PROBABILISTIC GEODESIC FLOW MATCHING ON LOCATION-SCALE FAMILIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Flow matching (FM) has recently emerged as a promising framework for generative modeling due to its conceptual simplicity and strong empirical performance. In FM, samples are transported along a vector field parameterized by a neural network, inducing a probability path that evolves from a simple noise distribution to the target data distribution, governed by an ordinary differential equation (ODE). However, existing FM approaches predominantly rely on probability paths derived from optimal transport (OT) between Gaussian distributions, which may be suboptimal for capturing complex data with inhomogeneous structures such as heavy tail or sharp contrast. In this work, we generalize FM to the broader class of location–scale families for handling data inhomogeneity and introduce a novel class of probability paths defined as geodesics on the manifold of probability distributions. We name this approach *probabilistic geodesic flow matching* to distinguish it from prior geodesic (Riemannian) FM methods defined in input space. We argue that Euclidean OT-based paths are not necessarily optimal in probability space and may limit modeling flexibility. Through synthetic benchmarks and scientific datasets at different scales, we demonstrate that the proposed method more effectively captures complex distributions, leading to improved or comparable performance compared with SOTA geometry-motivated generative models.

1 INTRODUCTION

Diffusion models and flow matching methods are two classes of popular generative modeling approaches. Denoising diffusion probabilistic model (Ho et al., 2020) is a discrete Markov model that converts a complex data distribution into a simple tractable distribution by gradually adding noise in a forward process and then trains a generative model to remove the noise in a reverse process. Song et al. (2021) generalize it to a continuum of distributions and reformulate these two processes using SDEs. In the SDE framework, the Fokker–Planck equation governs the evolution of the density path, which induces an ODE that enables a deterministic flow with marginal densities the same as SDE (Maoutsa et al., 2020). The continuous normalizing flow (CNF Chen et al., 2018) directly models the vector field of the probabilistic flow that generates the probability path from noise to data distributions, bypassing the noise-adding forward process. Recently, flow matching (FM Lipman et al., 2023) introduces a diffusion-style training objective to predict the vector field and improves efficiency by avoiding expensive ODE simulation in CNF training.

Despite the connection to diffusion processes, dominant FM methods are based on probability paths as an optimal transport (OT) interpolant (McCann, 1997) between Gaussian distributions. OT paths are known to be geodesics in the 2-Wasserstein space (Benamou & Brenier, 2000; Villani, 2009). They may not necessarily be optimal for complex data distributions, e.g. non-Gaussian distributions with heavy tails. Heavy-tailed priors using student-t distributions have recently been studied in diffusion models (Pandey et al., 2025). In this work, we generalize FM by deviating from the default choice of Gaussian distributions and explore a broader location-scale family. In particular, we focus on exponential power distributions (Gómez et al., 1998), which have the flexibility of imposing regularization and modeling tail probabilities through a parameter $q > 0$ and embrace Gaussian distribution as a special case $q = 2.0$. To search for the optimal probability path on the manifold of exponential power distributions, we adopt the Fisher-Rao metric and solve the geodesic under this metric. The proposed FM with this geodesic path is hence named *probabilistic geodesic flow matching*, to differentiate it from the existing Riemannian (geodesic) FM (Chen & Lipman, 2024).

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.

Connection to Existing Literature Our work directly generalizes the FM method (Lipman et al., 2023). In addition to adopting broader location-scale distributions, our framework, developed from probability path to vector field and then sample flow, is reverse to Lipman et al. (2023), but provides a natural foundation of manifold perspective of probability paths. We want to highlight that the Riemannian FM (Chen & Lipman, 2024) has its geodesics defined in the input space of most applications. However, our *probabilistic* geodesics are defined on the manifold of exponential power and more general location-scale distributions. Fisher-Rao metric is also considered in Cheng et al. (2024) and Davis et al. (2024) for discrete data. Other manifold based FM includes Kapusniak et al. (2024) with a learned metric and Atanackovic et al. (2025) with a 2-Wasserstein. Despite the different applications, their constructions either take advantage of generic spherical geometry or learn from data, while our derivation is more native to exponential power distributions with **analytic** solution (see Table .4 for a more comprehensive comparison). Our work makes multi-fold contributions:

1. It generalizes FM with a broader location-scale family beyond Gaussian distributions.
2. It interprets optimal probability paths as energy-minimizing trajectories in the probability space with appropriate metrics.
3. It proposes probabilistic geodesic FM demonstrating improved or comparable performance.

The remainder of the paper is organized as follows. Section 2 reviews the background on FM and exponential power distributions. Section 3 introduces the flexible FM framework in location-scale family. Then we derive the probabilistic geodesics on the manifold of exponential power distributions in Section 4. We demonstrate the numerical advantages in Section 5 and conclude in Section 6 with some discussion on limitations and future directions.

2 Background Review

2.1 (Conditional) Flow Matching

To learn the distribution $q(\cdot)$ of a given dataset and generate synthetic samples $\mathbf{x} \sim q(\cdot)$, flow-based approaches evolve noisy data $\mathbf{x}_0 \sim p_0(\cdot)$ to clean data $\mathbf{x}_1 \sim p_1(\cdot) \approx q(\cdot)$ along a flow mapping $\phi_t : \mathbb{R}^d \to \mathbb{R}^d$ defined by the following ODE with given vector field $\mathbf{v}_t : \mathbb{R}^d \to \mathbb{R}^d$ for $t \in [0, 1]$:

$$\begin{aligned} \frac{d}{dt}\phi_t(\mathbf{x}) &= \mathbf{v}_t(\phi_t(\mathbf{x})), \\ \phi_0(\mathbf{x}) &= \mathbf{x}_0, \end{aligned} \tag{1}$$

where $\mathbf{x}_t := \phi_t(\mathbf{x}_0)$ follows a probability path $p_t(\cdot)$ given by the push-forward equation

$$p_t = [\phi_t]_\sharp p_0, \quad [\phi_t]_\sharp p_0(\mathbf{x}) = p_0(\phi_t^{-1}(\mathbf{x})) \det \left[\frac{\partial \phi_t}{\partial \mathbf{x}}\right]^{-1}. \tag{2}$$

A vector field $\mathbf{v}_t$ is said to generate the probability path p_t if and only if the continuity equation holds:

$$\partial_t p_t(\mathbf{x}) = -\nabla_{\mathbf{x}} \cdot (\mathbf{v}_t(\mathbf{x}) p_t(\mathbf{x})), \tag{3}$$

which induces the following equation evolving the logarithm of density:

$$\frac{d}{dt} \log p_t(\mathbf{x}_t) = -\mathrm{tr}\left(\frac{\partial \mathbf{v}_t}{\partial \mathbf{x}_t}\right) = -\nabla \cdot \mathbf{v}_t(\mathbf{x}_t). \tag{4}$$

Once the vector field $\mathbf{v}_t$ is learned from the data, a synthetic sample can be generated by solving the ODE (1) up to time $t = 1$: $\mathbf{x}_1 = \phi_1(\mathbf{x}_0) \sim p_1(\cdot) \approx q(\cdot)$.

The FM considers a conditional probability path $p_t(\cdot|\mathbf{x}_1)$ for a fixed data point $\mathbf{x}_1 \in \mathbb{R}^d$ such that $p_0(\cdot) = p_0(\cdot|\mathbf{x}_1)$ and $p_1(\cdot) = \int p_1(\cdot|\mathbf{x}_1) q(\mathbf{x}_1) d\mathbf{x}_1 \approx q(\cdot)$. A common choice is $p_0(\cdot|\mathbf{x}_1) = \mathcal{N}(\cdot; \mathbf{0}, \mathbf{I})$ and $p_1(\cdot|\mathbf{x}_1) = \mathcal{N}(\cdot; \mathbf{x}_1, \sigma_{\min}^2 \mathbf{I})$ for some small $\sigma_{\min} > 0$. A conditional vector field $\mathbf{u}_t(\cdot|\mathbf{x}_1)$ is then constructed to generate this conditional probability path $p_t(\cdot|\mathbf{x}_1)$ by the flow $\frac{d}{dt}\psi_t(\mathbf{x}) = \mathbf{u}_t(\psi_t(\mathbf{x})|\mathbf{x}_1)$. Hence, conditional FM (CFM) trains a neural network $\mathbf{v}_t(\cdot; \theta)$ to match $\mathbf{u}_t(\cdot|\mathbf{x}_1)$ by minimizing the loss:

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t, q(\mathbf{x}_1), p_t(\mathbf{x}|\mathbf{x}_1)} \|\mathbf{v}_t(\mathbf{x}; \theta) - \mathbf{u}_t(\mathbf{x}|\mathbf{x}_1)\|_2^2. \tag{5}$$

where $t \sim \mathcal{U}[0,1]$, $\mathbf{x}_1 \sim q(\cdot)$, and $\mathbf{x} \sim p_t(\cdot|\mathbf{x}_1)$.

Starting with the conditional flow ψ_t as a canonical affine transformation in the family of Gaussian distributions $\mathcal{N}(\boldsymbol{\mu}_t, \sigma_t^2\mathbf{I})$:

$$\psi_t(\mathbf{x}) = \boldsymbol{\mu}_t(\mathbf{x}_1) + \sigma_t(\mathbf{x}_1)\mathbf{x}, \tag{6}$$

Lipman et al. (2023) prove that the corresponding conditional vector field $\mathbf{u}_t(\cdot|\mathbf{x}_1)$ in the following unique format generates the desired conditional probability path $p_t(\mathbf{x}|\mathbf{x}_1)$:

$$\mathbf{u}_t(\mathbf{x}|\mathbf{x}_1) = \dot{\boldsymbol{\mu}}_t(\mathbf{x}_1) + \frac{\dot{\sigma}_t(\mathbf{x}_1)}{\sigma_t(\mathbf{x}_1)}(\mathbf{x} - \boldsymbol{\mu}_t(\mathbf{x}_1)). \tag{7}$$

In particular, Lipman et al. (2023) propose and advocate the conditional flow based on the optimal transport (OT) displacement map ψ pushing from $p_0(\cdot|\mathbf{x}_1)$ to $p_1(\cdot|\mathbf{x}_1)$ (McCann, 1997):

$$\psi_t = (1-t)\mathrm{id} + t\psi. \tag{8}$$

The OT displacement between $\mathcal{N}(\boldsymbol{\mu}_0, \sigma_0^2\mathbf{I})$ and $\mathcal{N}(\boldsymbol{\mu}_1, \sigma_1^2\mathbf{I})$ is known to be $\psi(\mathbf{x}) = \boldsymbol{\mu}_1 + \frac{\sigma_1}{\sigma_0}(\mathbf{x} - \boldsymbol{\mu}_0)$ (Villani, 2009). With the previous Gaussian conditionals $p_0(\cdot|\mathbf{x}_1) = \mathcal{N}(\cdot; \mathbf{0}, \mathbf{I})$ and $p_1(\cdot|\mathbf{x}_1) = \mathcal{N}(\cdot; \mathbf{x}_1, \sigma_{\min}^2\mathbf{I})$, Lipman et al. (2023) develop OT based conditional *sample flow* (ψ_t) $\rightarrow$ *vector field* ($\dot{\psi}_t$) $\rightarrow$ *probability path* ($p_t(\cdot|\mathbf{x}_1)$) in turn as follows:

$$\begin{aligned} \psi_t(\mathbf{x}) = (1 - (1-\sigma_{\min})t)\mathbf{x} + t\mathbf{x}_1, \quad \mathbf{u}_t(\mathbf{x}|\mathbf{x}_1) = \frac{\mathbf{x}_1 - (1-\sigma_{\min})\mathbf{x}}{1 - (1-\sigma_{\min})t}, \\ p_t(\cdot|\mathbf{x}_1) = \mathcal{N}(\cdot; \boldsymbol{\mu}_t, \sigma_t\mathbf{I}), \quad \boldsymbol{\mu}_t = (1-t)\boldsymbol{\mu}_0 + t\boldsymbol{\mu}_1, \ \sigma_t = (1-t)\sigma_0 + t\sigma_1, \end{aligned} \tag{9}$$

where $\boldsymbol{\mu}_0 = \mathbf{0}$, $\boldsymbol{\mu}_1 = \mathbf{x}_1$, $\sigma_0 = 1$ and $\sigma_1 = \sigma_{\min} > 0$.

2.2 Exponential Power Distribution

Although Gaussian distribution is the default choice of many FM models, it is just a member of a larger *location-scale* family that also include Cauchy, Laplace, and Student's-t distributions.

Definition 1 (location-scale family)**.** *A location-scale family is a family of probability distributions parametrized by a location parameter* $\boldsymbol{\mu} \in \mathbb{R}^d$ *and a positive scale parameter* $\sigma > 0$*. Suppose* Z *is a random variable following distribution in this family corresponding to* $\boldsymbol{\mu} = \mathbf{0}$ *and* $\sigma = 1$ *with cumulative distribution function (CDF)* $F_0(z)$*. Then* $X = \boldsymbol{\mu} + \sigma Z$ *has CDF* $F(X) = F_0((X - \boldsymbol{\mu})/\sigma)$*.*

In this paper, we introduce another more flexible member, the *exponential power distribution*, a.k.a. generalized normal distribution, to flow matching methods for broader choices beyond Gaussian. Gómez et al. (1998) has a multivariate version of the exponential power distribution defined below.

Definition 2. *The* exponential power distribution*, denoted as* $\mathcal{P}_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, has the following probability density*

$$p(\mathbf{x}) = \frac{q\Gamma(\frac{d}{2})}{2\Gamma(\frac{d}{q})} 2^{-\frac{d}{q}} \pi^{-\frac{d}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{r^{\frac{q}{2}}}{2}\right\}, \quad r(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}). \tag{10}$$

Remark 1. *When* $q = 2$*, the exponential power distribution reduces to multivariate normal distribution, i.e.* $\mathcal{P}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*. Generally, the parameter* q *controls the regularization around center and tail behavior: the smaller* q*, the sharper regularization and the heavier tail it induces, as illustrated in Figure 1.*

As a special elliptic contoured distribution (Johnson, 1987; Fang & Zhang, 1990), the exponential power distribution is location-scale, i.e. if $Z \sim \mathcal{P}_q(\mathbf{0}, \mathbf{I})$, then $X = \boldsymbol{\mu} + \sigma Z \sim \mathcal{P}_q(\boldsymbol{\mu}, \sigma^2\mathbf{I})$. In this work, we focus on the isotropic case $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ and anisotropic case $\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\sigma}^2)$.

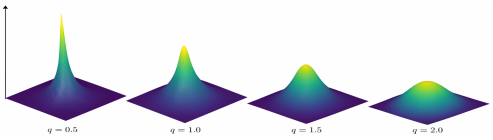

Figure 1: Probability densities of exponential power distributions $\mathcal{P}_q(\mathbf{0}, \mathbf{I})$ for selected q's.

The following proposition synthesizes the properties of exponential power distribution (Gómez et al., 1998; Li et al., 2023) useful for the development of new flow matching methods.

Proposition 2.1. *For a d-dimensional random variable $X \sim \mathcal{P}_q(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, we have*

1. $\mathrm{E}(X) = \boldsymbol{\mu}$, $\mathrm{Cov}(X) = s_{q,d}\boldsymbol{\Sigma}$ *with a scaling factor* $s_{q,d} = \frac{2^{\frac{2}{q}}\Gamma(\frac{d+2}{q})}{d\Gamma(\frac{d}{q})}$.
2. $X \overset{L}{=} \boldsymbol{\mu} + \sqrt{r}\mathbf{L}U_S$, *where* $r = (X-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(X-\boldsymbol{\mu})$ *and* $r^{\frac{q}{2}} \sim \Gamma(\alpha = \frac{d}{q}, \beta = \frac{1}{2})$, $\mathbf{L}\mathbf{L}^\mathsf{T} = \boldsymbol{\Sigma}$, *and* $\sqrt{r} \perp U_S$ *with* $U_S \sim \mathcal{U}(\mathcal{S}^{d-1})$ *uniformly distributed on the sphere.*

3 Flexible Flow Path in Location-Scale Family

In this section, we develop a series of novel flow matching methods in the broad location-scale family and propose new algorithms based on exponential power distributions. In addition, we also explore different probability paths by minimizing flow energies under various metric tensors, including Euclidean and Fisher-Rao (Section 4).

In the OT theory (Villani, 2009), the OT displacement between two univariate probability distributions $p_0(\cdot|x_1)$ and $p_1(\cdot|x_1)$ is $\psi(x) = F_1^{-1}(F_0(x))$ with F_0 and F_1 being their CDFs respectively. We immediately have OT interpolant path based on the conditional flow (8) in the one dimensional case. For general dimensions, we may have the following OT-based conditional flow

$$\psi_t(\mathbf{x}) = (1-t)\mathbf{x} + t\psi(\mathbf{x}), \quad \psi(\mathbf{x}) = [\psi(x_i)]_{i=1}^d, \quad \psi(x_i) = F_1^{-1}(F_0(x_i)). \tag{11}$$

Because the exponential power distribution is location-scale, we have $F_1(x) = F_0((x-x_1)/\sigma_{\min})$. Substituting it into (11) yields the same OT map $\psi(\mathbf{x}) = \mathbf{x}_1 + \sigma_{\min}\mathbf{x}$.

