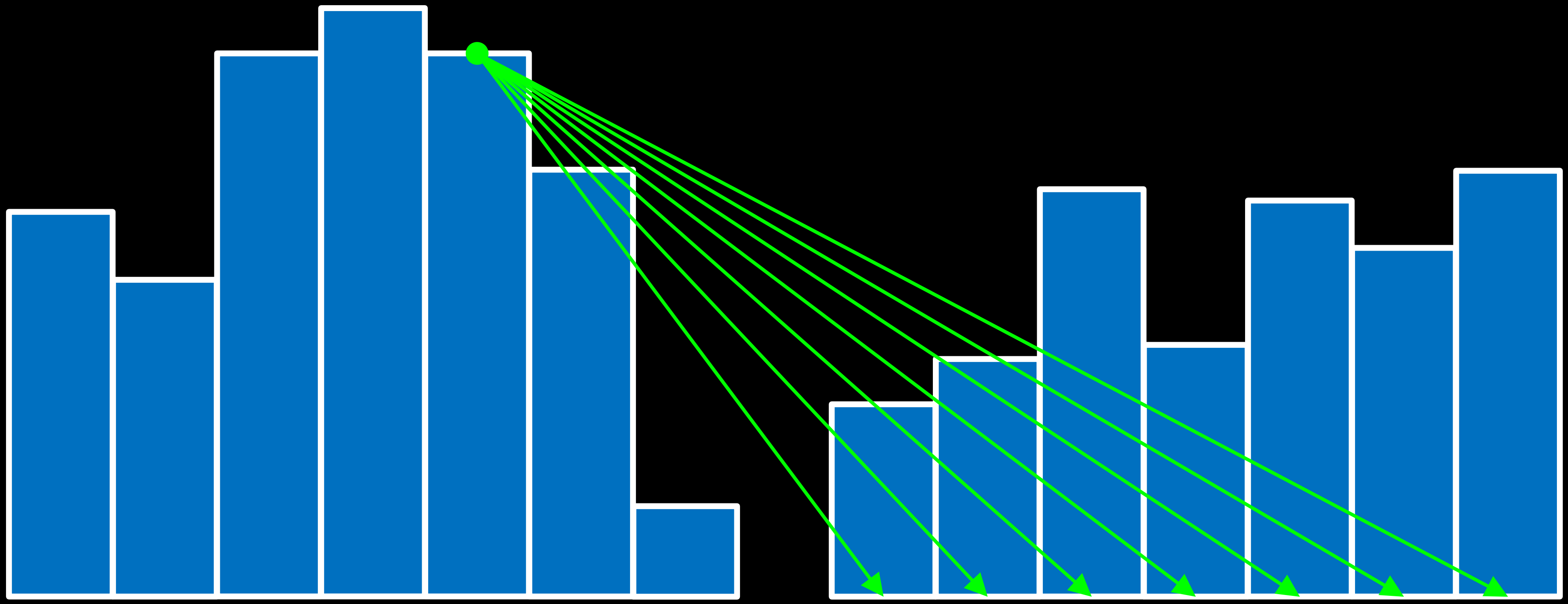


Optimal Transport
on Discrete Domains

Justin Solomon

MIT EECS


WARNING
INFORMAL
AND NOT
COMPREHENSIVE


WARNING
FEWER MESHES

What is Optimal Transport?

A geometric way to compare probability measures.

Nobel prize

Fields medal
(and French politician)


Monge


Kantorovich


Dantzig


Wasserstein

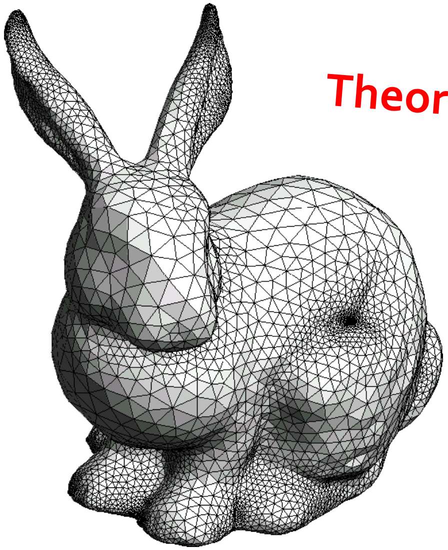

Brenier


Otto


McCann

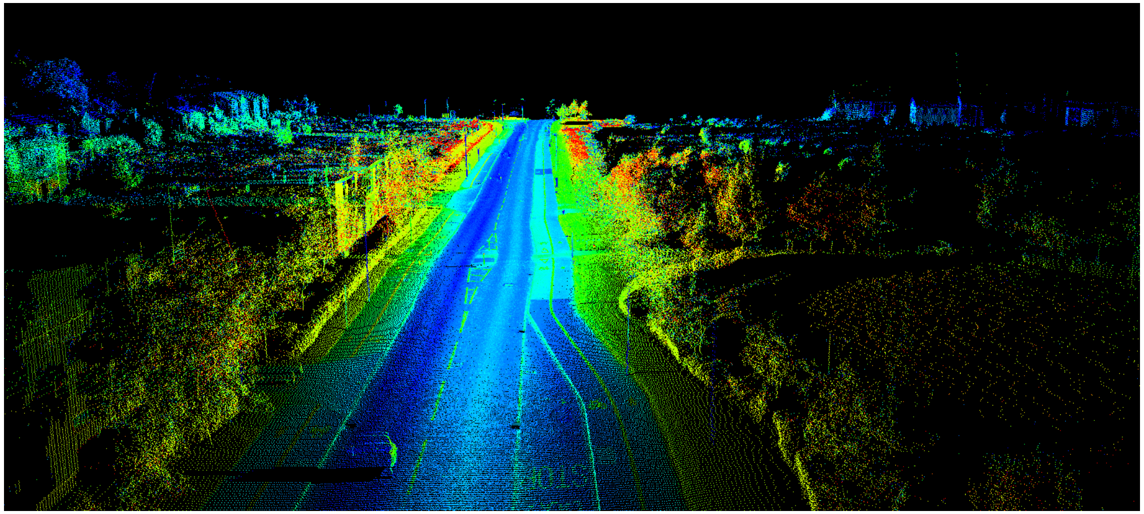

Villani

Motivation



Theory

Practice



What's Missing?

Resilience to

noise and uncertainty.

Hardly an "implementation detail!"

Plan For Today

1. Introduction to optimal transport
    - 1D examples
    - Many formulas
2. Applications
3. Discrete/discretized transport
    - Entropic regularization
    - Eulerian transport
    - Semidiscrete transport
4. Extensions

Plan For Today

1. Introduction to optimal transport
   - 1D examples
   - Many formulas
2. Applications
3. Discrete/discretized transport
   - Entropic regularization
   - Eulerian transport
   - Semidiscrete transport
4. Extensions

Our Approach

Understand geometry from a

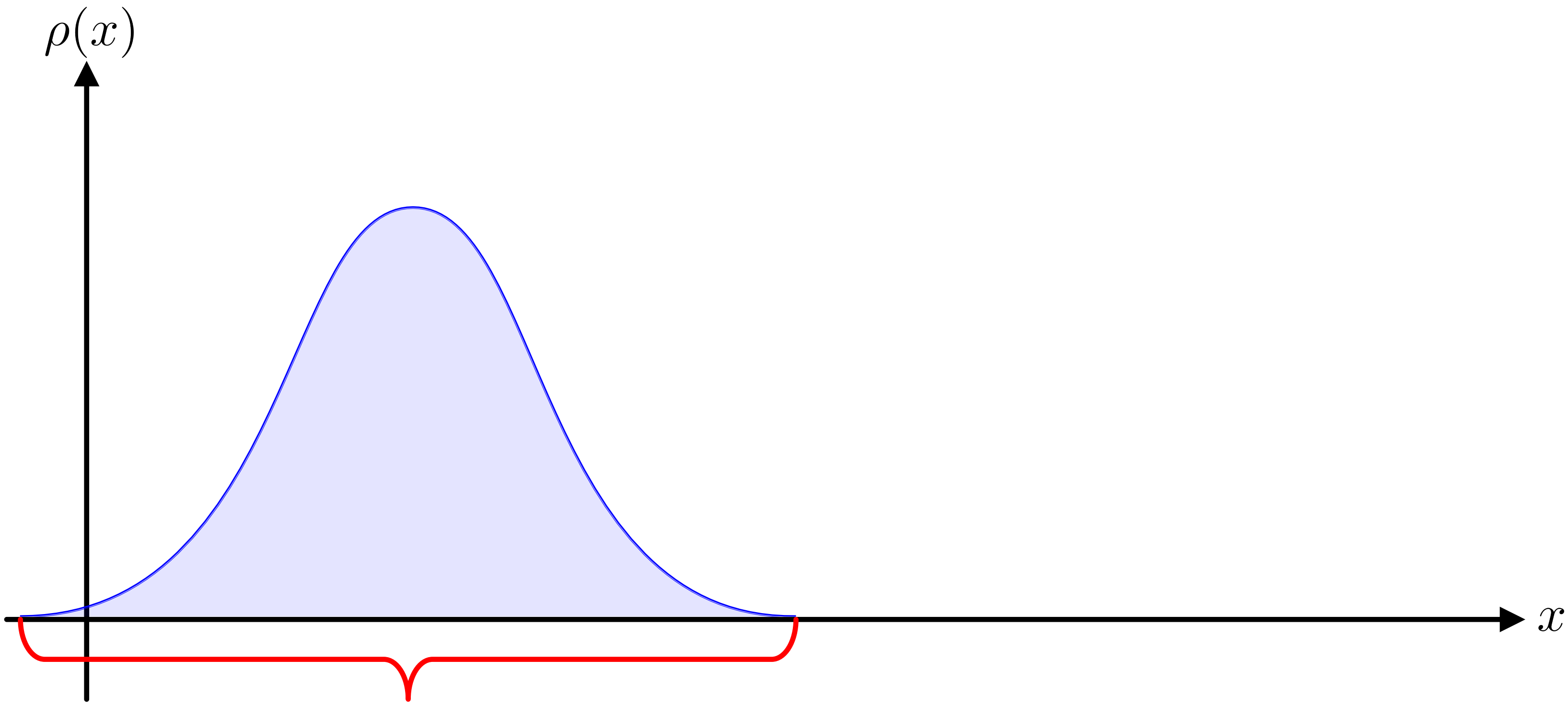
"softened" probabilistic

standpoint.

Secondary goal:

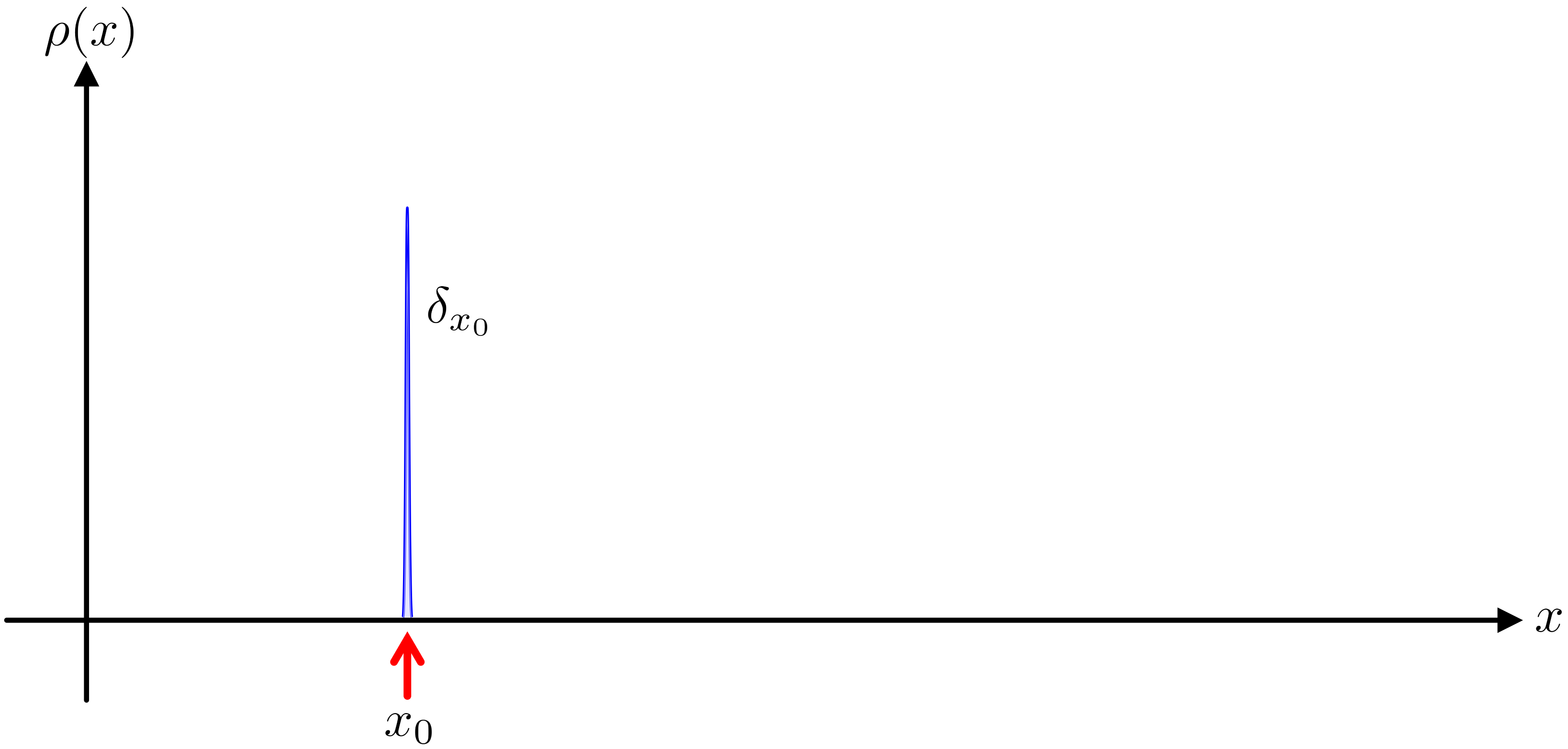
Application of machinery from previous talks
(vector fields, geodesics, meshes…)

Probability as Geometry



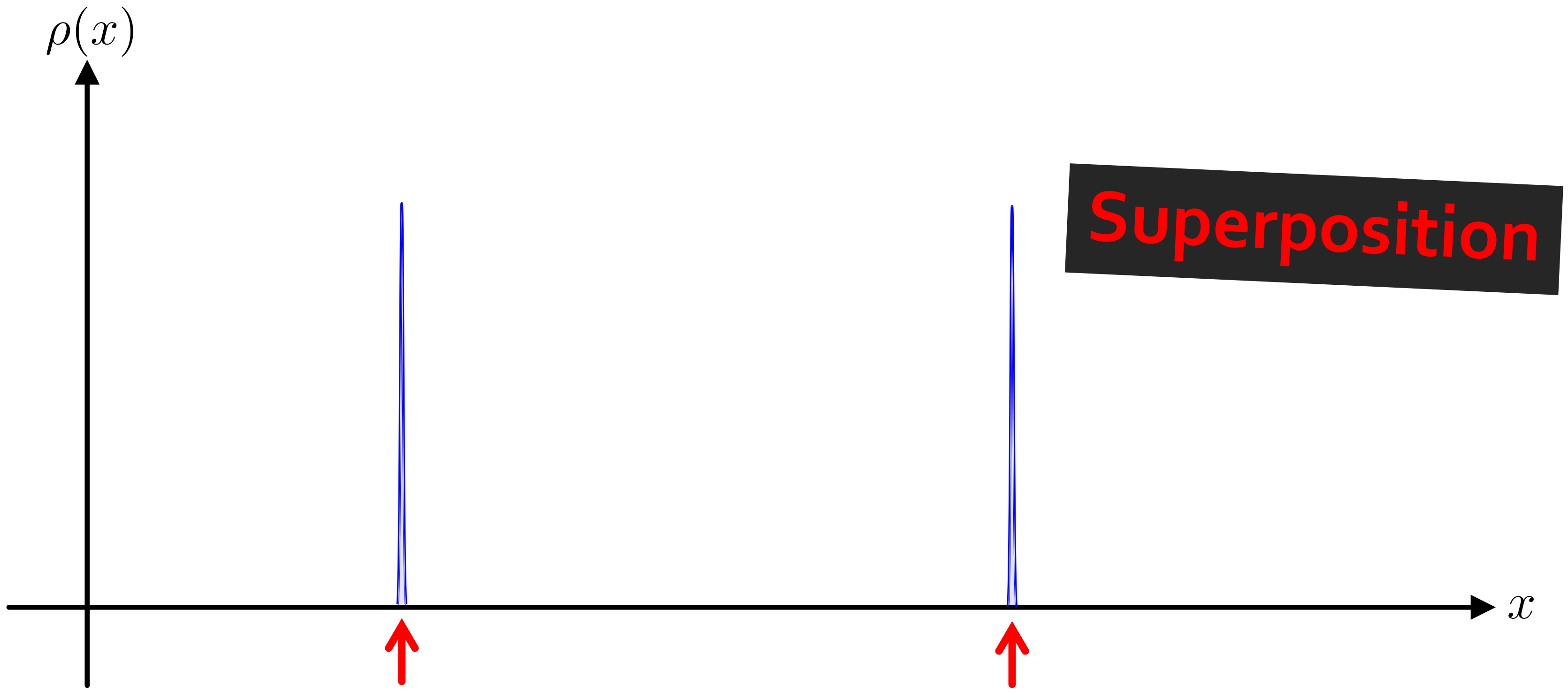
“Somewhere over here.”

Probability as Geometry



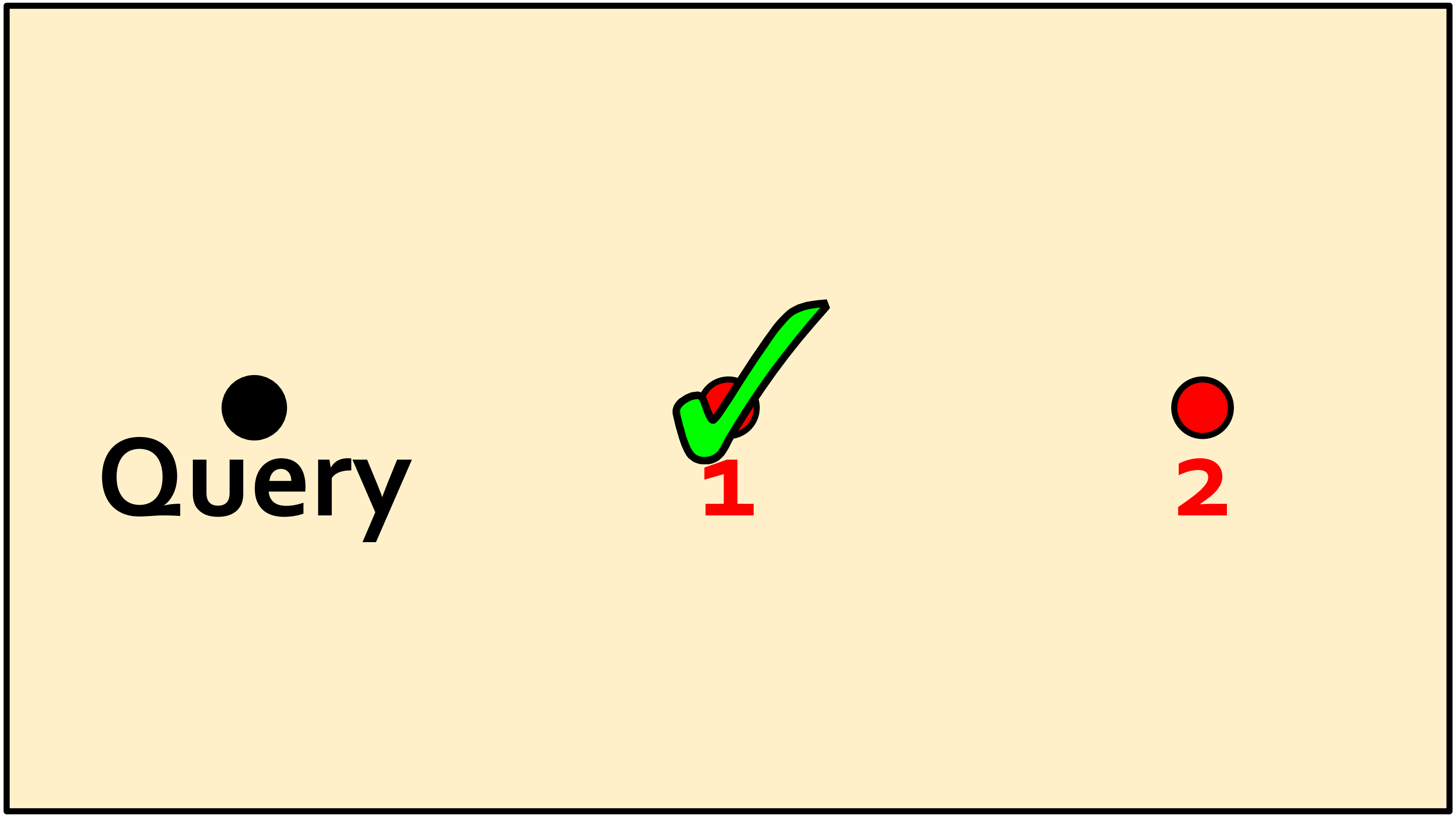
"Exactly here."

Probability as Geometry



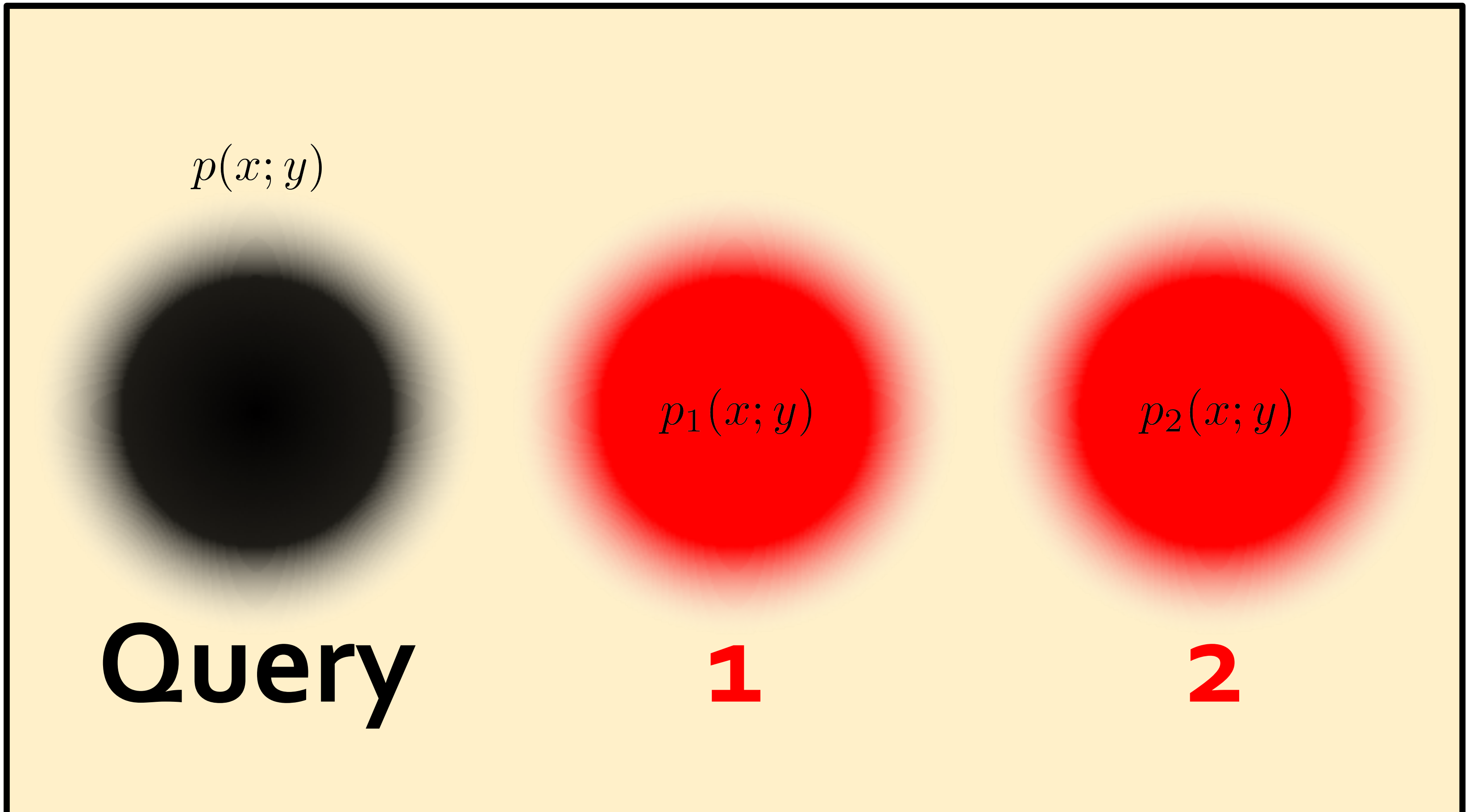
"One of these two places."

How to Incorporate Geometry?



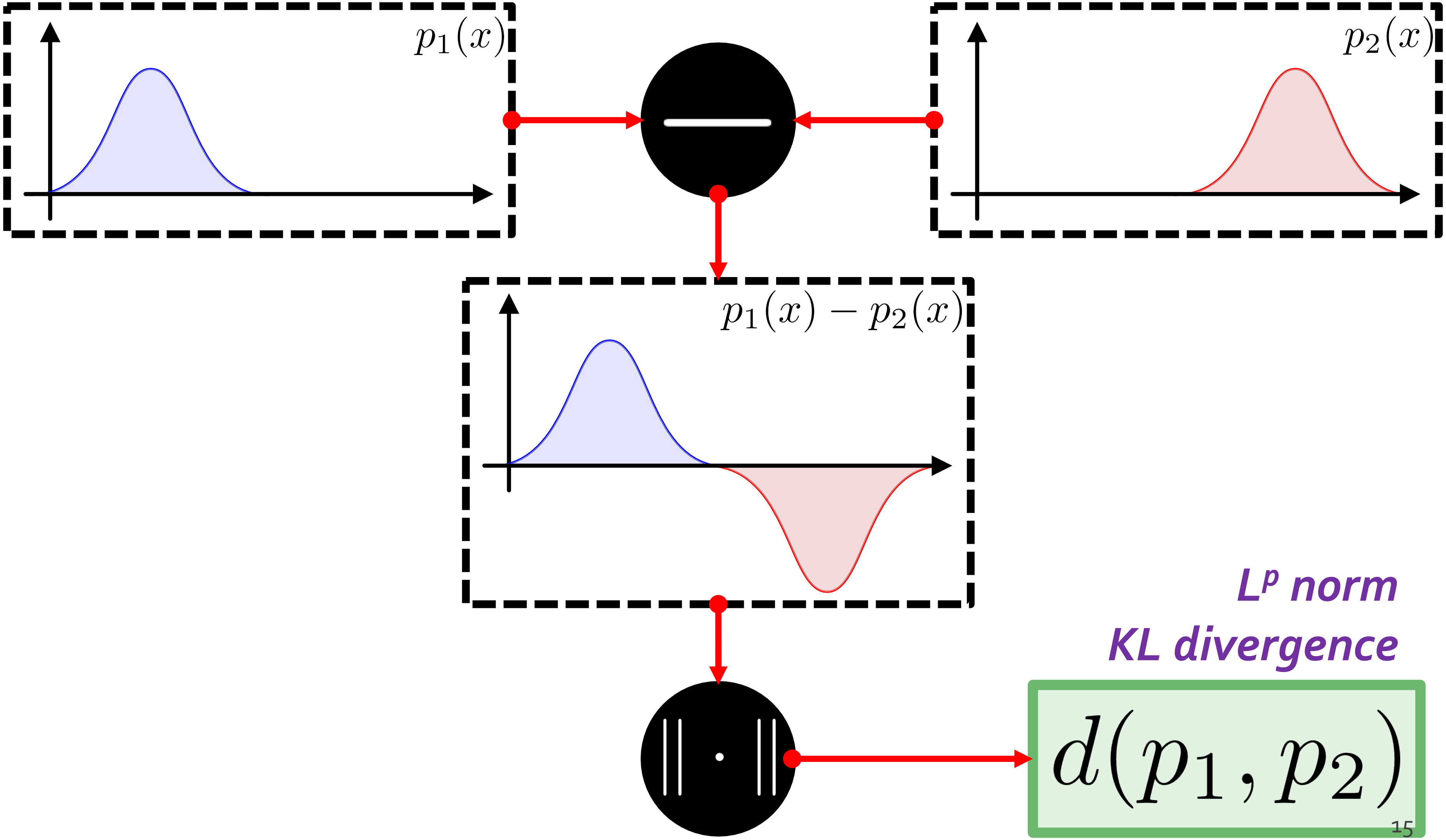
Which is closer, 1 or 2?

