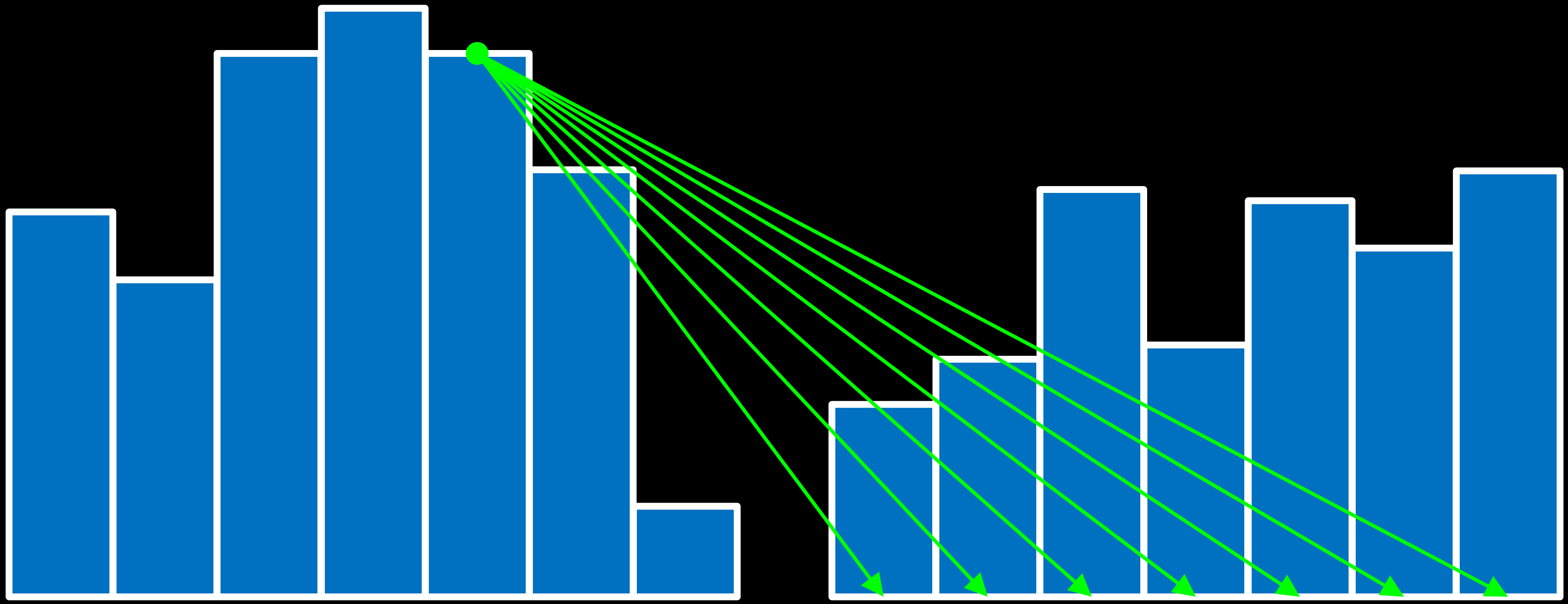


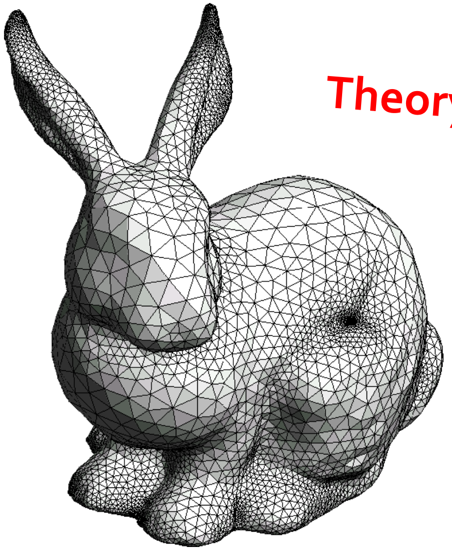
Tutorial on

Optimal Transport

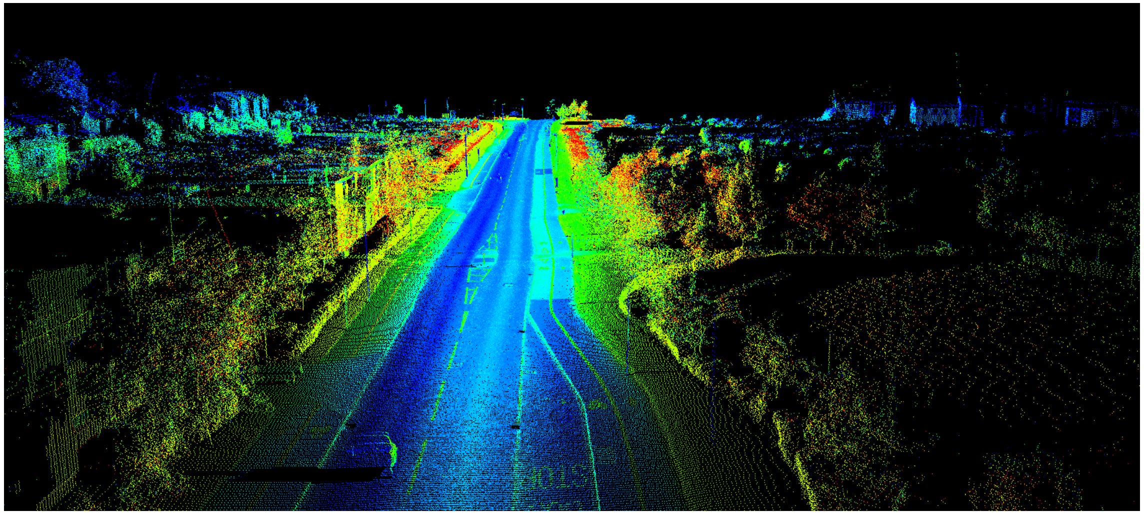
Justin Solomon

Motivation

Theory



Practice



Potential VRDI Motivation

- Reasonable notion of compactness:
  Travel time to voting booth
- Connection to Voronoi and power diagrams
- Natural measure of geometric similarity

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

Plan For Today

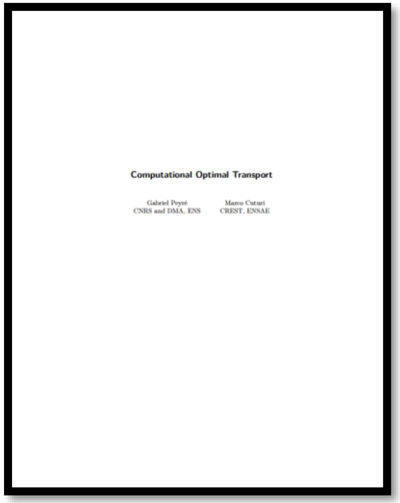
1. Introduction to optimal transport
   - 1D examples
   - Many formulas

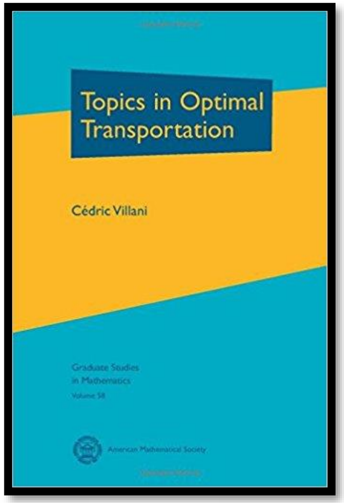
2. Applications

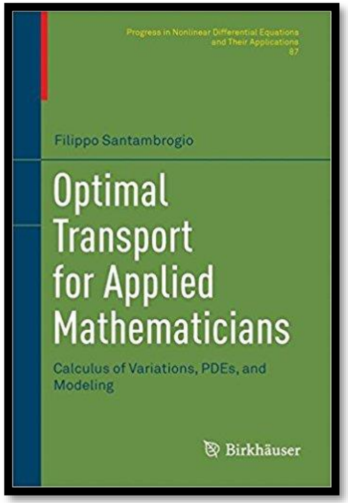
3. Discrete/discretized transport
   - Entropic regularization
   - Eulerian transport
   - Semidiscrete transport

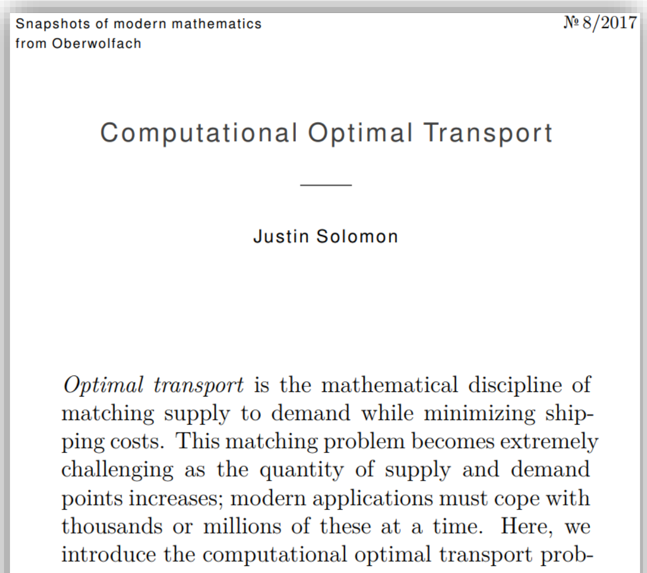
4. Extensions & frontiers

Useful References





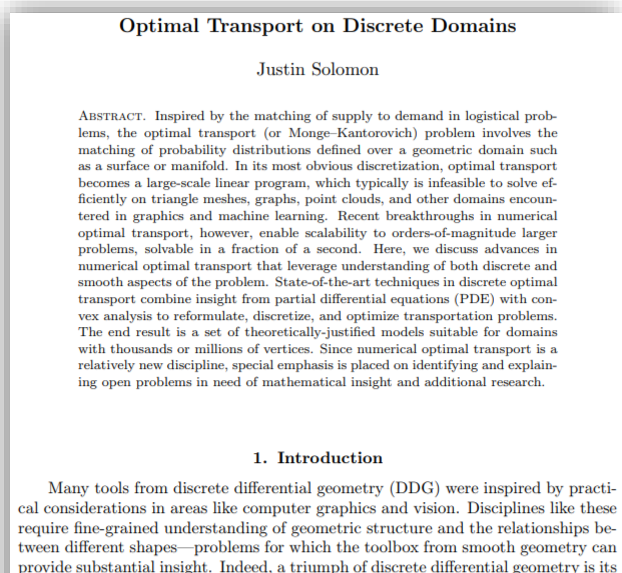



Snapshots of modern mathematics from Oberwolfach №8/2017

Computational Optimal Transport

Justin Solomon

Optimal transport is the mathematical discipline of matching supply to demand while minimizing shipping costs. This matching problem becomes extremely challenging as the quantity of supply and demand points increases; modern applications must cope with thousands or millions of these at a time. Here, we introduce the computational optimal transport prob-


Optimal Transport on Discrete Domains

Justin Solomon

Abstract. Inspired by the matching of supply to demand in logistical problems, the optimal transport (or Monge–Kantorovich) problem involves the matching of probability distributions defined over a geometric domain such as a surface or manifold. In its most obvious discretization, optimal transport becomes a large-scale linear program, which typically is infeasible to solve efficiently on triangle meshes, graphs, point clouds, and other domains encountered in graphics and machine learning. Recent breakthroughs in numerical optimal transport, however, enable scalability to orders-of-magnitude larger problems, solvable in a fraction of a second. Here, we discuss advances in numerical optimal transport that leverage understanding of both discrete and smooth aspects of the problem. State-of-the-art techniques in discrete optimal transport combine insight from partial differential equations (PDE) with convex analysis to reformulate, discretize, and optimize transportation problems. The end result is a set of theoretically-justified models suitable for domains with thousands or millions of vertices. Since numerical optimal transport is a relatively new discipline, special emphasis is placed on identifying and explaining open problems in need of mathematical insight and additional research.

1. Introduction

Many tools from discrete differential geometry (DDG) were inspired by practical considerations in areas like computer graphics and vision. Disciplines like these require fine-grained understanding of geometric structure and the relationships between different shapes—problems for which the toolbox from smooth geometry can provide substantial insight. Indeed, a triumph of discrete differential geometry is its

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 & frontiers

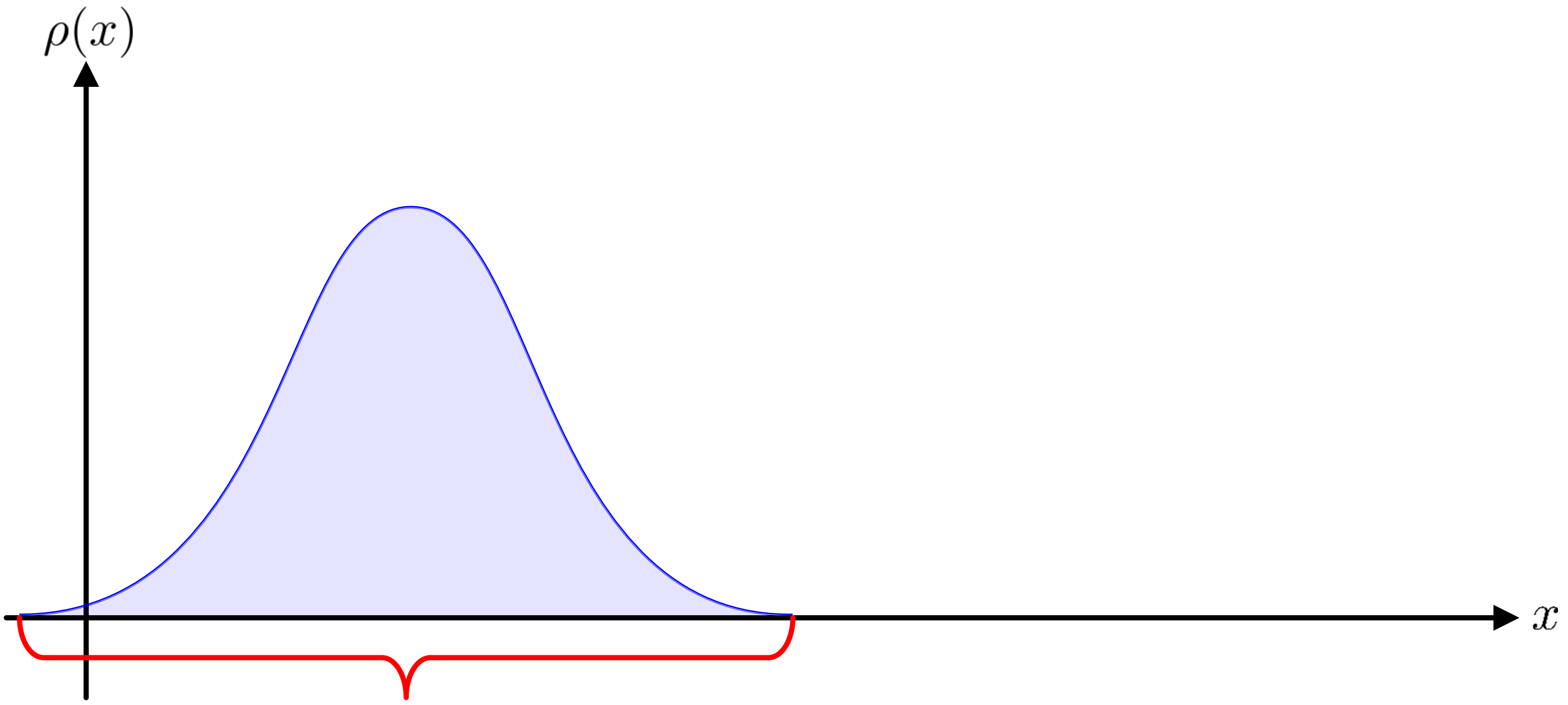
Transport Philosophy

Understand geometry from a

"softened" probabilistic

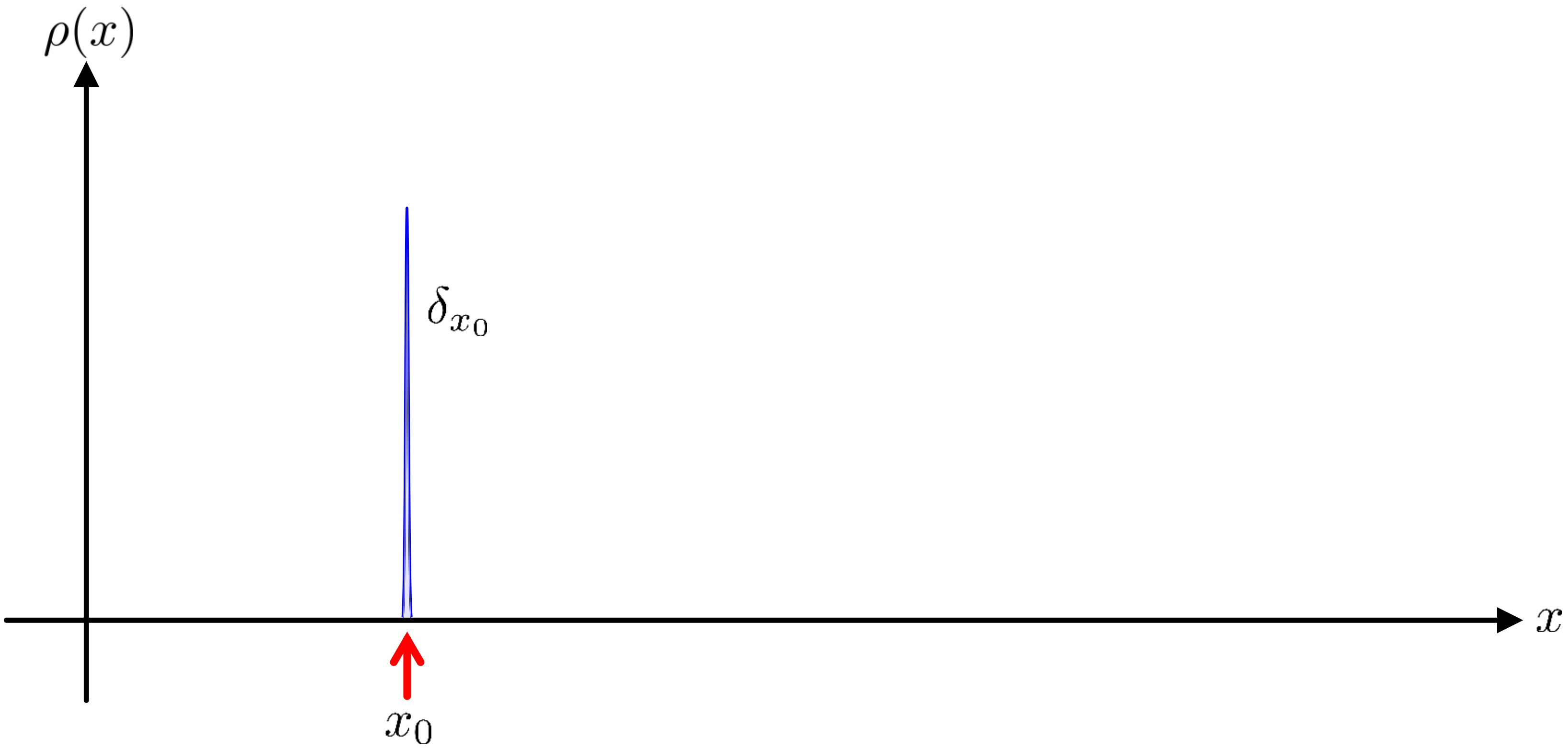
standpoint.

