Normalizing Flows on Tori and Spheres

ICML 2020

NYU

Collaborators

DeepMind

Danilo Rezende

George Papamakarios

Sébastien Racanière

Center for Theoretical Physics, MIT

Gurtej Kanwar

Phiala Shanahan

Center for Cosmology and Particle Physics, NYU

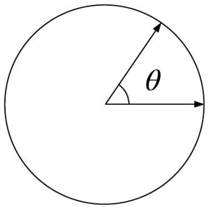
Michael Albergo

Kyle Cranmer

Overview

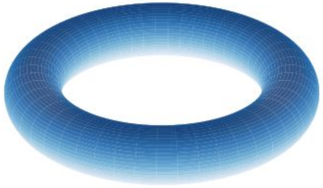
- Probability distributions on:

$\mathbb{S}^1$



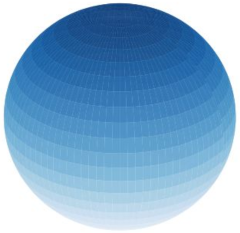
Circles

$\mathbb{T}^D$



Tori

$\mathbb{S}^D$

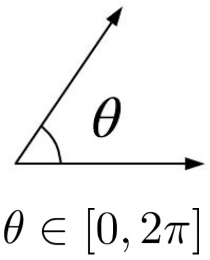


Spheres

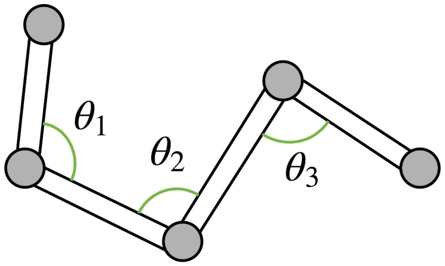
- As flexible as we like
- Any dimension D we like
- With efficient and exact density evaluation and sampling

Why circles, tori and spheres?

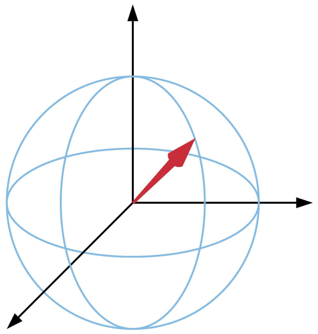
- Not all data are Euclidean!



Angles



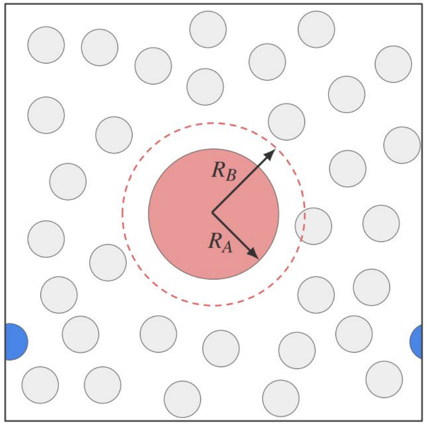
Joint configurations of molecules / robot arms



Directions

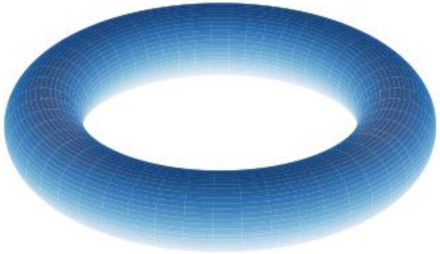
Physics application: Estimating free energy

