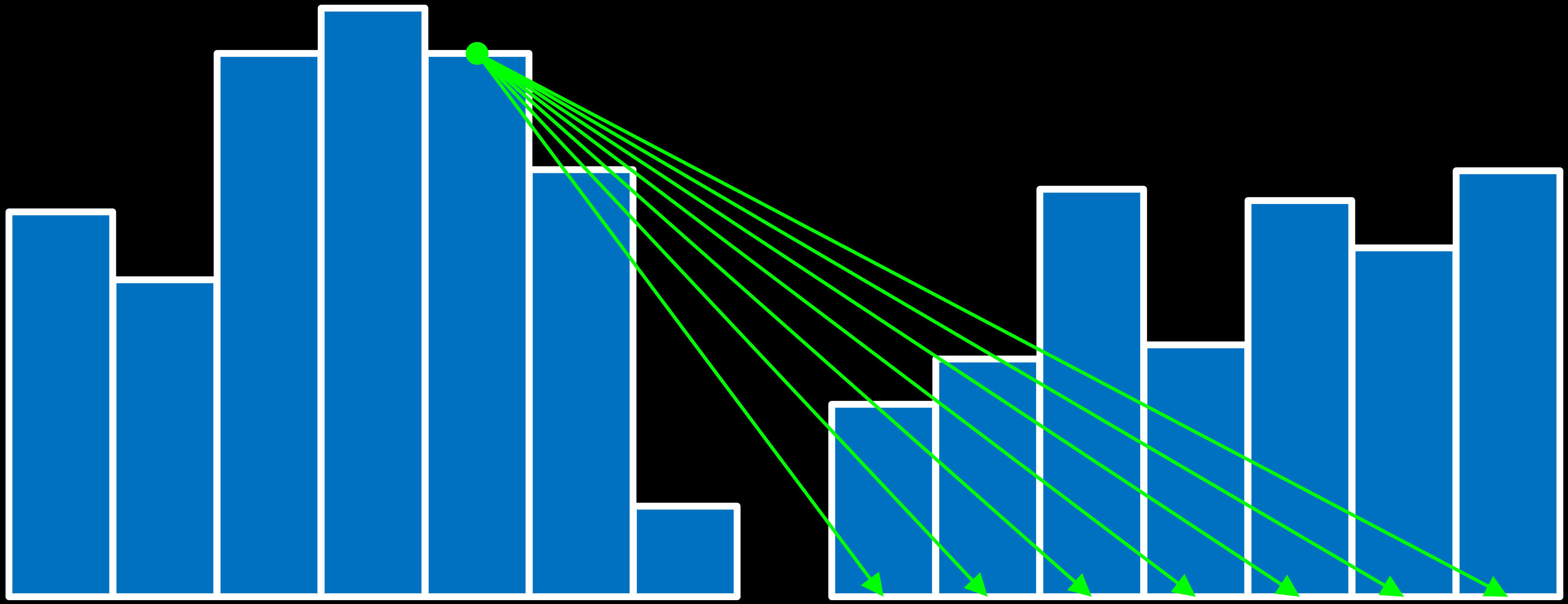


Transportation Distances

An Informal Tutorial

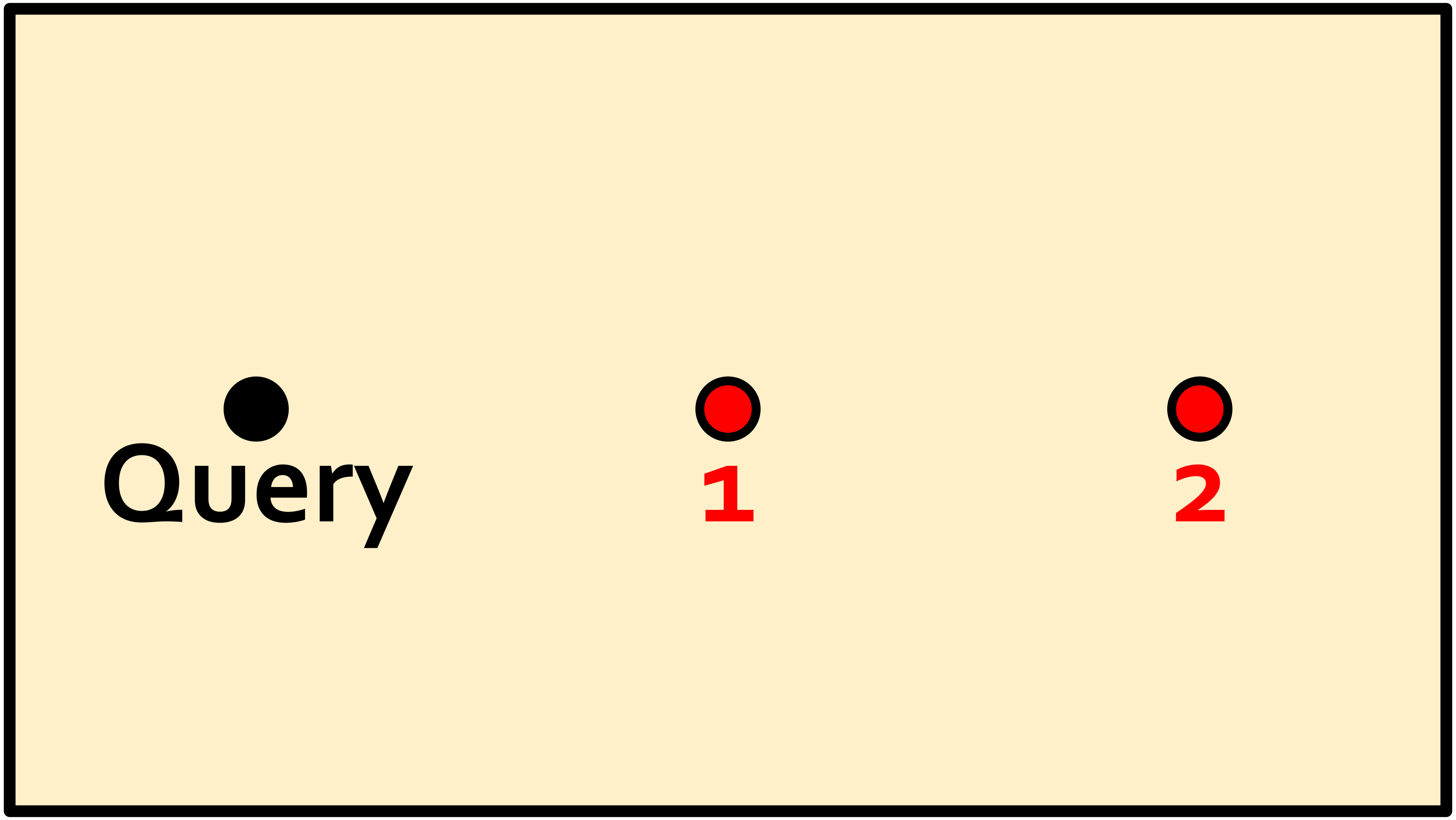