Fuzzy Version



Which is closer, 1 or 2?

Typical Measurement



L^p norm

KL divergence

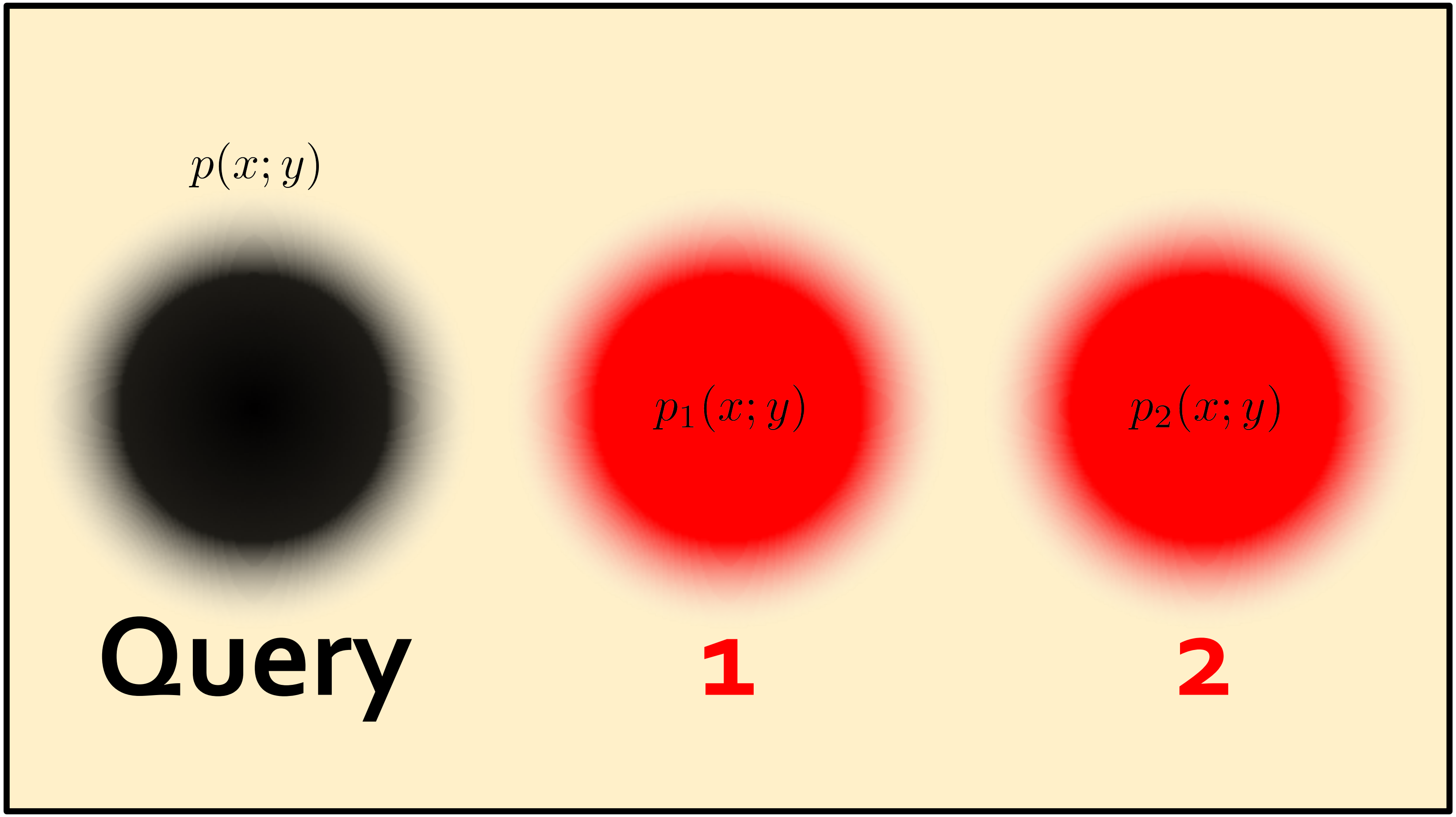
$$d(p_1, p_2)$$

Examples

$$d_{L_1}(\rho_0, \rho_1) := \int_{-\infty}^{\infty} |\rho_0(x) - \rho_1(x)| \, dx$$

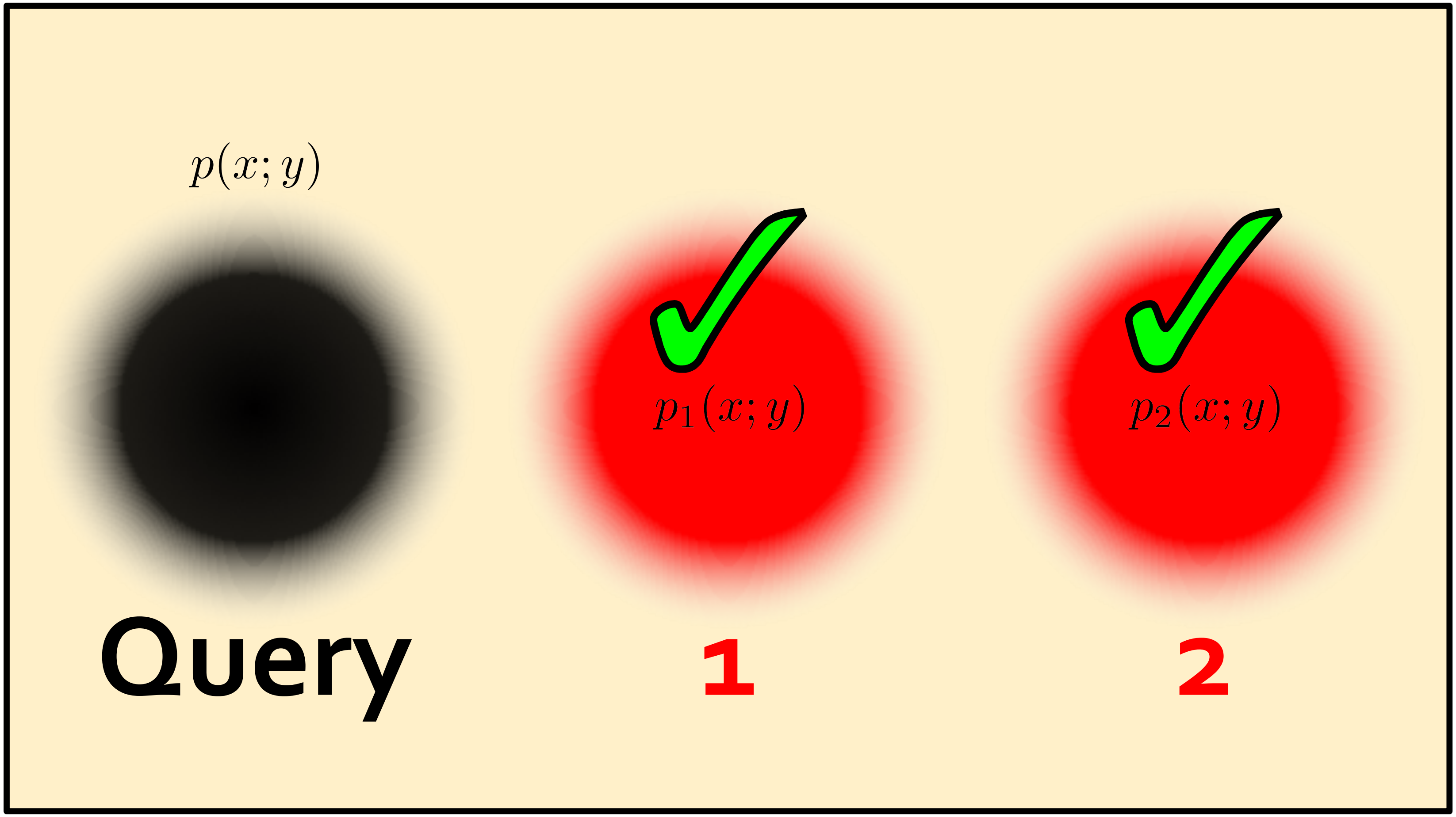
$$\mathrm{KL}(\rho_0 || \rho_1) := \int_{-\infty}^{\infty} \rho_0(x) \log \frac{\rho_0(x)}{\rho_1(x)} \, dx$$

Returning to the Question



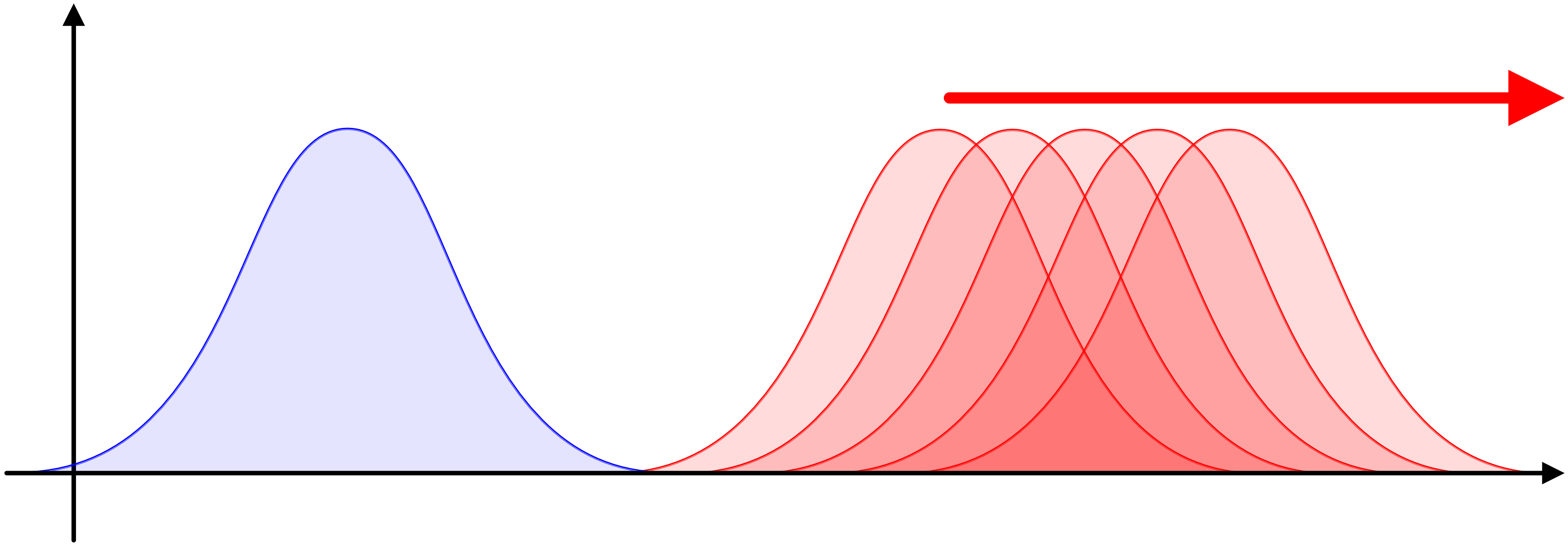
Which is closer, 1 or 2?

Returning to the Question



Neither! Equidistant.

What's Wrong?



Measured overlap, not displacement.

Put Another Way

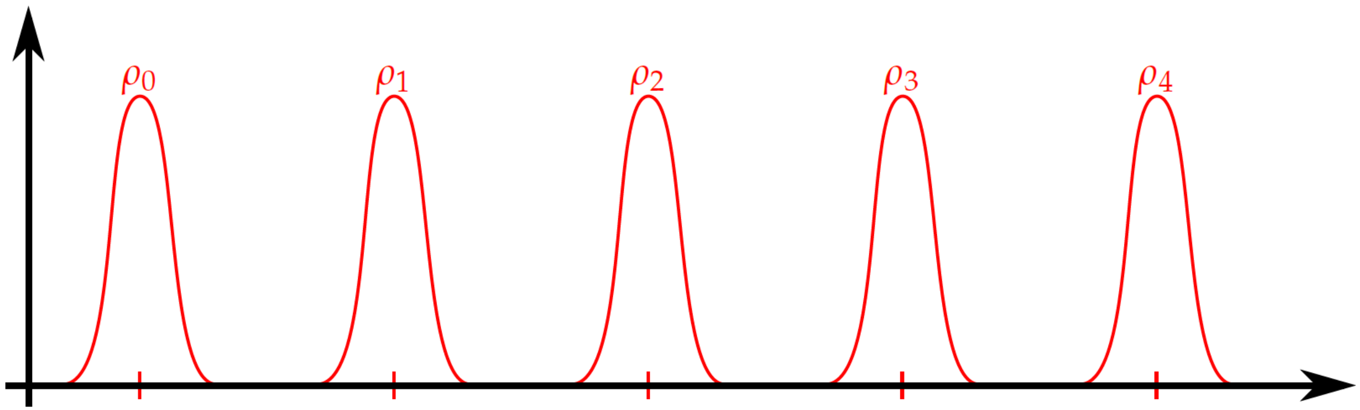
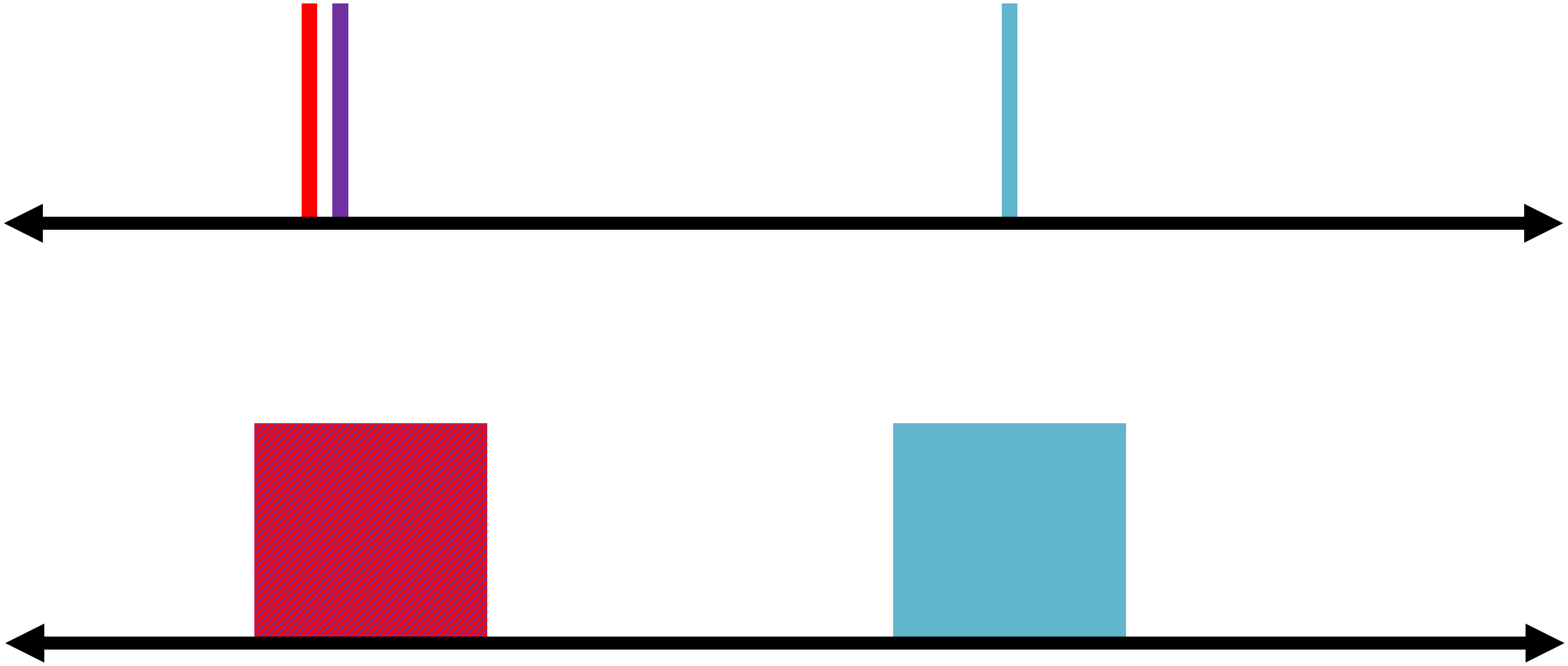


Figure 2: The distributions $\rho_0, \ldots, \rho_4$ are equidistant with respect to the L_1 and KL divergences, while the Wasserstein distance from optimal transport increases linearly with distance over $\mathbb{R}$.

Related Issue



Smaller bins worsen histogram distances

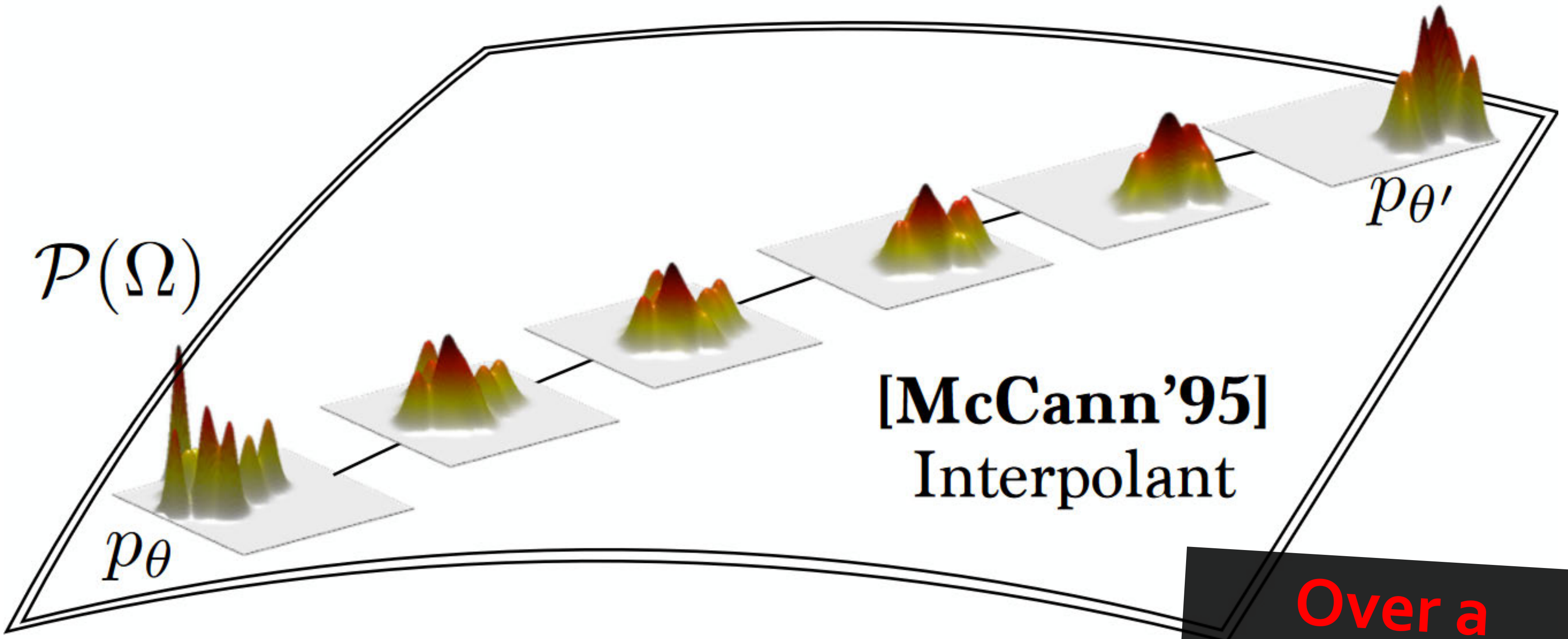
The Root Cause

Permuting histogram bins has

no effect

on these distances.

Optimal Transport

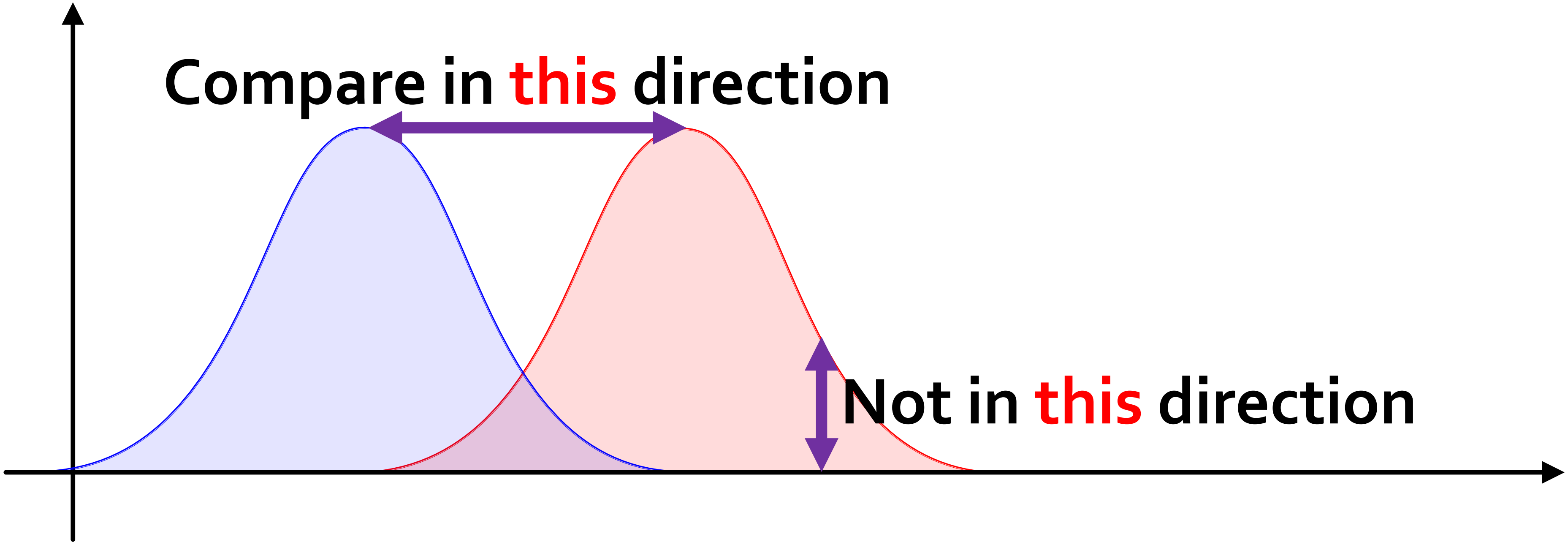


Over a metric space

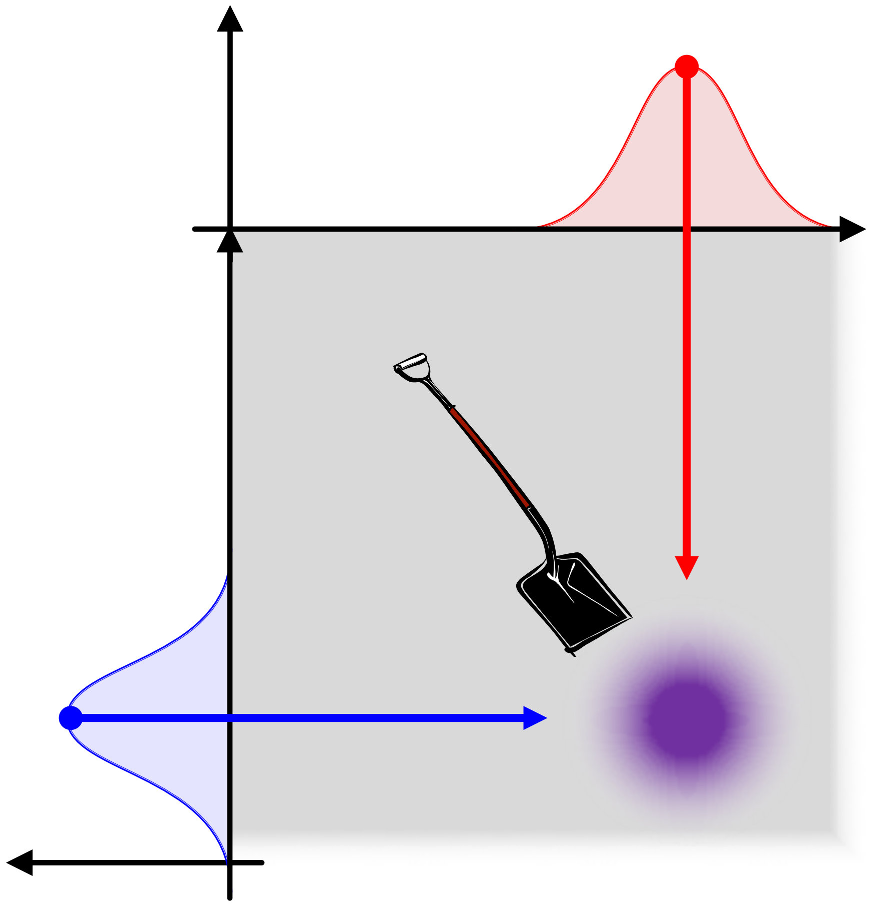
Image courtesy M. Cuturi

Geometric theory of probability

Alternative Idea

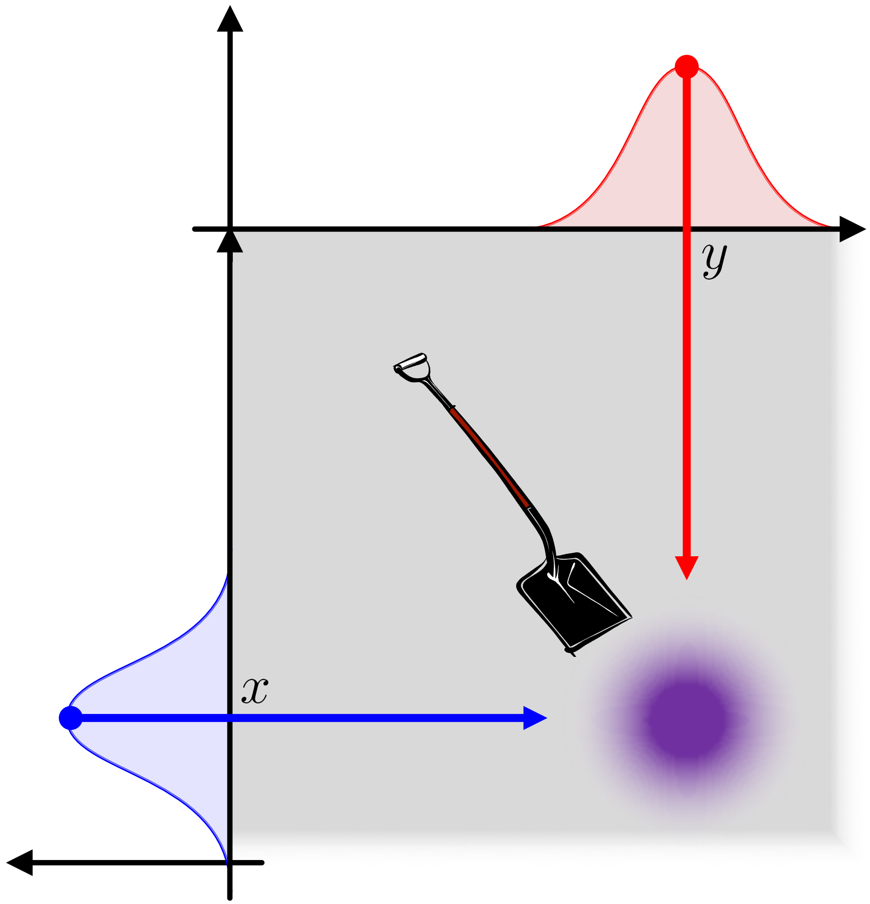


Alternative Idea



Match mass from the distributions

Alternative Idea



Cost to move mass m from x to y:

$$m \cdot d(x, y)$$

Match mass from the distributions

Observation

Even the laziest shoveler
must do some work.

Property of the distributions themselves!

My house last week!