This observation motivates us to *start the search with any probability path in the broad location-scale family*, instead of being restricted to Gaussian distributions and the canonical affine path (6). Note that probability paths $\{p_t\}$ in the location-scale family parametrized by location and scale coordinates $(\boldsymbol{\mu}_t, \sigma_t)$ form a Riemannian manifold endowed with an information-geometry metric (Amari & Nagaoka, 2000), which will be explored in Section 4. The following theorem states that any probability path $p_t(\cdot; \boldsymbol{\mu}_t, \sigma_t)$ from the location-scale family can be generated by a determined vector field.

Theorem 3.1. *Any probability path $p_t(\cdot; \boldsymbol{\mu}_t, \sigma_t)$ in the location-scale family with pdf p_0 for the standard random variable $\mathbf{Z} = (\mathbf{X} - \boldsymbol{\mu}_t)/\sigma_t$ has the following vector field $\mathbf{v}_t$ that satisfies the continuity equation* (3)*:*

$$\mathbf{v}_t(\mathbf{x}) = \dot{\boldsymbol{\mu}}_t + \frac{\dot{\sigma}_t}{\sigma_t}(\mathbf{x} - \boldsymbol{\mu}_t). \tag{12}$$

Proof. See Appendix A . □

Remark 2. *The proof of this theorem is distribution-agnostic in the location-scale family. Though the format of the vector field $\mathbf{v}_t$ to generate probability p_t is the same as in (Theorem 3 of Lipman et al., 2023), our theorem generalizes it to the location-scale family from which we can use the exponential power distribution $\mathcal{P}_q$ beyond Gaussian $\mathcal{P}_2 = \mathcal{N}$.*

Starting with the probability path $p_t(\cdot; \boldsymbol{\mu}_t, \sigma_t)$, Theorem 3.1 determines the conservation law (3) (4) respecting vector field. The following theorem states that such vector field (12) prescribes the flow.

Theorem 3.2. *The flow equation* (1) *with vector field $\mathbf{v}_t$* (12) *has the unique solution as an affine transformation:*

$$\phi_t(\mathbf{x}) = \boldsymbol{\mu}_t + \sigma_t\mathbf{x}. \tag{13}$$

Proof. See Appendix A . □

Remark 3. *Although we end up with the same affine flow, unlike Lipman et al. (2023), which searches for proper flow first, we can start with any probability path in the vast location-scale family, e.g. Normal, Laplace, Student's t, elliptical distributions, then specify the vector field* (12) *compatible with the continuity law* (3)*, and determine the flow* (13) *at last. Supported by Theorem 3.1 and Theorem 3.2, the workflow* probability path $(p_t) \to$ vector field $(\mathbf{v}_t) \to$ sample flow (ϕ_t) *is reverse to that of Lipman et al. (2023), which lends us greater flexibility in choosing flow in the FM framework.*

We emphasize the novelty and importance of the order **probability path** $(p_t) \xrightarrow{Th3.1}$ **vector field** $(\mathbf{v}_t) \xrightarrow{Th3.2}$ **sample flow** (ϕ_t), which would otherwise no longer guarantee $p_t = [\phi_t]_\sharp p_0$ to have tractable geodesics when p_0 moves away from Gaussian, a property lacking in Kapusniak et al. (2024); Davis et al. (2024). Because the conditional probability path $p_t(\cdot|\mathbf{x}_1)$ in the location-scale family is parametrized by location $\boldsymbol{\mu}_t$ and scale σ_t, what remains in the CFM framework is to determine them with boundary conditions $(\boldsymbol{\mu}_0, \sigma_0) = (\mathbf{0}, 1)$ and $(\boldsymbol{\mu}_1, \sigma_1) = (\mathbf{x}_1, \sigma_{\min})$ to drive the conditional vector field $\mathbf{u}_t(\cdot|\mathbf{x}_1)$ (7) and the conditional flow ψ_t (6). Then a neural network is trained to regress $\mathbf{u}_t(\cdot|\mathbf{x}_1)$ along the flow $\psi_t(\mathbf{x}_0)$, i.e. $\mathbf{u}_t(\psi_t(\mathbf{x}_0)|\mathbf{x}_1) = \frac{d}{dt}\psi_t(\mathbf{x}_0)$, with the following loss:

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t,q(\mathbf{x}_1),p(\mathbf{x}_0)} \|\mathbf{v}_t(\psi_t(\mathbf{x}_0); \theta) - \frac{d}{dt}\psi_t(\mathbf{x}_0)\|_2^2, \tag{14}$$

where $t \sim \mathcal{U}[0, 1]$, $\mathbf{x}_1 \sim q(\cdot)$, and $\mathbf{x}_0 \sim p_0(\cdot) = \mathcal{P}_q(\cdot; \mathbf{0}, \mathbf{I})$.

In the following, we specify the coordinate path $\{(\boldsymbol{\mu}_t, \sigma_t) : t \in [0, 1]\}$ on the manifold of location-scale probability distributions $\mathcal{M} = \{p(\cdot; \boldsymbol{\mu}, \sigma)\}$ by minimizing the flow energy. We will revisit and introduce three probability paths in the family of exponential power distributions: i) minimal energy path; ii) non-linear interpolant path; and iii) (probabilistic) geodesic path.

3.1 Minimal Energy Path

OT with quadratic cost is known to be an energy minimization problem under the Wasserstein-2 metric (Benamou & Brenier, 2000). To determine the best conditional vector field, we adopt the energy criterion for the flow:

$$e(\mathbf{v}) = \mathbb{E}_{p_t(\mathbf{x})} \int_0^1 \|\dot{\mathbf{x}}_t\|^2 dt = \int_{\mathbb{R}^d} \int_0^1 p_t(\mathbf{x}) \|\mathbf{v}_t(\mathbf{x})\|_2^2 dt d\mathbf{x}. \tag{15}$$

Minimizing the energy (15) leads to geodesic in the Euclidean space, as stated below.

Proposition 3.1. *The following OT interpolant path minimizes the flow energy* (15)*:*

$$\boldsymbol{\mu}_t = (1-t)\boldsymbol{\mu}_0 + t\boldsymbol{\mu}_1 = t\mathbf{x}_1, \quad \sigma_t = (1-t)\sigma_0 + t\sigma_1 = 1 - t + t\sigma_{\min}. \tag{16}$$

The minimal energy attained is $e_{\min} = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|^2 + d|\sigma_1 - \sigma_0|^2$.

Proof. See Appendix A . □

Therefore, we have the conditional flow and vector field along the minimal energy path for loss (14):

$$\psi_t(\mathbf{x}) = (1 - (1 - \sigma_{\min})t)\mathbf{x} + t\mathbf{x}_1, \quad \frac{d}{dt}\psi_t(\mathbf{x}) = \mathbf{x}_1 - (1 - \sigma_{\min})\mathbf{x}.$$

3.2 Nonlinear Interpolant Path

The OT-interpolant (16) is a linear interpolation between the starting $(\boldsymbol{\mu}_0, \sigma_0)$ and ending $(\boldsymbol{\mu}_1, \sigma_1)$ points on the manifold. A popular nonlinear alternative involves the functions $\sin$ and $\cos$ (Lipman et al., 2024):

$$\mathbf{x}_t = \psi_t(\mathbf{x}) = \mathbf{x}\cos(\pi t/2) + \mathbf{x}_1 \sin(\pi t/2). \tag{17}$$

By $\frac{d}{dt}\psi_t(\mathbf{x}) = \mathbf{u}_t(\psi_t(\mathbf{x})|\mathbf{x}_1)$, we can derive the conditional vector field as $\mathbf{u}_t(\mathbf{x}|\mathbf{x}_1) = \frac{\pi}{2}\frac{\mathbf{x}_1 - \mathbf{x}\sin(\pi t/2)}{\cos(\pi t/2)}$, which generates the conditional probability path $p_t(\cdot; \boldsymbol{\mu}_t, \sigma_t)$ in the location-scale family with coordinates: $\boldsymbol{\mu}_t(\mathbf{x}_1) = \mathbf{x}_1 \sin(\pi t/2), \quad \sigma_t(\mathbf{x}_1) = \cos(\pi t/2)$, although it does not minimize the energy (15). Later we will show that it is actually a special case of the probabilistic geodesic path on the manifold of exponential power distributions with Fisher-Rao metric (Section 4).

4 Probabilistic Geodesic Flow

To search for the optimal probability path $\{p_t(\cdot; \boldsymbol{\mu}_t, \sigma_t) : t \in [0, 1]\}$ in the location-scale family, it is natural to adopt the Riemannian manifold of probability distributions $\mathcal{M} = \{p(\cdot; \boldsymbol{\mu}, \sigma)\}$ with Fisher-Rao metric g (Amari & Nagaoka, 2000). From the manifold perspective, we identify $(\boldsymbol{\mu}_t, \sigma_t) \leftrightarrow$

$p_t(\cdot; \boldsymbol{\mu}_t, \sigma_t)$. Although the methodology introduced applies to general location-scale distributions, we focus on the exponential power distribution $\mathcal{P}_q(\boldsymbol{\mu}, \sigma^2\mathbf{I})$. Lemma A.1 shows that the Fisher-Rao metric (23) for this manifold is $ds^2 = \frac{c_\mu \|d\boldsymbol{\mu}\|^2 + c_\sigma d\sigma^2}{\sigma^2}$ with $c_\mu(d,q) = \frac{2^{-\frac{2}{q}}}{d}\frac{\Gamma(\frac{d-2}{q}+2)}{\Gamma(\frac{d}{q})}q^2$, and $c_\sigma = qd$. Note that if $X \sim \mathcal{P}_q(\boldsymbol{\mu}_t, \sigma_t^2\mathbf{I})$, then $\mathbf{v}_t(X) \sim \mathcal{P}_q(\dot{\boldsymbol{\mu}}_t, \dot{\sigma}_t^2\mathbf{I})$ by (12). The flow energy (15) can be redefined by replacing the Euclidean metric with the Fisher-Rao metric (23):

$$e(\mathbf{v}) = \mathbb{E}_{p_t(\mathbf{x})}\int_0^1 \|\dot{\mathbf{x}}_t\|_g^2 dt = \int_{\mathbb{R}^d}\int_0^1 p_t(\mathbf{x})\|\mathbf{v}_t(\mathbf{x})\|_g^2 dt d\mathbf{x} = \int_0^1 \frac{c_\mu\|\dot{\boldsymbol{\mu}}_t\|^2 + c_\sigma\dot{\sigma}_t^2}{\sigma_t^2}dt. \tag{18}$$

By the calculus of variation, we get the geodesic as a semi-circle on the manifold of exponential power distributions $\mathcal{P}_q(\boldsymbol{\mu}_t, \sigma_t^2\mathbf{I})$ in the following theorem.

Theorem 4.1. *The following geodesic path minimizes the variational energy* (18)*:*

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + x(t)\mathbf{e}, \quad x(t) = \mu_c + R\cos\theta(t), \quad R = \sqrt{\mu_c^2 + \lambda^2\sigma_0^2}; \tag{19a}$$

$$\sigma_t = \frac{R}{\lambda}\sin\theta(t), \quad \theta(t) = 2\arctan\left(e^{Lt}\tan\frac{\theta_0}{2}\right), \quad L = \log\frac{\tan(\theta_1/2)}{\tan(\theta_0/2)}, \tag{19b}$$

where $\lambda = \sqrt{c_\sigma/c_\mu}$, $\mathbf{e} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)/\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|_2$, $\mu_c = \frac{\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_0\|^2 + \lambda^2(\sigma_1^2-\sigma_0^2)}{2\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_0\|}$, *and* $\theta_0 = \operatorname{atan2}(\lambda\sigma_0, -\mu_c)$, $\theta_1 = \operatorname{atan2}(\lambda\sigma_1, \|\boldsymbol{\mu}_1-\boldsymbol{\mu}_0\| - \mu_c)$. *The minimal energy attained is* $e_{\min} = c_\sigma L^2$*, where* $\sqrt{c_\sigma}|L|$ *is the Fisher-Rao distance between* $\mathcal{P}_q(\boldsymbol{\mu}_0, \sigma_0^2\mathbf{I})$ *and* $\mathcal{P}_q(\boldsymbol{\mu}_1, \sigma_1^2\mathbf{I})$*, for which* $|L|$ *can be rewritten in the original parameters as*

$$|L| = \operatorname{arcosh}\left(1 + \frac{c_\mu\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|^2 + c_\sigma|\sigma_1 - \sigma_0|^2}{2c_\sigma\sigma_0\sigma_1}\right). \tag{20}$$

Proof. See Appendix A. □

Although mathematically complicated, this **analytically tractable** formula has numerical complexity $\mathcal{O}(d)$, which becomes negligible compared to network training in higher dimensions, as evidenced by the training time reported in Section 5 and Section B. We also have an anisotropic version that decouples the isotropic geodesic into d independent one-dimensional geodesics (Appendix C).

In Theorem A.1, the CFM loss (14) along the geodesic in Theorem 4.1 decomposes into the FM loss and the variance of conditional vector field and is bounded from below.

In particular, we get the vector field $\mathbf{v}_t$ (12) that generates the exponential power probability path $\mathcal{P}_q(\boldsymbol{\mu}_t, \sigma_t^2\mathbf{I})$, and the corresponding flow $\phi_t(\mathbf{x})$ (13). Letting $\boldsymbol{\mu}_0 = \mathbf{0}, \boldsymbol{\mu}_1 = \mathbf{x}_1, \sigma_0 = 1, \sigma_1 = \sigma_{\min}$, we have the conditional flow and vector field along the probabilistic geodesic path to feed in (14):

$$\begin{aligned}\psi_t(\mathbf{x}) &= \boldsymbol{\mu}_t(\mathbf{x}_1) + \sigma_t(\mathbf{x}_1)\mathbf{x} = x(t)\mathbf{x}_1/\|\mathbf{x}_1\| + \frac{R}{\lambda}\sin\theta(t)\mathbf{x},\\ \frac{d}{dt}\psi_t(\mathbf{x}) &= \mathbf{u}_t(\psi_t(\mathbf{x})|\mathbf{x}_1) = \left[-\sin\theta(t)\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|} + \frac{1}{\lambda}\cos\theta(t)\mathbf{x}\right]RL\sin\theta(t).\end{aligned} \tag{21}$$

Interestingly, when $R = \lambda = \|\mathbf{x}_1\|$, letting $\sigma_{\min} \downarrow 0$, we have $\mu_c = \frac{\|\mathbf{x}_1\|^2 + \lambda^2(\sigma_{\min}^2 - 1)}{2\|\mathbf{x}_1\|} \to 0$. Therefore $x(t) \to R\cos\theta(t)$. The conditional flow along this special geodesic path reduces to

$$\psi_t(\mathbf{x}) = x(t)\mathbf{x}_1/\|\mathbf{x}_1\| + \frac{R}{\lambda}\sin\theta(t)\mathbf{x} \to \cos\theta(t)\mathbf{x}_1 + \sin\theta(t)\mathbf{x},$$

which is exactly the sinusoidal path (17) with $\theta(t) = \pi(1-t)/2$.

FM methods transfer samples such that the distribution they represent changes from pure noise (standard normal or exponential power) to the data distribution. The ultimate goal is to evolve distributions, not to transport samples. *OT based approaches move samples along straight lines in the Euclidean space. However, they do not necessarily travel the shortest distances in the space of probability distributions.* The leftmost panel of Figure A.1 shows that the Fisher-Rao distance may be smaller than the Euclidean distance between two exponential power distributions $\mathcal{P}_q(\mathbf{0}, \mathbf{I})$ and $\mathcal{P}_q(\mathbf{1}, \sigma_{\min}^2\mathbf{I})$ for certain q's. The other panels of Figure A.1 demonstrate selected probabilistic geodesic paths (second from left) and contrast their corresponding flow paths for $q = 1.0$ vs $q = 2.0$ (rightmost two) – smaller q leads to a more stringent movement than larger q.