Probability as Geometry



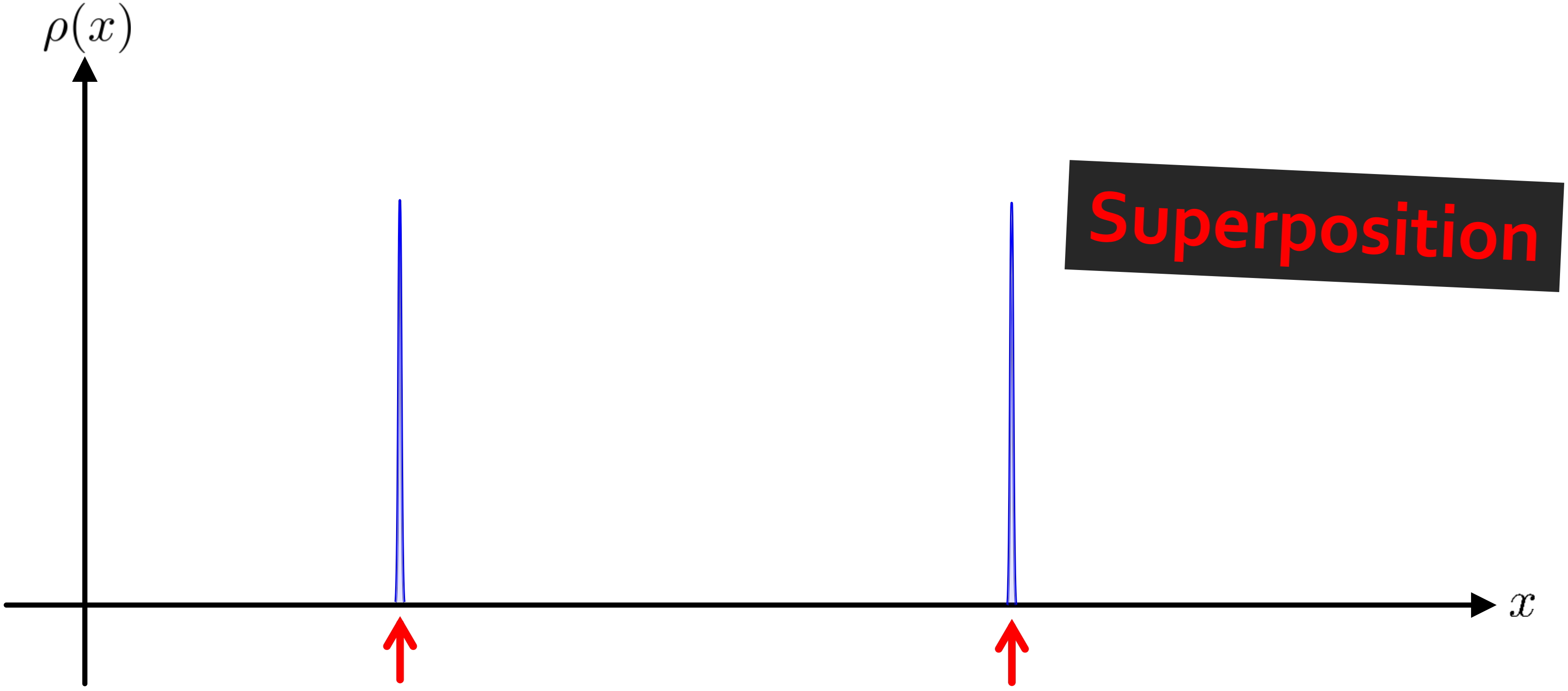
"Somewhere over here."

Probability as Geometry



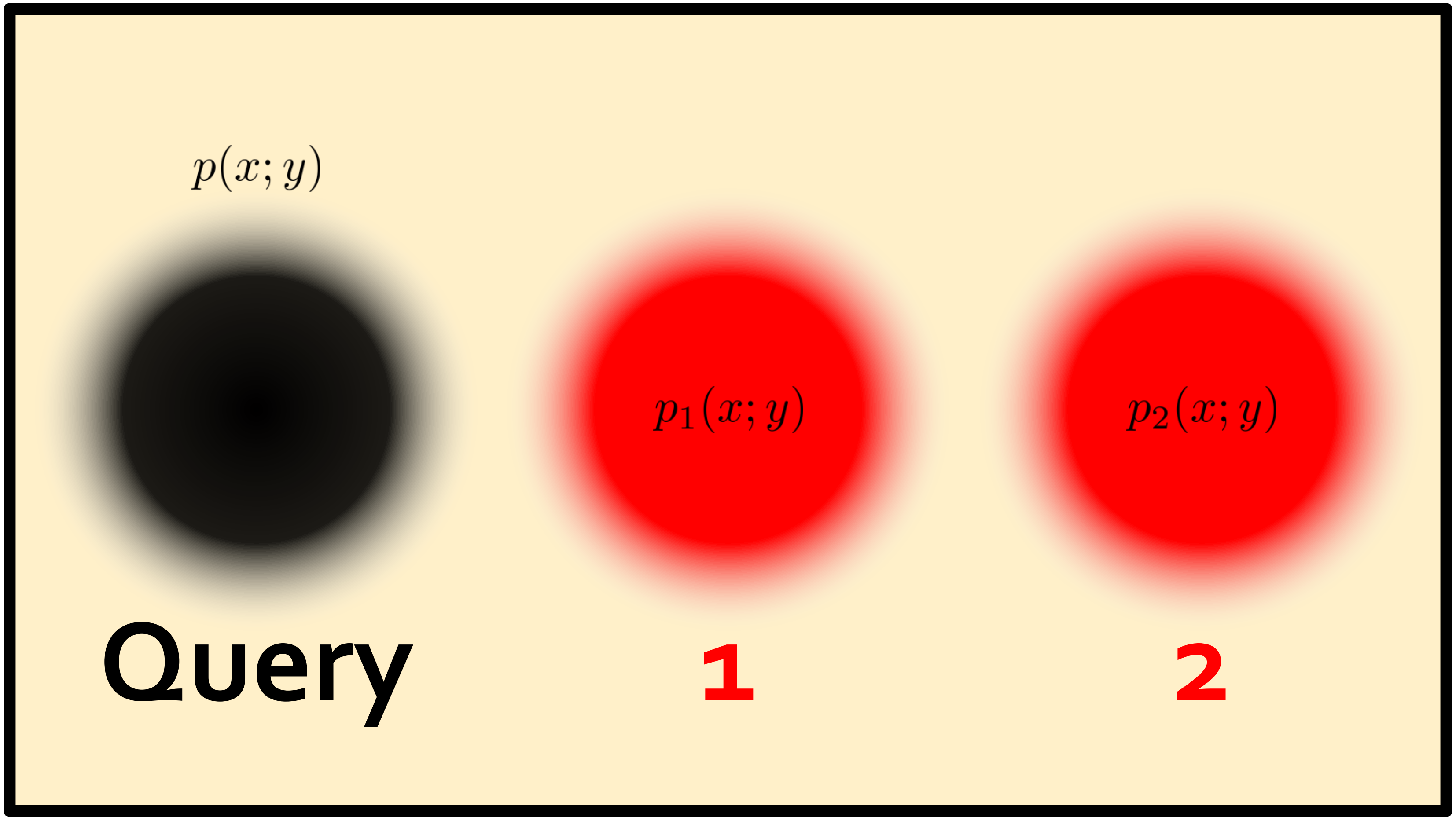
"Exactly here."

Probability as Geometry



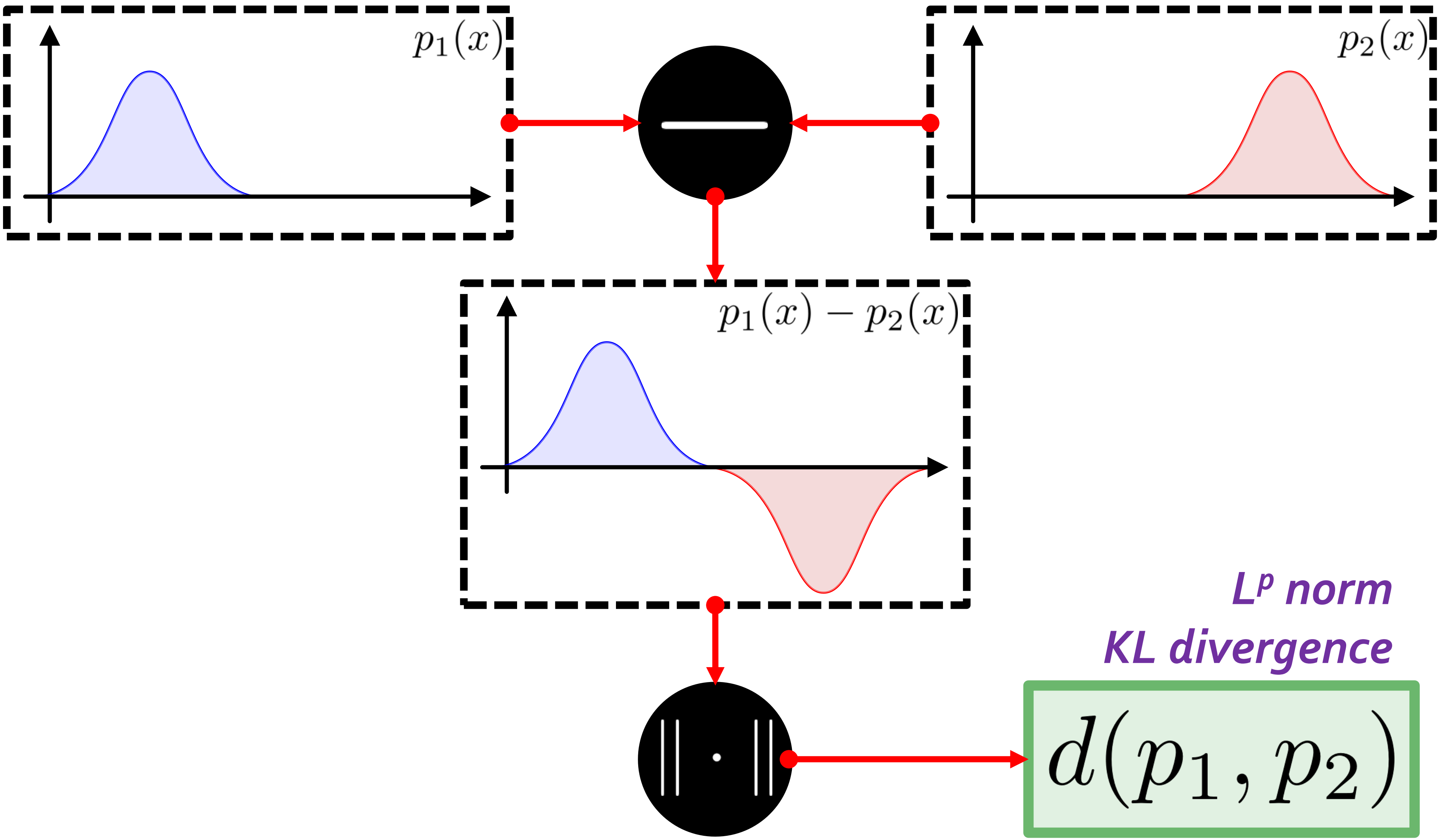
"One of these two places."

Fuzzy Geometry

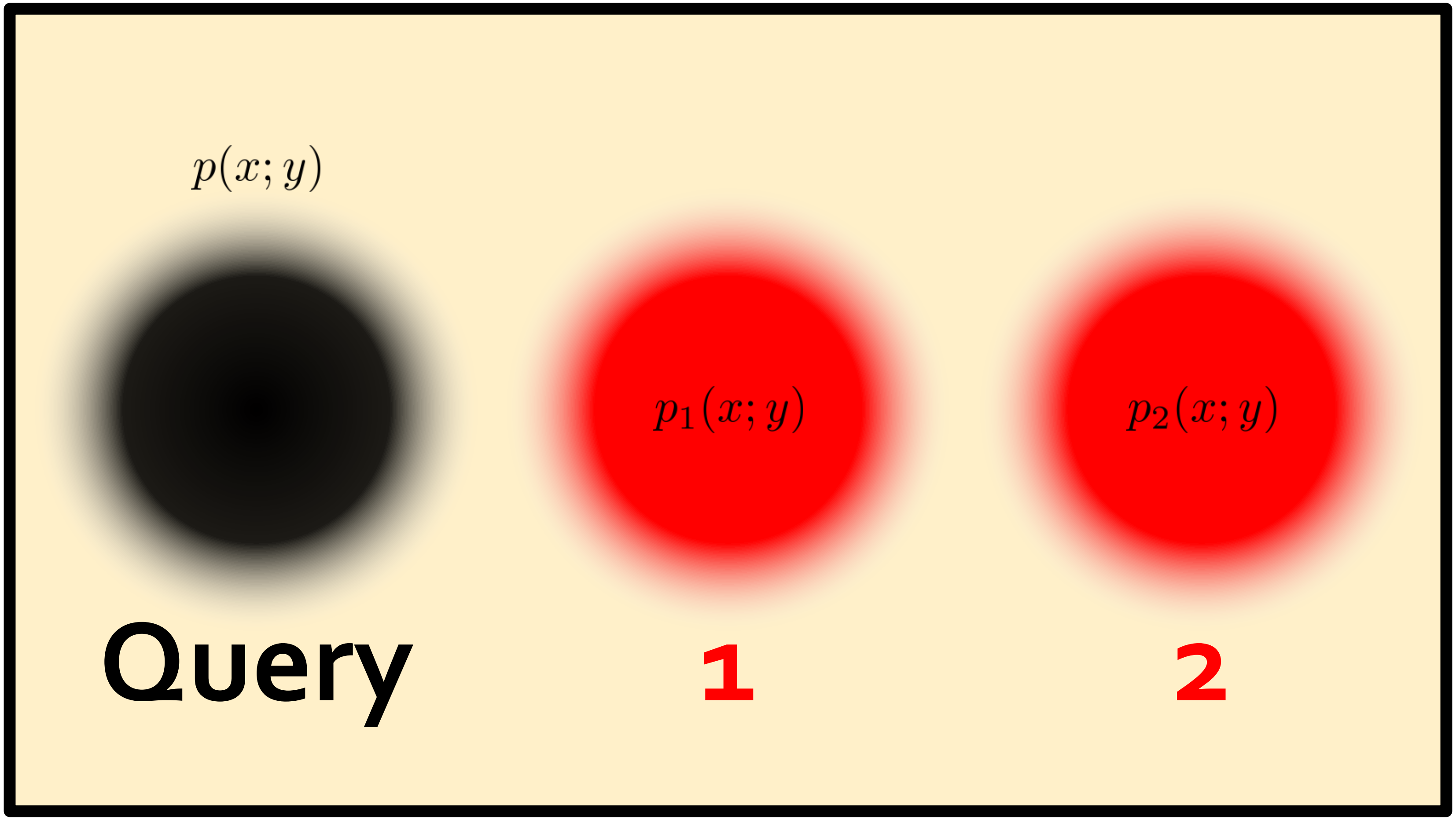


Which is closer, 1 or 2?