The Setup: Transport in 1D

$$\pi(x, y) := \text{Amount moved from } x \text{ to } y$$

$$\pi(x, y) \geq 0 \ \forall x, y \in \mathbb{R}$$

Mass is positive

$$\int_{\mathbb{R}} \pi(x, y) \, dy = \rho_0(x) \ \forall x \in \mathbb{R}$$

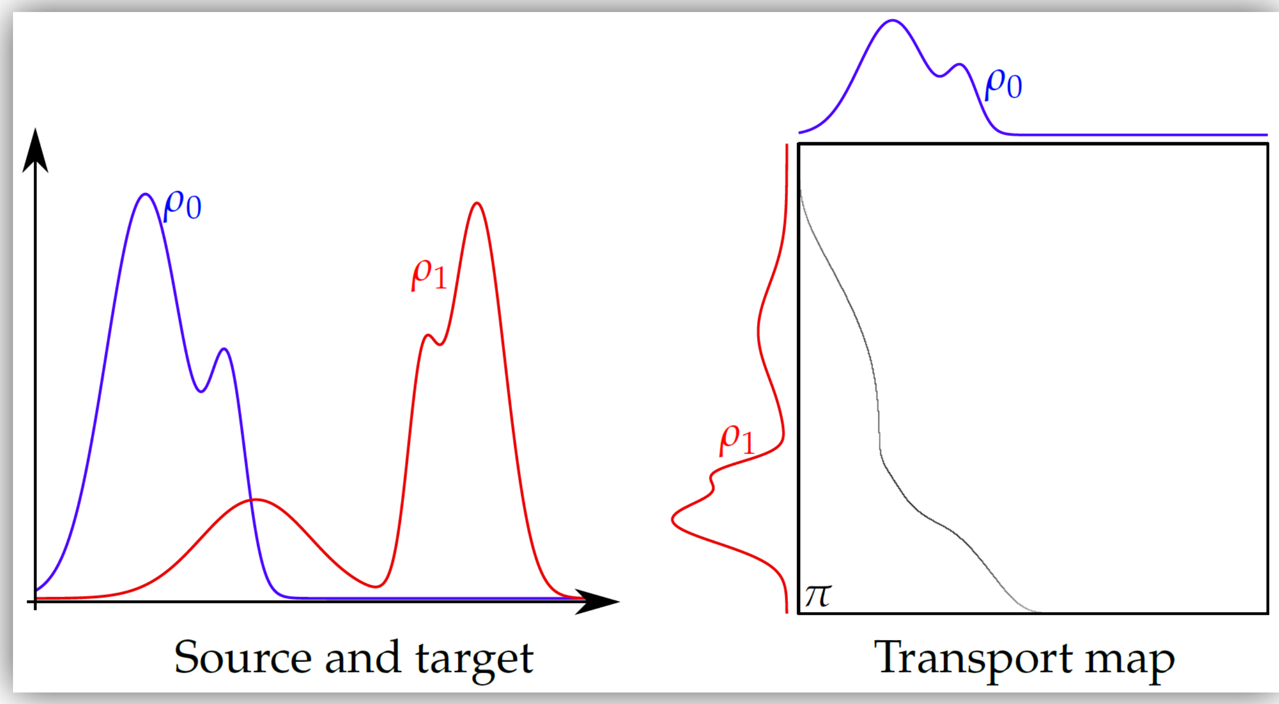
Must scoop everything up

$$\int_{\mathbb{R}} \pi(x, y) \, dx = \rho_1(y) \ \forall y \in \mathbb{R}$$

Must cover the target

1-Wasserstein in 1D

$$\mathcal{W}_1(\rho_0,\rho_1) := \left\{ \begin{array}{rll} \min_\pi & \iint_{\mathbb{R}\times\mathbb{R}} \pi(x,y)|x-y|\,dx\,dy & \text{Minimize total work} \\ \text{s.t.} & \pi \geq 0\ \forall x,y \in \mathbb{R} & \text{Nonnegative mass} \\ & \int_{\mathbb{R}} \pi(x,y)\,dy = \rho_0(x)\ \forall x \in \mathbb{R} & \text{Starts from } \rho_0 \\ & \int_{\mathbb{R}} \pi(x,y)\,dx = \rho_1(y)\ \forall y \in \mathbb{R} & \text{Ends at } \rho_1 \end{array} \right.$$



p-Wasserstein in 1D

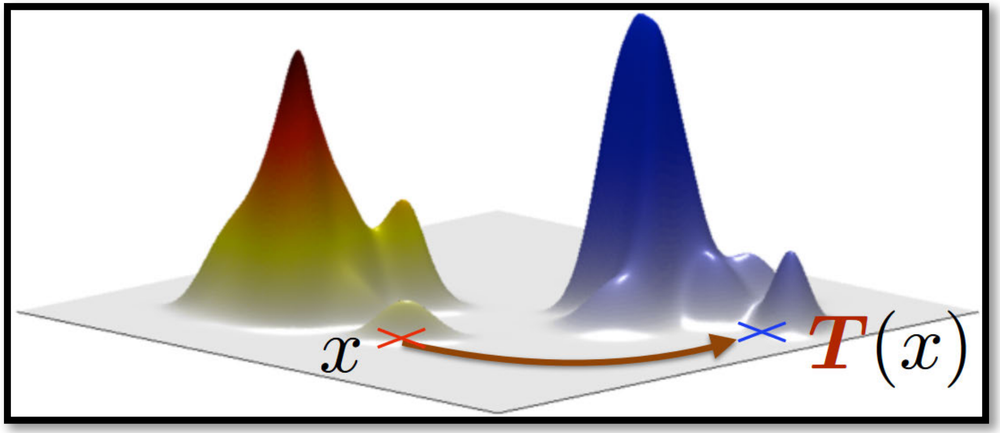
Kantorovich problem:
Replace with $c(x, y)$.

$$[\mathcal{W}_p(\rho_0, \rho_1)]^p := \begin{cases} \min_\pi & \iint_{\mathbb{R}\times\mathbb{R}} \pi(x,y)|x-y|^p \, dx \, dy \\ \text{s.t.} & \pi \geq 0 \; \forall x, y \in \mathbb{R} \\ & \int_{\mathbb{R}} \pi(x,y) \, dy = \rho_0(x) \; \forall x \in \mathbb{R} \\ & \int_{\mathbb{R}} \pi(x,y) \, dx = \rho_1(y) \; \forall y \in \mathbb{R} \end{cases}$$

Triangle inequality when $p \geq 1$.

Monge Formulation

$$\inf_{\phi_\sharp \rho_0 = \rho_1} \int_{-\infty}^{\infty} c(x, \phi(x)) \rho_0(x)\, dx$$



Not always well-posed!

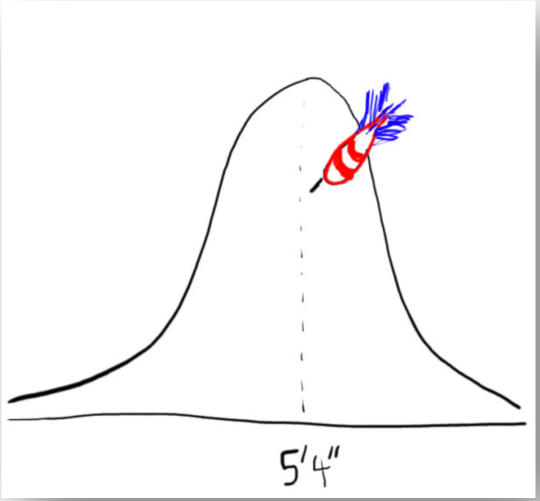
Q:

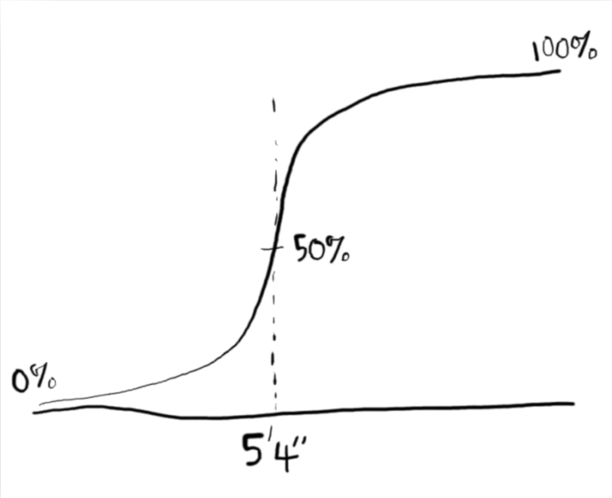
When is transport **computable**?

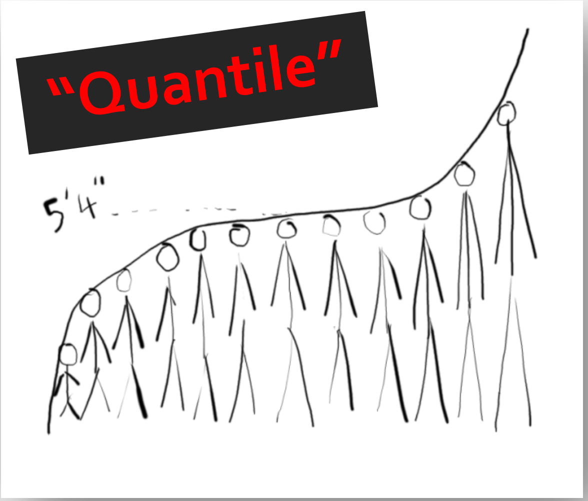
Needed: Finite number of unknowns.

In One Dimension: Closed-Form

http://realgl.blogspot.com/2013/01/pdf-cdf-inv-cdf.html







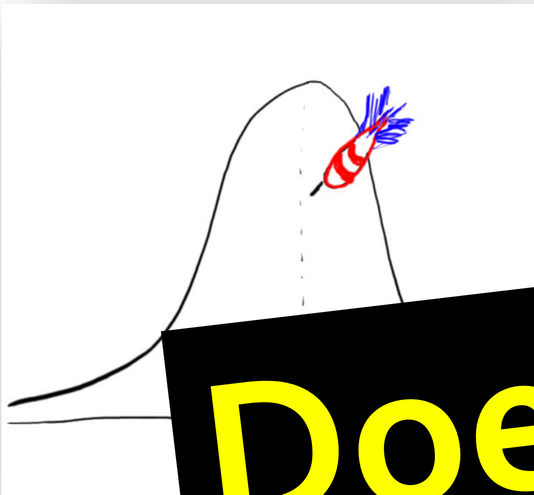
PDF → **[CDF]** → **CDF^{-1}**

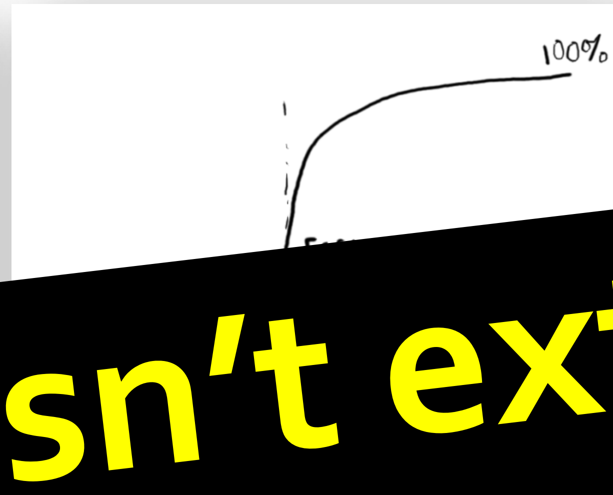
$$\mathcal{W}_1(\mu,\nu)=\int_{-\infty}^{\infty}|\mathrm{CDF}(\mu)-\mathrm{CDF}(\nu)|\,d\ell$$

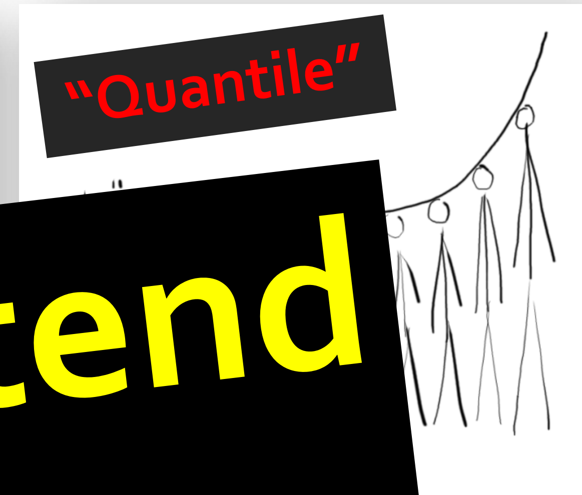
$$\mathcal{W}_2^2(\mu,\nu)=\int_{-\infty}^{\infty}\left(\mathrm{CDF}^{-1}(\mu)-\mathrm{CDF}^{-1}(\nu)\right)^2\,d\ell$$

In One Dimension: Closed-Form

http://realgl.blogspot.com/2013/01/pdf-cdf-inv-cdf.html







Doesn't extend past 1D!

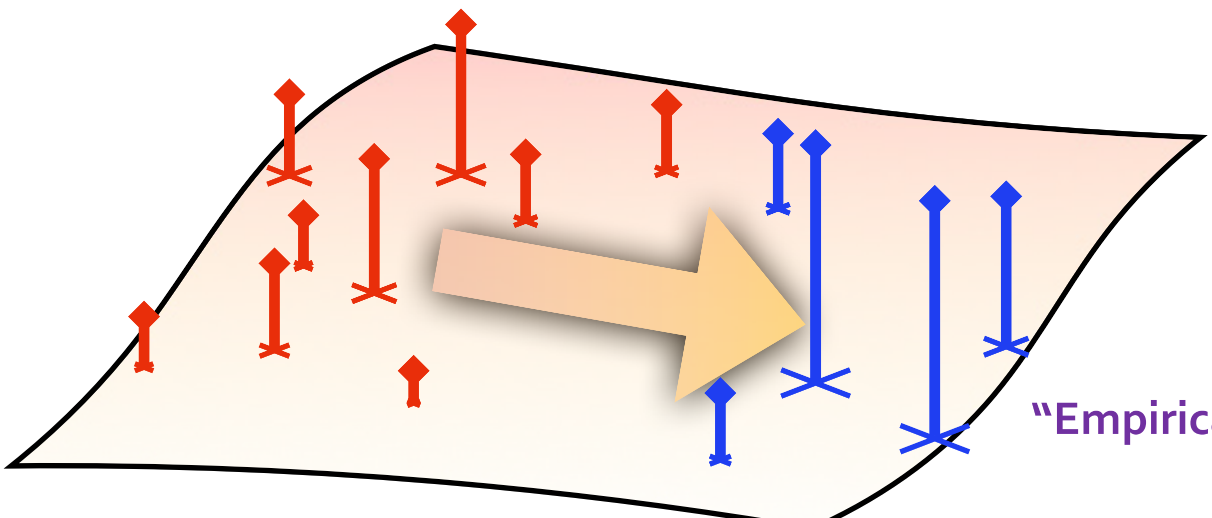
$$\int_{-\infty} |\mathrm{CDF}(\mu) - \mathrm{CDF}(\nu)|\, d\ell$$

$$\mathcal{W}_2^2(\mu,\nu) = \int_{-\infty}^{\infty} \left(\mathrm{CDF}^{-1}(\mu) - \mathrm{CDF}^{-1}(\nu)\right)^2 \, d\ell$$

Fully-Discrete Transport

$$\delta_x(S) := \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_0 := \sum_{i=1}^{k_0} a_{0i}\delta_{x_{0i}} \qquad \mu_0 := \sum_{i=1}^{k_1} a_{1i}\delta_{x_{1i}}$$



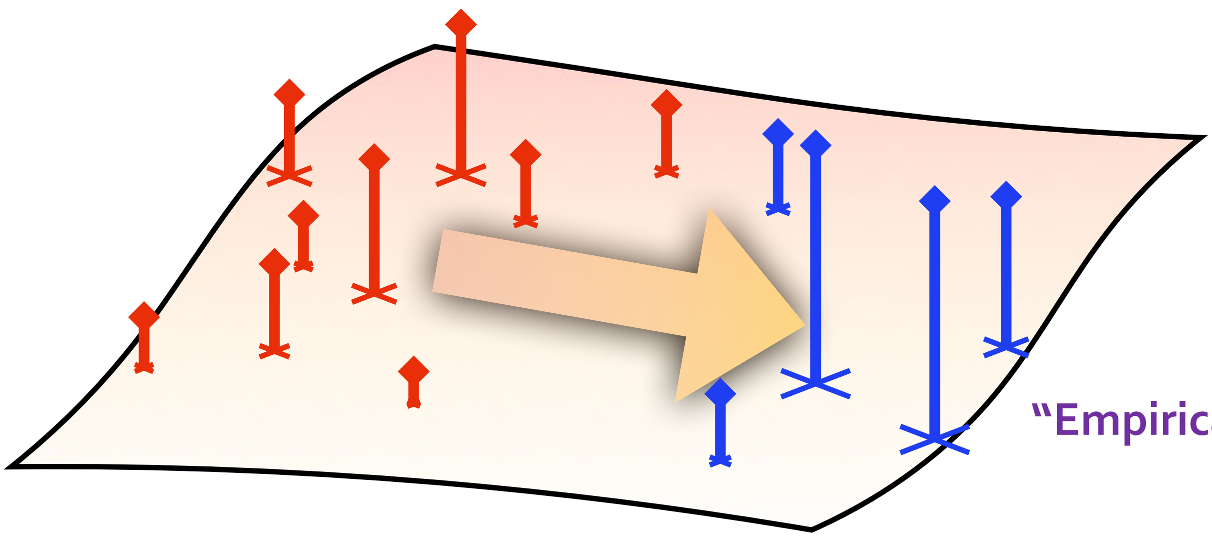
"Empirical measures"

Fully-Discrete Transport

$$[\mathcal{W}_p(\mu_0,\mu_1)]^p = \begin{cases} \min_{T\in\mathbb{R}^{k_0\times k_1}} & \sum_{ij} T_{ij}|x_{0i}-x_{1j}|^p \\ \text{s.t.} & T \geq 0 \\ & \sum_j T_{ij} = a_{0i} \\ & \sum_i T_{ij} = a_{1j} \end{cases}$$

Linear program: Finite number of variables

Algorithms: Simplex, interior point, auction, …



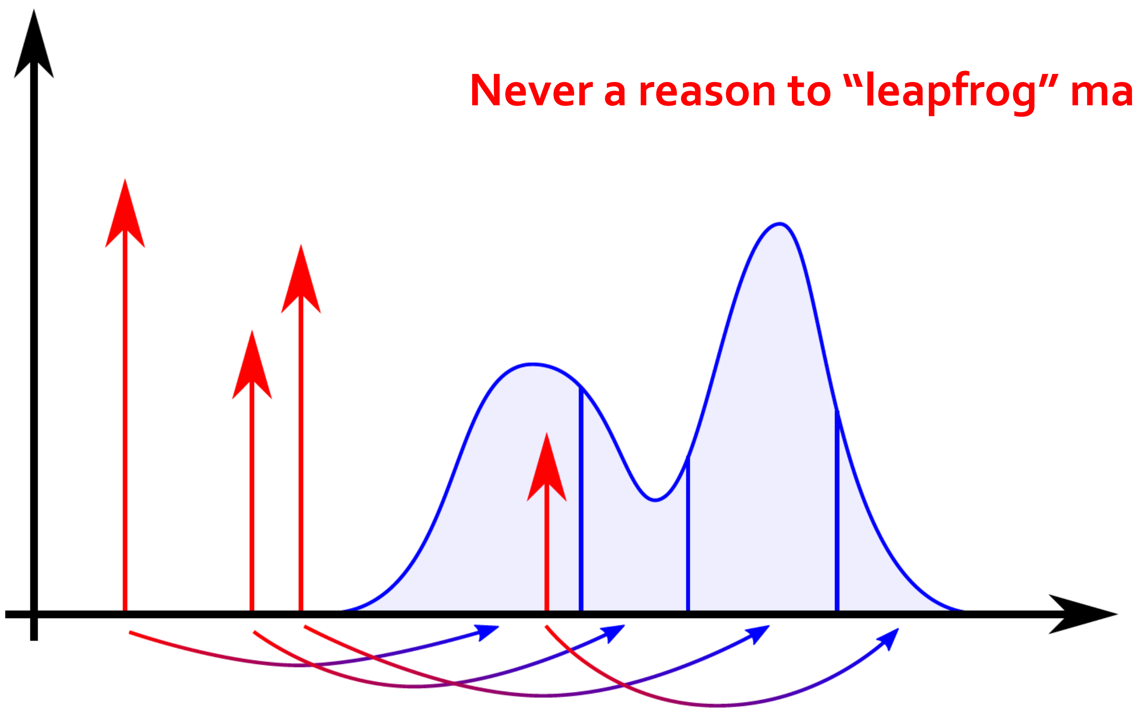
"Empirical measures"

Semidiscrete Transport

$$\mu_0 := \sum_{i=1}^{k_0} a_{0i}\delta_{x_{0i}}$$

$$\mu_1(S) := \int_S \rho_1(x)\,dx$$

Never a reason to "leapfrog" mass!



Plan For Today

1. Introduction to optimal transport
   - 1D examples
   - Many formulas

2. Applications

3. Discrete/discretized transport
   - Entropic regularization
   - Eulerian transport
   - Semidiscrete transport

4. Extensions

More General Formulation

Monge-Kantorovich Problem

Probability Measure

$$\mu(X) = 1$$

$$\mu(S \subseteq X) \in [0, 1]$$

$$\mu\left(\cup_{i \in I} E_i\right) = \sum_{i \in I} \mu(E_i)$$

when E_i disjoint,

I countable

X is the domain

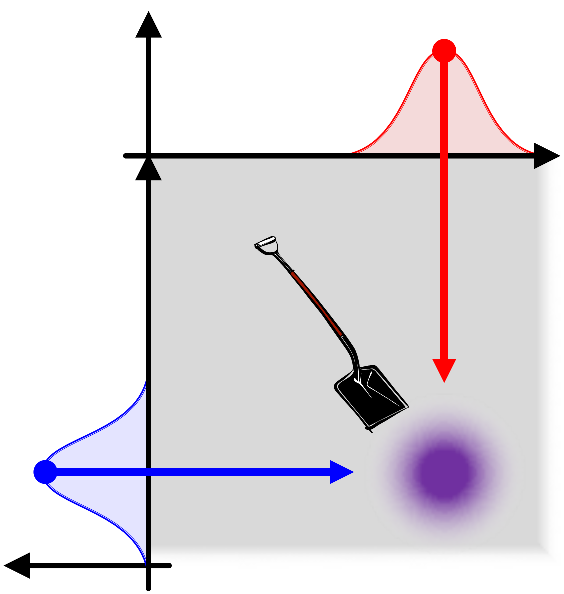
"Prob(X)"

Function from sets to probability

Measure Coupling

$$\mu \in \mathrm{Prob}(X), \nu \in \mathrm{Prob}(Y)$$

$$\Pi(\mu,\nu) := \left\{ \pi \in \mathrm{Prob}(X \times Y) : \begin{pmatrix} \pi(A \times Y) = \mu(A) \\ \pi(X \times B) = \nu(B) \end{pmatrix} \right\}$$



Analog of transportation matrix

Kantorovich Problem

$$\mathrm{OT}(\mu, \nu; c) := \min_{\pi \in \Pi(\mu,\nu)} \iint_{X \times Y} c(x, y)\, d\pi(x, y)$$

General transport problem!

Example: Discrete Transport

$$X = \{1, 2, \ldots, k_1\}, Y = \{1, 2, \ldots, k_2\}$$

"Earth Mover's Distance"

$$\mathrm{OT}(v, w; C) = \begin{cases} \min_{T \in \mathbb{R}^{k_1 \times k_2}} & \sum_{ij} T_{ij} c_{ij} \\ \text{s.t.} & T \geq 0 \\ & \sum_j T_{ij} = v_i \ \forall i \in \{1, \ldots, k_1\} \\ & \sum_i T_{ij} = w_j \ \forall j \in \{1, \ldots, k_2\}. \end{cases}$$

Metric when d(x,y) satisfies the triangle inequality.

"The Earth Mover's Distance as a Metric for Image Retrieval"
Rubner, Tomasi, and Guibas; IJCV 40.2 (2000): 99—121.

Revised in:

"Ground Metric Learning"
Cuturi and Avis; JMLR 15 (2014)

p-Wasserstein Distance

$$\mathcal{W}_p(\mu,\nu) \equiv \min_{\pi \in \Pi(\mu,\nu)} \left(\iint_{X \times X} d(x,y)^p \, d\pi(x,y) \right)^{1/p}$$

Shortest path distance

Expectation

Geodesic distance d(x,y)

http://www.sciencedirect.com/science/article/pii/S152407031200029X#

Kantorovich Duality

$$\mathrm{OT}(\mu,\nu;c) := \begin{cases} \min_\pi \ \iint_{X\times Y} c(x,y)\, d\pi(x,y) \\ \quad \text{s.t.} \ \ \pi \in \Pi(\mu,\nu) \end{cases}$$

Primal

$$= \begin{cases} \max_{\phi,\psi} \ \int_X \phi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) \\ \qquad \text{s.t.} \ \ \phi(x) + \psi(y) \le c(x,y) \text{ for a.e. } x \in X, y \in Y \end{cases}$$

Dual

On the board (time-permitting):
Motivation for discrete duality

Flow-Based W_2

$$\mathcal{W}_2^2(\rho_0,\rho_1)=\begin{cases}\inf_{\rho,v} \iint_{M\times[0,1]} \frac{1}{2}\rho(x,t)\|v(x,t)\|^2\,dx\,dt \\ \quad \text{s.t. } \nabla\cdot(\rho(x,t)v(x,t)) = \frac{\partial\rho(x,t)}{dt} \\ \quad\quad v(x,t)\cdot\hat{n}(x) = 0\ \forall x\in\partial M \\ \quad\quad \rho(x,0) = \rho_0(x) \\ \quad\quad \rho(x,1) = \rho_1(x)\end{cases}$$

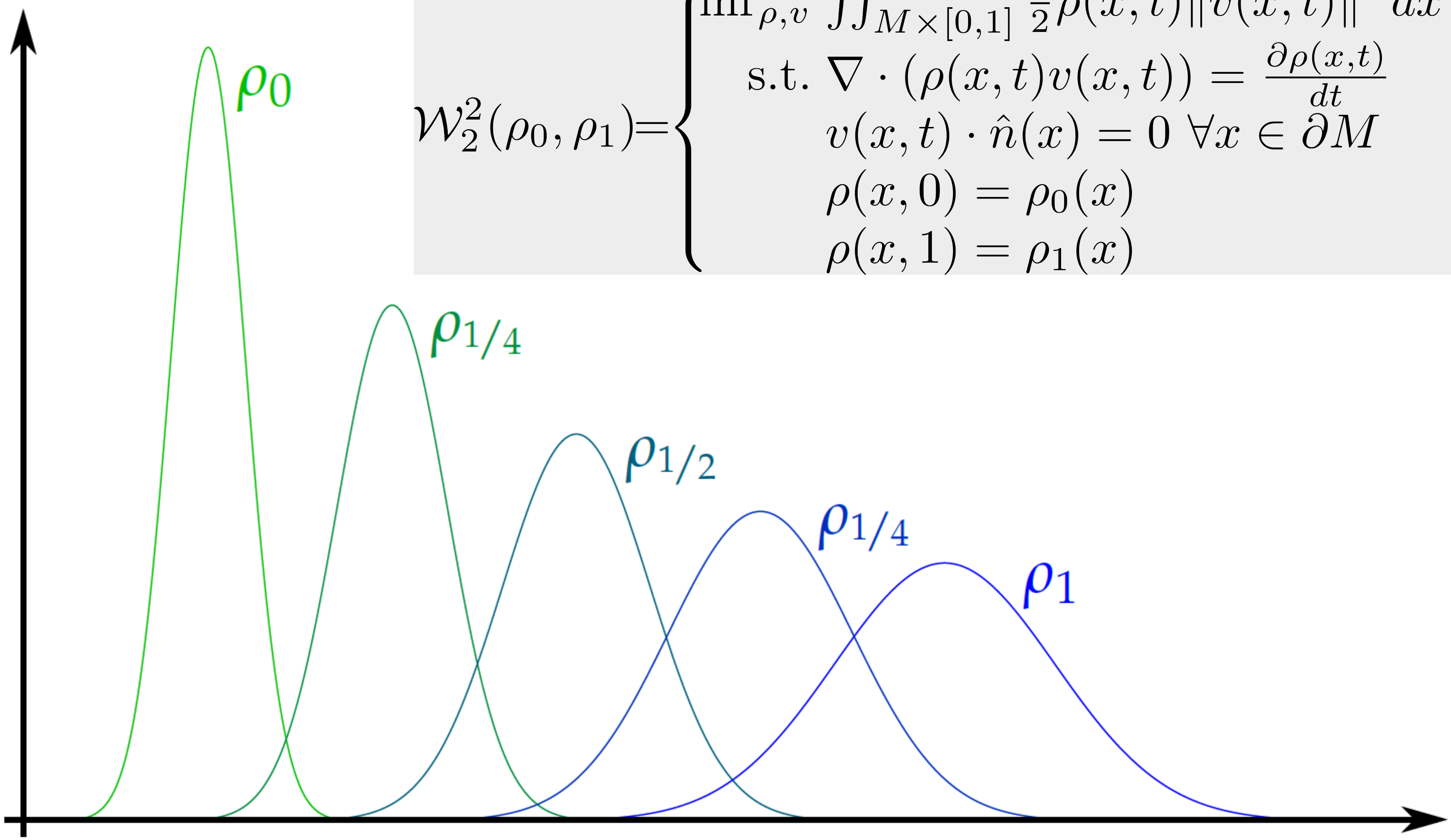
Benamou & Brenier

"A computational fluid mechanics solution of the Monge-Kantorovich mass transfer problem"

Numer. Math. 84 (2000), pp. 375-393

Displacement Interpolation

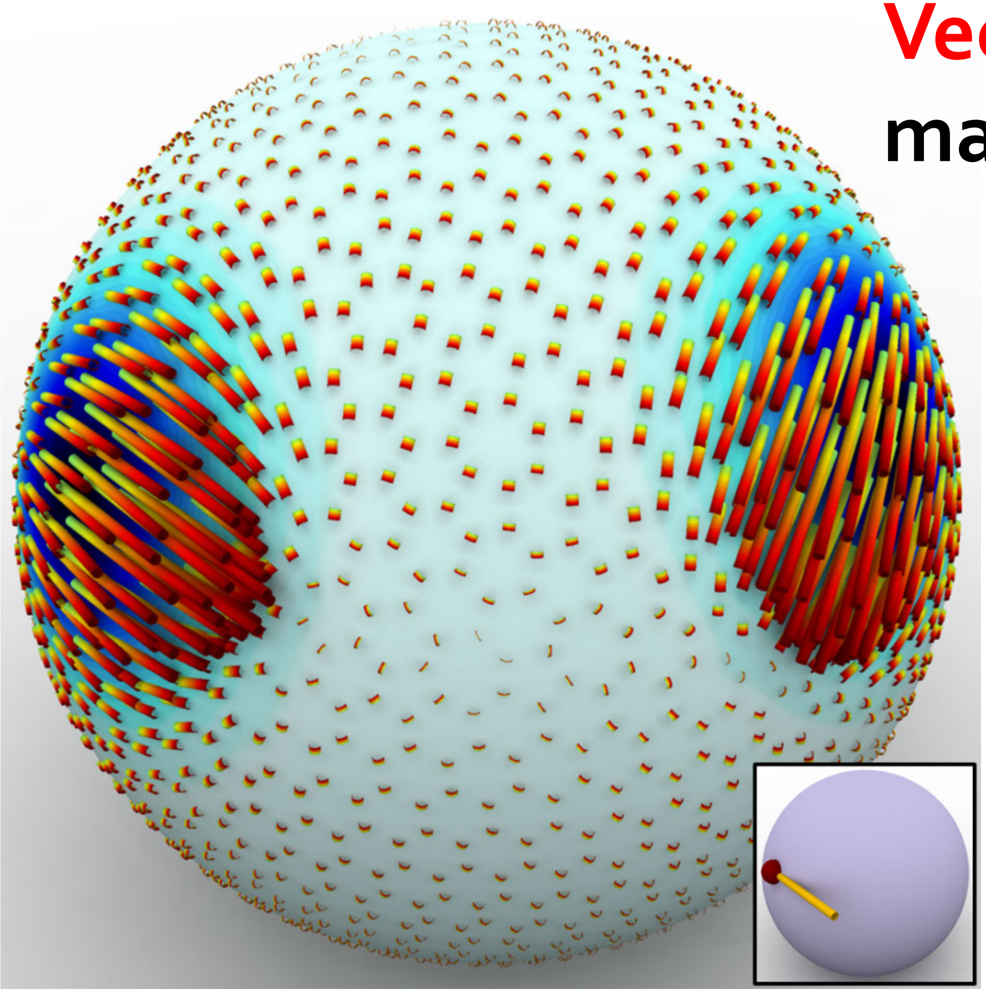
$$\mathcal{W}_2^2(\rho_0,\rho_1) = \begin{cases} \inf_{\rho,v} \iint_{M\times[0,1]} \frac{1}{2}\rho(x,t)\|v(x,t)\|^2\,dx\,dt \\ \quad \text{s.t. } \nabla\cdot(\rho(x,t)v(x,t)) = \frac{\partial\rho(x,t)}{dt} \\ \quad\quad v(x,t)\cdot\hat{n}(x) = 0\ \forall x\in\partial M \\ \quad\quad \rho(x,0) = \rho_0(x) \\ \quad\quad \rho(x,1) = \rho_1(x) \end{cases}$$



Interpretation

- Consider set of distributions as an **infinite-dimensional manifold**
- **Tangent spaces** from advection
- **Geodesics** from displacement interpolation

“Tangent Space”



Vector field moving mass around

Continuity equation:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \cdot v)$$

Fun fact:
v is curl-free!

$$v = \nabla \psi$$

Image from [Solomon, Guibas, & Butscher 2013]

Plan For Today

1. Introduction to optimal transport
   - 1D examples
   - Many formulas

2. Applications

3. Discrete/discretized transport
   - Entropic regularization
   - Eulerian transport
   - Semidiscrete transport

4. Extensions

Wassersteinization

[wos-ur-stahyn-ahy-sey-sh*uh*-n]
noun.