System of N particles with
periodic boundary conditions



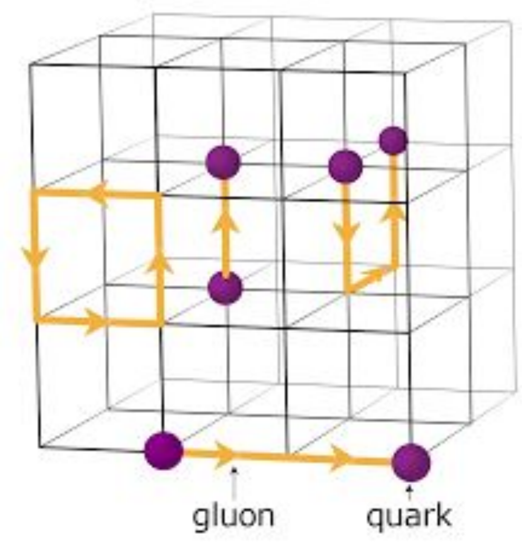
$\cong$

$\mathbb{T}^{3N}$



Wirnsberger & Ballard et al., *Targeted free energy estimation via learned mappings*, arxiv.org/abs/2002.04913, 2020

Physics application: Simulating quantum fields on a lattice



$$\mathrm{U}(1) \cong \mathbb{S}^1$$

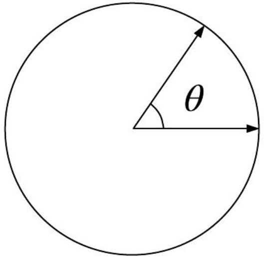
$$\mathrm{SU}(2) \cong \mathbb{S}^3$$

$$\text{Diagonal SU}(3) \cong \mathbb{T}^2$$

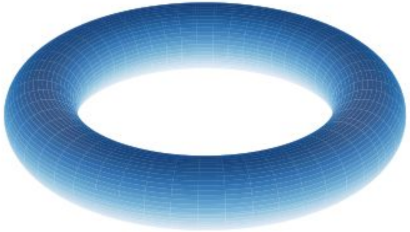
Kanwar et al., *Equivariant flow-based sampling for lattice gauge theory*, arxiv.org/abs/2003.06413, 2020

Directional statistics

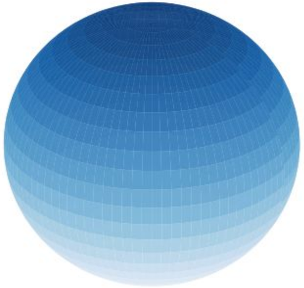
$\mathbb{S}^1$

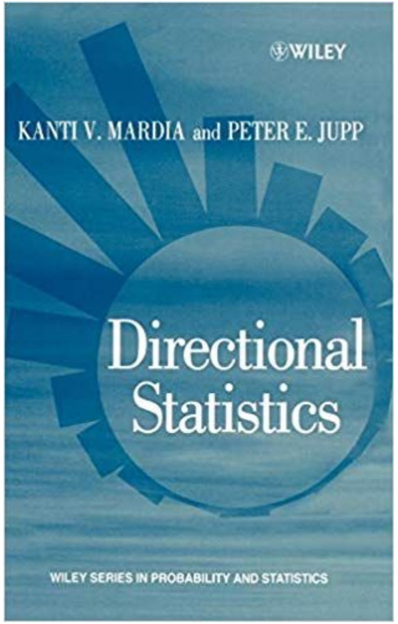


$\mathbb{T}^D$



$\mathbb{S}^D$

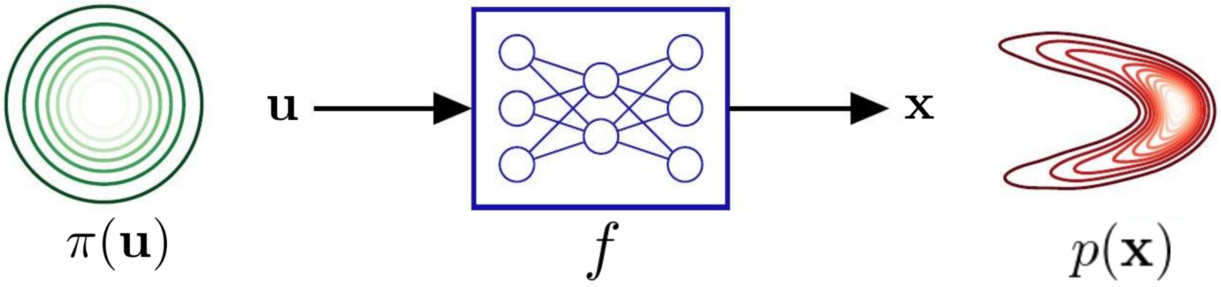




Common techniques:

