RW-FLOW: ONE-STEP GENERATION ON COMPACT MANIFOLDS VIA WASSERSTEIN GRADIENT FLOWS

Anonymous authors
Paper under double-blind review

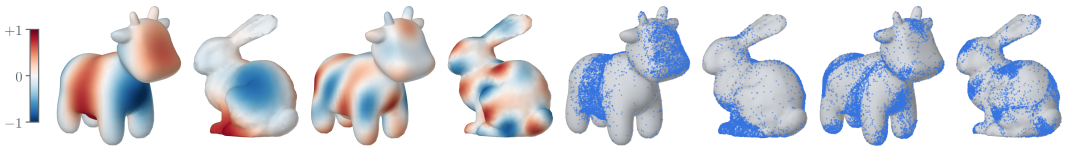


Figure 1: **One-step samples on general manifolds.** Left: target eigenfunctions on the Bunny and Spot meshes for different eigenfunction indices k. Right: samples generated by our method in a single function evaluation (NFE $= 1$). Samples concentrate according to the geometry-induced target distributions on the underlying meshes.

ABSTRACT

Manifold-valued data, and consequently the distributions they induce, are prevalent across many domains, ranging from the locations of geospatial events, such as earthquakes, to biomolecular torsion angles that encode information about three-dimensional structure. While diffusion and flow-based generative models have been successfully extended to compact manifolds, sampling typically requires tens or hundreds of sequential network evaluations. We introduce RW-FLOW, a theoretically grounded framework for learning one-step generative models on compact manifolds via Wasserstein gradient flows. The main challenge is *identifiability*: driving the velocity field to zero should guarantee that the model distribution matches the target distribution. We establish a necessary and sufficient condition for identifiability on compact, connected Riemannian manifolds. We specifically show that, for a symmetric, Lipschitz-continuous cost function, the velocity field induced by the Sinkhorn divergence is identifiable if and only if the associated Gibbs kernel is nondegenerate. This characterization provides a general principle for designing identifiable costs on compact manifolds. It also reveals that the squared geodesic distance, the natural manifold analogue of the squared Euclidean distance, does not always guarantee identifiability. Across benchmarks involving geospatial events, protein side chain torsion angles, RNA backbone torsion angles, and general manifolds discretized as triangular meshes, RW-FLOW outperforms existing one-step methods in nearly all settings under fair comparison conditions.

1 INTRODUCTION

Diffusion and flow-based generative models have been successfully extended from Euclidean data to distributions supported on Riemannian manifolds (Huang et al., 2022; De Bortoli et al., 2022; Chen & Lipman, 2024), enabling applications to manifold-valued data such as geospatial events on spheres and molecular torsion angles on tori. However, sampling remains costly, typically requiring many network evaluations. Recent work has therefore begun adapting few-step acceleration techniques, including flow-map distillation, to Riemannian settings (Davis et al., 2026; Woo et al., 2026; Cheng et al., 2025). These methods reduce sampling cost but are more difficult to optimize due to redundancy in their parameterizations: they learn maps across multiple pairs of timesteps rather than directly optimizing the one-step map used at inference.

In contrast, Han et al. (2026) recently proposed W-Flow, a framework for learning one-step generative models based on Wasserstein gradient flows (WGFs). W-Flow directly learns a one-step pushforward map by regressing model outputs toward moving targets induced by the Wasserstein gradient of an energy functional that measures the discrepancy between the model and target distributions. A

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

key requirement is *identifiability*: the velocity field should vanish only when the model distribution matches the target distribution. Han et al. (2026) considered various divergences as the energy functional and showed that the Sinkhorn divergence (Feydy et al., 2019) with the squared Euclidean distance cost performs best and satisfies the identifiability property in Euclidean spaces.

Given the strong empirical performance of W-Flow (Han et al., 2026) in Euclidean domains, a natural question is whether its one-step formulation can be extended to compact manifolds. A key challenge is that characterizing cost functions that induce identifiable velocity fields is not straightforward, in contrast to the Euclidean setting examined by Han et al. (2026), where the squared Euclidean distance guarantees identifiability. This leads to our central question: *which cost functions induce identifiable velocity fields on compact manifolds?*

For compact, connected Riemannian manifolds, we show that there exist many cost functions which induce an identifiable velocity field, opening a new axis for designing one-step models by modifying the cost function. We characterize these cost functions by showing that for a symmetric, Lipschitz cost function, the induced velocity field is identifiable if and only if the associated Gibbs kernel is nondegenerate. Based on this, we explore broad classes of costs that induce identifiable velocity fields, including spectral costs on general compact manifolds and distance-based costs under suitable geometric conditions. We also analyze the squared geodesic cost, the natural manifold analogue of the squared Euclidean cost, and show that it does not always guarantee identifiability.

On standard benchmarks, including Earth and climate datasets (National Geophysical Data Center / World Data Service (NGDC/WDS), 2020a;b; Brakenridge, 2017; NASA Earth Observing System Data and Information System (EOSDIS), 2020) on spherical domains and protein and RNA torsion-angle datasets (Lovell et al., 2003; Murray et al., 2003) on tori, we demonstrate that our approach is not only theoretically sound but also empirically effective, surpassing existing methods in nearly all settings under matched model capacity and training time. We further show that spectral costs that are applicable to *any* compact manifold by construction, enable accelerated sampling on manifolds represented as triangular meshes (Fig. 1). In these settings, our method achieves performance comparable to that of a multi-step baseline while sampling two to three orders of magnitude faster.

2 Background: W-Flow

W-Flow (Han et al., 2026) is a powerful framework for learning one-step generative models by evolving the model distribution toward the target along a Wasserstein gradient flow (WGF) of an energy functional that measures the discrepancy between the two. Given an energy functional $\mathcal{F} : \mathcal{P}(\mathbb{R}^n) \to \mathbb{R}$ on the space of probability distributions, its WGF is defined by

$$\partial_t q_t = \nabla \cdot \left(q_t \nabla \frac{\delta \mathcal{F}}{\delta q}(q_t) \right), \tag{1}$$

where $\frac{\delta \mathcal{F}}{\delta q}(q_t)$ denotes the first variation of $\mathcal{F}$ at q_t. This induces a time-dependent probability path q_t in the Wasserstein space following the steepest descent direction of $\mathcal{F}$. Equivalently, Eqn. 1 can be viewed as a continuity equation with velocity field $V_{q_t,p} = -\nabla \frac{\delta \mathcal{F}}{\delta q}(q_t)$.

Choosing $\mathcal{F}(q) = \mathcal{D}(q, p)$ for a divergence $\mathcal{D}$ and target distribution p yields the velocity field $V_{q,p} = -\nabla \frac{\delta \mathcal{D}(\cdot,p)}{\delta q}(q)$, which transports q_t toward p. Han et al. (2026) translate this flow into a model-training objective: they transport generated samples along $V_{q_\theta,p}$ and regress the generator outputs toward the transported samples.

$$\mathcal{L} = \mathbb{E}_{z\sim q_0} \left[\| f_\theta(z) - \texttt{stopgrad}(f_\theta(z) + \eta V_{q_\theta,p}(f_\theta(z))) \|^2 \right], \tag{2}$$

where f_θ is a one-step transport map, $q_\theta = (f_\theta)_{\#} q_0$ is the model distribution, and $\eta > 0$ is the step size. Note that Eqn. 2 is equivalent to $\eta^2 \mathbb{E}_{x\sim q_\theta} \| V_{q_\theta,p}(x) \|^2$: it vanishes exactly when the velocity field does. Driving the velocity field to zero, however, does not necessarily guarantee that the model converges to the target distribution (*i.e.,*$V_{q_\theta,p} \equiv 0$ implies $q_\theta = p$); such a guarantee requires the field to be *identifiable*.

For identifiability, Han et al. (2026) instantiate $\mathcal{D}$ with the *Sinkhorn divergence* (Feydy et al., 2019; Ramdas et al., 2017; Genevay et al., 2018; Salimans et al., 2018),

$$\mathcal{S}_{\varepsilon,c}(q, p) = \mathrm{OT}_{\varepsilon,c}(q, p) - \tfrac{1}{2}\mathrm{OT}_{\varepsilon,c}(q, q) - \tfrac{1}{2}\mathrm{OT}_{\varepsilon,c}(p, p), \tag{3}$$

where $\mathrm{OT}_{\varepsilon,c}$ is an entropy-regularized optimal transport cost:

$$\mathrm{OT}_{\varepsilon,c}(q,p) = \min_{\pi\in\Pi(q,p)} \int c(x,y)\,\mathrm{d}\pi(x,y) + \varepsilon D_{\mathrm{KL}}(\pi \,\|\, q\otimes p). \tag{4}$$

Here, $\Pi(q,p)$ is the set of couplings with marginals q and p and $c(x,y)$ specifies the cost of transporting mass from x to y. The minimizer π^{qp} is unique and is typically computed in practice by Sinkhorn iterations (Cuturi, 2013). With $c(x,y) = \frac{1}{2}\|x-y\|^2$, Han et al. (2026) show that the Sinkhorn divergence is identifiable.

Although W-Flow has achieved state-of-the-art one-step performance on Euclidean domains, including image generation, its extension to compact manifolds remains unexplored. Moreover, while identifiability holds for many cost functions in Euclidean domains, it remains unclear whether any such cost function exists on compact manifolds, motivating the following question: *Which cost functions guarantee identifiability on compact manifolds, and how can we characterize them?*

3 RW-Flow

We introduce Riemannian W-Flow (RW-Flow), which extends W-Flow (Han et al., 2026) to compact, connected Riemannian manifolds. In the following, we first derive the WGF of the Sinkhorn divergence defined with a general cost function on a Riemannian manifold (Sec. 3.1). We then provide our main theorem, which characterizes cost functions by connecting identifiability to the nondegeneracy of the associated Gibbs kernels, and present two approaches for verifying this nondegeneracy (Sec. 3.2). Finally, we explore a variety of cost functions by applying these criteria, including the squared geodesic cost and both spectral and distance-based costs (Sec. 3.3).

3.1 Velocity Fields on Compact Manifolds

Let $\mathcal{M}$ denote a compact, connected, smooth manifold equipped with a Riemannian metric g, so that $\mathcal{M}$ is complete and the exponential map $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ is defined on the whole tangent space $T_x\mathcal{M}$. We write $\langle\cdot,\cdot\rangle_g$ and $\|\cdot\|_g$ for the metric and its norm on $T_x\mathcal{M}$, d_g for the geodesic distance, $\log_x$ for the inverse of $\exp_x$ on the domain where it is defined, and ∇_g, div_g for the Riemannian gradient and divergence. Given a base distribution $q_0 \in \mathcal{P}(\mathcal{M})$ and a target distribution $p \in \mathcal{P}(\mathcal{M})$, our goal is to learn a transport map $f_\theta : \mathcal{M} \to \mathcal{M}$ such that $(f_\theta)_{\#}q_0 = p$.

Building on this machinery, we first extend the W-Flow training objective (Han et al., 2026) in Eqn. 2 to Riemannian manifolds. In particular, Eqn. 2 comprises three components specific to $\mathbb{R}^n$: (1) the additive update $f_\theta(z) + \eta V_{q_\theta,p}(f_\theta(z))$, (2) the squared Euclidean distance between the generator output and the target, and (3) the field V itself. On a Riemannian manifold $(\mathcal{M}, g)$, the first two components admit natural intrinsic counterparts: a point x is moved along a tangent vector $V(x) \in T_x\mathcal{M}$ by the exponential map $\exp_x(\eta V(x))$ with $\eta > 0$ denoting a step size, while the discrepancy between two points x and y is measured by the squared geodesic distance $d_g^2(x,y)$. Accordingly, for Riemannian manifolds, Eqn. 2 can be rewritten as

$$\mathcal{L} = \mathbb{E}_{z\sim q_0}\left[d_g^2\left(f_\theta(z),\ \texttt{stopgrad}\left(\exp_{f_\theta(z)}(\eta\, V_{q_\theta,p}(f_\theta(z)))\right)\right)\right]. \tag{5}$$

For the exponential map in Eqn. 5 to be well defined, the field evaluated at $x = f_\theta(z)$ must lie in the tangent space at $f_\theta(z)$. This condition is readily satisfied when V is derived from a Wasserstein gradient flow on $\mathcal{M}$. Specifically, the continuity equation in Eqn. 1 naturally extends to $\mathcal{M}$ by replacing the Euclidean operators with their Riemannian counterparts (Villani, 2008)

$$\partial_t q_t = \mathrm{div}_g\left(q_t \nabla_g \frac{\delta\mathcal{F}}{\delta q}(q_t)\right). \tag{6}$$

Analogous to W-Flow (Han et al., 2026), we choose the debiased Sinkhorn divergence as our energy functional: $\mathcal{F}(q) = \mathcal{S}_{\varepsilon,c}(q,p)$. Here, $\mathrm{OT}_{\varepsilon,c}$ and $\mathcal{S}_{\varepsilon,c}$ are defined as in Eqn. 4 and Eqn. 3, respectively. Then, the resulting velocity field, obtained as the Wasserstein gradient of $\mathcal{S}_{\varepsilon,c}$, is given below.

Theorem 3.1. *Let $\mathcal{F}(q) = \mathcal{S}_{\varepsilon,c}(q,p)$ on a compact $(\mathcal{M}, g)$ and c be a differentiable in its first argument. Then the Wasserstein gradient flow velocity $V = -\nabla_g \frac{\delta\mathcal{F}}{\delta q}$ is*

$$V_{q,p}^{\varepsilon,c}(x) \;=\; -\int_{\mathcal{M}} \nabla_g^{(1)} c(x,y)\; \pi^{qp}(\,\mathrm{d}y \mid x) \;+\; \int_{\mathcal{M}} \nabla_g^{(1)} c(x,z)\; \pi^{qq}(\,\mathrm{d}z \mid x) \;\in\; T_x\mathcal{M}, \tag{7}$$

where $\pi^{qp}(\,\mathrm{d}y \mid x)$ denotes the disintegration with respect to x of the optimal coupling π^{qp} of the entropy-regularized $\mathrm{OT}_{\varepsilon,c}(q,p)$ in Eqn. 4, and $\nabla_g^{(1)}$ is the Riemannian gradient in the first argument.

The derivation can be found in App. B.1. Intuitively, the first term shifts x toward the data points to which it is coupled, whereas the second term induces repulsion among the generated samples and cancels the entropic bias, ensuring that the field vanishes when $q = p$.

An immediate consequence of Thm. 3.1 is that the cost function determines the form of the velocity field and, consequently, its identifiability. In the next subsection, we present a theoretical criterion for determining whether a given cost function induces an identifiable velocity field.

3.2 Identifiability via Nondegeneracy

Recall from Sec. 2 that a velocity field $V_{q_\theta,p}$ is *identifiable* if $V_{q_\theta,p} \equiv 0$ implies $q_\theta = p$. Since the training objective in Eqn. 5 drives the norm of $V_{q_\theta,p}$ to zero, identifiability ensures that the generator does not reach a spurious fixed point even when $q_\theta \neq p$. While the squared Euclidean cost is sufficient to ensure identifiability in Euclidean spaces (Han et al., 2026), we extend this result by characterizing a broader class of cost functions that preserve identifiability.

Our main result, presented below, shows that for a symmetric, Lipschitz continuous cost function c, identifiability of the velocity field induced by the Sinkhorn divergence is determined by the *nondegeneracy* of the corresponding Gibbs kernel $k = e^{-c/\varepsilon}$, defined below.

Definition 3.1. *Let $(k * \sigma)(x) := \int k(x,y)\,\mathrm{d}\sigma(y)$ be an integral operator for $\sigma \in \mathfrak{M}_b(\mathcal{M})$. We call k nondegenerate if $\sigma \mapsto k * \sigma$ is injective on finite signed measures $\mathfrak{M}_b(\mathcal{M})$. Otherwise, it is degenerate.*

With this notion in place, we state our main theorem, which provides a criterion for determining which cost functions induce identifiable velocity fields and which do not.

Theorem 3.2. *Let $(\mathcal{M}, g)$ be a compact, connected Riemannian manifold with a symmetric, Lipschitz cost function $c : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$. Denote its Gibbs kernel $k := e^{-c/\varepsilon}$. Then $V^{\varepsilon,c}_{q,p}$ is identifiable if and only if k is nondegenerate.*

Thm. 3.2 implies a broader class of admissible cost functions. Specifically, if a cost function is symmetric and Lipschitz continuous and its associated Gibbs kernel is nondegenerate, then the corresponding velocity field is identifiable, making Eqn. 5 suitable for learning the generator. Motivated by this result, we next explore a broad range of cost functions in Sec. 3.3. Here, as a preliminary step, we introduce two approaches to verify the nondegeneracy of their associated Gibbs kernels: *spectral analysis* and *universality*.

Verifying Nondegeneracy via Spectral Analysis. For kernels that are diagonalizable in the Laplace-Beltrami eigenbasis, their nondegeneracy can be determined based on their spectra.

