Riemannian Convex Potential Maps

Samuel Cohen*, **Brandon Amos***, Yaron Lipman

UCL FACEBOOK AI מכון ויצמן למדע WEIZMANN INSTITUTE OF SCIENCE

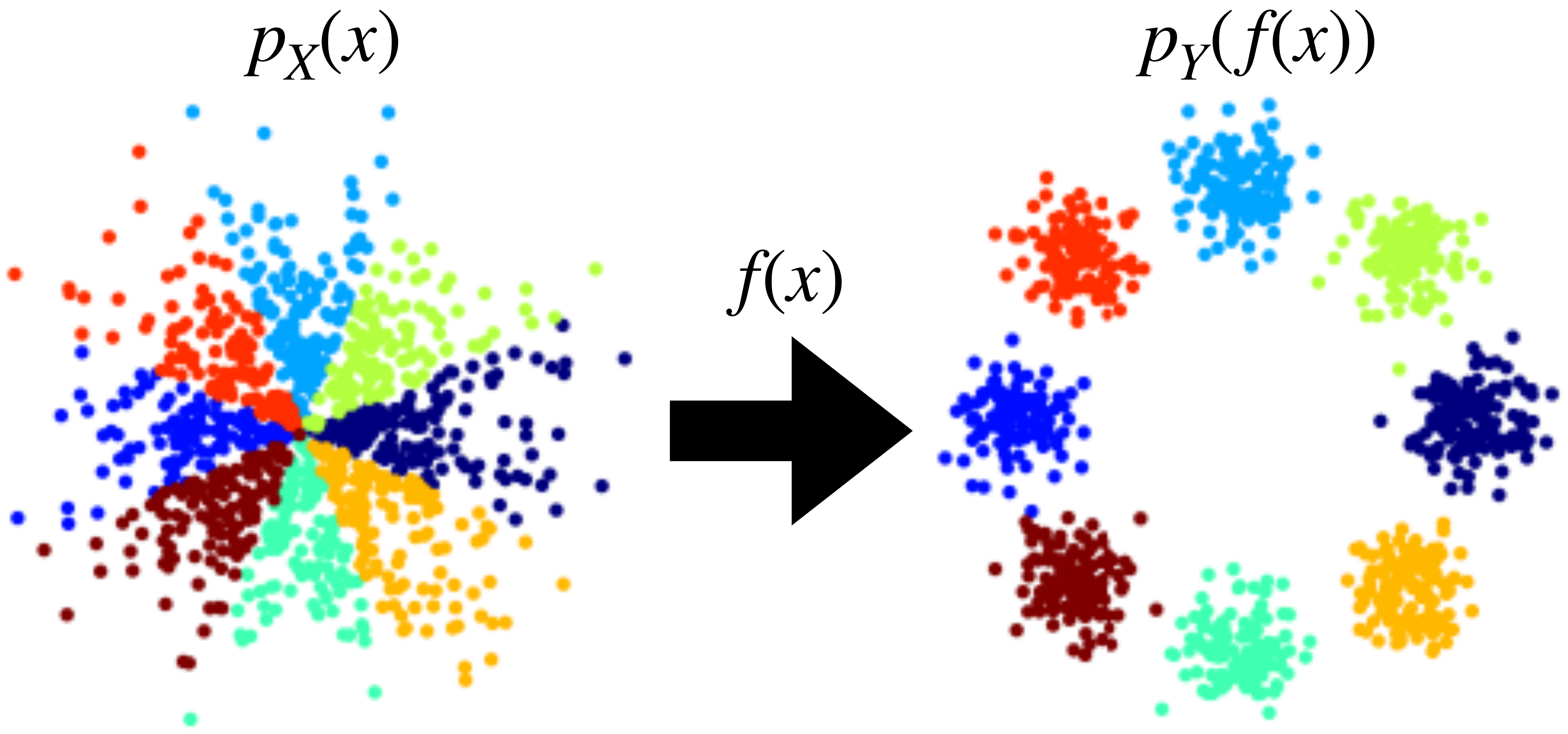
Normalizing flows are powerful models

Open problem: How to best-model f?

- Model a **differentiable, invertible flow** f_θ such that

$$p_Y(f_\theta(x)) = p_X(x) \left| \frac{\partial f_\theta(x)}{\partial x} \right|^{-1}$$

Convex Potential Maps



Glow

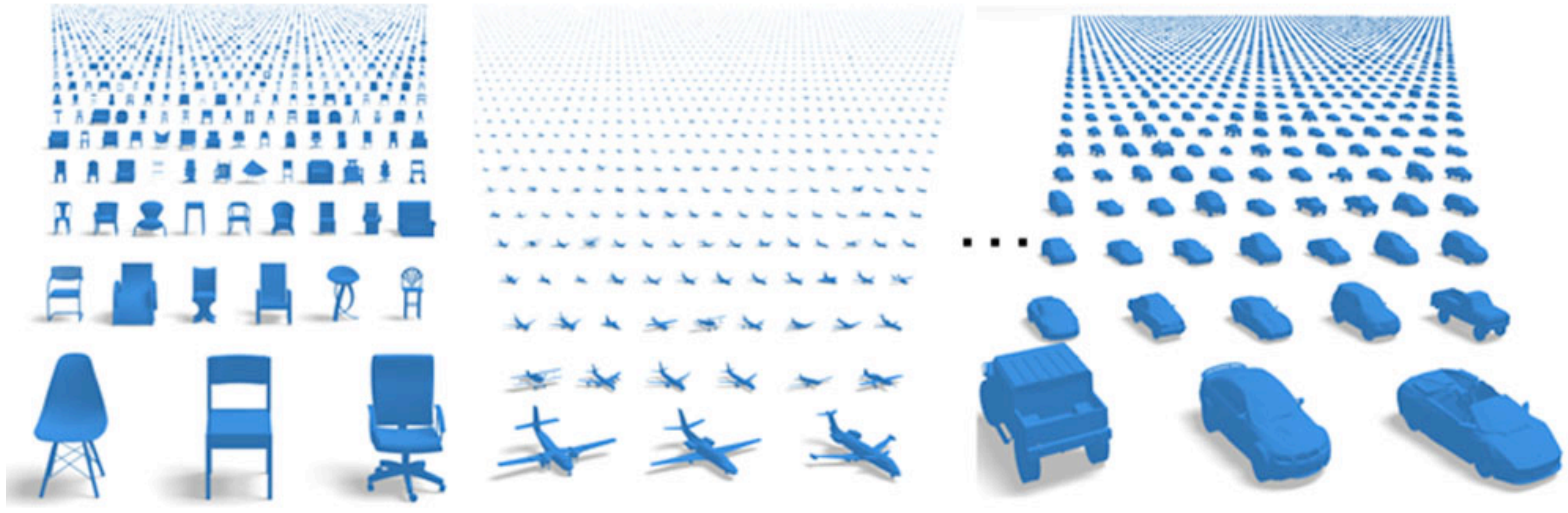
How to best-model the flow?

