TOWARDS VARIATIONAL FLOW MATCHING ON GENERAL GEOMETRIES

Olga Zaghen[1], **Floor Eijkelboom**[1], **Alison Pouplin**[2], **Erik J. Bekkers**[1]
[1] University of Amsterdam, [2] Aalto University
o.zaghen@uva.nl

ABSTRACT

We introduce Riemannian Gaussian Variational Flow Matching (RG-VFM), an extension of Variational Flow Matching (VFM) that leverages Riemannian Gaussian distributions for generative modeling on structured manifolds. We derive a variational objective for probability flows on manifolds with closed-form geodesics, making RG-VFM comparable – though fundamentally different to Riemannian Flow Matching (RFM) in this geometric setting. Experiments on a checkerboard dataset wrapped on the sphere demonstrate that RG-VFM captures geometric structure more effectively than Euclidean VFM and baseline methods, establishing it as a robust framework for manifold-aware generative modeling.

1 INTRODUCTION

Generative modeling has become a fundamental task in machine learning, with different frameworks achieving remarkable success across various data modalities (Ramesh et al., 2022; Rombach et al., 2022). While diffusion models have shown exceptional effectiveness (Ho et al., 2020; Song et al., 2020), they rely on constrained probability paths requiring specialized techniques. Continuous normalizing flows (CNFs) (Chen et al., 2018) offer greater flexibility (Song et al., 2021), but remain computationally expensive due to solving high-dimensional ODEs during training and sampling (Ben-Hamu et al., 2022; Rozen et al., 2021; Grathwohl et al., 2019). Flow Matching (FM) (Lipman et al., 2023) addresses these challenges by expressing the transport field through conditional fields, enabling simulation-free learning of ODEs through assumed dynamics towards specific endpoints.

Recent developments have extended flow matching in two key directions: Variational Flow Matching (VFM) (Eijkelboom et al., 2024) reframes the problem as posterior inference over trajectories, providing a probabilistic perspective with flexible modeling choices, while other work has demonstrated flow matching's potential on geometric structures (Chen & Lipman, 2024). This geometric extension is particularly relevant for data on non-Euclidean spaces, where underlying geometry crucially influences probability paths.

In this paper, we develop Riemannian Gaussian VFM, bridging these directions by extending VFM to Riemannian manifolds with closed-form metrics. We derive a variational objective for Gaussian distributions in these spaces, establishing a geometric extension of VFM, comparable yet fundamentally different from RFM. Through experiments on a spherical checkerboard dataset, we demonstrate that our model effectively leverages manifold structure, outperforming geometry-unaware methods.

2 BACKGROUND

Flow Matching. Modern generative modeling interprets sampling from a target distribution p_1 as transporting a base distribution p_0 by learning dynamics. Typically, p_0 is a standard Gaussian, and the transformation follows a time-dependent mapping $\varphi_t \colon [0,1] \times \mathbb{R}^D \to \mathbb{R}^D$ where φ_0 is the identity and φ_1 pushes p_0 onto p_1, e.g. normalizing flows (Chen et al., 2018) use an ODE governed by some time-dependent velocity field u_t. Though likelihood training is possible through the change of variables formula, solving an ODE during training is expensive. Flow Matching (FM) (Lipman et al., 2023) bypasses this by directly learning the velocity field:

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}_{t,x}\big[||u_t(x) - v_t^\theta(x)||^2\big]. \tag{1}$$

This is made computationally feasible by reformulating u_t with a conditional velocity field (i.e. assumed dynamics towards a given x_1), giving rise to Conditional Flow Matching (CFM):

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t,x_1,x}\left[||u_t(x \mid x_1) - v_t^\theta(x)||^2\right]. \tag{2}$$

Minimizing eq. (2) provides an unbiased estimate of $\nabla_\theta \mathcal{L}_{\text{FM}}$, allowing efficient per-sample training.

Riemannian Flow Matching. Riemannian Flow Matching (RFM) (Chen & Lipman, 2024) extends FM to Riemannian manifolds. Given a smooth Riemannian manifold $\mathcal{M}$ with closed-form geodesics and metric $\mathbf{g}$, RFM learns a vector field v_t:

$$\mathcal{L}_{\text{RFM}}(\theta) = \mathbb{E}_{t,x_1,x}\left[\left\|v_t^\theta(x) - \log_x(x_1)/(1-t)\right\|_{\mathbf{g}}^2\right], \tag{3}$$

with $\log_x(x_1)$ denoting the Riemannian log map, which returns the initial velocity vector of the geodesic connecting x to x_1 (more details on Riemannian manifolds are in appendix A.1). Unlike Euclidean Flow Matching, RFM respects the curvature and geodesics of the underlying space $\mathcal{M}$. Through geodesic or spectral distances, it enables simulation-free training when manifold operations are available, and can utilize approximate distances when closed-form geodesics are intractable, maintaining theoretical guarantees while enabling efficient generative modeling.

Variational Flow Matching. Variational Flow Matching (VFM) (Eijkelboom et al., 2024) reformulates FM by introducing a variational distribution $q_t^\theta(x_1 \mid x)$ to approximate the unknown posterior $p_t(x_1 \mid x)$, where the learned velocity v_t^θ is expressed as the expectation of the condition velocity under this variational approximation over trajectories. Then, the VFM objective is to minimize the KL divergence between joint distributions, i.e.:

$$\mathcal{L}_{\text{VFM}}(\theta) = \mathbb{E}_t\left[\text{KL}\left(p_t(x_1, x) \mid\mid q_t^\theta(x_1, x)\right)\right] = -\mathbb{E}_{t,x_1,x}\left[\log q_t^\theta(x_1 \mid x)\right] + \text{const.} \tag{4}$$

When $u_t(x \mid x_1)$ is linear in x_1 – e.g. a straight-line interpolation – the expectation depends only on marginal distributions, implying this objective reduces to a series of D univariate tasks:

$$\mathcal{L}_{\text{VFM}}(\theta) = -\mathbb{E}_{t,x_1,x}\left[\sum_{d=1}^{D} \log q_t^\theta(x_1^d \mid x)\right], \text{ e.g. } \mathcal{L}_{\text{VFM}}(\theta) = \mathbb{E}_{t,x_1,x}\left[\|\mu_t^\theta(x) - x_1\|^2\right], \tag{5}$$

if q_t^θ is Gaussian, relating VFM directly back to FM (see Eijkelboom et al. (2024) for details). A key feature of VFM is its flexibility in choosing q_t^θ, as different choices of q_t^θ allow adaptation to various geometries and data types, improving efficiency and expressiveness.