Justin Solomon

Stanford University


WARNING
NOT
COMPREHENSIVE

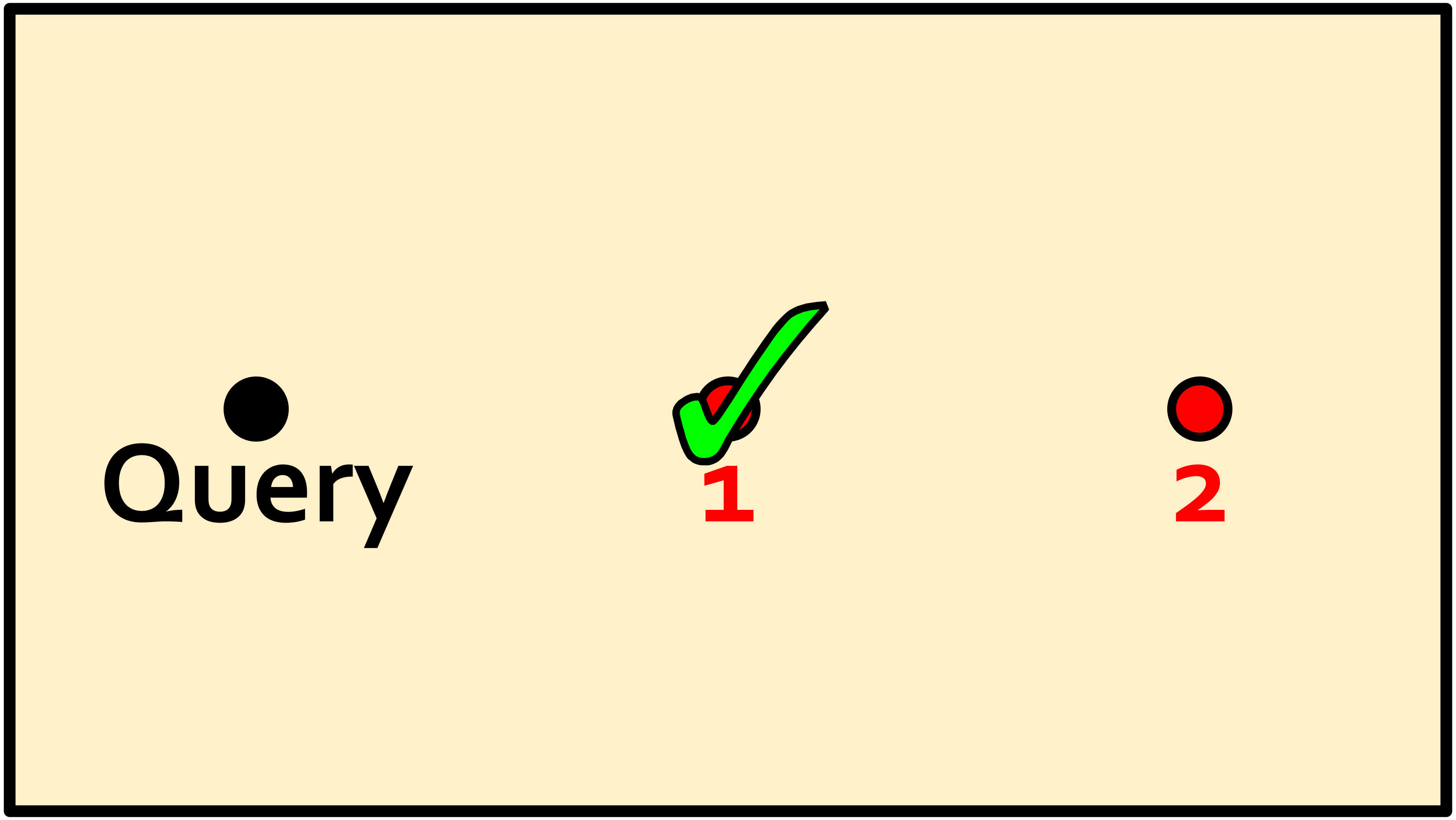
Biased toward computational applications