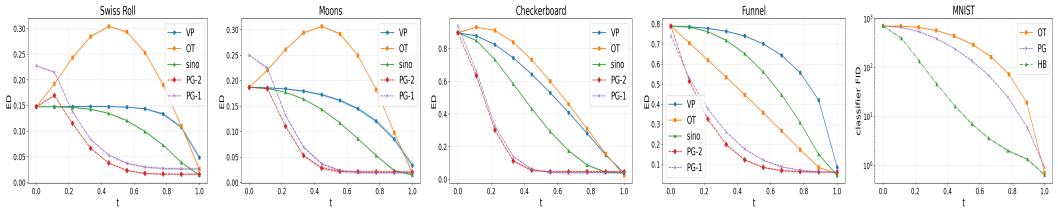

Figure 2: Contrasting trajectories of energy distance (ED) in the Swiss roll, moons, checkerboard, and funnel examples (left four) and class Frechet inception distance (clf-FID) in MNIST (rightmost).

Limitation There is a numerical issue on the conditional vector field that aggravates with increasing dimensions. When $\sigma_1 \to 0$, the geodesic length $|L| \to \infty$ as in (20). As $t \to 1$, $\sigma_t \to \sigma_{\min} \ll 1$, and the conditional vector field $\dot{\psi}_t$ in (21) fluctuates greatly in magnitude as the flow moves towards the data end. Therefore, the vector field learned by neural network would cause the resulted ODE stiff and numerically difficult to solve. Luckily, this limitation can be alleviated by a hybrid (HB):

$$\psi_t(\mathbf{x}) = \begin{cases} \boldsymbol{\mu}_t(\mathbf{x}_1) + \sigma_t(\mathbf{x}_1)\mathbf{x}, & t < t_c, \\ t\mathbf{x}_1 + (1-t)\mathbf{x}, & t \geq t_c. \end{cases}$$

where we use $t_c = 0.85$ in numerical experiments.

5 Numerical Results

In this section, we compare the proposed probabilistic geodesic (PG) FM and its hybrid variant (HB) against baseline models with variance preserving (VP), optimal transport (OT) interpolant, and sinusoidal (sino) probability paths. To fully test their performance, we also include the state-of-the-art models Riemannian FM (RFM Chen & Lipman, 2024), Metric FM (MFM Kapusniak et al., 2024), and Fisher FM (FFM Davis et al., 2024) in the comparison. Three 2d benchmark examples including Swiss roll, moons, and checkerboard, higher dimensional simulations of an augmented eight-Gaussian and Neal's funnel distribution, handwritten digits, four scientific datasets (Appendix B.4), and a molecular generation problem are used in the test. We refer to Wasserstein-2 distance (W_2), energy distance (ED), and maximum mean discrepancy (MMD) to measure the discrepancy between synthetic sample distribution and real data distribution. One can refer to Appendix B.1 for more details of the setup. All the computer code will be released.

5.1 Simulations

5.1.1 2d Benchmarks: Swiss Roll, Moons, and Checkerboard

First, we test VP, OT, sino, and PG-2 (on Gaussian distributions) and PG-1 (on exponential power distributions $\mathcal{P}_1(\boldsymbol{\mu}_t, \sigma_t)$) on three 2d benchmark datasets: Swiss roll, moons and checkerboard (Grathwohl et al., 2019). Figure B.1 demonstrates their sample paths on the Swiss roll dataset. It is clear that PG methods are more effective in transporting samples in which the rolling patterns emerge in earlier stages. Starting with the more concentrated standard exponential power noise, PG-1 learns the data feature even faster than the others based on standard Gaussian noise. Similar quick feature learning also shows in the other two examples (Figures B.1 and B.2). The metrics in terms of W_2, ED, and MMD in Tables B.1 and B.2 also support the numerical advantage (smaller generative discrepancy) of PG methods, of which some are one order of magnitude smaller. The time overhead (time/epoch) fades off as the dimension increases (see Tables 1, B.5, and B.4).

Figure 2 plots the trajectories of ED on these benchmark datasets and clf-FID on MNIST. OT interpolant with independent coupling ($\mathbf{x}_0 \perp \mathbf{x}_1$) features an "energy hump" (refer to Remark 7) in the middle of the path due to the well-known midpoint (marginal) variance contraction (Albergo & Vanden-Eijnden, 2023; Tong et al., 2024). This does not present an issue for our PG paths as their energy distances to the target distribution decrease quickly and monotonically.

Ablation on q. As illustrated in Figures 1 and Figure A.1, the parameter q in the exponential power distribution plays a role of regularization: smaller q induces a more strict movement of the sample.

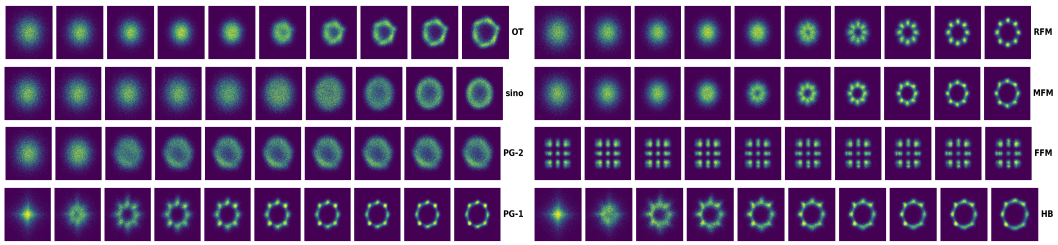

Figure 3: Sample paths in the 8-Gaussians (36d) example. Only the first two dimensions are plotted.

However, that does not necessarily lead to better performance when training becomes more difficult in the non-convex regime ($q < 1.0$). In Table B.3, these discrepancy metrics generally decrease and then increase as q changes from above 2 to below 1.

5.1.2 Higher Dimensions: 36d eight-Gaussians, 100d Funnel Distribution

To test the robustness of PG methods to higher dimensionality, we extend the classical eight-Gaussians example (Grathwohl et al., 2019) from 2d to $d = 36$ dimensions by filling in the remaining $d - 2$ coordinates with isotropic Gaussian noise. To test the capability of PG methods of handling heavy-tail data, we investigate the more challenging Neal's funnel distribution (Neal, 2003) whose density $f(\mathbf{x}, \nu) = \mathcal{N}_d(\mathbf{x}; \mathbf{0}, e^{\nu/2}\mathbf{I})\mathcal{N}(\nu; 0, 3)$ features a funnel (Figure B.2).

Table 1: Neal's funnel (100d): comparison in W_2, ED, and MMD (mean $\pm$ std over 10 seeds).

DIST	Method	$W_2 \downarrow$	ED $\downarrow$	MMD $\downarrow$	time/epoch
q=1.0	OT	$3.73e0 \pm 2.49e{-1}$	$1.43e{-1} \pm 1.35e{-2}$	$3.52e{-3} \pm 1.07e{-3}$	$5.90e{-2} \pm 6.85e{-5}$
q=2.0	OT	$3.74e0 \pm 2.00e{-1}$	$1.41e{-1} \pm 1.18e{-2}$	$3.32e{-3} \pm 7.60e{-4}$	$5.76e{-2} \pm 8.21e{-5}$
q=1.0	RFM	$3.65e0 \pm 2.01e{-1}$	$1.27e{-1} \pm 1.33e{-2}$	$2.30e{-3} \pm 7.17e{-4}$	$7.68e{-2} \pm 6.59e{-4}$
q=2.0	RFM	$3.69e0 \pm 2.12e{-1}$	$1.29e{-1} \pm 1.10e{-2}$	$2.48e{-3} \pm 5.29e{-4}$	$7.11e{-2} \pm 5.74e{-4}$
q=1.0	MFM	$4.05e0 \pm 2.39e{-1}$	$1.26e{-1} \pm 1.16e{-2}$	$2.65e{-3} \pm 7.57e{-4}$	$4.36e{-1} \pm 1.57e{-2}$
q=2.0	MFM	$4.17e0 \pm 1.83e{-1}$	$1.27e{-1} \pm 1.31e{-2}$	$2.80e{-3} \pm 5.95e{-4}$	$4.26e{-1} \pm 1.62e{-2}$
q=1.0	FFM	$6.18e0 \pm 5.90e{-1}$	$1.00e0 \pm 7.25e{-2}$	$1.16e{-1} \pm 8.27e{-3}$	$1.00e{-2} \pm 2.52e{-5}$
q=2.0	FFM	$5.55e0 \pm 6.64e{-1}$	$8.99e{-1} \pm 7.95e{-2}$	$1.01e{-1} \pm 8.49e{-3}$	$8.69e{-3} \pm 1.38e{-5}$
q=1.0	PG	$3.76e0 \pm 3.06e{-1}$	$1.82e{-1} \pm 1.04e{-2}$	$5.94e{-3} \pm 8.64e{-4}$	$6.51e{-2} \pm 1.39e{-4}$
q=2.0	PG	$4.58e0 \pm 6.98e{-1}$	$1.62e{-1} \pm 2.64e{-2}$	$4.48e{-3} \pm 1.34e{-3}$	$6.33e{-2} \pm 1.08e{-4}$
q=1.0	HB	$2.99e0 \pm 2.38e{-1}$	$1.26e{-1} \pm 1.69e{-2}$	$2.39e{-3} \pm 6.59e{-4}$	$6.92e{-2} \pm 2.69e{-4}$
q=2.0	HB	$3.67e0 \pm 3.27e{-1}$	$1.84e{-1} \pm 1.25e{-1}$	$8.34e{-3} \pm 1.55e{-2}$	$6.61e{-2} \pm 3.42e{-4}$

Why non-Gaussianity matters? Figure 3 compares the sample paths of the baseline (OT, sino) and SOTA (RFM, MFM, FFM) models with the proposed PG and HB models. It challenges all the methods to discover eight 2d Gaussian blobs in 36 dimensional ambient space, yet PG-1 and HB are the fastest to identify all the modes. What is more, the exponential power distribution's heavier tail property (with smaller q as in PG-1) helps identify and isolate those modes faster and better than Gaussian ($q = 2$ as in PG-2) which tends to smooth and blur the pattern. This explains why most $q = 1.0$ rows have lower discrepancy metrics than $q = 2.0$ (except RFM and MFM) in Table B.4 for which PG-1 achieves the lowest ED and MMD. In Table 1, except for a few cells, $q = 1.0$ almost unanimously beats $q = 2.0$, which again justifies the adoption of exponential power distribution. HB attains the lowest W_2 and ED for $q = 1.0$. Note that MFM reaches the same lowest ED at a much higher time cost ($4.36e - 1$ vs $6.92e - 2$ seconds per epoch for HB).

5.2 MNIST Handwritten Digits

In this subsection, we consider the MNIST dataset (Lecun et al., 1998) consisting of 60,000 training handwritten digits and 10,000 testing handwritten digits with $d = 28 \times 28 = 784$. The sharp contrast along the edge of these digits challenges FM models. Here we mainly focus on comparing the OT, PG, and HB paths.

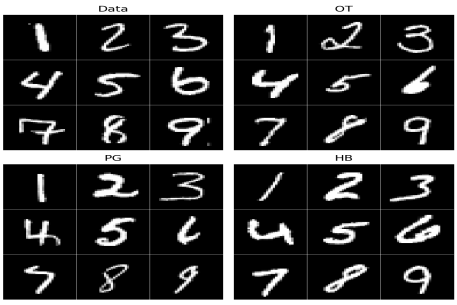

Figure 4: sample handwritten digits.

Table 2: Comparison of generation quality metrics across different path methods. Values are reported as mean $\pm$ standard deviation over $n = 5$ runs. Higher IS is better, while lower FID and clf-FID are better.

DIST	PATH	IS ↑	FID ↓	clf-FID ↓
q=1.0	OT	2.099 ± 0.008	1.606 ± 0.037	0.718 ± 0.108
q=2.0	OT	2.107 ± 0.003	1.597 ± 0.019	0.770 ± 0.081
q=1.0	PG	2.095 ± 0.009	1.671 ± 0.013	0.909 ± 0.139
q=2.0	PG	2.083 ± 0.006	2.281 ± 0.097	1.843 ± 0.279
q=1.0	HB	2.107 ± 0.008	1.727 ± 0.029	0.567 ± 0.119
q=2.0	HB	2.102 ± 0.008	1.803 ± 0.040	0.556 ± 0.071

In Table 2, we evaluate the sample quality of OT, PG, and HB under both $q = 1$ and $q = 2$ in terms of the inception score (IS) and the Fréchet inception distance (FID), based on ImageNet-pretrained Inception-V3 activations. Additionally, we report a classifier FID (clf-FID), computed identically but on penultimate-layer features from a LeNet-5 Lecun et al. (1998) classifier trained on MNIST, which makes the metric sensitive to domain-specific structure that inception features do not capture. While OT achieves the highest IS and the lowest FID, HB attains the highest IS (same) and the lowest clf-FID. Figure 4 illustrates some generated digits.

5.3 Molecular Generation

Lastly, we consider a real data example of thousands of dimensions – de novo molecule generation (Section 4.4 of Davis et al., 2024), which aims to generate molecules positions unconditionally based on QM9 data (Ruddigkeit et al., 2012; Ramakrishnan et al., 2014). Refer to Appendix B.5 for a detailed description of the problem. Table 3 reports the percentage of stable atoms within molecules (Atoms S), valid molecules (Mols Val), and stable molecules (Mols. S) based on 1,000 molecules (first 4 rows quoted from Table 3 of Davis et al. (2024)). For a fare comparison, all the source/noise distributions are Gaussian. PG/HB is slightly lower in the percentage of valid molecules (Mols Val), but has higher percentages of stable atoms within molecules (Atoms S) and stable molecules (Mols. S) than OT, FFM and FlowMol all together. Figure 5 shows synthetic molecules generated by PG.

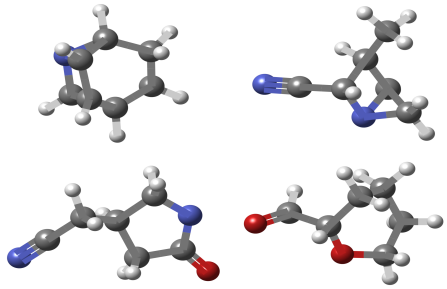

Figure 5: Generated molecules by PG-2.

Table 3: Results of molecular generation on QM9: the percentage of stable atoms within molecules (Atoms S), valid molecules (Mols Val), and stable molecules (Mols. S) based on 1,000 molecules.

Method	**Atoms S. (%) ↑**	**Mols. Val. (%) ↑**	**Mols. S. (%) ↑**
Fisher-Flow	98.6	95.3	88.2
JODO	**99.4**	**98.9**	**98.7**
EquiFM	99.4	94.4	93.2
FlowMol	98.9	**96.9**	84.2
OT	99.35	89.80	91.60
PG	**99.49**	90.63	**93.75**
HB	99.39	89.84	92.97

6 Conclusion

In this paper, we extend the conditional flow matching from Gaussian to a broader location-scale family with a particular focus on exponential power distributions that embrace Gaussian as a special case $q = 2$. We propose a class of novel probability paths as geodesics on a manifold of probability distributions with Fisher-Rao metric. The proposed methods demonstrate fast feature learning capability with numerical performance superior or comparable to SOTA FM models.

The probabilistic geodesic path is not as conceptually straightforward as the OT interpolant path. The computational overhead diminishes as the dimension increases, compared to the cost of network training. Future directions include integration of the proposed probabilistic geodesic path into recent one-step methods (Frans et al., 2025; Geng et al., 2026) to further boost the efficiency and effectiveness of FM.

AI USE STATEMENT

In this work, we used generative AI tools for polishing presentation of the paper (refining sentences and paragraphs). We have not used generative AI tools for writing the whole paper. In addition, we used generative AI tools for assistance in coding and debuging. We have reviewed all AI-assisted work. For example, the LLM-generated code was verified and tested for correctness by all co-authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

All results are reproducible using `python` codes that will be published after publication. We include a `jupyter` notebook of a demonstrative example in the submitted supplementary materials.

REFERENCES

Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023.

M.S. Albergo. Flow-based generative models for markov chain monte carlo in lattice field theory. *Physical Review D*, 100(3), 2019.

S. Amari and H. Nagaoka. *Methods of Information Geometry*, volume 191 of *Translations of Mathematical monographs*. Oxford University Press, 2000.

Lazar Atanackovic, Xi Zhang, Brandon Amos, Mathieu Blanchette, Leo J Lee, Yoshua Bengio, Alexander Tong, and Kirill Neklyudov. Meta flow matching: Integrating vector fields on the wasserstein manifold. In *The Thirteenth International Conference on Learning Representations*, 2025.

Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, Jan 2000. ISSN 0945-3245.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *The Twelfth International Conference on Learning Representations*, 2024.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Chaoran Cheng, Jiahan Li, Jian Peng, and Ge Liu. Categorical flow matching on statistical manifolds. In *Advances in Neural Information Processing Systems 37*, NeurIPS 2024, pp. 54787–54819. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024.