Typical Measurement

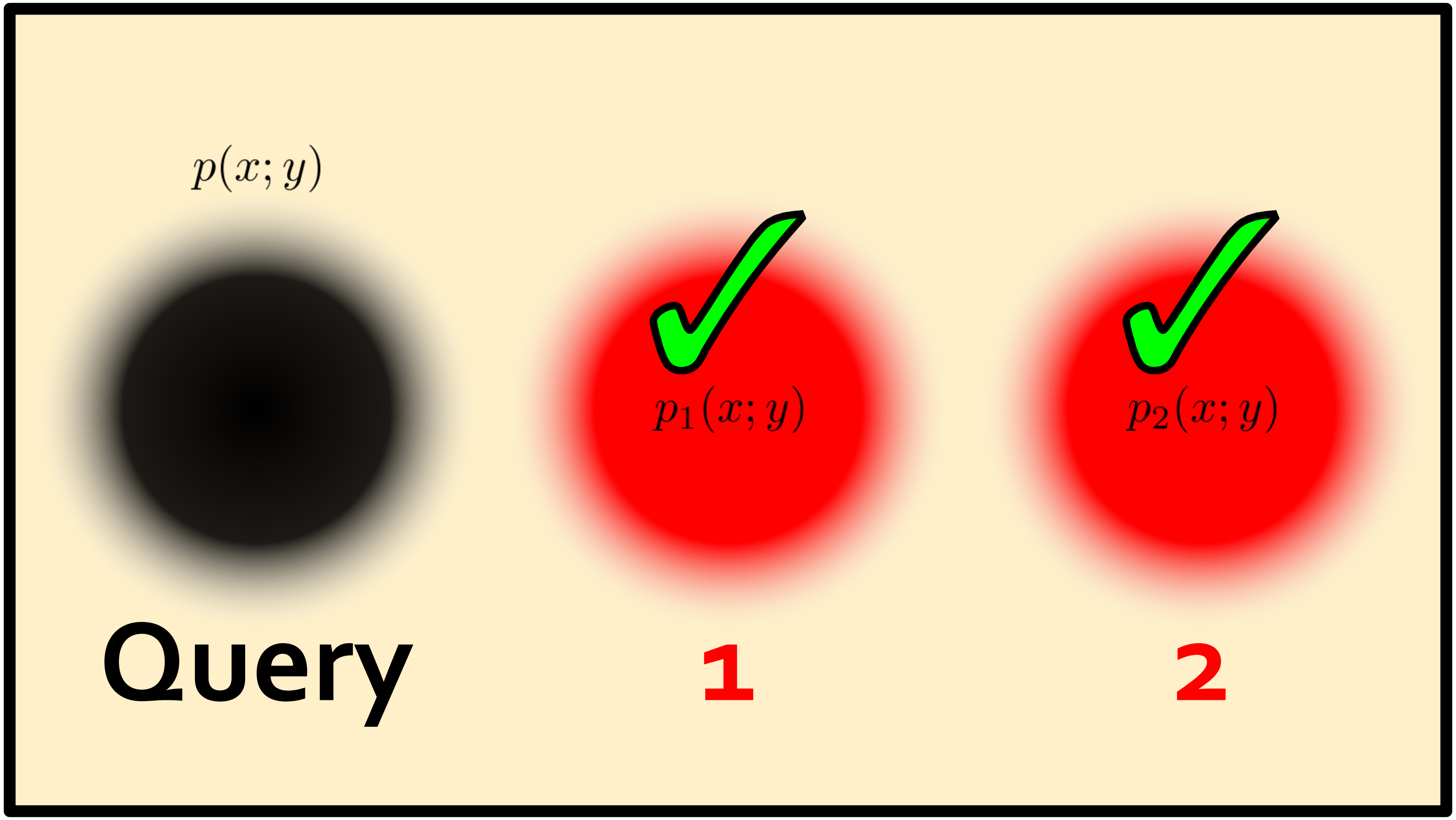


Returning to the Question



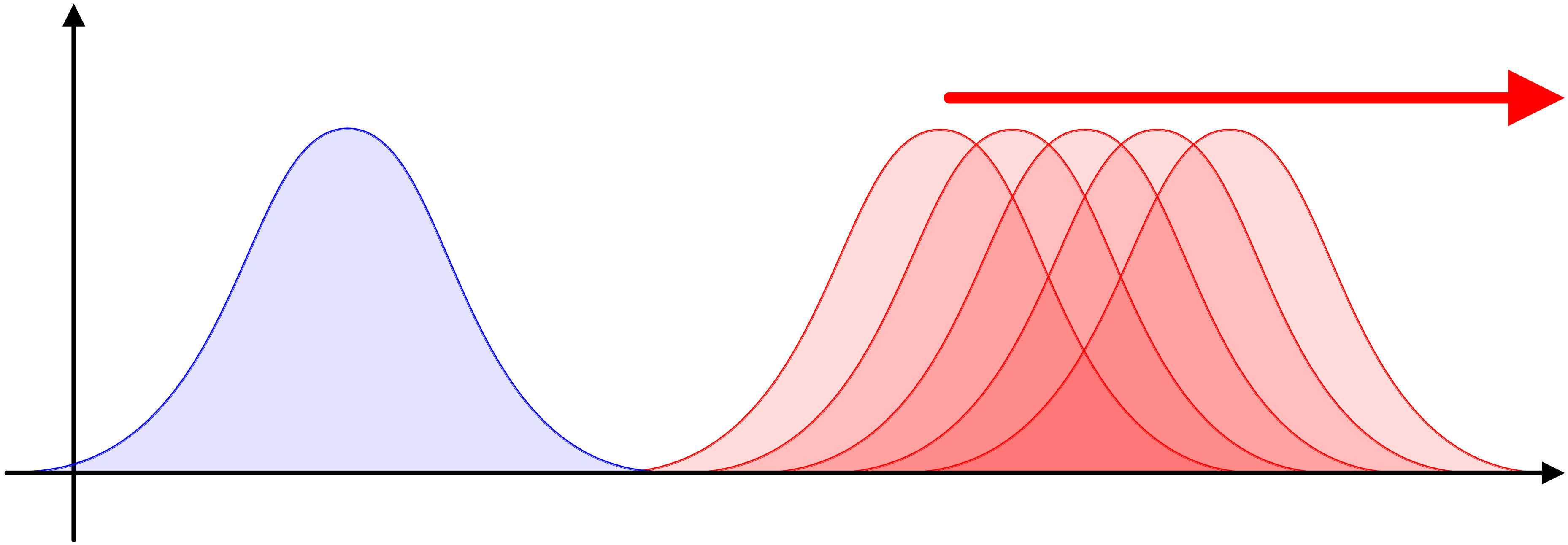
Which is closer, 1 or 2?

Returning to the Question



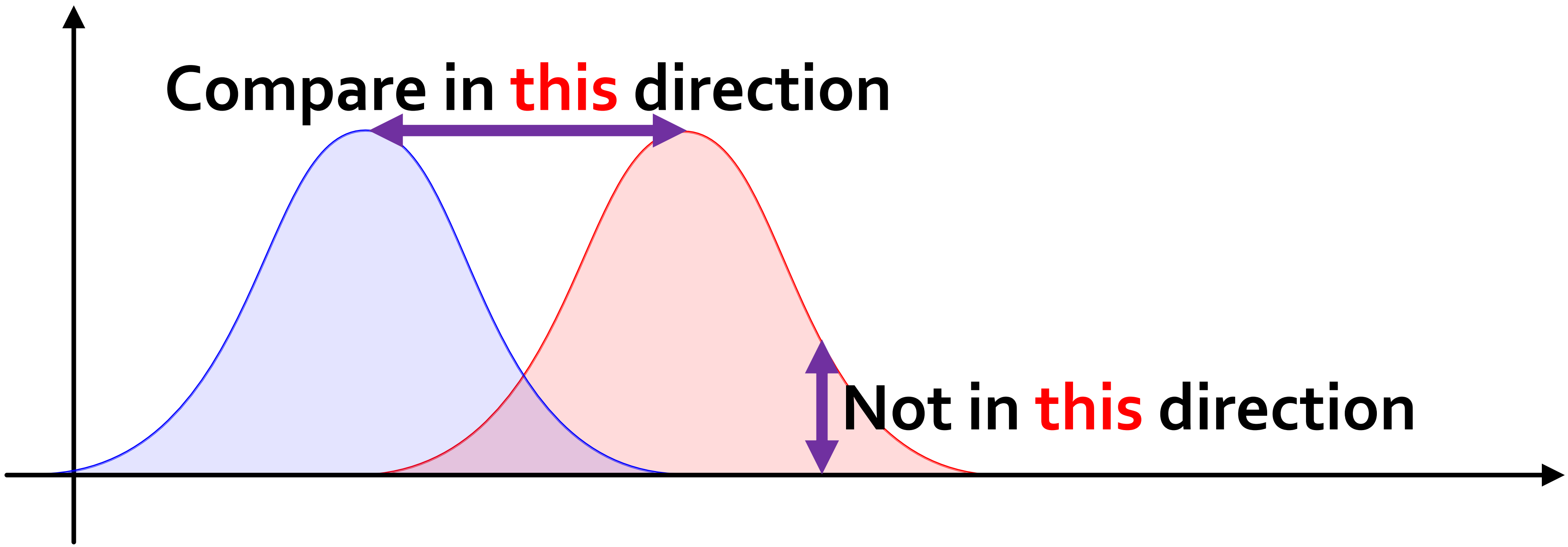
Neither! Equidistant.

What’s Wrong?



Measured overlap, not displacement.

Alternative Idea



Observation

Even the laziest shoveler **must do some work.**

Property of the distributions themselves!

My house!

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

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

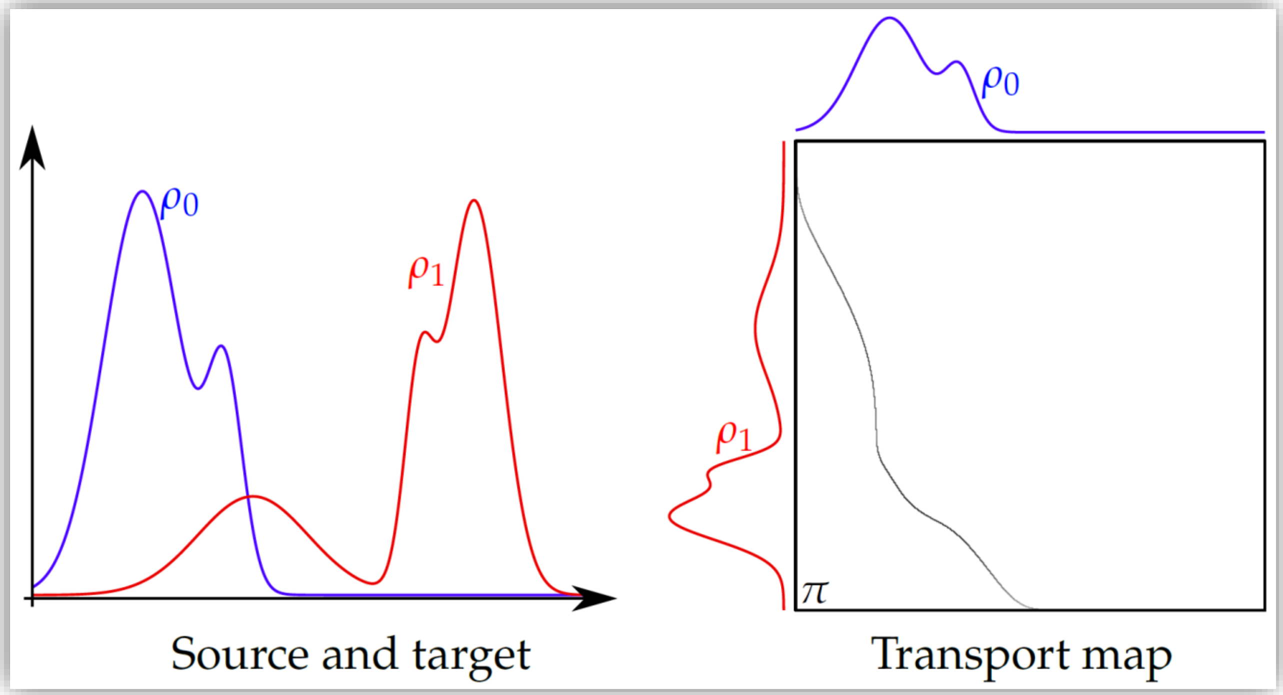
Must scoop everything up

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

Must cover the target

1-Wasserstein in 1D

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



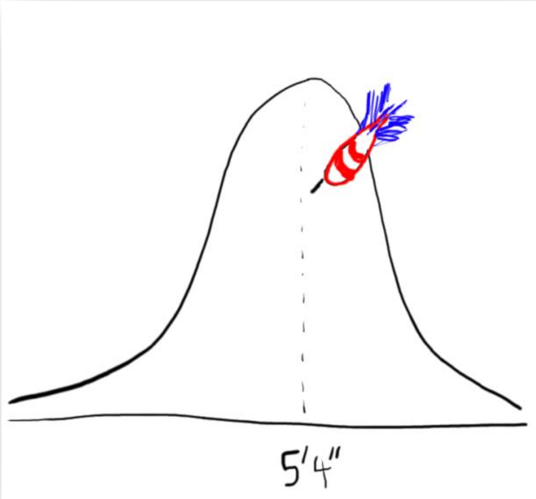
Q:

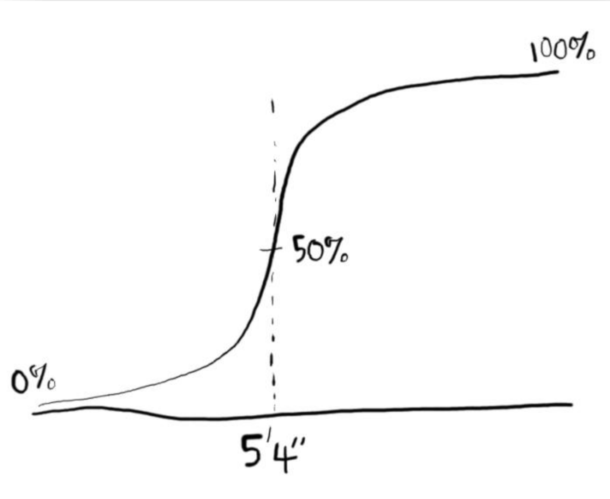
When is transport computable?

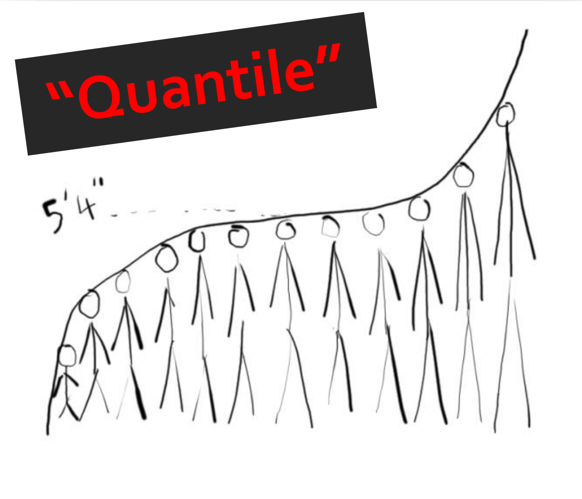
Needed: Finite number of unknowns.

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!

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

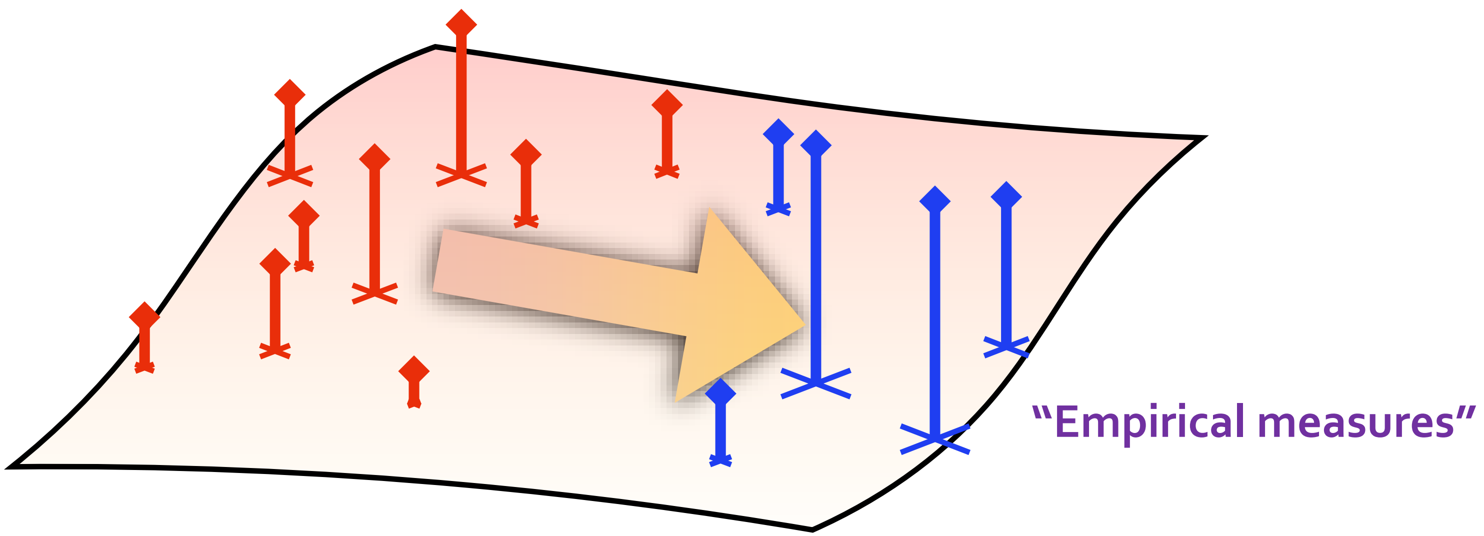
$$\mathcal{W}_2^2(\mu,\nu) = \int_{-\infty}^{\infty} \left(\mathrm{CDF}^{-1}(\mu) - \mathrm{CDF}^{-1}(\nu)\right)^2\, d\ell$$