3 Riemannian Gaussian Variational Flow Matching

The geometric generalization of the VFM framework stems from the observation that learning the posterior probability $p_t(x_1 \mid x)$ implicitly encodes the geometry of the distribution's support. For example, in CatFlow (Eijkelboom et al., 2024), defining $q_t^\theta(x_1 \mid x)$ as a categorical distribution ensures that the velocities point towards the probability simplex. This raises the question of whether other geometric information about the support of p_1 can be similarly encoded in $q_t^\theta(x_1 \mid x)$.

To investigate this, we consider the case where $p_t(x_1 \mid x)$ is defined as a Gaussian distribution, but with its support on the manifold $\mathcal{M} := \text{supp}(p_1)$ instead of the ambient Euclidean space. In this setting, the Riemannian Gaussian distribution naturally arises as a generalization of the Gaussian to a Riemannian manifold. We will refer to CFM and RFM as *vanilla* models, minimizing velocity differences, while VFM and the proposed model are *variational*, matching endpoints in their loss.

3.1 Model Definition

Riemannian Gaussian. Let $\mathcal{M}$ be a Riemannian manifold with metric $\mathbf{g}$. The Riemannian Gaussian (RG) distribution (Pennec, 2006) is defined as

$$\mathcal{N}_{\text{Riem}}(z \mid \sigma, \mu) = \frac{1}{C}\exp\left(-\frac{\text{dist}_{\mathbf{g}}(z, \mu)^2}{2\sigma^2}\right), \tag{6}$$

where $z, \mu \in \mathcal{M}$ (with μ as the mean), $\sigma > 0$ is a scale parameter, and $\text{dist}_{\mathbf{g}}(z, \mu)$ denotes the geodesic distance determined by $\mathbf{g}$. The constant C depends on both z and μ, and it normalizes the distribution over $\mathcal{M}$. A more detailed geometric explanation can be found in Appendix A.2.

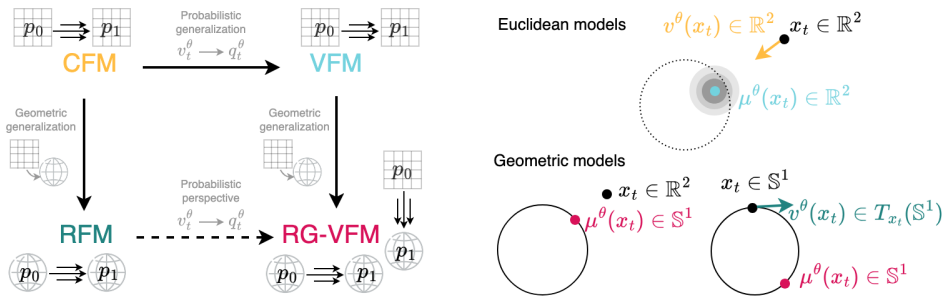


Figure 1: **Left**: Overview of the main Flow Matching models relevant to our proposed approach, illustrating their relationships. **Right**: Visualization of each model's predictions (color-coded to match the corresponding model on the left) for a target distribution p_1 supported on the sphere $\mathbb{S}_1$.

RG-VFM objective. We define the Riemannian Gaussian-VFM (RG-VFM) objective by using the Riemannian Gaussian as our variational approximation, i.e.

$$\mathcal{L}_{\text{RG-VFM}}(\theta) = -\mathbb{E}_{t,x_1,x}\left[\log \mathcal{N}_{\text{Riem}}(x_1 \mid \mu_t^\theta(x), \sigma_t(x))\right]. \tag{7}$$

In the Gaussian VFM case, this setting reduces to a straightforward MSE optimization, so it is natural to wonder whether a similar simplification holds here. In fact, such a simplification exists under two assumptions: (1) the manifold is homogeneous – that is, any point can be transformed into any other by a distance-preserving symmetry (a formal definition is provided in appendix A.1); and (2) we have access to a closed-form expression for its geodesics. Formally, the following holds (see appendix B.1 for details):

Proposition 3.1. *Let $\mathcal{M}$ be a homogeneous manifold with closed-form geodesics. Then, the VFM objective reduces to*

$$\mathcal{L}_{\textit{RG-VFM}}(\theta) = \mathbb{E}_{t,x_1,x}\left[||\log_{x_1}(\mu_t^\theta(x))||_{\mathbf{g}}^2\right]. \tag{8}$$

Minimizing this loss is equivalent to computing the Fréchet mean of the distribution, that is, compute $\mu^\star = \arg\min_{\mu_\theta \in \mathcal{M}} \int_M \text{dist}_{\mathbf{g}}(x_1, \mu_\theta)^2 \, dx_1$, i.e. the point μ_θ that minimizes the expected squared geodesic distance to the data points (Fréchet, 1948). Intuitively, this can be viewed as a generalization of the mean squared error from the Euclidean setting to a Riemannian framework.