Oscar Davis, Samuel Kessler, Mircea Petrache, Ismail Ilkan Ceylan, Michael M. Bronstein, and Joey Bose. Fisher flow matching for generative modeling over discrete data. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

K. Fang and Y.T. Zhang. *Generalized Multivariate Analysis*. Science Press, 1990. ISBN 9780387176512.

Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut models. In *The Thirteenth International Conference on Learning Representations*, 2025.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026.

E. Gómez, M.A. Gomez-Viilegas, and J.M. Marín. A multivariate generalization of the power exponential family of distributions. *Communications in Statistics - Theory and Methods*, 27(3):589–600, jan 1998.

Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, and David Duvenaud. Scalable reversible generative models with free-form continuous dynamics. In *International Conference on Learning Representations*, 2019.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020.

Mark E. Johnson. *Multivariate Statistical Simulation*, chapter 6 Elliptically Contoured Distributions, pp. 106–124. Probability and Statistics. John Wiley & Sons, Ltd, 1987. ISBN 9781118150740.

Kacper Kapusniak, Peter Potaptchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael Bronstein, Avishek Bose, and Francesco Di Giovanni. Metric flow matching for smooth interpolations on the data manifold. In *Advances in Neural Information Processing Systems 37*, NeurIPS 2024, pp. 135011–135042. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Shuyi Li, Michael O' Connor, and Shiwei Lan. Bayesian learning via q-exponential process. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Proceedings of the 37th Conference on Neural Information Processing Systems*, volume 36, pp. 72867–72887. Curran Associates, Inc., 2023.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky T. Q. Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow matching guide and code. *arXiv:2412.06264*, 12 2024.

Dimitra Maoutsa, Sebastian Reich, and Manfred Opper. Interacting particle solutions of fokker–planck equations through gradient–log–density estimation. *Entropy*, 22(8):802, 2020. ISSN 1099-4300.

Robert J. McCann. A convexity principle for interacting gases. *Advances in Mathematics*, 128(1):153–179, 1997. ISSN 0001-8708.

Radford M. Neal. Slice sampling. *The Annals of Statistics*, 31(3), jun 2003.

Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.

Kushagra Pandey, Jaideep Pathak, Yilun Xu, Stephan Mandt, Michael Pritchard, Arash Vahdat, and Morteza Mardani. Heavy-tailed diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025.

Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1), August 2014. ISSN 2052-4463.

Lars Ruddigkeit, Ruud van Deursen, Lorenz C. Blum, and Jean-Louis Reymond. Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17. *Journal of Chemical Information and Modeling*, 52(11):2864–2875, November 2012. ISSN 1549-960X.

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

Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. Expert Certification.

Cédric Villani. *Optimal Transport*. Springer Berlin Heidelberg, 2009. ISBN 9783540710509.

F. Alexander Wolf, Philipp Angerer, and Fabian J. Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1):15, Feb 2018. ISSN 1474-760X.

Table .4: Comparison of flow-matching methods.

Method	Manifold of	Metric	Geodesic	Data type
Riemannian FM (Chen & Lipman, 2024)	general (mainly data)	standard Riemannian	general	non-Euclidean
Metric FM (Kapusniak et al., 2024)	Data	learned	general	general
Meta FM (Atanackovic et al., 2025)	probability densities	2-Wasserstein	general	gmbedded in graph
Categorical FM (Cheng et al., 2024)	probability measures	Fisher-Rao	circle	discrete
Fisher FM (Davis et al., 2024)	probability densities	Fisher-Rao	circle	discrete
PG FM (ours)	probability densities	Fisher-Rao	Poincaré semicircle	continuous, Euclidean

A PROOFS

Theorem 3.1. *Any probability path $p_t(\cdot;\boldsymbol{\mu}_t,\sigma_t)$ in the location-scale family with pdf p_0 for the standard random variable $\mathbf{Z}=(\mathbf{X}-\boldsymbol{\mu}_t)/\sigma_t$ has the following vector field $\mathbf{v}_t$ that satisfies the continuity equation* (3)*:*

$$\mathbf{v}_t(\mathbf{x})=\dot{\boldsymbol{\mu}}_t+\frac{\dot{\sigma}_t}{\sigma_t}(\mathbf{x}-\boldsymbol{\mu}_t). \tag{12}$$

Proof of Theorem 3.1. Because $\mathbf{Z}\sim p_0$, we have

$$\mathbf{X}_t=\boldsymbol{\mu}_t+\sigma_t\mathbf{Z}\sim p_t,\quad p_t(\mathbf{x})=p_0((\mathbf{x}-\boldsymbol{\mu}_t)/\sigma_t)\sigma_t^{-d}. \tag{22}$$

Along this trajectory, we have

$$\dot{\mathbf{X}}_t=\dot{\boldsymbol{\mu}}_t+\dot{\sigma}_t\mathbf{Z}=\dot{\boldsymbol{\mu}}_t+\frac{\dot{\sigma}_t}{\sigma_t}(\mathbf{X}_t-\boldsymbol{\mu}_t).$$

Let $\mathbf{v}_t(\mathbf{x})=\dot{\boldsymbol{\mu}}_t+\frac{\dot{\sigma}_t}{\sigma_t}(\mathbf{x}-\boldsymbol{\mu}_t)$. The flow (1) with $\mathbf{v}_t$ then follows a probability path p_t (22) given by the pushforward (2). This is equivalent to the continuity equation (3), which can also be verified.

On the one hand,

$$\begin{aligned}\partial_t p_t(\mathbf{x})&=-d\sigma_t^{-d-1}\dot{\sigma}_t p_0(\mathbf{z})+\sigma_t^{-d}\nabla p_0(\mathbf{z})\cdot(-\dot{\boldsymbol{\mu}}_t\sigma_t^{-1}-(\mathbf{x}-\boldsymbol{\mu}_t)\sigma_t^{-2}\dot{\sigma}_t)\\&=-d\frac{\dot{\sigma}_t}{\sigma_t}p_t(\mathbf{x})-\sigma_t^{-d-1}\nabla p_0(\mathbf{z})\cdot(\dot{\boldsymbol{\mu}}_t+\dot{\sigma}_t\mathbf{z}),\end{aligned}$$

On the other hand,

$$\begin{aligned}-\nabla\cdot(\mathbf{v}_t p_t)&=-(\nabla\cdot\mathbf{v}_t)p_t(\mathbf{x})-\mathbf{v}_t\cdot\nabla p_t(\mathbf{x})\\&=-d\frac{\dot{\sigma}_t}{\sigma_t}p_t(\mathbf{x})-(\dot{\boldsymbol{\mu}}_t+\dot{\sigma}_t\mathbf{z})\cdot(\sigma_t^{-d-1}\nabla p_0(\mathbf{z}))\end{aligned}$$

Therefore, the two sides of (3) are equal. Hence, the proof is completed. □

Theorem 3.2. *The flow equation* (1) *with vector field $\mathbf{v}_t$* (12) *has the unique solution as an affine transformation:*

$$\phi_t(\mathbf{x})=\boldsymbol{\mu}_t+\sigma_t\mathbf{x}. \tag{13}$$

Proof of Theorem 3.2. Let $\mathbf{u}_t=\phi_t(\mathbf{x})-\boldsymbol{\mu}_t$. With the vector field $\mathbf{v}_t$ (12), we substitute it into (1) to get

$$\dot{\mathbf{u}}_t=\dot{\phi}_t-\dot{\boldsymbol{\mu}}_t=\mathbf{v}_t(\phi_t)-\dot{\boldsymbol{\mu}}_t=\dot{\boldsymbol{\mu}}_t+\frac{\dot{\sigma}_t}{\sigma_t}(\phi_t-\boldsymbol{\mu}_t)-\dot{\boldsymbol{\mu}}_t=\frac{\dot{\sigma}_t}{\sigma_t}\mathbf{u}_t,$$

which already decouples across coordinates. Integrating it as a vector ODE yields

$$\mathbf{u}_t=\frac{\sigma_t}{\sigma_0}\mathbf{u}_0.$$

Since $\phi_0(\mathbf{x})=\mathbf{x}$, $\boldsymbol{\mu}_0=\mathbf{0}$ and $\sigma_0=1$, we get $\mathbf{u}_0=\mathbf{x}$. Therefore $\mathbf{u}_t=\sigma_t\mathbf{x}$ and $\phi_t(\mathbf{x})=\boldsymbol{\mu}_t+\sigma_t\mathbf{x}$. Hence the proof is completed. □

Proposition 3.1. *The following OT interpolant path minimizes the flow energy* (15)*:*

$$\boldsymbol{\mu}_t=(1-t)\boldsymbol{\mu}_0+t\boldsymbol{\mu}_1=t\mathbf{x}_1,\quad \sigma_t=(1-t)\sigma_0+t\sigma_1=1-t+t\sigma_{\min}. \tag{16}$$

The minimal energy attained is $e_{\min}=\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_0\|^2+d|\sigma_1-\sigma_0|^2$.

Proof of Proposition 3.1. We compute the energy (15) of the flow path

$$\begin{aligned} e(\mathbf{v}) &= \sum_{i=1}^{d} \int_{\mathbb{R}^d} \int_0^1 p_t(\mathbf{x}) v_t^2(x_i) dt d\mathbf{x} \\ &= \sum_{i=1}^{d} \int_{\mathbb{R}} \int_0^1 p_t(x_i) [\dot{\mu}_{t,i} + (x_i - \mu_{t,i}) \dot{\sigma}_t / \sigma_t]^2 dt dx_i \\ &= \int_0^1 \sum_{i=1}^{d} [(\dot{\mu}_{t,i})^2 + (\dot{\sigma}_t)^2] dt \\ &= \int_0^1 L(\boldsymbol{\mu}_t, \sigma_t, \dot{\boldsymbol{\mu}}_t, \dot{\sigma}_t) dt, \quad L(\boldsymbol{\mu}_t, \sigma_t, \dot{\boldsymbol{\mu}}_t, \dot{\sigma}_t) = \|\dot{\boldsymbol{\mu}}_t\|^2 + d(\dot{\sigma}_t)^2 \end{aligned}$$

By the calculus of variation, the Euler–Lagrange equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{\mu}}_t} - \frac{\partial L}{\partial \boldsymbol{\mu}_t} = 0, \quad \frac{d}{dt}\frac{\partial L}{\partial \dot{\sigma}_t} - \frac{\partial L}{\partial \sigma_t} = 0$$

Solving these equations yields the desired geodesic path affine in t.

Substituting (16) into the above energy, we get the minimal value

$$\min e(\mathbf{v}) = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|^2 + d|\sigma_1 - \sigma_0|^2.$$

□

Lemma A.1. *The exponential power distribution $\mathcal{P}_q(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ has the density function $p(\mathbf{x}) = \frac{q\Gamma(\frac{d}{2})}{2\Gamma(\frac{d}{q})} 2^{-\frac{d}{q}} \pi^{-\frac{d}{2}} \sigma^{-d} \exp\left\{-\frac{r^{\frac{q}{2}}}{2}\right\}$, with $r(\mathbf{x}) = \|\mathbf{x} - \boldsymbol{\mu}\|^2/\sigma^2$. Its Fisher information matrix and Fisher-Rao metric are*

$$g(\boldsymbol{\mu}, \sigma) = \frac{1}{\sigma^2} \begin{bmatrix} c_\mu \mathbf{I} & \mathbf{0}^\mathsf{T} \\ \mathbf{0} & c_\sigma \end{bmatrix}, \quad \textit{and} \quad ds^2 = \frac{c_\mu \|d\boldsymbol{\mu}\|^2 + c_\sigma d\sigma^2}{\sigma^2}, \tag{23}$$

where $c_\mu(d, q) = \frac{2^{-\frac{2}{q}}}{d} \frac{\Gamma(\frac{d-2}{q}+2)}{\Gamma(\frac{d}{q})} q^2$, and $c_\sigma = qd$.

Proof. By property 2 of Proposition 2.1, $S = r^{\frac{q}{2}} = (\|\mathbf{x} - \boldsymbol{\mu}\|/\sigma)^q \sim \Gamma(\alpha = \frac{d}{q}, \beta = \frac{1}{2})$ (Gómez et al., 1998; Li et al., 2023). The log-density and their derivatives are

$$\begin{aligned} l(\boldsymbol{\mu}, \sigma) =& -d \log \sigma - \frac{1}{2} r^{\frac{q}{2}} + \text{const.} \\ \frac{\partial l}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}, \sigma) =& \frac{q}{2\sigma} S^{1-\frac{1}{q}} \mathbf{u}, \quad \mathbf{u} = \frac{\mathbf{x} - \boldsymbol{\mu}}{\|\mathbf{x} - \boldsymbol{\mu}\|} \sim \text{unif}(\mathcal{S}^{d-1}) \\ \frac{\partial l}{\partial \sigma}(\boldsymbol{\mu}, \sigma) =& \frac{1}{\sigma}\left[-d + \frac{q}{2} S\right] \\ \frac{\partial^2 l}{\partial \boldsymbol{\mu} \partial \sigma} =& -\frac{q^2}{2\sigma^2} S^{1-\frac{1}{q}} \mathbf{u} \\ \frac{\partial^2 l}{\partial \sigma^2}(\boldsymbol{\mu}, \sigma) =& \frac{1}{\sigma^2}\left[d - \frac{q}{2}(1+q) S\right] \end{aligned}$$

By the moments of gamma distribution $\mathbb{E}[S^\alpha] = 2^\alpha \frac{\Gamma(\frac{d}{q}+\alpha)}{\Gamma(\frac{d}{q})}$, we have the following Fisher information matrix $\mathcal{I}$:

$$\begin{aligned} \mathcal{I}_{\boldsymbol{\mu}\boldsymbol{\mu}} &= \mathbb{E}\left[\frac{\partial l}{\partial \boldsymbol{\mu}} \frac{\partial l}{\partial \boldsymbol{\mu}}^\mathsf{T}\right] = \frac{c_\mu}{\sigma^2}\mathbf{I}, \quad c_\mu(d, q) = \frac{2^{-\frac{2}{q}}}{d} \frac{\Gamma(\frac{d-2}{q}+2)}{\Gamma(\frac{d}{q})} q^2 \\ \mathcal{I}_{\boldsymbol{\mu}\sigma} &= -\mathbb{E}\left[\frac{\partial^2 l}{\partial \boldsymbol{\mu} \partial \sigma}\right] = \mathbf{0}, \quad \mathcal{I}_{\sigma\sigma} = \mathbb{E}\left[\left(\frac{\partial l}{\partial \sigma}\right)^2\right] = -\mathbb{E}\left[\frac{\partial^2 l}{\partial \sigma^2}\right] = \frac{c_\sigma}{\sigma^2},\ c_\sigma = qd \end{aligned}$$

Therefore, the Fisher metric as written below is a hyperbolic metric

$$g(\boldsymbol{\mu}, \sigma) = \frac{1}{\sigma^2} \begin{bmatrix} c_\mu \mathbf{I} & \mathbf{0}^\top \\ \mathbf{0} & c_\sigma \end{bmatrix}, \quad or \quad ds^2 = \frac{c_\mu \|d\boldsymbol{\mu}\|^2 + c_\sigma d\sigma^2}{\sigma^2}.$$

□

Theorem 4.1. *The following geodesic path minimizes the variational energy* (18)*:*

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + x(t)\mathbf{e}, \quad x(t) = \mu_c + R\cos\theta(t), \quad R = \sqrt{\mu_c^2 + \lambda^2\sigma_0^2}; \tag{19a}$$

$$\sigma_t = \frac{R}{\lambda}\sin\theta(t), \quad \theta(t) = 2\arctan\left(e^{Lt}\tan\frac{\theta_0}{2}\right), \quad L = \log\frac{\tan(\theta_1/2)}{\tan(\theta_0/2)}, \tag{19b}$$

where $\lambda = \sqrt{c_\sigma/c_\mu}$, $\mathbf{e} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)/\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|_2$, $\mu_c = \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|^2 + \lambda^2(\sigma_1^2 - \sigma_0^2)}{2\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|}$, *and* $\theta_0 = \operatorname{atan2}(\lambda\sigma_0, -\mu_c)$, $\theta_1 = \operatorname{atan2}(\lambda\sigma_1, \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\| - \mu_c)$. *The minimal energy attained is* $e_{\min} = c_\sigma L^2$*, where* $\sqrt{c_\sigma}|L|$ *is the Fisher-Rao distance between* $\mathcal{P}_q(\boldsymbol{\mu}_0, \sigma_0^2\mathbf{I})$ *and* $\mathcal{P}_q(\boldsymbol{\mu}_1, \sigma_1^2\mathbf{I})$*, for which* $|L|$ *can be rewritten in the original parameters as*

$$|L| = \operatorname{arcosh}\left(1 + \frac{c_\mu\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|^2 + c_\sigma|\sigma_1 - \sigma_0|^2}{2c_\sigma\sigma_0\sigma_1}\right). \tag{20}$$

Remark 4. *When* $q = 2$, $c_\mu \equiv 1$ *and* $c_\sigma = 2d$*. The above Fisher-Rao formula yields the well-known geodesic distance between two Gaussians* $\mathcal{N}(\boldsymbol{\mu}_0, \sigma_0^2\mathbf{I})$ *and* $\mathcal{N}(\boldsymbol{\mu}_1, \sigma_1^2\mathbf{I})$*:*

$$d_{FR} = \sqrt{2d}\operatorname{arcosh}\left(1 + \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|^2 + 2d|\sigma_1 - \sigma_0|^2}{4d\sigma_0\sigma_1}\right).$$

Proof of Theorem 4.1. The geodesic equation under the hyperbolic metric $ds^2 = \frac{c_\mu\|d\boldsymbol{\mu}\|^2 + c_\sigma d\sigma^2}{\sigma^2}$ (23) can be solved in either semi-circle parametrization or hyperbolic parametrization, which are detailed as follows.