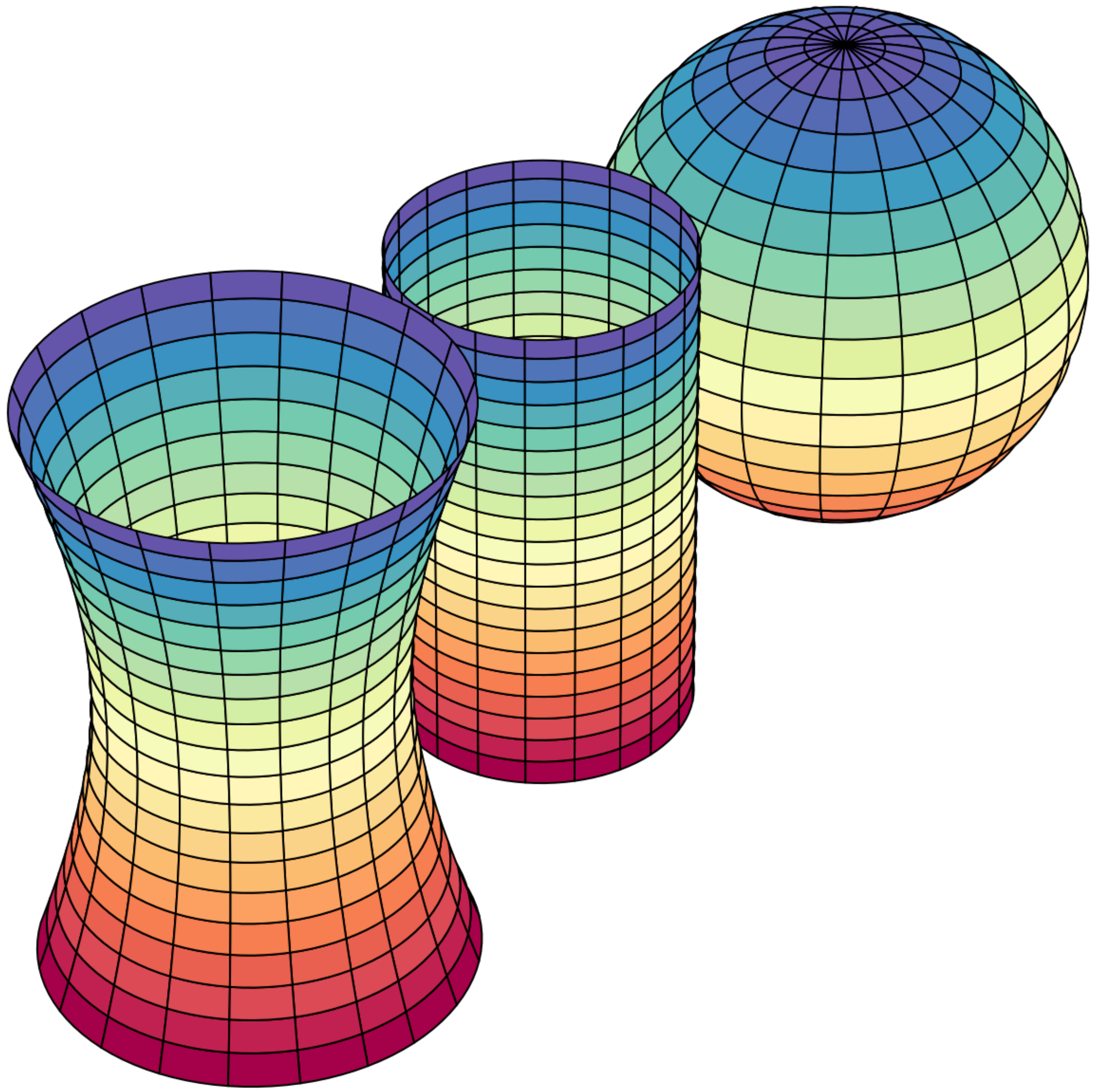
- The **Jacobian determinant** $|\partial f(x)/\partial x|$ needs to be **invertible**
  - Leads to specialized architectures (RealNVP, NICE, Glow, MAF, IAF)
  - **Challenge:** Ensuring the flow is universal
    - Does the model have the capacity to model **any** distribution?

Motivation: Surfaces and Riemannian Manifolds

- Many physical phenomena live in **non-Euclidean geometries**
  - **Riemannian manifolds** are locally-Euclidean surfaces
  - Let's model and learn **distributions** on them!

ShapeNet





This talk: Convex optimization and flows

- **Convex Potential Flows** with **Input-Convex Neural Networks**
- **Riemannian Optimal Transportation**
- **Riemannian Convex Potential Maps**

Convex Potential Flows: Universal Probability Distributions with Optimal Transport and Convex Optimization

Chin-Wei Huang
University of Montreal & Mila
chin-wei.huang@umontreal.ca

Ricky T. Q. Chen
University of Toronto & Vector Institute
rtqichen@cs.toronto.edu

Christos Tsirigotis
University of Montreal & Mila
christos.tsirigotis@umontreal.ca

Aaron Courville
University of Montreal, Mila & CIFAR Fellow
aaron.courville@umontreal.ca