Fully-Discrete Transport

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

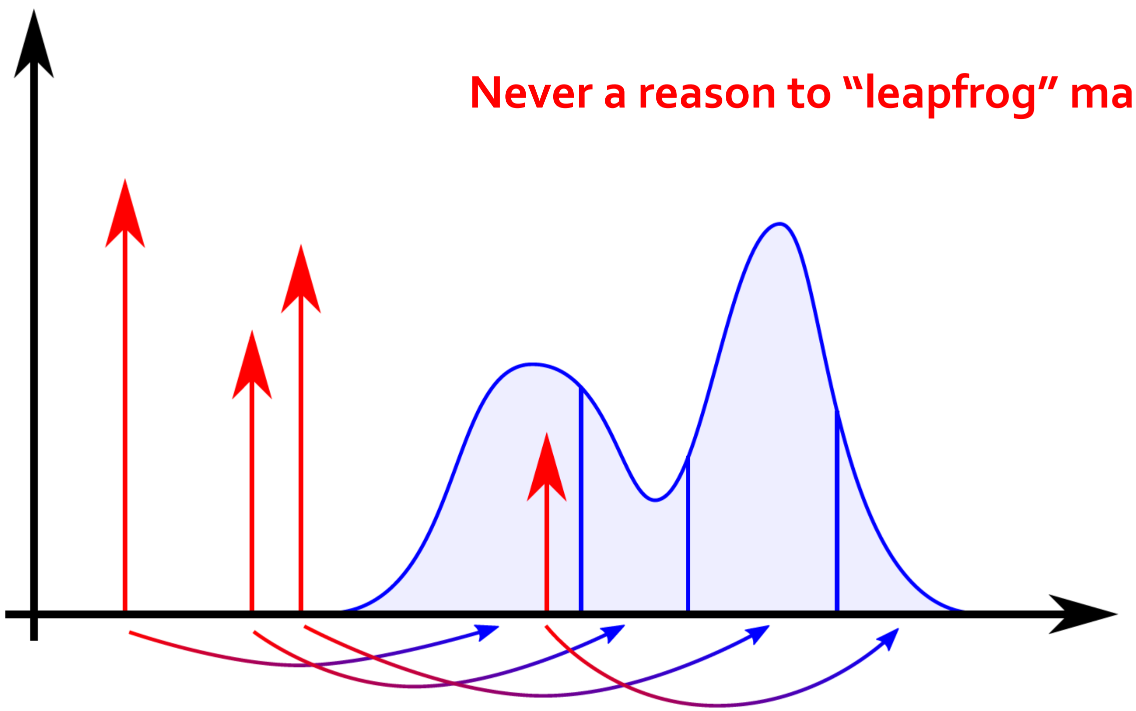
Linear program: Finite number of variables

Algorithms: Simplex, interior point, auction, …



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!

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 & frontiers

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

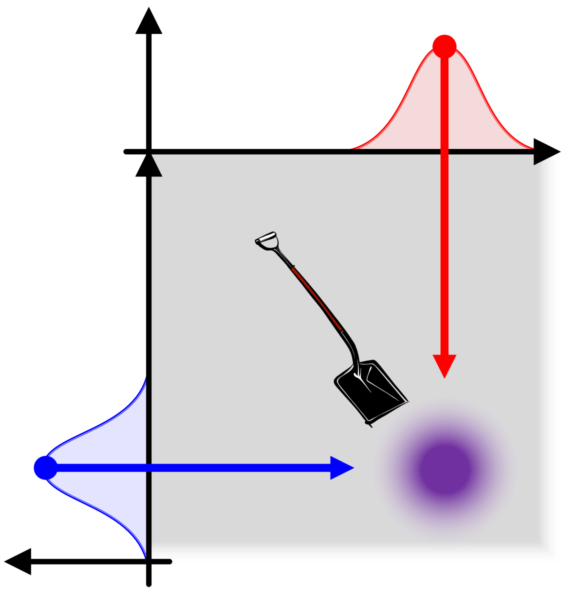
"$\mathrm{Prob}(X)$"

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

Flow-Based W_2

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

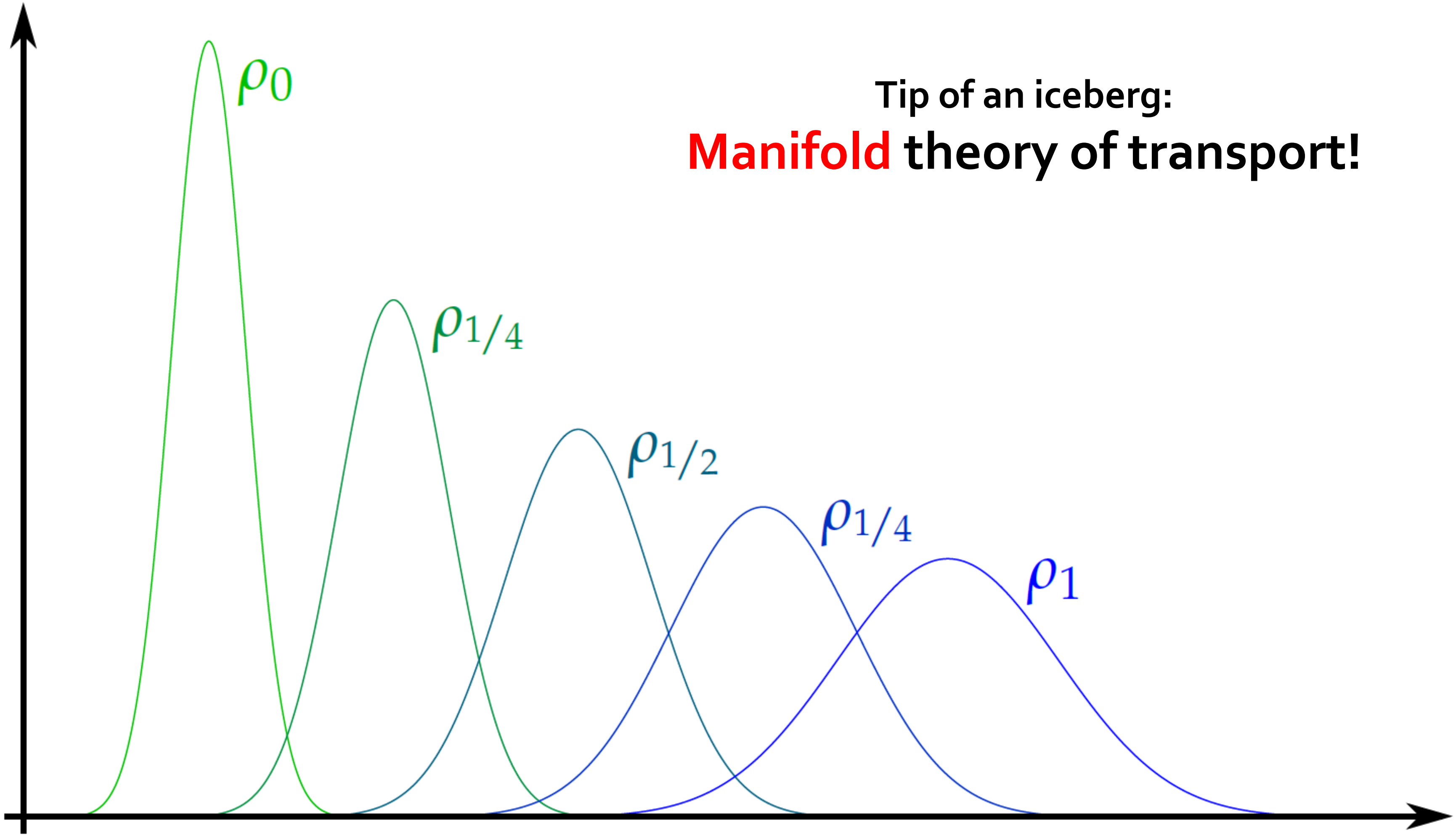
Benamou & Brenier

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

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

Displacement Interpolation

Tip of an iceberg:
Manifold theory of transport!



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 & frontiers

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.

Redistricting?

Balanced power diagrams for redistricting

Vincent Cohen-Addad* Philip N. Klein† Neal E. Young‡

January 6, 2018

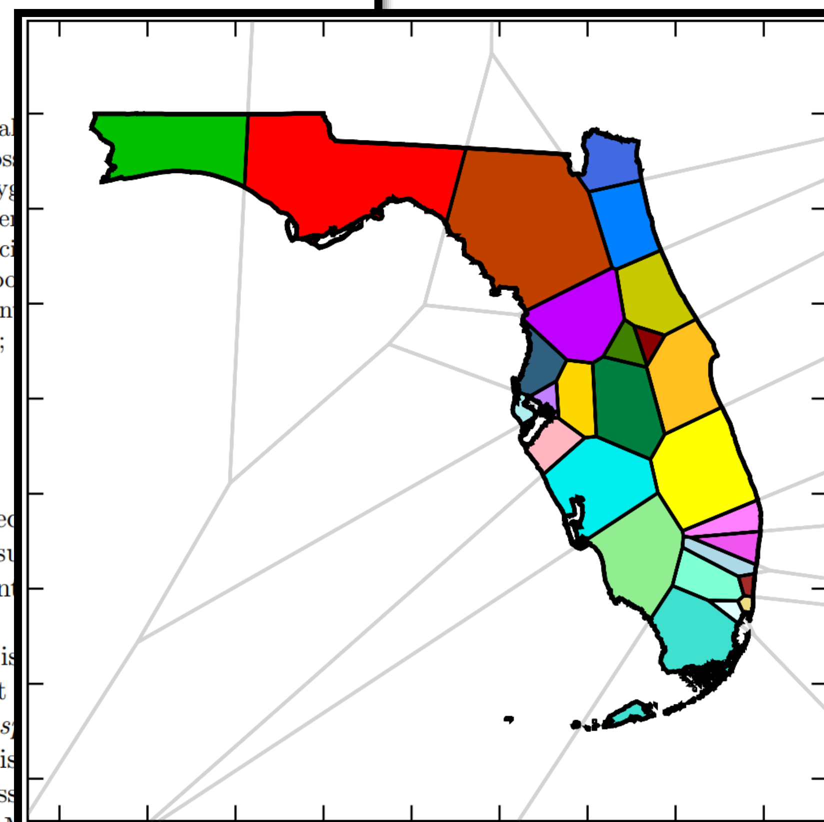
Abstract

We explore a method for *redistricting*, decomposing a geographical
districts, so that the populations of the districts are as close as pos
compact and contiguous. Each district is the intersection of a poly
area. The polygons are convex and the average number of sides pe
The polygons tend to be quite compact. With each polygon is associ
is the centroid of the locations of the residents associated with the po
be viewed as a heuristic for finding centers and a balanced assignmen
as to minimize the sum of squared distances of residents to centers;
said to have low dispersion.

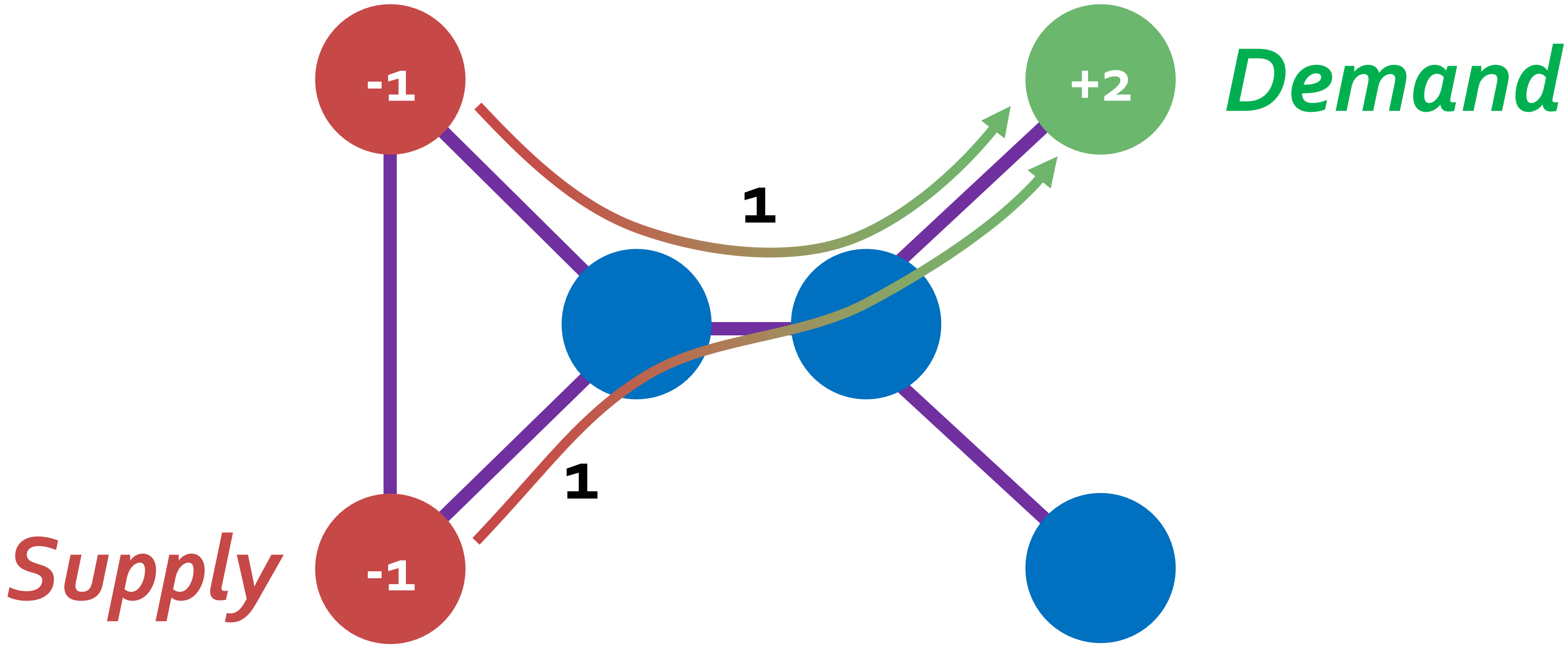
1 Introduction

Redistricting. *Redistricting*, in the context of elections refers to de
into subareas such that all subareas have the same population. The su
most US states, districts are supposed to be *contiguous* to the exten
can reasonably be interpreted to mean *connected*.

In most states, districts are also supposed to be *compact*. This is
Some measures of compactness are based on boundaries; a district
are simpler rather than contorted. Some measures are based on *dis*
the district spreads from a central core" [17]. Idaho directs its redis
drawing districts that are oddly shaped." Other states loosely address
"Arizona and Colorado focus on contorted boundaries; California, M
on dispersion; and Iowa embraces both" [17].

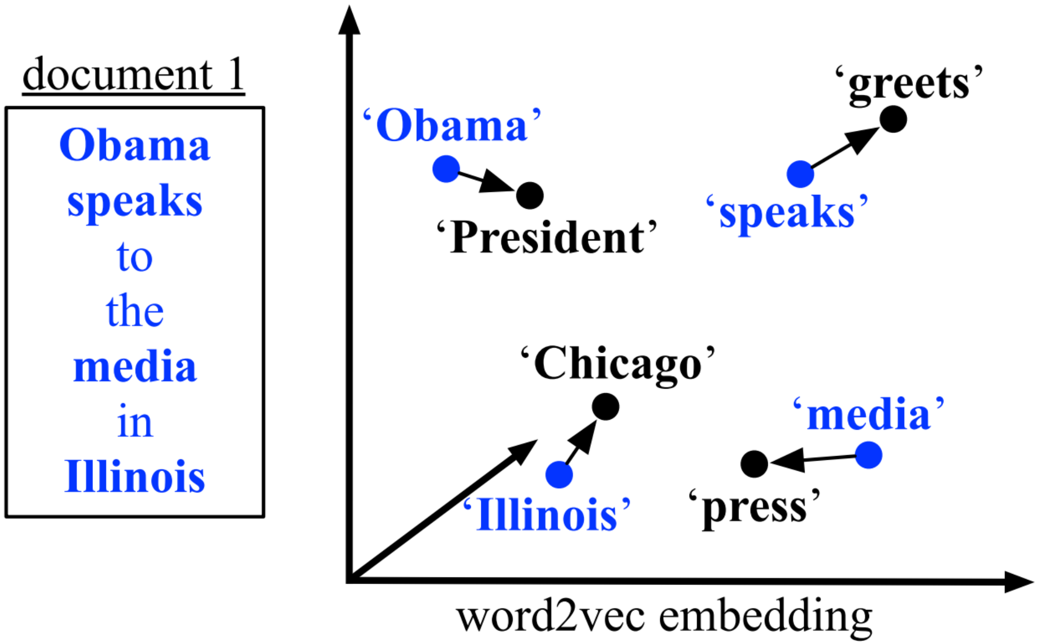


Operations and Logistics



Minimum-cost flow

Histograms and Descriptors

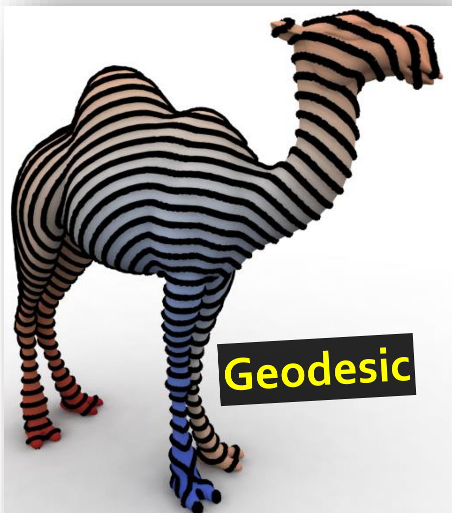


Use deep network embedding

[Kusner et al. 2015]