(and things I know about)

Motivating Question



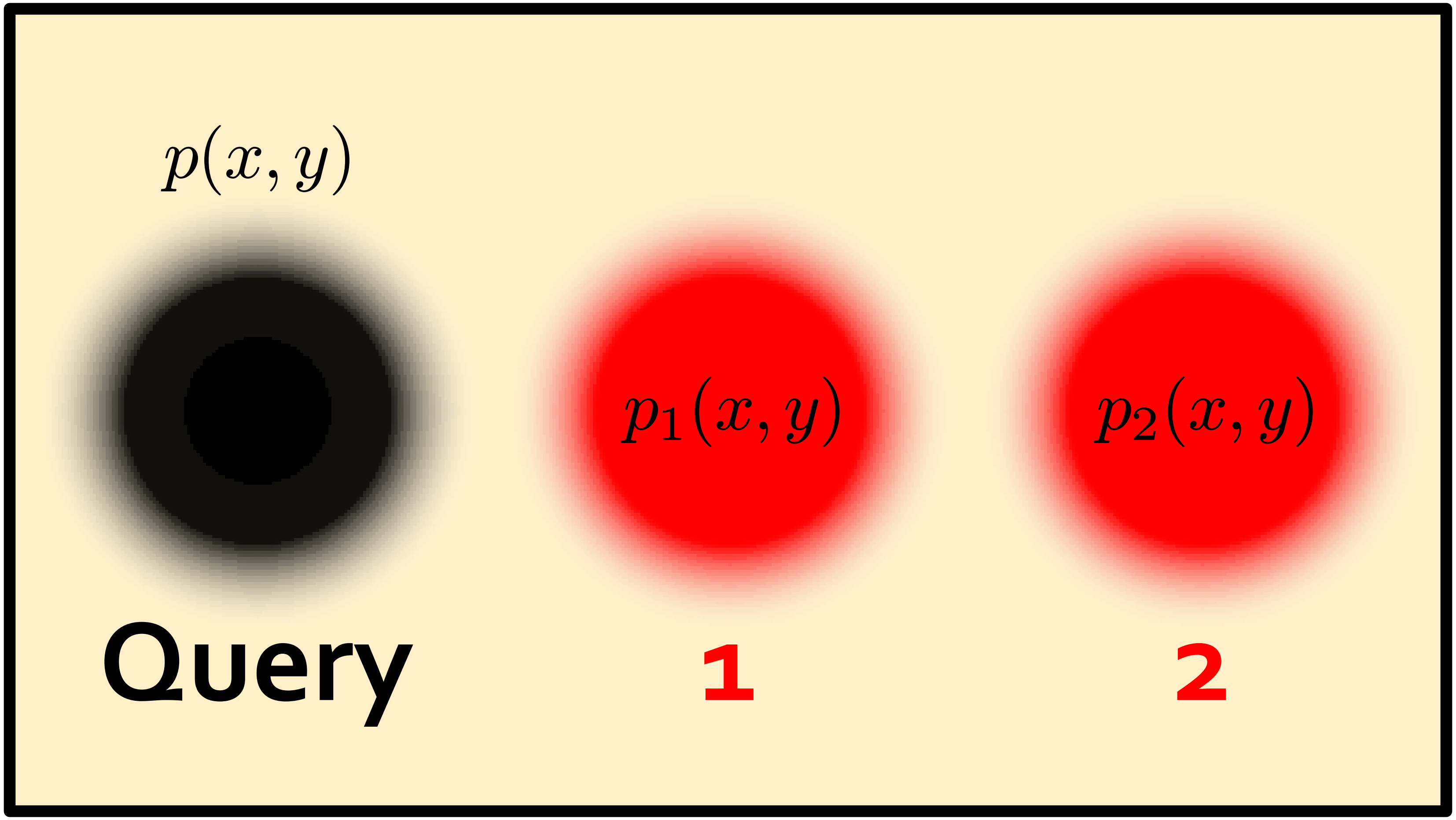
Which is closer, 1 or 2?