Introduction of optimal transport into a computational problem.

cf. least-squarification, L_1ification, deep-netification, kernelization

Key Ingredients

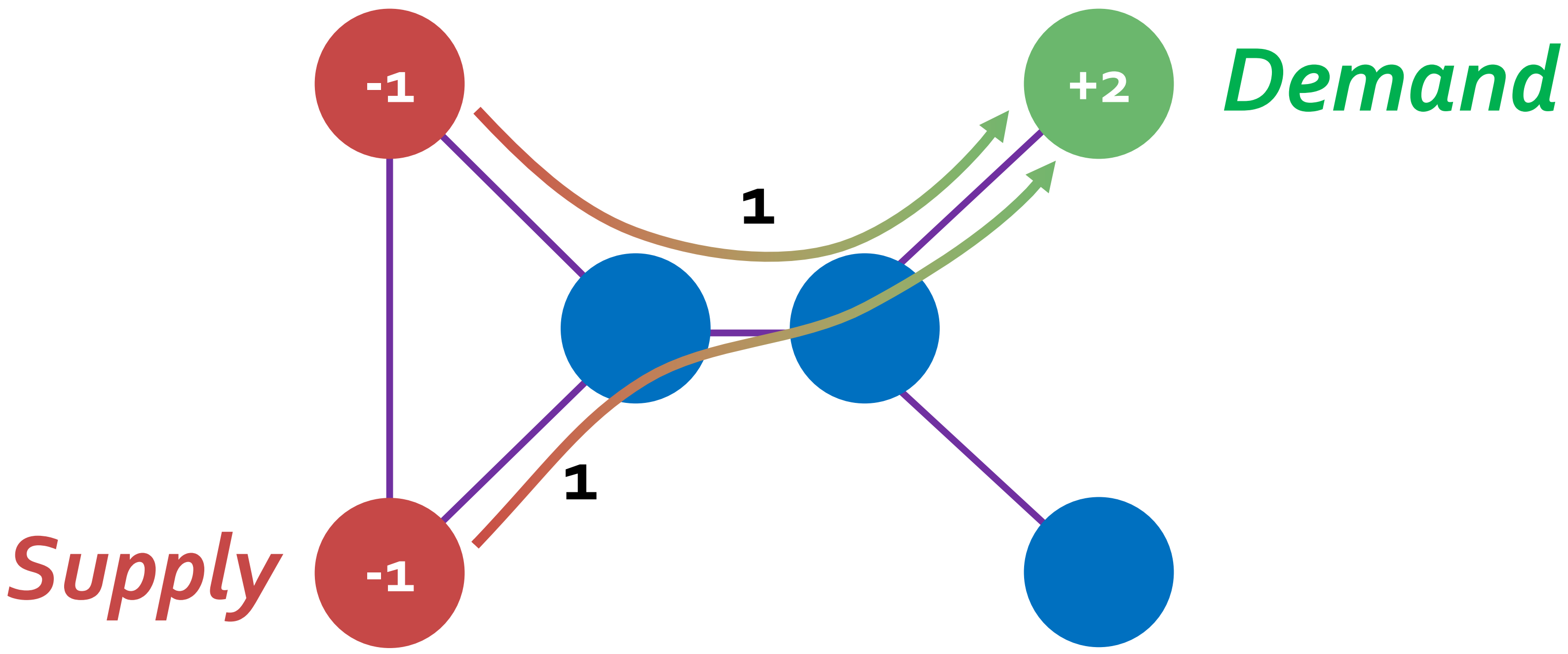
We have tools to

- Solve optimal transport problems numerically
- Differentiate transport distances in terms of their input distributions

Bonus:

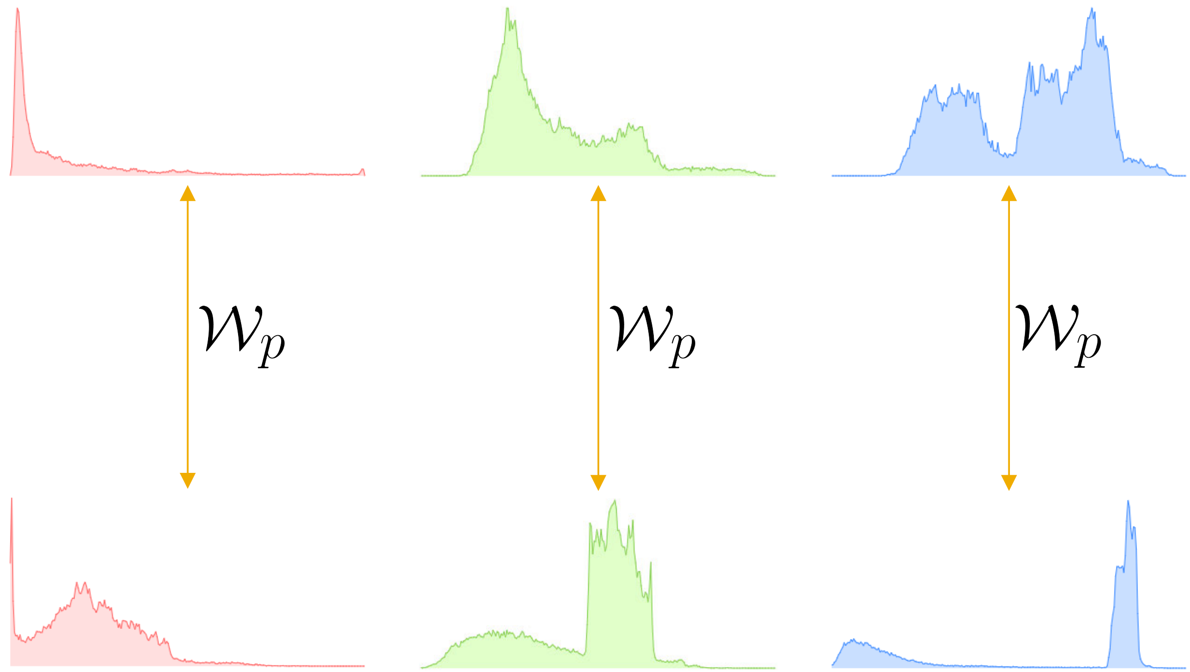
Transport cost from μ to ν is a convex function of μ and ν.

Operations and Logistics



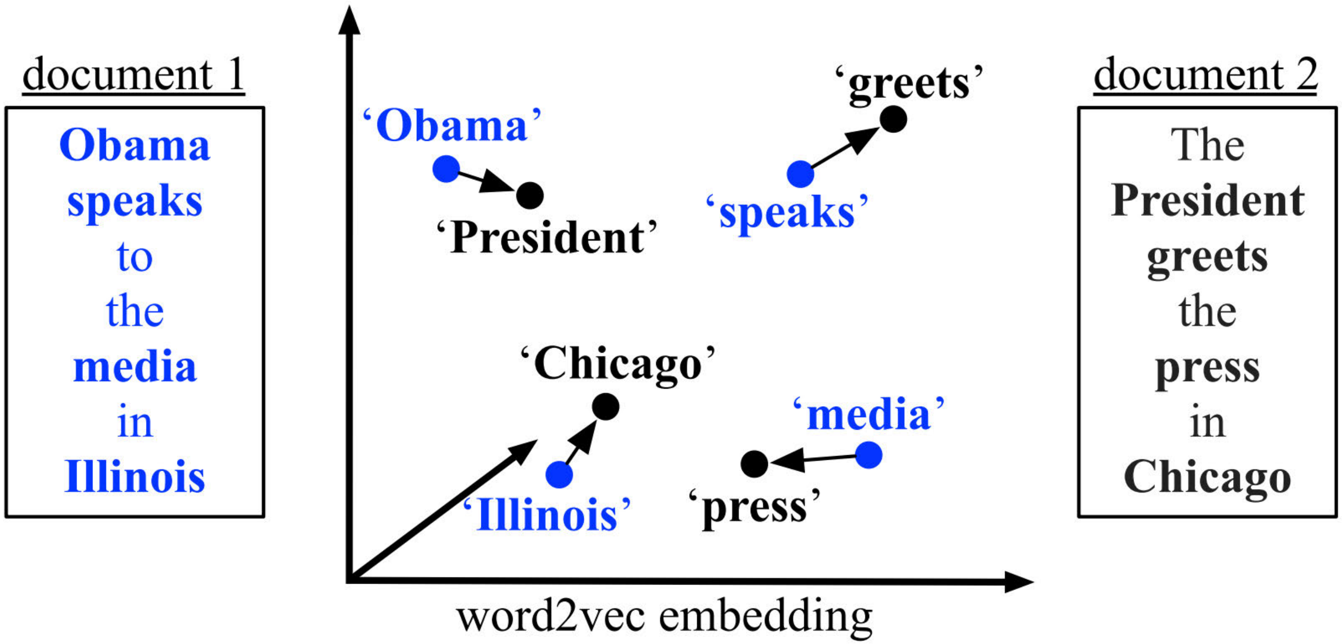
Minimum-cost flow

Histograms and Descriptors



[Rubner, Tomasi, & Guibas 2000]

Histograms and Descriptors

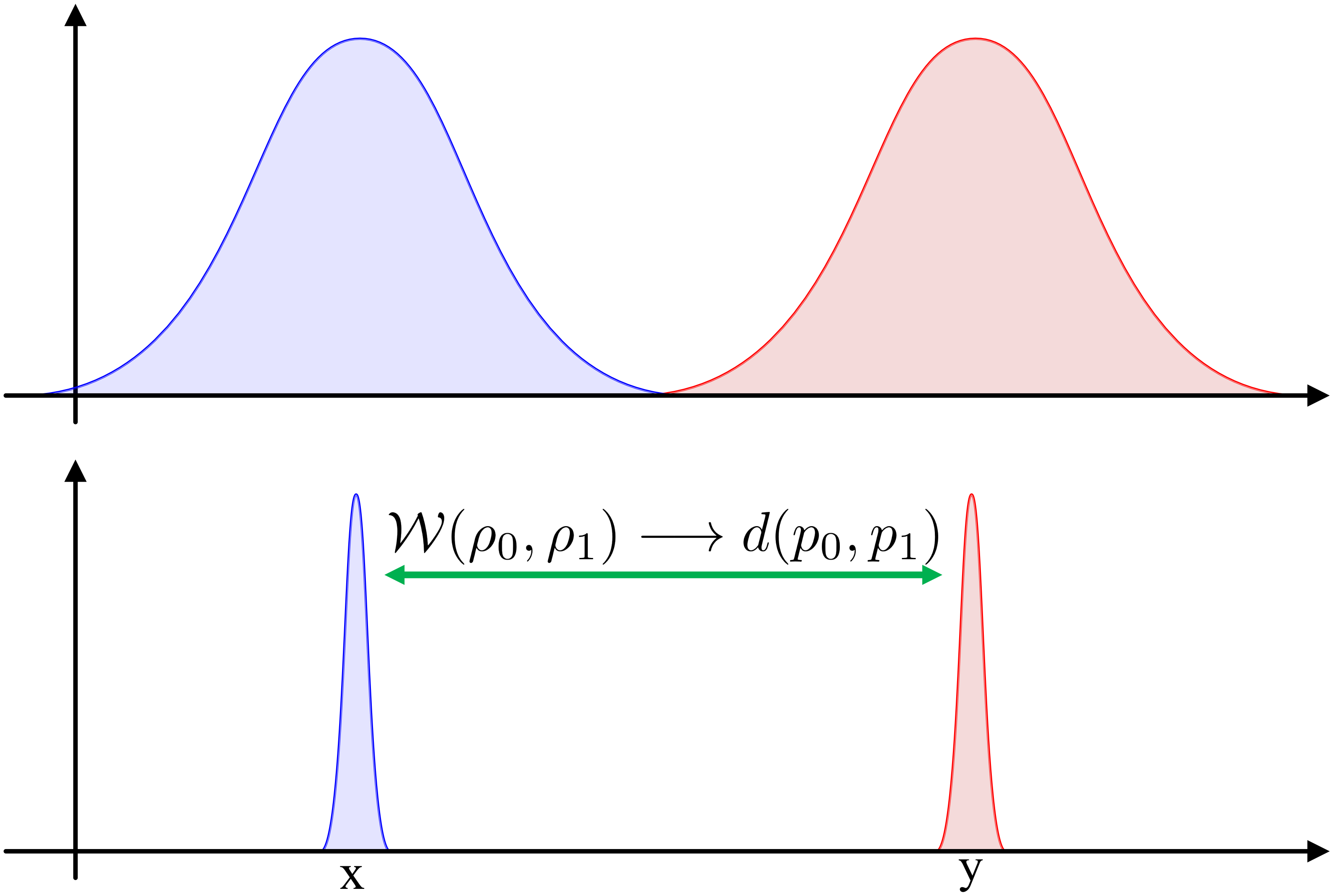


Use deep network embedding

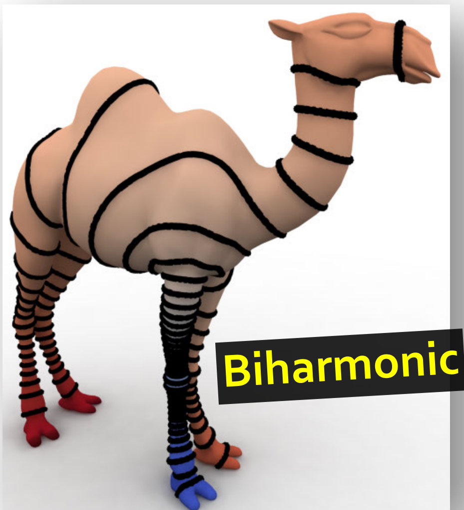
[Kusner et al. 2015]

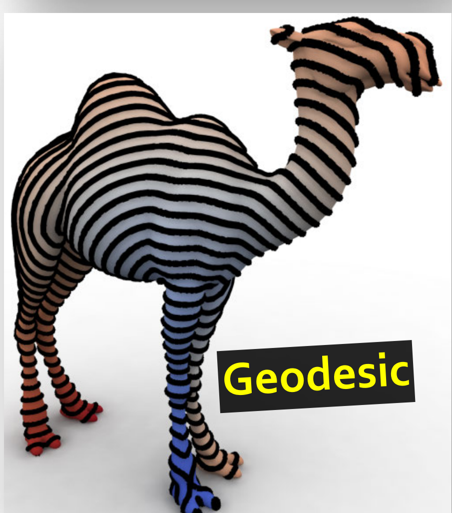
Word Mover's Distance (WMD)

Pointwise Distance

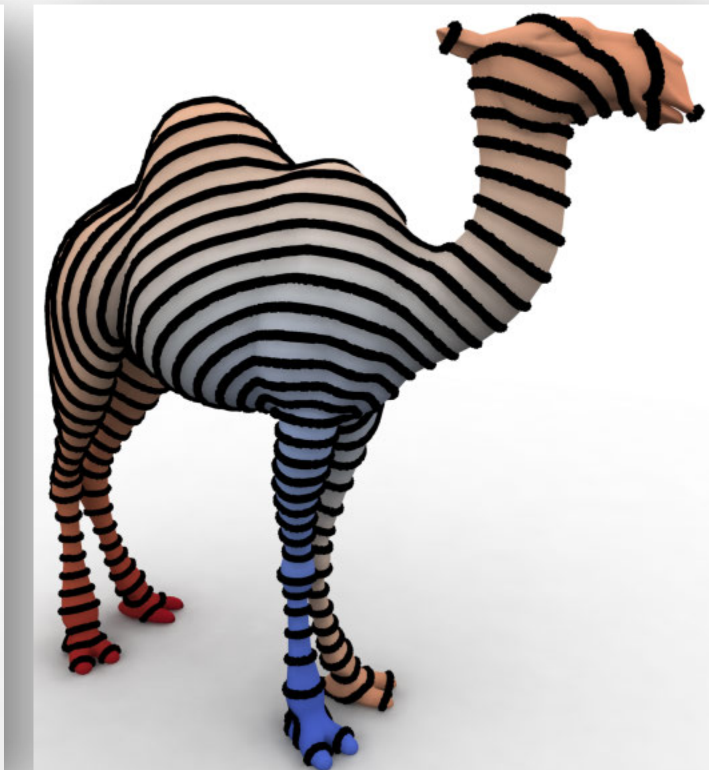


Distance Approximation





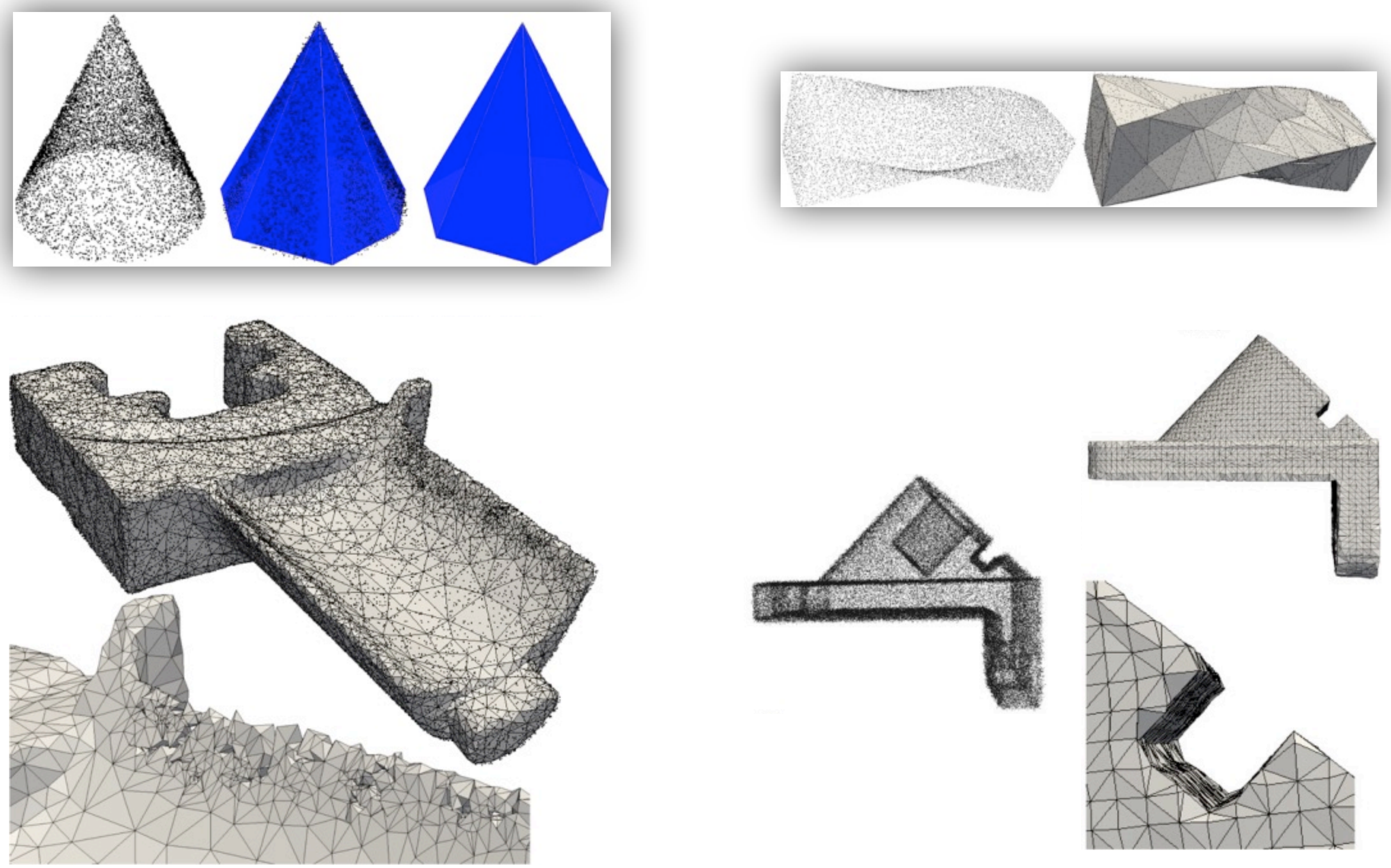
0 eigenfunctions



100 eigenfunctions

Proposition: Satisfies triangle inequality.

Registration and Reconstruction



[Digne et al. 2014]

Distance from point cloud to mesh

Blue Noise and Stippling

$$\min_{x_1,\ldots,x_n} \mathcal{W}_2^2\left(\mu, \frac{1}{n}\sum_i \delta_{x_i}\right)$$

Image courtesy F. de Goes; photo by F. Durand

Political Redistricting (?)