Word Mover's Distance (WMD)

Distance Approximation

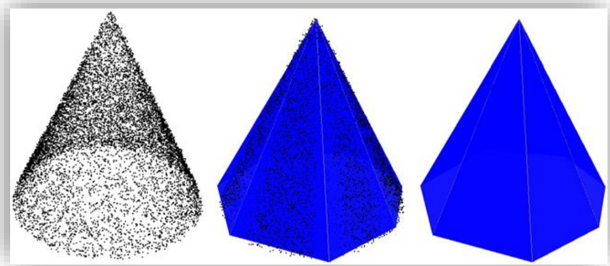


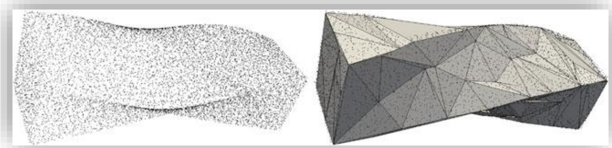
0 eigenfunctions

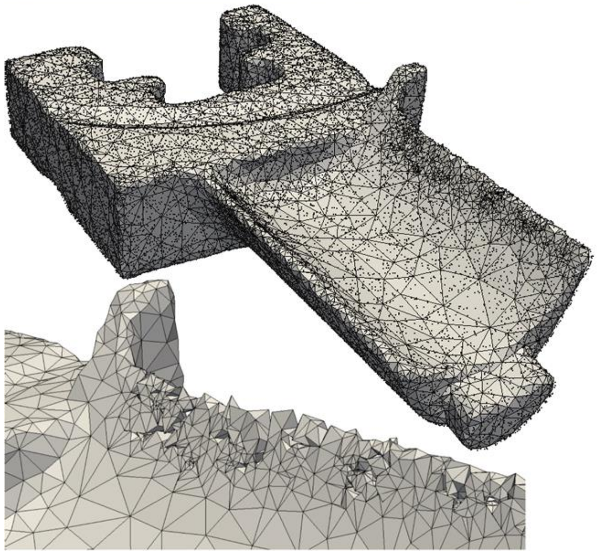
100 eigenfunctions

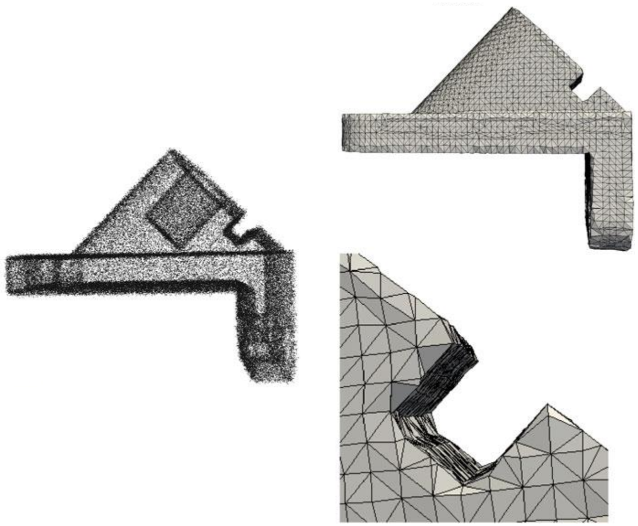
Proposition: Satisfies triangle inequality.

Registration and Reconstruction









[Digne et al. 2014]

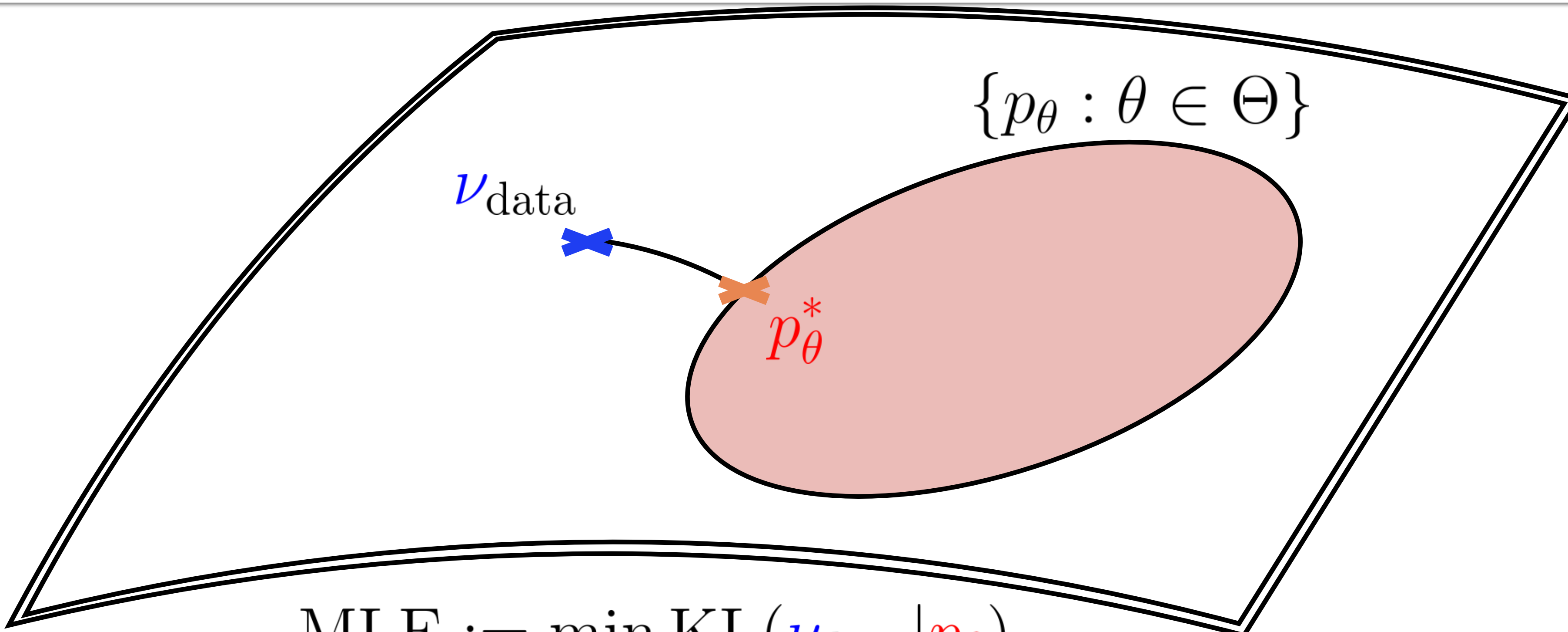
Distance from point cloud to mesh

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

Statistical Estimation



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

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

[Bassetti 2006]

Minimum Kantorovich Estimator

Distributionally Robust Optimization

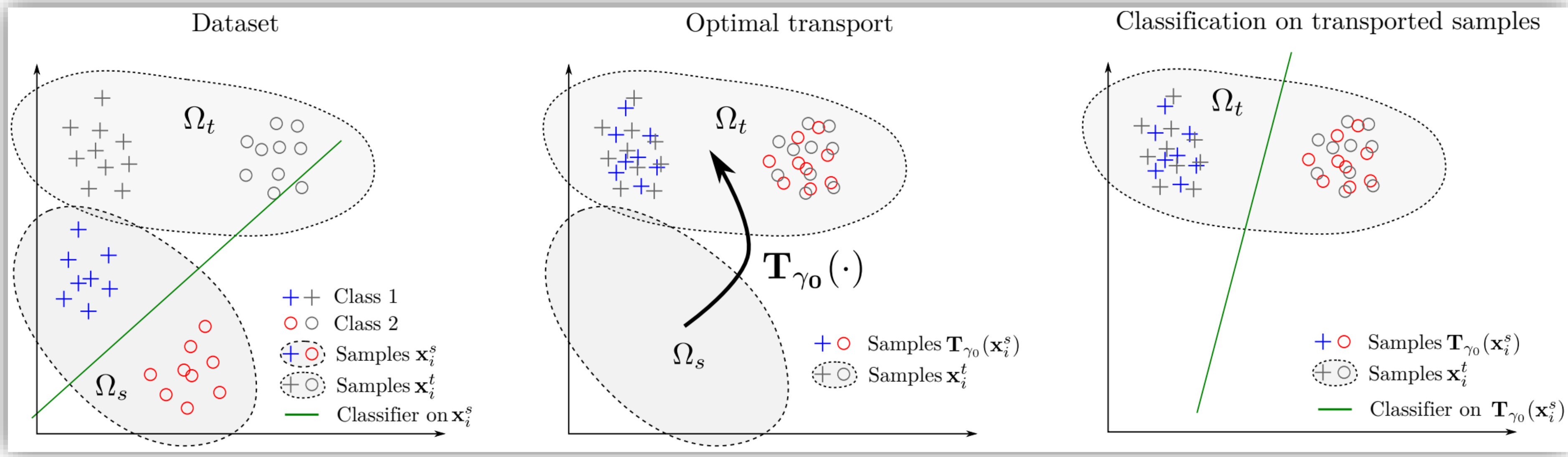
$$\inf_{x \in \mathbb{X}} \sup_{\mathbb{Q} \in \hat{\mathcal{P}}_N} \mathbb{E}_{\xi \sim \mathbb{Q}}[h(x,\xi)]$$

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

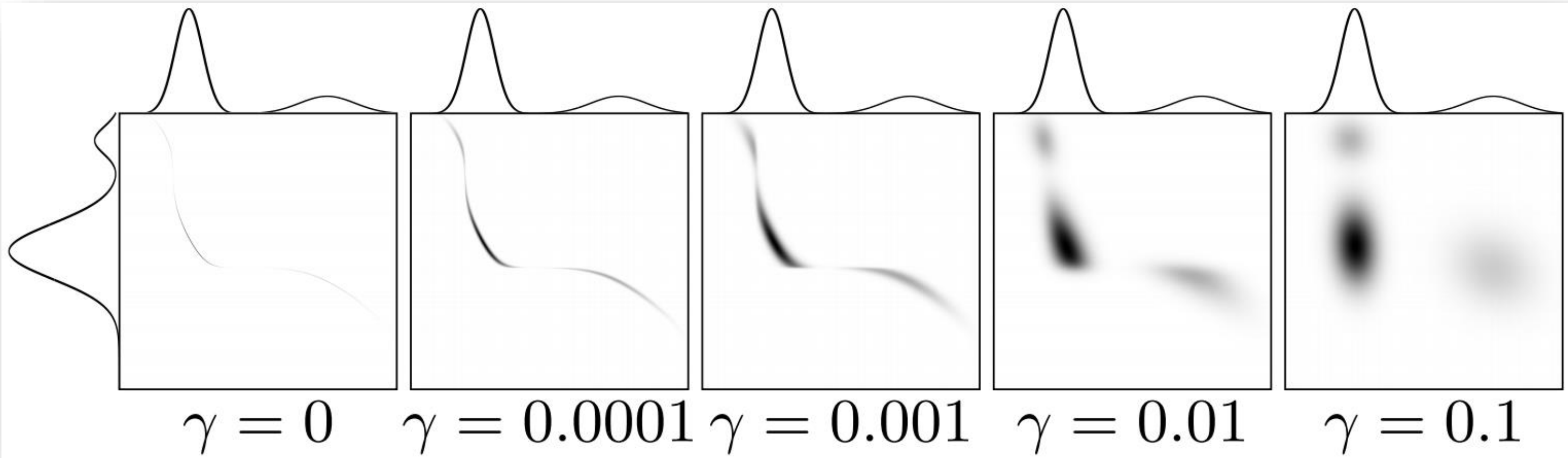
1. Introduction to optimal transport
    - 1D examples
    - Many formulas

2. Applications

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

4. Extensions & frontiers

Entropic Regularization



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

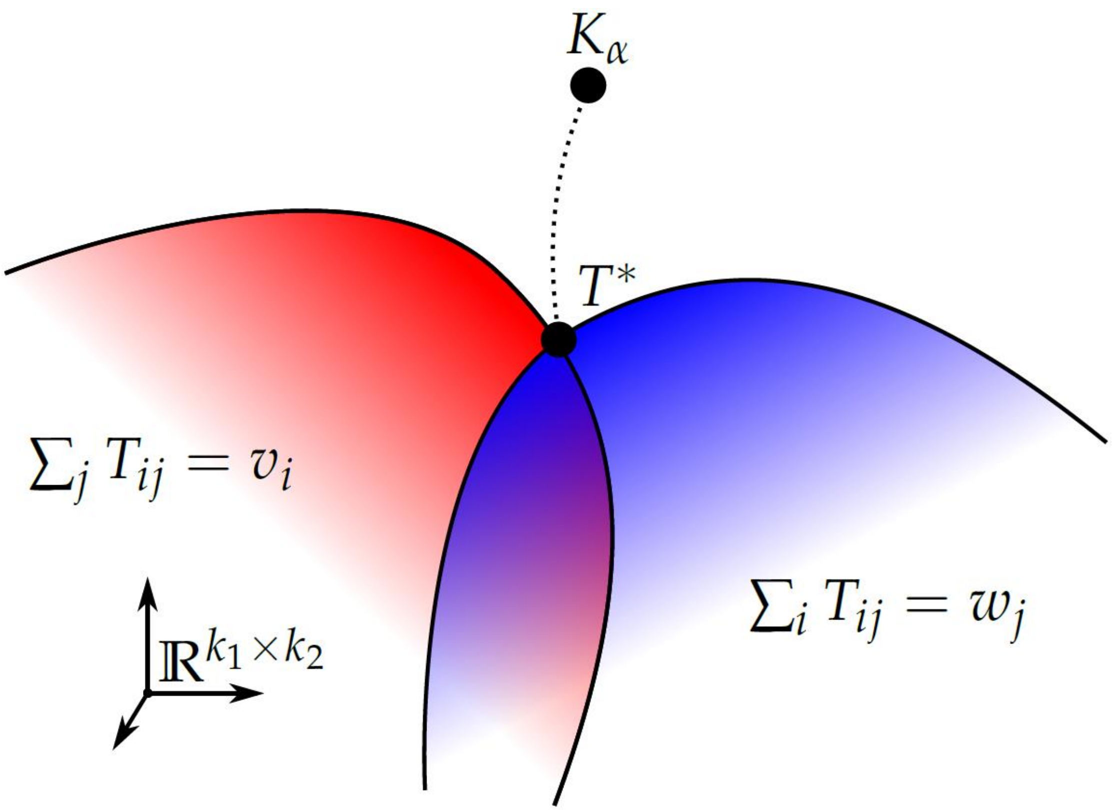
OK to drop nonnegative constraint!

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

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

Interpretation as Projection

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

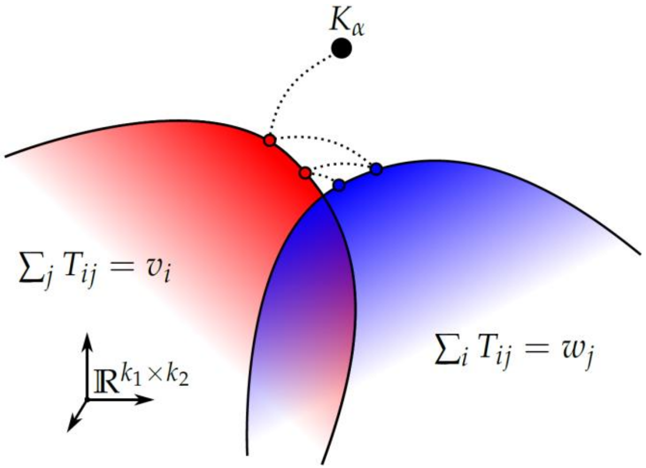


"Iterative Bregman Projections for Regularized Transportation Problems" (Benamou et al. 2014)

Sinkhorn Algorithm

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

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

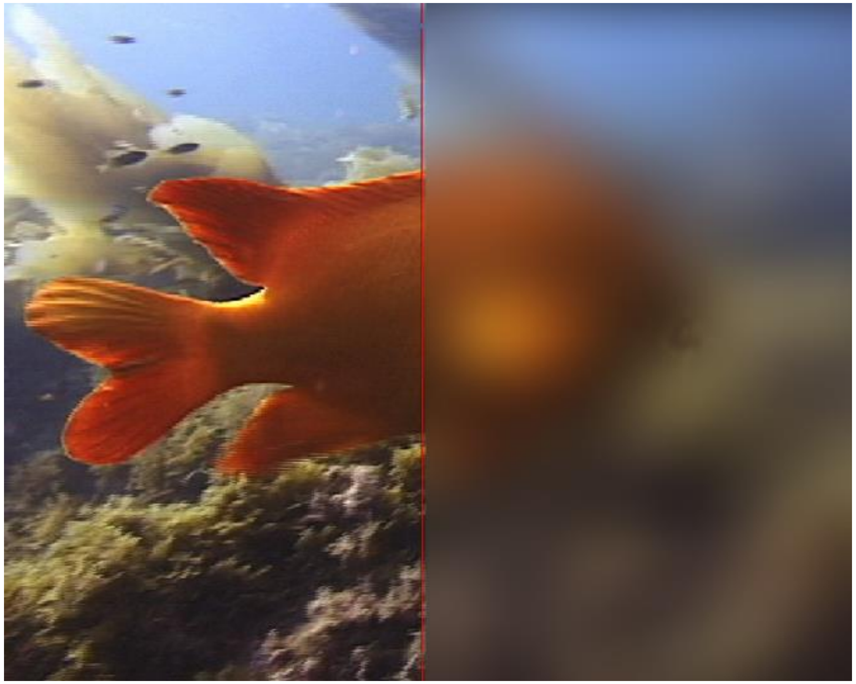


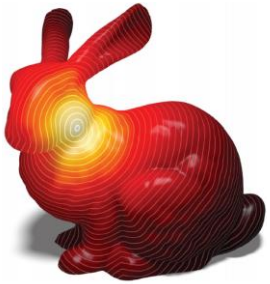
Sinkhorn & Knopp. "Concerning nonnegative matrices and doubly stochastic matrices". Pacific J. Math. 21, 343–348 (1967).

