FACEBOOK AI

Riemannian Continuous Normalizing Flows

Emile Mathieu†, Maximilian Nickel‡

†Department of Statistics, University of Oxford, United Kingdom, ‡Facebook Artificial Intelligence Research, New York, USA

UNIVERSITY OF OXFORD

Overview

- We build on **continuous normalizing flows** [1] and introduce **Riemannian continuous normalizing flows**, a model which admits the parametrization of **flexible probability measures on smooth manifolds** by defining flows as the solution to **ordinary differential equations**.

Motivation

- **Normalizing flows** (NFs) have shown great promise for **modelling flexible probability distributions**.
- **Data** is often naturally described on **Riemannian manifolds** such as spheres, tori, and hyperbolics pace.
- Most NFs implicitly assume a **flat geometry**, making them either **misspecified or ill-suited** in these situations.

Contributions

- We give **sufficient conditions** for a **vector field** to generate a **diffeomorphic flow** mapping a manifold onto itself.
- We derive the **continuous change of variables** for **manifold-valued random variables**.
- We propose a **neural network architecture** that outputs vector fields decomposed on the **coordinates vector fields**.
- We **empirically demonstrate** the advantages of our method on **constant curvature manifolds** – ie., the Poincaré disk and the sphere – compared to **non-Riemannian and projected method**.

Vector flows

- We parametrize flows through the time-evolution of manifold-valued particles $\boldsymbol{z}$, as described by the ordinary differential equation (ODE)

$$\frac{d\boldsymbol{z}(t)}{dt} = \boldsymbol{f}_\theta(\boldsymbol{z}(t), t) \qquad (1)$$

where $\boldsymbol{f}_\theta$ denotes a *vector field*.

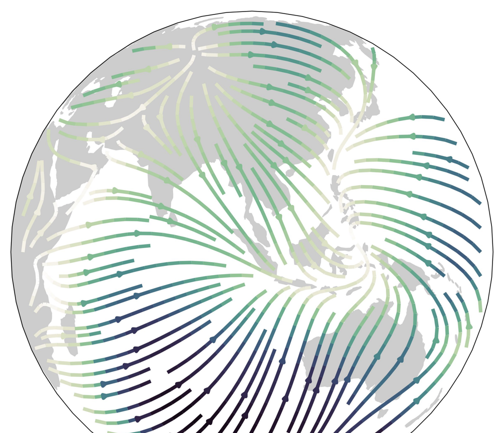


Figure 1: Trajectories generated on the sphere to model volcano eruptions.

Proposition 1: Vector flows

Let $\mathcal{M}$ be a *smooth* complete manifold. Furthermore, let $\boldsymbol{f}_\theta$ be a C^1-bounded time-dependent *vector field*. Then there exists a *global flow* $\phi : \mathcal{M} \times \mathbb{R} \mapsto \mathcal{M}$ such that for each $t \in \mathbb{R}$, the map $\phi(\cdot, t) : \mathcal{M} \mapsto \mathcal{M}$ is a C^1-diffeomorphism (i.e. C^1 bijection with C^1 inverse).

Likelihood computation

Proposition 2: Instantaneous change of variables

Let $\boldsymbol{z}(t)$ be a continuous manifold-valued random variable given in local coordinates, which is described by the ODE from Eq. (1) with probability density $p_\theta(\boldsymbol{z}(t))$. The change in log-probability then also follows a differential equation given by

$$\frac{\partial \log p_\theta(\boldsymbol{z}(t))}{\partial t} = -\operatorname{div}(\boldsymbol{f}_\theta(\boldsymbol{z}(t), t)) = -|G(\boldsymbol{z}(t))|^{-\frac{1}{2}} \operatorname{tr}\left(\frac{\partial \sqrt{|G(\boldsymbol{z}(t))|}\boldsymbol{f}_\theta(\boldsymbol{z}(t), t)}{\partial \boldsymbol{z}}\right). \qquad (2)$$

- This can efficiently be estimated with Hutchinson's trace estimator, as

$$\operatorname{div}(\boldsymbol{f}_\theta(\boldsymbol{z}(t), t)) = |G(\boldsymbol{z}(t))|^{-\frac{1}{2}} \, \mathbb{E}_{p(\epsilon)}\left[\epsilon^\top \frac{\partial \sqrt{|G(\boldsymbol{z}(t))|}\boldsymbol{f}_\theta(\boldsymbol{z}(t), t)}{\partial \boldsymbol{z}} \epsilon\right]. \qquad (3)$$

Hyperbolic geometry and limits of standard and wrapped methods

- Standard NFs are unaware of the underlying geometry or border of the manifold.
- The *wrapped* model defined as $\exp_{\mu_\sharp} P$ with P a standard NF on $\mathbb{R}^2$, is competitive but struggles when the target is located far from the origin.