Motivating Question



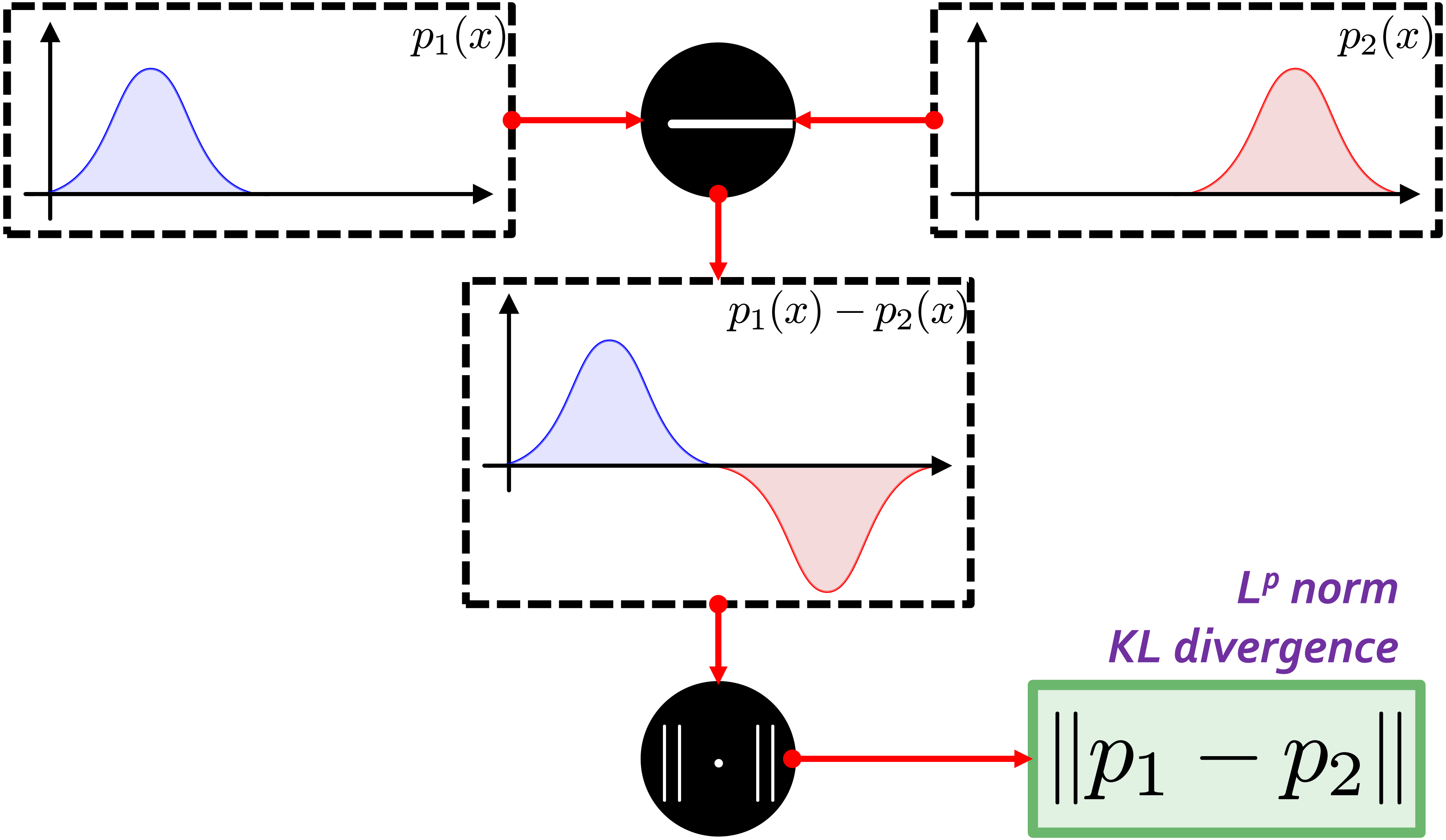
Which is closer, 1 or 2?