Poincaré semi-circle parametrization The metric (23) can be rescaled to a standard Poincaré metric which leads to a semi-circle in the Poincaré upper half-plane. Denote $\lambda = \sqrt{c_\sigma/c_\mu}$ and $\mathbf{e} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)/\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|_2$. Use the scalar projection $x(t) = \langle\boldsymbol{\mu}(t) - \boldsymbol{\mu}_0, \mathbf{e}\rangle$ and rescale $y(t) = \lambda\sigma(t)$. The metric (23) reduces to

$$ds^2 = \frac{c_\sigma}{y^2}(dx^2 + dy^2).$$

By calculus of variation ($L(x, y, \dot{x}, \dot{y}) = \frac{c_\sigma(\dot{x}^2 + \dot{y}^2)}{y^2}$), we can get

$$(x - \mu_c)^2 + y^2 = R^2, \quad \mu_c = \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|^2 + \lambda^2(\sigma_1^2 - \sigma_0^2)}{2\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|}, \quad R = \sqrt{\mu_c^2 + \lambda^2\sigma_0^2}.$$

where μ_c and R are determined by the boundary conditions $(x_0, y_0) = (0, \lambda\sigma_0)$ and $(x_1, y_1) = (\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|, \lambda\sigma_1)$. Then the semi-circle can be parametrized by angle $\theta(t) \in (0, \pi)$: $x(t) = \mu_c + R\cos\theta(t), y(t) = R\sin\theta(t)$ with two end points $\theta(0) = \theta_0$ and $\theta(1) = \theta_1$ corresponding to $(x_0, y_0) = (0, \lambda\sigma_0)$ and $(x_1, y_1) = (\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|, \lambda\sigma_1)$ respectively:

$$\theta_0 = \operatorname{atan2}(\lambda\sigma_0, -\mu_c), \quad \theta_1 = \operatorname{atan2}(\lambda\sigma_1, \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\| - \mu_c),$$

where $\operatorname{atan2}(y, x) = \arg(x + iy) \in (-\pi, \pi)$.

Note the arc-length along the Poincaré semi-circle has constant speed L: $ds = \frac{\sqrt{dx^2 + dy^2}}{y} = \frac{R|d\theta|}{R\sin\theta} = Ldt$. Therefore, we have

$$\theta(t) = 2\arctan\left(e^{Lt}\tan\frac{\theta_0}{2}\right), \quad L = \log\frac{\tan(\theta_1/2)}{\tan(\theta_0/2)}.$$

Therefore, the final geodesic equation in the semi-circle parameterization of Poincaré becomes

$$\boldsymbol{\mu}(t) = \boldsymbol{\mu}_0 + x(t)\mathbf{e}, \quad x(t) = \mu_c + R\cos\theta(t);$$
$$\sigma(t) = \frac{R}{\lambda}\sin\theta(t).$$

Substituting the geodesic solution back to the energy equation, we have

$$e(\mathbf{v}) = \int_\gamma ds^2 = c_\sigma \int_0^1 \frac{\dot{x}^2 + \dot{y}^2}{y^2} dt = c_\sigma \int_0^1 \frac{\dot{\theta}^2(t)}{\sin^2\theta(t)} dt = c_\sigma L^2,$$

where $|L|$ is the Fisher-Rao distance between two distributions, which can be rewritten in the original parameters as

$$|L| = \operatorname{arcosh}\left(1 + \frac{c_\mu \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|^2 + c_\sigma |\sigma_1 - \sigma_0|^2}{2c_\sigma \sigma_0 \sigma_1}\right).$$

Hyperbolic parametrization With Einstein's notation, $g_{\mu_i\mu_j} = \frac{c_\mu}{\sigma^2}\delta_{ij}$, $g_{\sigma\sigma} = \frac{c_\sigma}{\sigma^2}$, we compute the Christoffel symbols $\Gamma^\lambda_{\mu\nu} = \frac{1}{2}g^{\lambda\rho}[\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu}]$ with inverse metric $g^{\mu_i\mu_j} = \frac{\sigma^2}{c_\mu}\delta_{ij}$, $g^{\sigma\sigma} = \frac{\sigma^2}{c_\sigma}$. The only nonzero partial derivatives come from $\partial_\sigma g = -\frac{2}{\sigma}g$, giving

$$\Gamma^{\mu_i}_{\mu_j\sigma} = -\frac{1}{\sigma}\delta_{ij}, \quad \Gamma^\sigma_{\mu_i\mu_j} = \frac{c_\mu}{c_\sigma\sigma}\delta_{ij}, \quad \Gamma^\sigma_{\sigma\sigma} = -\frac{1}{\sigma}.$$

Therefore, the geodesic equations $\ddot{x}^\lambda + \Gamma^\lambda_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = 0$ become

$$\ddot{\mu}_i - \frac{2}{\sigma}\dot{\mu}_i\dot{\sigma} = 0, \tag{24a}$$

$$\ddot{\sigma} + \frac{c_\mu}{c_\sigma\sigma}\|\dot{\boldsymbol{\mu}}\|^2 - \frac{1}{\sigma}\dot{\sigma}^2 = 0. \tag{24b}$$

Solving (24a) yields $\dot{\boldsymbol{\mu}} = \mathbf{c}\sigma^2$. Substituting it into (24b) gives $\ddot{\sigma} + \frac{c_\mu}{c_\sigma}\|\mathbf{c}\|^2\sigma^3 - \frac{1}{\sigma}\dot{\sigma}^2 = 0$. By changing the variable $\sigma = e^\tau$, (24b) becomes a standard autonomous second-order ODE

$$\ddot{\tau} + \frac{c_\mu}{c_\sigma}\|\mathbf{c}\|^2 e^{2\tau} = 0,$$

which implies the following conserved energy:

$$\dot{\tau}^2 + \frac{c_\mu}{c_\sigma}\|\mathbf{c}\|^2 e^{2\tau} = \frac{E}{c_\sigma}, \quad or \quad \frac{c_\mu\|\mathbf{c}\|^2\sigma^4 + c_\sigma\dot{\sigma}^2}{\sigma^2} = E. \tag{25}$$

Denote $\omega = \sqrt{E/c_\sigma}$, $\kappa = \|\mathbf{c}\|\sqrt{c_\mu/c_\sigma}$. This energy equation (25) can be rewritten as a separable ODE $\dot{\sigma}^2 = \omega^2\sigma^2 - \kappa^2\sigma^4$ which has the following solution

$$\sigma(t) = \frac{\omega}{\kappa}\operatorname{sech}(\omega t + \alpha),$$

where ω, κ and α can be determined by the boundary conditions. Using (24a), or equivalent $\dot{\boldsymbol{\mu}} = \mathbf{c}\sigma^2$, one can solve

$$\boldsymbol{\mu}(t) = \boldsymbol{\mu}_0 + \mathbf{c}\frac{\omega}{\kappa^2}(\tanh(\omega t + \alpha) - \tanh(\alpha)),$$

where $\mathbf{c} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\frac{\kappa^2}{\omega(\tanh(\omega+\alpha)-\tanh(\alpha))}$. We also have the conserved energy $E = c_\sigma\omega^2$ along the final geodesic:

$$\boldsymbol{\mu}(t) = \boldsymbol{\mu}_0 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\frac{\tanh(\omega t + \alpha) - \tanh(\alpha)}{\tanh(\omega + \alpha) - \tanh(\alpha)}; \tag{26a}$$

$$\sigma(t) = \frac{\omega}{\kappa}\operatorname{sech}(\omega t + \alpha), \tag{26b}$$

where ω, κ, α can be numerically solved from the following algebraic equations:

$$\frac{\omega}{\kappa}\operatorname{sech}(\alpha) = \sigma_0, \quad \frac{\omega}{\kappa}\operatorname{sech}(\omega + \alpha) = \sigma_1, \quad \kappa = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|\frac{\kappa^2\sqrt{c_\mu/c_\sigma}}{\omega(\tanh(\omega + \alpha) - \tanh(\alpha))}.$$

□

Remark 5. *Note, the Poincaré semi-circle parametrization is equivalent to this hyperbolic parametrization through the following identities:*

$$\sin(2\arctan e^{-s}) = \operatorname{sech}(s), \quad \cos(2\arctan e^{-s}) = \tanh(s), \quad s = \omega t + \alpha, \quad R = \frac{\omega}{\kappa}\lambda.$$

However, the former has fully explicit constants, and yet the latter contains implicit constants that need to be resolved numerically.

In particular, we get the vector field that generates the probability path by (12)*:*

$$\begin{aligned}\mathbf{v}_t(\mathbf{x}) &= \dot{\boldsymbol{\mu}}_t + \frac{\dot{\sigma}_t}{\sigma_t}(\mathbf{x} - \boldsymbol{\mu}_t) = \mathbf{c}\sigma_t^2 - \omega\tanh(\omega t + \alpha)[\mathbf{x} - \boldsymbol{\mu}_0 - \mathbf{c}\frac{\omega}{\kappa^2}(\tanh(\omega t + \alpha) - \tanh(\alpha))] \\ &= \omega\left[\frac{1 - \tanh(\omega t + \alpha)\tanh(\alpha)}{\tanh(\omega + \alpha) - \tanh(\alpha)}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) - \tanh(\omega t + \alpha)(\mathbf{x} - \boldsymbol{\mu}_0)\right].\end{aligned}$$

Letting $\boldsymbol{\mu}_0 = \mathbf{0}, \boldsymbol{\mu}_1 = \mathbf{x}_1$, $\sigma_0 = 1, \sigma_1 = \sigma_{\min}$*, we have the conditional vector field*

$$\mathbf{u}_t(\mathbf{x}|\mathbf{x}_1) = \omega\left[\frac{1 - \tanh(\omega t + \alpha)\tanh(\alpha)}{\tanh(\omega + \alpha) - \tanh(\alpha)}\mathbf{x}_1 - \tanh(\omega t + \alpha)\mathbf{x}\right],$$

and the conditional flow

$$\psi_t(\mathbf{x}) = \mathbf{x}_1\frac{\tanh(\omega t + \alpha) - \tanh(\alpha)}{\tanh(\omega + \alpha) - \tanh(\alpha)} + \frac{\omega}{\kappa}\operatorname{sech}(\omega t + \alpha)\mathbf{x}.$$

Let $p_t(\cdot|\mathbf{x}_1) \sim \mathcal{P}_q(\boldsymbol{\mu}_t(\mathbf{x}_1), \sigma_t^2(\mathbf{x}_1)\mathbf{I})$ with $(\boldsymbol{\mu}_t, \sigma_t)$ as in Theorem 4.1. We have the marginal probability path $p_t(\mathbf{x}) = \int p_t(\mathbf{x}|\mathbf{x}_1)q(\mathbf{x}_1)d\mathbf{x}_1$, the *marginal* vector field and the FM loss:

$$\mathbf{u}_t(\mathbf{x}) = \mathbb{E}[\mathbf{u}_t(\mathbf{x}|\mathbf{x}_1)|\mathbf{x}_t = \mathbf{x}] = \frac{\int \mathbf{u}_t(\mathbf{x}|\mathbf{x}_1)p_t(\mathbf{x}|\mathbf{x}_1)q(\mathbf{x}_1)d\mathbf{x}_1}{p_t(\mathbf{x})}, \ \mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{t,p_t(\mathbf{x})}\|\mathbf{v}_t(\mathbf{x};\theta) - \mathbf{u}_t(\mathbf{x})\|_2^2.$$

The following theorem A.1 states that the CFM loss (14) along the geodesic in Theorem 4.1 decomposes into the FM loss and the variance of conditional vector field and is bounded from below.

Theorem A.1. *For every parameter* θ *in the neural network* $\mathbf{v}_t(\cdot;\theta)$*, we have*

$$\begin{aligned}\mathcal{L}_{CFM}(\theta) &= \mathcal{L}_{FM}(\theta) + C, \quad C := \mathbb{E}_{t,p_t(\mathbf{x})}\text{Var}_{\mathbf{x}_1|\mathbf{x}_t=\mathbf{x}}[\mathbf{u}_t(\mathbf{x}|\mathbf{x}_1)], \\ C &= \inf_\theta \mathcal{L}_{CFM}(\theta) \le \sigma_{\max}^2\kappa_{q,d}\mathbb{E}_{q(\mathbf{x}_1)}L^2(\mathcal{P}_q(\mathbf{0},\mathbf{I}), \mathcal{P}_q(\mathbf{x}_1, \sigma_{\min}^2\mathbf{I})),\end{aligned}$$

where C *does not depend on the neural network* $\mathbf{v}_t(\cdot;\theta)$*,* $\sigma_{\max}$*,* $\kappa_{q,d}$ *are constants depending on the geodesic, and* $L^2(\mathcal{P}_q(\mathbf{0},\mathbf{I}), \mathcal{P}_q(\mathbf{x}_1, \sigma_{\min}^2\mathbf{I}))$ *is the scaled square Fisher-Rao distance between* $\mathcal{P}_q(\mathbf{0},\mathbf{I})$ *and* $\mathcal{P}_q(\mathbf{x}_1, \sigma_{\min}^2\mathbf{I})$*.*

Proof. Denote $\mathbf{x}_t := \psi_t(\mathbf{x}_0)$ as the conditional flow (6). With $q(\mathbf{x}_1)p_t(\mathbf{x}_t|\mathbf{x}_1) = p_t(\mathbf{x}_t)p(\mathbf{x}_1|\mathbf{x}_t)$, the CFM loss (14) can be rewritten

$$\begin{aligned}\mathcal{L}_{\text{CFM}}(\theta) &= \mathbb{E}_t\mathbb{E}_{q(\mathbf{x}_1)}\mathbb{E}_{p_t(\mathbf{x}_t|\mathbf{x}_1)}\|\mathbf{v}_t(\mathbf{x}_t;\theta) - \mathbf{u}_t(\mathbf{x}_t|\mathbf{x}_1)\|_2^2 \\ &= \mathbb{E}_t\mathbb{E}_{p_t(\mathbf{x}_t)}\mathbb{E}_{p(\mathbf{x}_1|\mathbf{x}_t)}\|\mathbf{v}_t(\mathbf{x}_t;\theta) - \mathbf{u}_t(\mathbf{x}_t|\mathbf{x}_1)\|_2^2 \\ &= \underbrace{\mathbb{E}_t\mathbb{E}_{p_t(\mathbf{x}_t)}\|\mathbf{v}_t(\mathbf{x}_t;\theta) - \mathbf{u}_t(\mathbf{x}_t)\|_2^2}_{\mathcal{L}_{\text{FM}}(\theta)} + \underbrace{\mathbb{E}_t\mathbb{E}_{p_t(\mathbf{x}_t)}\mathbb{E}_{p(\mathbf{x}_1|\mathbf{x}_t)}\|\mathbf{u}_t(\mathbf{x}_t|\mathbf{x}_1) - \mathbf{u}_t(\mathbf{x}_t)\|_2^2}_{C}.\end{aligned}$$

where $C = \mathbb{E}_{t,p_t(\mathbf{x})}\text{Var}_{\mathbf{x}_1|\mathbf{x}_t=\mathbf{x}}[\mathbf{u}_t(\mathbf{x}|\mathbf{x}_1)]$ based on the definition of marginal vector field $\mathbf{u}_t(\mathbf{x})$. Since $\mathcal{L}_{\text{FM}}(\theta) \ge 0$, we get $\mathcal{L}_{\text{CFM}}(\theta) \ge C$, with equality achieved iff $\|\mathbf{v}_t(\mathbf{x}_t;\theta) - \mathbf{u}_t(\mathbf{x}_t)\| = 0$ for p_t-a.e $(\mathbf{x}, t)$. Therefore, C is the lower bound of the CFM loss (14), regardless of the network $\mathbf{v}_t(\cdot;\theta)$.