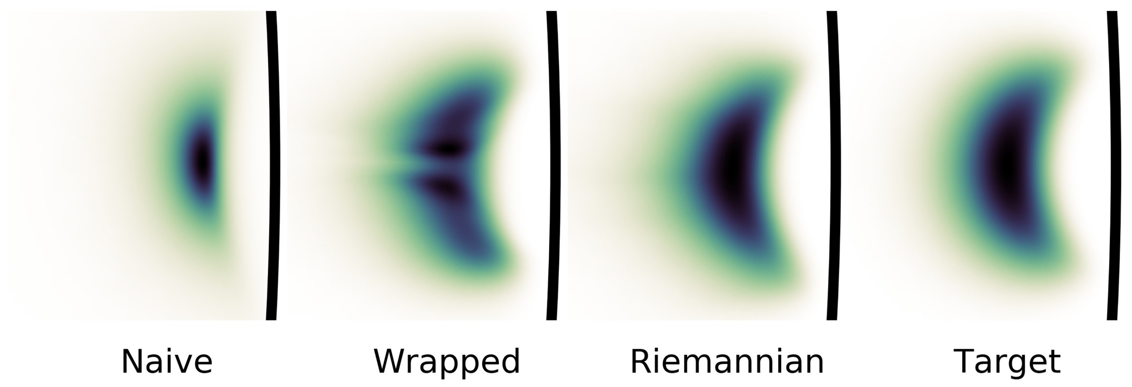


Figure 2: Probability densities on $\mathbb{B}^2$ (zoomed). Models have been trained by maximum likelihood to fit $\mathcal{N}^{\mathrm{W}}(\exp_{\mathbf{0}}(2\,\partial x), \Sigma)$.

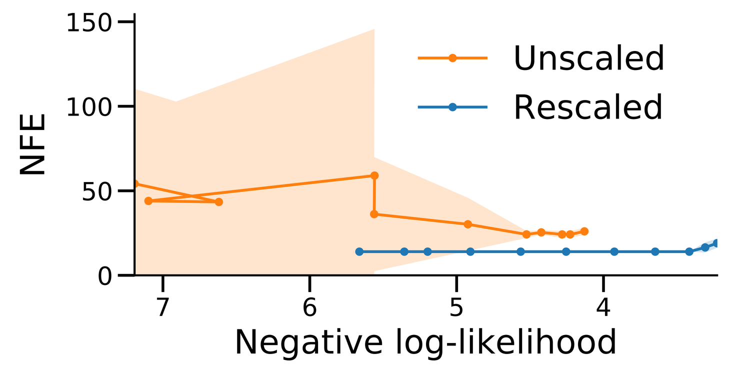


Figure 3: Ablation study of the vector field architecture for the *Riemannian* model.

Spherical geometry and limits of the stereographic projection

- *Stereographic model*: $P^{\mathrm{S}}_\theta = \rho_\sharp^{-1} P$ with ρ the stereographic projection [2].
- We investigate the failure point of the *stereographic* model by targeting a simple and unimodal distribution: a von Mises-Fisher distribution.