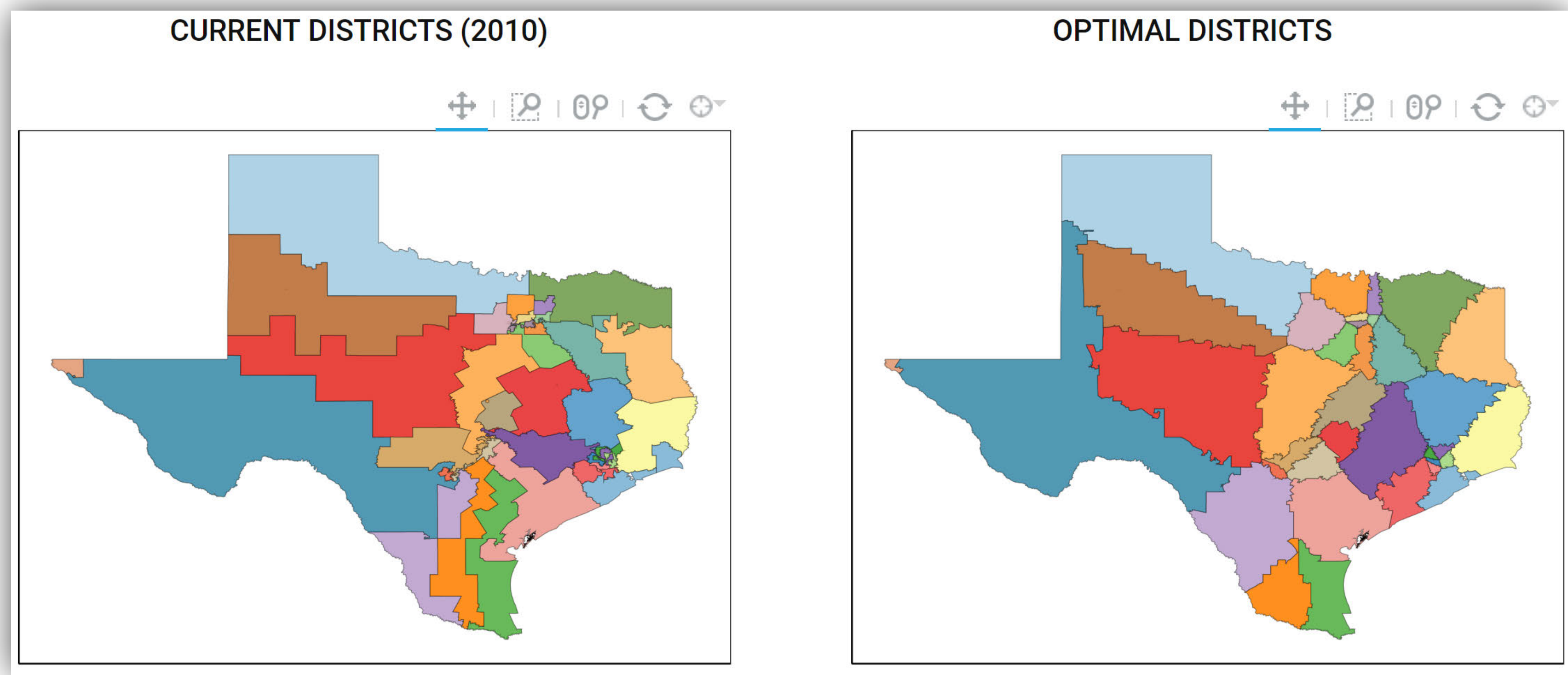
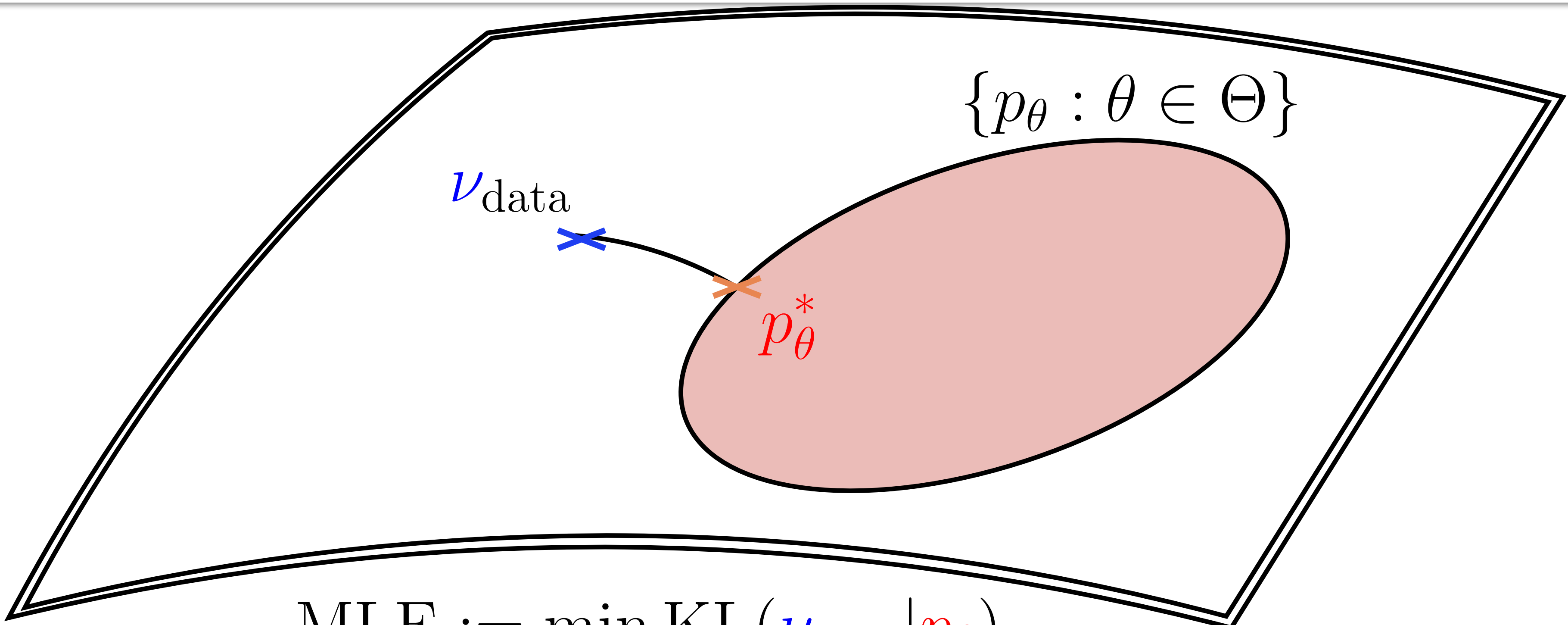


Image from optimaldistricts.org

Statistical Estimation



$$\text{MLE} := \min_{\theta \in \Theta} \text{KL}(\nu_{\text{data}} | p_\theta)$$

$$\longrightarrow \text{MKE} := \min_{\theta \in \Theta} \mathcal{W}_2(\nu_{\text{data}}, p_\theta)$$

[Bassetti 2006]

Minimum Kantorovich Estimator

Distributionally Robust Optimization

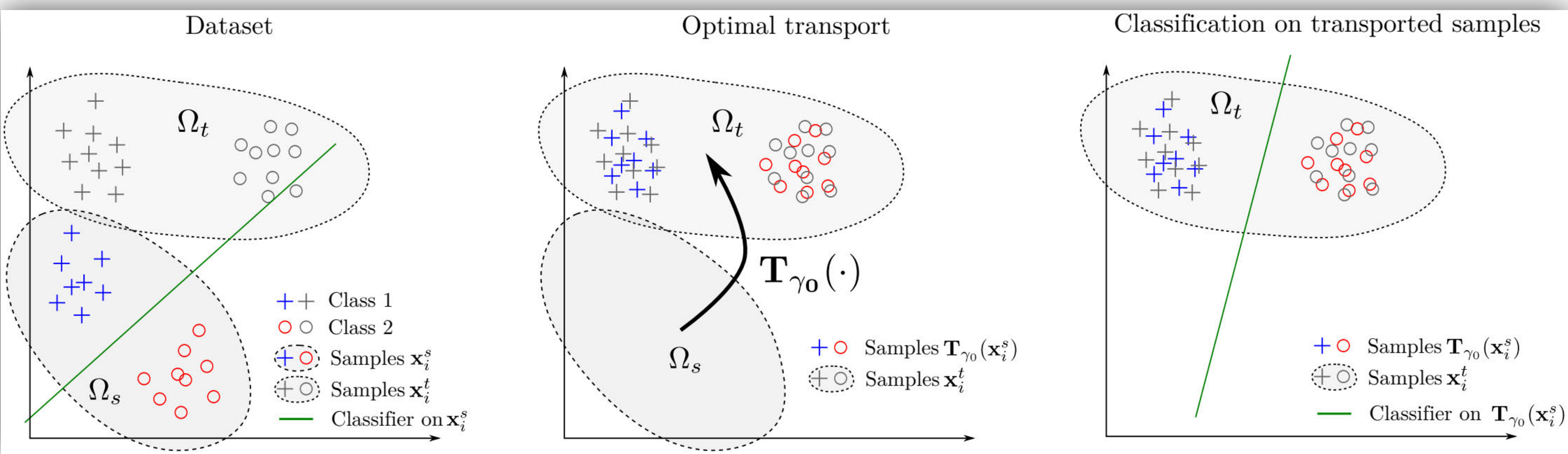
$$\inf_{x \in \mathbb{X}} \sup_{\mathbb{Q} \in \hat{\mathcal{P}}_N} \mathbb{E}_{\xi \sim \mathbb{Q}}[h(x, \xi)]$$

Wasserstein ball around empirical distribution

Loss function

[Esvahani & Kuhn 2017]

Domain Adaptation



1. Estimate transport map
2. Transport labeled samples to new domain
3. Train classifier on transported labeled samples

[Courty et al. 2017]

Engineering Design

EPFL Computer Graphics and Geometry Laboratory; Rayform SA

Plan For Today

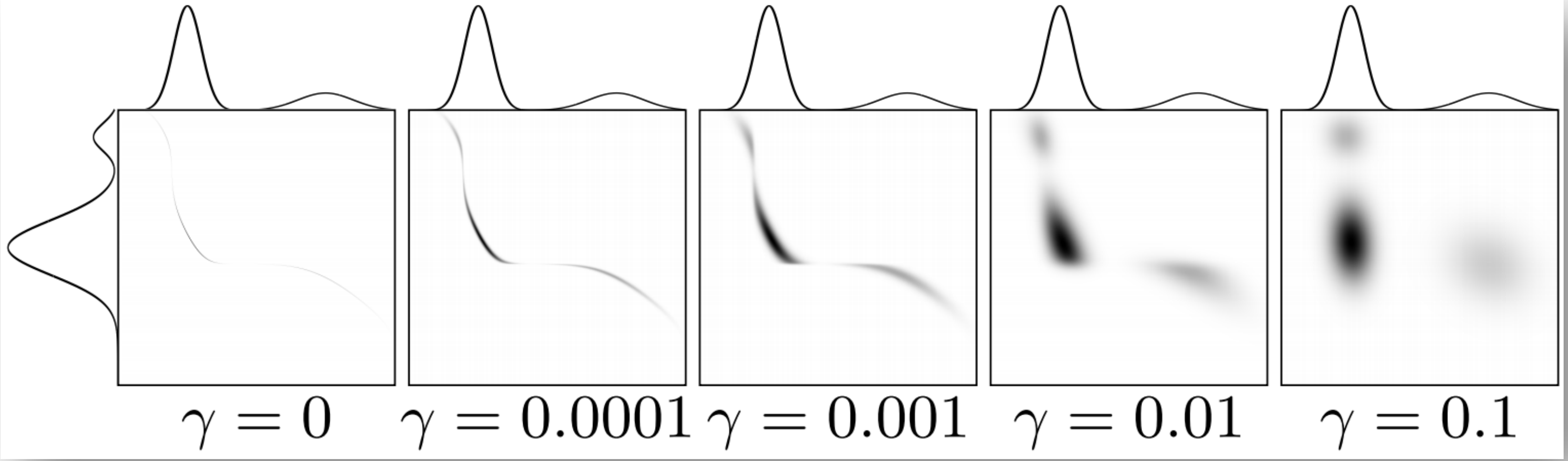
1. Introduction to optimal transport
    - 1D examples
    - Many formulas

2. Applications

3. Discrete/discretized transport
    - Entropic regularization
    - Eulerian transport
    - Semidiscrete transport

4. Extensions

Entropic Regularization



$$\begin{aligned} \min_T \quad & \textstyle\sum_{ij} T_{ij} c_{ij} - \alpha H(T) \\ \text{s.t.} \quad & \textstyle\sum_j T_{ij} = v_i \\ & \textstyle\sum_i T_{ij} = w_j \end{aligned}$$

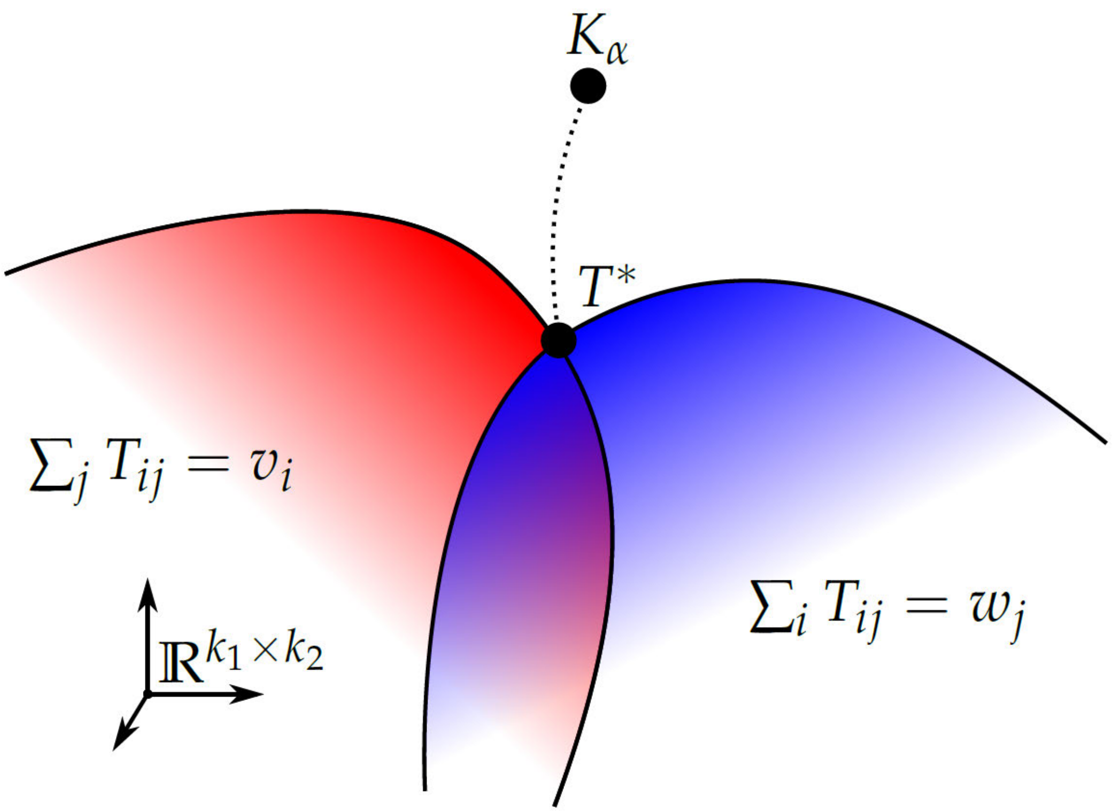
OK to drop nonnegative constraint!

$$H(T) := -\sum_{ij} T_{ij} \log T_{ij}$$

Cuturi. "Sinkhorn distances: Lightspeed computation of optimal transport" (NIPS 2013)

Interpretation as Projection

$$\sum_{ij} T_{ij}c_{ij} - \alpha H(T) = \mathrm{KL}(T|K_\alpha) \text{ where } K_\alpha := \exp(-C_\alpha)$$



Key Lemma

Prove on the board (time-permitting):

$$T = \operatorname{diag}(p) K_\alpha \operatorname{diag}(q),$$

$$\text{where } K_\alpha := \exp(-C/\alpha)$$

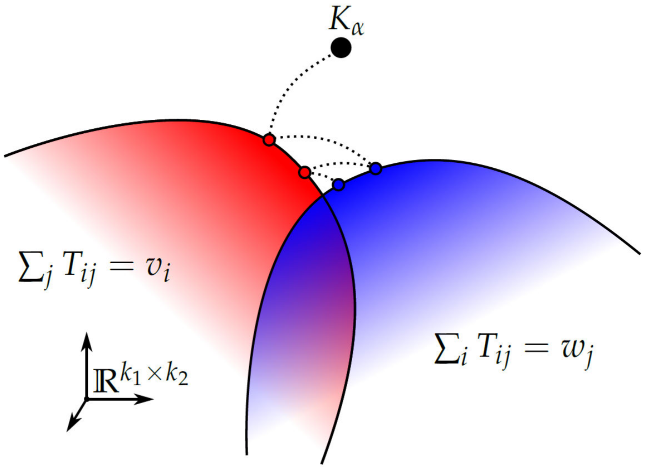
$$\begin{aligned} \min_T \quad & \textstyle\sum_{ij} T_{ij} c_{ij} - \alpha H(T) \\ \text{s.t.} \quad & \textstyle\sum_j T_{ij} = v_i \\ & \textstyle\sum_i T_{ij} = w_j \end{aligned}$$

$$H(T) := -\sum_{ij} T_{ij} \log T_{ij}$$

Sinkhorn Algorithm

$$T = \mathrm{diag}(p) K_\alpha \mathrm{diag}(q),$$
$$\text{where } K_\alpha := \exp(-C/\alpha)$$

$$p \leftarrow v \oslash (K_\alpha q)$$
$$q \leftarrow w \oslash (K_\alpha^\top p)$$



Sinkhorn & Knopp. "Concerning nonnegative matrices and doubly stochastic matrices". Pacific J. Math. 21, 343–348 (1967).

Alternating projection

Ingredients for Sinkhorn

1. Supply vector p
2. Demand vector q
3. Multiplication by K

$$K_{ij} = e^{-c_{ij}/\alpha}$$

W2 distance: Fast *K* Product

$$(Kv)_{ij} = \sum_{k\ell} g_\sigma(\|(i,j) - (k,\ell)\|_2) v_{k\ell}$$

Fish image from borisfx.com

Gaussian convolution

Sinkhorn on a Grid

$$p \leftarrow v \oslash (K_\alpha q)$$

$$q \leftarrow w \oslash (K_\alpha^\top p)$$

No need to store K_α

Sinkhorn on a Grid

$$p \leftarrow v \oslash (K_\alpha q)$$

$$q \leftarrow w \oslash (K_\alpha^\top p)$$

What about surfaces?

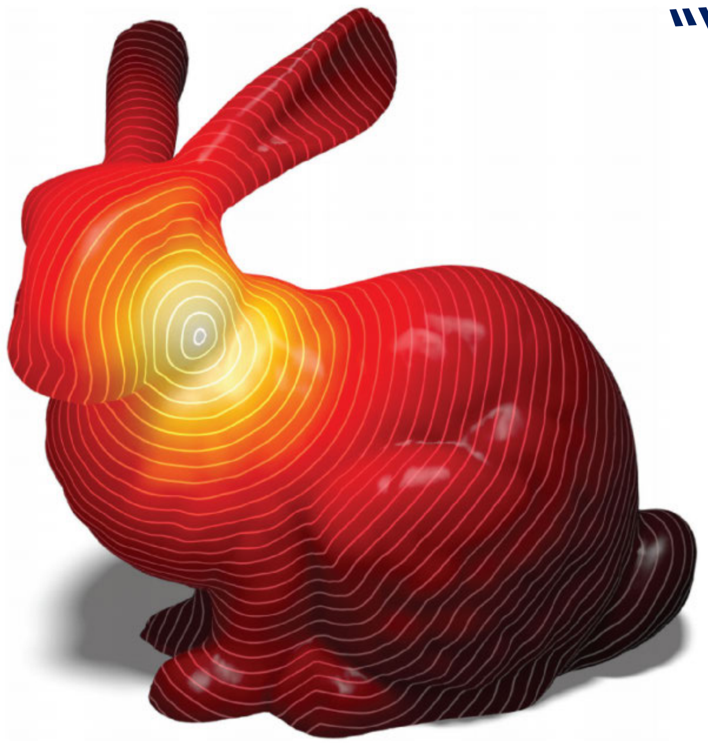
No need to store K_α

Recall:

Geodesic Distances

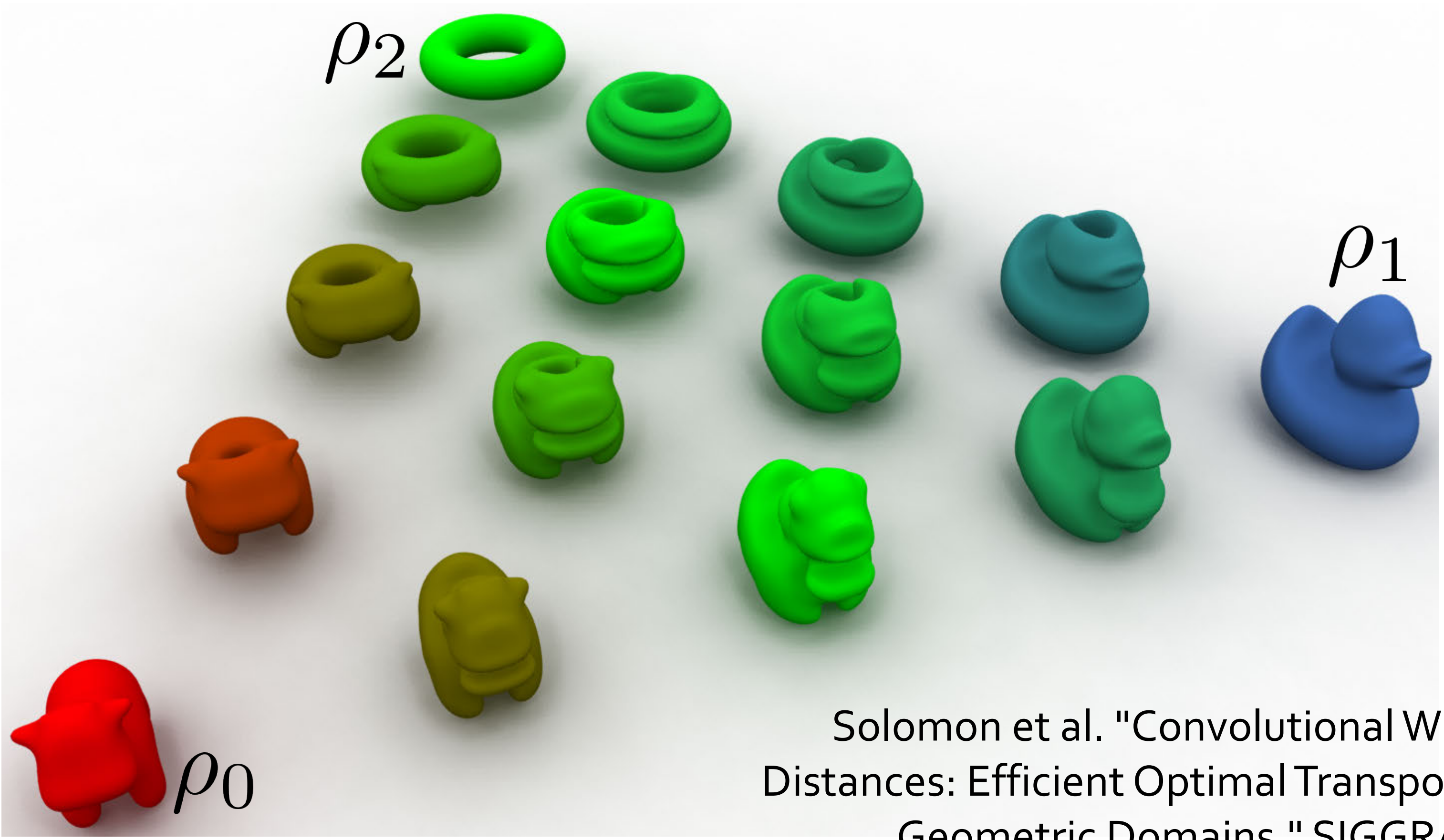
$$d_g(p,q) = \lim_{t\to 0} \sqrt{-4t \log k_{t,p}(q)}$$

"Varadhan's Theorem"



"Geodesics in heat"
Crane, Weischedel, and Wardetzky; TOG 2013

Approximate Sinkhorn



Solomon et al. "Convolutional Wasserstein Distances: Efficient Optimal Transportation on Geometric Domains." SIGGRAPH 2015.

Replace K_α with heat kernel

Plan For Today

1. Introduction to optimal transport
   - 1D examples
   - Many formulas

2. Applications

3. Discrete/discretized transport
   - Entropic regularization
   - Eulerian transport
   - Semidiscrete transport

4. Extensions

Recall:

Flow-Based W_2

$$\mathcal{W}_2^2(\rho_0,\rho_1)=\begin{cases}\inf_{\rho,v} \iint_{M\times[0,1]} \frac{1}{2}\rho(x,t)\|v(x,t)\|^2\,dx\,dt \\ \quad \text{s.t. } \nabla\cdot(\rho(x,t)v(x,t)) = \frac{\partial\rho(x,t)}{dt} \\ \qquad v(x,t)\cdot\hat{n}(x) = 0\ \forall x\in\partial M \\ \qquad \rho(x,0) = \rho_0(x) \\ \qquad \rho(x,1) = \rho_1(x)\end{cases}$$

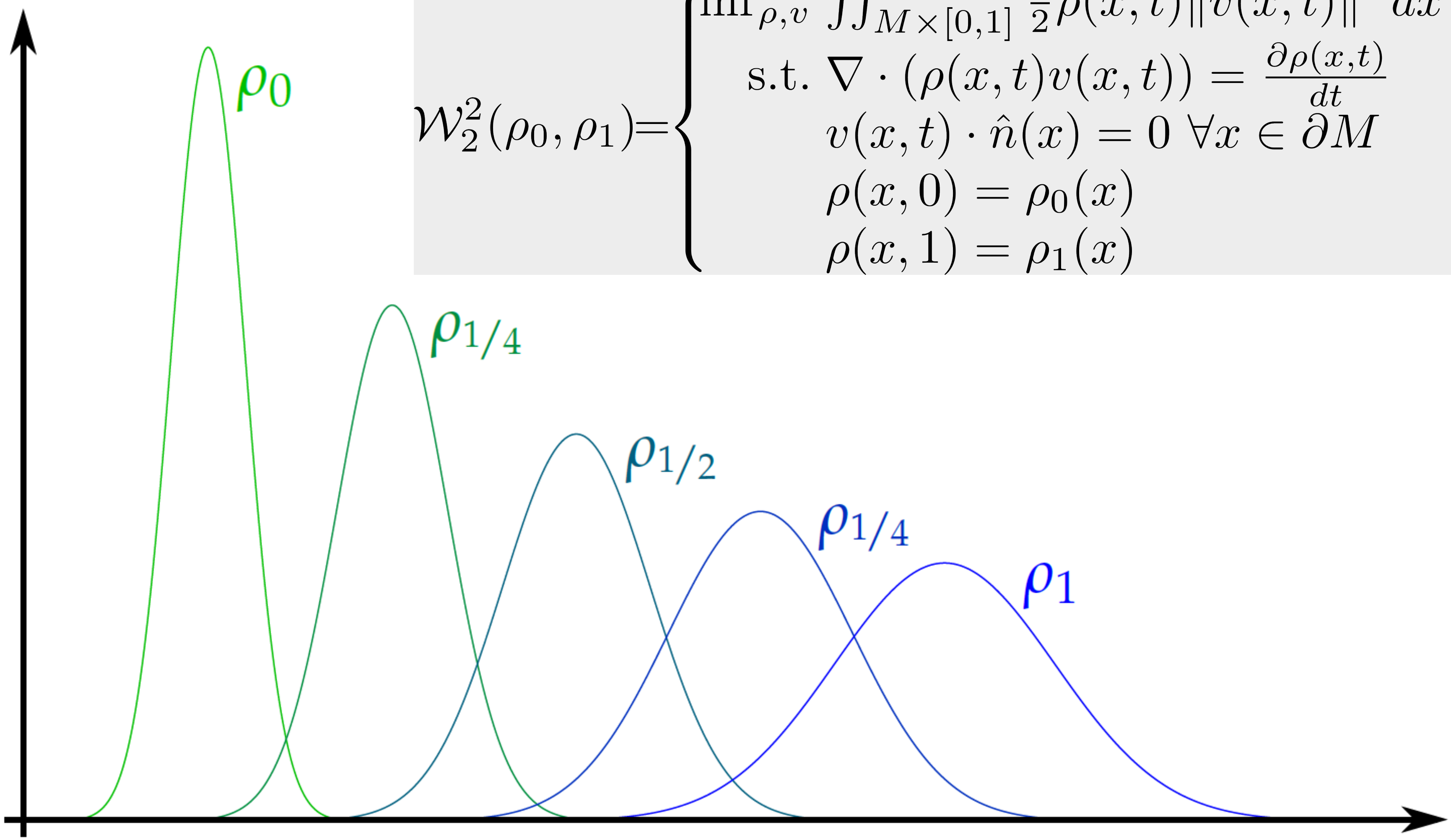
Benamou & Brenier

"A computational fluid mechanics solution of the Monge-Kantorovich mass transfer problem"

Numer. Math. 84 (2000), pp. 375-393

Recall: Displacement Interpolation

$$\mathcal{W}_2^2(\rho_0,\rho_1) = \begin{cases} \inf_{\rho,v} \iint_{M\times[0,1]} \frac{1}{2}\rho(x,t)\|v(x,t)\|^2\,dx\,dt \\ \quad \text{s.t. } \nabla\cdot(\rho(x,t)v(x,t)) = \frac{\partial\rho(x,t)}{dt} \\ \qquad v(x,t)\cdot\hat{n}(x) = 0\ \forall x\in\partial M \\ \qquad \rho(x,0) = \rho_0(x) \\ \qquad \rho(x,1) = \rho_1(x) \end{cases}$$



Convex Formulation

Velocity

$$\mathcal{W}_2^2(\rho_0,\rho_1)=\begin{cases}\inf_{\rho,v} \iint_{M\times[0,1]} \frac{1}{2}\rho(x,t)\|v(x,t)\|^2\,dx\,dt\\ \quad \text{s.t. } \nabla\cdot(\rho(x,t)v(x,t)) = \frac{\partial\rho(x,t)}{dt}\\ \qquad v(x,t)\cdot\hat{n}(x) = 0\ \forall x\in\partial M\\ \qquad \rho(x,0) = \rho_0(x)\\ \qquad \rho(x,1) = \rho_1(x)\end{cases}$$

Momentum

$$\mathcal{W}_2^2(\rho_0,\rho_1)=\begin{cases}\inf_{\rho,v} \iint_{M\times[0,1]} \frac{\|J(x,t)\|_2^2}{2\rho(x,t)}\,dx\,dt\\ \quad \text{s.t. } \nabla\cdot J(x,t) = \frac{\partial\rho(x,t)}{dt}\\ \qquad J(x,t)\cdot\hat{n}(x) = 0\ \forall x\in\partial M\\ \qquad \rho(x,0) = \rho_0(x)\\ \qquad \rho(x,1) = \rho_1(x)\end{cases}$$

$$J := \rho\cdot v$$

Lemma: Minimax Formulation

$$\frac{\|J\|_2^2}{2\rho} = \begin{cases} \sup_{a,b} & a\rho + b^\top J \\ \text{s.t.} & a + \frac{\|b\|_2^2}{2} \leq 0 \end{cases}$$

$$\begin{aligned} \inf_{J,\rho} \sup_{a,b,\phi} \quad & \int_0^1 \int_{\mathbb{R}^n} \left[a(x,t)\rho(x,t) + b(x,t)^\top J(x,t) + \phi(x,t)\left(\frac{\partial \rho(x,t)}{\partial t} + \nabla \cdot J(x,t)\right)\right] dA(x)\, dt \\ & + \int_{\mathbb{R}^n} \left[\phi(x,1)(\rho_1(x) - \rho(x,1)) - \phi(x,0)(\rho_0(x) - \rho(x,0))\right] dA(x) \\ \text{s.t.} \quad & a(x,t) + \frac{\|b(x,t)\|_2^2}{2} \leq 0 \;\forall x \in \mathbb{R}^n, t \in (0,1). \end{aligned}$$

Momentum

Boundary conditions

Advection equation

Simplification

$$z := \{\rho, J\}$$
$$q := \{a, b\}$$

$$F(q) := \begin{cases} 0 & \text{if } a(x,t) + \frac{\|b(x,t)\|_2^2}{2} \leq 0 \ \forall x \in \mathbb{R}^n, t \in (0,1) \\ \infty & \text{otherwise.} \end{cases}$$

$$G(\phi) := \int_{\mathbb{R}^n} (\phi(x,0)\rho_0(x) - \phi(x,1)\rho_1(x)) \, dA(x)$$

Convex!