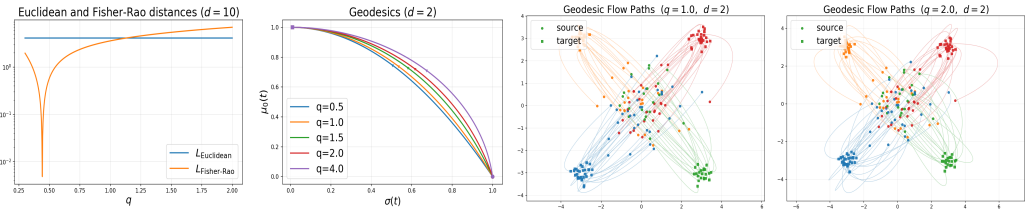

Figure A.1: Euclidean vs Fisher-Rao distances, (probabilistic) geodesic probability paths, flow paths for $q = 1.0$ and $q = 2.0$ (from left to right) on the manifold of exponential power distributions.

To get the upper bound of C for the geodesic path, we notice that with conditional vector field (7)

$$\begin{aligned}
C &\leq \mathbb{E}_t \mathbb{E}_{q(\mathbf{x}_1)} \mathbb{E}_{p_t(\mathbf{x}|\mathbf{x}_1)} \|\mathbf{u}_t(\mathbf{x}|\mathbf{x}_1)\|^2 \\
&= \mathbb{E}_t \mathbb{E}_{q(\mathbf{x}_1)} [\|\dot{\boldsymbol{\mu}}_t(\mathbf{x}_1)\|^2 + (\dot{\sigma}_t(\mathbf{x}_1)/\sigma_t(\mathbf{x}_1))^2 \mathbb{E}_{p_t(\mathbf{x}|\mathbf{x}_1)} \|\mathbf{x} - \boldsymbol{\mu}_t(\mathbf{x}_1)\|^2] \\
&\overset{prop 2.1}{=} \mathbb{E}_{q(\mathbf{x}_1)} \int_0^1 [\|\dot{\boldsymbol{\mu}}_t(\mathbf{x}_1)\|^2 + s_{q,d} d \dot{\sigma}_t^2(\mathbf{x}_1)] dt \\
&\leq \mathbb{E}_{q(\mathbf{x}_1)} \int_0^1 \sigma_t^2 \max\left\{\frac{1}{c_\mu}, \frac{s_{q,d} d}{c_\sigma}\right\} \frac{c_\mu \|\dot{\boldsymbol{\mu}}_t\|^2 + c_\sigma \dot{\sigma}_t^2}{\sigma_t^2} dt \\
&\leq \sigma_{\max}^2 \kappa_{q,d} \mathbb{E}_{q(\mathbf{x}_1)} L^2(\mathcal{P}_q(0, \mathbf{I}), \mathcal{P}_q(\mathbf{x}_1, \sigma_{\min}^2 \mathbf{I})),
\end{aligned}$$

where $\sigma_{\max} := \sup_{t \in [0,1], \mathbf{x}_1 \in \text{supp}(q(\mathbf{x}_1))} \sigma_t^2(\mathbf{x}_1)$, and $\kappa_{q,d} = \max\left\{\frac{1}{c_\mu}, \frac{s_{q,d} d}{c_\sigma}\right\}$. □

Remark 6. *Note that the CFM loss bound C depends only on the probability path $p_t(\cdot|\mathbf{x})$ and the data distribution $q(\cdot)$, not the neural network $\mathbf{v}(\cdot, \theta)$.*

Remark 7. *With the OT-interpolant path* (9)*, the CFM loss bound collapses to $C = \mathbb{E}_{t, p_t(\mathbf{x})} \text{Var}_{\mathbf{x}_1|\mathbf{x}_t}[\mathbf{x}_1] \left(1 + \frac{(1-\sigma_{\min})t}{\sigma_t}\right)^2$, which peaks around $t \approx \frac{1}{2}$, where the posterior $p(\mathbf{x}_1|\mathbf{x}_t)$ is most uncertain. Correspondingly, an "energy hump" may appear, as illustrated in Figure 2.*

Remark 8. *The proof suggests a more natural choice of loss based on the energy* (18)*, i.e. replacing the L_2 norm with the Fisher-Rao norm in* (14)*. That would yield a tighter bound for C.*

B More Numerical Results

B.1 Setup for Numerical Experiments

We follow the framework of (Lipman et al., 2023, CFM). For all simulations, we use a 5-layer dense neural network with sinusoidal (Fourier) time encoding and 512 hidden dimensions and train it with the ADAM optimizer with learning rate $1e-4$. We set $\sigma_{\min} = 1e-3$ for PG/HB and $1e-3$ for other baselines. In the sampling/inference stage, we adopt the 'midpoint' method to solve the flow ODE (1) with step size 0.05.

For image applications (MNIST), we use a classic UNet for MNIST (64 base channels, 2 residual blocks) with sinusoidal time embedding trained by AdamW at a learning rate $2e-4$. The exponential moving average (ema) is also adopted with decay rate 0.999. In sampling, the same 'midpoint' method is used for solving ODE with 100 discretized steps. FID scores are evaluated based on 50k samples.

We refer to Wasserstein-2 distance ($W_2^2(P, Q) = \inf_{\pi \in \Pi(P,Q)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|_2^2 d\pi(x, y)$), energy distance ($\text{ED}^2(P, Q) = 2\mathbb{E}\|X-Y\| - \mathbb{E}\|X-X'\| - \mathbb{E}\|Y-Y'\|$ for $X, X' \overset{iid}{\sim} P$ and $Y, Y' \overset{iid}{\sim} Q$ independent), and maximum mean discrepancy ($\text{MMD}^2(P, Q; k) = \mathbb{E}_{X,X'} k(X, X') + \mathbb{E}_{Y,Y'} k(Y, Y') - 2\mathbb{E}_{X,Y} k(X, Y)$) with radial basis function (rbf) kernel k to measure the discrepancy between synthetic sample distribution and real data distribution. The training time per epoch, 'time/epoch', is reported in Tables 1, B.1, B.4, and B.5. The sampler for the exponential power distribution is based on Proposition 2.1.

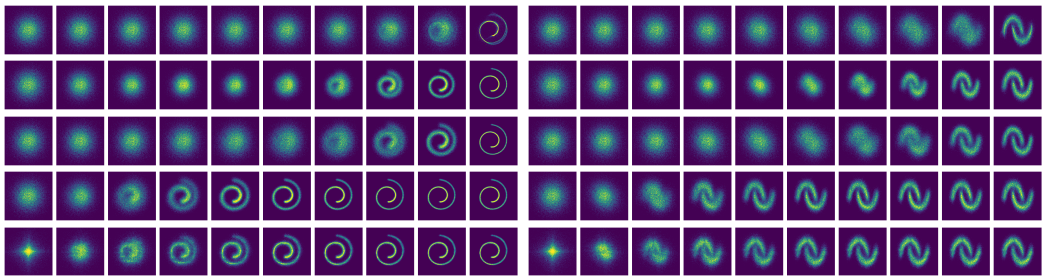

Figure B.1: Sample paths in the Swiss roll (left) and moons (right) examples: VP, OT, sino, PG-2, and PG-1 (from top to bottom).

B.2 2D BENCHMARKS

Table B.1: Swiss roll dataset comparison in W_2, ED, and MMD (mean $\pm$ std over 10 seeds).

DIST	PATH	$W_2 \downarrow$	ED $\downarrow$	MMD $\downarrow$	time/epoch
q=1.0	VP	$1.49e{-1} \pm 3.5e{-2}$	$3.69e{-2} \pm 1.5e{-2}$	$3.33e{-4} \pm 2.2e{-4}$	$5.84e{-2} \pm 6.1e{-4}$
q=2.0	VP	$1.42e{-1} \pm 4.5e{-2}$	$2.93e{-2} \pm 2.3e{-2}$	$3.06e{-4} \pm 3.6e{-4}$	$6.24e{-2} \pm 5.3e{-4}$
q=1.0	OT	$1.14e{-1} \pm 2.3e{-2}$	$1.56e{-2} \pm 1.1e{-2}$	$7.50e{-5} \pm 9.1e{-5}$	$5.66e{-2} \pm 5.4e{-4}$
q=2.0	OT	$1.12e{-1} \pm 3.0e{-2}$	$1.69e{-2} \pm 1.6e{-2}$	$9.03e{-5} \pm 1.7e{-4}$	$5.40e{-2} \pm 6.2e{-4}$
q=1.0	sino	$1.33e{-1} \pm 2.5e{-2}$	$2.53e{-2} \pm 1.3e{-2}$	$1.87e{-4} \pm 1.2e{-4}$	$5.74e{-2} \pm 3.7e{-4}$
q=2.0	sino	$1.17e{-1} \pm 1.4e{-2}$	$1.24e{-2} \pm 8.3e{-3}$	$6.05e{-5} \pm 5.0e{-5}$	$5.46e{-2} \pm 4.2e{-4}$
q=1.0	PG	$1.05e{-1} \pm 2.4e{-2}$	$9.56e{-3} \pm 7.8e{-3}$	$5.27e{-5} \pm 8.4e{-5}$	$1.05e{-1} \pm 3.2e{-4}$
q=2.0	PG	$1.08e{-1} \pm 2.0e{-2}$	$8.66e{-3} \pm 8.0e{-3}$	$3.55e{-5} \pm 6.3e{-5}$	$1.10e{-1} \pm 9.9e{-3}$

Table B.2: Comparison on Moons and Checkerboard datasets in W_2, ED, and MMD (mean $\pm$ std over 10 seeds).

		Moons			Checkerboard		
DIST	PATH	$W_2 \downarrow$	ED $\downarrow$	MMD $\downarrow$	$W_2 \downarrow$	ED $\downarrow$	MMD $\downarrow$
q=1.0	VP	$1.33e{-1} \pm 2.4e{-2}$	$2.56e{-2} \pm 1.5e{-2}$	$2.14e{-4} \pm 1.8e{-4}$	$3.58e{-1} \pm 9.6e{-2}$	$4.45e{-2} \pm 2.9e{-2}$	$3.01e{-4} \pm 3.0e{-4}$
q=2.0	VP	$1.06e{-1} \pm 1.7e{-2}$	$1.00e{-2} \pm 1.0e{-2}$	$5.69e{-5} \pm 6.5e{-5}$	$3.53e{-1} \pm 6.2e{-2}$	$4.29e{-2} \pm 1.6e{-2}$	$2.05e{-4} \pm 1.2e{-4}$
q=1.0	OT	$1.17e{-1} \pm 2.1e{-2}$	$1.94e{-2} \pm 1.2e{-2}$	$1.08e{-4} \pm 1.2e{-4}$	$3.15e{-1} \pm 5.3e{-2}$	$3.23e{-2} \pm 1.7e{-2}$	$1.63e{-4} \pm 1.1e{-4}$
q=2.0	OT	$9.83e{-2} \pm 1.3e{-2}$	$9.39e{-3} \pm 8.6e{-3}$	$5.56e{-5} \pm 6.0e{-5}$	$2.96e{-1} \pm 4.1e{-2}$	$2.21e{-2} \pm 1.6e{-2}$	$9.36e{-5} \pm 7.4e{-5}$
q=1.0	sino	$1.25e{-1} \pm 1.8e{-2}$	$2.26e{-2} \pm 1.1e{-2}$	$1.59e{-4} \pm 1.1e{-4}$	$3.59e{-1} \pm 9.1e{-2}$	$4.76e{-2} \pm 2.4e{-2}$	$3.38e{-4} \pm 2.5e{-4}$
q=2.0	sino	$1.06e{-1} \pm 1.5e{-2}$	$1.22e{-2} \pm 1.0e{-2}$	$6.25e{-5} \pm 5.5e{-5}$	$3.01e{-1} \pm 4.5e{-2}$	$2.32e{-2} \pm 1.6e{-2}$	$1.06e{-4} \pm 7.0e{-5}$
q=1.0	PG	$1.02e{-1} \pm 1.3e{-2}$	$9.26e{-3} \pm 9.2e{-3}$	$6.94e{-5} \pm 8.4e{-5}$	$3.17e{-1} \pm 3.8e{-2}$	$2.74e{-2} \pm 1.6e{-2}$	$1.24e{-4} \pm 8.3e{-5}$
q=2.0	PG	$1.00e{-1} \pm 1.4e{-2}$	$5.62e{-3} \pm 8.0e{-3}$	$3.68e{-5} \pm 7.3e{-5}$	$3.05e{-1} \pm 3.2e{-2}$	$2.56e{-2} \pm 6.8e{-3}$	$1.04e{-4} \pm 3.0e{-5}$

B.3 HIGHER DIMENSIONAL SIMULATIONS

Next, we investigate the more challenging Neal's funnel distribution (Neal, 2003):

$$f(\mathbf{x}, \nu) = \mathcal{N}_d(\mathbf{x}; \mathbf{0}, e^{\nu/2}\mathbf{I})\mathcal{N}(\nu; 0, 3),$$

whose density features a cusp on one end and a fat tail on the other (refer to the right panel of Figure B.2, which also demonstrates PG methods' faster feature learning), posing difficulties for generative models to sample data from it. We consider $d = 10/100$ and generate 10,000 samples to train these FM models.

Tables 1 and B.5 further support the advantage and dimension robustness of PG methods as they achieve the lowest or comparable discrepancy metrics. Except a few cells in these tables, $q = 1.0$ almost unanimously beats $q = 2.0$ which again justifies the adoption of exponential power distribution when handling heavy-tail and inhomogeneity in the data. Note in Table 1, MFM reaches the same lowest ED at a much higher time cost ($4.36e-1$ vs $6.92e-2$ seconds per epoch for HB).

B.4 SCIENTIFIC DATASETS

Next, we investigate the following four datasets from molecular science, genetics, and quantum field theory: `Alanine dipeptide` (30d) is the standard small-molecule benchmark popular for

Table B.3: PG q-sweep comparison across Swiss roll, moons, and checkerboard in W_2 and ED (mean $\pm$ std over 10 seeds).

	Swiss Roll		Moons		Checkerboard	
DIST	$W_2 \downarrow$	ED $\downarrow$	$W_2 \downarrow$	ED $\downarrow$	$W_2 \downarrow$	ED $\downarrow$
q=0.5	$3.70e{-1} \pm 2.1e{-2}$	$6.76e{-2} \pm 5.8e{-3}$	$3.18e{-1} \pm 1.7e{-2}$	$6.14e{-2} \pm 4.6e{-3}$	$3.92e{-1} \pm 4.2e{-2}$	$3.56e{-2} \pm 1.3e{-2}$
q=0.8	$1.11e{-1} \pm 1.9e{-2}$	$1.02e{-2} \pm 8.8e{-3}$	$1.09e{-1} \pm 1.6e{-2}$	$1.41e{-2} \pm 9.6e{-3}$	$3.19e{-1} \pm 4.1e{-2}$	$2.80e{-2} \pm 1.3e{-2}$
q=1.0	$1.05e{-1} \pm 2.4e{-2}$	$9.56e{-3} \pm 7.8e{-3}$	$1.02e{-1} \pm 1.3e{-2}$	$9.26e{-3} \pm 9.2e{-3}$	$3.17e{-1} \pm 3.8e{-2}$	$2.74e{-2} \pm 1.6e{-2}$
q=1.5	$1.06e{-1} \pm 1.4e{-2}$	$9.13e{-3} \pm 8.6e{-3}$	$9.95e{-2} \pm 1.2e{-2}$	$9.93e{-3} \pm 9.3e{-3}$	$3.19e{-1} \pm 5.4e{-2}$	$3.07e{-2} \pm 1.8e{-2}$
q=2.0	$1.08e{-1} \pm 2.0e{-2}$	$8.66e{-3} \pm 8.0e{-3}$	$1.00e{-1} \pm 1.4e{-2}$	$5.62e{-3} \pm 8.0e{-3}$	$3.05e{-1} \pm 3.2e{-2}$	$2.56e{-2} \pm 6.8e{-3}$
q=2.5	$1.15e{-1} \pm 1.8e{-2}$	$1.29e{-2} \pm 8.0e{-3}$	$9.80e{-2} \pm 8.5e{-3}$	$7.68e{-3} \pm 6.9e{-3}$	$3.37e{-1} \pm 4.5e{-2}$	$3.66e{-2} \pm 1.6e{-2}$

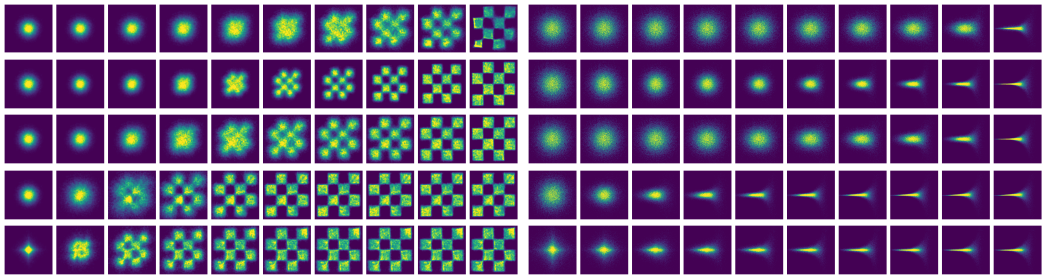

Figure B.2: Sample paths in the checkerboard (2d, left) and Neal's funnel (10d, right) examples: VP, OT, sino, PG-2, and PG-1 (from top to bottom).

generative modeling (Noé et al., 2019). `QM9` (87d) contains 133,885 stable small organic molecules with up to nine heavy atoms (C, N, O, F) from GDB-17 (Ruddigkeit et al., 2012; Ramakrishnan et al., 2014). `Single-cell (sc) RNA` (50d) contains 2,700 PBMCs from a healthy donor, freely available from 10x Genomics (Wolf et al., 2018). `Lattice` ϕ^4 (64d) is from the 2D scalar field theory, used as a benchmark for generative models (Albergo, 2019).