Abstract

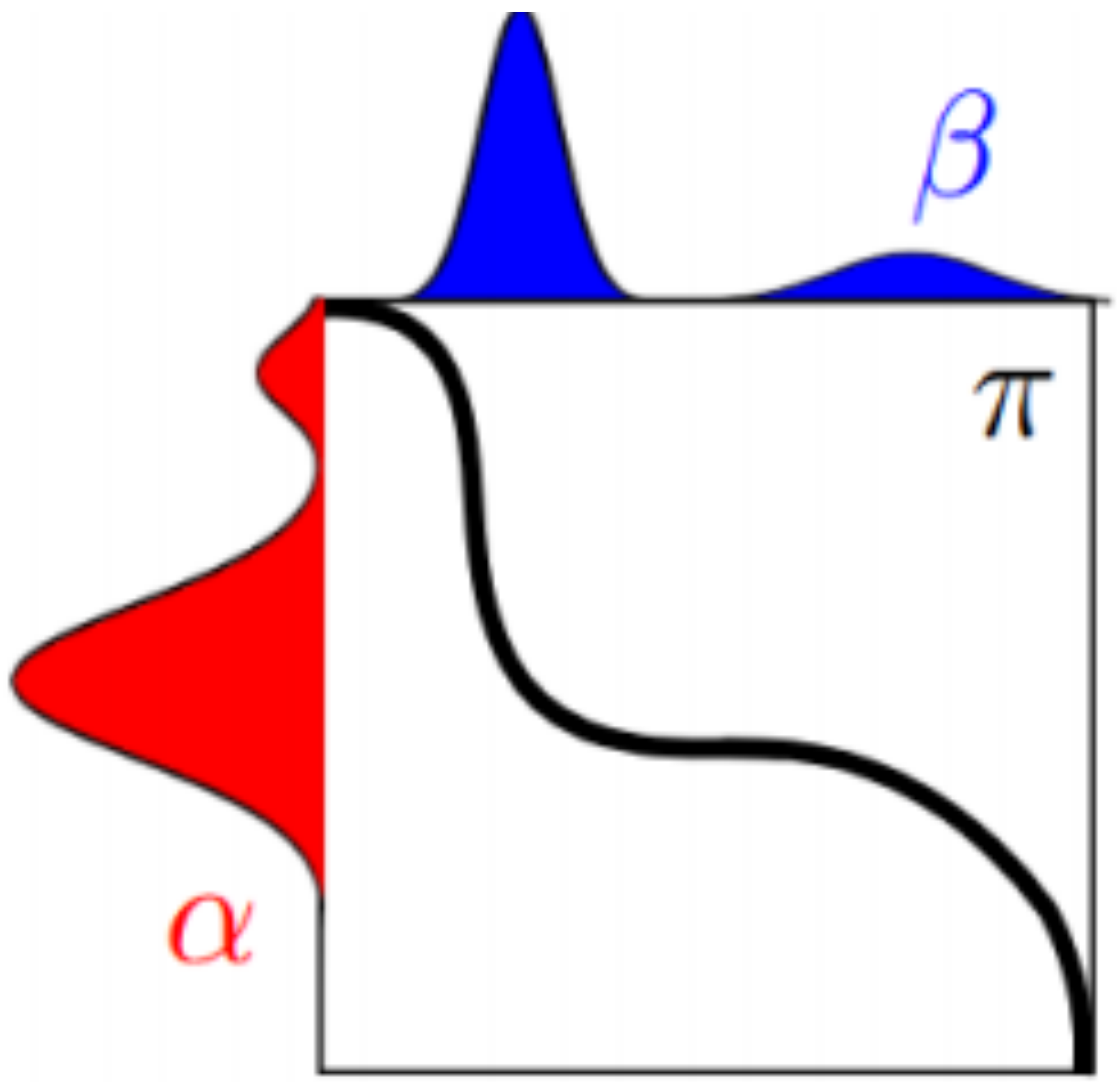
Flow-based models are powerful tools for designing probabilistic models with tractable density. This paper introduces Convex Potential Flows (CP-Flow), a natural and efficient parameterization of invertible models inspired by the optimal transport (OT) theory. CP-Flows are the gradient map of a strongly convex neural potential function. The convexity implies invertibility and allows us to resort to convex optimization to solve the convex conjugate for efficient inversion. To enable maximum likelihood training, we derive a new gradient estimator of the log-determinant of the Jacobian, which involves solving an inverse-Hessian vector product using the conjugate gradient method. The gradient estimator has *constant-memory* cost, and can be made effectively *unbiased* by reducing the error tolerance level of the convex optimization routine. Theoretically, we prove that CP-Flows are *universal* density approximators and are *optimal* in the OT sense. Our empirical results show that CP-Flow performs competitively on standard benchmarks of density estimation and variational inference.

Background: Optimal Transport

- Optimal transport seeks to find an **optimal coupling** π between measures α and β

- **Monge's formulation:** Represent the coupling as a map π and find the minimum cost one:

$$\min_{\pi:\pi(\alpha)\sim p_\beta} \mathbb{E}_{\alpha\sim p_\alpha}\left[c(\alpha,\pi(\alpha))\right]$$



(Source: Computational Optimal Transport)

Brenier's Theorem

Theorem 1 (**Brenier's Theorem**, Theorem 1.22 of Santambrogio (2015)). *Let μ, ν be probability measures with a finite second moment, and assume μ has a Lebesgue density p_X. Then there exists a convex potential G such that the gradient map $g := \nabla G$ (defined up to a null set) uniquely solves the Monge problem in eq. (2) with the quadratic cost function $c(x, y) = ||x - y||^2$.*