- Wrapping (e.g. wrapped Gaussian)
- Projecting (e.g. projected Gaussian)
- Conditioning (e.g. von Mises-Fisher)

Normalizing flows



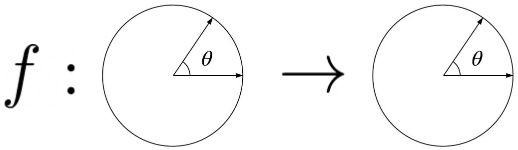
Sampling

$$\mathbf{u} \sim \pi(\mathbf{u})$$
$$\mathbf{x} = f(\mathbf{u})$$

Density evaluation

$$p(\mathbf{x}) = \pi(\mathbf{u}) \left| \det \frac{\partial f}{\partial \mathbf{u}} \right|^{-1}$$

Flows on the circle



Parameterize using angle:

$$f : [0, 2\pi] \to [0, 2\pi]$$

Fix endpoints:

$$f(0) = 0$$
$$f(2\pi) = 2\pi$$

Positive derivative:

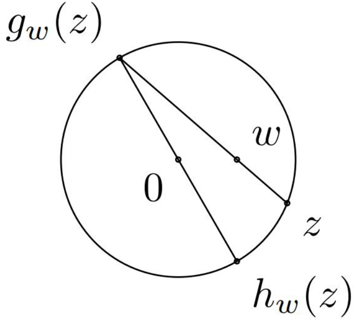
$$\nabla f(\theta) > 0$$

Match endpoint derivatives:

$$\nabla f(0) = \nabla f(2\pi)$$

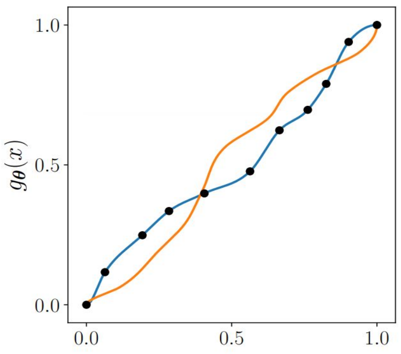
Flows on the circle: Three ways

Möbius transforms

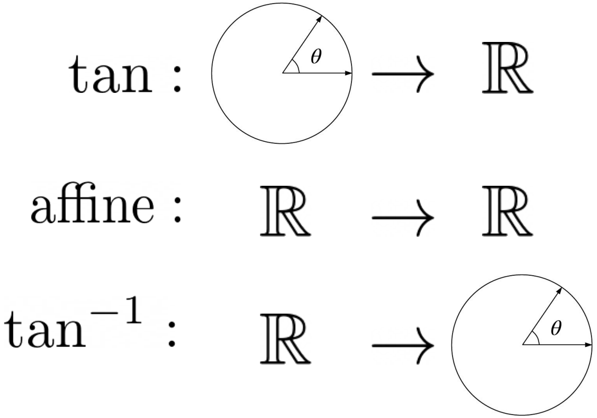


+ rotation to fix endpoints

Circular splines (CS)



Non-compact projections (NCP)



Expressive flows on circle: Mixtures

Composing Möbius (or NCP) transformations does not increase expressivity since they form a group. Instead, we propose an efficient method to create mixtures of them.