Lemma 3.3. *Let $\mathcal{M}$ be compact and k a continuous symmetric kernel on $\mathcal{M}$. Let $\{\phi_n\}_{n\geq 0}$ be Laplace-Beltrami eigenfunctions such that they form a basis for $L^2(\mathcal{M})$ and*

$$\int k(x,y)\,\phi_n(y)\,\mathrm{dvol}_g(y) = \hat{k}_n\,\phi_n(x) \qquad \text{for all } n \geq 0. \tag{8}$$

Then k is nondegenerate if and only if $\hat{k}_n \neq 0$ for all $n \geq 0$.

The proof is deferred to App. B. Notably, the condition in Lem. 3.3 applies to every spectral kernel on any compact manifold (Sec. 3.3.1), as well as to every kernel that depends only on the geodesic distance on $\mathbb{S}^d$ or $\mathbb{T}^d$. Importantly, Lem. 3.3 does not require k to be positive definite: the coefficients $\hat{k}_n$ may be negative, provided that none of them vanishes.

Verifying Nondegeneracy via Universality. When the spectrum is not available in closed form but the kernel is positive definite, nondegeneracy can instead be verified through *universality* (refer to App. C for definition), following Micchelli et al. (2006); Sriperumbudur et al. (2011).

Proposition 3.3.1 (Micchelli et al. (2006))**.** *Let $\mathcal{M}$ be compact and k a continuous positive definite kernel on $\mathcal{M}$. Then k is nondegenerate if and only if it is universal.*

Hence, when the Gibbs kernel associated with a cost function is positive definite, it suffices to verify its universality to establish nondegeneracy. Together with Thm. 3.2, the spectral and universality approaches provide complementary criteria for determining whether a given cost function induces an identifiable velocity field. With these results, we explore various cost functions in the following.

3.3 Cost Functions on Compact Manifolds

In the Euclidean case, the squared Euclidean cost guarantees identifiability (Han et al., 2026). It is therefore natural to ask whether its manifold analogue, the squared geodesic cost $c(x, y) = \frac{1}{2}d_g^2(x, y)$, also induces an identifiable velocity field. This cost satisfies the assumptions of Thm. 3.2; consequently, it suffices to determine whether the associated Gibbs kernel $\exp\big(-\frac{1}{2\varepsilon}d_g^2\big)$ is nondegenerate.

Proposition 3.3.2. *For the geodesic Gaussian kernel $k_\varepsilon = \exp\big(-\frac{1}{2\varepsilon}d_g^2\big)$ on $\mathbb{S}^d$ or $\mathbb{T}^d$, the set*

$$\Gamma = \{\varepsilon > 0 \mid k_\varepsilon \text{ is degenerate}\} \tag{9}$$

is at most countable. Particularly, for $\mathbb{T}^d$, Γ is countably infinite.

The proof is deferred to App. B. Since Gaussian kernels built on the geodesic distance are not positive definite for any $\varepsilon > 0$ on many manifolds (Feragen et al., 2015; Da Costa et al., 2023; 2025; Li, 2024), Prop. 3.3.1 does not apply directly; we instead rely on Lem. 3.3. On both $\mathbb{S}^d$ and $\mathbb{T}^d$, any kernel that depends only on the geodesic distance is diagonalizable as in Lem. 3.3, and the spectral coefficient associated with each mode vanishes for at most countably many values of ε. Therefore, on $\mathbb{S}^d$ and $\mathbb{T}^d$, the squared geodesic cost induces an identifiable velocity field for almost every $\varepsilon > 0$. As we show in Sec. 3.3.2, dropping the square resolves this on $\mathbb{S}^d$.

However, our argument does not extend to general compact manifolds, motivating us to seek cost functions whose Gibbs kernels are provably nondegenerate for every $\varepsilon > 0$. Through the lens of Thm. 3.2, we explore two such classes. The first, which we call *spectral costs*, is defined using the Laplace-Beltrami eigenfunctions (Sec. 3.3.1). The second, which we call *distance-based costs*, includes the chordal and geodesic costs whose guarantees hold under conditions on the manifold (Sec. 3.3.2). In the following, we introduce these cost functions and prove the identifiability of their induced velocity fields using Thm. 3.2, together with the spectral and universality criteria developed previously. We organize the cost families along with their guarantees and formulas in Tab. 1.

3.3.1 Spectral Costs

A *spectral cost* is a cost function induced by a spectral kernel $k_\rho : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$, defined on any compact Riemannian manifold $\mathcal{M}$. We begin by defining spectral kernels below.

Definition 3.2 (Spectral Kernels)**.** *Let $\{(\lambda_n, \phi_n)\}_{n=0}^{\infty}$ denote the eigenpairs of the Laplace-Beltrami operator on a compact Riemannian manifold $(\mathcal{M}, g)$, with $\{\phi_n\}$ orthonormal in $L^2(\mathcal{M})$. Spectral kernels are defined by*

$$k_\rho(x, y) = \sum_{n=0}^{\infty} \rho(\lambda_n)\phi_n(x)\phi_n(y), \tag{10}$$

where the spectral density $\rho\colon [0, \infty) \to [0, \infty)$ decays fast enough that the series converges uniformly.

A key property of spectral kernels is that their nondegeneracy is characterized by their spectral density ρ, as stated in the following proposition.

Proposition 3.3.3 (Nondegeneracy of Spectral Kernels)**.** *A spectral kernel k_ρ is nondegenerate if and only if $\rho(\lambda_n) \neq 0$ for all λ_n.*

Proof. Let $\{\lambda_n, \phi_n\}_{n\geq 0}$ be the eigenpairs of the Laplace-Beltrami operator on $\mathcal{M}$. Consider

$$k * (\phi_n \mathrm{vol}_g) = \int k_\rho(x, y)\phi_n(y)\,\mathrm{dvol}_g(y) = \rho(\lambda_n)\phi_n(x) \tag{11}$$

Now by Lem. 3.3, k_ρ is nondegenerate if and only if $\rho(\lambda_n) \neq 0$ for all λ_n. □

Using this property, we can show that nondegenerate spectral kernels induce cost functions that, when used to define the Sinkhorn divergence in Eqn. 4, yield identifiable velocity fields, as stated in the following result.

Table 1: **Overview of cost functions considered.** P_x denotes the orthogonal projection onto $T_x\mathcal{M}$ for $\mathcal{M} \subset \mathbb{R}^D$. The last column states for which ε and on which manifolds the Gibbs kernel is nondegenerate, so that the velocity field is identifiable by Thm. 3.2; "a.e." means all but countably many. On manifolds not listed, no guarantee is established.

Cost	$c(x, y)$	Gibbs kernel $e^{-c/\varepsilon}$	$\nabla_g^{(1)} c(x, y)$	Identifiable for
Squared Geodesic	$\frac{1}{2} d_g(x, y)^2$	$e^{-d_g(x,y)^2/2\varepsilon}$	$-\log_x(y)$	a.e. ε on $\mathbb{S}^d$, $\mathbb{T}^d$
Spectral	$-\varepsilon \log k_\rho(x, y)$	$\sum_n \rho(\lambda_n)\, \phi_n(x)\phi_n(y)$	$-\varepsilon\, \nabla_g^{(1)} \log k_\rho(x, y)$	all ε, any $\mathcal{M}$
Chordal	$\|x - y\|^2$	$e^{-\|x-y\|^2/\varepsilon}$	$2P_x(x - y)$	all ε, any $\mathcal{M} \subset \mathbb{R}^D$
Geodesic	$d_g(x, y)$	$e^{-d_g(x,y)/\varepsilon}$	$-\log_x(y)/d_g(x, y)$	all ε on $\mathbb{S}^d$

Corollary 3.3.1. *Let k_ρ be a spectral kernel defined on a compact, connected manifold $\mathcal{M}$ with a nonvanishing spectral density ρ with $k_\rho > 0$ and $\log k_\rho$ is Lipschitz. Then, the velocity V in Eqn. 7 with the cost $c_{k_\rho} = -\varepsilon \log k_\rho$ is identifiable.*

We identify several kernels that satisfy the condition in Cor. 3.3.1 and summarize them in Tab. 2. Accordingly, the Riemannian gradient $\nabla_g^{(1)} c_{k_\rho}$ is obtained by differentiating the truncated sum. Furthermore, the eigenfunctions are spherical harmonics on the sphere and Fourier modes on flat tori, both of which yield closed-form expressions.

3.3.2 Distance-Based Costs

Distance-based costs also fit within our framework: under suitable conditions on the manifold, they too guarantee identifiability. We discuss two such examples: the chordal cost and the geodesic cost.

Table 2: **Spectral densities of nondegenerate spectral kernels.** When used as costs in the Sinkhorn divergence Eqn. 4, these kernels yield identifiable velocity fields. See App. D for details.

Kernel	$\rho(\lambda_n)$	Parameters
Matérn	$\sigma^2 \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^{-\nu - d/2}$	$\nu > \frac{1}{2},\ \kappa, \sigma^2 > 0$
Heat	$e^{-t\lambda_n}$	$t > 0$
Sub. Heat	$e^{-t\lambda_n^\alpha}$	$t > 0,\ \alpha \in (0, 1]$

Chordal Cost. When $\mathcal{M}$ is embedded in $\mathbb{R}^D$, the chordal kernels $\exp\left(-\lambda \|x - y\|^2\right)$ (Steinwart, 2001) are known to be positive definite and universal for any compact subset of $\mathbb{R}^D$. In particular, they are nondegenerate by Prop. 3.3.1. Using the relation between the cost and its Gibbs kernel in Thm. 3.2, we identify the induced cost as the squared chordal distance $\|x - y\|^2$ restricted to $\mathcal{M}$, which is Lipschitz on a compact manifold and symmetric, making it identifiable.

Geodesic Cost. The Laplacian kernel $\exp(-\lambda d_g(x, y))$ is universal on the sphere ($\mathbb{S}^d$) for every $\lambda > 0$ (Zhang et al., 2024). By Prop. 3.3.1, it is nondegenerate. With $\lambda = 1/\varepsilon$, it is the Gibbs kernel of the geodesic distance itself,

$$c(x, y) = d_g(x, y), \tag{12}$$

which is a symmetric, Lipschitz cost on $\mathbb{S}^d$. Therefore, identifiability follows from Thm. 3.2, in contrast to the squared geodesic cost.

4 Related Work

Generative Modeling on Manifolds. Early work extended normalizing flows to non-Euclidean domains, including hyperbolic spaces, spheres, and tori (Bose et al., 2020; Rezende et al., 2020; Mathieu & Nickel, 2020). More recent work adapted diffusion and flow-based models to Riemannian manifolds (Huang et al., 2022; De Bortoli et al., 2022; Chen & Lipman, 2024). However, as with their Euclidean counterparts, sampling from these models remains computationally expensive because it requires multiple sequential steps to integrate an SDE or ODE.

Few-Step Generative Models. To reduce the sampling cost of diffusion and flow models, several methods distill pretrained multi-step models into few-step generators (Salimans & Ho, 2022; Yin et al., 2024; Zhou et al., 2024). However, these approaches require an additional distillation stage on top of teacher pretraining. Other methods instead train few-step generators from scratch by incorporating the underlying SDE or ODE dynamics directly into the training objective (Song & Dhariwal, 2024;

Boffi et al., 2025; Geng et al., 2025). Drifting models (Deng et al., 2026) follow this paradigm to learn one-step generators from scratch. During training, they evolve the model distribution along a velocity field that characterizes its discrepancy from the target distribution. W-Flow (Han et al., 2026) further introduces an identifiable velocity field, improving empirical performance while providing stronger guarantees that the learned distribution converges to the target.

Few-step Generative Modeling on Manifolds. Several recent works accelerate generation on manifolds by adapting few-step techniques originally developed for Euclidean settings. Woo et al. (2026) and Davis et al. (2026) learn shortcut mappings that avoid repeatedly integrating an ODE or SDE, while Cheng et al. (2025) extend consistency models (Song et al., 2023) to manifolds. Concurrently, Esteban-Casadevall et al. (2026) generalize drifting models (Deng et al., 2026) by representing the drifting field as a weighted average of kernel gradients, extending them to arbitrary manifolds. They show that this field corresponds to the WGF of the KL divergence between smoothed probability densities. In contrast, our work adopts the Sinkhorn divergence as the primary energy functional and additionally investigates the squared Maximum Mean Discrepancy (MMD) in App. C.

5 Experiments

5.1 Experiment Setup

Benchmarks. To demonstrate the effectiveness of our method, we conduct extensive experiments on standard benchmarks spanning several manifolds. Specifically, we evaluate RW-Flow on the task of modeling distributions over a sphere ($\mathbb{S}^2$), using geospatial event data over the Earth (Mathieu & Nickel, 2020). We also consider distributions over torus geometry, with representative examples being protein side chain ($\mathbb{T}^2$) (Lovell et al., 2003), and RNA backbone torsion angles ($\mathbb{T}^7$) (Murray et al., 2003). Furthermore, we consider modeling distributions on general manifolds using mesh-based examples and distributions from Chen & Lipman (2024).

Evaluation Metrics. Across all benchmarks, we assess sample quality using the following metrics: kMMD (Maximum Mean Discrepancy), MMD (Minimum Matching Distance), 1-NNA (1-Nearest-Neighbor Accuracy), and COV (Coverage). Each metric is computed between the test data and the generated samples.

Baselines. We compare our method against state-of-the-art one- and few-step generative models designed for distributions on manifolds. Specifically, we consider Riemannian Flow Matching (RFM) (Chen & Lipman, 2024), Generalised Flow Map (GFM) (Davis et al., 2026), Riemannian MeanFlow (RMF) (Woo et al., 2026), and Riemannian Consistency Model (RCM) (Cheng et al., 2025). We additionally include the concurrent work Kernel-Gradient Drifting Models (KGD) (Esteban-Casadevall et al., 2026) as a baseline. To ensure a fair comparison, we match the number of model parameters, training wall-clock time, and optimizer configurations across all methods. We use the official implementations of each baseline, making only minor modifications to integrate them with our standalone evaluation script. Full experimental details are provided in App. E.

5.2 Sphere – $\mathbb{S}^2$

We evaluate RW-Flow on the task of modeling distributions on spheres. Following prior work (De Bortoli et al., 2022; Chen & Lipman, 2024; Davis et al., 2026), we use the geospatial datasets introduced by Mathieu & Nickel (2020), collected from a variety of sources, including National Geophysical Data Center / World Data Service (NGDC/WDS) (2020a;b), Brakenridge (2017), and NASA Earth Observing System Data and Information System (EOSDIS) (2020). The main quantitative results are reported in Tab. 3. As shown, our method outperforms current state-of-the-art methods across nearly all benchmarks in the one-step setting, with the geodesic and subordinated heat cost variants achieving the strongest overall performance. Notably, both costs induce identifiable velocity fields on the sphere.

Alongside the quantitative evaluation, Fig. 2 presents a qualitative comparison between RW-Flow and GFM (Davis et al., 2026), the strongest baseline. The visual results reflect the gap observed in the quantitative metrics: on benchmarks such as 'Wildfire' and 'Flood', GFM produces overly diffuse samples, whereas RW-Flow closely captures the reference distributions.

5.3 Flat Torus – $\mathbb{T}^d$

We also consider distributions on flat tori, the natural state spaces of torsion angles in protein and RNA that encode their structures. Specifically, we use protein side-chain torsion angles represented

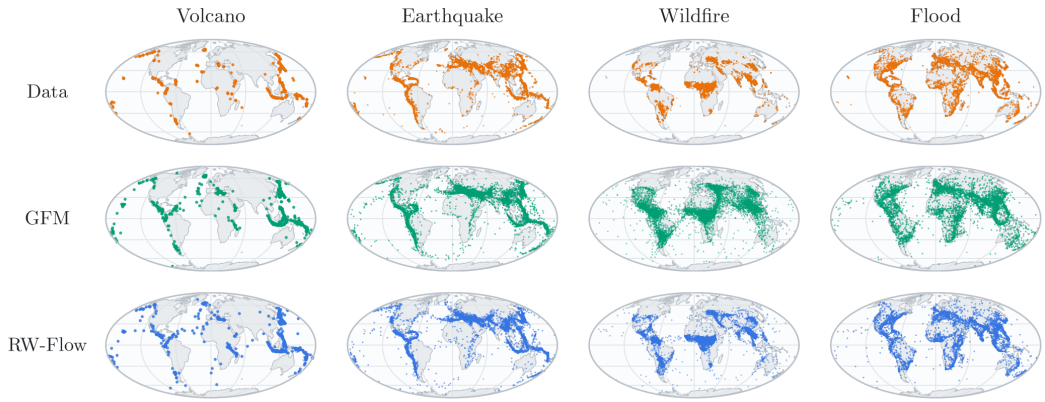


Figure 2: **Generated samples on geospatial event data benchmarks at 1 NFE.** Each column visualizes the dataset (top) and samples generated by GFM (Davis et al., 2026) (middle) and RW-FLOW-S (bottom). RW-FLOW-S more accurately captures the underlying data distributions, whereas GFM produces diffuse samples that deviate substantially from the ground-truth distributions.