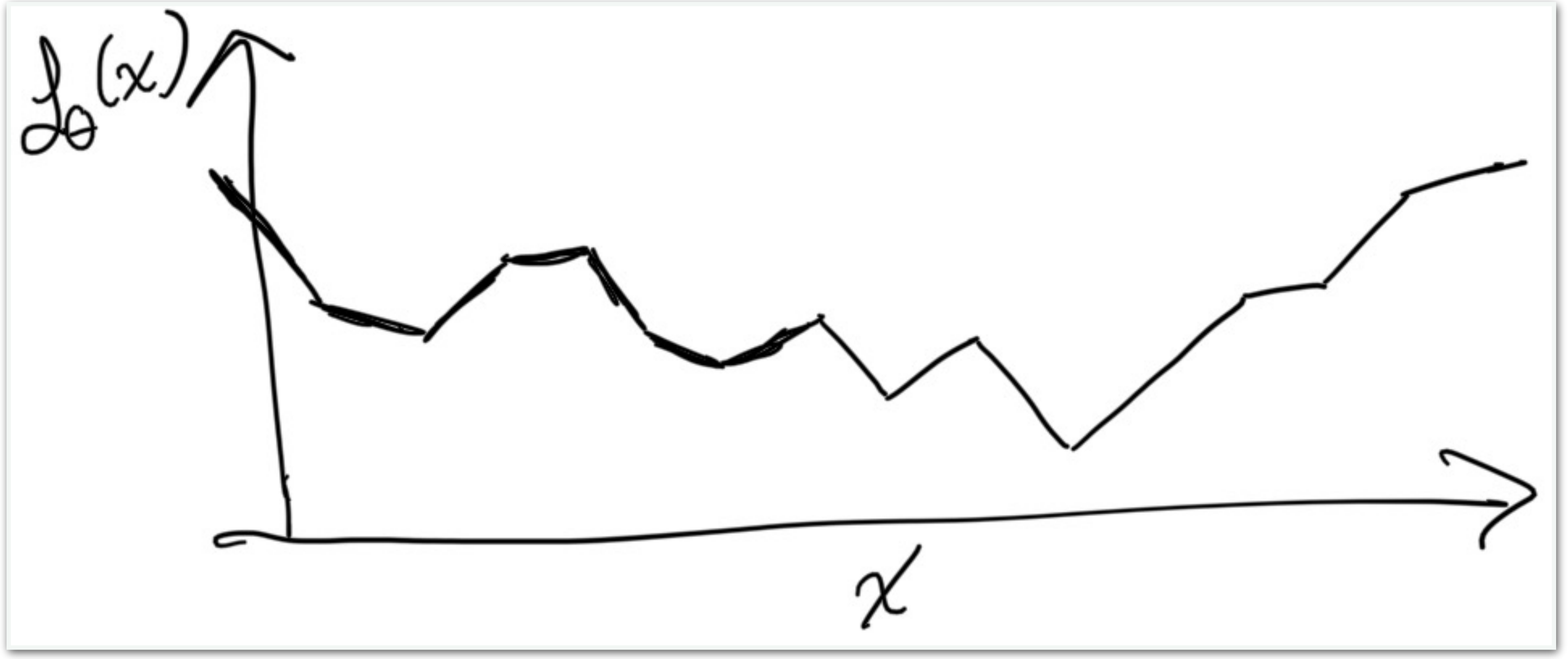
(Source: Convex potential flows)

- Celebrated result in optimal transport 🎉
- **Monge problems** can be solved using **gradients of a convex function**
  - I.e., $\pi(x) = \nabla G(x)$
- **Idea:** Construct a flow using (gradients of) **convex functions**

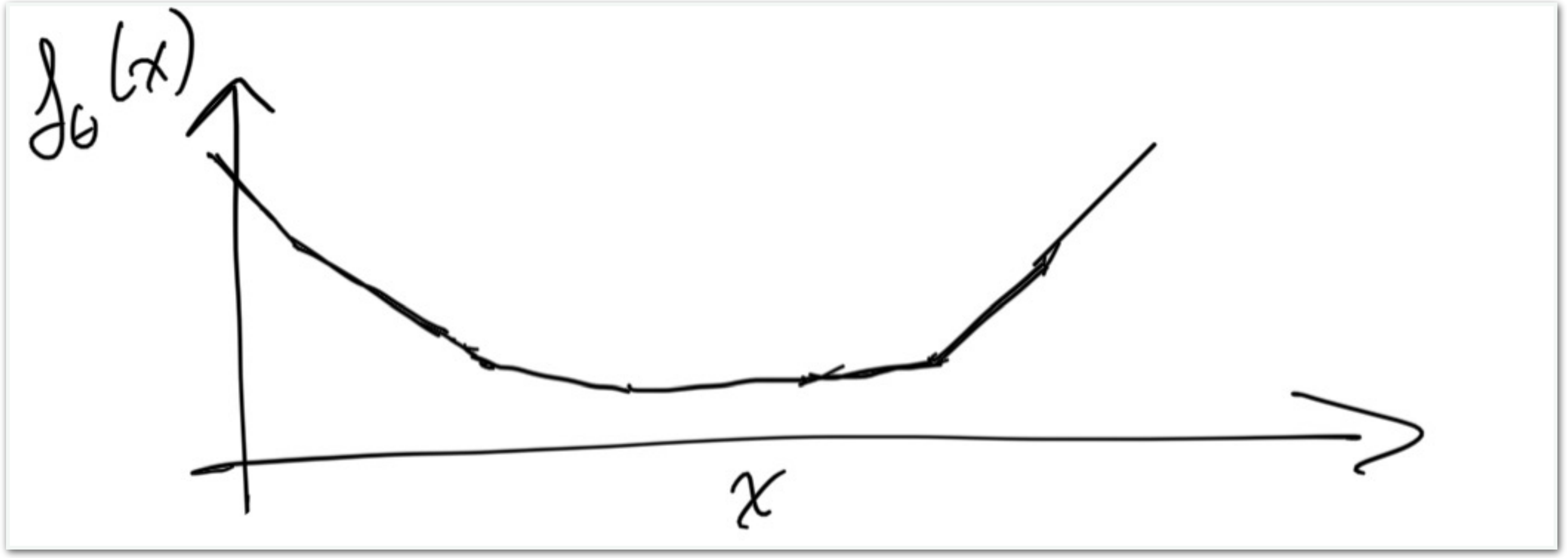
Model with input-convex neural networks

Input-Convex Neural Networks

- **Fact:** ReLU neural nets represent non-convex piecewise linear function



- **Idea:** Constrain them to (universally) represent convex functions



How to achieve input convexity?

- Most networks can be "trivially" modified to **guarantee input-convexity**
- Consider a simple feedforward k-layer **ReLU network:** (for $i = 1, \ldots, k$)

$$z_{i+1} = \max\left\{0, W_i z_i + b_i\right\} \qquad f(x;\theta) = z_k + 1 \qquad z_1 = x$$

- **Theorem.** f is convex in y provided that the W_i are non-negative for $i > 1$
- Any **convex and non-decreasing activation function** has this property

