metric.

Upon completion, the algorithm returns the initial geometric state U_0, representing a sample drawn approximately from the reverse-time law of the horizontal diffusion process. This scheme thereby enables efficient simulation-based generation of samples from complex data distributions defined over the manifold and its frame bundle.

Impact Statement

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

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 introduction state the main contributions of HDM, including the frame-bundle formulation, horizontal diffusion processes, and empirical validation on synthetic and scientific manifolds; these claims are supported by the theory and experiments in the paper.

   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: The paper discusses important practical limitations of the approach, including the reliance on local differential-geometric quantities such as the metric and Levi–Civita connection, as well as geometry-dependent sample complexity and curvature effects.

   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: The paper states formal propositions and theorems and provides assumptions, proof sketches, and appendix derivations for the theoretical results.

   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: The manuscript provides the proposed formulation, algorithmic details, datasets, architecture, and evaluation setup in the main paper and appendix, which together provide a reproducible description of the reported results.

   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: [N/A]

   Justification: The submission provides detailed methodological and experimental descriptions, and the code will be released upon acceptance.

   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: The experimental setup, datasets, architecture, hyperparameters, evaluation metrics, and additional ablations are described in the Experiments and Implementation Details sections and the appendix.

   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: The paper reports quantitative comparisons using sliced Wasserstein metrics across multiple datasets and settings, providing sufficient empirical detail to support the comparative conclusions presented in the manuscript.

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: The paper reports implementation-level computational information, including the lifted state size, runtime-related discussion, and GPU memory usage trends in the ablation section.

   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 believe that this research conforms to it.

   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: [N/A]

    Justification: This work is a foundational study on geometric generative modeling, and we do not identify a direct societal impact that requires a dedicated broader-impact discussion for this submission.

    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: The paper does not release high-risk assets such as large-scale generative systems or scraped datasets that would require special misuse safeguards.

    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: The paper credits the datasets, software, and prior methods used in the experiments through citations in the main paper and appendix.

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: [N/A]

Justification: This submission does not release a new public asset alongside the anonymized paper submission.

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 research with 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 crowdsourcing or research with human subjects and therefore does not require IRB approval.

    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: The core methodology of this work does not use LLMs as an important, original, or non-standard component.

    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.