The RG-VFM objective (eq. 8) minimizes the geodesic distance on $\mathcal{M}$ between target and predicted endpoints, so it only needs to learn the local geometry of $\mathcal{M}$ around p_1. In contrast, vanilla RFM requires that $\text{supp}(p_0)$ lie entirely on $\mathcal{M}$ because its vector fields depend on the manifold's intrinsic geometry; off-manifold points can lead to unstable estimates of tangents and vector fields.

For this study, we consider two versions: (1) RG-VFM-$\mathbb{R}^3$, where p_0 is Euclidean, $\mathcal{M} \subsetneq \text{supp}(p_0)$, and we use linear interpolation for conditional velocities, and (2) RG-VFM-$\mathcal{M}$, where $\text{supp}(p_0) = \mathcal{M}$, we use geodesic interpolation, and define velocities on the tangent spaces to $\mathcal{M}$.

3.2 How does RG-VFM fit in the existing Flow Matching framework?

Figure 1 (left) illustrates how RG-VFM fits within the framework of related FM models. In VFM, a *probabilistic* generalization of CFM is obtained by making the posterior distribution explicit and customizable, obtaining standard CFM under the choice of a specific Gaussian (see Eijkelboom et al. (2024)). In contrast, RFM serves as a *geometric* generalization of CFM, where the model and its objective take into account the intrinsic properties and metric of the underlying Riemannian manifold. The same happens for the variational models: VFM with a Gaussian posterior is a particular instance of RG-VFM when the geometry is Euclidean. In Euclidean space, $\|\log_{x_1}(\mu_t^\theta(x))\|_{\mathbf{g}}^2$ reduces to $\|\mu_t^\theta(x) - x_1\|_2^2$, thereby recovering the VFM objective.

A further comparison can be made between the simplified version of RFM and RG-VFM-$\mathcal{M}$, where $\mathcal{M}$ is a homogeneous manifold with closed-form geodesics. The variational model (RG-VFM) is *not* a direct generalization of vanilla RFM because, unlike in Euclidean space, *tangent spaces at different points on a manifold do not coincide*. This difference is reflected in the models' outputs (fig. 1):

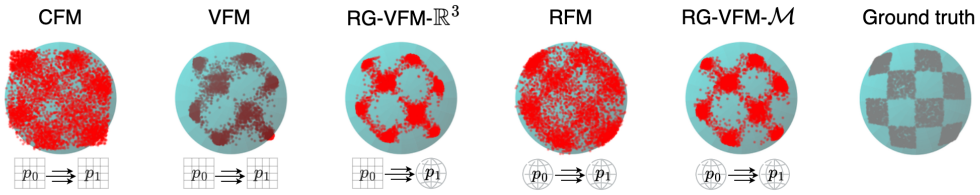


Figure 2: Comparison of the spherical checkerboard distribution generated with CFM, VFM, RFM and our methods RG-VFM-$\mathbb{R}^3$ and RG-VFM-$\mathcal{M}$.

vanilla models predict velocity fields, which are integrated as ODEs to construct flows, whereas variational models predict endpoint distributions, ideally aligning with the target distribution p_1.

In Euclidean space, the difference between two vectors starting at x and pointing to different endpoints is simply the vector between those endpoints, leading to identical L_2 terms in the objectives, i.e. $\|\mu_t^\theta(x) - x_1\|_2^2$ for VFM and $\|u_t(x \mid x_1) - v_t^\theta(x)\|^2$ for CFM. However, since $T_x\mathcal{M} \neq T_{x_1}\mathcal{M}$ in general, in their geometric counterparts this equivalence no longer holds: indeed, the difference vector in the RFM objective, $v_t^\theta(x) - \frac{\log_x(x_1)}{(1-t)}$, is in $T_x\mathcal{M}$, while $\log_{x_1}(\mu_t^\theta(x))$ is in $T_{x_1}\mathcal{M}$. This fundamental distinction separates RG-VFM from RFM. More details are in appendix B.2.

4 Experiments

Inspired by the planar checkerboard benchmark in generative modeling (Grathwohl et al., 2018), we introduce a spherical checkerboard distribution as our target p_1, whose support is $\mathbb{S}^2 \subset \mathbb{R}^3$. The noisy distribution p_0 is defined differently for each model: for CFM, VFM, and RG-VFM-$\mathbb{R}^3$, p_0 is the uniform distribution on $[-1, 1]^3 \subset \mathbb{R}^3$, while for RG-VFM-$\mathcal{M}$ and RFM, it is the uniform distribution on $\mathbb{S}^2$.

We conduct two sets of experiments: we (1) compare Euclidean to geometric models in capturing the correct geometry – assessed by the norms of generated samples (ideally unit norm, since the points should lie on the sphere) – and (2) evaluate vanilla versus variational models in reproducing the target distribution. Figure 2 displays the generated distributions alongside the ground truth. All the experimental details and results are provided in appendix C.

Comparing the Euclidean and geometric models, the norm statistics reveals key differences. The Euclidean models (CFM and VFM) show slight norm deviation from the unit sphere (CFM: mean = 1.00, std = 0.094; VFM: mean = 1.00, std = 0.021), while the geometric models maintain a near-perfect unit norm. The geometric extensions effectively capture the underlying geometry well.

Comparing vanilla and variational models, we observe that vanilla models produce more blurred distributions, whereas variational models better patches contrasts. As shown in fig. 2, RG-VFM-$\mathbb{R}^3$ and RG-VFM-$\mathcal{M}$ exhibit the best visual performance, with minimal differences. This is likely because RG-VFM's loss functions is both geometrically informed and focused on minimizing the distance between predictions and distribution endpoints rather than matching intermediate flow velocities. More interestingly, standard VFM also demonstrates strong performance, which we speculate is due to its emphasis on endpoint minimization. In essence, emphasizing endpoint accuracy enables variational models to capture the fine details of the target distribution's shape, and the additional geometric awareness of RG-VFM further enhances the result.