Summary: Convex Potential Flows

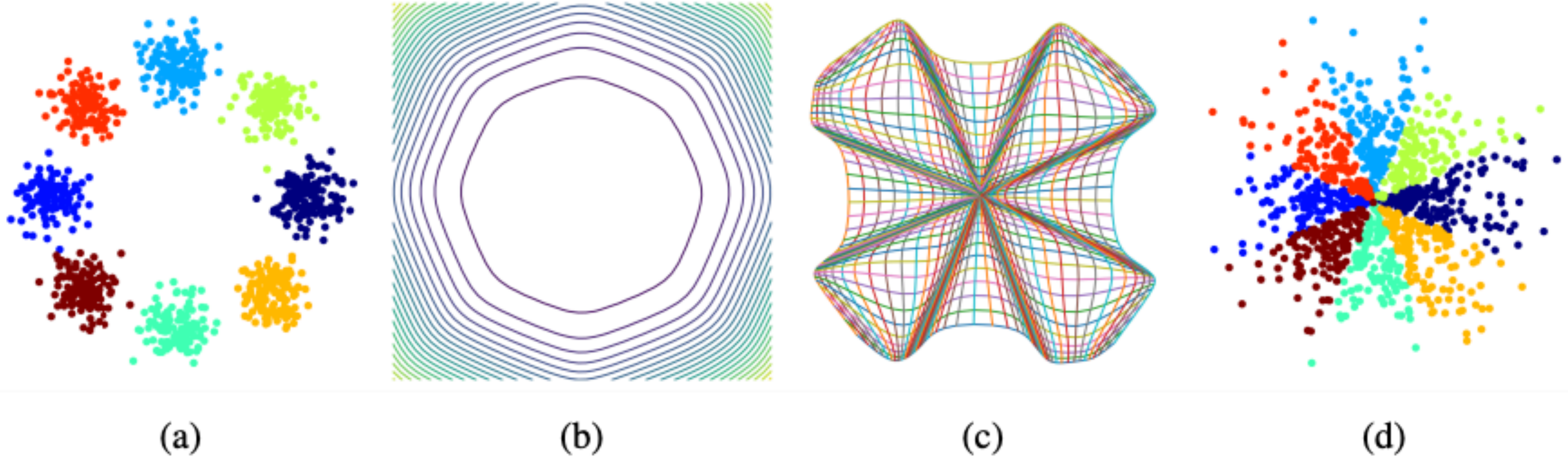


Figure 1: Illustration of Convex Potential Flow. (a) Data x drawn from a mixture of Gaussians. (b) Learned convex potential F. (c) Mesh grid distorted by the gradient map of the convex potential $f = \nabla F$. (d) Encoding of the data via the gradient map $z = f(x)$. Notably, the encoding is the *value of the gradient* of the convex potential. When the curvature of the potential function is locally flat, gradient values are small and this results in a contraction towards the origin.

Convex Potential Flows are Universal

1. **ICNNs** model the gradient of any convex function

2. Apply **Brenier's theorem** (any flow is the gradient of a convex function)

Related work on Euclidean convex potential flows

1. Korotin et al. "**Wasserstein-2 Generative Networks.**" 2019.

2. Taghvaei & Jalali. "**2-wasserstein approximation via restricted convex potentials with application to improved training for gans**." 2019.

3. Makkuva et al. "**Optimal transport mapping via input convex neural networks.**" 2019.

4. Finlay et al. "**Learning normalizing flows from Entropy-Kantorovich potentials.**" 2020.

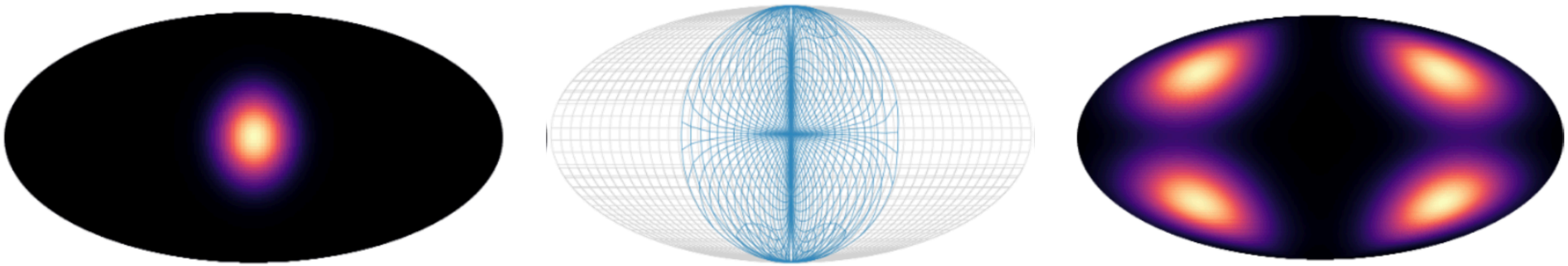
This talk: Convex optimization and flows

- Convex Potential Flows with Input-Convex Neural Networks
- **Riemannian Optimal Transportation**
- Riemannian Convex Potential Maps

Riemannian Optimal Transport

- Given source μ and target ν measures on manifolds find an (OT) map pushing source to target.

$$\underset{t:t_{\#}\mu=\nu}{\operatorname{argmin}} \int_{\mathcal{M}} c[x, t(x)]d\mu$$



μ ν

c-convexity

- **Standard convexity** is just for Euclidean spaces
- **c-convexity** is an extension that can be applied to Riemannian manifolds [Villani 2009]
  - The **cost** $c : \mathcal{X} \times \mathcal{Y} \to (-\infty, +\infty]$ can be, e.g., a manifold distance
- **Definitions.** Let ψ be a function and $\mathcal{X}, \mathcal{Y}$ be sets.
  - ψ **is c-convex** if it can be written as $\psi(x) = \sup_y \big(\zeta(y) - c(x, y)\big)$ for all x
  - The c**-transform** of ψ is $\psi^c(y) = \inf_x \big(\psi(x) + c(x, y)\big)$

Connecting c-convexity and Euclidean convexity

- **Captures Euclidean convexity** with $c(x, y) = -x^\top y$
- The c-transform becomes the **Legendre** transform $\psi^c(y) = \inf_x \left(\psi(x) - x^\top y\right)$
- c-convexity definition: ψ is c-convex if it can be represented as the convex conjugate of another function ζ

McCann's Extension to Brenier's Theorem

- **Brenier's theorem** was originally for Euclidean spaces with quadratic costs
    - **Monge transport map** can be represented as $t(x) = \nabla\phi$ with ϕ convex
- **McCann's result** extends it to **Riemannian spaces** using c-convexity
    - $t(x) = \exp(\nabla\phi)$ with ϕ c-convex

POLAR FACTORIZATION OF MAPS ON RIEMANNIAN MANIFOLDS

Robert J. McCann*
Department of Mathematics
University of Toronto, Toronto Ontario Canada M5S 3G3
mccann@math.toronto.edu