Fuzzy Version

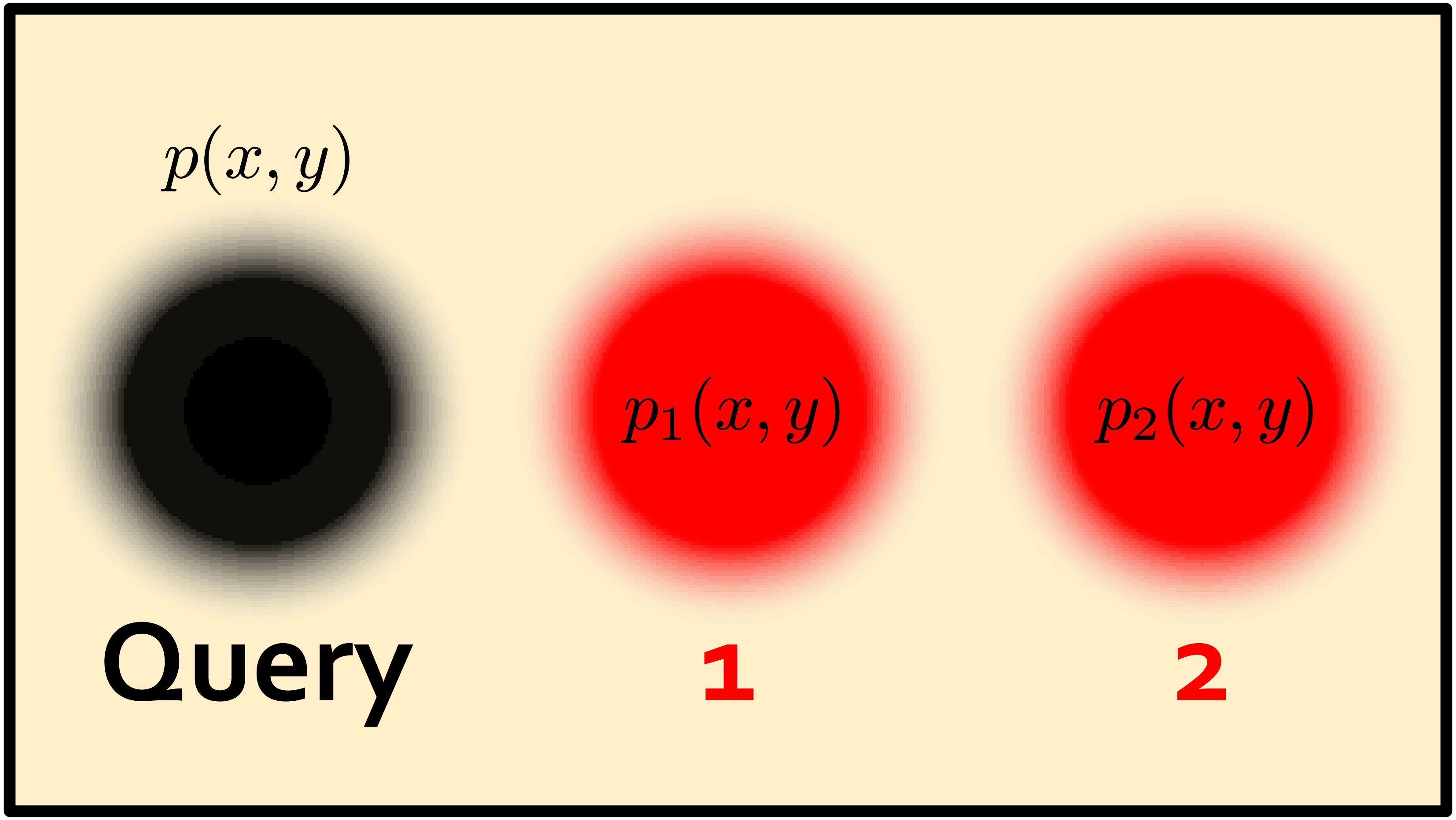


Which is closer, 1 or 2?