Table 3: **Results on geospatial event data benchmarks.** KMMD, COV, and 1-NNA (closest to 0.5 is best) on the volcano, earthquake, flood, and fire datasets (NFE = 1; **bold** = best, underline = second best; the held-out row is the val-vs-test reference and is excluded from ranking.) Suffixes denote variants: CT (Consistency Training) sCT (simplified Consistency Training) L (Lagrangian), E (Eulerian), S (semigroup), MF (mean flow); GG (geodesic Gaussian), GL (geodesic Laplacian), M (Matérn); v denotes v-prediction; RW-FLOW costs: SG (squared geodesic), C (chordal), G (geodesic), H (heat), M (Matérn), S (subordinated). The metrics are averaged over 3 runs. Full results with standard deviations are in Tab. 9.

	Volcano			Earthquake			Flood			Fire		
Method	KMMD↓	COV↑	1-NNA	KMMD↓	COV↑	1-NNA	KMMD↓	COV↑	1-NNA	KMMD↓	COV↑	1-NNA
Held-out	0.153	0.451	0.555	0.026	0.570	0.471	0.025	0.548	0.509	0.024	0.543	0.504
RFM	0.292	0.431	0.835	0.291	0.278	0.845	0.307	0.345	0.790	0.357	0.176	0.913
GFM-L	0.099	0.516	0.553	0.040	**0.570**	0.562	0.062	0.569	0.547	0.029	0.520	0.643
GFM-E	0.199	0.431	0.823	0.174	0.339	0.806	0.191	0.350	0.795	0.190	0.201	0.900
GFM-S	**0.094**	0.512	0.677	0.042	0.478	0.693	0.060	0.498	0.651	0.031	0.392	0.768
GFM-MF	0.119	0.512	0.711	0.043	0.510	0.658	0.059	0.515	0.642	0.033	0.408	0.751
RMF-Lv	0.119	0.504	0.663	0.041	0.535	0.640	0.047	0.545	0.613	0.034	0.453	0.715
RMF-Ev	0.142	0.496	0.689	0.038	0.502	0.653	0.045	0.494	0.661	0.032	0.428	0.740
RMF-Sv	0.122	0.512	0.687	0.040	0.464	0.703	0.045	0.498	0.661	0.033	0.380	0.770
KGD-GG	0.237	0.447	0.799	0.353	0.252	0.866	0.163	0.453	0.700	0.398	0.154	0.926
KGD-GL	0.254	0.435	0.732	0.252	0.353	0.772	0.158	0.460	0.657	0.272	0.291	0.816
KGD-M	0.131	0.488	0.699	0.051	0.485	0.689	0.059	0.521	0.631	0.138	0.332	0.801
RCM-CT	0.147	0.496	0.691	0.055	0.520	0.655	0.052	0.511	0.659	0.045	0.405	0.744
RCM-scT	0.142	0.496	0.811	0.092	0.376	0.796	0.115	0.417	0.751	0.109	0.236	0.871
RW-FLOW-SG	0.110	0.533	0.563	**0.031**	0.552	0.565	0.061	0.558	0.533	**0.024**	0.530	0.643
RW-FLOW-C	0.112	**0.545**	0.612	0.036	0.538	0.594	0.053	0.553	0.580	0.031	0.475	0.688
RW-FLOW-G	0.103	0.484	0.510	0.047	0.556	**0.520**	0.044	0.551	**0.530**	0.027	**0.554**	**0.583**
RW-FLOW-H	0.114	0.516	0.587	0.040	0.527	0.618	0.052	**0.587**	0.591	0.025	0.426	0.730
RW-FLOW-M	0.122	0.524	0.638	0.033	0.529	0.624	0.054	0.522	0.639	**0.024**	0.439	0.738
RW-FLOW-S	**0.094**	0.512	**0.496**	0.040	0.568	0.537	**0.043**	0.563	0.535	0.028	0.553	0.592

as points on $\mathbb{T}^2$ (Lovell et al., 2003), and RNA backbone torsion angles on $\mathbb{T}^7$ (Murray et al., 2003). The quantitative results are summarized in Tab. 4. Across nearly all settings, variants of RW-FLOW outperform both the Lagrangian and other self-distillation variants of GFM. Interestingly, RW-FLOW-SG performs comparably to RW-FLOW-H, even though our framework does not guarantee identifiability of the squared geodesic cost for every $\varepsilon > 0$. This is consistent with Prop. 3.3.2: on $\mathbb{T}^d$ the squared geodesic cost is identifiable for all but countably many ε, so a generic choice of ε behaves like an identifiable cost in practice.

Table 4: **Results on protein and RNA torsion-angle datasets.** KMMD, COV, and 1-NNA (closest to 0.5 is best) on four protein and RNA torsion-angle datasets (NFE = 1; **bold** = best, underline = second best; the held-out row is the val-vs-test reference and is excluded from ranking.) Variant suffixes as in Tab. 3. The metrics are averaged over 3 runs. Full results with standard deviations are in Tab. 10.

	General			Glycine			Proline			Prepro			RNA		
Method	KMMD↓	COV↑	1-NNA	KMMD↓	COV↑	1-NNA	KMMD↓	COV↑	1-NNA	KMMD↓	COV↑	1-NNA	KMMD↓	COV↑	1-NNA
Held-out	0.007	0.585	0.499	0.023	0.590	0.491	0.020	0.602	0.489	0.027	0.572	0.514	0.036	0.588	0.504
RFM	0.469	0.191	0.831	0.281	0.314	0.741	0.478	0.160	0.880	0.486	0.225	0.820	0.559	0.205	0.942
GFM-L	**0.018**	0.515	0.561	0.030	0.562	0.523	0.026	0.547	0.545	0.043	0.551	0.531	0.102	0.571	0.572
GFM-E	0.327	0.257	0.777	0.134	0.389	0.681	0.388	0.162	0.878	0.324	0.288	0.794	0.540	0.313	0.888
GFM-S	0.067	0.362	0.690	0.056	0.450	0.633	0.059	0.378	0.699	0.080	0.423	0.672	0.183	0.473	0.756
GFM-MF	0.094	0.405	0.654	0.044	0.500	0.581	0.094	0.421	0.659	0.122	0.469	0.624	0.558	0.248	0.926
RMF-Lv	0.059	0.453	0.610	0.049	0.494	0.592	0.063	0.498	0.586	0.080	0.507	0.594	0.195	0.545	0.669
RMF-Ev	0.105	0.386	0.672	0.047	0.489	0.604	0.091	0.401	0.676	0.106	0.453	0.643	0.336	0.449	0.790
RMF-Sv	0.037	0.415	0.643	0.047	0.474	0.611	0.031	0.369	0.701	0.049	0.511	0.584	0.094	0.498	0.644
KGD-GG	0.181	0.398	0.656	0.119	0.459	0.619	0.029	0.540	0.554	0.071	0.552	0.540	0.560	0.204	0.948
KGD-GL	0.337	0.283	0.754	0.213	0.371	0.693	0.026	**0.576**	0.526	0.380	0.294	0.768	0.559	0.213	0.940
RCM-CT	0.155	0.342	0.707	0.267	0.284	0.768	0.138	0.377	0.715	0.332	0.287	0.791	0.416	0.407	0.809
RCM-sCT	0.402	0.210	0.819	0.331	0.243	0.807	0.464	0.139	0.904	0.351	0.283	0.776	0.544	0.311	0.889
RW-FLOW-SG	0.021	0.547	0.532	0.032	0.570	0.523	0.023	0.571	**0.508**	**0.027**	**0.580**	**0.510**	0.105	0.551	0.519
RW-FLOW-C	0.086	0.483	0.582	0.038	0.567	0.535	**0.018**	0.563	0.524	0.042	0.568	0.516	**0.046**	0.565	0.532
RW-FLOW-G	0.022	**0.562**	**0.519**	0.117	0.472	0.598	0.029	0.574	0.536	0.110	0.526	0.562	0.062	**0.580**	**0.508**
RW-FLOW-H	0.020	0.542	0.533	**0.028**	**0.574**	**0.517**	0.023	0.567	0.532	0.062	0.562	0.518	0.104	0.552	0.520

5.4 GENERAL MANIFOLDS

The spectral costs discussed in Sec. 3.3.1 offer a key advantage over alternative choices: they are applicable to arbitrary compact manifolds. This allows RW-FLOW to extend naturally beyond specific manifold geometries. To demonstrate this capability, we consider the task of learning distributions on triangular meshes, following the setup of Chen & Lipman (2024). As shown in Tab. 5, RW-FLOW matches the target distributions comparably to RFM (Chen & Lipman, 2024) using only a single step. In contrast, RFM requires 1,000 Euler integration steps (RFM_{1000}), resulting in a sampling time that is two orders of magnitude longer. Qualitative results from RW-FLOW are provided in Fig. 1.

Table 5: **Results on mesh datasets.** KMMD, COV, 1-NNA (closest to 0.5 is best), and sampling time on the Bunny and Spot datasets (**bold** = best, underline = second best; the held-out row is the val-vs-test reference and is excluded from ranking). RFM is sampled with 1,000 steps of its projected Euler integrator (RFM_{1000}); RW-FLOW-H uses the heat cost and is sampled with NFE = 1.

	Bunny ($k = 10$)				Bunny ($k = 50$)				Spot ($k = 10$)				Spot ($k = 50$)			
Method	KMMD↓	COV↑	1-NNA	Time (s)↓	KMMD↓	COV↑	1-NNA	Time (s)↓	KMMD↓	COV↑	1-NNA	Time (s)↓	KMMD↓	COV↑	1-NNA	Time (s)↓
Held-out	0.020	0.583	0.493	—	0.025	0.583	0.498	—	0.012	0.587	0.500	—	0.014	0.584	0.515	—
RFM_{1000}	0.048	**0.569**	**0.514**	49.5	**0.028**	**0.565**	**0.516**	48.7	**0.016**	**0.561**	**0.522**	54.8	0.023	0.540	**0.549**	54.7
RW-FLOW-H	**0.030**	0.558	0.537	**0.16**	0.034	0.544	0.550	**0.13**	0.018	0.493	0.601	**0.24**	**0.015**	**0.549**	0.559	**0.24**

6 CONCLUSION

We propose RW-FLOW, a one-step generative model guided by the Wasserstein gradient flow of Sinkhorn divergence for learning distributions on compact Riemannian manifolds. RW-FLOW builds on a rigorous theoretical foundation for characterizing cost functions that guarantee identifiability of the induced velocity fields. Our main theoretical result establishes a broad class of such cost functions, providing principled choices for the training objective. Experiments further show that RW-FLOW surpasses recent few-step baselines under matched model capacity and training cost, demonstrating strong empirical performance alongside its theoretical guarantees.

AI USE STATEMENT

In this work, AI tools were used to help the authors in proving mathematical claims, implementing methods, supporting qualitative and thematic analysis, and providing feedback on the research methodology or experiments. We have not used generative AI tools to assist in translation and interpretation of the results. The generation of synthetic datasets and the cleaning or reformatting of datasets were not applicable to this work. We have reviewed all AI-assisted work. The authors have manually verified and refined the code, mathematical claims, as well as parts of the manuscript that are written or edited using AI tools. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

All authors have read and adhere to the ICLR Code of Ethics. This work develops generative modeling methodology and its theory. It involves no human subjects, crowdsourcing, or personally identifiable information. All experiments use publicly available datasets.

REPRODUCIBILITY STATEMENT

To support reproducibility, we will release our code publicly upon acceptance, together with instructions for running it locally and the exact configuration files used for every experiment.

REFERENCES

G.E. Andrews, R. Askey, and R. Roy. *Special Functions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1999. URL https://books.google.co.kr/books?id=kGshpCa3eYwC.

Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 1950.

Iskander Azangulov, Andrei Smolensky, Alexander Terenin, and Viacheslav Borovitskiy. Stationary kernels and Gaussian processes on lie groups and their homogeneous spaces i: the compact case. *Journal of Machine Learning Research*, 25(280):1–52, 2024.

M. Berger, P. Gauduchon, and E. Mazet. *Le spectre d'une variété riemannienne*. Lecture notes in mathematics. Springer-Verlag, 1971. ISBN 9783540054375.

Salomon Bochner. *Harmonic analysis and the theory of probability*. Courier Corporation, 2005.

Nicholas Boffi, Michael Albergo, and Eric Vanden-Eijnden. How to build a consistency model: Learning flow maps via self-distillation. In *NeurIPS*, 2025.

Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, et al. Matérn Gaussian processes on Riemannian manifolds. In *NeurIPS*, 2020.

Joey Bose, Ariella Smofsky, Renjie Liao, Prakash Panangaden, and Will Hamilton. Latent variable modelling with hyperbolic normalizing flows. In *ICML*, 2020.

G. R. Brakenridge. Global active archive of large flood events. Dartmouth Flood Observatory, University of Colorado, 2017. Accessed via http://floodobservatory.colorado.edu/Archives/index.html.

Ricky TQ Chen and Yaron Lipman. Flow matching on general geometries. In *ICLR*, 2024.

Chaoran Cheng, Yusong Wang, Yuxin Chen, Xiangxin Zhou, Nanning Zheng, and Ge Liu. Riemannian consistency model. In *NeurIPS*, 2025.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *NeurIPS*, 2013.

Nathaël Da Costa, Cyrus Mostajeran, and Juan-Pablo Ortega. The Gaussian kernel on the circle and spaces that admit isometric embeddings of the circle. In *International Conference on Geometric Science of Information*. Springer, 2023.

Nathael Da Costa, Cyrus Mostajeran, Juan-Pablo Ortega, and Salem Said. Invariant kernels on Riemannian symmetric spaces: a harmonic-analytic approach. *SIAM Journal on Mathematics of Data Science*, 2025.

Oscar Davis, Michael Albergo, Nicholas Boffi, Michael Bronstein, and Joey Bose. Generalised flow maps for few-step generative modelling on Riemannian manifolds. In *ICLR*, 2026.

Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. In *NeurIPS*, 2022.

Mingyang Deng, He Li, Tianhong Li, Yilun Du, and Kaiming He. Generative modeling via drifting. *arXiv preprint arXiv:2602.04770*, 2026.

Maria Esteban-Casadevall, Jorge Carrasco-Pollo, Max Welling, Jan-Willem van de Meent, Erik J Bekkers, and Floor Eijkelboom. Kernel-gradient drifting models. *arXiv preprint arXiv:2605.10727*, 2026.

Aasa Feragen, Francois Lauze, and Soren Hauberg. Geodesic exponential kernels: When curvature and linearity conflict. In *CVPR*, 2015.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and mmd using Sinkhorn divergences. In *AISTATS*, 2019.

Aude Genevay, Gabriel Peyré, and Marco Cuturi. Learning generative models with Sinkhorn divergences. In *AISTATS*, 2018.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *NeurIPS*, 2025.

A. Grigoryan. *Heat Kernel and Analysis on Manifolds*. AMS/IP studies in advanced mathematics. American Mathematical Society, 2009. ISBN 9780821893937.

Jiaqi Han, Puheng Li, Qiushan Guo, Renyuan Xu, Stefano Ermon, and Emmanuel J Candès. One-step generative modeling via Wasserstein gradient flows. *arXiv preprint arXiv:2605.11755*, 2026.

Elton P. Hsu. Estimates of derivatives of the heat kernel on a compact riemannian manifold. *Proceedings of the American Mathematical Society*, 127(12), 1999.

Chin-Wei Huang, Milad Aghajohari, Joey Bose, Prakash Panangaden, and Aaron Courville. Riemannian diffusion models. In *NeurIPS*, 2022.

John M Lee. *Introduction to Riemannian manifolds*, volume 2. Springer Cham, 2018.

Siran Li. Gaussian kernels on nonsimply connected closed riemannian manifolds are never positive definite. *Bulletin of the London Mathematical Society*, 2024.

Simon C Lovell, Ian W Davis, W Bryan Arendall III, Paul IW De Bakker, J Michael Word, Michael G Prisant, Jane S Richardson, and David C Richardson. Structure validation by cα geometry: ϕ, ψ and cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, 2003.

Emile Mathieu and Maximilian Nickel. Riemannian continuous normalizing flows. *NeurIPS*, 2020.

Charles A. Micchelli, Yuesheng Xu, and Haizhang Zhang. Universal kernels. *Journal of Machine Learning Research*, 2006.

Laura JW Murray, W Bryan Arendall III, David C Richardson, and Jane S Richardson. RNA backbone is rotameric. *Proceedings of the National Academy of Sciences*, 2003.

NASA Earth Observing System Data and Information System (EOSDIS). Active fire data. Land, Atmosphere Near real-time Capability for EOS (LANCE) system operated by NASA's Earth Science Data and Information System (ESDIS), 2020. Accessed via `https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/active-fire-data`.