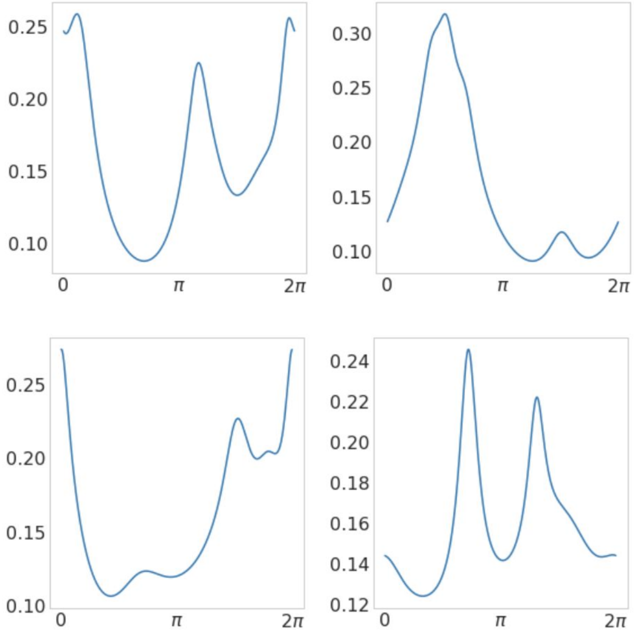
Assume

- N flows f_k on S^1
-

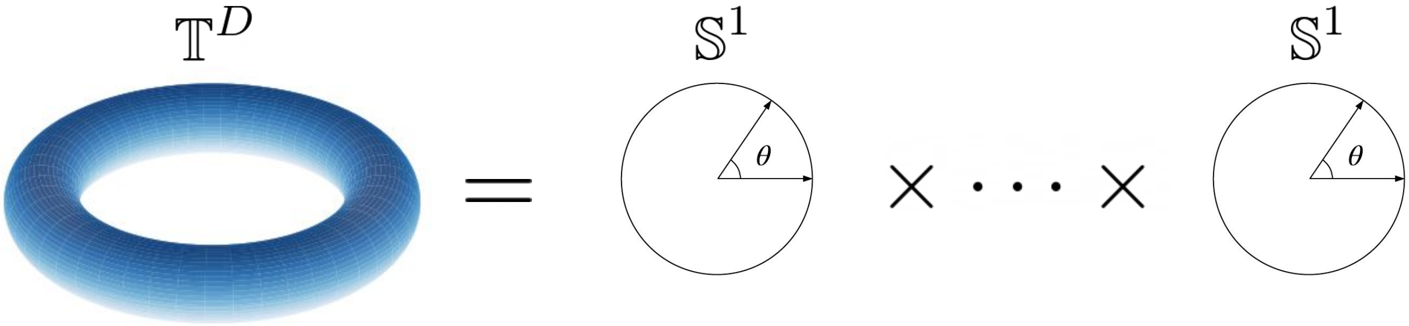
Define $f_k(0) = 0, f_k(2\pi) = 2\pi, \nabla f_k(0) = \nabla f_k(2\pi)$

$$f(\theta) = \frac{1}{N} \sum_k f_k(\theta)$$

Still a valid diffeomorphism of S^1 with tractable Jacobian!