Typical Measurement

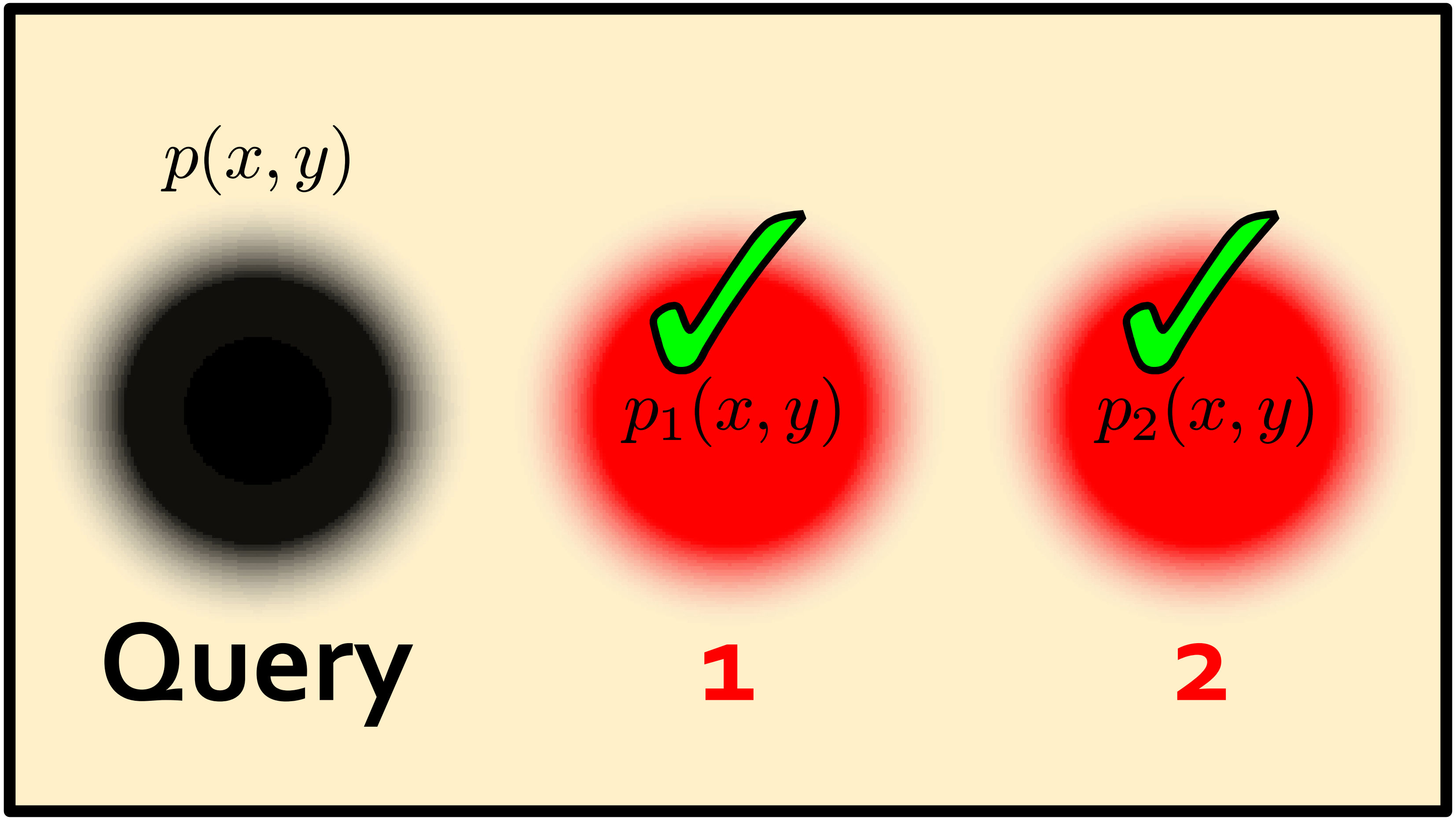


Returning to the Question



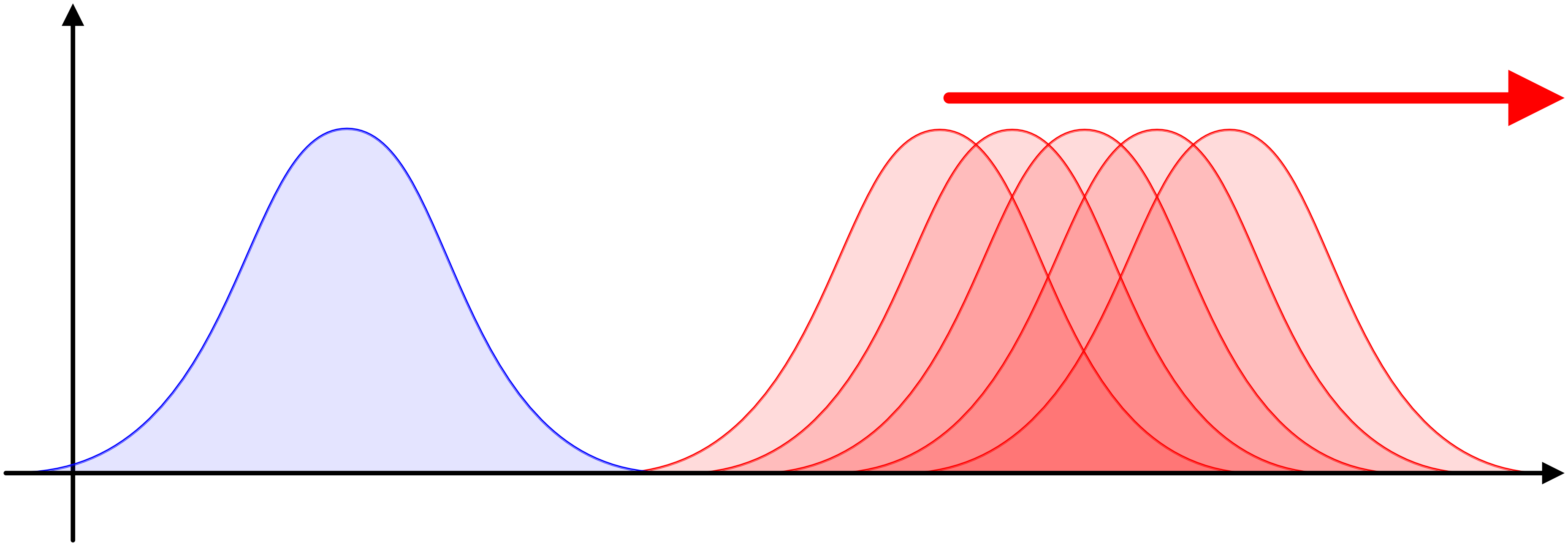