5 Conclusion

We introduced Riemannian Gaussian Variational Flow Matching (RG-VFM), extending VFM to structured manifolds through Riemannian Gaussian distributions. Our approach unifies RFM and VFM under a common probabilistic framework, with experiments on spherical data demonstrating superior performance over Euclidean-based methods. These results establish RG-VFM as a promising framework for modeling distributions on complex geometries, with applications in molecular modeling and physics-informed generative tasks.

6 ACKNOWLEDGMENTS

This project was funded by the CaLiForNIA project (Marie Skłodowska-Curie Actions Doctoral Network 2022), by the Bosch Center for Artificial Intelligence and by the Finnish Center for Artificial Intelligence (FCAI).

REFERENCES

Heli Ben-Hamu, Samuel Cohen, Joey Bose, Brandon Amos, Maximillian Nickel, Aditya Grover, Ricky T. Q. Chen, and Yaron Lipman. Matching normalizing flows and probability paths on manifolds. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 1749–1763. PMLR, 17–23 Jul 2022.

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

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Floor Eijkelboom, Grigory Bartosh, Christian Andersson Naesseth, Max Welling, and Jan-Willem van de Meent. Variational flow matching for graph generation. *arXiv preprint arXiv:2406.04843*, 2024.

Maurice Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré*, volume 10, pp. 215–310, 1948.

Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

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. *Advances in neural information processing systems*, 33:6840–6851, 2020.

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.

Xavier Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25:127–154, 2006.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022.

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

Noam Rozen, Aditya Grover, Maximilian Nickel, and Yaron Lipman. Moser flow: Divergence-based generative modeling on manifolds. *Advances in Neural Information Processing Systems*, 34:17669–17680, 2021.

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*, 2020.

Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. *Advances in neural information processing systems*, 34:1415–1428, 2021.

A GEOMETRIC BACKGROUND

A.1 RIEMANNIAN MANIFOLDS

In this section, we provide a concise review of Riemannian manifolds in the setting of complete, connected, and smooth manifolds. Let $\mathcal{M}$ be such a manifold endowed with a Riemannian metric $\mathbf{g}$. At each point $x \in \mathcal{M}$, the tangent space is denoted by $T_x\mathcal{M}$, and $\mathbf{g}$ induces an inner product on each $T_x\mathcal{M}$, which we write as $\langle u, v\rangle_{\mathbf{g}}$ for $u, v \in T_x\mathcal{M}$. By collecting all tangent spaces across $\mathcal{M}$, we obtain the tangent bundle:

$$T\mathcal{M} \;=\; \bigcup_{x\in\mathcal{M}} \{x\} \times T_x\mathcal{M}.$$

We will also consider time-dependent vector fields on $\mathcal{M}$. Specifically, let $\{u_t\}_{t\in[0,1]}$ be a family of smooth vector fields, with each u_t mapping a point $x \in \mathcal{M}$ to an element $u_t(x) \in T_x\mathcal{M}$. The operator $\mathrm{div}_{\mathbf{g}}(u_t)$ denotes the Riemannian divergence of u_t with respect to the spatial variable x.

Furthermore, the volume form on $\mathcal{M}$ induced by $\mathbf{g}$ is denoted by $d\mathcal{M}_z$. For any real-valued function $f : \mathcal{M} \to \mathbb{R}$, we write

$$\int_{\mathcal{M}} f(x)\, d\mathcal{M}_z$$

to denote its integral over $\mathcal{M}$.

Homogeneous Manifold. A Riemannian manifold $\mathcal{M}$ is homogeneous if its isometry group acts transitively on $\mathcal{M}$, i.e., for any two points $x, y \in \mathcal{M}$, there exists an isometry $f : \mathcal{M} \to \mathcal{M}$ such that $f(x) = y$.

A.2 RIEMANNIAN GAUSSIAN DISTRIBUTIONS

We describe the construction of the Riemannian Gaussian (RG) distribution, which generalizes the familiar Gaussian distribution to the setting of a Riemannian manifold. The definition is inspired from Pennec (2006).

Riemannian Gaussian. Let $\mathcal{M}$ be a Riemannian manifold endowed with the metric tensor $\mathbf{g}$. The RG distribution is defined by

$$\mathcal{N}_{\text{Riem}}(z \mid \sigma, \mu) = \frac{1}{C} \exp\Big(-\frac{\mathrm{dist}_{\mathbf{g}}(z, \mu)^2}{2\sigma^2}\Big),$$

where $z \in \mathcal{M}$ is a point on the manifold, $\mu \in \mathcal{M}$ plays the role of the mean, and $\sigma > 0$ is a scale parameter controlling the spread of the distribution. Here, $\mathrm{dist}_{\mathbf{g}}(z, \mu)$ denotes the geodesic distance between z and μ as determined by the metric $\mathbf{g}$, and C is a normalization constant chosen so that the total probability integrates to 1 over $\mathcal{M}$:

$$C = \int_{\mathcal{M}} \exp\Big(-\frac{\mathrm{dist}_{\mathbf{g}}(z, \mu)^2}{2\sigma^2}\Big)\, d\mathcal{M}_z.$$

The measure $d\mathcal{M}_z$ represents the Riemannian volume element, which in local coordinates takes the form

$$d\mathcal{M}_z = \sqrt{\det \mathbf{g}(z)}\, dz,$$

with dz being the standard Lebesgue measure in the coordinate chart and $\mathbf{g}(z)$ is the Riemannian metric tensor at the point z. This formulation ensures that the probability density is adapted to the geometric structure of the manifold.