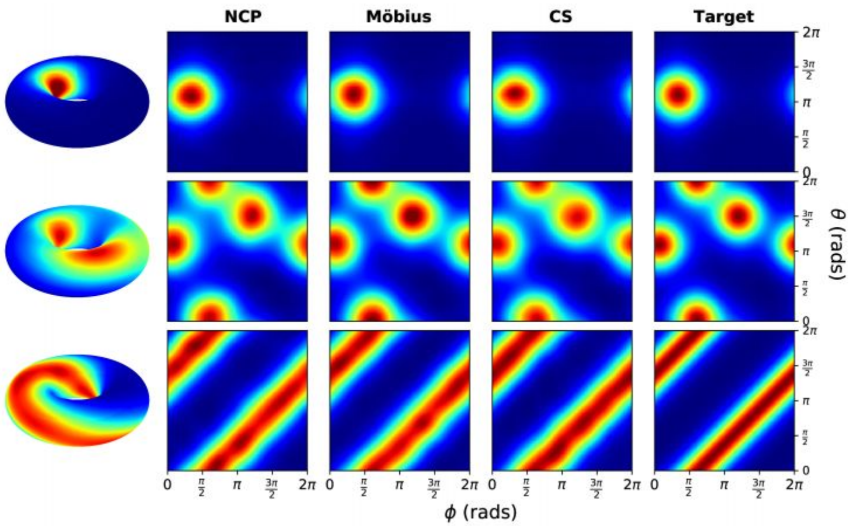
Flows on tori

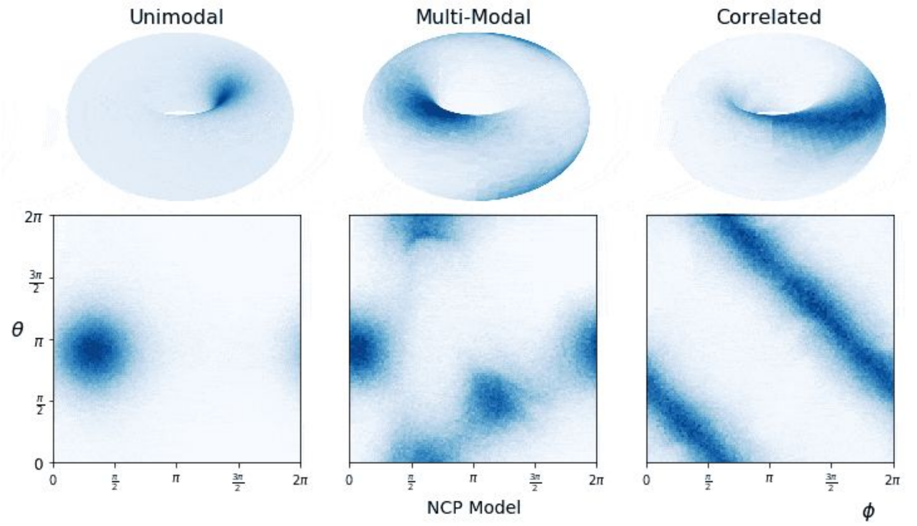


$$p(\theta_1, \ldots, \theta_D) = \prod_i p(\theta_i \mid \theta_1, \ldots, \theta_{i-1})$$

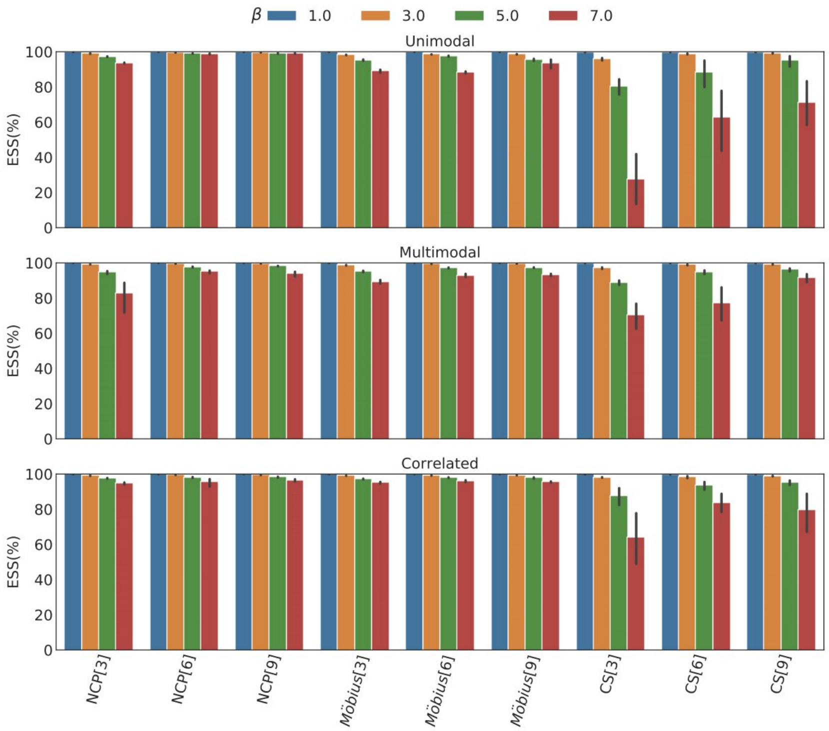
Autoregressive flow whose conditionals are circle flows (Möbius, CS or NCP)