National Geophysical Data Center / World Data Service (NGDC/WDS). NCEI/WDS global significant earthquake database. NOAA National Centers for Environmental Information, 2020a. Accessed via https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ngdc.mgg.hazards:G012153.

National Geophysical Data Center / World Data Service (NGDC/WDS). NCEI/WDS global significant volcanic eruptions database. NOAA National Centers for Environmental Information, 2020b. Accessed via https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ngdc.mgg.hazards:G10147.

Giuseppe Patané. wfem heat kernel: Discretization and applications to shape analysis and retrieval. *Computer Aided Geometric Design*, 30(3):276–295, 2013. ISSN 0167-8396.

Gabriel Peyré and Marco Cuturi. Computational optimal transport with applications to data sciences. *Foundations and Trends in Machine Learning*, 2019.

Aaditya Ramdas, Nicolás García Trillos, and Marco Cuturi. On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy*, 19, 2017.

Danilo Jimenez Rezende, George Papamakarios, Sébastien Racaniere, Michael Albergo, Gurtej Kanwar, Phiala Shanahan, and Kyle Cranmer. Normalizing flows on tori and spheres. In *ICML*, 2020.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *ICLR*, 2022.

Tim Salimans, Han Zhang, Alec Radford, and Dimitris Metaxas. Improving GANs using optimal transport. In *ICLR*, 2018.

René L. Schilling, Renming Song, and Zoran Vondracek. *Bernstein Functions*. De Gruyter, Berlin, New York, 2010. ISBN 9783110215311. doi: doi:10.1515/9783110215311.

Alex Smola, Arthur Gretton, Le Song, and Bernhard Schölkopf. A Hilbert space embedding for distributions. In *International conference on algorithmic learning theory*. Springer, 2007.

Yang Song and Prafulla Dhariwal. Improved techniques for training consistency models. In *ICLR*, 2024.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. *ICML*, 2023.

Bharath K. Sriperumbudur, Kenji Fukumizu, and Gert R.G. Lanckriet. Universality, characteristic kernels and RKHS embedding of measures. *Journal of Machine Learning Research*, 2011. URL http://jmlr.org/papers/v12/sriperumbudur11a.html.

Ingo Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of machine learning research*, 2001.

C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008. ISBN 9783540710509. URL https://books.google.co.kr/books?id=hV8o5R7_5tkC.

Dongyeop Woo, Marta Skreta, Seonghyun Park, Kirill Neklyudov, and Sungsoo Ahn. Riemannian MeanFlow. In *ICML*, 2026.

Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Fredo Durand, William T Freeman, and Taesung Park. One-step diffusion with distribution matching distillation. In *CVPR*, 2024.

Qi Zhang, Lingzhou Xue, and Bing Li. Dimension reduction for Fréchet regression. *Journal of the American Statistical Association*, 2024.

Mingyuan Zhou, Huangjie Zheng, Zhendong Wang, Mingzhang Yin, and Hai Huang. Score identity distillation: Exponentially fast distillation of pretrained diffusion models for one-step generation. In *ICML*, 2024.

APPENDIX

This appendix is organized as follows. App. A summarizes the notation used throughout the paper, and App. B contains the proofs deferred from the main text. App. C discusses alternative energy functionals, such as the squared Maximum Mean Discrepancy (MMD) and the KL divergence over smoothed densities as in Esteban-Casadevall et al. (2026). App. D examines the spectral kernels used in this work (Heat, Matérn, and Subordinated Heat) in greater depth. Finally, App. E details our experiments, including the experimental setup, training and inference algorithms, and full results with standard deviations.

A NOTATION

Tab. 6 collects the notation used throughout the paper. For an in-depth introduction to Riemannian Geometry, refer to Lee (2018) or the appendix of Woo et al. (2026). For optimal transport and Wasserstein gradient flows on Riemannian manifolds we refer to Villani (2008), and for the entropy-regularized transport and the Sinkhorn divergence to Feydy et al. (2019).

Table 6: **Notations used in the paper**.

Symbol	Meaning
Manifold and geometry	
$(\mathcal{M}, g)$	compact, connected, smooth Riemannian manifold with metric g; $\mathbb{S}^d$ and $\mathbb{T}^d$ denote the unit sphere and the flat torus of dimension d
$\mathbb{R}^D$	ambient Euclidean space in which $\mathcal{M}$ is embedded for the network parametrization
$T_x\mathcal{M}$	tangent space of $\mathcal{M}$ at x; $\langle\cdot,\cdot\rangle_g$ and $\|\cdot\|_g$ are the metric and its norm on $T_x\mathcal{M}$
$\exp_x, \log_x$	exponential map $T_x\mathcal{M} \to \mathcal{M}$ at x and its inverse on the domain where it is defined
$d_g(x, y)$	geodesic distance on $(\mathcal{M}, g)$
∇_g, div_g	Riemannian gradient and divergence; $\nabla_g^{(1)}$ is the gradient with respect to the first argument of a function on $\mathcal{M} \times \mathcal{M}$
vol_g	Riemannian volume measure
P_x	orthogonal projection $\mathbb{R}^D \to T_x\mathcal{M}$
(λ_n, ϕ_n)	eigenpairs of the Laplace–Beltrami operator, with $\{\phi_n\}$ orthonormal in $L^2(\mathcal{M})$
$\hat{k}_j$	spectral coefficient of a kernel k diagonalized by $\{\phi_j\}$ (Lem. 3.3)
Measures and transport	
$\mathcal{P}(\mathcal{M})$	probability measures on $\mathcal{M}$
$\mathfrak{M}_b(\mathcal{M})$	finite signed measures on $\mathcal{M}$; $\|\sigma\|$ is the total-variation measure of σ
p, q_0, q_θ	target distribution, base distribution, and model distribution $q_\theta = (f_\theta)_\# q_0$, where $\#$ denotes the pushforward
$c(x, y)$	cost function on $\mathcal{M} \times \mathcal{M}$, assumed symmetric and Lipschitz
ε	entropic regularization strength
$k = e^{-c/\varepsilon}$	Gibbs kernel of the cost c
$k * \sigma$	kernel convolution operator, $(k * \sigma)(x) = \int k(x, y)\ \mathrm{d}\sigma(y)$
$\Pi(q, p)$	couplings with marginals q and p
$\mathrm{OT}_{\varepsilon,c}$, $\mathcal{S}_{\varepsilon,c}$	entropy-regularized optimal transport cost Eqn. 4 and Sinkhorn divergence Eqn. 3
π^{qp}	optimal coupling of $\mathrm{OT}_{\varepsilon,c}(q, p)$; $\pi^{qp}(\,\mathrm{d}y \mid x)$ is its disintegration with respect to x
u, v, w	dual potentials of $\mathrm{OT}_{\varepsilon,c}(q, p)$ in the first and second argument, and the symmetric dual potential of $\mathrm{OT}_{\varepsilon,c}(q, q)$
D_{KL}	Kullback–Leibler divergence
$\mathcal{F}$, $\frac{\delta\mathcal{F}}{\delta q}$	energy functional on $\mathcal{P}(\mathcal{M})$ and its first variation
$V_{q,p}$, $V_{q,p}^{\varepsilon,c}$	velocity field of the Wasserstein gradient flow of $\mathcal{F}$
k_ρ, ρ	spectral kernel Eqn. 10 and its spectral density; $c_{k_\rho} = -\varepsilon \log k_\rho$ is the induced spectral cost
Γ	set of ε for which the geodesic Gaussian kernel is degenerate (Prop. 3.3.2)
Training	
f_θ, v_θ	one-step generator $\mathcal{M} \to \mathcal{M}$ and the velocity network $\mathcal{M} \to \mathbb{R}^D$, with parameters θ
z	latent sample, $z \sim q_0$
η	step size of the update $\exp_x(\eta V(x))$
$\mathcal{L}$	training loss Eqn. 5

Symbol	Meaning
`stopgrad`, sg	stop-gradient operator
x_i, x_l', y_j	generated samples, a second independent batch of generated samples, and data samples
$\hat{q}_\theta$, $\hat{q}'_\theta$, $\hat{p}$	empirical measures of the three batches, with weights a_i, a'_l, b_j
App. C	
$\mathcal{H}_k$	reproducing kernel Hilbert space of a positive definite kernel k
m_μ	kernel mean embedding of a measure μ
$\tilde{p}_k$	kernel-smoothed density of p

B PROOFS

Unless otherwise stated, we assume $\mathcal{M}$ to be a compact, connected Riemannian manifold, c to be symmetric and Lipschitz continuous, and set $k := e^{-c/\varepsilon}$.

B.1 PROOF OF THEOREM 3.1

Here we derive the main formula following well-established results. The general cost function and the manifold setting barely change the computation from the Euclidean case.

Theorem 3.1. *Let $\mathcal{F}(q) = \mathcal{S}_{\varepsilon,c}(q,p)$ on a compact $(\mathcal{M}, g)$ and c be a differentiable in its first argument. Then the Wasserstein gradient flow velocity $V = -\nabla_g \frac{\delta \mathcal{F}}{\delta q}$ is*

$$V^{\varepsilon,c}_{q,p}(x) \;=\; -\int_{\mathcal{M}} \nabla_g^{(1)} c(x,y)\, \pi^{qp}(\,\mathrm{d}y \mid x) \;+\; \int_{\mathcal{M}} \nabla_g^{(1)} c(x,z)\, \pi^{qq}(\,\mathrm{d}z \mid x) \;\in\; T_x\mathcal{M}, \tag{7}$$

where $\pi^{qp}(\,\mathrm{d}y \mid x)$ denotes the disintegration with respect to x of the optimal coupling π^{qp} of the entropy-regularized $\mathrm{OT}_{\varepsilon,c}(q,p)$ in Eqn. 4, and $\nabla_g^{(1)}$ is the Riemannian gradient in the first argument.

Proof. Start by considering the dual problem of the entropy-regularized OT (Peyré & Cuturi, 2019),

$$\mathrm{OT}_{\varepsilon,c}(q,p) \;=\; \max_{u,v \in C(\mathcal{M})} \underbrace{\int u \,\mathrm{d}q + \int v \,\mathrm{d}p - \varepsilon \iint \left(e^{(u \oplus v - c)/\varepsilon} - 1\right) \mathrm{d}q \,\mathrm{d}p}_{=: D(u,v;q,p)}. \tag{13}$$

where $(u \oplus v)(x,y) = u(x) + v(y)$, and the dual potentials u and v satisfy

$$u(x) = -\varepsilon \log \int_{\mathcal{M}} e^{(v(y) - c(x,y))/\varepsilon} \,\mathrm{d}p(y), \qquad v(y) = -\varepsilon \log \int_{\mathcal{M}} e^{(u(x) - c(x,y))/\varepsilon} \,\mathrm{d}q(x). \tag{14}$$

Furthermore, the optimal transport plan solving Eqn. 4 is given by $\pi = \exp\left(\frac{1}{\varepsilon}(u \oplus v - c)\right) \cdot (q \otimes p)$ and that $\int_{\mathcal{M}} \exp\left(\frac{1}{\varepsilon}(u \oplus v - c)\right) \mathrm{d}p(y) = 1$ for p-a.e. x. Then the envelope theorem lets us differentiate at the optimum holding (u, v) fixed:

$$\int \frac{\delta \mathrm{OT}_{\varepsilon,c}(\cdot, p)}{\delta q} \,\mathrm{d}\chi = \frac{\mathrm{d}}{\mathrm{d}s}\Big|_0 D(u, v; q + s\chi, p) \tag{15}$$

$$= \int u \,\mathrm{d}\chi - \varepsilon \int \Big[\int e^{(u \oplus v - c)/\varepsilon} \,\mathrm{d}p(y) - 1 \Big] \,\mathrm{d}\chi(x) \tag{16}$$

$$= \int u \,\mathrm{d}\chi \tag{17}$$

for a signed measure χ with $\chi(\mathcal{M}) = 0$. Thus, the first variation of the Sinkhorn divergence is

$$\frac{\delta \mathcal{S}_{\varepsilon,c}(\cdot, p)}{\delta q} = u - w, \tag{18}$$

where w is the symmetric dual potential of $\mathrm{OT}_{\varepsilon,c}$. Now taking the Riemannian gradient, we have

$$\nabla_g u(x) = -\varepsilon \cdot \frac{\int e^{(v(y)-c(x,y))/\varepsilon}\Big(-\frac{1}{\varepsilon}\nabla_g^{(1)} c(x,y)\Big)\,\mathrm{d}p(y)}{\int e^{(v(y)-c(x,y))/\varepsilon}\,\mathrm{d}p(y)} \tag{19}$$

$$= \frac{\int e^{(v(y)-c(x,y))/\varepsilon}\,\nabla_g^{(1)} c(x,y)\,\mathrm{d}p(y)}{\int e^{(v(y)-c(x,y))/\varepsilon}\,\mathrm{d}p(y)} = \int \nabla_g^{(1)} c(x,y)\pi(\,\mathrm{d}y \mid x) \tag{20}$$

where the third equality comes from multiplying the numerator and denominator by $e^{u(x)/\varepsilon}$ as the denominator becomes 1, whereas the numerator becomes an integral against $e^{(u(x)+v(y)-c(x,y))/\varepsilon}\,\mathrm{d}p(y) = \pi(\,\mathrm{d}y \mid \mathrm{d}x)$. Hence the final result,

$$\nabla_g \frac{\delta \mathcal{S}_{\varepsilon,c}(\cdot,p)}{\delta q} = \nabla_g u - \nabla_g w = \int \nabla_g^{(1)} c(x,y)\ \pi^{qp}(\,\mathrm{d}y \mid x) - \int \nabla_g^{(1)} c(x,z)\ \pi^{qq}(\,\mathrm{d}z \mid x). \tag{21}$$

□

B.2 Proof of Theorem 3.2

We prove the characterization of identifiable velocity fields under mild conditions.

Theorem 3.2. *Let $(\mathcal{M}, g)$ be a compact, connected Riemannian manifold with a symmetric, Lipschitz cost function $c : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$. Denote its Gibbs kernel $k := e^{-c/\varepsilon}$. Then $V_{q,p}^{\varepsilon,c}$ is identifiable if and only if k is nondegenerate.*

We first prove the following useful lemma.

Lemma B.1 (Zero velocity is a kernel equation)**.** *For $p, q \in \mathcal{P}(\mathcal{M})$,*

$$V_{q,p}^{\varepsilon,c} = 0 \ \text{ on } \mathcal{M} \iff k * \big(e^{w/\varepsilon}(p-q)\big) = 0 \ \text{on } \mathcal{M}.$$

Proof. By the proof of Thm. 3.1, $\frac{\delta \mathcal{S}(\cdot,p)}{\delta q} = u - w$, where u and w are the first-argument dual potentials of $\mathrm{OT}_{\varepsilon,c}(q,p)$ and $\mathrm{OT}_{\varepsilon,c}(q,q)$, respectively, and

$$V_{q,p}^{\varepsilon,c} = -\nabla_g \frac{\delta \mathcal{S}(\cdot,p)}{\delta q} = \nabla_g(w-u) = 0 \tag{22}$$

if and only if $w - u = K$ for some constant $K \in \mathbb{R}$ since $\mathcal{M}$ is connected.

($\Rightarrow$) As the dual potentials are unique up to constants (Feydy et al., 2019) we can consider the shifted potentials $(u + K, v - K)$ and assume $w - u = 0$. The dual potentials satisfy (Eqn. 14)

$$e^{-u(x)/\varepsilon} = \int e^{(v(y)-c(x,y))/\varepsilon}\,\mathrm{d}p(y) \tag{23}$$

$$= \int e^{-c(x,y)/\varepsilon} e^{v(y)/\varepsilon}\,\mathrm{d}p(y) = (k * (e^{v/\varepsilon}p))(x). \tag{24}$$

Analogously, since $c(x,y)$, and hence $k(x,y)$, is symmetric,

$$e^{-v(y)/\varepsilon} = \int e^{(u(x)-c(x,y))/\varepsilon}\,\mathrm{d}q(x) = \int k(x,y) e^{u(x)/\varepsilon}\,\mathrm{d}q(x) \tag{25}$$

$$= \int k(y,x) e^{u(x)/\varepsilon}\,\mathrm{d}q(x) = (k * (e^{u/\varepsilon}q))(y) \tag{26}$$

$$e^{-w(x)/\varepsilon} = (k * (e^{w/\varepsilon}q))(x) \tag{27}$$

Furthermore, $u = w$ implies

$$e^{-u/\varepsilon} - e^{-w/\varepsilon} = k * (e^{v/\varepsilon}p) - k * (e^{w/\varepsilon}q) = k * (e^{v/\varepsilon}p - e^{w/\varepsilon}q) = 0, \tag{28}$$

as $*$ is a linear operation. Moreover, we have

$$e^{-v(y)/\varepsilon} = (k * (e^{u/\varepsilon} q))(y) = (k * (e^{w/\varepsilon} q))(y) \tag{29}$$

$$= \int k(y,x) e^{w(x)/\varepsilon} \, \mathrm{d}q(x) = e^{-w(y)/\varepsilon} \tag{30}$$

Hence, we have $v = w$ as well. This makes Eqn. 28

$$k * (e^{w/\varepsilon}(p - q)) = 0. \tag{31}$$

($\Leftarrow$) Suppose $k * \left(e^{w/\varepsilon}(p-q)\right) = 0$. The dual w satisfies

$$e^{-w/\varepsilon} = k * (e^{w/\varepsilon} q) = k * (e^{w/\varepsilon} p). \tag{32}$$

By Prop. 11 of Feydy et al. (2019), (w, w) solves the $\mathrm{OT}_{\varepsilon,c}(q,p)$ and $w = u + K$ for some constant $K \in \mathbb{R}$. □

Proof of Theorem 3.2. ($\Leftarrow$) Suppose k is nondegenerate. By Lem. B.1 and the fact that k is nondegenerate,

$$k * \left(e^{w/\varepsilon}(p-q)\right) = 0 \Rightarrow e^{w/\varepsilon}(p-q) = 0 \Rightarrow p = q \tag{33}$$

as $e^{w/\varepsilon} > 0$.