$$\mathcal{W}_2^2(\rho_0, \rho_1) = -\sup_{z} \inf_{q,\phi} \Big[\underbrace{F(q) + G(\phi) + \langle z, \nabla_{x,t}\phi - q \rangle}_{L_r(\phi, q, z)} \Big]$$

Benamou-Brenier Algorithm

$$\mathcal{W}_2^2(\rho_0,\rho_1) = -\sup_z \inf_{q,\phi} \Big[\underbrace{F(q) + G(\phi) + \langle z, \nabla_{x,t}\phi - q\rangle}_{L_r(\phi,q,z)} \Big]$$

$$F(q) := \begin{cases} 0 & \text{if } a(x,t) + \frac{\|b(x,t)\|_2^2}{2} \le 0 \ \forall x \in \mathbb{R}^n, t \in (0,1) \\ \infty & \text{otherwise.} \end{cases}$$

$$G(\phi) := \int_{\mathbb{R}^n} \left(\phi(x,0)\rho_0(x) - \phi(x,1)\rho_1(x)\right) dA(x)$$

$$\phi^{\ell+1} \leftarrow \arg\min_\phi L_r(\phi, q^\ell, z^\ell)$$

Global linear PDE

$$\Delta\phi = \nabla \cdot (z - rq)$$

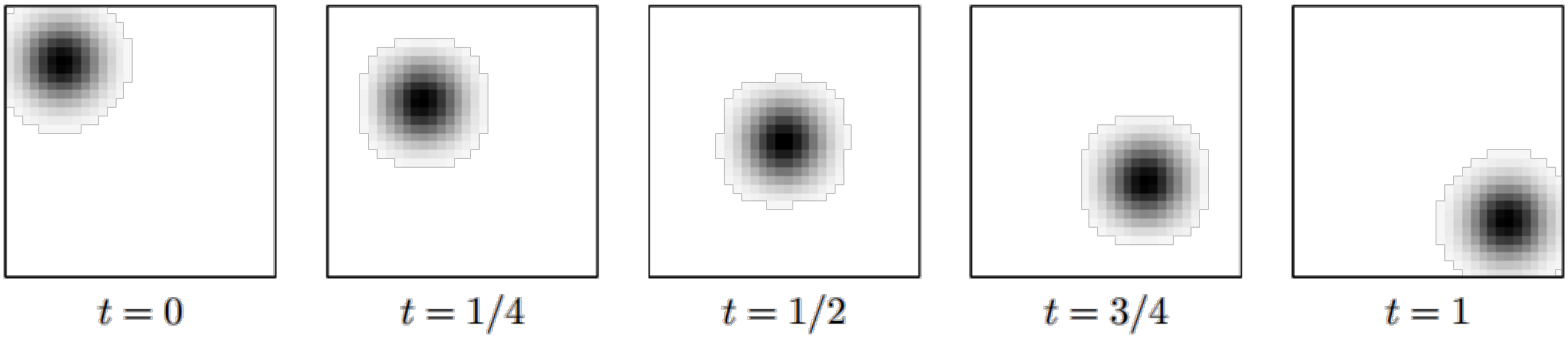
$$q^{\ell+1} \leftarrow \arg\min_q L_r(\phi^{\ell+1}, q, z^\ell)$$

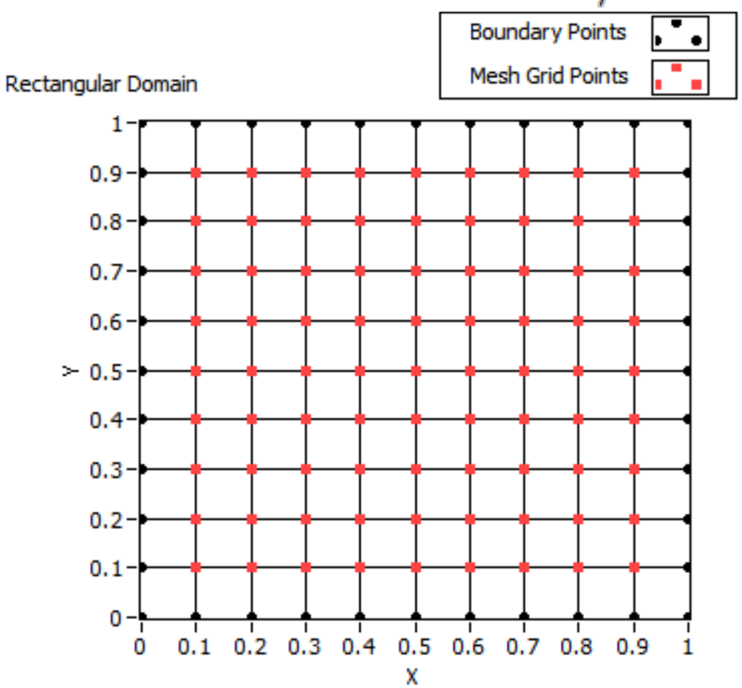
Local projection

$$z^{\ell+1} \leftarrow z^\ell + r(q - \nabla_{x,t}\phi)$$

Dual ascent

Discretization





Benamou-Brenier:
- Store (J, ρ) on each grid point
- Approximate objective function using FEM or divided differences

Open problem:
- Discretize on a triangle mesh
- Preserve triangle inequality

Talk to me if you're interested!

Transport image from "Optimal Transport with Proximal Splitting" (Papadakis, Peyré, and Oudet) • Grid from http://zone.ni.com/

Beckmann Problem: Linear Cost

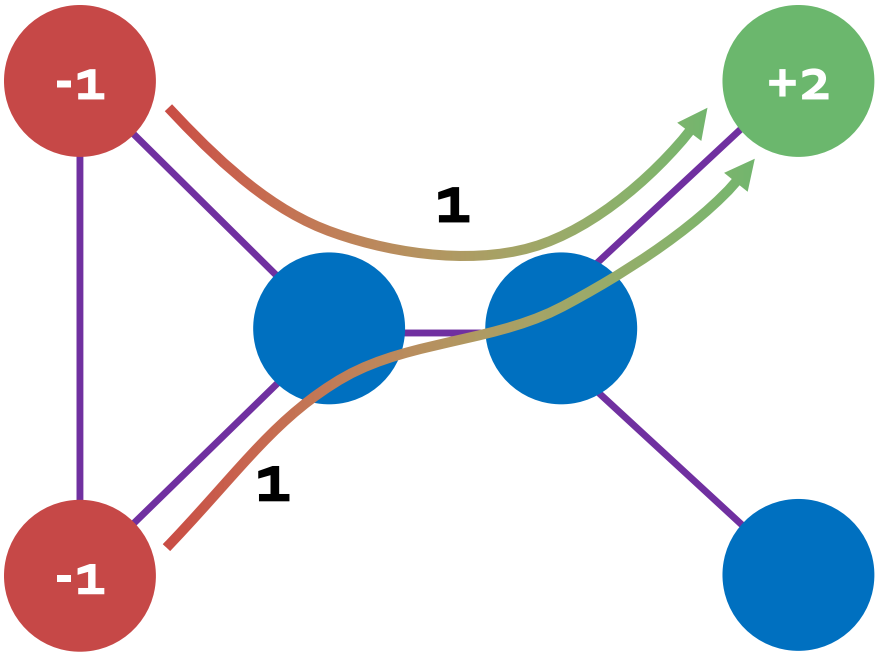
$$\min_T \langle T, D \rangle$$

$$\text{s.t. } T \geq 0$$

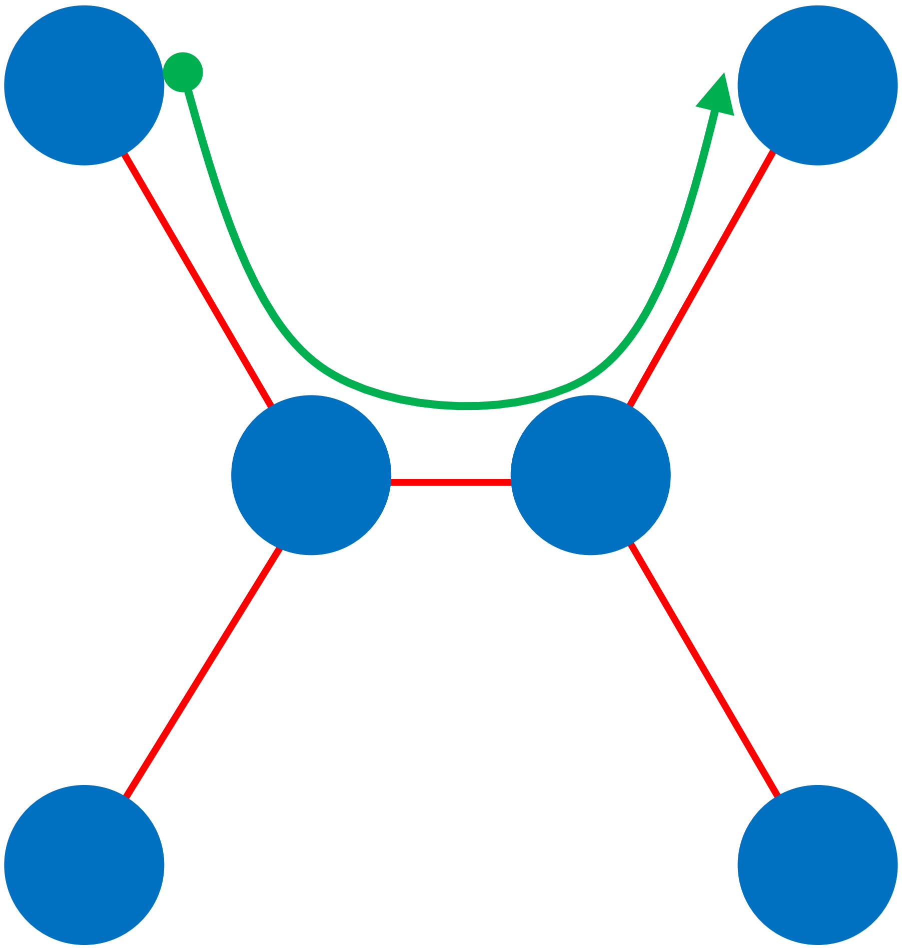
$$T\mathbf{1} = p_0$$

$$T^\top \mathbf{1} = p_1$$

D contains shortest path length.

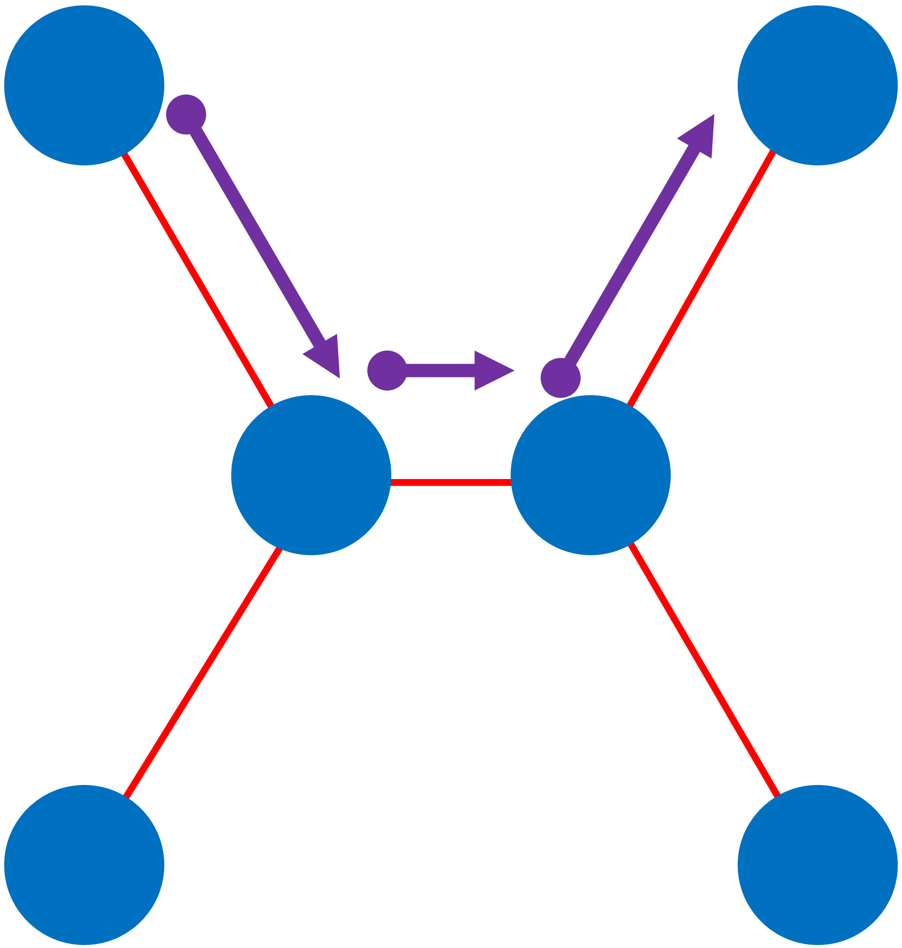


Simplification for Linear Cost



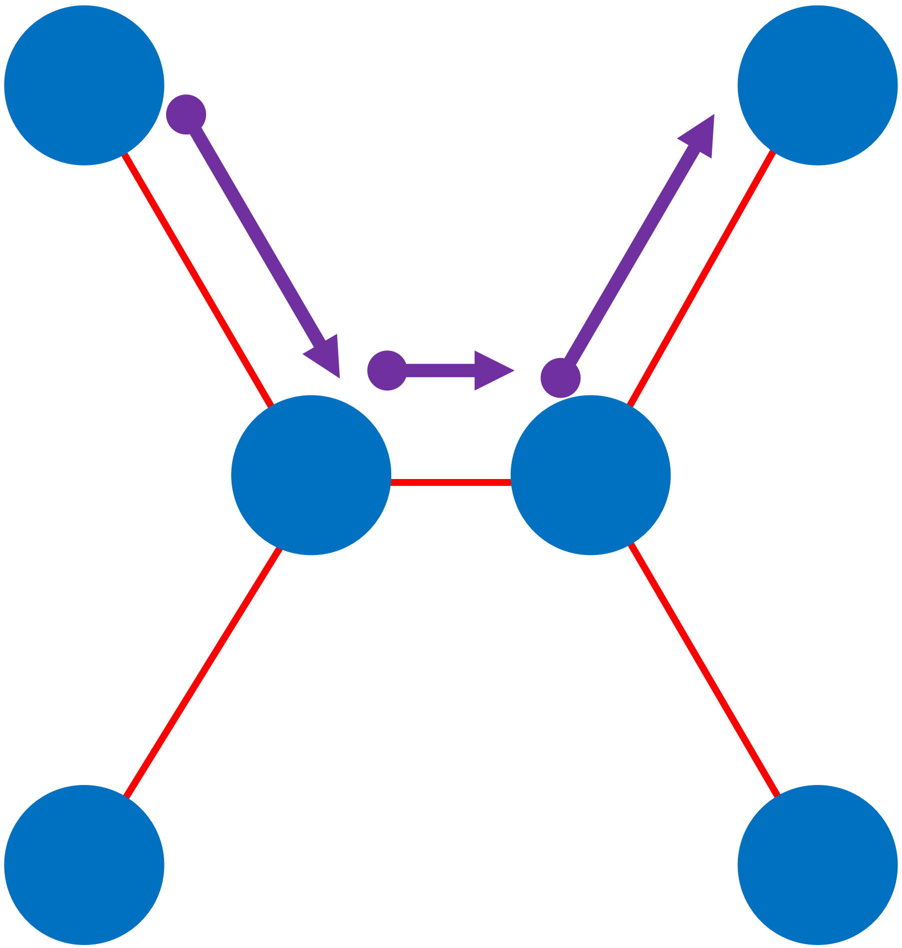
$$\begin{aligned} \min_T \quad & \textstyle\sum_{ij} T_{ij} d(x_i, x_j) \\ \text{s.t.} \quad & \textstyle\sum_j T_{ij} = p_i \\ & \textstyle\sum_i T_{ij} = q_j \\ & T \geq 0 \end{aligned}$$

Simplification for Linear Cost



$$\begin{aligned} \min_T \quad & \sum_{ij} T_{ij} d(x_i, x_j) \\ \text{s.t.} \quad & \sum_j T_{ij} = p_i \\ & \sum_i T_{ij} = q_j \\ & T \geq 0 \end{aligned}$$

Beckmann Formulation



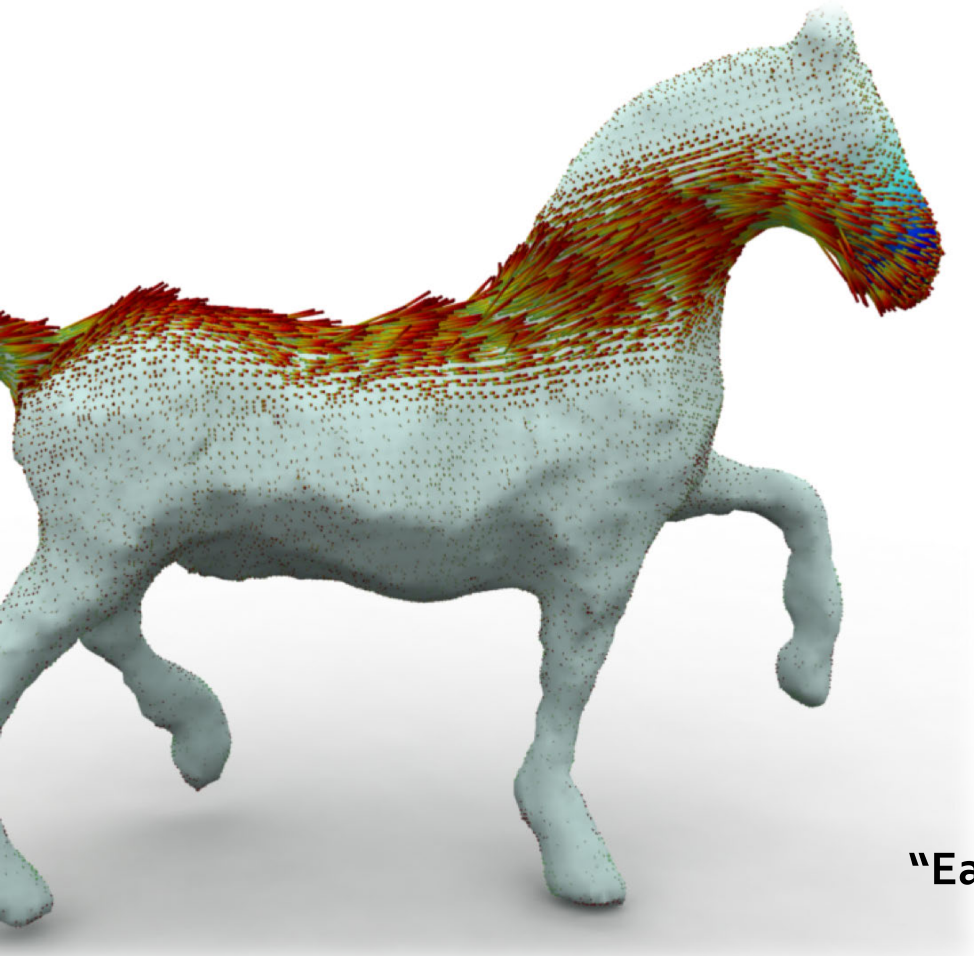
Better scaling for sparse graphs!

$$\begin{aligned} \min_T \quad & \textstyle\sum_e c_e |J_e| \\ \text{s.t.} \quad & D^\top J = \underbrace{p_1 - p_0}_{f} \end{aligned}$$

In computer science:

Network flow problem

Continuous Analog: Beckmann



Probabilities *advect* along the surface

"Eulerian"

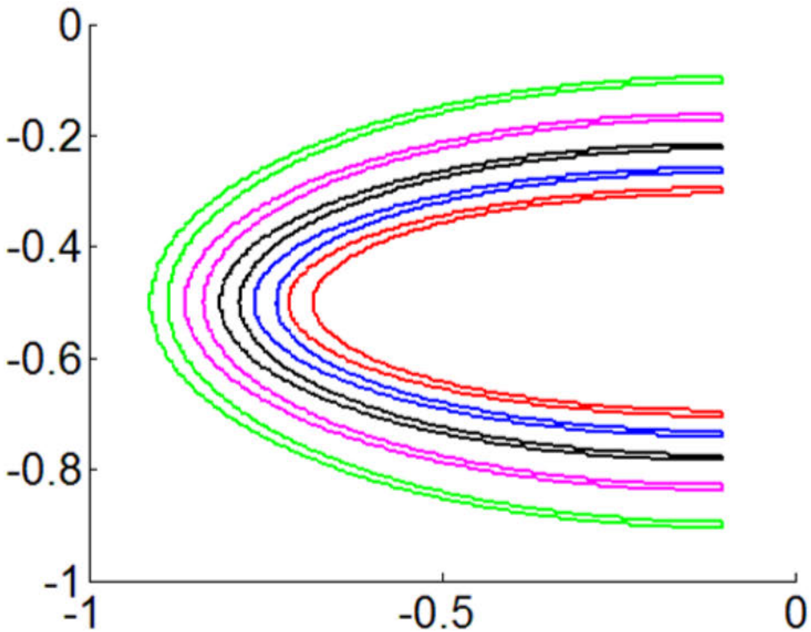
$$\mathcal{W}_1(\rho_0,\rho_1)=\begin{cases}\inf_J \int_M \|J(x)\|\,dx\\ \text{s.t. } \nabla\cdot J(x)=\rho_1(x)-\rho_0(x)\\ \qquad J(x)\cdot n(x)=0\ \forall x\in\partial M\end{cases}$$

Solomon, Rustamov, Guibas, and Butscher. "Earth Mover's Distances on Discrete Surfaces." SIGGRAPH 2014

Monge-Ampère PDE

$$\det(H\Psi(x))\rho_1(\nabla\Psi(x)) = \rho_0(x)$$

Monge problem solved by gradient of a convex function [Brenier 1991]
→ second-order nonlinear elliptic PDE



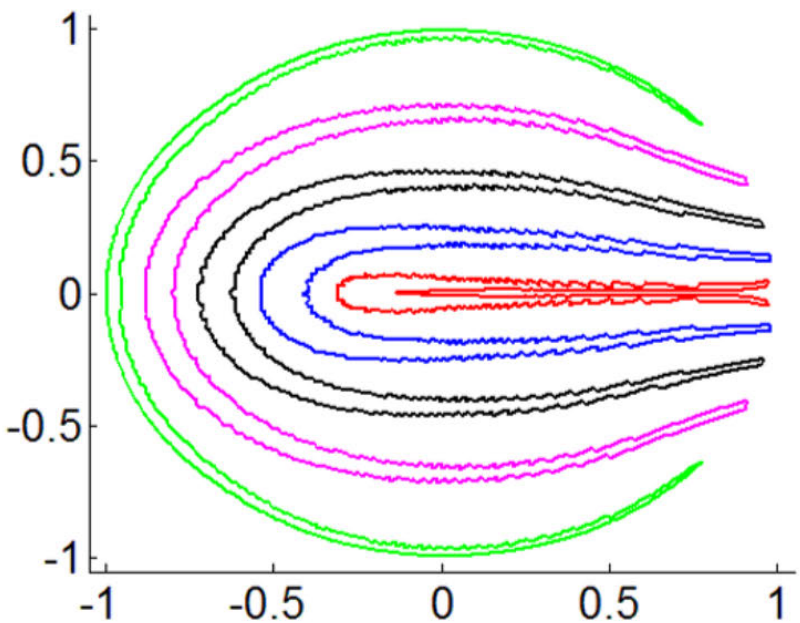


Image from [Benamou, Froese, & Oberman 2012]

Plan For Today

1. Introduction to optimal transport
   - 1D examples
   - Many formulas

2. Applications

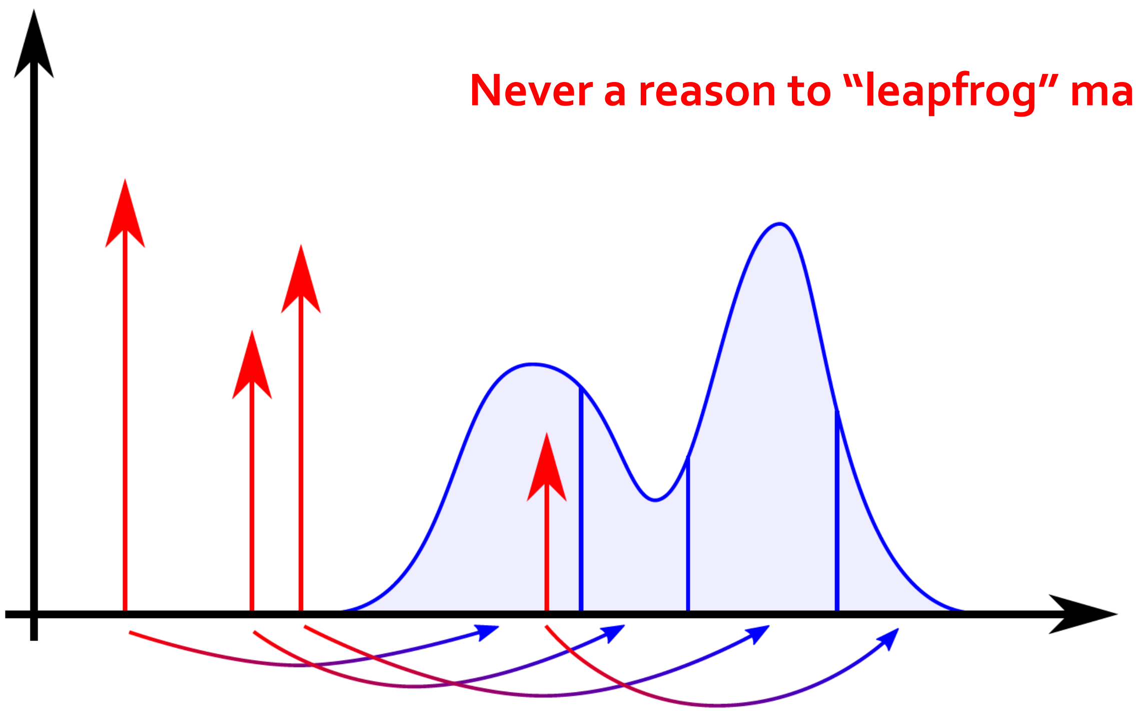
3. Discrete/discretized transport
   - Entropic regularization
   - Eulerian transport
   - Semidiscrete transport

4. Extensions

Recall:

Semidiscrete Transport

$$\mu_0 := \sum_{i=1}^{k_0} a_{0i}\delta_{x_{0i}} \qquad \mu_1(S) := \int_S \rho_1(x)\,dx$$



Never a reason to "leapfrog" mass!

Recall:

Kantorovich Duality

$$\mathrm{OT}(\mu,\nu;c) := \begin{cases} \min_\pi \ \iint_{X\times Y} c(x,y)\, d\pi(x,y) \\ \quad \text{s.t. } \ \pi \in \Pi(\mu,\nu) \end{cases}$$

Primal

$$= \begin{cases} \max_{\phi,\psi} \ \int_X \phi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) \\ \qquad \text{s.t. } \ \phi(x) + \psi(y) \le c(x,y) \text{ for a.e. } x \in X, y \in Y \end{cases}$$

Dual

General Case

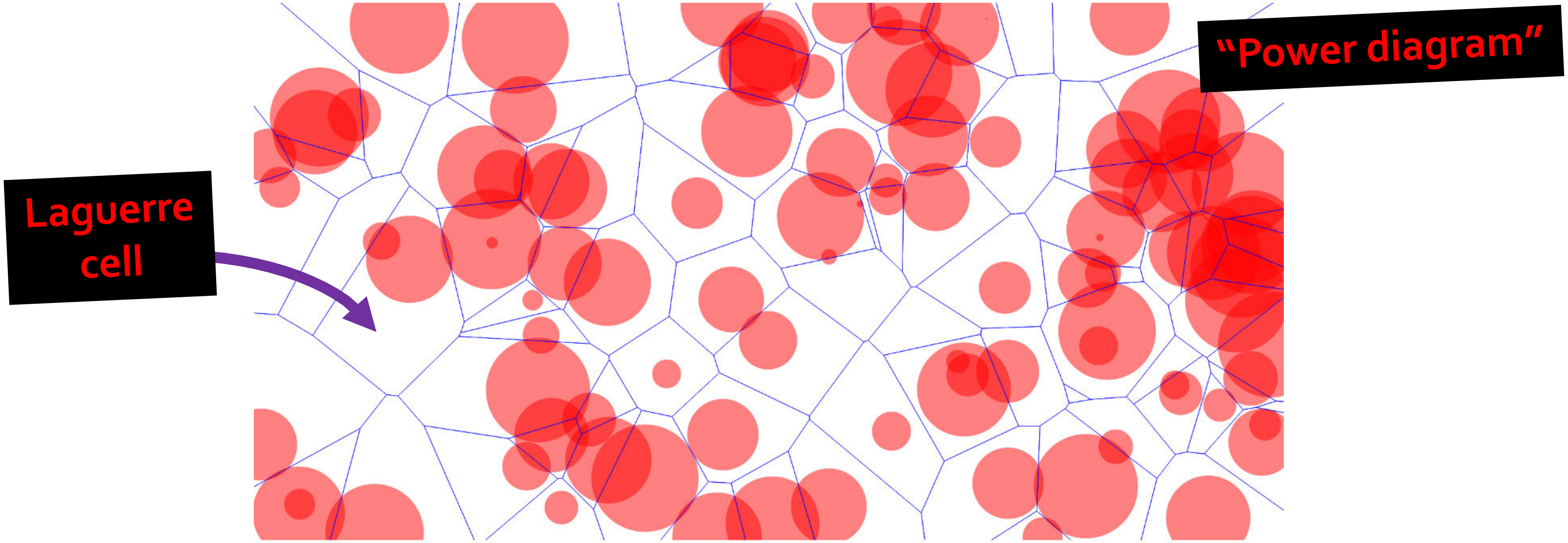
$$\mu_0 := \sum_{i=1}^{k} a_i \delta_{x_i}$$

$$\nu(S) := \int_S \rho(x)\,dx$$

On the board:

$$\mathcal{W}_2^2(\mu,\nu) = \sup_{\phi\in\mathbb{R}^k} \sum_i \left[a_i\phi_i + \int_{\mathrm{Lag}_\phi^c(x_i)} \rho(y)[c(x_i,y)-\phi_i]\,dA(y) \right]$$

$$\mathrm{Lag}_\phi^c(x_i) := \{y\in\mathbb{R}^n : c(x_i,y)-\phi_i \le c(x_j,y)-\phi_j\ \forall j\neq i\}$$



https://www.jasondavies.com/power-diagram/

Semidiscrete Algorithm

$$F(\phi) := \sum_i \left[a_i \phi_i + \int_{\mathrm{Lag}^c_\phi(x_i)} \rho(y)[c(x_i, y) - \phi_i]\, dA(y) \right]$$

$$\frac{\partial F}{\partial \phi_i} = a_i - \int_{\mathrm{Lag}^c_\phi(x_i)} \rho(y)\, dA(y)$$

Concave in ϕ!