Observation. In the special case where $\mathcal{M} = \mathbb{R}^d$ and the metric is Euclidean (i.e., $\mathbf{g}(z) = \mathbf{I}$), the geodesic distance reduces to the usual Euclidean distance, and the RG distribution becomes the standard multivariate Gaussian with covariance matrix $\sigma^2\mathbf{I}$. On more general manifolds, however, the curvature and topology are taken into account through the geodesic distance and the volume element, leading to a natural extension of the Gaussian concept. This construction can be applied to spaces such as hyperbolic manifolds, where one can define the distribution in the tangent space at a point μ and then use the exponential map to project it onto the manifold.

Comparison to vMF. A closely related distribution is the von Mises–Fisher (vMF) distribution, which is traditionally defined on the sphere S^n by

$$\mathrm{vMF}(z \mid \mu, \kappa) \propto \exp\big(\kappa \langle z, \mu \rangle\big),$$

with $\mu \in S^n$ and $\langle \cdot, \cdot \rangle$ denoting the standard dot product. The vMF distribution is based on the notion of directional data and an inner product structure that measures alignment. In contrast, the RG distribution is inherently tied to the Riemannian metric, making it applicable to a much wider class of manifolds. Generalizing the idea behind the vMF distribution to other geometries often requires embedding the manifold into a larger ambient space and defining a suitable bilinear form (such as the Minkowski inner product in hyperbolic geometry). In this sense, the RG approach offers a more natural and geometrically intrinsic formulation.

In summary, the Riemannian Gaussian distribution is defined in terms of the geodesic distance and the corresponding volume element, and it adapts to the underlying geometry of any Riemannian manifold.

B R-VFM AND LINK WITH RFM

B.1 DETAILED DERIVATION OF RG-VFM OBJECTIVE

Proposition 3.1. *Let $\mathcal{M}$ be a homogeneous manifold with closed-form geodesics. Then, the VFM objective reduces to*

$$\mathcal{L}_{\textit{RG-VFM}}(\theta) = \mathbb{E}_{t,x_1,x}\left[||\log_{x_1}(\mu_t^\theta(x))||_{\mathbf{g}}^2\right]. \tag{8}$$

Proof. The objective of VFM is defined as

$$\mathcal{L}_{\mathrm{VFM}}(\theta) = -\mathbb{E}_{t,x_1,x}\left[\log q_t^\theta(x_1|x)\right].$$

We define the objective function of RG-VFM by setting the posterior probability as the Riemannian Gaussian, i.e.,

$$q_t^\theta(x_1|x) = \mathcal{N}_{\mathrm{Riem}}(x_1 \mid \mu_t^\theta(x), \sigma(x)),$$

so that

$$\mathcal{L}_{\text{RG-VFM}}(\theta) = -\mathbb{E}_{t,x_1,x}\left[\log \mathcal{N}_{\mathrm{Riem}}(x_1 \mid \mu_t^\theta(x), \sigma(x))\right].$$

More explicitly, we have

$$\begin{aligned}
\mathcal{L}_{\text{RG-VFM}}(\theta) &= -\mathbb{E}_{t,x_1,x}\left[\log q_t^\theta(x_1|x)\right] \\
&= -\mathbb{E}_{t,x_1,x}\left[\log \mathcal{N}_{\mathrm{Riem}}(x_1 \mid \mu_t^\theta(x), \sigma(x))\right] \\
&= -\mathbb{E}_{t,x_1,x}\left[\log\left(\frac{1}{C(\mu_t^\theta(x))}\exp\left(-\frac{\mathrm{dist}_{\mathbf{g}}(x_1, \mu_t^\theta(x))^2}{2\sigma(x)^2}\right)\right)\right] \\
&= -\mathbb{E}_{t,x_1,x}\left[\log\left(\frac{1}{C(\mu_t^\theta(x))}\right) - \frac{\mathrm{dist}_{\mathbf{g}}(x_1, \mu_t^\theta(x))^2}{2\sigma(x)^2}\right] \\
&= -\mathbb{E}_{t,x_1,x}\left[\log\left(\frac{1}{C(\mu_t^\theta(x))}\right)\right] + \mathbb{E}_{t,x_1,x}\left[\frac{\mathrm{dist}_{\mathbf{g}}(x_1, \mu_t^\theta(x))^2}{2\sigma(x)^2}\right],
\end{aligned}$$

where $\mathrm{dist}_{\mathbf{g}}()$ denotes the geodesic distance induced by the Riemannian metric $\mathbf{g}$.

Without any regularity assumptions on $\mathcal{M}$, no further simplification is possible. However, under the following assumptions the objective becomes more tractable:

1. **Homogeneity:** If the manifold $(\mathcal{M}, \mathbf{g})$ is homogeneous, the normalization constant

$$C = \int_{\mathcal{M}} \exp\left(-\frac{\mathrm{dist}_{\mathbf{g}}(z, \mu)^2}{2\sigma^2}\right) d\mathcal{M}_z$$

is independent of the mean μ. Hence, defining

$$K := -\mathbb{E}_{t,x_1,x}\left[\log\left(\frac{1}{C(\mu_t^\theta(x))}\right)\right],$$

which is constant with respect to θ, we obtain

$$\mathcal{L}_{\text{RG-VFM}}(\theta) = K + \mathbb{E}_{t,x_1,x}\left[\frac{\text{dist}_{\mathbf{g}}(x_1, \mu_t^\theta(x))^2}{2\sigma(x)^2}\right].$$

Since K is a constant that is independent of the model's parameters θ, the minimization objective becomes

$$\mathcal{L}_{\text{RG-VFM}}(\theta) = \mathbb{E}_{t,x_1,x}\left[\frac{\text{dist}_{\mathbf{g}}(x_1, \mu_t^\theta(x))^2}{2\sigma(x)^2}\right].$$