($\Rightarrow$) To prove the contrapositive, suppose k is degenerate. There exists a finite signed measure $\sigma \neq 0$ such that $k * \sigma = 0$. Let $q = (|\sigma| + \mathrm{vol})/Z \in \mathcal{P}(\mathcal{M})$ where $Z = \mathrm{vol}(\mathcal{M}) + |\sigma|(\mathcal{M})$. Then define $p := q + te^{-w/\varepsilon}\sigma$ and notice

$$\int e^{-w/\varepsilon} d\sigma = \int k * (e^{w/\varepsilon} q) \, \mathrm{d}\sigma = \iint k(x,y) e^{w(y)/\varepsilon} \, \mathrm{d}q(y) \, \mathrm{d}\sigma(x) \tag{34}$$

$$= \iint k(x,y) \, \mathrm{d}\sigma(x) e^{w(y)/\varepsilon} \, \mathrm{d}q(y) = \int (k * \sigma) e^{w/\varepsilon} \, \mathrm{d}q(y) = 0. \tag{35}$$

Since $e^{-w/\varepsilon}$ is continuous on a compact $\mathcal{M}$, it is bounded, so we can choose small $t > 0$ such that $t \sup e^{-w/\varepsilon} < \frac{1}{Z}$ and $p = q + te^{-w/\varepsilon}\sigma = \frac{1}{Z}|\sigma| + \frac{1}{Z}\mathrm{vol} + te^{-w/\varepsilon}\sigma > \frac{1}{Z}\mathrm{vol}$. Hence, we have constructed $p, q \in \mathcal{P}(\mathcal{M})$ such that $k * \left(e^{w/\varepsilon}(p-q)\right) = 0$. By Lem. B.1, $V^{\varepsilon,c}_{q,p} = 0$ but $q \neq p$, so $V^{\varepsilon,c}_{q,p}$ is not identifiable. □

B.3 Proof of Lemma 3.3

Lemma 3.3. *Let $\mathcal{M}$ be compact and k a continuous symmetric kernel on $\mathcal{M}$. Let $\{\phi_n\}_{n\geq 0}$ be Laplace-Beltrami eigenfunctions such that they form a basis for $L^2(\mathcal{M})$ and*

$$\int k(x,y) \, \phi_n(y) \, \mathrm{dvol}_g(y) = \hat{k}_n \, \phi_n(x) \qquad \textit{for all } n \geq 0. \tag{8}$$

Then k is nondegenerate if and only if $\hat{k}_n \neq 0$ for all $n \geq 0$.

Proof. ($\Rightarrow$) Suppose $\hat{k}_j = 0$ for some j. Then,

$$k * (\phi_j \mathrm{vol}_g) = \int k(x,y) \phi_j(y) \, \mathrm{dvol}_g(y) = \hat{k}_j \phi_j = 0. \tag{36}$$

k is degenerate.

($\Leftarrow$) Suppose $\hat{k}_j \neq 0$ for all j. Let $\sigma \in \mathfrak{M}_b(\mathcal{M})$ be a finite signed measure such that $k * \sigma = 0$. By Fubini and the symmetry of k,

$$0 = \int (k * \sigma) \phi_j \, \mathrm{dvol}_g = \iint k(x,y) \, \mathrm{d}\sigma(x) \phi_j(y) \, \mathrm{dvol}_g(y) \tag{37}$$

$$= \iint k(y,x) \phi_j(y) \, \mathrm{dvol}_g(y) \, \mathrm{d}\sigma(x) = \int \hat{k}_j \phi_j \, \mathrm{d}\sigma = \hat{k}_j \int \phi_j \, \mathrm{d}\sigma \tag{38}$$

Thus $\int \phi_j \, \mathrm{d}\sigma = 0$ for every eigenfunction ϕ_j. The span of eigenfunctions is dense in $C^\infty(\mathcal{M})$ (Berger et al., 1971), which is dense in $C(\mathcal{M})$. Now Hahn-Banach density theorem and Riesz Representation theorem implies $\sigma = 0$, so k must be nondegenerate. □

B.4 Proof of Proposition 3.3.2

Proposition 3.3.2. *For the geodesic Gaussian kernel $k_\varepsilon = \exp\big(-\frac{1}{2\varepsilon}d_g^2\big)$ on $\mathbb{S}^d$ or $\mathbb{T}^d$, the set*

$$\Gamma = \{\varepsilon > 0 \mid k_\varepsilon \text{ is degenerate}\} \tag{9}$$

is at most countable. Particularly, for $\mathbb{T}^d$, Γ is countably infinite.

We prove the above for $\mathbb{S}^d$ ($d \geq 2$) and $\mathbb{T}^d$ ($d \geq 1$); the case $\mathbb{S}^1$ will be covered by the second proposition.

Proposition B.1.1 (Sphere)**.** *Let $d \geq 2$ and $k_\varepsilon = e^{-d_g^2/2\varepsilon}$ on $\mathbb{S}^d$. The set Γ of $\varepsilon > 0$ for which k_ε is degenerate is at most countable.*

Proof. Set $\beta := 1/(2\varepsilon)$ and $k_\beta := e^{-\beta d_g^2}$.

Since $d_g(x, y) = \arccos\langle x, y\rangle$, the Funk-Hecke formula (Andrews et al., 1999) gives, for every spherical harmonic Y_ℓ of degree ℓ and with $t = \cos\theta$,

$$k_\beta * (Y_\ell \mathrm{vol}_g) = \gamma_\ell(\beta)\, Y_\ell, \qquad \gamma_\ell(\beta) = |\mathbb{S}^{d-1}| \int_0^\pi e^{-\beta\theta^2}\, g_\ell(\theta)\, \mathrm{d}\theta, \tag{39}$$

$$g_\ell(\theta) := P_{\ell,d+1}(\cos\theta)\, \sin^{d-1}\theta, \tag{40}$$

where $P_{\ell,d+1}$ is the Legendre polynomial of degree ℓ in dimension $d+1$. Spherical harmonics are continuous, so by Lem. 3.3, k_β is nondegenerate if and only if $\gamma_\ell(\beta) \neq 0$ for all $\ell \geq 0$.

Since the integrand is entire in β and jointly continuous on $\mathbb{C} \times [0, \pi]$, γ_ℓ is entire.

Notice that γ_ℓ is the Laplace-transform of the function $G_\ell(s) := \frac{g_\ell(\sqrt{s})}{2\sqrt{s}} 1_{[0,\pi^2]}(s)$ with $s = \theta^2$. By Lerch's theorem, $\gamma_\ell \not\equiv 0$ since $G_\ell(s)$ is not an identically zero function.

Since $\gamma_\ell \not\equiv 0$, by the identity theorem, $Z_\ell := \{\beta > 0 : \gamma_\ell(\beta) = 0\}$ is at most countable. By Eqn. 39 and Lem. 3.3, k_β is degenerate exactly for $\beta \in \bigcup_{\ell \geq 0} Z_\ell$, which is countable. □

Proposition B.1.2 (Torus)**.** *Let $d \geq 1$ and $k_\varepsilon = e^{-d_g^2/2\varepsilon}$ on $\mathbb{T}^d = (\mathbb{R}/2\pi\mathbb{Z})^d$. The set Γ of $\varepsilon > 0$ for which k_ε is degenerate is countably infinite.*

Proof. Set $\beta := 1/(2\varepsilon)$ and $k_\beta := e^{-\beta d_g^2}$ as before.

Here $d_g^2(x, y) = \sum_{i=1}^d |x_i - y_i|_{\mathrm{w}}^2$, with $|\cdot|_{\mathrm{w}}$ the difference wrapped to $[-\pi, \pi]$, so $k_\beta(x, y) = \prod_i e^{-\beta|x_i - y_i|_{\mathrm{w}}^2}$. For $n \in \mathbb{Z}^d$, Fubini's theorem and the substitution $y_i = x_i - \theta_i$ (after centering $x - \pi \leq y \leq x + \pi$) give

$$\int_{\mathbb{T}^d} k_\beta(x, y)\, e^{in\cdot y}\, \mathrm{d}y = e^{in\cdot x} \prod_{i=1}^d \int_{-\pi}^\pi e^{-\beta\theta^2} e^{-in_i\theta}\, \mathrm{d}\theta = e^{in\cdot x} \prod_{i=1}^d h_{n_i}(\beta), \tag{41}$$

$$h_m(\beta) := \int_{-\pi}^\pi e^{-\beta\theta^2} e^{im\theta}\, \mathrm{d}\theta. \tag{42}$$

Taking real and imaginary parts, $\cos(n \cdot x)$ and $\sin(n \cdot x)$ are eigenfunctions of $k_\beta *$ with eigenvalue $\prod_i h_{n_i}(\beta)$. These functions are continuous. Since $h_0 > 0$ and $h_{-m} = h_m$, Lemma 3.3 shows that k_β is nondegenerate if and only if $h_m(\beta) \neq 0$ for all $m \geq 1$.

Substitute $s = \theta\sqrt{\beta} + \frac{mi}{2\sqrt{\beta}}$, and complete the square of the integrand exponent,

$$h_m(\beta) = \frac{e^{-\frac{m^2}{4\beta}}}{\sqrt{\beta}} \int_{-\bar{z}}^z e^{-s^2}\, \mathrm{d}s = \frac{1}{2}\sqrt{\frac{\pi}{\beta}} e^{-\frac{m^2}{4\beta}} \big(\mathrm{erf}(z) - \mathrm{erf}(-\bar{z})\big) = \sqrt{\frac{\pi}{\beta}} e^{-\frac{m^2}{4\beta}}\, \mathrm{Re}\,\mathrm{erf}(z), \tag{43}$$

where $z = \pi\sqrt{\beta} + \frac{mi}{2\sqrt{\beta}}$. Let $H_m(\beta) := \mathrm{Re}\,\mathrm{erf}(z(\beta))$, and since $z^2 = \pi^2\beta - \frac{m^2}{4\beta} + i\pi m$,

$$H_m'(\beta) = \sqrt{\frac{\pi}{\beta}} (-1)^m e^{\frac{m^2}{4\beta} - \pi^2\beta}. \tag{44}$$

Therefore, $H_m(\beta)$ is a monotone function and $H_m(\beta) \to 1$ as $\beta \to \infty$,

$$H_m(\beta) = 1 - (-1)^m \int_\beta^\infty \sqrt{\frac{\pi}{s}} e^{m^2/4s - \pi^2 s} \, \mathrm{d}s =: 1 - (-1)^m J_m(\beta) \tag{45}$$

When m is odd, $H_m(\beta) > 1$. When m is even, H_m increases strictly from $-\infty$ to 1 due to $J_m(\beta) \to \infty$ as $\beta \to 0^+$. Hence, $H_m(\beta)$ has exactly one zero for even m, which in turn leaves $h_m(\beta)$ with exactly one zero.

This implies $|Z_m| := |\{\beta > 0 : h_m(\beta) = 0\}| = 0$ or 1 depending on the parity of m. Moreover, for different $m > m'$, $J_m(\beta) > J_{m'}(\beta)$, so each h_m has unique β_m as their root.

k_β is degenerate exactly for $\beta \in \bigcup_{m \geq 1} Z_m$, which is countably infinite set. □

C Other Energy Functionals

Although our analysis has focused on the Sinkhorn divergence, other energy functionals can also serve as the guiding discrepancy field. In this section, we describe how the framework changes under alternative choices of energy functional and compare their resulting forms and identifiability conditions. Tab. 7 summarizes the energy functionals, their velocity fields, and their identifiability conditions.

We begin by fixing the relevant definitions.

Definition C.1 (Reproducing Kernel Hilbert Space (RKHS))**.** *Let $\mathcal{M}$ be a compact Riemannian manifold. Let $k : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ be a symmetric positive definite kernel. By the Moore-Aronszajn theorem (Aronszajn, 1950), there exists a unique Hilbert space $\mathcal{H}_k$ of functions on $\mathcal{M}$, namely the completion of* $\mathrm{span}\{k(x, \cdot) : x \in \mathcal{M}\}$*, such that $k(x, \cdot) \in \mathcal{H}_k$ for all $x \in \mathcal{M}$ and the reproducing property holds:*

$$\langle f, k(x, \cdot) \rangle_{\mathcal{H}_k} = f(x) \qquad \forall f \in \mathcal{H}_k, \; x \in \mathcal{M}. \tag{46}$$

Definition C.2 (Universal kernel)**.** *Let $\mathcal{M}$ be a compact Riemannian manifold. A continuous positive definite kernel k is* universal *if $\mathcal{H}_k$ is dense in $\mathcal{C}(\mathcal{M})$ with respect to $\| \cdot \|_\infty$ (Steinwart, 2001).*

Definition C.3 (Characteristic kernel)**.** *A measurable positive definite kernel k is* characteristic *if the kernel mean embedding*

$$\mu \mapsto m_\mu := \int_{\mathcal{M}} k(\cdot, y) \, \mathrm{d}\mu(y) \tag{47}$$

is injective on $\mathcal{P}(\mathcal{M})$, where $\int \sqrt{k(y, y)} \, \mathrm{d}\mu(y) < \infty$ is assumed so that $m_\mu \in \mathcal{H}_k$ (Smola et al., 2007).

Remark C.1. The definition of characteristic kernel is similar to the definition of nondegeneracy, but contrary to characteristic kernels, nondegenerate kernels need not be positive definite. Moreover, characteristic kernels operate on $\mathcal{P}(\mathcal{M})$, whereas nondegeneracy is defined on $\mathfrak{M}_b(\mathcal{M})$.

C.1 Squared Maximum Mean Discrepancy (MMD)

Definition C.4. *The squared Maximum Mean Discrepancy (MMD) between two distributions p and q is given by $\frac{1}{2}\,\mathrm{MMD}^2(q, p) = \frac{1}{2}\|m_p - m_q\|_{\mathcal{H}}^2 = \frac{1}{2}\iint k(x, y) \, \mathrm{d}r(x) \, \mathrm{d}r(y)$ for $r = p - q$ and $p, q \in \mathcal{P}(\mathcal{M})$.*

Proposition C.1.1. *The velocity field $V_{q,p}$ induced by $\frac{1}{2}\,\mathrm{MMD}^2$ is identifiable if and only if k is characteristic.*

Proof. Let $\mathcal{F}_{\mathrm{MMD}}(q) = \frac{1}{2}\mathrm{MMD}^2(q, p)$. It can be shown (Han et al., 2026) that

$$\frac{\delta \mathcal{F}_{\mathrm{MMD}}}{\delta q}(q) = m_q - m_p \tag{48}$$

Hence, $V_{q,p}$ has the form

$$V_{q,p} = -\nabla_g(m_q - m_p) = \int \nabla_g k(x, y) \, \mathrm{d}p(y) - \int \nabla_g k(x, y) \, \mathrm{d}q(y). \tag{49}$$

($\Leftarrow$) Suppose k is characteristic. We have $V_{qp} \equiv 0 \Rightarrow m_q - m_p \equiv \text{const}$. Letting $r = q - p$,

$$\|m_q - m_p\|_{\mathcal{H}}^2 = \|m_r\|_{\mathcal{H}}^2 = \iint k(x,y)\,\mathrm{d}r(x)\,\mathrm{d}r(y) = \int m_r\,\mathrm{d}r(y) = m_r \int \mathrm{d}r(y) = 0 \quad (50)$$

since r has a zero mass on the space being the difference of the two probability measures. This gives $m_p = m_q$ and as the mean embedding is injective, we must have $p = q$.

($\Rightarrow$) Suppose identifiability and $m_q = m_p$ for some $q, p \in \mathcal{P}(\mathcal{M})$. We further have

$$V_{q,p} = -\nabla_g(m_q - m_p) \equiv 0 \;\Rightarrow\; q = p. \quad (51)$$

This establishes k is characteristic.

□

C.2 Smoothed KL Divergence

This case is treated by Esteban-Casadevall et al. (2026), but for the sake of completeness, we include the details here.