May 27, 1999

Abstract

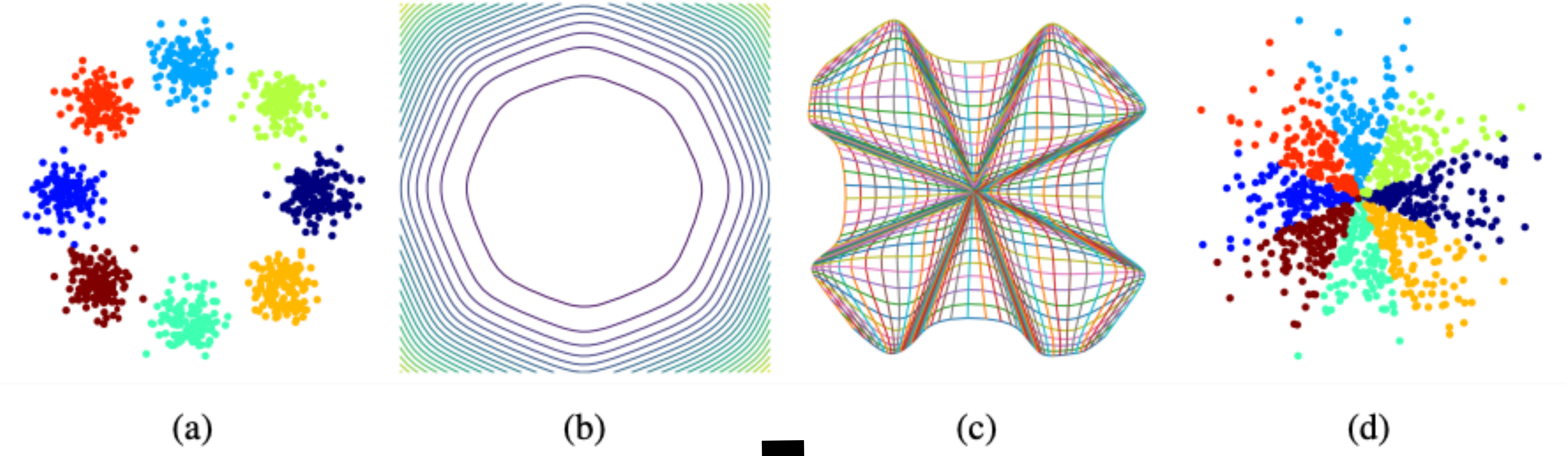
Let (M, g) be a connected compact manifold, C^3 smooth and without boundary, equipped with a Riemannian distance $d(x, y)$. If $s : M \longrightarrow M$ is merely Borel and never maps positive volume into zero volume, we show $s = t \circ u$ factors uniquely a.e. into the composition of a map $t(x) = \exp_x[-\nabla\psi(x)]$ and a volume-preserving map $u : M \longrightarrow M$, where $\psi : M \longrightarrow \mathbf{R}$ is an infimal convolution with $c(x, y) = d^2(x, y)/2$. Like the factorization it generalizes from Euclidean space, this non-linear decomposition can be linearized around the identity to yield the Hodge decomposition of vector fields.

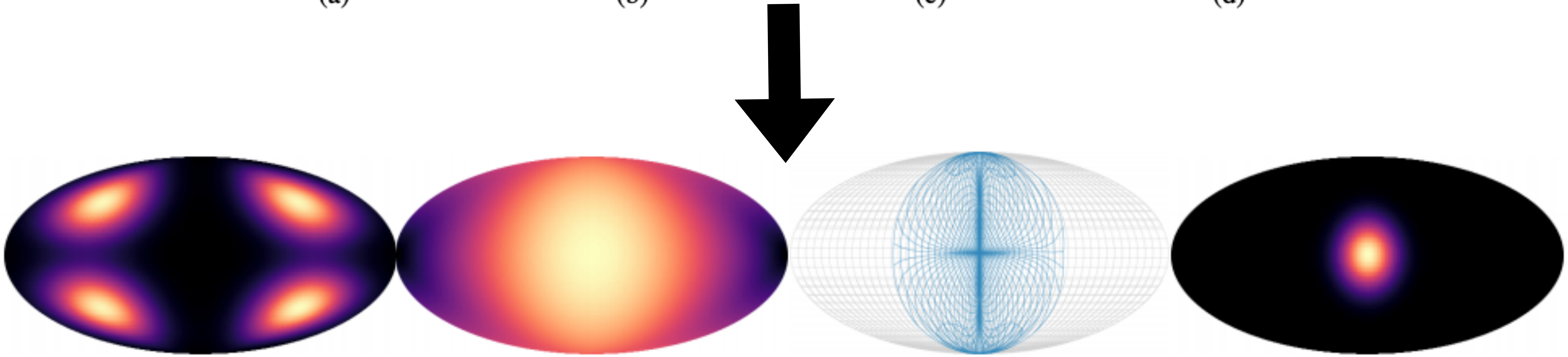
The results are obtained by solving a Riemannian version of the Monge-Kantorovich problem, which means minimizing the expected value of the cost $c(x, y)$ for transporting one distribution $f \geq 0$ of mass in $L^1(M)$ onto another. A companion article extends this solution to strictly convex or concave cost functions $c(x, y) \geq 0$ of the Riemannian distance on non-compact manifolds.

This talk: Convex optimization and flows

- Convex Potential Flows with Input-Convex Neural Networks
- Riemannian Optimal Transportation
- **Riemannian Convex Potential Maps**