2. **Closed-form Geodesics:** If the geometry allows closed-form expressions for geodesics, namely

$$\gamma(t) = \exp_x\Big(t \cdot \log_x(y)\Big),$$

then the geodesic distance between two points is given by:

$$\text{dist}_{\mathbf{g}}(z, \mu) = \| \log_z(\mu)\|_{\mathbf{g}}.$$

In this setting, we can write

$$\text{dist}_{\mathbf{g}}(x_1, \mu_t^\theta(x))^2 = \| \log_{x_1}(\mu_t^\theta(x))\|_{\mathbf{g}}^2,$$

so that the objective becomes

$$\mathcal{L}_{\text{RG-VFM}}(\theta) = -\mathbb{E}_{t,x_1,x}\left[\log\left(\frac{1}{C(\mu_t^\theta(x))}\right)\right] + \mathbb{E}_{t,x_1,x}\left[\frac{1}{2\sigma(x)^2}\| \log_{x_1}(\mu_t^\theta(x))\|_{\mathbf{g}}^2\right].$$

3. **Combined Assumptions:** If both conditions hold, the objective simplifies to

$$\mathcal{L}_{\text{RG-VFM}}(\theta) = \mathbb{E}_{t,x_1,x}\left[\frac{1}{2\sigma(x)^2}\| \log_{x_1}(\mu_t^\theta(x))\|_{\mathbf{g}}^2\right].$$

If we further assume that $\sigma(x)$ is constant, this reduces to

$$\mathcal{L}_{\text{RG-VFM}}(\theta) = \mathbb{E}_{t,x_1,x}\left[\| \log_{x_1}(\mu_t^\theta(x))\|_{\mathbf{g}}^2\right].$$

□

Examples of simple geometries. A homogeneous manifold does not necessarily imply that geodesics admit closed-form expressions. Conversely, the simple geometries with closed-form geodesics considered in the RFM setting—such as hyperspheres $\mathbb{S}^n$, hyperbolic spaces $\mathbb{H}^n$, flat tori $T^n = [0, 2\pi]^n$, and the space of SPD matrices $\mathcal{S}_d^+$ with the affine-invariant metric—are homogeneous. Thus, when restricting to these geometries for comparison with RFM, we are in the combined case.

Euclidean space. In the Euclidean case (which also falls into the combined case), the objective simplifies further to

$$\mathcal{L}_{\text{RG-VFM}}(\theta) = \mathbb{E}_{t,x_1,x}\left[\|\mu_t^\theta(x) - x_1\|^2\right].$$

B.2 RG-VFM vs RFM on homogeneous spaces with closed-form geodesics

The objective of RG-VFM is defined as

$$\mathcal{L}_{\text{RG-VFM}}(\theta) = \mathbb{E}_{t,x_1,x}\left[\| \log_{x_1}(\mu_t^\theta(x))\|_{\mathbf{g}}^2\right],$$

while the objective of RFM, in the case of closed-form geodesics, is given by

$$\mathcal{L}_{\text{RFM}}(\theta) = \mathbb{E}_{t,x_1,x}\left[\left\|v_t^\theta(x) - \log_x(x_1)/(1-t)\right\|_{\mathbf{g}}^2\right],$$

with $\mathbf{g}$ being the metric tensor at $x \sim p_t(x|x_1)$.

Ignoring multiplicative constants that depend only on t and x, comparing the two losses reduces to comparing the quantities

$$\| \log_{x_1}(\mu_t^\theta(x))\|_{\mathbf{g}}^2 \quad \text{and} \quad \|v_t^\theta(x) - \log_x(x_1)\|_{\mathbf{g}}^2.$$

Euclidean space. In Euclidean space $\mathbb{R}^n$, the tangent space at each point is naturally identified with $\mathbb{R}^n$. In this setting,

$$\log_{x_1}(\mu_t^\theta(x)) = \mu_t^\theta(x) - x_1 \quad \text{and} \quad \log_x(x_1) = x_1 - x.$$

Notice that

$$\mu_t^\theta(x) - x_1 = \mu_t^\theta(x) - x + x - x_1 = (\mu_t^\theta(x) - x) - (x_1 - x) = (\mu_t^\theta(x) - x) - \log_x(x_1).$$

If we define (ignoring multiplicative constants such as $1/(1-t)$)

$$v_t^\theta(x) = \log_x(\mu_t^\theta(x)) = \mu_t^\theta(x) - x,$$

then it follows that

$$\log_{x_1}(\mu_t^\theta(x)) = \log_x(\mu_t^\theta(x)) - \log_x(x_1),$$

implying

$$\|\log_{x_1}(\mu_t^\theta(x))\|_{\mathbf{g}}^2 = \|v_t^\theta(x) - \log_x(x_1)\|_{\mathbf{g}}^2.$$

Thus, $\mathcal{L}_{\text{RG-VFM}}(\theta)$ and $\mathcal{L}_{\text{RFM}}(\theta)$ are equivalent up to an additive constant. This result is consistent with the known equivalence between $\mathcal{L}_{\text{VFM}}(\theta)$ and $\mathcal{L}_{\text{CFM}}(\theta)$.

General geometries. In non-Euclidean spaces, however, the quantities

$$\|\log_{x_1}(\mu_t^\theta(x))\|_{\mathbf{g}}^2 \quad \text{and} \quad \|v_t^\theta(x) - \log_x(x_1)\|_{\mathbf{g}}^2$$

are not necessarily equal. This is because $\log_{x_1}(\mu_t^\theta(x))$ is a vector in $T_{x_1}\mathcal{M}$, while $\log_x(\mu_t^\theta(x)) - \log_x(x_1)$ lies in $T_x\mathcal{M}$, and in general $T_{x_1}\mathcal{M} \neq T_x\mathcal{M}$. Establishing a relation between these vectors is not straightforward and can be illustrated by comparing the law of cosines in Euclidean, hyperbolic spaces, and on hyperspheres.

C Experiments

In this section, we present further results from the experiments described in Section 4.