Definition C.5 (Smoothing)**.** *Let k be a positive normalized kernel such that $\int k(x,y)\,\mathrm{dvol}_g(x) = 1$. Denote the kernel-smoothed density by*

$$\tilde{p}(x) = \int k(x,y)\,\mathrm{d}p(y). \quad (52)$$

Remark C.2. As a function, $\tilde{p}$ is the same as the mean embedding of p to RKHS of k. However, here we treat $\tilde{p}(x)$ as a density of a probability measure with respect to the base measure dvol_g, which is true since the kernel is normalized.

In contrast with our framework, Esteban-Casadevall et al. (2026) considers a Wasserstein Gradient Flow from $\tilde{q}$. In particular, the Wasserstein Gradient Flow corresponding to their setting is

$$\partial_t r_t = \mathrm{div}_g\left(r_t \nabla_g \frac{\delta \mathcal{F}^{KL}}{\delta r}(r_t)\right), \quad r_0 := \tilde{q} \quad (53)$$

where $\mathcal{F}^{KL}(q) := D_{\mathrm{KL}}(q \,\|\, \tilde{p})$, i.e., their guiding functional is the regular KL divergence between the smoothed densities $\tilde{q}$ and $\tilde{p}$.

Proposition C.2.1. *Let k be a normalized kernel such that $k > 0$. The velocity field $V_{\tilde{q},\tilde{p}}$ induced by $\mathcal{F}^{KL}$ is identifiable if and only if k is characteristic.*

Proof. Define $\mathcal{F}^{\mathrm{KL}}$ as in the above. It can be shown that (Han et al., 2026)

$$\frac{\delta \mathcal{F}^{\mathrm{KL}}}{\delta q}(\tilde{q}) = -\log \tilde{p} + \log \tilde{q}. \quad (54)$$

Thus, we get

$$V_{\tilde{q},\tilde{p}}(x) = -\nabla_g \frac{\delta \mathcal{F}^{\mathrm{KL}}}{\delta q}(\tilde{q})(x) = \nabla_g \log \frac{\tilde{p}(x)}{\tilde{q}(x)}. \quad (55)$$

($\Leftarrow$) Suppose k is characteristic. This means when $V_{q,p} \equiv 0$,

$$\nabla_g \log \tilde{p}(x) = \nabla_g \log \tilde{q}(x). \quad (56)$$

Since k is a normalized kernel,

$$\tilde{p}(x) = \tilde{q}(x). \quad (57)$$

As k is characteristic, we have $q = p$.

($\Rightarrow$) Assume identifiability. Let p, q be probability measures such that $\tilde{p} = \tilde{q}$. Then,

$$V_{\tilde{q},\tilde{p}} = \nabla_g \log \frac{\tilde{p}(x)}{\tilde{q}(x)} = 0. \quad (58)$$

By identifiability, $q = p$, establishing k is characteristic.

□

Table 7: **Summary of velocity terms and identifiability conditions for different energy functionals.** The velocity field of an energy functional is $V_{q,p} = T_{q,p} - T_{q,q}$.

Energy Functional	$\boldsymbol{T_{q,p}(x)}$	**Identifiability Condition**
Sinkhorn	$-\int \nabla_g^{(1)} c(x,y)\, \pi^{qp}(\,\mathrm{d}y \mid x)$	k nondegenerate, c symmetric Lipschitz
MMD	$\int \nabla_g^{(1)} k(x,y)\ \mathrm{d}p(y)$	k characteristic
Smoothed KL (KGD)	$\frac{\int \nabla_g^{(1)} k(x,y)\ \mathrm{d}p(y)}{\int k(x,y)\ \mathrm{d}p(y)}$	k characteristic, normalized

We summarize the relationships between proposed energy functionals in Tab. 7. Overall, both $\frac{1}{2}\mathrm{MMD}^2$ and smoothed KL divergence can achieve identifiability whenever k is characteristic, whereas the Sinkhorn divergence requires the Gibbs kernel to be nondegenerate (Thm. 3.2), which for positive definite kernels is universality (Prop. 3.3.1). Smola et al. (2007) proves that if k is a universal kernel, it must be characteristic.

D Discussion on Spectral Kernels

Recall that given a compact manifold $\mathcal{M}$ and its eigenpairs $\{\lambda_n, \phi_n\}_{n\geq 0}$ of the Laplace-Beltrami operator such that they form an orthogonal basis for $L^2(\mathcal{M})$, the spectral kernels are defined as

$$k_\rho(x,y) = \sum_{n=0}^{\infty} \rho(\lambda_n)\phi_n(x)\phi_n(y), \tag{59}$$

where $\rho(\lambda_n)$ is the corresponding spectral density.

Heat Kernel. If we set $\rho_H(\lambda) = e^{-\lambda t}$, the resulting expression is the heat kernel, the fundamental solution of the heat equation,

$$\frac{\partial}{\partial t} K(x,y,t) = \Delta_g^{(1)} K(x,y,0), \quad K(x,y,0) = \delta(x-y) \tag{60}$$

where $\Delta_g^{(1)}$ is the Laplace-Beltrami operator acting on the second argument, and the initial condition $\delta(x-y)$ is the Dirac delta function.

Heat kernels are always positive due to the strong maximum/minimum principle. Moreover, the score of heat kernel is bounded (Hsu, 1999),

$$\|\nabla_g^{(1)} \log K(x,y,t)\|_g \leq C_1 \left(\frac{d_g(x,y)}{t} + \frac{1}{\sqrt{t}} \right) < \infty \tag{61}$$

and hence $\log K(x,y,t)$ is a Lipschitz function with respect to x and y. This is why heat kernel is a valid kernel in the lens of Thm. 3.2.

Matérn Kernel. When $\rho_M(\lambda) = \sigma^2(\frac{2\nu}{\kappa^2} + \lambda)^{-\nu - d/2}$, we obtain the Matérn kernel (Borovitskiy et al., 2020) with the three parameters σ^2, ν and κ. Here d is the dimension of the underlying space. We may relate Matérn kernels to heat kernels by

$$\text{Matérn}(\sigma^2, \nu, \kappa^2) \to \text{Heat}(\kappa^2/2) \quad \text{as} \quad \nu \to \infty, \tag{62}$$

that is, the heat kernel can be viewed as the limit of the Matérn kernel.

The Matérn kernels are strictly positive because they have a strict positive gamma mixture representation (Azangulov et al., 2024):

$$Q(x,y,t) := k_{\rho_M}(x,y) = \sigma^2 \int_0^\infty t^{\nu-1+d/2} e^{-\frac{2\nu}{\kappa^2}t} K(x,y,t)\,\mathrm{d}t > 0. \tag{63}$$

Recalling the following upper bound for heat kernel (Grigoryan, 2009),

$$K(x,y,t) \le C_2 t^{-d/2} e^{-d_g^2/5t}, \tag{64}$$

we may compute the following bound for the gradient of the heat kernel

$$|\nabla_g^{(1)} K(x,y,t)| = K(x,y,t)|\nabla_g^{(1)} \log K(x,y,t)| \le C_1 C_2 t^{-(d+1)/2}\,(r+1)\,e^{-r^2/5} \le C_3 t^{-(d+1)/2}, \tag{65}$$

where $r = \frac{d_g}{\sqrt{t}}$. Furthermore, letting $Q_t := Q(t,\cdot,\cdot)$,

$$\nabla_g^{(1)} Q_t(x,y) = \frac{\sigma^2}{C_{\nu,\kappa}} \int_0^\infty t^{\nu-1+d/2} e^{-\frac{2\nu}{\kappa^2}t} \nabla_g^{(1)} K(x,y,t)\,\mathrm{d}t \tag{66}$$

$$\le C_4 \int_0^\infty t^{\nu-3/2} e^{-\frac{2\nu}{\kappa^2}t}\,\mathrm{d}t \tag{67}$$

The RHS converges exactly when $\nu - 3/2 > -1 \Leftrightarrow \nu > 1/2$ by the Gamma function's integral representation. As Q_t is a positive continuous function, it attains a nonzero minimum value m_1 on the compact set $\mathcal{M} \times \mathcal{M}$. We finally have

$$|\nabla_g^{(1)} \log Q_t| \le \frac{\sup |\nabla_g^{(1)} Q_t|}{m_1} < \infty. \tag{68}$$

Therefore, Q_t is indeed Lipschitz continuous for $\nu > 1/2$. This confirms Matérn kernel is a valid kernel for Thm. 3.2.

Subordinated Heat Kernel. When $\rho_S(\lambda) = e^{-t\lambda^\alpha}$ with $t > 0$ and $\alpha \in (0,1]$, the resulting kernel is the *subordinated heat kernel* (Bochner, 2005). Let $(S_t)_{t\ge 0}$ be the increasing Lévy process with $\mathbb{E}[e^{-\lambda S_t}] = e^{-t\lambda^\alpha}$, and subordination of the heat semigroup (Schilling et al., 2010) gives the mixture representation

$$P(x,y,t) = \int_0^\infty K(x,y,s)\ \mathrm{d}\zeta_t(s) = \sum_{n\ge 0} \left(\int_0^\infty e^{-s\lambda_n}\ \mathrm{d}\zeta_t(s) \right) \phi_n(x)\phi_n(y) \tag{69}$$

$$= \sum_{n\ge 0} \mathbb{E}[e^{-\lambda_n S_t}]\phi_n(x)\phi_n(y) = \sum_{n\ge 0} e^{-t\lambda_n^\alpha} \phi_n(x)\phi_n(y), \tag{70}$$

where ζ_t is the law of S_t. Thus $P_t := P(t,\cdot,\cdot)$ is a positive mixture of heat kernels. Strict positivity of heat kernel K_s for every $s > 0$ gives $P_t > 0$, and by Eqn. 65 together with

$$\int_0^1 s^{-p}\ \mathrm{d}\zeta_t(s) \le \mathbb{E}[S_t^{-p}] = \frac{1}{\Gamma(p)} \int_0^\infty \lambda^{p-1} e^{-t\lambda^\alpha}\ \mathrm{d}\lambda < \infty \qquad \text{for every } p > 0,$$

so $|\nabla_g P_t|_\infty < \infty$. Finally $m_2 := \min_{M\times M} P_t > 0$ by compactness, hence $|\nabla_g \log P_t| \le \sup_{M\times M} |\nabla_g P_t|/m_2 =: L_t$, which is a Lipschitz constant of $\log P_t$. This confirms again that Thm. 3.2 is applicable.

Computation. When evaluating these kernels, we truncate the series in Eqn. 10 to the eigenpairs with $n \le n_{\max}$. Strictly, truncating the series makes the cost unidentifiable since the kernel then will have a finite rank. However, since the error is directly controlled by $n_{\max}$, we adaptively choose $n_{\max}$ based on the decay rate of the heat kernel being considered (larger t means quick convergence for heat kernels). This is a standard practice in spectral geometry processing and manifold kernel methods (Patané, 2013; Borovitskiy et al., 2020).

E More Experimental Details

E.1 Training & Inference Algorithms

We train the one-step generator $f_\theta : \mathcal{M} \to \mathcal{M}$ using the loss in Eqn. 5, with the velocity field $V_{q_\theta,p}$ in Thm. 3.1 being the target. Any of the costs in Sec. 3.3 can be used; the procedure is identical for all of them, since the cost enters only through the Sinkhorn iterations and the gradient $\nabla_g^{(1)} c$ in Eqn. 7.

Algorithm 1 Training loop for f_θ on $\mathcal{M}$

Require: data distribution p on $\mathcal{M}$, base distribution q_0 on $\mathcal{M}$, regularization ε, source batch size N, data batch size M, number of steps K, number of Sinkhorn iterations L, ground cost c, flow step η, learning rate γ

for $k = 1, \dots, K$ **do**
  Sample $\{z_i\}_{i=1}^N \sim q_0$, $\{z'_l\}_{l=1}^N \sim q_0$, $\{y_j\}_{j=1}^M \sim p$
  $x_i \leftarrow f_\theta(z_i)$, $x'_l \leftarrow \mathrm{sg}\big(f_\theta(z'_l)\big)$
  $V_i^{qp} \leftarrow \textsc{RiemSinkhorn}\big(\{\mathrm{sg}(x_i)\}, \{y_j\}, \varepsilon, L, c\big)$
  $V_i^{qq} \leftarrow \textsc{RiemSinkhorn}\big(\{\mathrm{sg}(x_i)\}, \{x'_l\}, \varepsilon, L, c\big)$
  $v_i \leftarrow V_i^{qp} - V_i^{qq}$
  $\tilde{x}_i \leftarrow \mathrm{sg}\big(\exp_{x_i}(\eta\, v_i)\big)$
  $\mathcal{L}(\theta) \leftarrow \frac{1}{N}\sum_{i=1}^N d_g\big(f_\theta(z_i), \tilde{x}_i\big)^2$
  $\theta \leftarrow \theta - \gamma\, \nabla_\theta \mathcal{L}(\theta)$
end for

Algorithm 2 RiemSinkhorn: coupling-weighted cost gradient under a ground cost

Require: sources $\{x_i\}_{i=1}^N$, targets $\{y_j\}_{j=1}^M$ on $\mathcal{M}$, ε, L, ground cost c

$C_{ij} \leftarrow c(x_i, y_j)$, $a_i \leftarrow 1/N$, $b_j \leftarrow 1/M$
$u \leftarrow 0$, $v \leftarrow 0$ ▷ dual potentials
for $\ell = 1, \dots, L$ **do** ▷ log-domain sinkhorn loop
  $v_j \leftarrow -\varepsilon \log \sum_i a_i \exp\big((u_i - C_{ij})/\varepsilon\big)$
  $u_i \leftarrow -\varepsilon \log \sum_j b_j \exp\big((v_j - C_{ij})/\varepsilon\big)$
end for
$\pi(j \mid i) \leftarrow b_j \exp\big((u_i + v_j - C_{ij})/\varepsilon\big)$
$V_i \leftarrow -\sum_j \pi(j \mid i)\, \nabla_g^{(1)} c(x_i, y_j)$
return $\{V_i\}_{i=1}^N$

In practice, instead of parameterizing the velocity in the tangent space of $\mathcal{M}$, we predict the velocity in the ambient space $\mathbb{R}^D$ and project it onto the tangent space, following (Chen & Lipman, 2024; Davis et al., 2026; Woo et al., 2026). Specifically, let $v_\theta \colon \mathcal{M} \to \mathbb{R}^D$ denote the velocity prediction network. For the network parametrization, we regard $\mathcal{M}$ as embedded in $\mathbb{R}^D$. We then define

$$f_\theta(z) = \exp_z\big(P_z\, v_\theta(z)\big), \tag{71}$$

where $P_z \colon \mathbb{R}^D \to T_z\mathcal{M}$ denotes the orthogonal projection onto the tangent space at z.

The training objective requires estimating the velocity, defined in Eqn. 7, at every training iteration. However, the velocity cannot be evaluated exactly, as this would require integration over all points from q_θ and p, which is computationally expensive. We therefore estimate the velocity field by solving entropy-regularized minibatch OT problems using the Sinkhorn algorithm. In particular, we independently sample $\{x_i\}_{i=1}^N \sim q_\theta$, $\{x'_l\}_{l=1}^N \sim q_\theta$, and sample the target points $\{y_j\}_{j=1}^M \sim p$. The velocity is then estimated as

$$V_{\hat{q}_\theta, \hat{p}}(x_i) = -\frac{1}{a_i}\sum_{j=1}^M \nabla_g^{(1)} c(x_i, y_j)\pi_{ij}^{\hat{q}_\theta \hat{p}} + \frac{1}{a_i}\sum_{l=1}^N \nabla_g^{(1)} c(x_i, x'_l)\pi_{il}^{\hat{q}_\theta \hat{q}'_\theta}, \tag{72}$$

where $\hat{q}_\theta = \sum_{i=1}^N a_i \delta_{x_i}$, $\hat{q}'_\theta = \sum_{l=1}^N a'_l \delta_{x'_l}$, and $\hat{p} = \sum_{j=1}^M b_j \delta_{y_j}$.

The network is updated via gradient descent on Eqn. 5. The complete training algorithm is presented in Alg. 1, which summarizes the full training loop of RW-Flow. We further detail the Sinkhorn routine, invoked within Alg. 1, for estimating the velocity field over a minibatch in Alg. 2.

E.2 EXPERIMENTAL SETUP

Every method (each baseline and each cost variant of RW-FLOW) is trained and evaluated under the protocol summarised in Tab. 8: the same network, the same optimisation settings, the same wall-clock budget, the same mechanical model-selection rule and the same standalone scorer.

Table 8: Experimental setup, shared by every method.