Flows on tori: Results



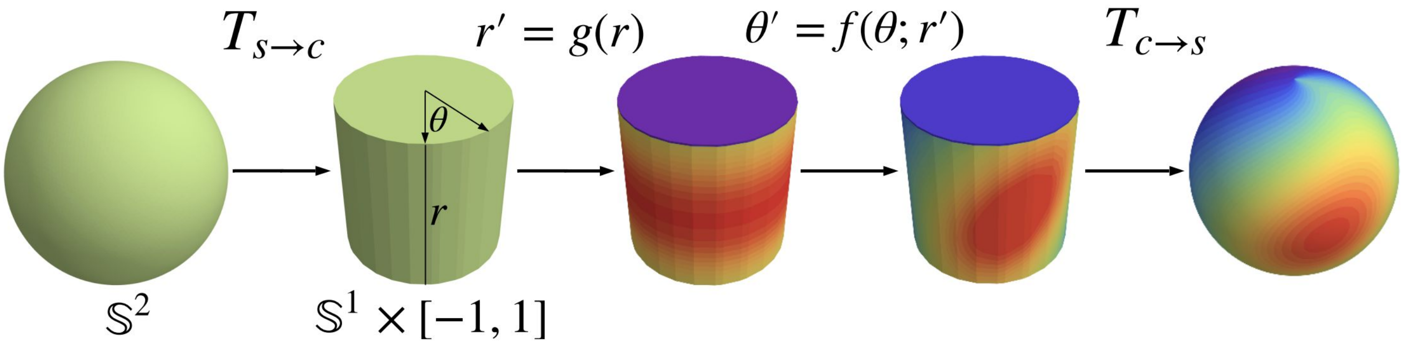


Flows on the circle: Results



Comparison of Möbius, CS & NCP

Flows on 2-spheres: Cylindrical coordinates



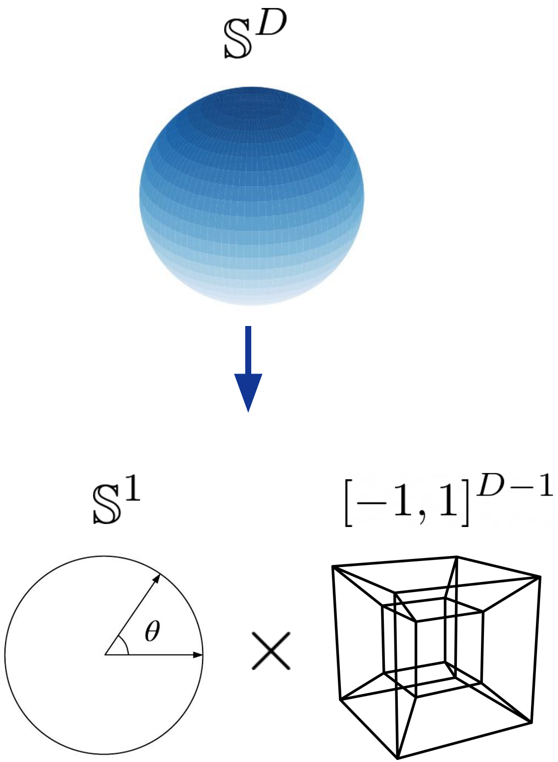
Sphere to cylinder