C.1 Experimental setup

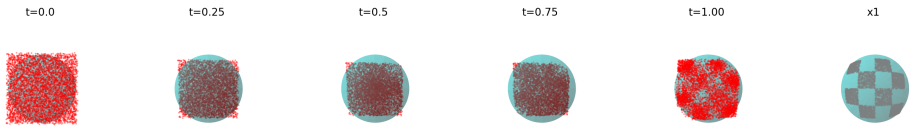
In all experiments, the target distribution p_1 is the spherical checkerboard, so its support is $\mathbb{S}^2$. The noisy distribution p_0 varies by model: for CFM, VFM, and RG-VFM-$\mathbb{R}^3$, p_0 is the uniform distribution on $[-1,1]^3 \subset \mathbb{R}^3$, while for RG-VFM-$\mathcal{M}$ and RFM, p_0 is the uniform distribution on $\mathbb{S}^2$. In every case, we train a five-layer MLP with 64/128 hidden features for 3000 epochs, that we use to generate 5000 samples using a Runge-Kutta ODE solver.

C.2 Results

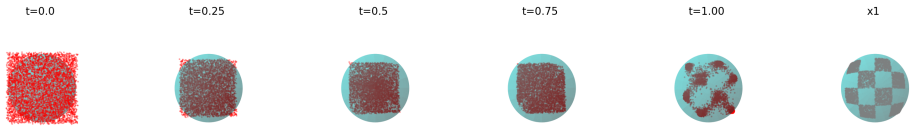
Figure 3 illustrates the generative flow trajectories over time, from the initial distribution p_0 to the generated distribution at $t = 1$.

Figure 4 displays the generated distributions unwrapped onto a flat surface for easier visualization and comparison. These results visually confirm the observations presented in Section 4.

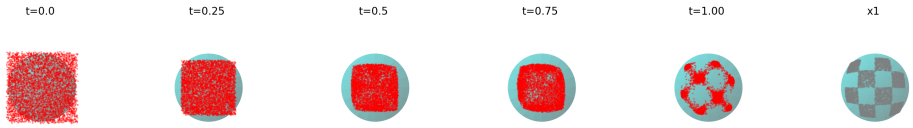
Finally, Figure 5 shows histograms of the norm values of the generated samples. As discussed in Section 4, this metric differentiates the Euclidean models (CFM and VFM) from the others. Ideally, points should have a norm of 1, since they lie on the sphere. However, because the Euclidean models lack explicit geometric information, their points deviate slightly from a norm of 1—with CFM exhibiting a larger divergence. In contrast, the geometric models consistently generate points that lie almost exactly on the sphere.



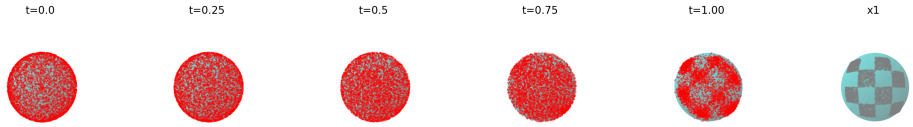
(a) **Model**: CFM; $\mathrm{supp}(p_0) := \mathbb{R}^3$, p_0: uniform distribution on $[-1, 1]^3 \subset \mathbb{R}^3$.



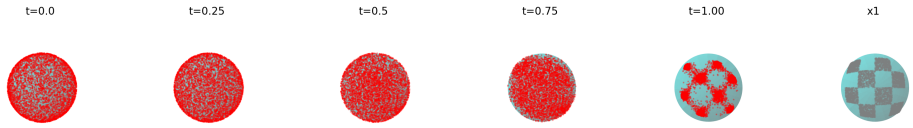
(b) **Model**: VFM; $\mathrm{supp}(p_0) := \mathbb{R}^3$; p_0: uniform distribution on $[-1, 1]^3 \subset \mathbb{R}^3$.



(c) **Model**: RG-VFM; $\mathrm{supp}(p_0) := \mathbb{R}^3$; p_0: uniform distribution on $[-1, 1]^3 \subset \mathbb{R}^3$.

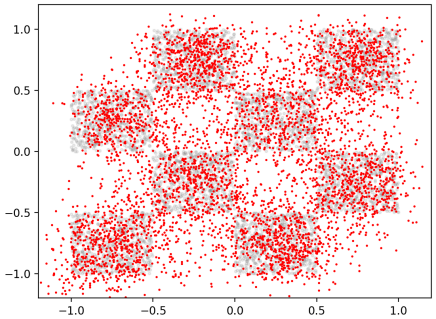


(d) **Model**: RFM; $\mathrm{supp}(p_0) := \mathbb{S}^2$; p_0: uniform distribution on $\mathbb{S}^2$.

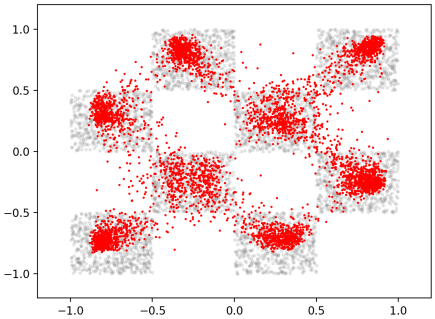


(e) **Model**: RG-VFM; $\mathrm{supp}(p_0) := \mathbb{S}^2$; p_0: uniform distribution on $\mathbb{S}^2$.

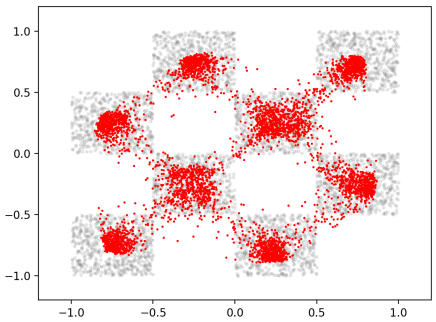
Figure 3: Flow trajectories of 5,000 samples, initially drawn from the noisy distribution p_0 at $t = 0$, evolving to reach their final configuration by $t = 1$.