Alternating projection

Ingredients for Sinkhorn

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





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

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

Plan For Today

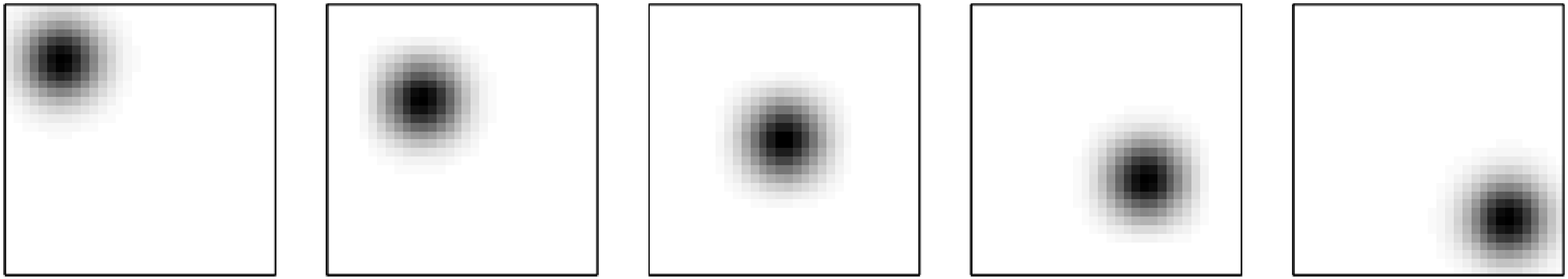
1. Introduction to optimal transport
   - 1D examples
   - Many formulas

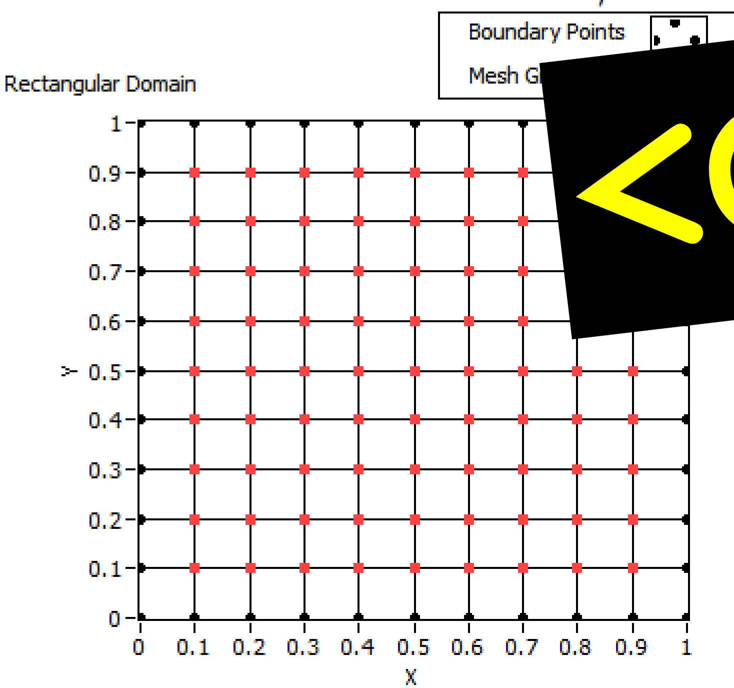
2. Applications

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

4. Extensions & frontiers

Discretization





<Omit>

[redacted] grid point
[redacted]tive function using
[redacted] differences

Open problem:

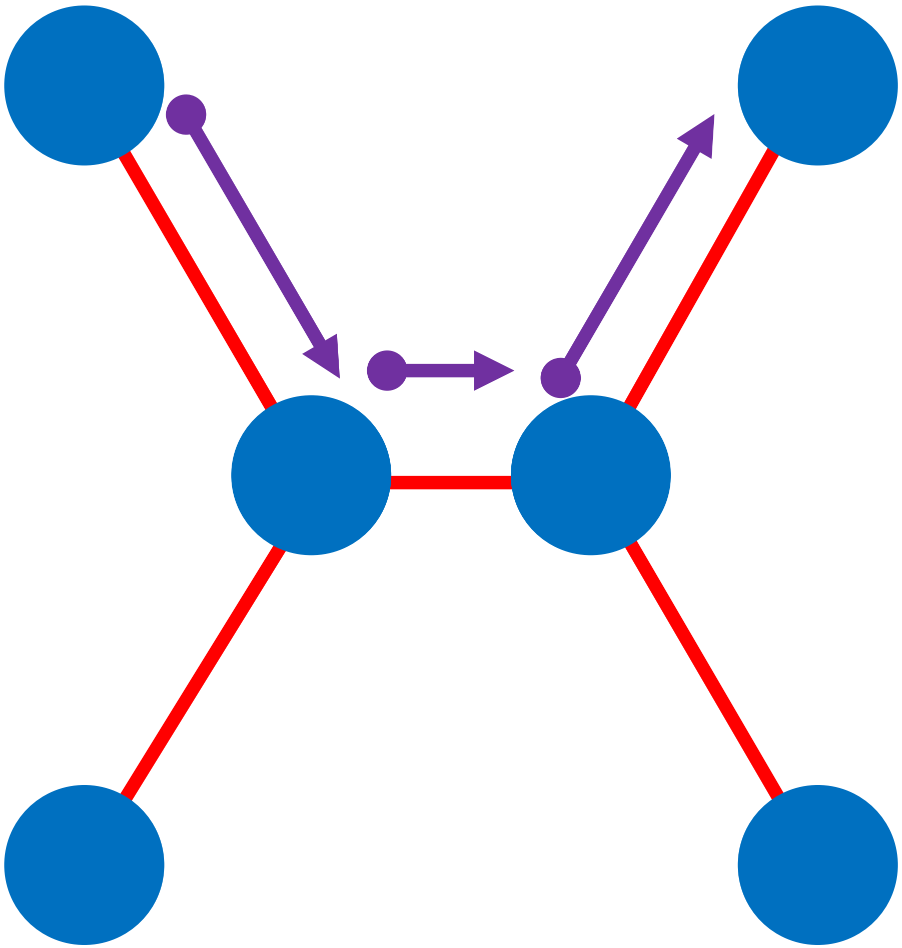
- Discretize on a triangle mesh
- Preserve triangle inequality

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 Formulation

Better scaling for sparse graphs!

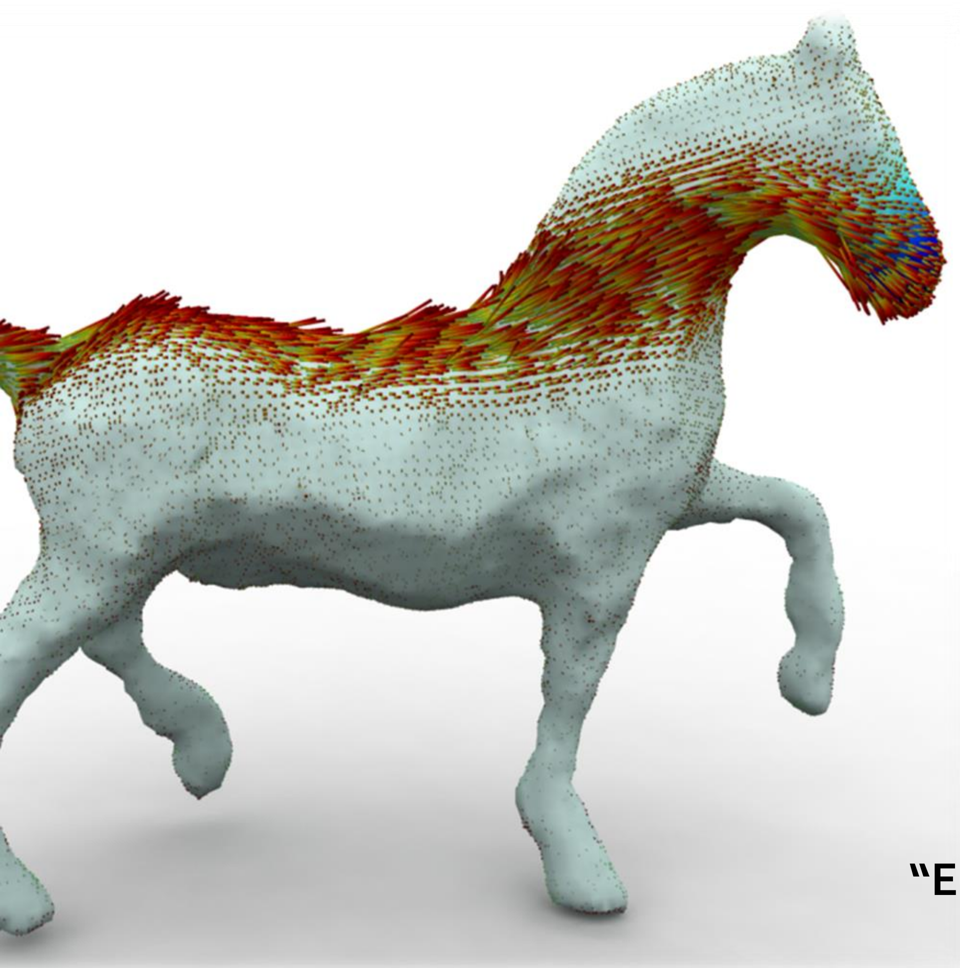


$$\begin{aligned} \min_T \quad & \textstyle\sum_e c_e |J_e| \\ \text{s.t.} \quad & D^\top J = \underbrace{p_1 - p_0}_{f} \end{aligned}$$

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

Plan For Today

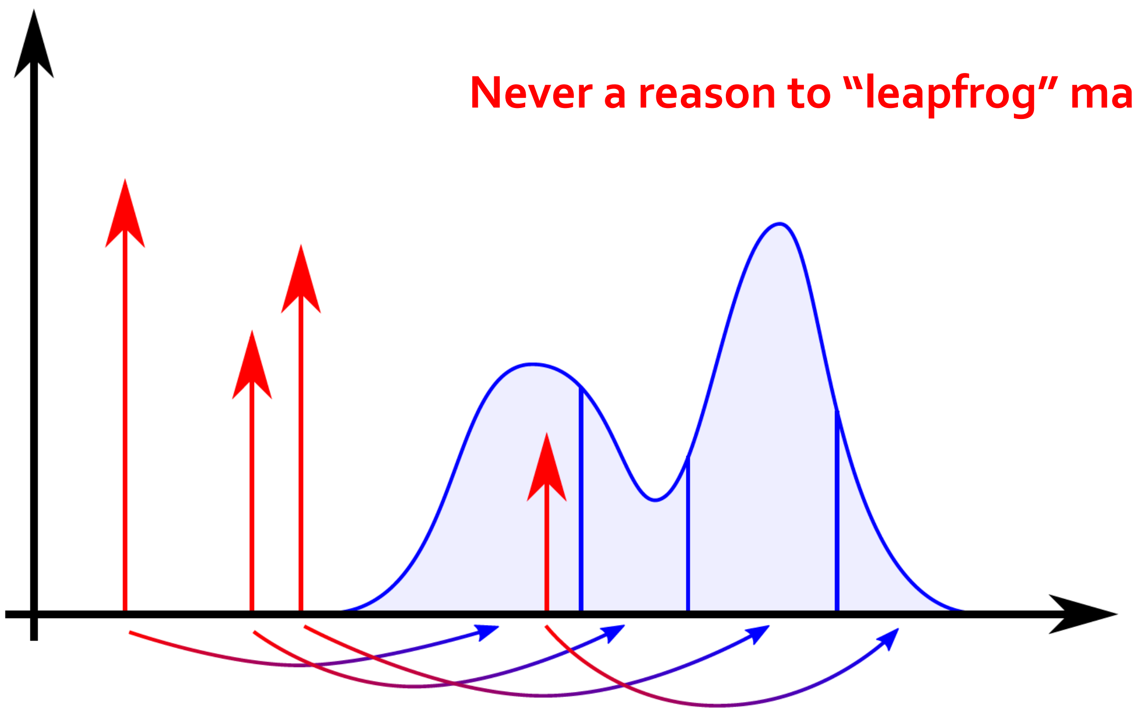
1. Introduction to optimal transport
    - 1D examples
    - Many formulas
2. Applications
3. Discrete/discretized transport
    - Entropic regularization
    - Eulerian transport
    - Semidiscrete transport
4. Extensions & frontiers

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!



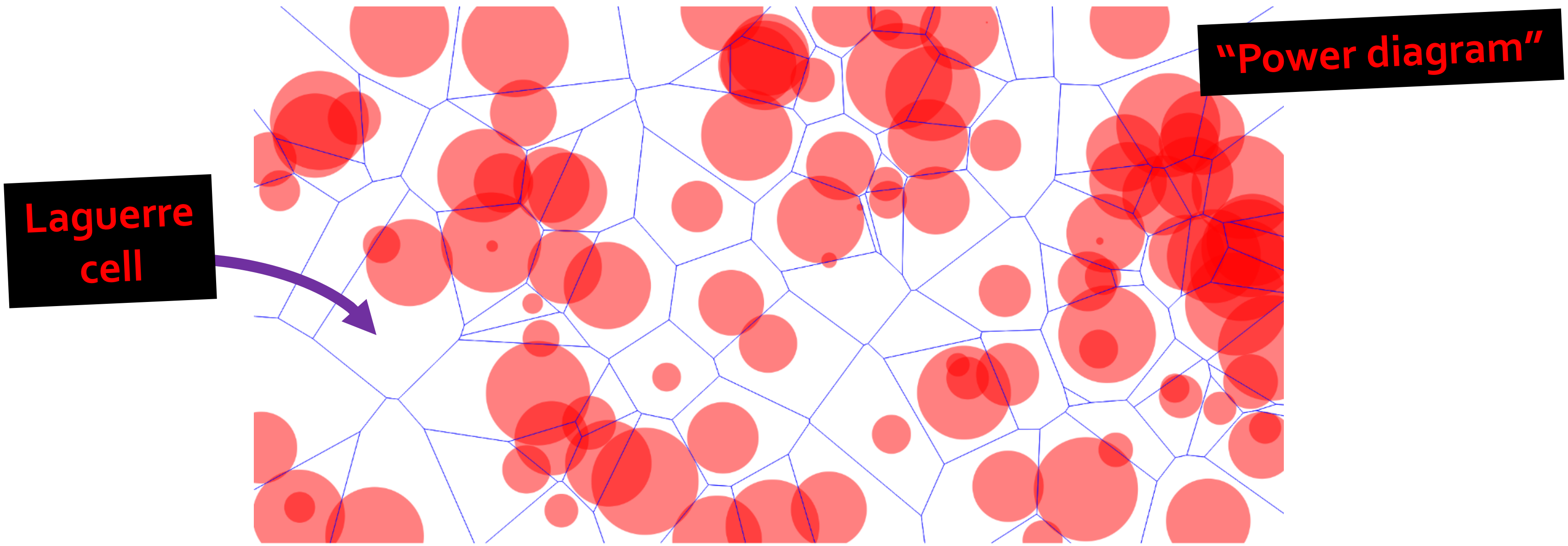
General Case: Semidiscrete

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

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

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

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



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!