Data	
Datasets	$\mathbb{S}^2$: volcano, earthquake, fire, flood (Mathieu & Nickel, 2020); $\mathbb{T}^2$: Top500 backbone torsion angles (ϕ, ψ) (Lovell et al., 2003) by amino-acid type (General, Glycine, Proline, Pre-Pro); $\mathbb{T}^7$: RNA torsion angles (Murray et al., 2003); all following the conventions of Chen & Lipman (2024).
Coordinates	ambient unit vector on $\mathbb{S}^2$; angles in $[0, 2\pi)$ on $\mathbb{T}^d$
Split	80/10/10 train/validation/test, computed once with a split seed independent of the training seed and read by every implementation; validation for model selection, test once for the reported number
Network	
Architecture	MLP, $\mathrm{Linear}(d_{\mathrm{in}}, H) \to \mathrm{SiLU} \to [\mathrm{Linear}(H, H) \to \mathrm{SiLU}]^{\times 3} \to \mathrm{Linear}(H, d)$; no normalisation, residual connections or time embedding; PyTorch default initialisation
Width H	1024 on $\mathbb{S}^2$; 512 on $\mathbb{T}^2$ and $\mathbb{T}^7$
Input	raw coordinate concatenated with the raw scalar time(s) the objective needs ($d_{\mathrm{in}} = d + n_{\mathrm{times}}$): none for RW-FLOW and KGD, one for RFM and RCM, two for GFM and RMF
Parameters	3.16M ($\mathbb{S}^2$), 0.79M ($\mathbb{T}^2$), 0.80M ($\mathbb{T}^7$); each time input adds H parameters, at most 0.13%
Optimisation	
Optimizer	AdamW, $\beta = (0.9, 0.999)$, $\epsilon = 10^{-8}$, weight decay 0.01
Learning rate	3×10^{-4}, cosine-annealed to 3×10^{-6} over the run, no warm-up
Batch size	1024
Gradient clipping	global norm 1.0
EMA	decay 0.999; EMA weights are used for selection and scoring
Precision	float32, TF32 matmuls
Budget	equal wall-clock time per (method, dataset)
Seeds	0, 1, 2 per (method, dataset)
Model selection	
Rule	No selection; no early stopping
Cadence	40 validations per run at equal wall-clock intervals
Statistic	against a fixed subsample of at most 2048 validation points, averaged over three sample draws at a fixed seed; multi-step methods validated at 100 network evaluations
Checkpoints	The last checkpoints are scored
Evaluation	
Scorer	one standalone script shared by all methods, importing nothing from any method repository; non-finite or off-manifold samples are rejected, not projected
Samples	$n = n_{\mathrm{test}}$ per setting, sampling seed shared by every method; all metrics use the geodesic distance d_g
KMMD	Maximum Mean Discrepancy; square root of the biased V-statistic of the maximum mean discrepancy with kernel $\exp(-d_g^2)$; the primary metric
1-NNA	1 Nearest Neighbor Accuracy; leave-one-out nearest-neighbour two-sample accuracy between samples and test points; 0.5 means indistinguishable; the only metric whose null value is calibrated and independent of n
COV	Coverage; fraction of test points that are the nearest neighbour of some sample; detects mode collapse; below 1 even for a perfect model
MMD	Minimum Matching Distance; mean geodesic distance from each test point to its nearest sample; blind to collapse, read together with COV
Floors	validation split scored against the test split at the same n
Compute	
Hardware	NVIDIA RTX 3090 (24 GB)

E.3 Full Results with Standard Deviations

Here we show more extensive table of the sphere, torus, and mesh runs in Tab. 9, Tab. 10, and Tab. 11.

The MMD flows are competitive on several datasets, but their performance depends strongly on the kernel and dataset: MMDF-L fails on General, and MMDF-G and MMDF-H show unstable performance on RNA for some seeds. Indeed, Han et al. observed that gradients may vanish when the model and target distributions are far apart, resulting in unstable optimization. On the other hand, Sinkhorn divergence does not suffer from this instability, performing well for all datasets.

Table 9: **Results on earth datasets (full).** Mean±s.d. over 3 seeds of κMMD, MMD, COV, and 1-NNA (closest to 0.5 is best) on the volcano, earthquake, flood, and fire datasets (NFE = 1; **bold** = best, underline = second best; the held-out row is the val-vs-test reference and is excluded from ranking.) Suffixes denote variants: CT (Consistency Training) sCT (simplified Consistency Training) L (Lagrangian), E (Eulerian), S (semigroup), MF (mean flow); x/v denote x- and v-prediction; GG (geodesic Gaussian), GL (geodesic Laplacian), M (Matérn); MMDF = Wasserstein gradient flow of $\frac{1}{2}\,\mathrm{MMD}^2$ with G (Gaussian), L (Laplacian), H (heat), M (Matérn) kernels; RW-FLOW costs: SG (squared geodesic), C (chordal), G (geodesic), H (heat), M (Matérn), S (subordinated). Means only are summarised in Table 3.

(a) Volcano and Earthquake

	Volcano				Earthquake			
Method	κMMD↓	MMD↓	COV↑	1-NNA	κMMD↓	MMD↓	COV↑	1-NNA
Held-out	0.153	0.046	0.451	0.555	0.026	0.023	0.570	0.471
RFM	0.292 ± 0.012	0.203 ± 0.026	0.431 ± 0.028	0.835 ± 0.027	0.291 ± 0.007	0.070 ± 0.008	0.278 ± 0.013	0.845 ± 0.016
GFM-L	0.099 ± 0.034	0.066 ± 0.013	0.516 ± 0.019	0.553 ± 0.044	0.040 ± 0.009	**0.028** ± 0.001	**0.570** ± 0.013	0.562 ± 0.010
GFM-E	0.199 ± 0.035	0.144 ± 0.009	0.431 ± 0.019	0.823 ± 0.011	0.174 ± 0.012	0.068 ± 0.006	0.339 ± 0.020	0.806 ± 0.012
GFM-S	0.094 ± 0.025	0.086 ± 0.007	0.512 ± 0.021	0.677 ± 0.037	0.042 ± 0.009	0.039 ± 0.002	0.478 ± 0.011	0.693 ± 0.024
GFM-MF	0.119 ± 0.043	0.091 ± 0.024	0.512 ± 0.012	0.711 ± 0.036	0.043 ± 0.009	0.037 ± 0.001	0.510 ± 0.015	0.658 ± 0.020
RMF-Lx	0.600 ± 0.097	0.567 ± 0.237	0.175 ± 0.093	0.945 ± 0.042	0.442 ± 0.040	0.217 ± 0.077	0.178 ± 0.017	0.898 ± 0.006
RMF-Lv	0.119 ± 0.054	0.084 ± 0.006	0.504 ± 0.028	0.663 ± 0.021	0.041 ± 0.008	0.035 ± 0.000	0.535 ± 0.011	0.640 ± 0.015
RMF-Ex	0.807 ± 0.133	1.004 ± 0.303	0.069 ± 0.031	0.990 ± 0.007	0.644 ± 0.176	0.725 ± 0.284	0.026 ± 0.007	0.990 ± 0.006
RMF-Ev	0.142 ± 0.060	0.088 ± 0.009	0.496 ± 0.025	0.689 ± 0.048	0.038 ± 0.005	0.037 ± 0.000	0.502 ± 0.011	0.653 ± 0.009
RMF-Sx	0.973 ± 0.078	1.671 ± 0.257	0.012 ± 0.000	1.000 ± 0.000	0.810 ± 0.070	1.230 ± 0.152	0.002 ± 0.001	0.999 ± 0.001
RMF-Sv	0.122 ± 0.063	0.085 ± 0.010	0.512 ± 0.012	0.687 ± 0.035	0.040 ± 0.005	0.043 ± 0.001	0.464 ± 0.007	0.703 ± 0.012
KGD-GG	0.237 ± 0.107	0.155 ± 0.057	0.447 ± 0.051	0.799 ± 0.066	0.353 ± 0.016	0.076 ± 0.005	0.252 ± 0.004	0.866 ± 0.004
KGD-GL	0.254 ± 0.128	0.151 ± 0.081	0.435 ± 0.077	0.732 ± 0.164	0.252 ± 0.182	0.065 ± 0.028	0.353 ± 0.174	0.772 ± 0.162
KGD-M	0.131 ± 0.025	0.092 ± 0.001	0.488 ± 0.024	0.699 ± 0.013	0.051 ± 0.003	0.041 ± 0.003	0.485 ± 0.027	0.689 ± 0.020
RCT	0.147 ± 0.060	0.112 ± 0.023	0.496 ± 0.074	0.691 ± 0.019	0.055 ± 0.010	0.036 ± 0.002	0.520 ± 0.004	0.655 ± 0.024
sRCT	0.142 ± 0.023	0.155 ± 0.008	0.496 ± 0.014	0.811 ± 0.022	0.092 ± 0.013	0.061 ± 0.004	0.376 ± 0.014	0.796 ± 0.009
MMDF-G	0.110 ± 0.004	0.073 ± 0.002	0.508 ± 0.007	0.677 ± 0.046	0.041 ± 0.002	0.038 ± 0.002	0.496 ± 0.011	0.657 ± 0.015
MMDF-L	**0.091** ± 0.022	0.057 ± 0.016	0.508 ± 0.014	0.510 ± 0.025	0.036 ± 0.004	0.030 ± 0.001	0.545 ± 0.019	0.571 ± 0.004
MMDF-H	0.102 ± 0.018	0.074 ± 0.006	0.508 ± 0.025	0.679 ± 0.023	0.043 ± 0.005	0.038 ± 0.000	0.506 ± 0.004	0.669 ± 0.010
MMDF-M	0.107 ± 0.007	0.071 ± 0.004	0.541 ± 0.031	0.630 ± 0.029	0.032 ± 0.003	0.038 ± 0.001	0.517 ± 0.005	0.653 ± 0.015
RW-FLOW-SG	0.110 ± 0.018	0.053 ± 0.003	0.533 ± 0.028	0.563 ± 0.023	**0.031** ± 0.007	0.029 ± 0.002	0.552 ± 0.010	0.565 ± 0.016
RW-FLOW-C	0.112 ± 0.033	0.063 ± 0.003	**0.545** ± 0.031	0.612 ± 0.041	0.036 ± 0.004	0.030 ± 0.001	0.538 ± 0.009	0.594 ± 0.003
RW-FLOW-G	0.103 ± 0.013	0.066 ± 0.022	0.484 ± 0.055	0.510 ± 0.030	0.047 ± 0.004	**0.028** ± 0.002	0.556 ± 0.015	**0.520** ± 0.017
RW-FLOW-H	0.114 ± 0.006	0.056 ± 0.001	0.516 ± 0.019	0.587 ± 0.037	0.040 ± 0.006	0.032 ± 0.000	0.527 ± 0.013	0.618 ± 0.017
RW-FLOW-M	0.122 ± 0.015	0.077 ± 0.017	0.524 ± 0.044	0.638 ± 0.042	0.033 ± 0.012	0.034 ± 0.000	0.529 ± 0.003	0.624 ± 0.005
RW-FLOW-S	0.094 ± 0.016	**0.049** ± 0.006	0.512 ± 0.021	**0.496** ± 0.031	0.040 ± 0.002	**0.028** ± 0.002	0.568 ± 0.006	0.537 ± 0.009

(b) Flood and Fire

	Flood				Fire			
Method	κMMD↓	MMD↓	COV↑	1-NNA	κMMD↓	MMD↓	COV↑	1-NNA
Held-out	0.025	0.033	0.548	0.509	0.024	0.009	0.543	0.504
RFM	0.307 ± 0.008	0.073 ± 0.004	0.345 ± 0.013	0.790 ± 0.010	0.357 ± 0.015	0.044 ± 0.004	0.176 ± 0.010	0.913 ± 0.010
GFM-L	0.062 ± 0.020	**0.035** ± 0.001	0.569 ± 0.016	0.547 ± 0.007	0.029 ± 0.003	0.014 ± 0.000	0.520 ± 0.004	0.643 ± 0.009
GFM-E	0.191 ± 0.017	0.085 ± 0.003	0.350 ± 0.008	0.795 ± 0.022	0.190 ± 0.009	0.050 ± 0.005	0.201 ± 0.013	0.900 ± 0.013
GFM-S	0.060 ± 0.022	0.048 ± 0.001	0.498 ± 0.018	0.651 ± 0.014	0.031 ± 0.007	0.022 ± 0.001	0.392 ± 0.004	0.768 ± 0.004
GFM-MF	0.059 ± 0.023	0.045 ± 0.001	0.515 ± 0.010	0.642 ± 0.020	0.033 ± 0.006	0.020 ± 0.001	0.408 ± 0.018	0.751 ± 0.012
RMF-Lx	0.539 ± 0.235	0.244 ± 0.038	0.166 ± 0.024	0.911 ± 0.022	0.613 ± 0.146	0.272 ± 0.245	0.076 ± 0.022	0.966 ± 0.010
RMF-Lv	0.047 ± 0.011	0.045 ± 0.002	0.545 ± 0.022	0.613 ± 0.019	0.034 ± 0.011	0.020 ± 0.002	0.453 ± 0.011	0.715 ± 0.009
RMF-Ex	0.831 ± 0.065	1.122 ± 0.333	0.031 ± 0.009	0.991 ± 0.010	0.868 ± 0.097	1.149 ± 0.096	0.009 ± 0.004	0.997 ± 0.002
RMF-Ev	0.045 ± 0.010	0.046 ± 0.001	0.494 ± 0.003	0.661 ± 0.012	0.032 ± 0.010	0.020 ± 0.000	0.428 ± 0.009	0.740 ± 0.010
RMF-Sx	0.934 ± 0.255	1.585 ± 0.434	0.002 ± 0.000	0.999 ± 0.001	0.841 ± 0.075	1.279 ± 0.026	0.002 ± 0.001	1.000 ± 0.000
RMF-Sv	0.045 ± 0.012	0.049 ± 0.001	0.498 ± 0.007	0.661 ± 0.003	0.033 ± 0.010	0.023 ± 0.001	0.380 ± 0.004	0.770 ± 0.005
KGD-GG	0.163 ± 0.181	0.063 ± 0.025	0.453 ± 0.113	0.700 ± 0.110	0.398 ± 0.008	0.051 ± 0.002	0.154 ± 0.002	0.926 ± 0.006
KGD-GL	0.158 ± 0.184	0.057 ± 0.034	0.460 ± 0.152	0.657 ± 0.153	0.272 ± 0.206	0.038 ± 0.021	0.291 ± 0.223	0.816 ± 0.169
KGD-M	0.059 ± 0.023	0.047 ± 0.001	0.521 ± 0.012	0.631 ± 0.020	0.138 ± 0.189	0.031 ± 0.015	0.332 ± 0.133	0.801 ± 0.091
RCM-CT	0.052 ± 0.017	0.050 ± 0.002	0.511 ± 0.015	0.659 ± 0.030	0.045 ± 0.010	0.023 ± 0.001	0.405 ± 0.012	0.744 ± 0.006
RCM-sCT	0.115 ± 0.021	0.072 ± 0.006	0.417 ± 0.035	0.751 ± 0.021	0.109 ± 0.023	0.036 ± 0.004	0.236 ± 0.011	0.871 ± 0.007
MMDF-G	0.041 ± 0.011	0.048 ± 0.001	0.516 ± 0.017	0.656 ± 0.004	0.022 ± 0.002	0.022 ± 0.002	0.397 ± 0.007	0.755 ± 0.009
MMDF-L	**0.040** ± 0.006	0.036 ± 0.002	0.556 ± 0.022	0.547 ± 0.020	0.024 ± 0.006	0.013 ± 0.000	0.547 ± 0.012	0.609 ± 0.009
MMDF-H	0.042 ± 0.009	0.049 ± 0.003	0.507 ± 0.008	0.648 ± 0.009	0.021 ± 0.005	0.021 ± 0.001	0.405 ± 0.006	0.756 ± 0.009
MMDF-M	0.052 ± 0.004	0.043 ± 0.001	0.534 ± 0.014	0.618 ± 0.015	**0.017** ± 0.003	0.020 ± 0.001	0.426 ± 0.007	0.736 ± 0.002
RW-FLOW-SG	0.061 ± 0.012	**0.035** ± 0.001	0.558 ± 0.017	0.533 ± 0.014	0.024 ± 0.002	0.013 ± 0.000	0.530 ± 0.003	0.643 ± 0.012
RW-FLOW-C	0.053 ± 0.015	0.039 ± 0.003	0.553 ± 0.009	0.580 ± 0.020	0.031 ± 0.002	0.015 ± 0.001	0.475 ± 0.007	0.688 ± 0.018
RW-FLOW-G	0.044 ± 0.001	**0.035** ± 0.002	0.551 ± 0.015	**0.530** ± 0.020	0.027 ± 0.002	**0.011** ± 0.000	**0.554** ± 0.008	**0.583** ± 0.008
RW-FLOW-H	0.052 ± 0.017	0.040 ± 0.001	**0.587** ± 0.004	0.591 ± 0.011	0.025 ± 0.007	0.017 ± 0.001	0.426 ± 0.007	0.730 ± 0.018
RW-FLOW-M	0.054 ± 0.009	0.047 ± 0.003	0.522 ± 0.009	0.639 ± 0.026	0.024 ± 0.004	0.020 ± 0.000	0.439 ± 0.004	0.738 ± 0.012
RW-FLOW-S	0.043 ± 0.012	**0.035** ± 0.001	0.563 ± 0.011	0.535 ± 0.018	0.028 ± 0.004	0.012 ± 0.000	0.553 ± 0.010	0.592 ± 0.012

Table 10: **Results on torus datasets (full).** Mean±s.d. over 3 seeds of κMMD, MMD, COV, and 1-NNA (closest to 0.5 is best) on the four $\mathbb{T}^2$ Ramachandran datasets and the $\mathbb{T}^7$ RNA dataset (NFE = 1; **bold** = best, underline = second best; the held-out row is the val-vs-test reference and is excluded from ranking.) Variant suffixes as in Table 3. Means only are summarised in Table 4.