These datasets of various dimensions contain complex data structures and serve as competitive benchmarks to test FM generative models. In Figure B.4, we visualize the sample paths of the FM methods on alanine dipeptide and QM9 datasets by projecting them to 2d subspaces using the t-SNE algorithm. Again PG methods maintain high speed in feature discovery. The PG methods demonstrate better (Table B.7) or comparable (Table B.6) performance.

B.5 De novo Molecule Generation

The task is to draw a complete molecule from scratch: no reference structure is supplied, and a single sample must specify the 3D coordinates, the atom type and the formal charge of every atom, together with the bond order of every atom *pair*, jointly and consistently. Writing n for the number of atoms and d_a, d_c, d_e for the number of atom types, charge states and bond types, one molecule is a joint event of dimension

$$D(n) = \underbrace{3(n-1)}_{\text{coordinates}} + \underbrace{d_a\, n}_{\text{atom types}} + \underbrace{d_c\, n}_{\text{charges}} + \underbrace{d_e \binom{n}{2}}_{\text{bonds}} = \frac{d_e}{2}\, n^2 + \Big(3 + d_a + d_c - \frac{d_e}{2}\Big) n - 3, \qquad (27)$$

where the coordinate block is $3(n-1)$ rather than $3n$ because translations are quotiented out, i.e. the flow acts on the zero-centre-of-mass subspace. For QM9 ($d_a = 5$, $d_c = 6$, $d_e = 5$) this is $D(n) = \frac{5}{2} n^2 + \frac{23}{2} n - 3$, dominated by the bond block, which grows quadratically in n:

Modality	Dimension	$n = 3$	$n = 9$	$n = 18$	$n = 29$
Coordinates (zero-COM)	$3(n-1)$	6	24	51	84
Atom types	$d_a n$	15	45	90	145
Charges	$d_c n$	18	54	108	174
Bonds (upper triangle)	$d_e \binom{n}{2}$	15	180	765	2030
Total	$D(n)$	54	303	1014	2433

Table B.4: 8-Gaussians (36d): comparison in W_2, ED, and MMD (mean $\pm$ std over 10 seeds).

DIST	Method	$W_2 \downarrow$	ED $\downarrow$	MMD $\downarrow$	time/epoch
q=1.0	OT	$5.660e0 \pm 1.43e{-2}$	$5.93e{-2} \pm 1.25e{-2}$	$1.39e{-4} \pm 5.09e{-5}$	$5.92e{-2} \pm 1.27e{-4}$
q=2.0	OT	$5.493e0 \pm 1.41e{-2}$	$6.62e{-2} \pm 7.93e{-3}$	$2.47e{-4} \pm 4.99e{-5}$	$5.70e{-2} \pm 1.41e{-4}$
q=1.0	Sino	$5.499e0 \pm 1.03e{-2}$	$4.06e{-2} \pm 6.28e{-3}$	$9.67e{-5} \pm 2.37e{-5}$	$5.82e{-2} \pm 1.11e{-4}$
q=2.0	Sino	$5.329e0 \pm 2.86e{-3}$	$1.29e{-1} \pm 2.66e{-3}$	$1.17e{-3} \pm 4.85e{-5}$	$5.69e{-2} \pm 8.26e{-5}$
q=1.0	RFM	$5.49e0 \pm 8.79e{-3}$	$7.28e{-2} \pm 6.73e{-3}$	$2.93e{-4} \pm 4.30e{-5}$	$7.47e{-2} \pm 3.16e{-3}$
q=2.0	RFM	$5.50e0 \pm 1.61e{-2}$	$6.41e{-2} \pm 7.89e{-3}$	$2.24e{-4} \pm 6.10e{-5}$	$6.89e{-2} \pm 3.31e{-3}$
q=1.0	MFM	$5.58e0 \pm 1.45e{-2}$	$6.98e{-2} \pm 1.11e{-2}$	$2.12e{-4} \pm 6.18e{-5}$	$4.10e{-1} \pm 1.49e{-2}$
q=2.0	MFM	$5.58e0 \pm 1.32e{-2}$	$5.68e{-2} \pm 8.40e{-3}$	$1.48e{-4} \pm 3.73e{-5}$	$3.66e{-1} \pm 1.18e{-2}$
q=1.0	FFM	$5.79e0 \pm 2.29e{-2}$	$4.53e{-2} \pm 3.48e{-2}$	$2.14e{-4} \pm 2.82e{-4}$	$8.23e{-3} \pm 1.13e{-4}$
q=2.0	FFM	$5.80e0 \pm 3.78e{-2}$	$5.56e{-2} \pm 7.05e{-2}$	$4.89e{-4} \pm 1.27e{-3}$	$6.76e{-3} \pm 1.31e{-4}$
q=1.0	PG-iso	$5.399e0 \pm 2.74e{-2}$	$1.05e{-1} \pm 1.08e{-2}$	$7.48e{-4} \pm 1.62e{-4}$	$8.00e{-2} \pm 1.98e{-3}$
q=2.0	PG-iso	$5.455e0 \pm 7.64e{-3}$	$7.67e{-2} \pm 5.65e{-3}$	$3.93e{-4} \pm 5.85e{-5}$	$6.75e{-2} \pm 1.61e{-3}$
q=1.0	PG-aniso	$5.597e0 \pm 5.51e{-3}$	$4.04e{-2} \pm 6.23e{-3}$	$6.57e{-5} \pm 1.73e{-5}$	$6.59e{-2} \pm 2.54e{-4}$
q=2.0	PG-aniso	$5.563e0 \pm 1.64e{-2}$	$5.08e{-2} \pm 4.69e{-3}$	$1.50e{-4} \pm 2.22e{-5}$	$6.42e{-2} \pm 4.77e{-4}$
q=1.0	HB	$5.638e0 \pm 1.63e{-2}$	$4.44e{-2} \pm 1.16e{-2}$	$7.33e{-5} \pm 3.59e{-5}$	$7.84e{-2} \pm 9.44e{-4}$
q=2.0	HB	$5.571e0 \pm 1.16e{-2}$	$5.21e{-2} \pm 5.00e{-3}$	$1.37e{-4} \pm 2.62e{-5}$	$6.85e{-2} \pm 3.84e{-4}$

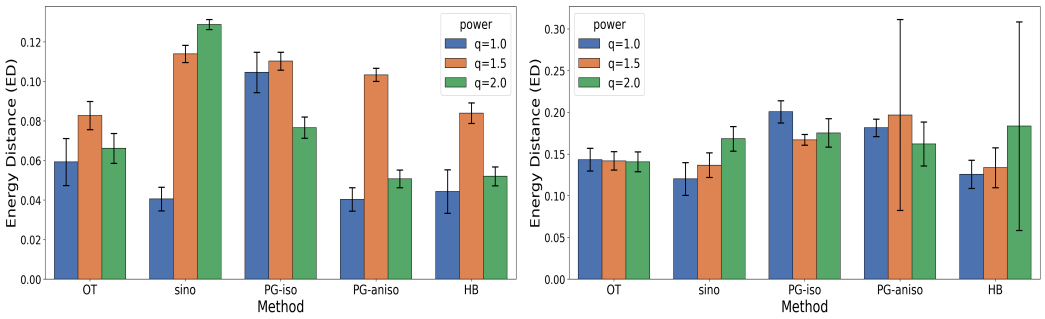

Figure B.3: Energy distance (ED) of OT, sino, PG-2, and PG-1 in the 8-Gaussians (36d, left) and funnel (100d, right) examples.

QM9 molecules contain at most nine heavy atoms and hence up to $n = 29$ atoms with hydrogens, averaging roughly $n \approx 18$: each molecule is an event of 54–2,433 dimensions, averaging roughly 10^3, of which the bond block alone accounts for 83% at $n = 29$. De novo generation on this benchmark is therefore a problem of order 10^3 dimensions per sample, despite QM9 being a small-molecule dataset.

Table B.5: Funnel example (10d) comparison in W_2, ED, and MMD (mean $\pm$ std over 10 seeds).

DIST	PATH	$W_2 \downarrow$	ED $\downarrow$	MMD $\downarrow$	time/epoch
q=1.0	VP	$2.55e0 \pm 1.0e{-1}$	$9.62e{-2} \pm 1.5e{-2}$	$3.04e{-3} \pm 1.2e{-3}$	$5.50e{-2} \pm 4.3e{-4}$
q=2.0	VP	$2.45e0 \pm 4.6e{-2}$	$9.48e{-2} \pm 3.3e{-2}$	$4.10e{-3} \pm 3.0e{-3}$	$5.27e{-2} \pm 1.2e{-4}$
q=1.0	OT	$2.47e0 \pm 9.2e{-2}$	$3.25e{-2} \pm 1.6e{-2}$	$2.36e{-4} \pm 1.5e{-4}$	$5.59e{-2} \pm 5.6e{-4}$
q=2.0	OT	$2.54e0 \pm 1.5e{-1}$	$2.69e{-2} \pm 1.3e{-2}$	$1.63e{-4} \pm 8.5e{-5}$	$5.37e{-2} \pm 4.0e{-4}$
q=1.0	sino	$2.51e0 \pm 8.7e{-2}$	$3.64e{-2} \pm 1.1e{-2}$	$2.76e{-4} \pm 1.7e{-4}$	$5.64e{-2} \pm 2.0e{-4}$
q=2.0	sino	$2.55e0 \pm 2.5e{-1}$	$3.07e{-2} \pm 8.6e{-3}$	$1.50e{-4} \pm 6.3e{-5}$	$5.30e{-2} \pm 1.3e{-3}$
q=1.0	PG-iso	$2.63e0 \pm 2.7e{-1}$	$3.54e{-2} \pm 1.1e{-2}$	$3.36e{-4} \pm 2.2e{-4}$	$8.00e{-2} \pm 4.8e{-4}$
q=2.0	PG-iso	$2.87e0 \pm 4.2e{-1}$	$3.44e{-2} \pm 1.6e{-2}$	$4.41e{-4} \pm 3.0e{-4}$	$7.78e{-2} \pm 6.5e{-4}$
q=1.0	PG-aniso	$2.61e0 \pm 1.9e{-1}$	$2.54e{-2} \pm 1.0e{-2}$	$2.53e{-4} \pm 1.4e{-4}$	$6.24e{-2} \pm 2.6e{-4}$
q=2.0	PG-aniso	$3.09e0 \pm 4.9e{-1}$	$4.54e{-2} \pm 1.2e{-2}$	$5.71e{-4} \pm 2.9e{-4}$	$6.04e{-2} \pm 2.2e{-4}$

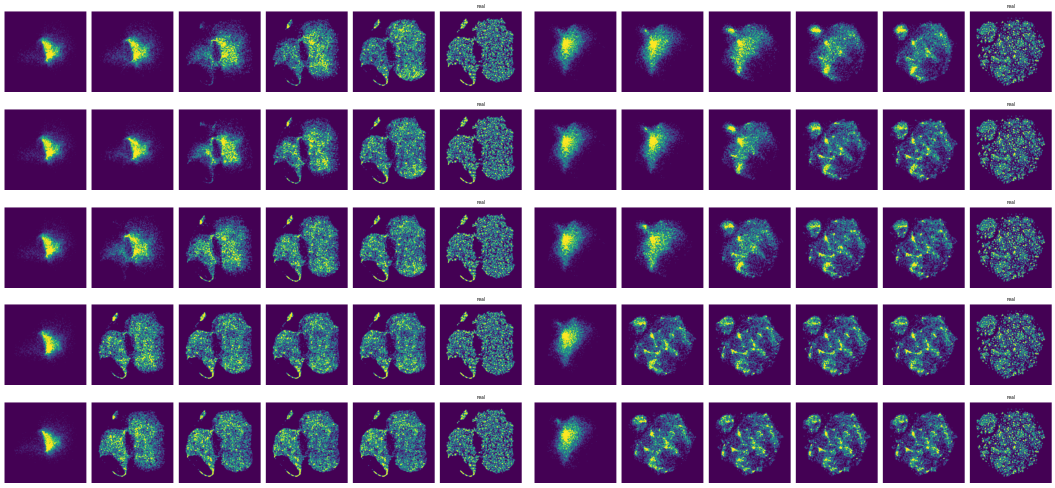

Figure B.4: Sample paths in the alanine dipeptide (30d, left) and QM9 (87d, right): VP, OT, sino, PG-2, and PG-1 (from top to bottom). The last column of each panel represents the real data. High dimensions have been projected to 2d using t-SNE for visualization.

Table B.6: Comparison across Alanine dipeptide (30d) and QM9 (87d) in W_2, ED, and MMD (mean $\pm$ std over 10 seeds).

		Alanine dipeptide (30d)			QM9 (87d)		
DIST	PATH	$W_2 \downarrow$	ED $\downarrow$	MMD $\downarrow$	$W_2 \downarrow$	ED $\downarrow$	MMD $\downarrow$
1.0	OT	$1.49e0 \pm 1.4e{-1}$	$7.75e{-2} \pm 3.1e{-2}$	$4.02e{-4} \pm 2.6e{-4}$	$7.27e0 \pm 8.8e{-1}$	$1.53e{-1} \pm 1.2e{-2}$	$8.65e{-4} \pm 1.3e{-4}$
2.0	OT	$1.41e0 \pm 1.1e{-1}$	$5.62e{-2} \pm 2.3e{-2}$	$1.71e{-4} \pm 1.2e{-4}$	$6.84e0 \pm 3.2e{-1}$	$1.31e{-1} \pm 1.6e{-2}$	$6.88e{-4} \pm 1.8e{-4}$
1.0	sino	$1.56e0 \pm 1.0e{-1}$	$8.86e{-2} \pm 3.6e{-2}$	$5.00e{-4} \pm 3.0e{-4}$	$7.05e0 \pm 2.5e{-1}$	$1.27e{-1} \pm 1.6e{-2}$	$6.29e{-4} \pm 1.7e{-4}$
2.0	sino	$1.45e0 \pm 1.3e{-1}$	$6.59e{-2} \pm 3.3e{-2}$	$2.88e{-4} \pm 2.1e{-4}$	$7.04e0 \pm 3.4e{-1}$	$1.27e{-1} \pm 3.0e{-2}$	$6.33e{-4} \pm 3.4e{-4}$
1.0	PG	$1.62e0 \pm 1.8e{-1}$	$1.05e{-1} \pm 3.7e{-2}$	$7.40e{-4} \pm 4.9e{-4}$	$6.94e0 \pm 3.3e{-1}$	$1.63e{-1} \pm 1.9e{-2}$	$9.05e{-4} \pm 2.0e{-4}$
2.0	PG	$1.51e0 \pm 9.4e{-2}$	$8.48e{-2} \pm 2.5e{-2}$	$4.40e{-4} \pm 2.1e{-4}$	$7.00e0 \pm 3.3e{-1}$	$1.56e{-1} \pm 1.8e{-2}$	$8.54e{-4} \pm 2.2e{-4}$
1.0	HB	$1.73e0 \pm 1.9e{-1}$	$2.00e{-1} \pm 3.4e{-2}$	$3.22e{-3} \pm 9.7e{-4}$	$7.27e0 \pm 2.9e{-1}$	$2.02e{-1} \pm 2.0e{-2}$	$1.58e{-3} \pm 3.4e{-4}$
2.0	HB	$1.66e0 \pm 7.2e{-2}$	$1.92e{-1} \pm 1.6e{-2}$	$3.06e{-3} \pm 5.8e{-4}$	$7.48e0 \pm 2.9e{-1}$	$2.17e{-1} \pm 1.5e{-2}$	$1.85e{-3} \pm 3.0e{-4}$

Table B.7: Comparison across scRNA (50d) and Lattice ϕ^4 (64d) in W_2, ED, and MMD (mean $\pm$ std over 10 seeds).