Summary: Riemannian Extension of Convex Potential Flows



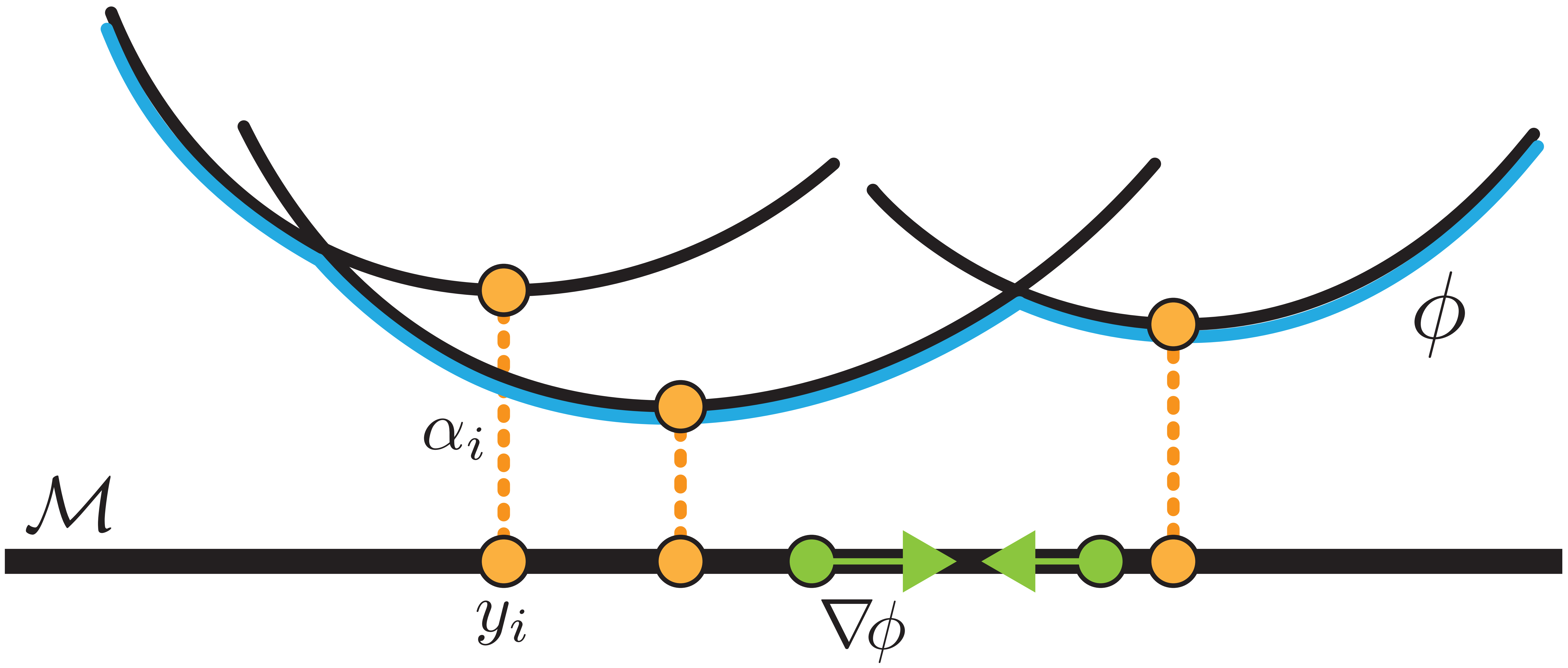


Our c-convex potential: Semidiscrete OT

- **Our semidiscrete OT on manifolds**

  Use discrete c-concave potentials of the form $\phi(x) = \min_{i \in [n]} c(x, y_i) + \alpha_i$

Learnable parameters



Theory: Universality

Theorem 1: For compact, boundaryless, smooth manifolds, $\left\{f \,|\, f(x) = \min_{i \in [n]} c(x, y_i) + \alpha_i \right\}$ is dense in $\{f \,|\, f$ is c-concave$\}$.

Theorem 2: If μ, ν are regular, there exists a sequence of discrete c-concave potentials ϕ_ϵ such that $\exp[-\nabla \phi_\epsilon] \xrightarrow{p} t$ where t is the OT map.

Implementation Details

- **Map architecture**: stack of multiple blocks of the form

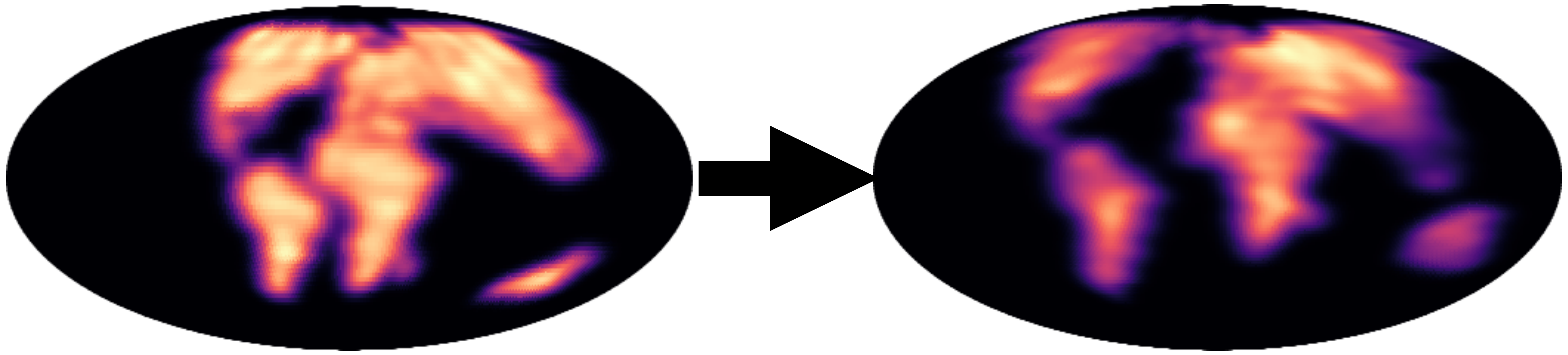
$$s_j(y_j) = \exp[-\nabla_{y_j}\phi_j(y_j)], \qquad j = 1,\ldots,T$$

- **Smoothing**: applied to discrete c-concave layers

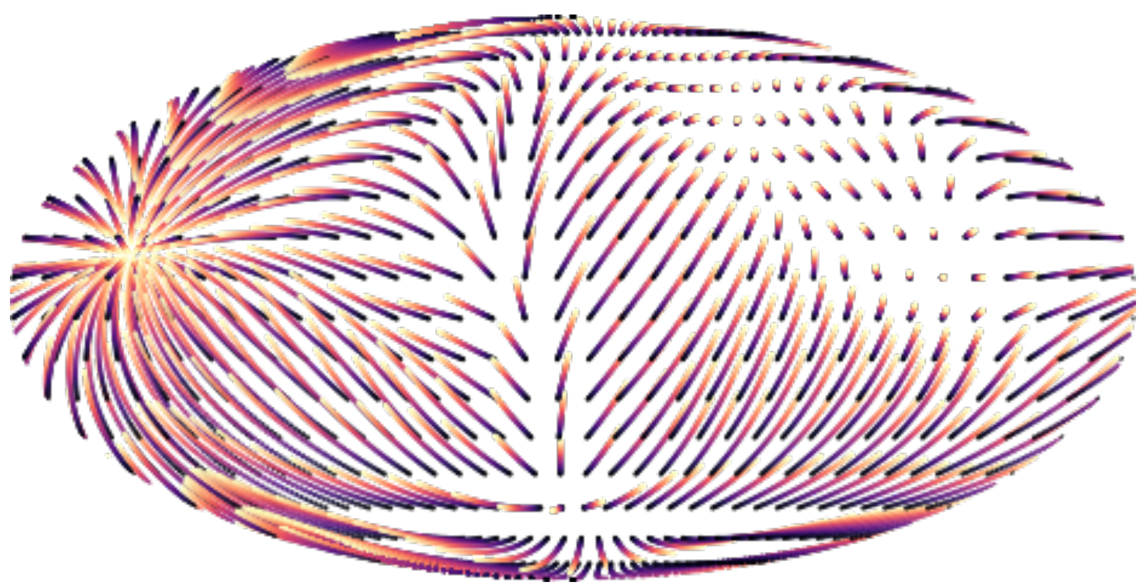
$$\min_\gamma(a_1, \ldots, a_n) = -\gamma \log \sum_{i=1}^{n} \exp -\frac{a_i}{\gamma}$$