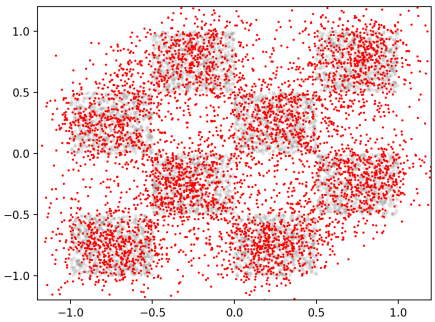
(a) **Model**: CFM; $\mathrm{supp}(p_0) := \mathbb{R}^3$, p_0: uniform distribution on $[-1, 1]^3 \subset \mathbb{R}^3$.



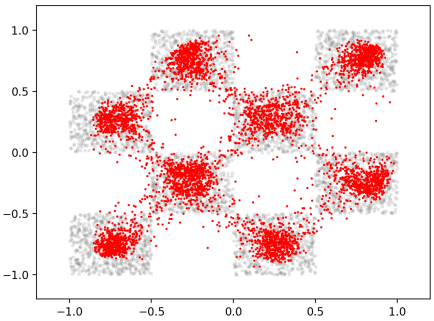
(b) **Model**: VFM; $\mathrm{supp}(p_0) := \mathbb{R}^3$; p_0: uniform distribution on $[-1, 1]^3 \subset \mathbb{R}^3$.



(c) **Model**: RG-VFM; $\mathrm{supp}(p_0) := \mathbb{R}^3$; p_0: uniform distribution on $[-1, 1]^3 \subset \mathbb{R}^3$.

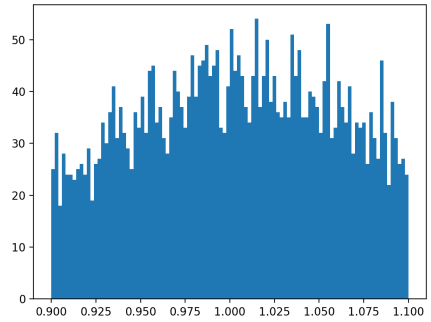


(d) **Model**: RFM; $\mathrm{supp}(p_0) := \mathbb{S}^2$; p_0: uniform distribution on $\mathbb{S}^2$.

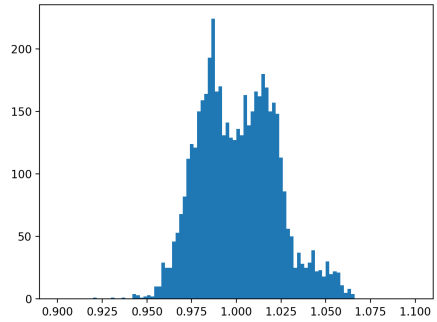


(e) **Model**: RG-VFM; $\mathrm{supp}(p_0) := \mathbb{S}^2$; p_0: uniform distribution on $\mathbb{S}^2$.

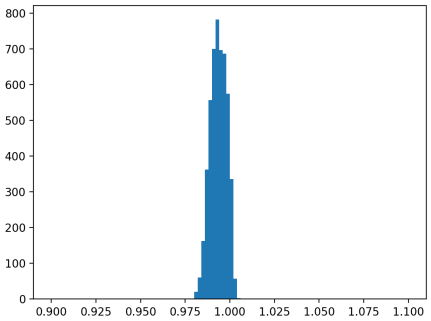
Figure 4: Sample distributions generated by different models (representing the flow configuration at $t = 1$) unwrapped from $\mathbb{S}^2$ to $\mathbb{R}^2$ for improved visualization. The true checkerboard distribution is shown in gray in the background.



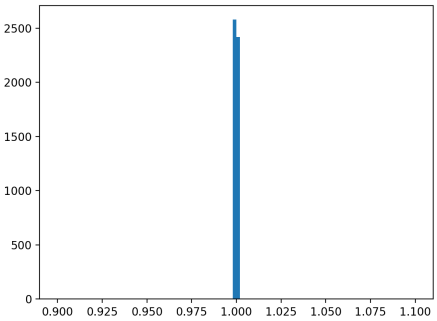
(a) **Model**: CFM; $\mathrm{supp}(p_0) := \mathbb{R}^3$, p_0: uniform distribution on $[-1, 1]^3 \subset \mathbb{R}^3$.



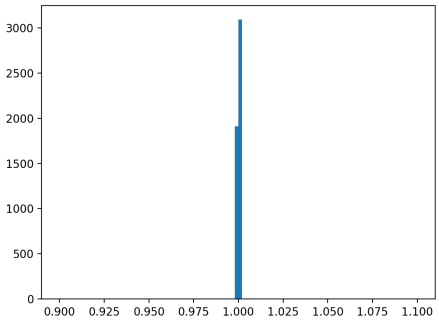
(b) **Model**: VFM; $\mathrm{supp}(p_0) := \mathbb{R}^3$; p_0: uniform distribution on $[-1, 1]^3 \subset \mathbb{R}^3$.



(c) **Model**: RG-VFM; $\mathrm{supp}(p_0) := \mathbb{R}^3$; p_0: uniform distribution on $[-1, 1]^3 \subset \mathbb{R}^3$.



(d) **Model**: RFM; $\mathrm{supp}(p_0) := \mathbb{S}^2$; p_0: uniform distribution on $\mathbb{S}^2$.



(e) **Model**: RG-VFM; $\mathrm{supp}(p_0) := \mathbb{S}^2$; p_0: uniform distribution on $\mathbb{S}^2$.

Figure 5: Histogram of the norm values of the 5000 samples describing the generated distribution.