Which is closer, 1 or 2?

Returning to the Question



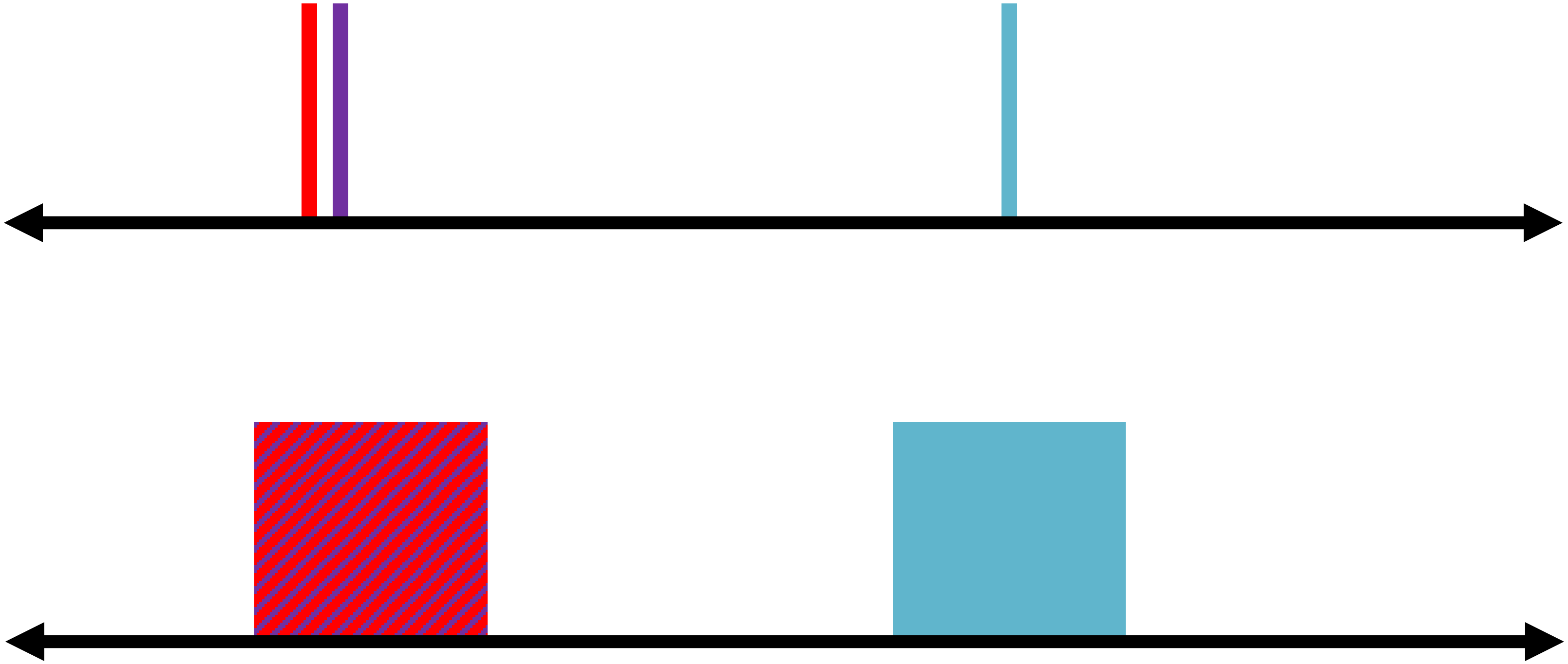
Neither!