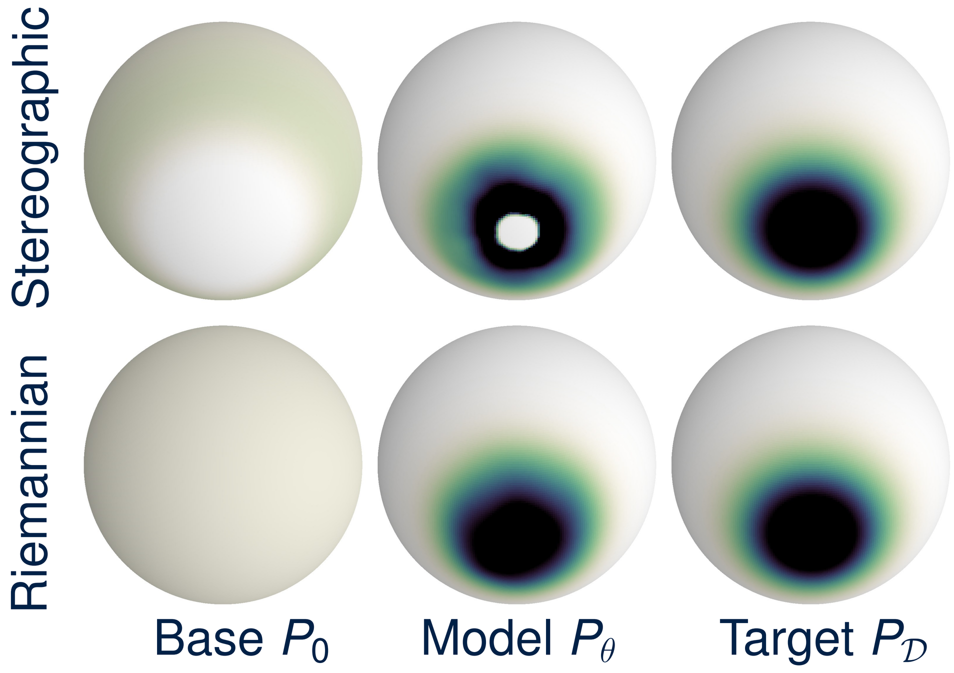


Figure 4: Probability distributions on $\mathbb{S}^2$. Models trained to fit a vMF($\boldsymbol{\mu} = -\boldsymbol{\mu}_0, \kappa = 10$).

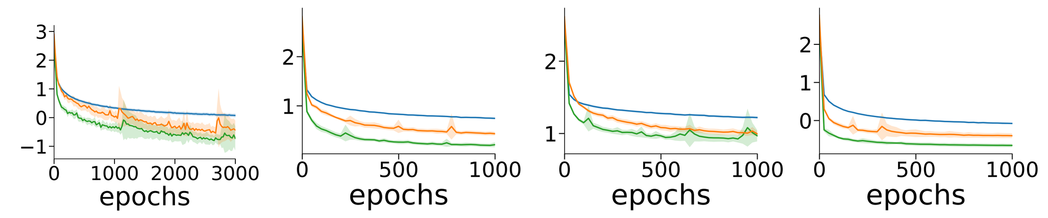
Loss	model / κ	**Stereographic**	**Riemannian**
$\mathcal{L}^{\mathrm{Like}}$	100	$63.60_{\pm 3.56}$	$\mathbf{-1.78}_{\pm 0.01}$
	50	$32.68_{\pm 3.15}$	$\mathbf{-1.09}_{\pm 0.01}$
	10	$6.45_{\pm 2.42}$	$\mathbf{0.52}_{\pm 0.01}$
$\mathcal{L}^{\mathrm{KL}}$	100	$1.56_{\pm 0.34}$	$\mathbf{0.04}_{\pm 0.02}$
	50	$0.68_{\pm 0.16}$	$\mathbf{0.03}_{\pm 0.02}$
	10	$0.12_{\pm 0.01}$	$\mathbf{0.01}_{\pm 0.00}$

Table 1: Performance of continuous flows on $\mathbb{S}^2$ with vMF($\boldsymbol{\mu} = -\boldsymbol{\mu}_0, \kappa$) targets (the smaller the better).

Density estimation of spherical data

- We assess the modelling capacity of our model against its *stereographic* counterpart and a mixture of vMF distributions.

Table 2: Negative test log-likelihood of spherical continuous normalizing flows on real spherical earth science data.

	Volcano	Earthquake	Flood	Fire
Mixture vMF	$-0.31_{\pm 0.07}$	$0.59_{\pm 0.01}$	$1.09_{\pm 0.01}$	$-0.23_{\pm 0.02}$
Stereographic	$-0.64_{\pm 0.20}$	$0.43_{\pm 0.04}$	$0.99_{\pm 0.04}$	$-0.40_{\pm 0.06}$
Riemannian	$-0.97_{\pm 0.15}$	$\mathbf{0.19}_{\pm 0.04}$	$\mathbf{0.90}_{\pm 0.03}$	$\mathbf{-0.66}_{\pm 0.05}$
Learning curves				
Data size	829	6124	4877	12810