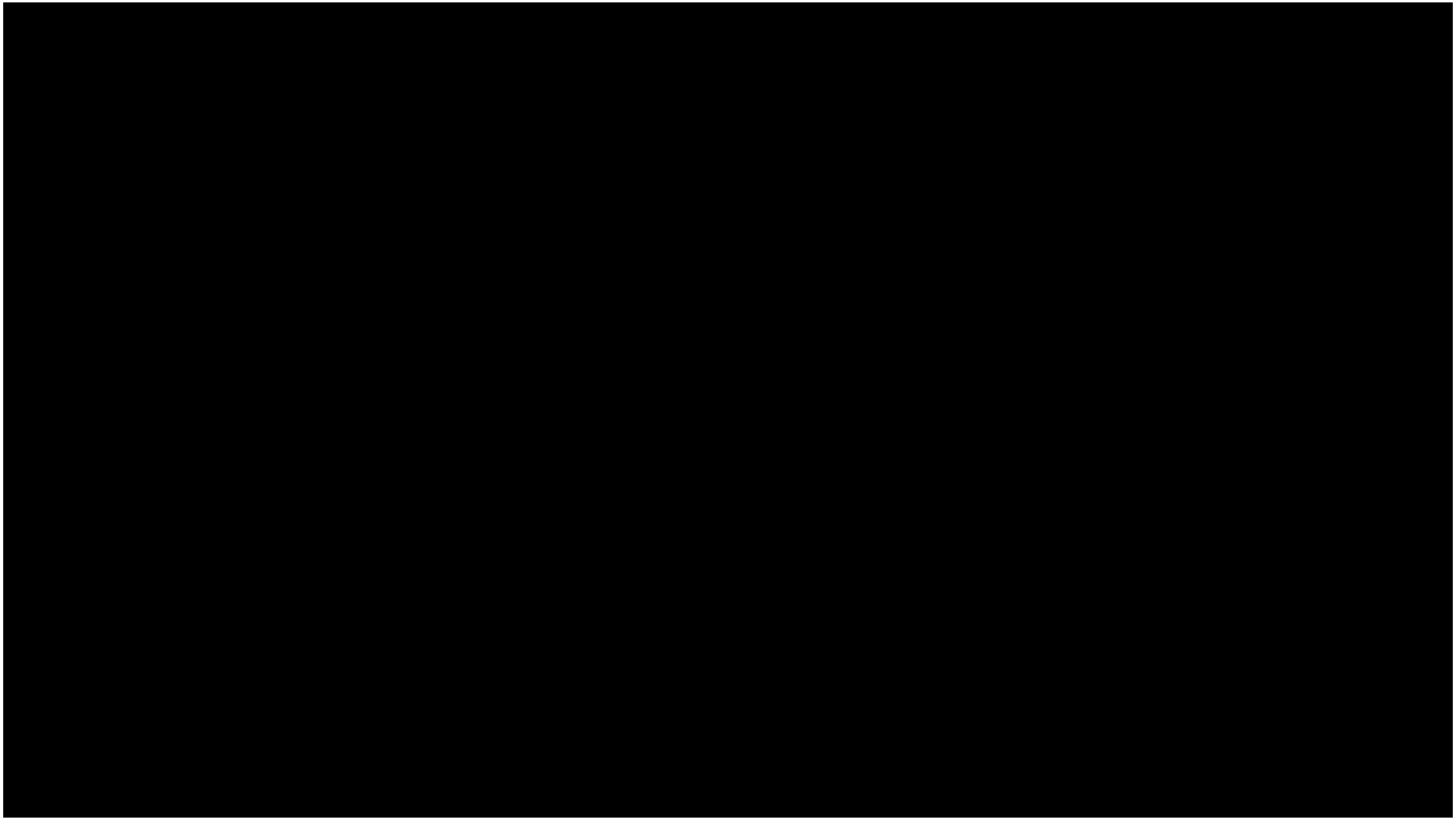
- **Simple algorithm: Gradient ascent**
  Ingredients: Power diagram
- **More complex: Newton's method**
  Converges globally [Kitagawa, Mérigot, & Thibert 2016]

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

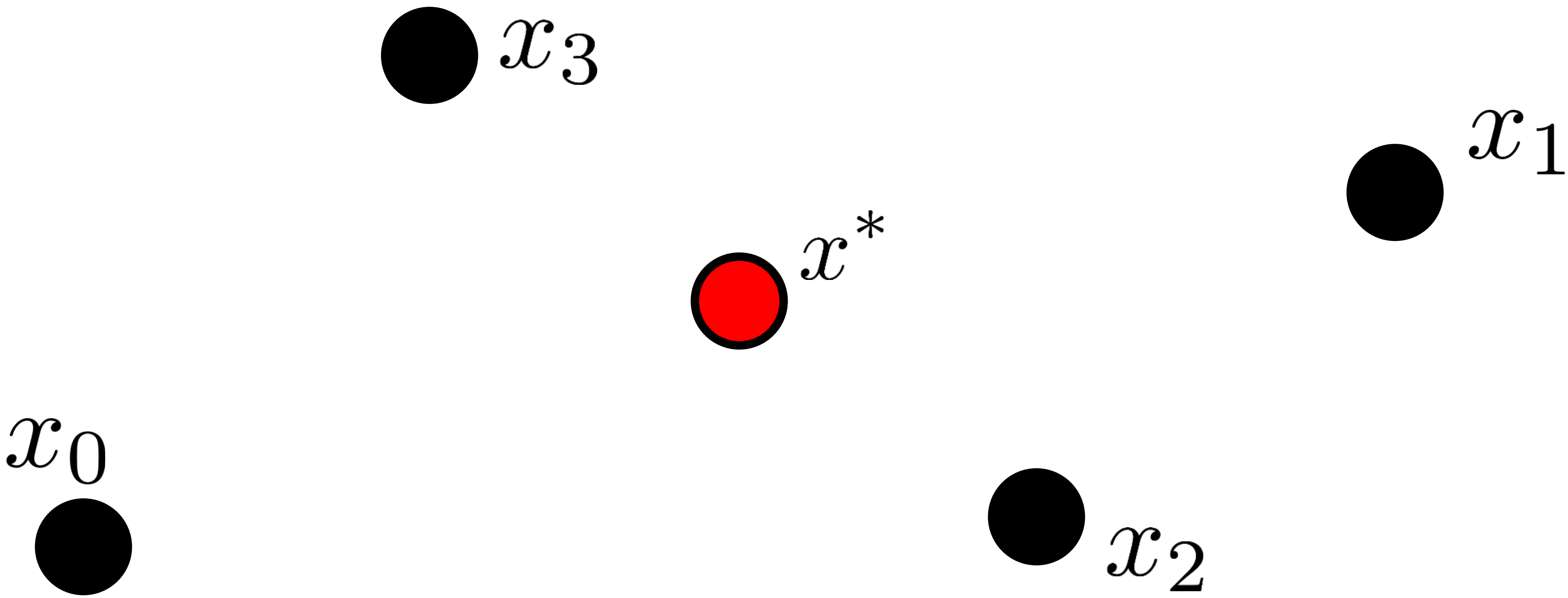
3. Discrete/discretized transport
   - Entropic regularization
   - Eulerian transport
   - Semidiscrete transport

4. Extensions & frontiers

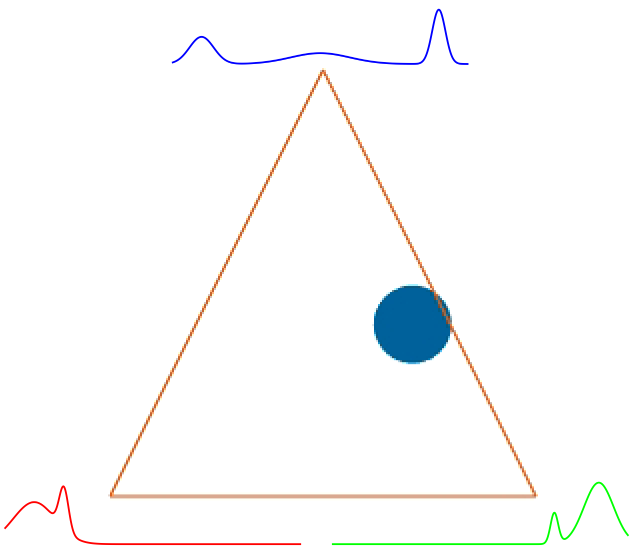
Application 1:

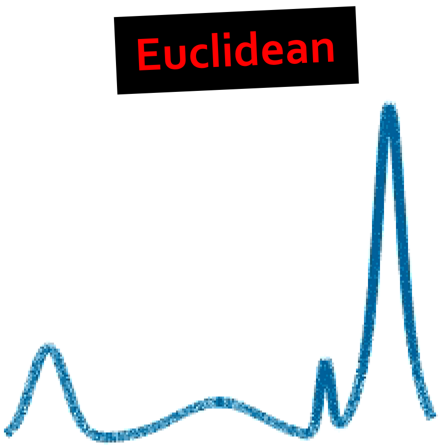
Example: Averaging

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



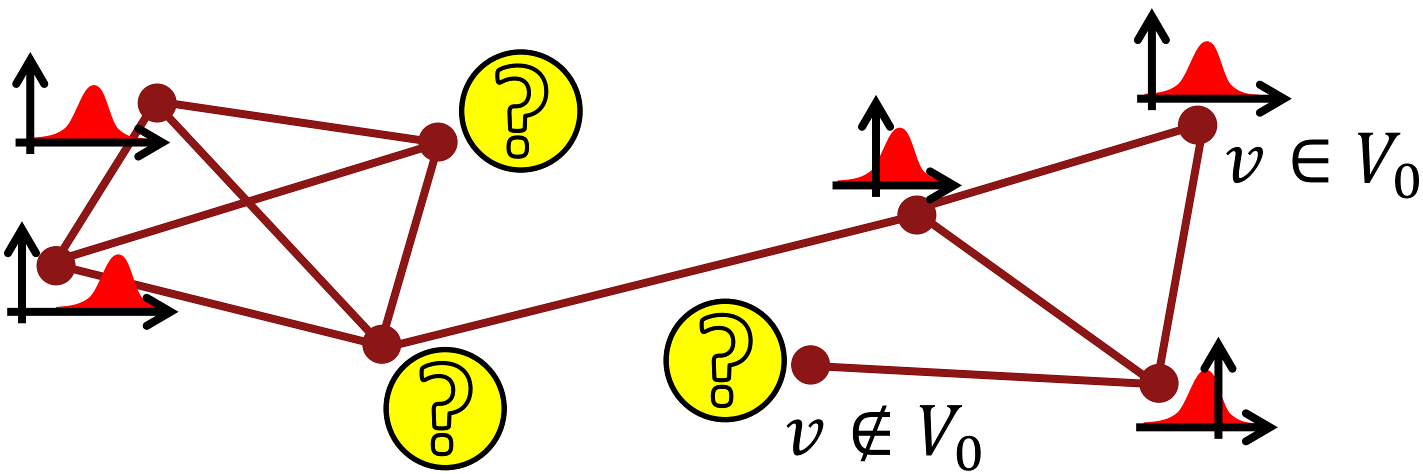
Barycenter Example



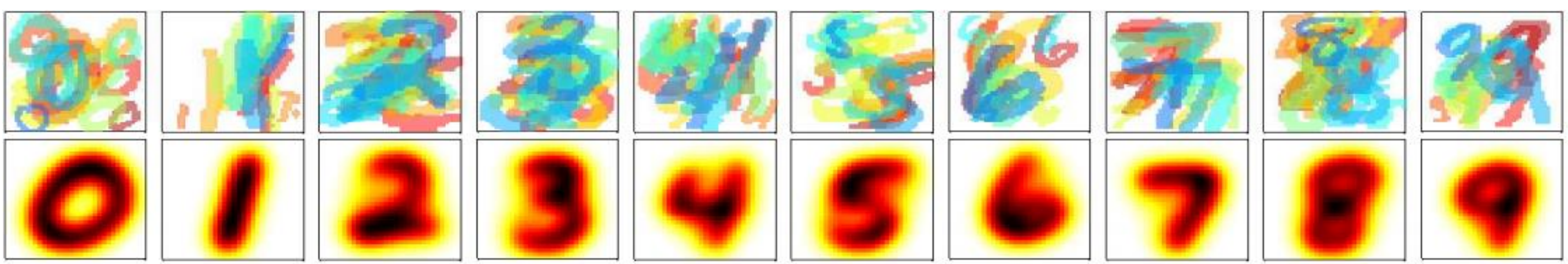


$$\text{Wasserstein:} \quad \mu^* := \left[\arg \min_{\mu \in \mathrm{Prob}(\mathbb{R}^n)} \sum_i \mathcal{W}_2^2(\mu, \mu_i) \right]$$

Barycenters in Machine Learning



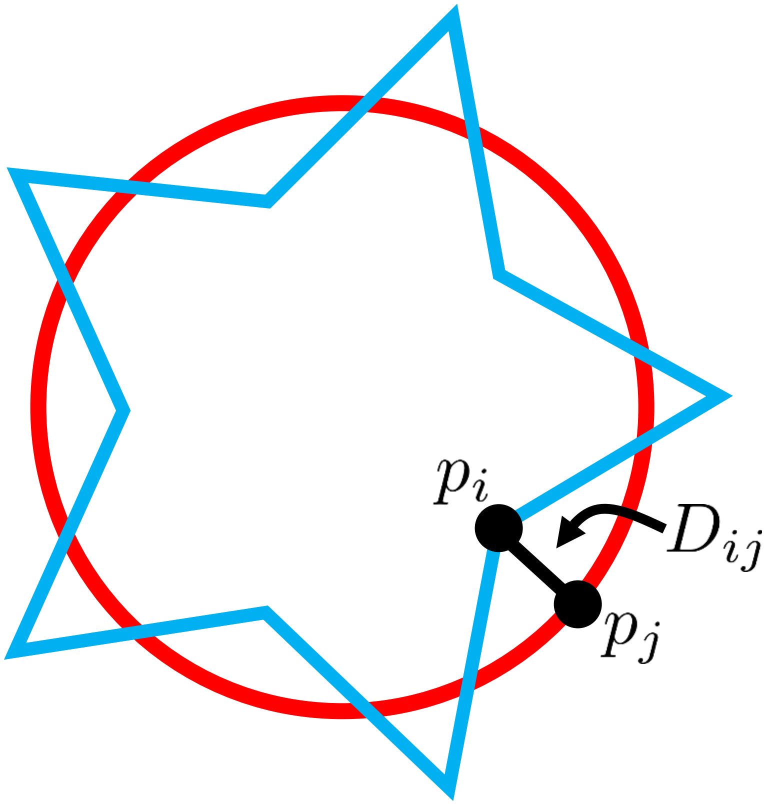
"Wasserstein Propagation for Semi-Supervised Learning" (Solomon et al.)



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

Application 2:

Model Problem: Linear Assignment



Between signals

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

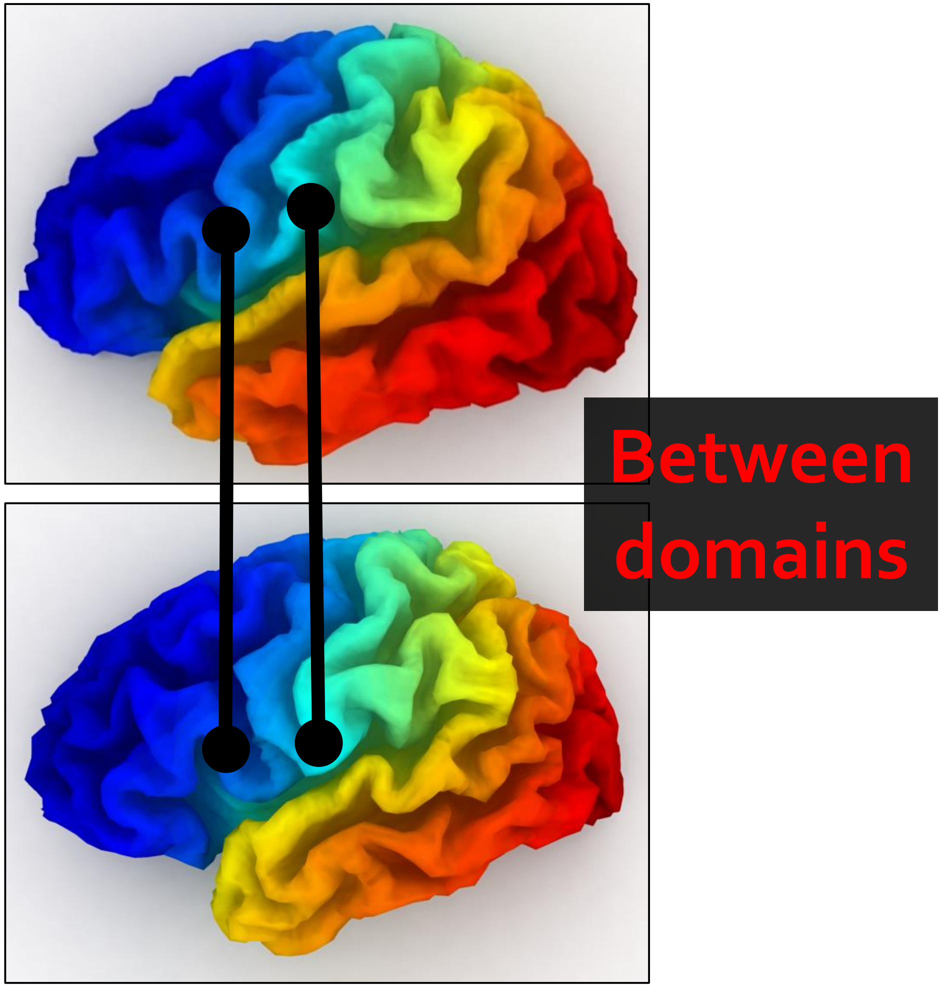
"No matched point should travel too far."