(a) General and Glycine

	General				Glycine			
Method	κMMD↓	MMD↓	COV↑	1-NNA	κMMD↓	MMD↓	COV↑	1-NNA
Held-out	0.007	0.011	0.585	0.499	0.023	0.046	0.590	0.491
RFM	0.469 ± 0.014	0.022 ± 0.000	0.191 ± 0.009	0.831 ± 0.007	0.281 ± 0.005	0.070 ± 0.005	0.314 ± 0.011	0.741 ± 0.007
GFM-L	0.018 ± 0.003	**0.012** ± 0.000	0.515 ± 0.011	0.561 ± 0.008	0.030 ± 0.008	**0.049** ± 0.000	0.562 ± 0.007	0.523 ± 0.002
GFM-E	0.327 ± 0.004	0.020 ± 0.000	0.257 ± 0.003	0.777 ± 0.002	0.134 ± 0.009	0.069 ± 0.005	0.389 ± 0.003	0.681 ± 0.008
GFM-S	0.067 ± 0.007	0.018 ± 0.002	0.362 ± 0.018	0.690 ± 0.014	0.056 ± 0.005	0.062 ± 0.003	0.450 ± 0.021	0.633 ± 0.017
GFM-MF	0.094 ± 0.016	0.019 ± 0.003	0.405 ± 0.030	0.654 ± 0.025	0.044 ± 0.002	0.058 ± 0.004	0.500 ± 0.011	0.581 ± 0.020
RMF-Lx	0.451 ± 0.076	0.025 ± 0.003	0.184 ± 0.061	0.836 ± 0.052	0.264 ± 0.008	0.077 ± 0.005	0.283 ± 0.015	0.771 ± 0.012
RMF-Lv	0.059 ± 0.004	0.014 ± 0.001	0.453 ± 0.011	0.610 ± 0.011	0.049 ± 0.004	0.053 ± 0.002	0.494 ± 0.011	0.592 ± 0.017
RMF-Ex	0.494 ± 0.001	0.026 ± 0.001	0.149 ± 0.002	0.868 ± 0.003	0.299 ± 0.001	0.089 ± 0.002	0.252 ± 0.007	0.792 ± 0.002
RMF-Ev	0.105 ± 0.014	0.021 ± 0.003	0.386 ± 0.017	0.672 ± 0.013	0.047 ± 0.007	0.059 ± 0.001	0.489 ± 0.010	0.604 ± 0.014
RMF-Sx	0.568 ± 0.077	0.809 ± 0.695	0.112 ± 0.029	0.900 ± 0.028	0.363 ± 0.113	0.552 ± 0.810	0.214 ± 0.091	0.822 ± 0.076
RMF-Sv	0.037 ± 0.003	0.016 ± 0.000	0.415 ± 0.007	0.643 ± 0.009	0.047 ± 0.004	0.057 ± 0.001	0.474 ± 0.018	0.611 ± 0.015
KGD-GG	0.181 ± 0.270	0.036 ± 0.030	0.398 ± 0.217	0.656 ± 0.180	0.119 ± 0.157	0.061 ± 0.021	0.459 ± 0.170	0.619 ± 0.150
KGD-GL	0.337 ± 0.274	0.022 ± 0.008	0.283 ± 0.235	0.754 ± 0.198	0.213 ± 0.143	0.071 ± 0.020	0.371 ± 0.178	0.693 ± 0.163
RCT	0.155 ± 0.011	0.031 ± 0.003	0.342 ± 0.013	0.707 ± 0.013	0.267 ± 0.016	0.084 ± 0.011	0.284 ± 0.014	0.768 ± 0.011
sRCT	0.402 ± 0.014	0.033 ± 0.009	0.210 ± 0.016	0.819 ± 0.014	0.331 ± 0.028	0.265 ± 0.032	0.243 ± 0.025	0.807 ± 0.018
MMDF-G	**0.016** ± 0.004	0.015 ± 0.001	0.509 ± 0.006	0.566 ± 0.005	0.036 ± 0.002	0.051 ± 0.002	0.552 ± 0.005	0.536 ± 0.004
MMDF-L	0.494 ± 0.002	0.026 ± 0.001	0.150 ± 0.003	0.865 ± 0.001	0.041 ± 0.002	**0.049** ± 0.001	0.572 ± 0.002	**0.516** ± 0.006
MMDF-H	**0.016** ± 0.005	0.015 ± 0.000	0.514 ± 0.008	0.561 ± 0.008	0.039 ± 0.002	0.051 ± 0.002	0.555 ± 0.002	0.532 ± 0.013
RW-Flow-SG	0.021 ± 0.006	0.015 ± 0.002	0.547 ± 0.005	0.532 ± 0.007	0.032 ± 0.007	0.050 ± 0.001	0.570 ± 0.003	0.523 ± 0.014
RW-Flow-C	0.086 ± 0.134	0.014 ± 0.002	0.483 ± 0.100	0.582 ± 0.083	0.038 ± 0.003	0.051 ± 0.002	0.567 ± 0.007	0.535 ± 0.018
RW-Flow-G	0.022 ± 0.008	0.014 ± 0.001	**0.562** ± 0.008	**0.519** ± 0.006	0.117 ± 0.146	0.058 ± 0.015	0.472 ± 0.151	0.598 ± 0.138
RW-Flow-H	0.020 ± 0.006	0.013 ± 0.000	0.542 ± 0.021	0.533 ± 0.021	**0.028** ± 0.006	0.050 ± 0.000	**0.574** ± 0.008	0.517 ± 0.008

(b) Proline and Prepro

	Proline				Prepro			
Method	κMMD↓	MMD↓	COV↑	1-NNA	κMMD↓	MMD↓	COV↑	1-NNA
Held-out	0.020	0.027	0.602	0.489	0.027	0.042	0.572	0.514
RFM	0.478 ± 0.038	0.062 ± 0.008	0.160 ± 0.023	0.880 ± 0.020	0.486 ± 0.016	0.082 ± 0.004	0.225 ± 0.005	0.820 ± 0.004
GFM-L	0.026 ± 0.004	0.028 ± 0.000	0.547 ± 0.006	0.545 ± 0.018	0.043 ± 0.014	0.042 ± 0.001	0.551 ± 0.019	0.531 ± 0.009
GFM-E	0.388 ± 0.015	0.065 ± 0.010	0.162 ± 0.013	0.878 ± 0.022	0.324 ± 0.017	0.085 ± 0.001	0.288 ± 0.010	0.794 ± 0.005
GFM-S	0.059 ± 0.011	0.044 ± 0.004	0.378 ± 0.015	0.699 ± 0.010	0.080 ± 0.017	0.060 ± 0.001	0.423 ± 0.023	0.672 ± 0.019
GFM-MF	0.094 ± 0.025	0.037 ± 0.003	0.421 ± 0.024	0.659 ± 0.025	0.122 ± 0.020	0.057 ± 0.004	0.469 ± 0.009	0.624 ± 0.008
RMF-Lx	0.560 ± 0.061	0.209 ± 0.144	0.090 ± 0.045	0.945 ± 0.038	0.346 ± 0.032	0.068 ± 0.002	0.311 ± 0.024	0.757 ± 0.024
RMF-Lv	0.063 ± 0.009	0.029 ± 0.002	0.498 ± 0.019	0.586 ± 0.017	0.080 ± 0.012	0.051 ± 0.002	0.507 ± 0.020	0.594 ± 0.007
RMF-Ex	0.580 ± 0.001	0.116 ± 0.009	0.094 ± 0.004	0.942 ± 0.005	0.561 ± 0.006	0.122 ± 0.008	0.146 ± 0.015	0.901 ± 0.014
RMF-Ev	0.091 ± 0.028	0.040 ± 0.003	0.401 ± 0.013	0.676 ± 0.019	0.106 ± 0.026	0.056 ± 0.003	0.453 ± 0.032	0.643 ± 0.017
RMF-Sx	0.612 ± 0.044	0.202 ± 0.125	0.091 ± 0.009	0.944 ± 0.016	0.637 ± 0.128	0.841 ± 1.257	0.113 ± 0.062	0.923 ± 0.040
RMF-Sv	0.031 ± 0.002	0.042 ± 0.002	0.369 ± 0.019	0.701 ± 0.012	0.049 ± 0.012	0.050 ± 0.003	0.511 ± 0.008	0.584 ± 0.022
KGD-GG	0.029 ± 0.011	0.028 ± 0.001	0.540 ± 0.019	0.554 ± 0.014	0.071 ± 0.048	0.123 ± 0.138	0.552 ± 0.012	0.540 ± 0.007
KGD-GL	0.026 ± 0.007	0.027 ± 0.001	**0.576** ± 0.016	0.526 ± 0.010	0.380 ± 0.301	0.100 ± 0.049	0.294 ± 0.248	0.768 ± 0.220
RCM-CT	0.138 ± 0.021	0.044 ± 0.007	0.377 ± 0.024	0.715 ± 0.013	0.332 ± 0.140	0.083 ± 0.009	0.287 ± 0.069	0.791 ± 0.063
RCM-sCT	0.464 ± 0.032	0.077 ± 0.011	0.139 ± 0.021	0.904 ± 0.011	0.351 ± 0.041	0.094 ± 0.017	0.283 ± 0.052	0.776 ± 0.037
MMDF-G	0.020 ± 0.001	0.029 ± 0.000	0.527 ± 0.018	0.560 ± 0.002	0.037 ± 0.005	0.048 ± 0.005	0.541 ± 0.009	0.553 ± 0.006
MMDF-L	0.021 ± 0.002	0.027 ± 0.001	0.564 ± 0.005	0.524 ± 0.002	**0.025** ± 0.001	**0.040** ± 0.001	**0.588** ± 0.015	**0.505** ± 0.022
MMDF-H	0.020 ± 0.001	0.030 ± 0.000	0.506 ± 0.016	0.580 ± 0.010	0.033 ± 0.004	0.046 ± 0.001	0.541 ± 0.000	0.545 ± 0.016
RW-Flow-SG	0.023 ± 0.001	**0.026** ± 0.001	0.571 ± 0.018	**0.508** ± 0.021	0.027 ± 0.008	0.042 ± 0.001	0.580 ± 0.011	0.510 ± 0.006
RW-Flow-C	**0.018** ± 0.004	0.027 ± 0.001	0.563 ± 0.014	0.524 ± 0.010	0.042 ± 0.005	0.042 ± 0.001	0.568 ± 0.008	0.516 ± 0.021
RW-Flow-G	0.029 ± 0.003	0.028 ± 0.001	0.574 ± 0.014	0.536 ± 0.019	0.110 ± 0.056	0.053 ± 0.008	0.526 ± 0.071	0.562 ± 0.069
RW-Flow-H	0.023 ± 0.003	**0.026** ± 0.001	0.567 ± 0.002	0.532 ± 0.009	0.062 ± 0.052	0.060 ± 0.032	0.562 ± 0.013	0.518 ± 0.013

(c) RNA

Method	RNA			
	κMMD↓	MMD↓	COV↑	1-NNA
Held-out	0.036	0.384	0.588	0.504
RFM	0.559 ± 0.000	1.665 ± 0.157	0.205 ± 0.002	0.942 ± 0.004
GFM-L	0.102 ± 0.013	0.433 ± 0.004	$\underline{0.571} \pm 0.010$	0.572 ± 0.017
GFM-E	0.540 ± 0.002	0.822 ± 0.074	0.313 ± 0.008	0.888 ± 0.006
GFM-S	0.183 ± 0.008	0.493 ± 0.002	0.473 ± 0.012	0.756 ± 0.006
GFM-MF	0.558 ± 0.001	1.195 ± 0.233	0.248 ± 0.021	0.926 ± 0.007
RMF-Lx	0.564 ± 0.007	2.219 ± 0.723	0.150 ± 0.088	0.971 ± 0.024
RMF-Lv	0.195 ± 0.008	0.473 ± 0.003	0.545 ± 0.001	0.669 ± 0.021
RMF-Ex	0.490 ± 0.007	0.660 ± 0.033	0.343 ± 0.007	0.872 ± 0.006
RMF-Ev	0.336 ± 0.008	0.526 ± 0.007	0.449 ± 0.011	0.790 ± 0.008
RMF-Sx	0.560 ± 0.000	1.760 ± 0.160	0.209 ± 0.011	0.939 ± 0.002
RMF-Sv	0.094 ± 0.006	0.441 ± 0.002	0.498 ± 0.001	0.644 ± 0.017
KGD-GG	0.560 ± 0.000	1.759 ± 0.052	0.204 ± 0.007	0.948 ± 0.004
KGD-GL	0.559 ± 0.000	1.580 ± 0.278	0.213 ± 0.003	0.940 ± 0.007
RCM-CT	0.416 ± 0.004	0.522 ± 0.006	0.407 ± 0.004	0.809 ± 0.016
RCM-sCT	0.544 ± 0.001	0.850 ± 0.011	0.311 ± 0.009	0.889 ± 0.001
MMDF-G	0.400 ± 0.276	1.360 ± 0.534	0.290 ± 0.123	0.861 ± 0.136
MMDF-L	$\mathbf{0.043} \pm 0.003$	0.449 ± 0.007	0.565 ± 0.006	0.535 ± 0.013
MMDF-H	0.396 ± 0.283	1.400 ± 0.632	0.308 ± 0.167	0.856 ± 0.146
RW-FLOW-SG	0.105 ± 0.001	0.437 ± 0.009	0.551 ± 0.006	$\underline{0.519} \pm 0.010$
RW-FLOW-C	$\underline{0.046} \pm 0.005$	$\mathbf{0.418} \pm 0.003$	0.565 ± 0.012	0.532 ± 0.012
RW-FLOW-G	0.062 ± 0.004	$\underline{0.428} \pm 0.008$	$\mathbf{0.580} \pm 0.006$	$\mathbf{0.508} \pm 0.011$
RW-FLOW-H	0.104 ± 0.003	0.438 ± 0.008	0.552 ± 0.010	0.520 ± 0.008

d

Table 11: **Mesh results under RFM's own training budget (full).** We report κMMD, MMD, COV, 1-NNA (closest to 0.5 is best), and sampling time on the Bunny and Spot datasets (**bold** = best, underline = second best; the held-out row is the val-vs-test reference and is excluded from ranking). RFM is trained for its shipped default of 100k iterations and sampled with 1,000 steps of its projected Euler integrator (RFM_{1000}); RW-FLOW-H uses the heat cost, and is sampled with NFE $= 1$. Time is the wall time to draw 2000 samples on one RTX 3090.

Method	Bunny ($k = 10$)					Bunny ($k = 50$)				
	κMMD↓	MMD↓	COV↑	1-NNA	Time (s)↓	κMMD↓	MMD↓	COV↑	1-NNA	Time (s)↓
Held-out	0.020	0.014	0.583	0.493	—	0.025	0.014	0.583	0.498	—
RFM_{1000}	$\underline{0.048}$	**0.015**	**0.569**	**0.514**	$\underline{49.5}$	**0.028**	**0.016**	**0.565**	**0.516**	$\underline{48.7}$
RW-FLOW-H	**0.030**	**0.015**	$\underline{0.558}$	$\underline{0.537}$	**0.16**	$\underline{0.034}$	$\underline{0.017}$	$\underline{0.544}$	$\underline{0.550}$	**0.13**

(a) Bunny ($k = 10$) and Bunny ($k = 50$)

Method	Spot ($k = 10$)					Spot ($k = 50$)				
	κMMD↓	MMD↓	COV↑	1-NNA	Time (s)↓	κMMD↓	MMD↓	COV↑	1-NNA	Time (s)↓
Held-out	0.012	0.016	0.587	0.500	—	0.014	0.016	0.584	0.515	—
RFM_{1000}	**0.016**	**0.017**	**0.561**	**0.522**	$\underline{54.8}$	$\underline{0.023}$	**0.017**	$\underline{0.540}$	**0.549**	$\underline{54.7}$
RW-FLOW-H	$\underline{0.018}$	$\underline{0.021}$	$\underline{0.493}$	$\underline{0.601}$	**0.24**	**0.015**	$\underline{0.019}$	**0.549**	$\underline{0.559}$	**0.24**

(b) Spot ($k = 10$) and Spot ($k = 50$)