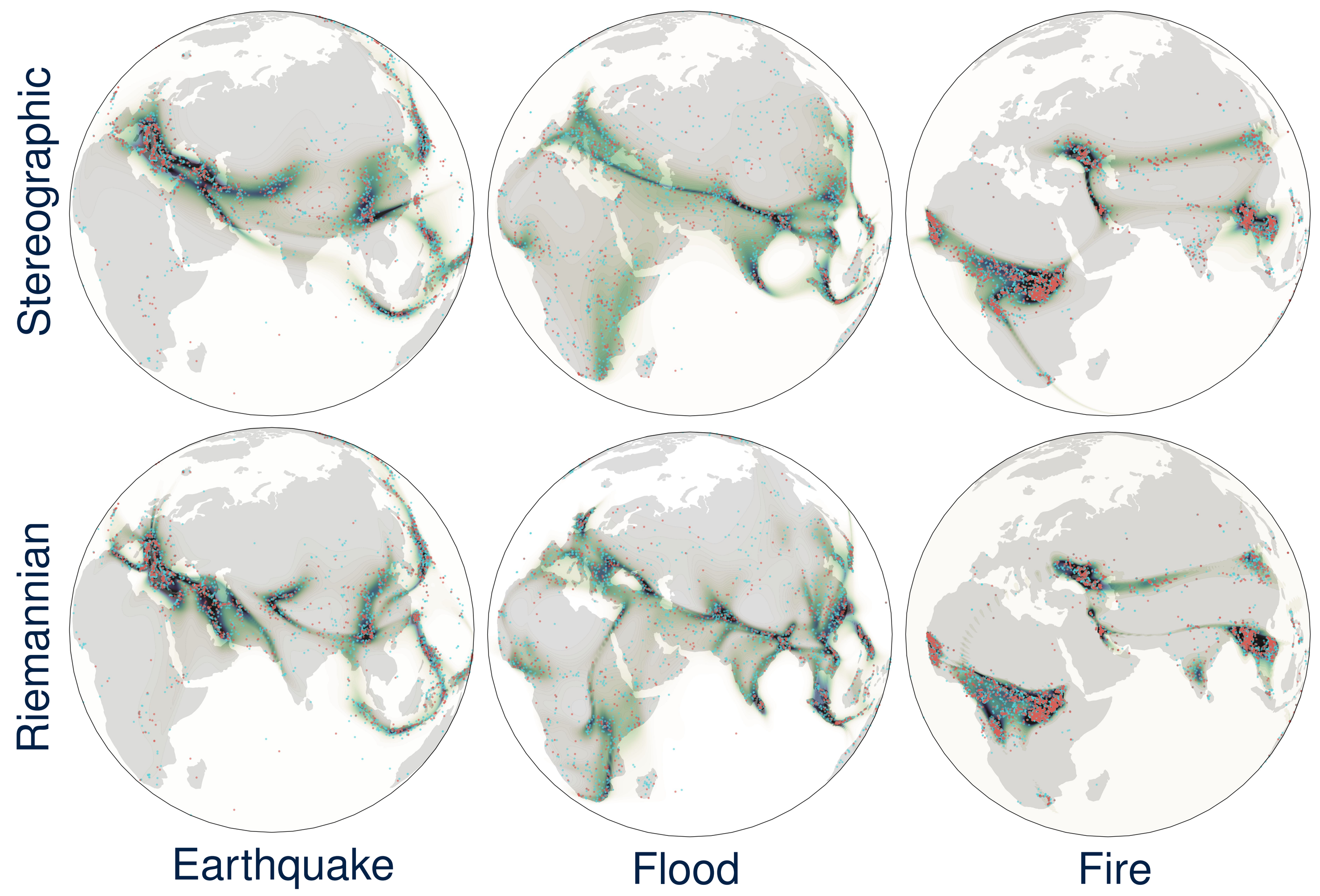


Figure 5: Blue and red dots represent training and testing datapoints, respectively. Heatmaps depict the log-likelihood of the trained models.

Future work and limitations

- Currently rely on a choice of local coordinates, e.g. (θ, ϕ) for $\mathbb{S}^2$, which are not defined on the full manifold.
- As a consequence, the stochastic estimator from Eq. (3) exhibits high variance in the spherical setting.
- Can bypass coordinates vector fields thanks to the existence of a family of divergence-free vector fields spanning the tangent bundle for homogeneous spaces.

References

[1] 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.

[2] Mevlana C. Gemici, Danilo Rezende, and Shakir Mohamed. Normalizing Flows on Riemannian Manifolds. *arXiv:1611.02304 [cs, math, stat]*, November 2016.