		scRNA (50d)			Lattice ϕ^4 (64d)		
DIST	PATH	$W_2 \downarrow$	ED $\downarrow$	MMD $\downarrow$	$W_2 \downarrow$	ED $\downarrow$	MMD $\downarrow$
1.0	OT	$5.32e0 \pm 4.0e{-2}$	$5.17e{-1} \pm 2.0e{-2}$	$1.88e{-2} \pm 1.5e{-3}$	$7.54e0 \pm 9.0e{-3}$	$2.08e{-1} \pm 4.0e{-3}$	$2.18e{-3} \pm 8.0e{-5}$
2.0	OT	$5.33e0 \pm 5.0e{-2}$	$5.10e{-1} \pm 2.0e{-2}$	$1.82e{-2} \pm 1.3e{-3}$	$7.53e0 \pm 1.1e{-2}$	$2.06e{-1} \pm 6.0e{-3}$	$2.14e{-3} \pm 1.2e{-4}$
1.0	sino	$5.41e0 \pm 5.0e{-2}$	$5.36e{-1} \pm 2.0e{-2}$	$2.04e{-2} \pm 1.5e{-3}$	$7.50e0 \pm 1.0e{-2}$	$2.25e{-1} \pm 4.0e{-3}$	$2.64e{-3} \pm 1.1e{-4}$
2.0	sino	$5.44e0 \pm 5.0e{-2}$	$5.46e{-1} \pm 2.0e{-2}$	$2.09e{-2} \pm 1.4e{-3}$	$7.47e0 \pm 1.0e{-2}$	$2.25e{-1} \pm 4.0e{-3}$	$2.74e{-3} \pm 9.0e{-5}$
1.0	PG	$5.71e0 \pm 4.0e{-2}$	$2.93e{-1} \pm 2.2e{-2}$	$5.07e{-3} \pm 6.0e{-4}$	$7.73e0 \pm 1.4e{-2}$	$1.55e{-1} \pm 7.0e{-3}$	$1.06e{-3} \pm 8.0e{-5}$
2.0	PG	$5.75e0 \pm 5.0e{-2}$	$2.73e{-1} \pm 1.3e{-2}$	$4.14e{-3} \pm 4.0e{-4}$	$7.69e0 \pm 2.0e{-2}$	$1.48e{-1} \pm 7.0e{-3}$	$9.99e{-4} \pm 1.1e{-4}$
1.0	HB	$5.72e0 \pm 5.1e{-2}$	$1.83e{-1} \pm 2.1e{-2}$	$1.42e{-3} \pm 3.7e{-4}$	$8.31e0 \pm 1.5e{-2}$	$1.78e{-1} \pm 6.5e{-3}$	$1.21e{-3} \pm 9.6e{-5}$
2.0	HB	$6.02e0 \pm 4.1e{-2}$	$1.58e{-1} \pm 1.6e{-2}$	$8.18e{-4} \pm 1.6e{-4}$	$8.28e0 \pm 2.4e{-2}$	$1.60e{-1} \pm 1.2e{-2}$	$1.07e{-3} \pm 1.4e{-4}$

C ANISOTROPIC PROBABILISTIC GEODESIC FLOW

We now extend the geodesic flow of Section 4 to the *anisotropic* location–scale family in which the scalar scale σ_t is replaced by a per-coordinate scale vector $\boldsymbol{\sigma}_t \in \mathbb{R}_+^d$, so that the covariance becomes $\mathrm{diag}(\boldsymbol{\sigma}_t^2)$. Following the component-wise independence assumed in Theorem 3.1, we take the anisotropic exponential power conditional density to be the product of d univariate marginals,

$$p_t(\mathbf{x}; \boldsymbol{\mu}_t, \boldsymbol{\sigma}_t) \;=\; \prod_{i=1}^{d} p_t\big(x_i; \mu_{t,i}, \sigma_{t,i}^2\big), \quad p_t(x_i; \mu_i, \sigma_i^2) \;=\; \frac{q}{2^{1+\frac{1}{q}}\Gamma(1/q)\,\sigma_i} \exp\Big\{-\tfrac{1}{2}\big|(x_i-\mu_i)/\sigma_i\big|^q\Big\}. \tag{28}$$

The endpoints of the conditional flow are unchanged in form, $p_0(\cdot \mid \mathbf{x}_1) = \prod_i \mathcal{P}_q(\cdot; 0, 1)$ and $p_1(\cdot \mid \mathbf{x}_1) = \prod_i \mathcal{P}_q(\cdot; x_{1,i}, \sigma_{\min}^2)$, i.e. $\boldsymbol{\mu}_0 = \mathbf{0},\ \boldsymbol{\sigma}_0 = \mathbf{1},\ \boldsymbol{\mu}_1 = \mathbf{x}_1,\ \boldsymbol{\sigma}_1 = \sigma_{\min}\mathbf{1}$.

Anisotropic Fisher metric. Because (28) factorizes across coordinates and the parameters (μ_i, σ_i) of distinct factors are decoupled, the joint Fisher information is block-diagonal, $\mathbf{I}(\boldsymbol{\mu}, \boldsymbol{\sigma}) = \bigoplus_{i=1}^{d} \mathbf{I}^{(1)}(\mu_i, \sigma_i)$, where each 2×2 block is the univariate exponential power Fisher information obtained by specializing the derivation of Section 4 to $d = 1$:

$$\mathbf{I}^{(1)}(\mu_i, \sigma_i) \;=\; \frac{1}{\sigma_i^2}\begin{pmatrix} c_\mu & 0 \\ 0 & c_\sigma \end{pmatrix}, \quad c_\mu \;=\; 2^{-2/q} q^2\, \frac{\Gamma\big(2-\frac{1}{q}\big)}{\Gamma(1/q)}, \quad c_\sigma \;=\; q. \tag{29}$$

(The vanishing of the off-diagonal $\mu-\sigma$ block follows from $\mathbb{E}[u_i] = \mathbb{E}[u_i u_j^2] = 0$ on S^{d-1}, exactly as in the isotropic case.) Consequently the anisotropic Fisher-Rao metric is the product of d rescaled Poincaré half-plane metrics,

$$ds^2 \;=\; \sum_{i=1}^{d} \frac{c_\mu\, d\mu_i^2 + c_\sigma\, d\sigma_i^2}{\sigma_i^2} \;=\; \sum_{i=1}^{d} ds_i^2. \tag{30}$$

Per-coordinate geodesic equations. Because (30) is a Riemannian product, geodesics in the joint manifold $(\mathbb{R}\times\mathbb{R}_+)^d$ are *Cartesian products of geodesics in each H^2 factor*; the Christoffel symbols are nonzero only *within* a coordinate. Repeating the computation of Section 4 in every factor yields, for each $i \in \{1, \dots, d\}$,

$$\ddot{\mu}_i \;-\; \tfrac{2}{\sigma_i}\,\dot{\mu}_i\,\dot{\sigma}_i = 0, \tag{31a}$$

$$\ddot{\sigma}_i \;+\; \tfrac{c_\mu}{c_\sigma \sigma_i}\,\dot{\mu}_i^2 \;-\; \tfrac{1}{\sigma_i}\,\dot{\sigma}_i^2 = 0. \tag{31b}$$

Each pair (31) is identical to the isotropic system (21a)–(21b) of Section 4, decoupled across i.

Closed-form solution: Poincaré semi-circle (per coordinate). Set $\lambda := \sqrt{c_\sigma / c_\mu}$ and, for each coordinate i, write $\Delta_i := \mu_{1,i} - \mu_{0,i}$. *Case* $\Delta_i \neq 0$. The geodesic in the ith H^2 factor is the unique semi-circle in the $(\mu_i, \lambda\sigma_i)$ upper half-plane connecting $(\mu_{0,i}, \lambda\sigma_{0,i})$ to $(\mu_{1,i}, \lambda\sigma_{1,i})$,

$$(\mu_i - \mu_{0,i} - \mu_{c,i})^2 + (\lambda\sigma_i)^2 = R_i^2, \tag{32}$$

with center and radius

$$\mu_{c,i} \;=\; \frac{\Delta_i^2 + \lambda^2(\sigma_{1,i}^2 - \sigma_{0,i}^2)}{2\Delta_i}, \qquad R_i \;=\; \sqrt{\mu_{c,i}^2 + \lambda^2\sigma_{0,i}^2}\,. \tag{33}$$

Parametrizing $(\mu_i(t) - \mu_{0,i}, \lambda\sigma_i(t)) = (\mu_{c,i} + R_i\cos\theta_i(t),\ R_i\sin\theta_i(t))$ with the constant-arc-length parametrization

$$\theta_i(t) = 2\arctan\big(e^{L_i t}\tan(\theta_{0,i}/2)\big), \qquad L_i = \log\frac{\tan(\theta_{1,i}/2)}{\tan(\theta_{0,i}/2)}, \tag{34}$$

where the boundary angles $\theta_{0,i} = \mathrm{atan2}(\lambda\sigma_{0,i}, -\mu_{c,i})$ and $\theta_{1,i} = \mathrm{atan2}(\lambda\sigma_{1,i}, \Delta_i - \mu_{c,i})$ both lie in $(0, \pi)$, gives the explicit geodesic

$$\boxed{\begin{aligned} \mu_i(t) &= \mu_{0,i} + \mu_{c,i} + R_i\cos\theta_i(t), \\ \sigma_i(t) &= \frac{R_i}{\lambda}\sin\theta_i(t). \end{aligned}} \tag{35}$$

Case $\Delta_i = 0$. The geodesic is the vertical $\mu_i(t) \equiv \mu_{0,i}$ together with the geometric scale interpolation $\sigma_i(t) = \sigma_{0,i}^{1-t}\sigma_{1,i}^t$, equivalently $\log\sigma_i$ linear in t.

Equivalent hyperbolic (sech/tanh) parametrization. The sech/tanh form of Section 4 carries over coordinate-wise. With $\omega_i, \kappa_i, \alpha_i$ determined from the boundary conditions,

$$\boxed{\mu_i(t) = \mu_{0,i} + (\mu_{1,i} - \mu_{0,i})\frac{\tanh(\omega_i t + \alpha_i) - \tanh(\alpha_i)}{\tanh(\omega_i + \alpha_i) - \tanh(\alpha_i)}, \quad \sigma_i(t) = \frac{\omega_i}{\kappa_i}\operatorname{sech}(\omega_i t + \alpha_i),} \tag{36}$$

where $\omega_i, \kappa_i, \alpha_i$ solve $\frac{\omega_i}{\kappa_i}\operatorname{sech}(\alpha_i) = \sigma_{0,i}$, $\frac{\omega_i}{\kappa_i}\operatorname{sech}(\omega_i + \alpha_i) = \sigma_{1,i}$, $\kappa_i = |\Delta_i|\sqrt{c_\mu/c_\sigma} \cdot \kappa_i^2/\big[\omega_i(\tanh(\omega_i + \alpha_i) - \tanh(\alpha_i))\big]$. The two parametrisations are related by $\sin(2\arctan e^{-s_i}) = \operatorname{sech} s_i$, and $\cos(2\arctan e^{-s_i}) = \tanh s_i$, with $s_i = \omega_i t + \alpha_i$ and $R_i = \lambda\,\omega_i/\kappa_i$.

Anisotropic Fisher–Rao distance and energy. The arc-length on the ith factor along the geodesic is

$$|L_i| \;=\; \operatorname{arcosh}\!\left(1 + \frac{c_\mu\,\Delta_i^2 + c_\sigma(\sigma_{1,i} - \sigma_{0,i})^2}{2\,c_\sigma\,\sigma_{0,i}\sigma_{1,i}}\right). \tag{37}$$

Because the metric is a Riemannian product, the joint squared Fisher–Rao distance and the geodesic energy are simply additive,

$$d_{\mathrm{FR}}^2(p_0, p_1) \;=\; c_\sigma \sum_{i=1}^{d} L_i^2, \qquad e(v) \;=\; c_\sigma \sum_{i=1}^{d} L_i^2. \tag{38}$$

For the symmetric anisotropic boundary $\boldsymbol{\sigma}_0 = \mathbf{1}$, $\boldsymbol{\sigma}_1 = \sigma_{\min}\mathbf{1}$, (37) reduces to $L_i = \operatorname{arcosh}\big(1 + [c_\mu x_{1,i}^2 + c_\sigma(\sigma_{\min} - 1)^2]/(2c_\sigma\sigma_{\min})\big)$.

Anisotropic vector field. By Theorem 3.1, applied component-wise, the vector field associated with the anisotropic geodesic path is

$$\boxed{\mathbf{v}_t(\mathbf{x}) \;=\; \dot{\boldsymbol{\mu}}_t \;+\; \frac{\dot{\boldsymbol{\sigma}}_t}{\boldsymbol{\sigma}_t} \odot (\mathbf{x} - \boldsymbol{\mu}_t),} \tag{39}$$

where $\odot$ and the division act component-wise. From (35) and $\dot\theta_i(t) = L_i \sin\theta_i(t)$ one gets, for every i with $\Delta_i \neq 0$,

$$\dot\mu_{t,i} = -R_i L_i \sin^2\theta_i(t), \qquad \frac{\dot\sigma_{t,i}}{\sigma_{t,i}} = L_i \cos\theta_i(t), \tag{40}$$

so that

$$v_{t,i}(x_i) \;=\; L_i\big[-R_i \sin^2\theta_i(t) + \cos\theta_i(t)\,(x_i - \mu_{t,i})\big]. \tag{41}$$

For coordinates with $\Delta_i = 0$, $\dot\mu_{t,i} = 0$ and $\dot\sigma_{t,i}/\sigma_{t,i} = \log\sigma_{1,i} - \log\sigma_{0,i}$ is constant in t, so $v_{t,i}(x_i) = (\log\sigma_{1,i} - \log\sigma_{0,i})(x_i - \mu_{0,i})$.

Anisotropic conditional vector field and flow. Letting $\boldsymbol{\mu}_0 = \mathbf{0}$, $\boldsymbol{\mu}_1 = \mathbf{x}_1$, $\boldsymbol{\sigma}_0 = \mathbf{1}$, $\boldsymbol{\sigma}_1 = \sigma_{\min}\mathbf{1}$, formula (41) specialises (per coordinate, for $x_{1,i} \neq 0$) to

$$u_{t,i}(x_i \mid x_{1,i}) \;=\; L_i\big[-\big(R_i \sin^2\theta_i(t) + \mu_{c,i}\cos\theta_i(t) + R_i\cos^2\theta_i(t)\big)\operatorname{sgn}(x_{1,i})\,1_{x_{1,i}\neq 0} + \cos\theta_i(t)\,x_i\big], \tag{42}$$

which, after collecting $R_i\sin^2\theta_i + R_i\cos^2\theta_i = R_i$, gives

$$u_{t,i}(x_i \mid x_{1,i}) \;=\; L_i\big[\cos\theta_i(t)\,x_i \;-\; (R_i + \mu_{c,i}\cos\theta_i(t))\operatorname{sgn}(x_{1,i})\big]. \tag{43}$$

Equivalently, in stacked vector form using component-wise scalars $L_i, R_i, \mu_{c,i}, \theta_i(t)$,

$$\mathbf{u}_t(\mathbf{x} \mid \mathbf{x}_1) \;=\; \mathbf{L} \odot \big[\cos\boldsymbol{\theta}(t) \odot \mathbf{x} \;-\; \big(\mathbf{R} + \boldsymbol{\mu}_c \odot \cos\boldsymbol{\theta}(t)\big) \odot \operatorname{sgn}(\mathbf{x}_1)\big]. \tag{44}$$

The associated conditional flow has the explicit form

$$\psi_t(\mathbf{x} \mid \mathbf{x}_1) \;=\; \big(\boldsymbol{\mu}_c + \mathbf{R} \odot \cos\boldsymbol{\theta}(t)\big) \odot \operatorname{sgn}(\mathbf{x}_1) \;+\; \frac{1}{\lambda}\,\mathbf{R} \odot \sin\boldsymbol{\theta}(t) \odot \mathbf{x}, \tag{45}$$

which reduces to the isotropic flow of Section 4 when $\boldsymbol{\sigma}_t = \sigma_t\mathbf{1}$ for some scalar σ_t and the angles $\theta_i(t)$ collapse to a common $\theta(t)$.

Remark (training). With (44) the FM loss (14) is unchanged in form, $\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t,q(\mathbf{x}_1),p_0(\mathbf{x}_0)}\|v_t(\psi_t(\mathbf{x}_0);\theta) - \frac{d}{dt}\psi_t(\mathbf{x}_0)\|_2^2$, but the per-coordinate scalars $\{L_i, R_i, \mu_{c,i}, \theta_i(t)\}_{i=1}^d$ must be precomputed once per sampled $\mathbf{x}_1$ (closed form, no ODE solve required). Hence the anisotropic PG path is no more expensive at training time than its isotropic counterpart.