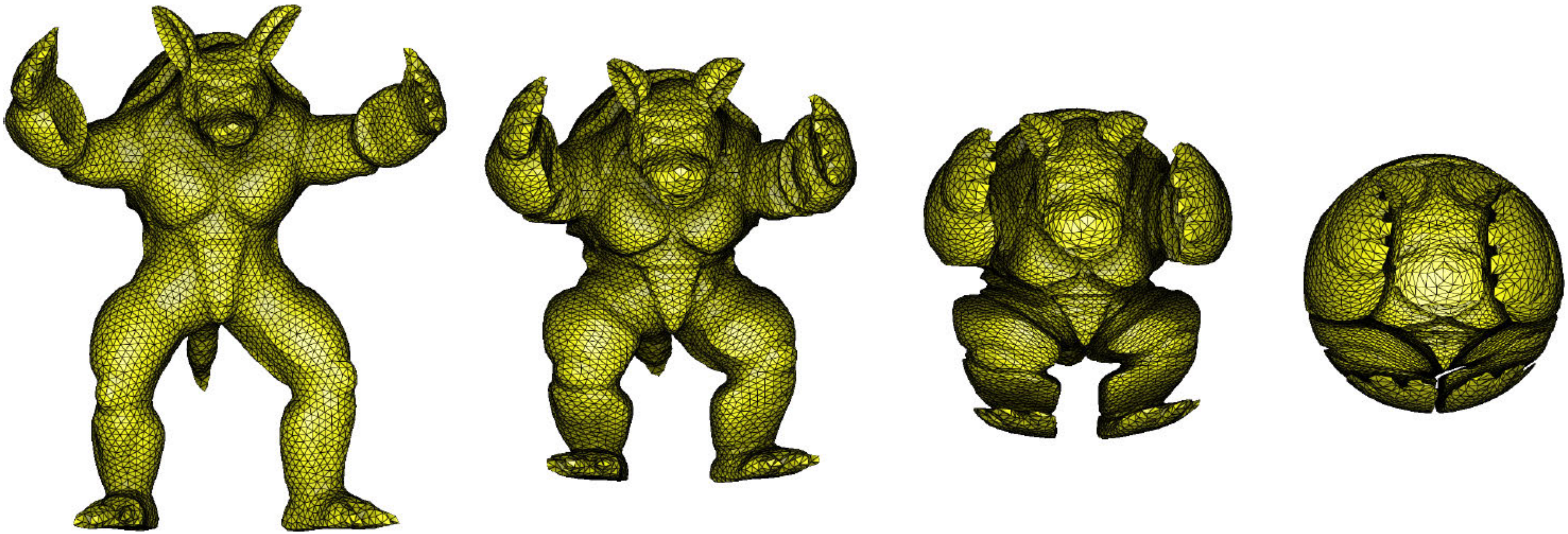
- **Simple algorithm: Gradient ascent**

  Ingredients: Power diagram

- **More complex: Newton's method**

  Converges globally [Kitagawa, Mérigot, & Thibert 2016]

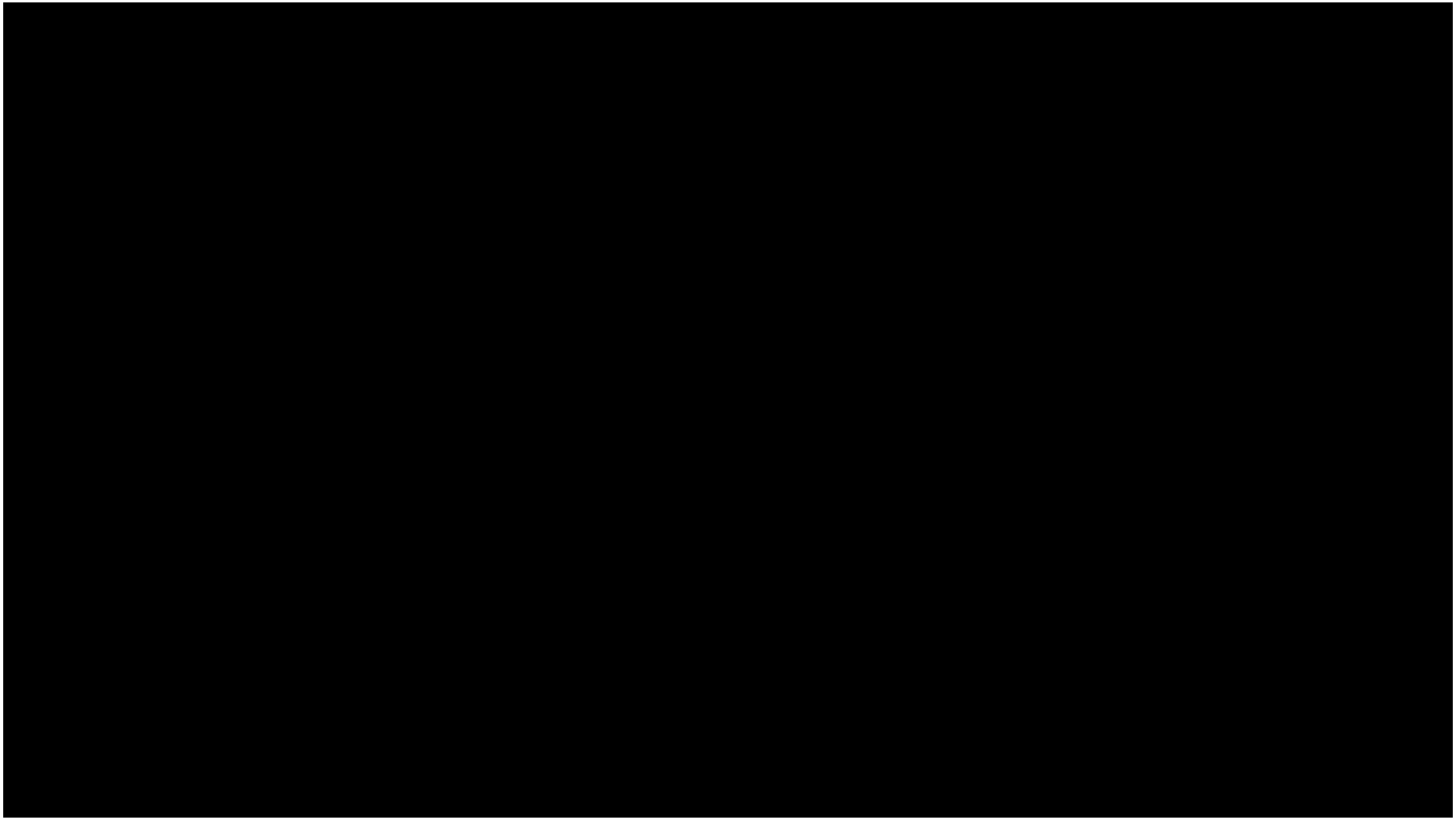
Application



Lévy. "A numerical algorithm for L2 semi-discrete optimal transport in 3D." (2014)

Points to tetrahedra

Application



Redux

Method	Advantages	Disadvantages
Entropic regularization	• Fast • Easy to implement • Works on mesh using heat kernel	• Blurry • Becomes singular as $\alpha \to 0$
Eulerian optimization	• Provides displacement interpolation • Connection to PDE	• Hard to optimize • Triangle mesh formulation unclear
Semidiscrete optimization	• No regularization • Connection to "classical" geometry	• Expensive computational geometry algorithms

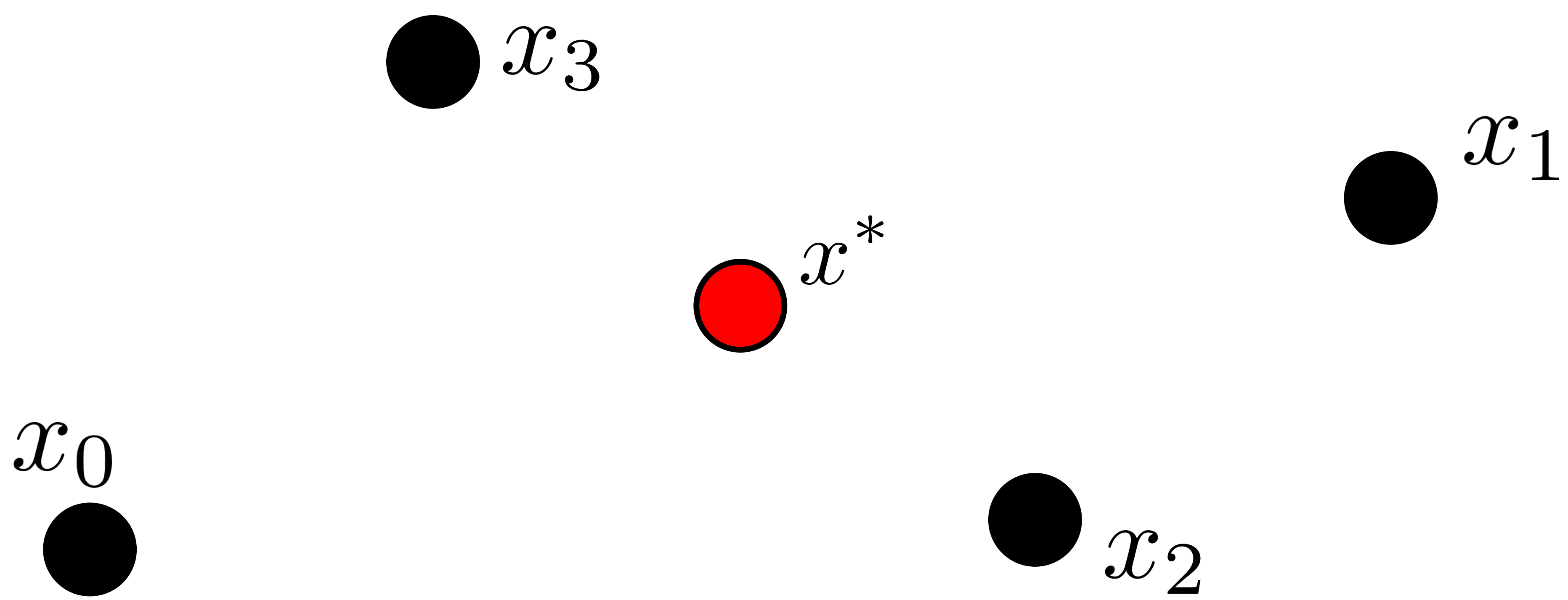
Many others:
Stochastic transport, dual ascent, Monge-Ampère PDE, …

Plan For Today

1. Introduction to optimal transport
   - 1D examples
   - Many formulas
2. Applications
3. Discrete/discretized transport
   - Entropic regularization
   - Eulerian transport
   - Semidiscrete transport
4. Extensions

Example: Averaging

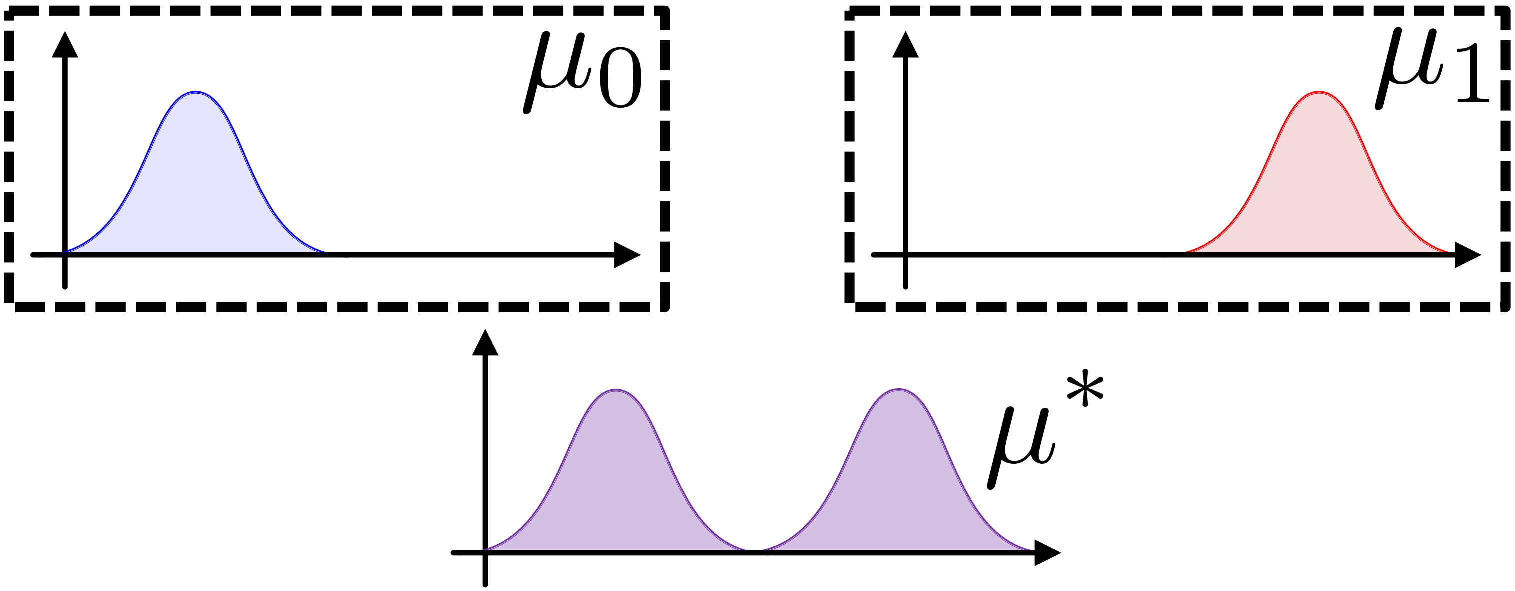
$$\text{Euclidean:} \quad x^* := \left[\arg\min_{x \in \mathbb{R}^n} \sum_i \|x - x_i\|_2^2\right] = \frac{1}{k}\sum_i x_i$$



Example: Averaging

$$\text{Euclidean:}\quad x^* := \left[\arg\min_{x\in\mathbb{R}^n} \sum_i \|x - x_i\|_2^2\right] = \frac{1}{k}\sum_i x_i$$

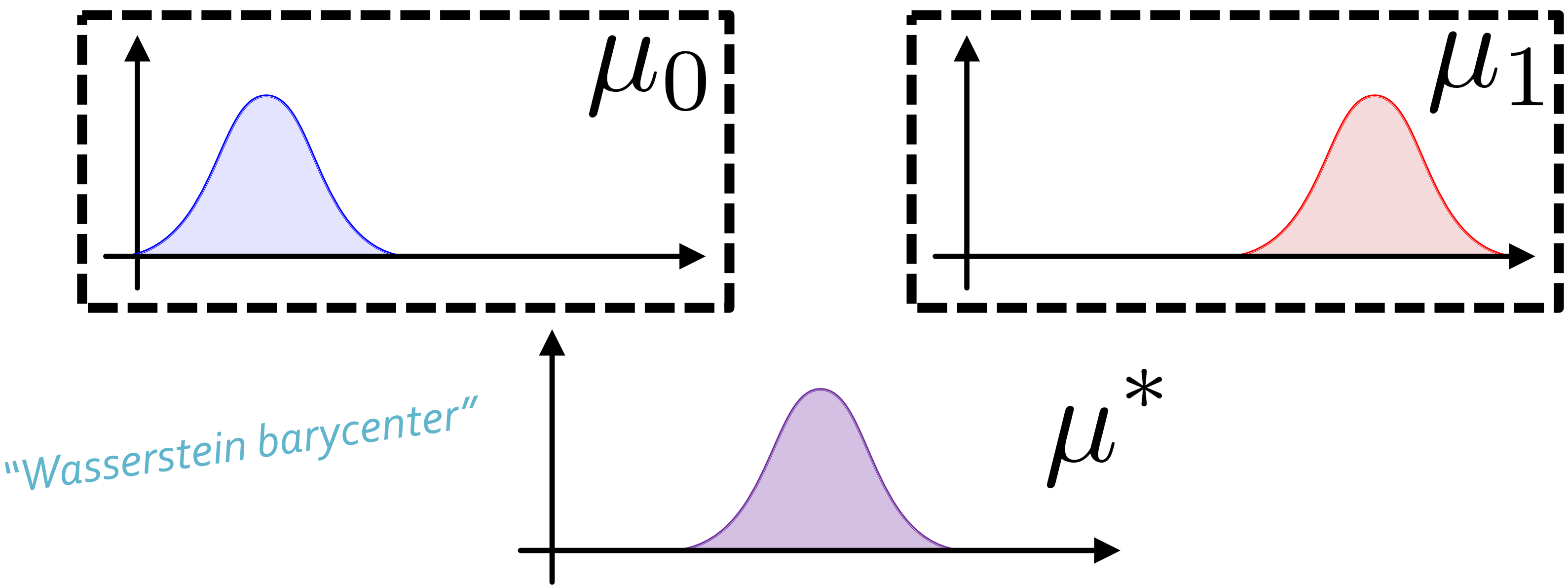
$$\text{Naïve distributional:}\quad \mu^* := \left[\arg\min_{\mu\in\mathrm{Prob}(\mathbb{R}^n)} \sum_i \|\mu - \mu_i\|_1\right] = \frac{1}{k}\sum_i \mu_i$$



Wassersteinized Averaging

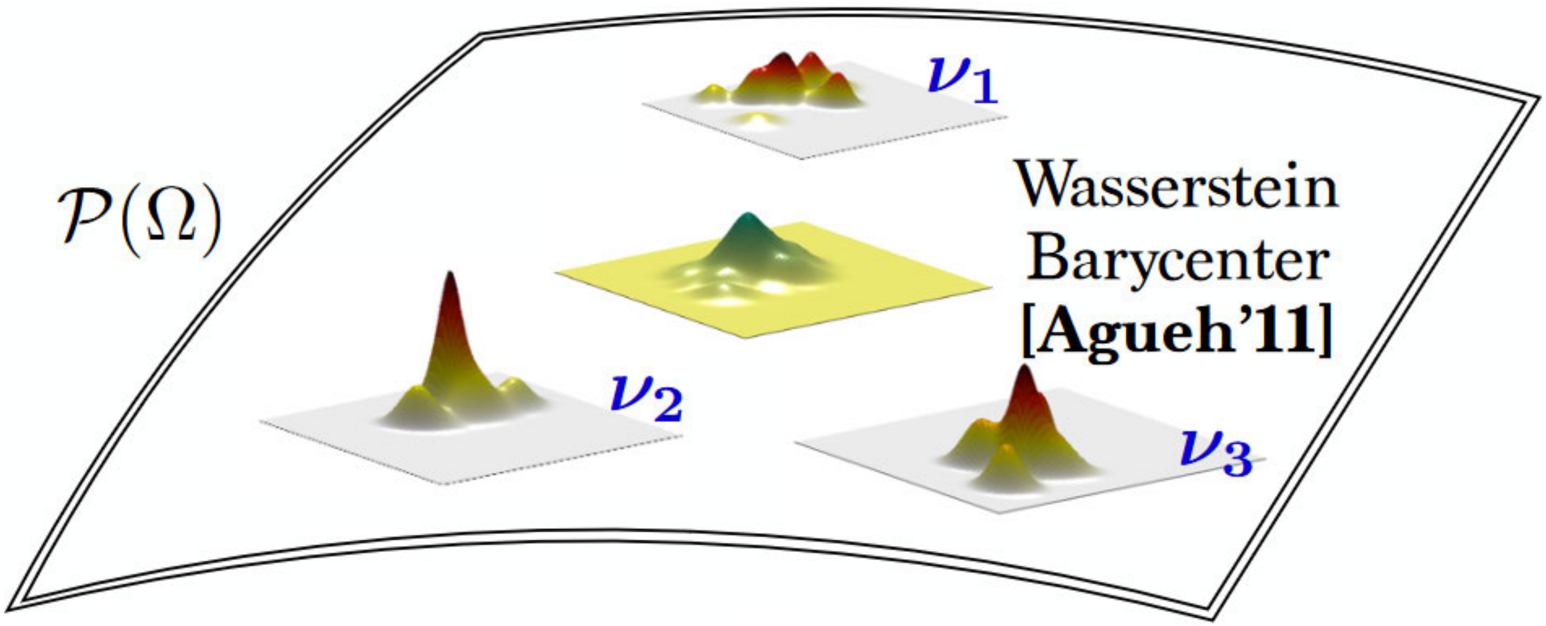
$$\text{Euclidean:} \quad x^* := \left[\arg\min_{x \in \mathbb{R}^n} \sum_i \|x - x_i\|_2^2 \right] = \frac{1}{k} \sum_i x_i$$

$$\text{Wasserstein:} \quad \mu^* := \left[\arg\min_{\mu \in \mathrm{Prob}(\mathbb{R}^n)} \sum_i \mathcal{W}_2^2(\mu, \mu_i) \right]$$



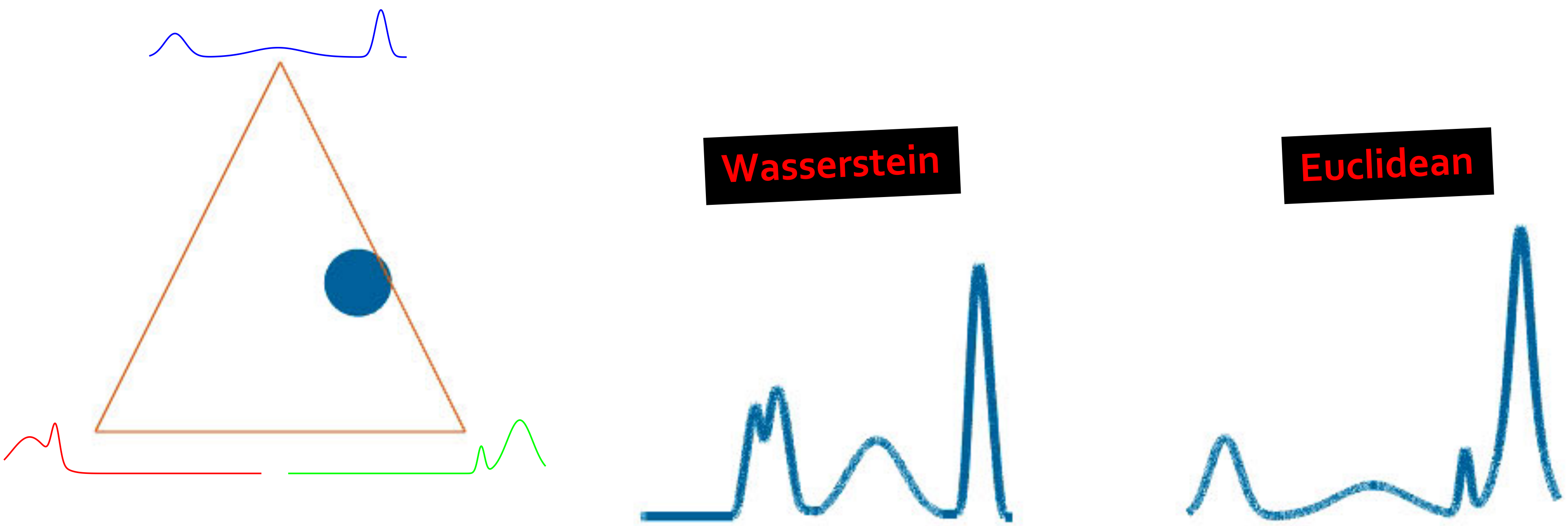
Barycenters

$$\min_{\mu \in \mathcal{P}(\Omega)} \sum_{i=1}^{N} \lambda_i W_p^p(\mu, \nu_i)$$

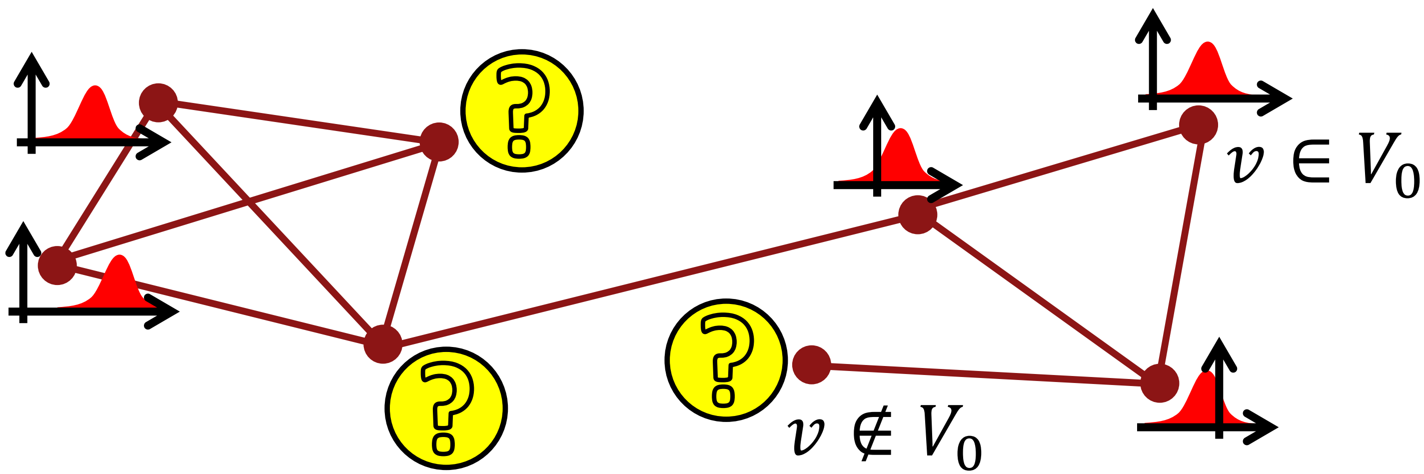


Slide courtesy M. Cuturi

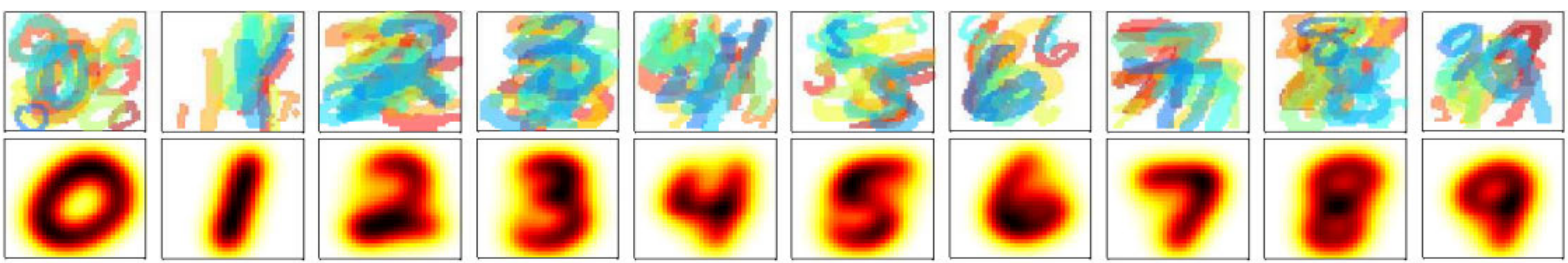
Barycenter Example



Barycenters in Machine Learning

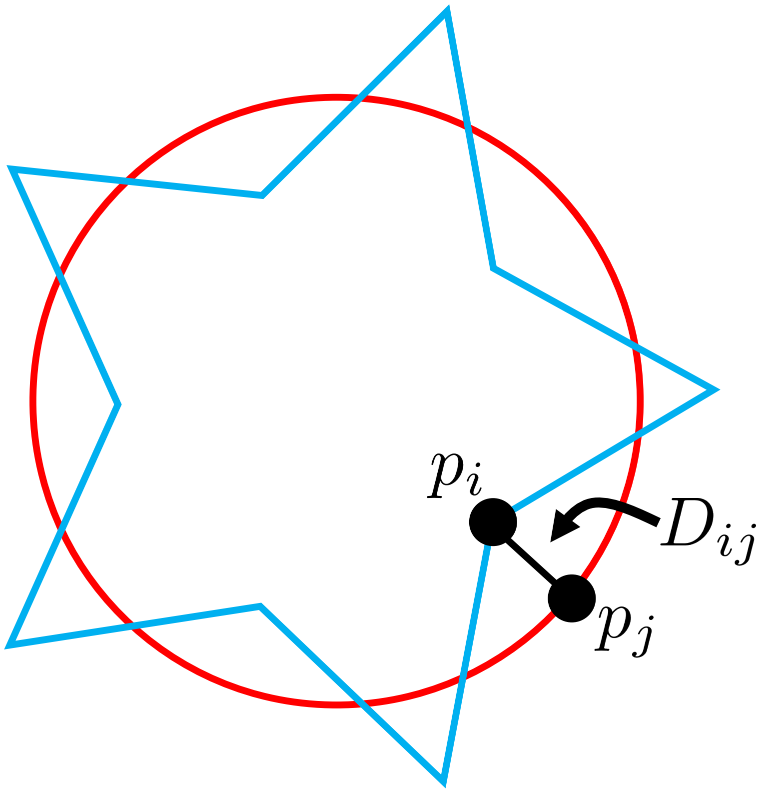


"Wasserstein Propagation for Semi-Supervised Learning" (Solomon et al.)