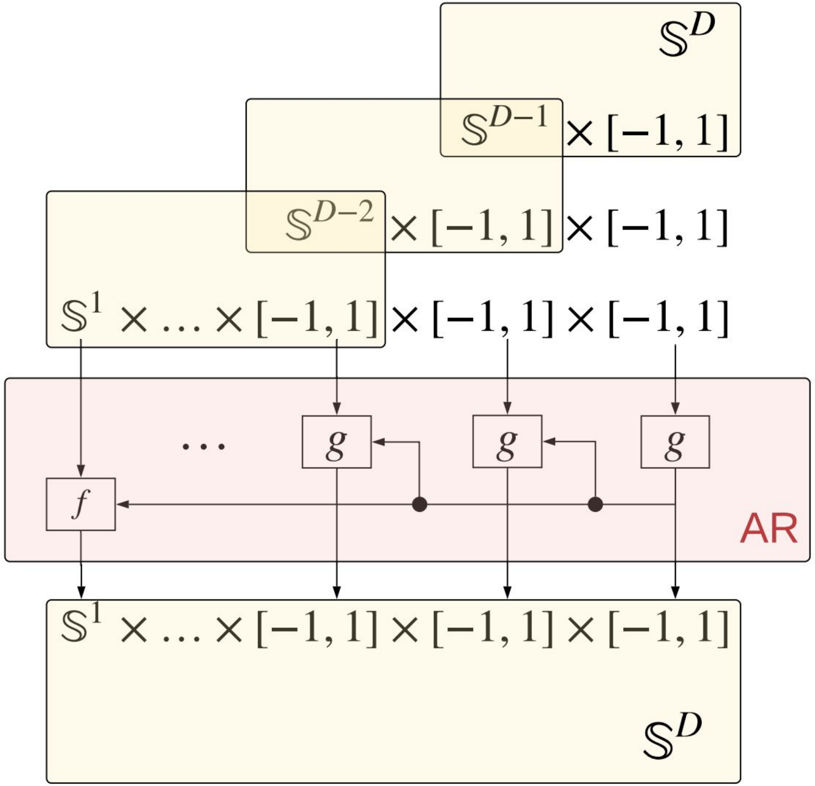
$$T_{s\to c}(x) = \left(\frac{x_{1:D}}{\sqrt{1 - x_{D+1}^2}}, x_{D+1} \right)$$

Cylinder to sphere

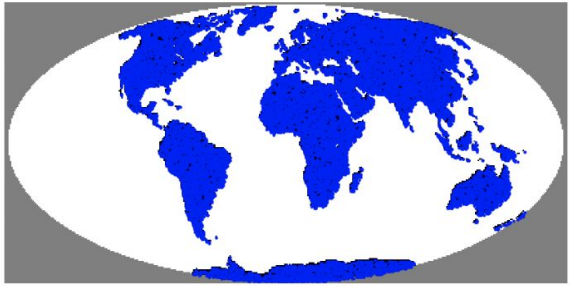
$$T_{c\to s}(z, r) = \left(z\sqrt{1 - r^2}, r \right)$$