- **Loss**: standard density estimation losses (NLL, KL)

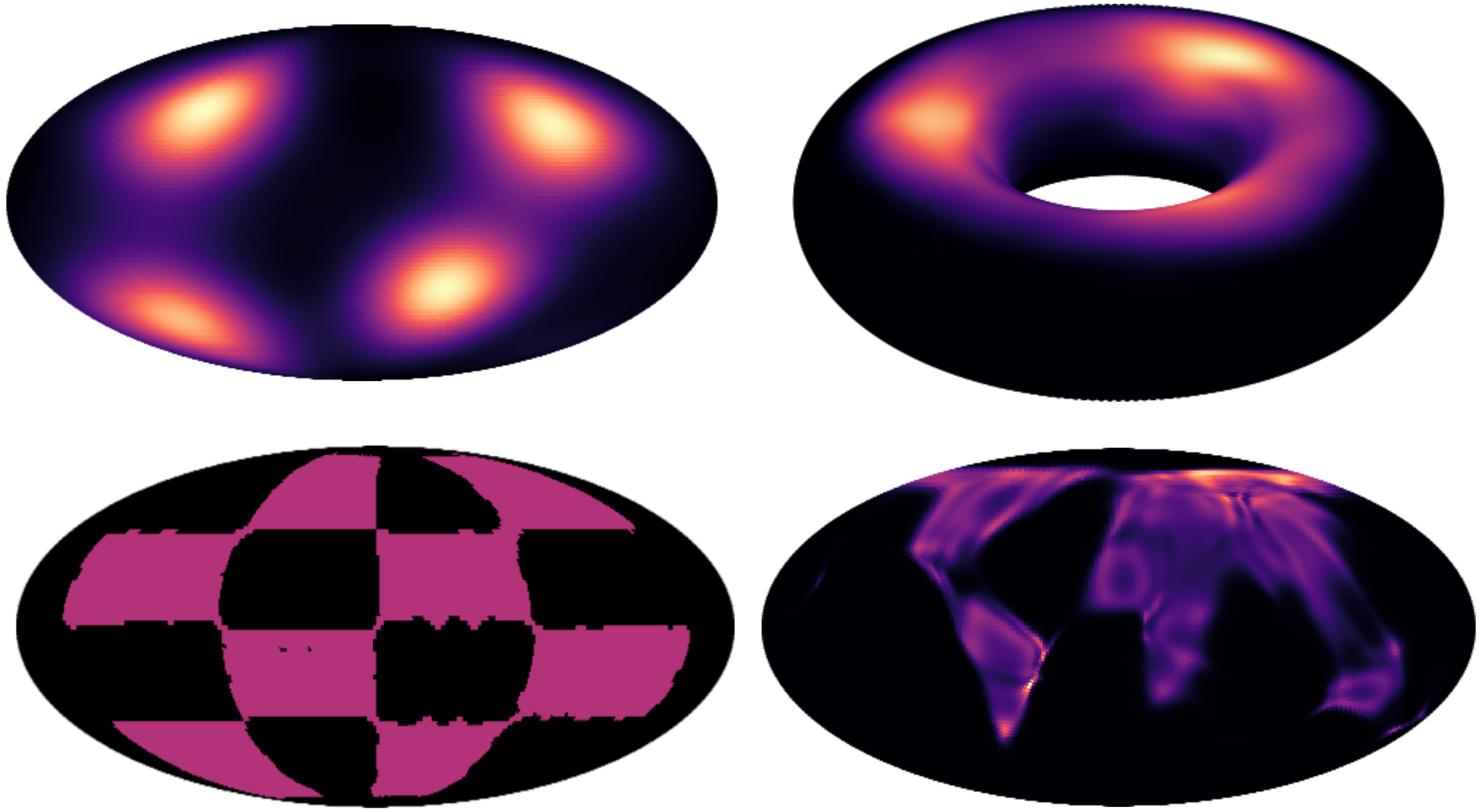
Results



Density Transportation



Geodesics Estimation



Density Estimation

Related work on exponential map flows

A Jacobian inequality for gradient maps on the sphere and its application to directional statistics

Tomonari SEI

September 16, 2018

Abstract

In the field of optimal transport theory, an optimal map is known to be a gradient map of a potential function satisfying cost-convexity. In this paper, the Jacobian determinant of a gradient map is shown to be log-concave with respect to a convex combination of the potential functions when the underlying manifold is the sphere and the cost function is the distance squared. The proof uses the non-negative cross-curvature property of the sphere recently established by Kim and McCann, and Figalli and Rifford. As an application to statistics, a new family of probability densities on the sphere is defined in terms of cost-convex functions. The log-concave property of the likelihood function follows from the inequality.

Normalizing Flows on Tori and Spheres

Danilo Jimenez Rezende, George Papamakarios, Sébastien Racanière, Michael S. Albergo, Gurtej Kanwar, Phiala E. Shanahan, Kyle Cranmer

Normalizing flows are a powerful tool for building expressive distributions in high dimensions. So far, most of the literature has concentrated on learning flows on Euclidean spaces. Some problems however, such as those involving angles, are defined on spaces with more complex geometries, such as tori or spheres. In this paper, we propose and compare expressive and numerically stable flows on such spaces. Our flows are built recursively on the dimension of the space, starting from flows on circles, closed intervals or spheres.

Riemannian Convex Potential Maps

Samuel Cohen*, **Brandon Amos***, Yaron Lipman

- **Convex Potential Flows** with **Input-Convex Neural Networks**
- **Riemannian Optimal Transportation**
- **Riemannian Convex Potential Maps**

UCL FACEBOOK AI מכון ויצמן למדע WEIZMANN INSTITUTE OF SCIENCE