"Fast Computation of Wasserstein Barycenters" (Cuturi and Doucet)

Model Problem: Linear Assignment



Between signals

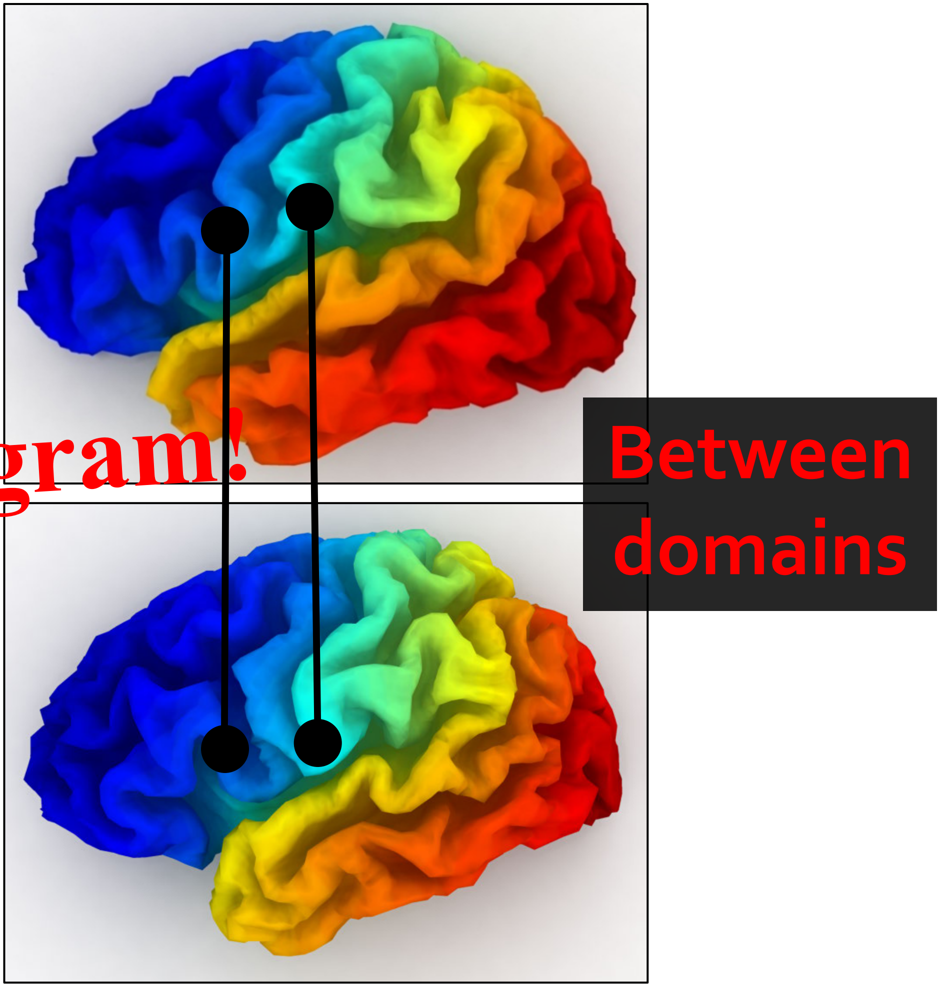
$$
\begin{aligned}
\min_T \quad & \langle T, D \rangle \\
\text{s.t.} \quad & T \geq 0 \\
& T\mathbf{1} = \mathbf{1} \\
& T^\top \mathbf{1} = \mathbf{1}
\end{aligned}
$$

"No matched point should travel too far."

Model Problem: Quadratic Matching

$$\begin{aligned} \min_T \quad & \langle M_0 T, T M_1 \rangle \\ \text{s.t.} \quad & T \geq 0 \\ & T\mathbf{1} = \mathbf{1} \\ & T^\top \mathbf{1} = \mathbf{1} \end{aligned}$$

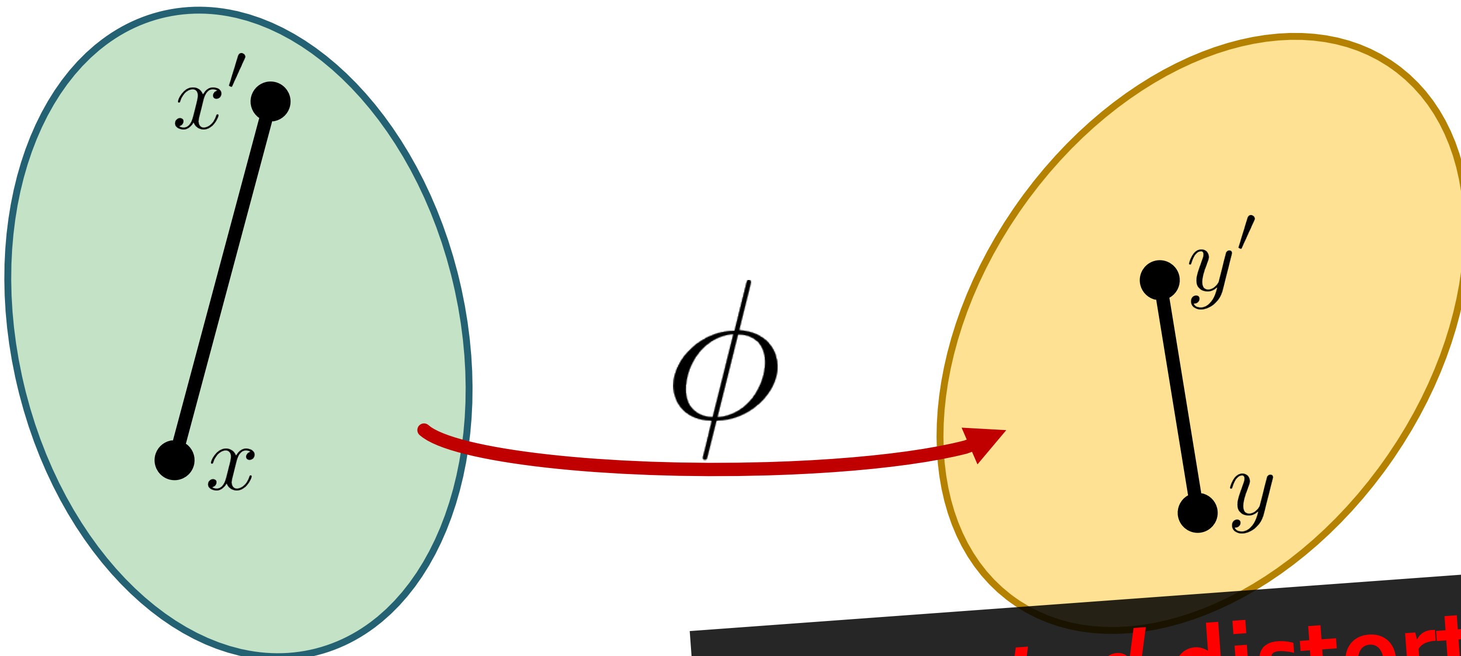
onvex quadratic program!



"Nearby points stay nearby."

Gromov-Wasserstein Distance

[Mémoli 2007]

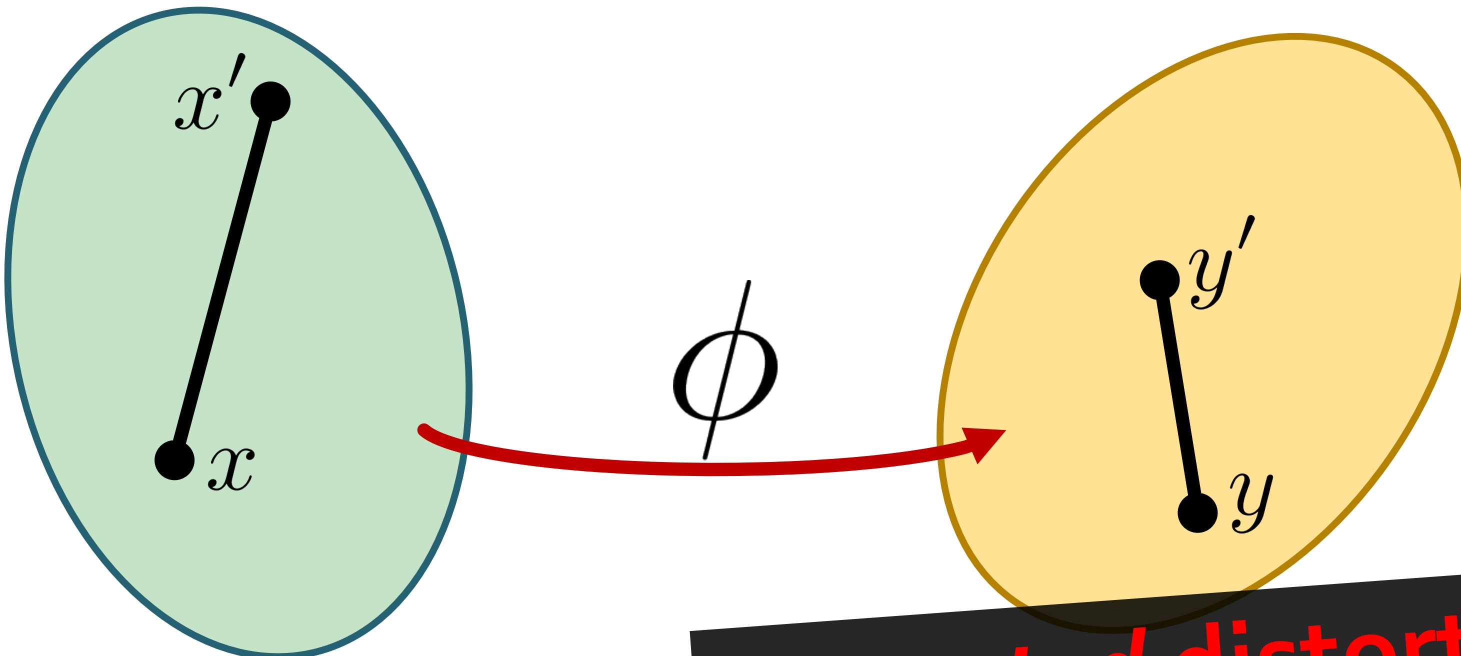


Expected distortion

$$\mathrm{GW}_2^2((\mu_0, d_0), (\mu, d)) := \min_{\gamma \in \mathcal{M}(\mu_0, \mu)} \iint_{\Sigma_0 \times \Sigma} [d_0(x, x') - d(y, y')]^2 \, d\gamma(x, y) \, d\gamma(x', y')$$

Quadratic Matching

[Mémoli 2007]



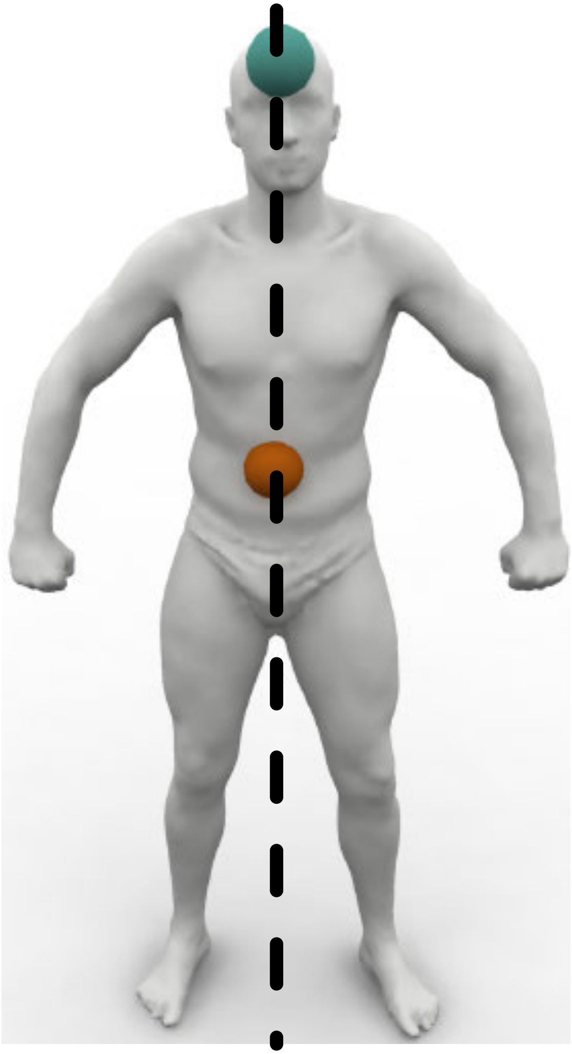
Expected distortion

$$\mathrm{GW}_2^2((\mu_0, d_0), (\mu, d)) :=$$

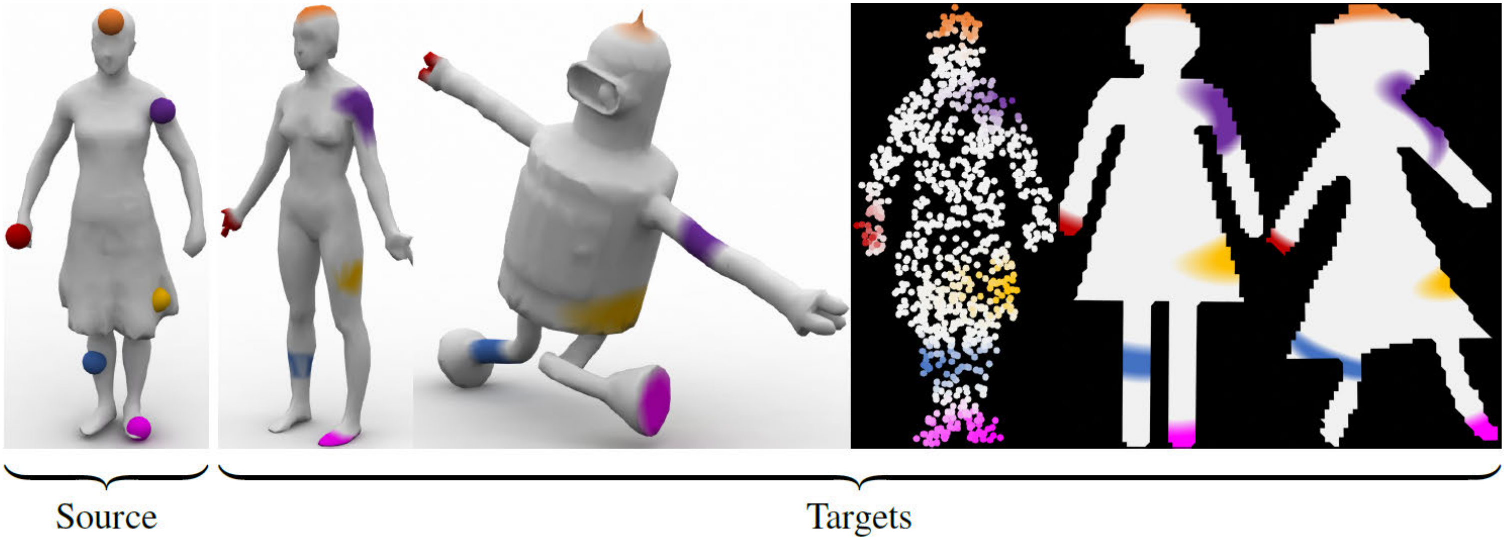
$$\min_{\gamma \in \mathcal{M}(\mu_0, \mu)} \iint_{\Sigma_0 \times \Sigma} [d_0(x, x') - d(y, y')]^2 \, \boxed{d\gamma(x, y)\, d\gamma(x', y')}$$

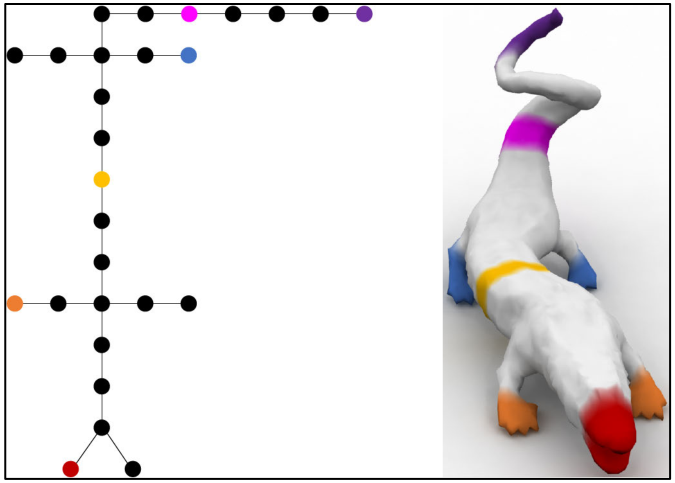
Fundamentally Nonconvex

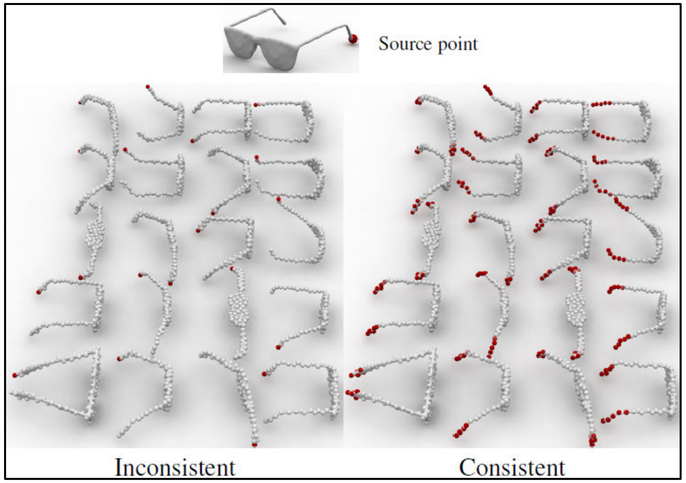
Convex relaxations *must* fail!



Variety of Correspondence Tasks

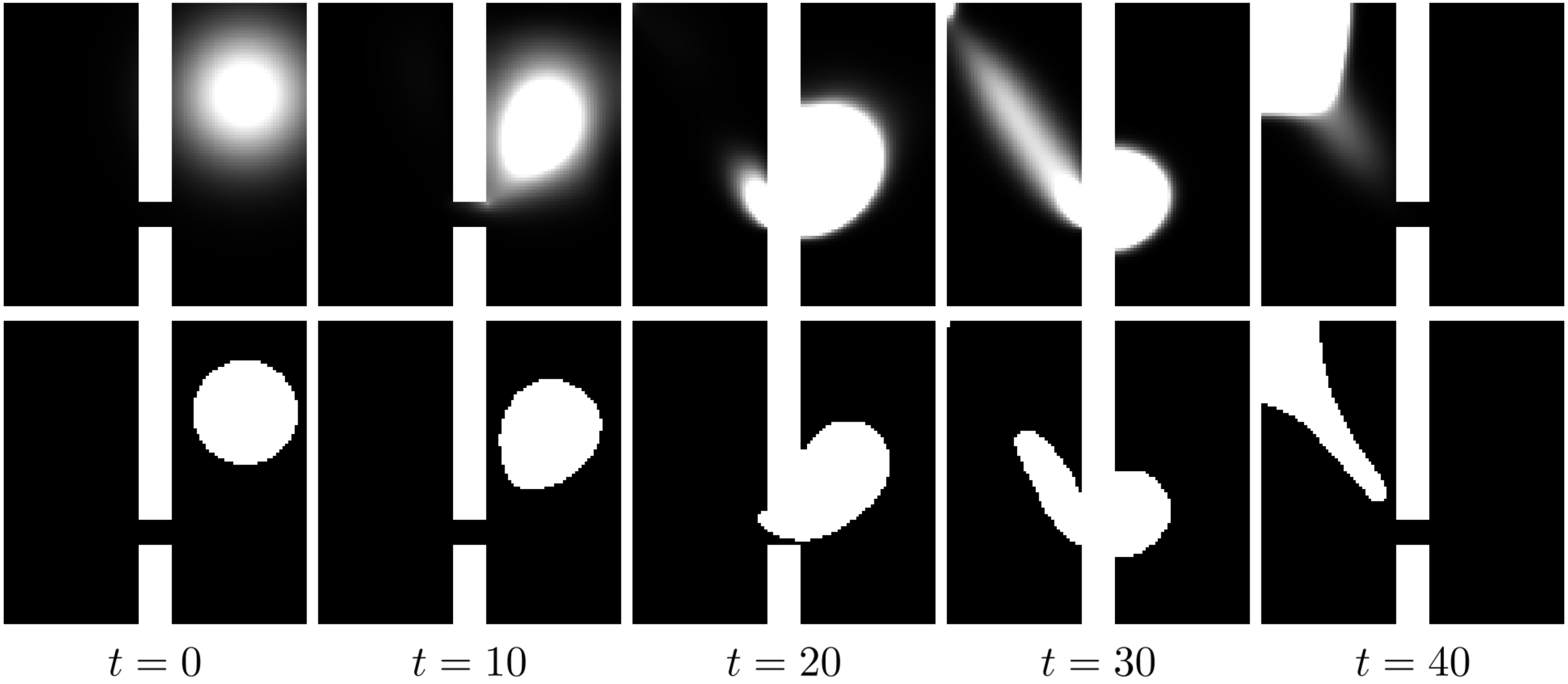






[Solomon et al. 2016]

Gradient Flows



"Entropic Wasserstein Gradient Flows" [Peyré 2015]

Matrix Fields and Vector Measures

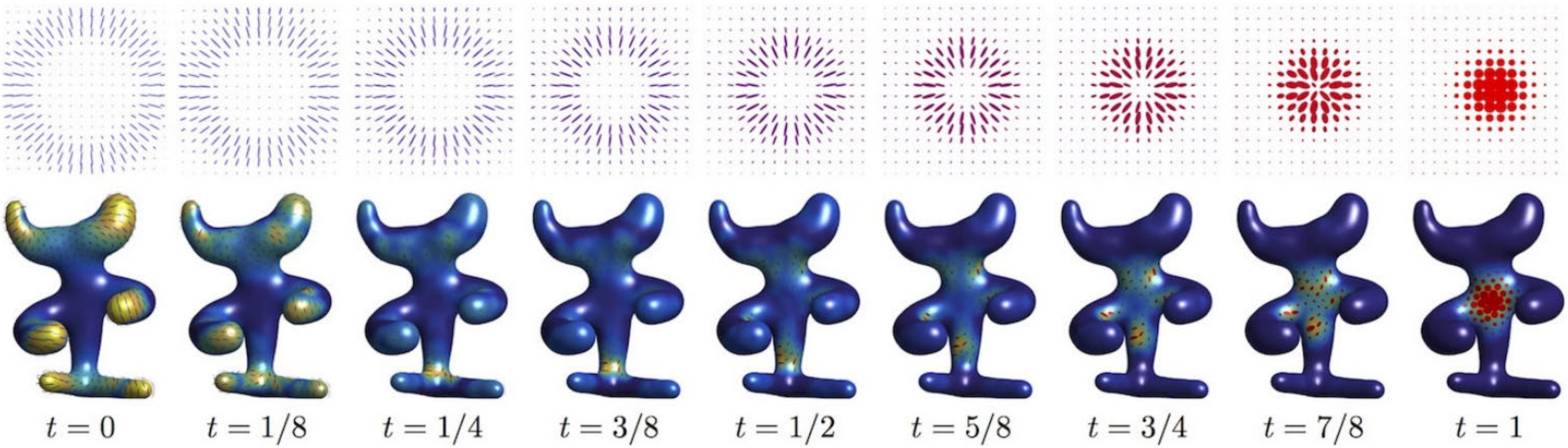


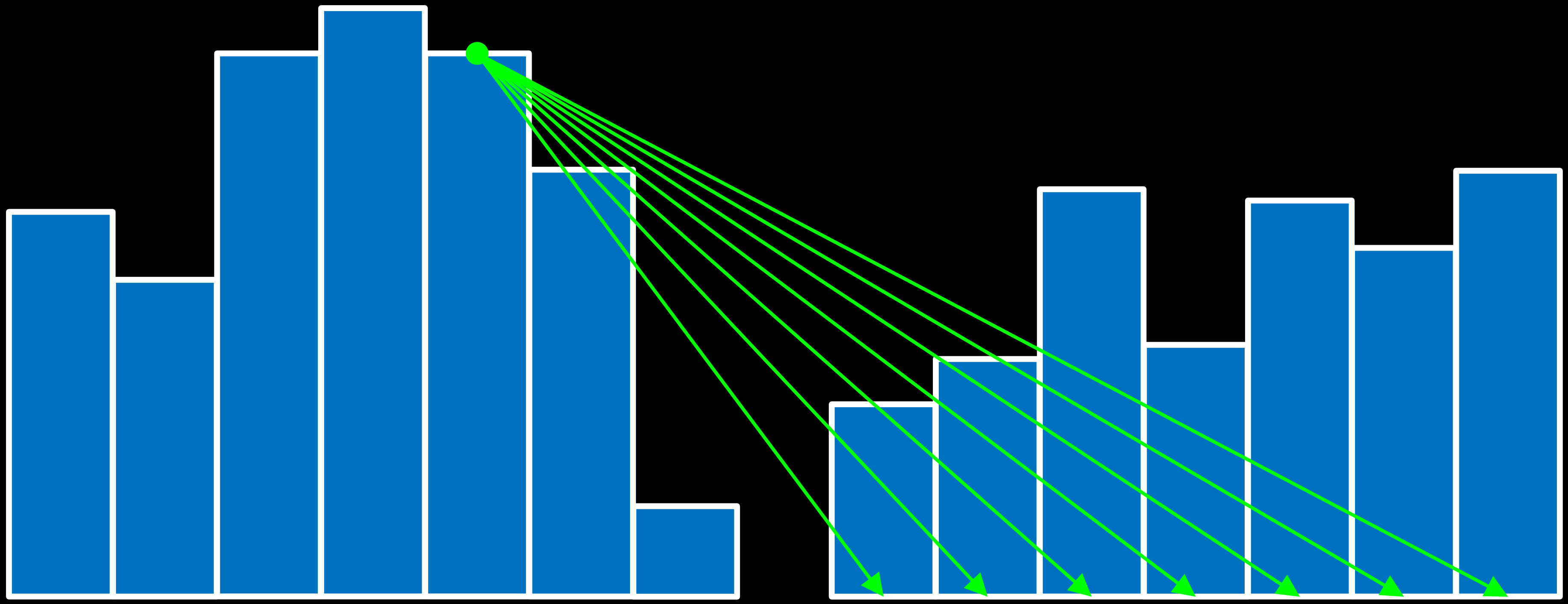
Image from

"Quantum Optimal Transport for Tensor Field Processing"

[Peyré et al. 2017]

Plan For Today

1. Introduction to optimal transport
   - 1D examples
   - Many formulas
2. Applications
3. Discrete/discretized transport
   - Entropic regularization
   - Eulerian transport
   - Semidiscrete transport
4. Extensions



Optimal Transport

on Discrete Domains

Justin Solomon

MIT EECS