Flows on spheres: Recursive D-dimensional model

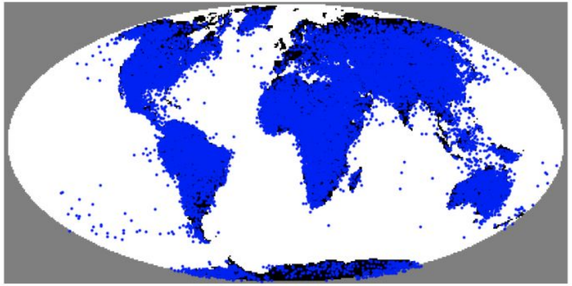




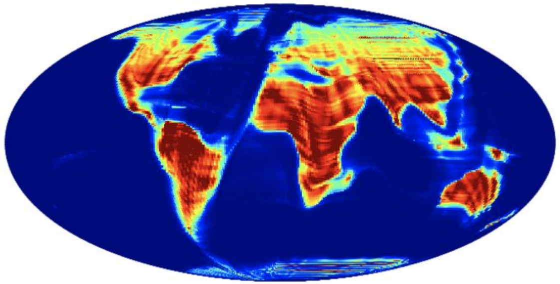
Flows on spheres: Results on S2



Target

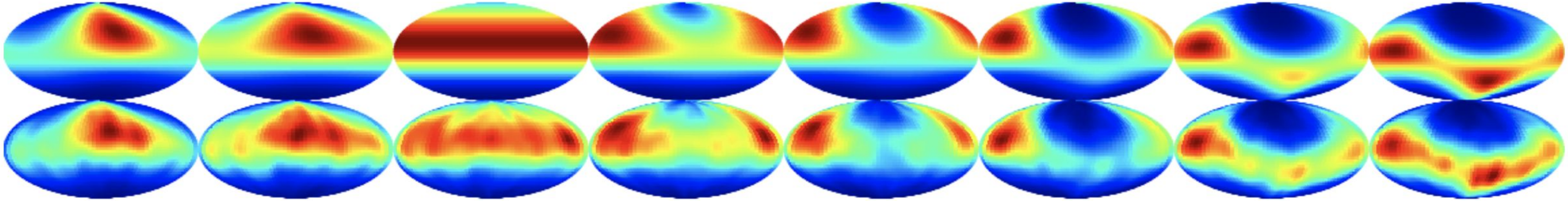


Flow samples



Flow density

Flows on spheres: Results on SU(2) ⇔ S3

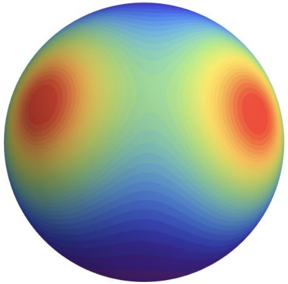


Target (top) vs flow (bottom) on a 3D sphere (shown are Mollweide projections)

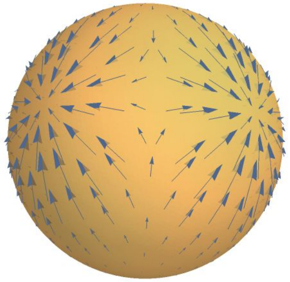
Flows on spheres: Exponential-Map flow

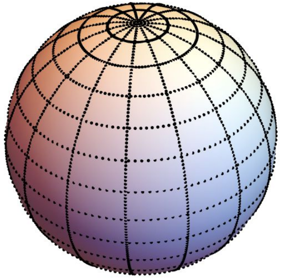
$$\exp_x : T_x S^n \to S^n \qquad \exp_x(v) = x \cos \|v\| + \frac{v}{\|v\|} \sin \|v\|$$

$$\phi(x) \qquad \nabla \phi(x) \in T_x \mathcal{M}$$

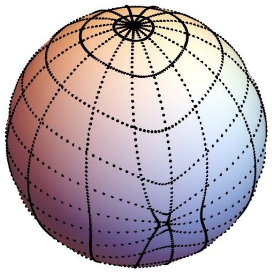


$$\Rightarrow$$





$$x \to \exp_x(\nabla \phi(x))$$



Autoregressive VS Exponential map flows on N-Spheres

	Model	KL [nats]	ESS
Auto-reg	MS ($N_T = 1, K_m = 12, K_s = 32$)	0.05 (0.01)	90%
Exp-map	EMP ($N_T = 1$)	0.50 (0.09)	43%
	EMSRE ($N_T = 1, K = 12$)	0.82 (0.30)	42%
	EMSRE ($N_T = 6, K = 5$)	0.19 (0.05)	75%
	EMSRE ($N_T = 24, K = 1$)	0.10 (0.10)	85%

Autoregressive VS Exponential map flows on N-Spheres

Method	Autoregressive	Exponential map
Pros	• Easy to scale ~O(N) • Modular	• Intrinsic to the sphere (does not require any particular coordinate system) • Defined everywhere on the sphere (no numerical instabilities) • Simpler to incorporate known symmetries
Cons	• Requires removing a set of measure-zero from the n-sphere, this may lead to numerical issues • Requires a particular coordinate system • Hard to combine with domain knowledge about density (e.g. symmetries)	• Hard to scale ~O(N^3) • More constrained family of flows

Takeaways

- Not all data are Euclidean!
- Directional statistics
- Normalizing flows on tori and spheres
  - As flexible as we like
  - Any-dimensional
  - Efficient and exact density evaluation and sampling
- Paper available at: arxiv.org/abs/2002.02428