Model Problem: Quadratic Matching

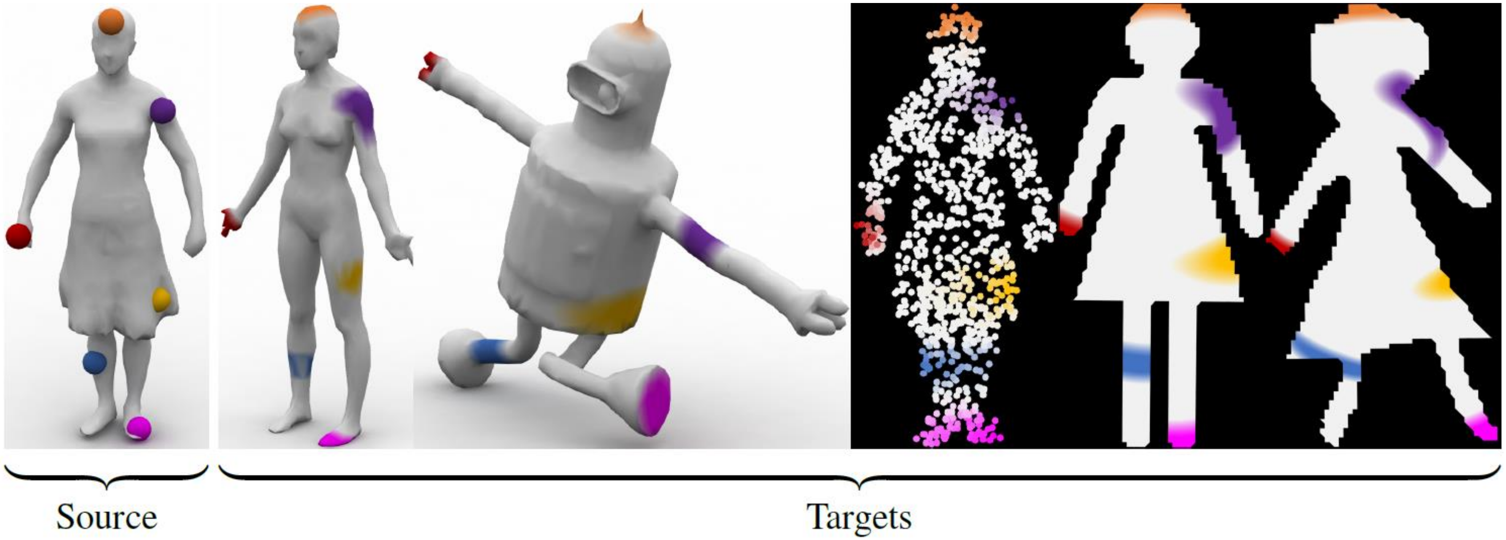
$$\begin{aligned} \min_T \quad & \langle M_0 T, T M_1 \rangle \\ \text{s.t.} \quad & T \geq 0 \\ & T\mathbf{1} = \mathbf{1} \\ & T^\top \mathbf{1} = \mathbf{1} \end{aligned}$$

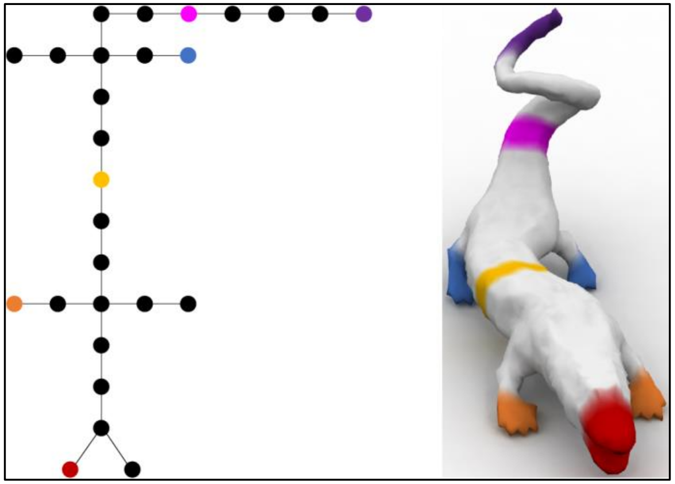
Nonconvex quadratic program!

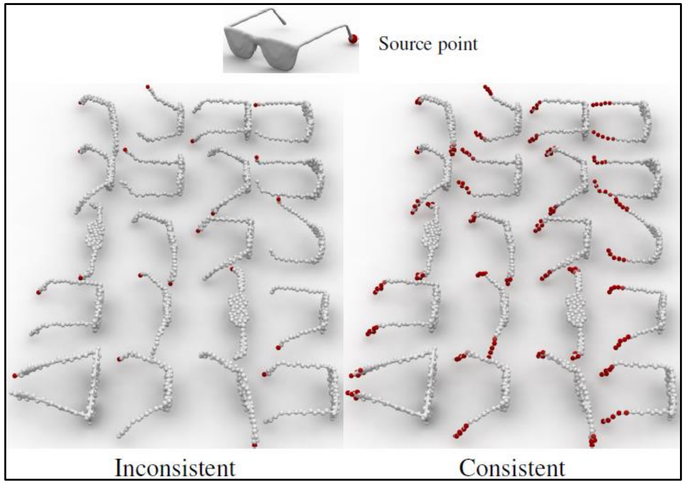
"Nearby points stay nearby."



Variety of Correspondence Tasks

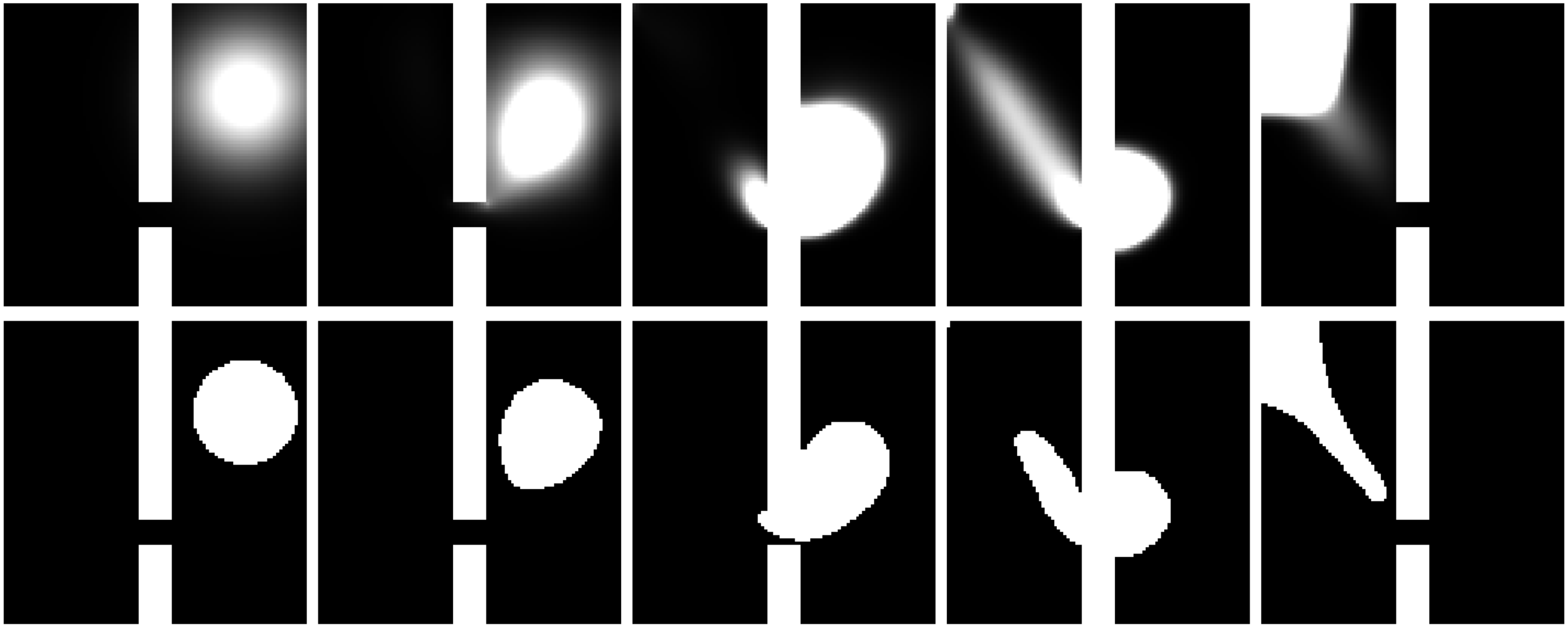






[Solomon et al. 2016]

Application 3: Gradient Flows



$t = 0$ $t = 10$ $t = 20$ $t = 30$ $t = 40$

"Entropic Wasserstein Gradient Flows" [Peyré 2015]

Interesting possibility for preserved structure!

Frontier:
Matrix Fields and Vector Measures

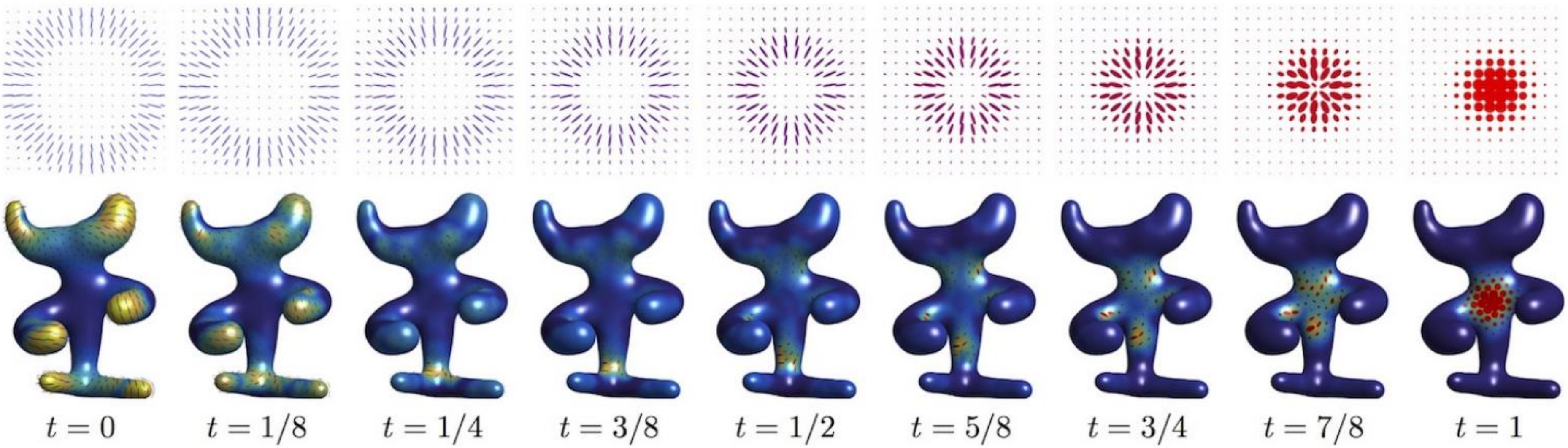


Image from
"Quantum Optimal Transport for Tensor Field Processing"
[Peyré et al. 2017]

Open problem: Dynamical version? Curved surfaces?

Frontier:

Sampling Problems

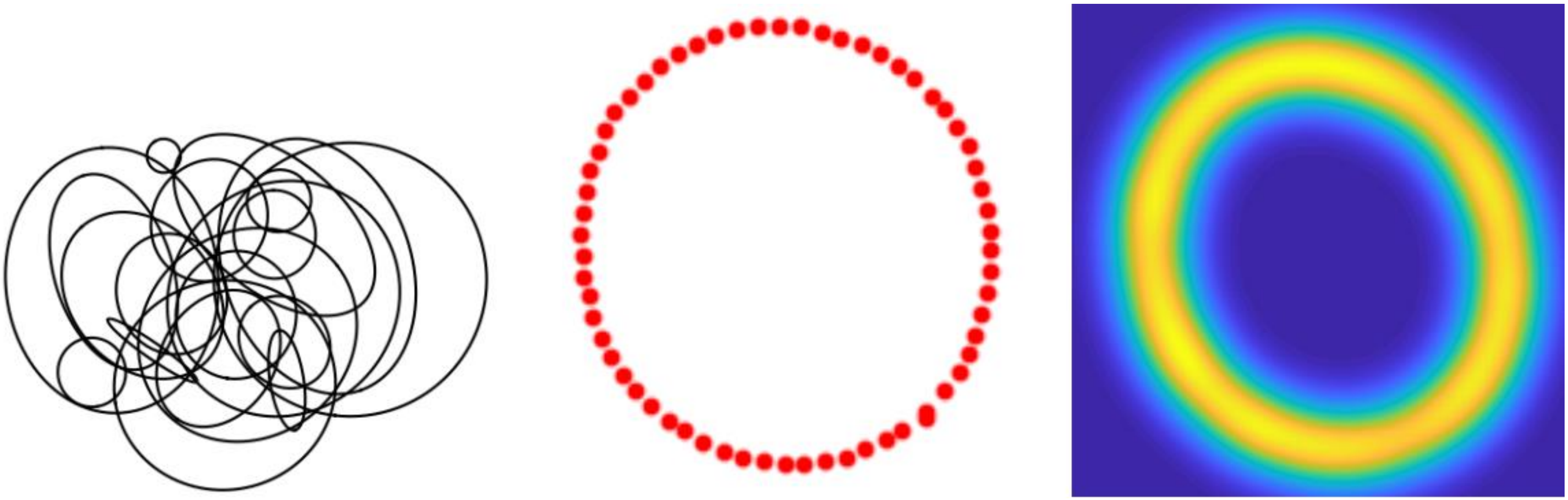


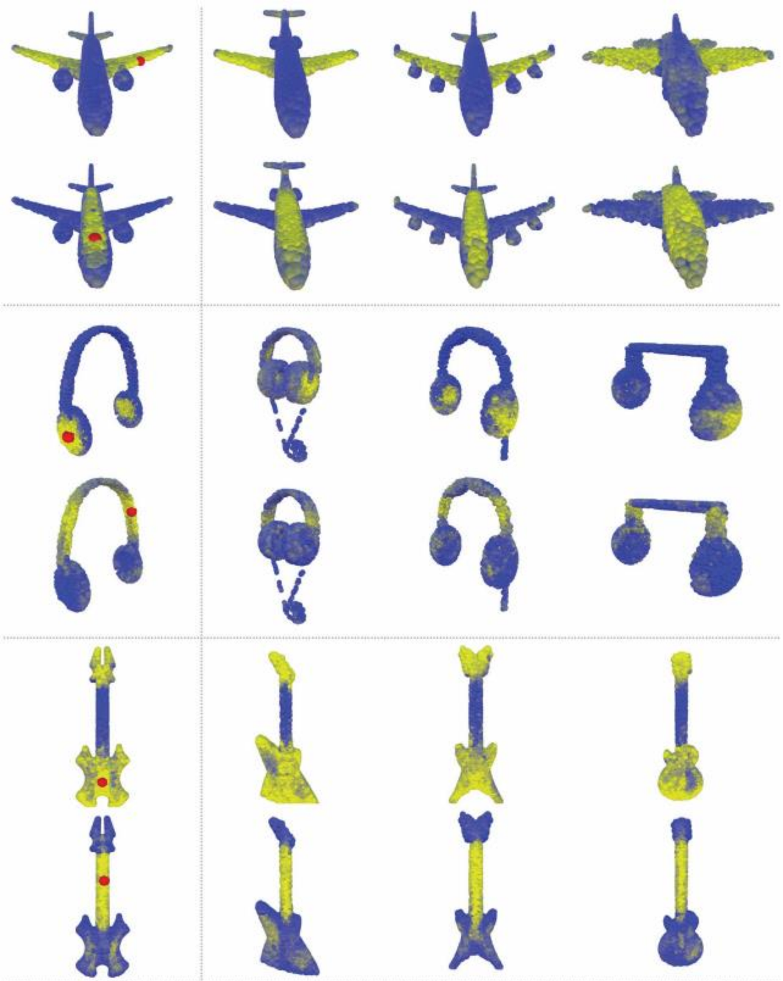
Image from

"Stochastic Wasserstein Barycenters"

[Claici et al. 2018]

Open problem: Sample from barycenter?

Frontier: Point Cloud Learning



Source points   Other point clouds from the same category

Image from
"Dynamic Graph CNN for Learning on Point Clouds"
[Wang et al. 2018]

Open problem: Many repeated instances of transport?

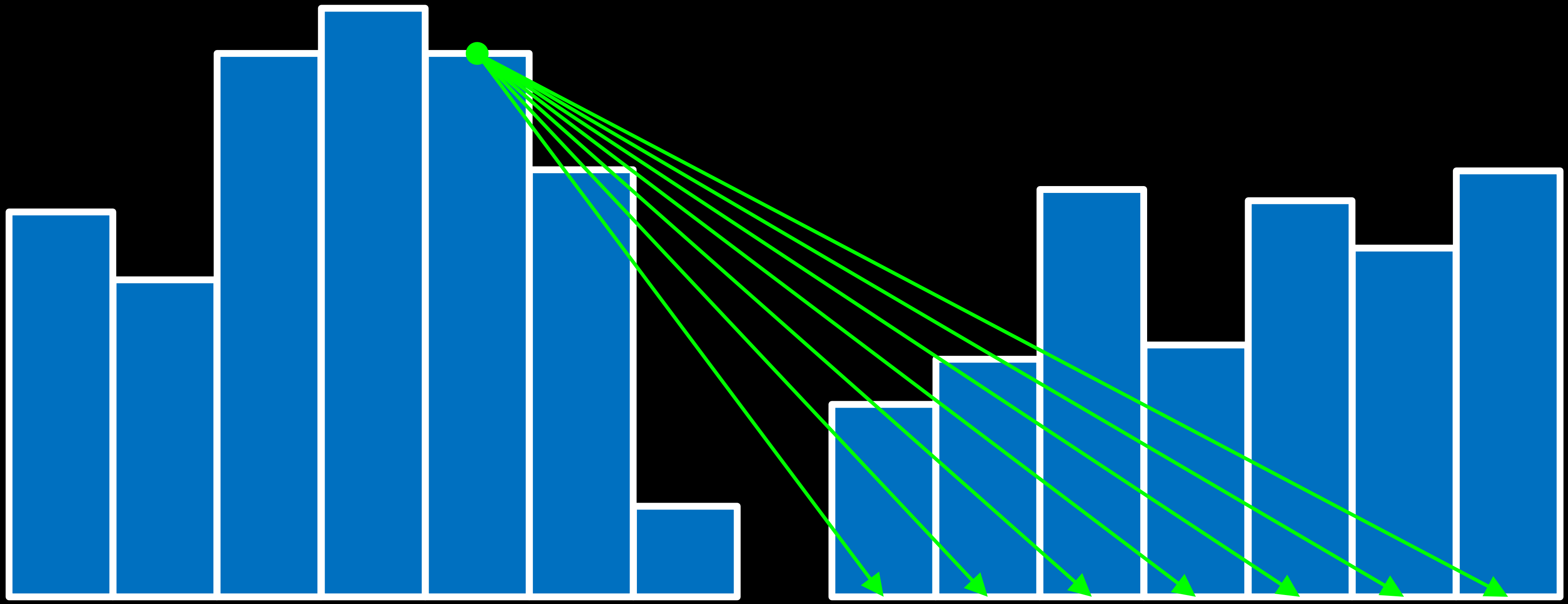
Plan For Today

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

2. Applications

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

4. Extensions & frontiers



Tutorial on

Optimal Transport

Questions?