What's Wrong?



Measured overlap, not displacement.

Related Issue



Smaller bins worsen histogram distances

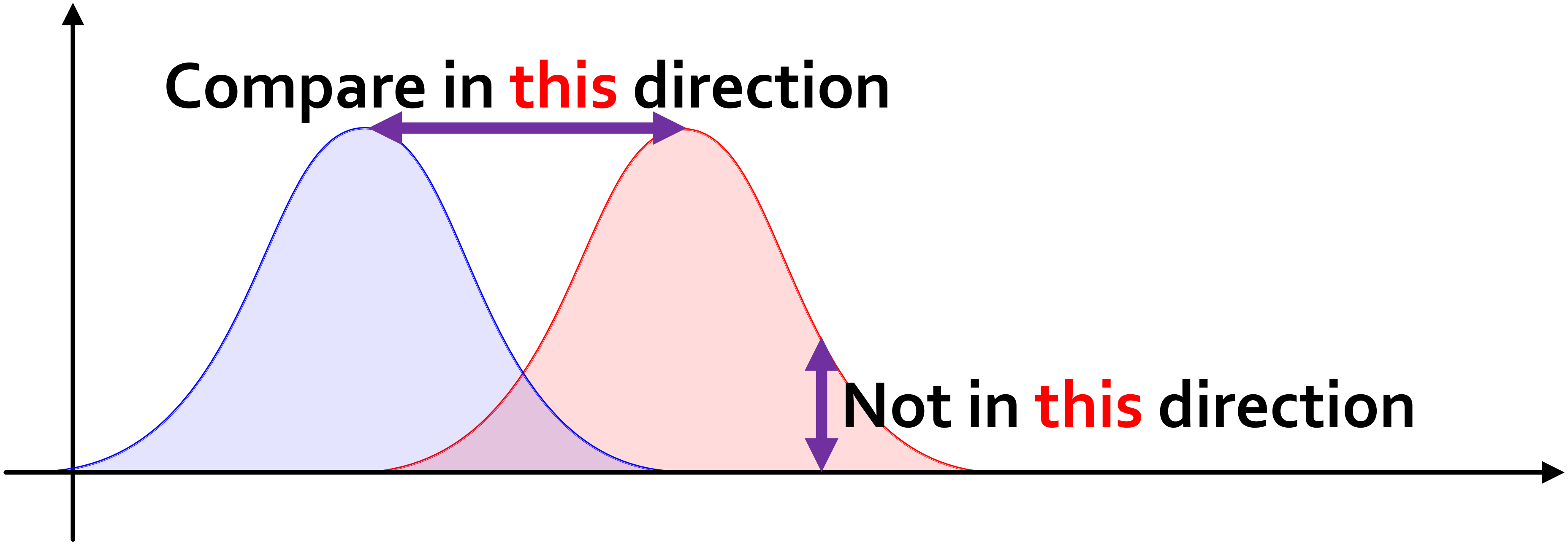
The Root Cause

Permuting histogram bins has

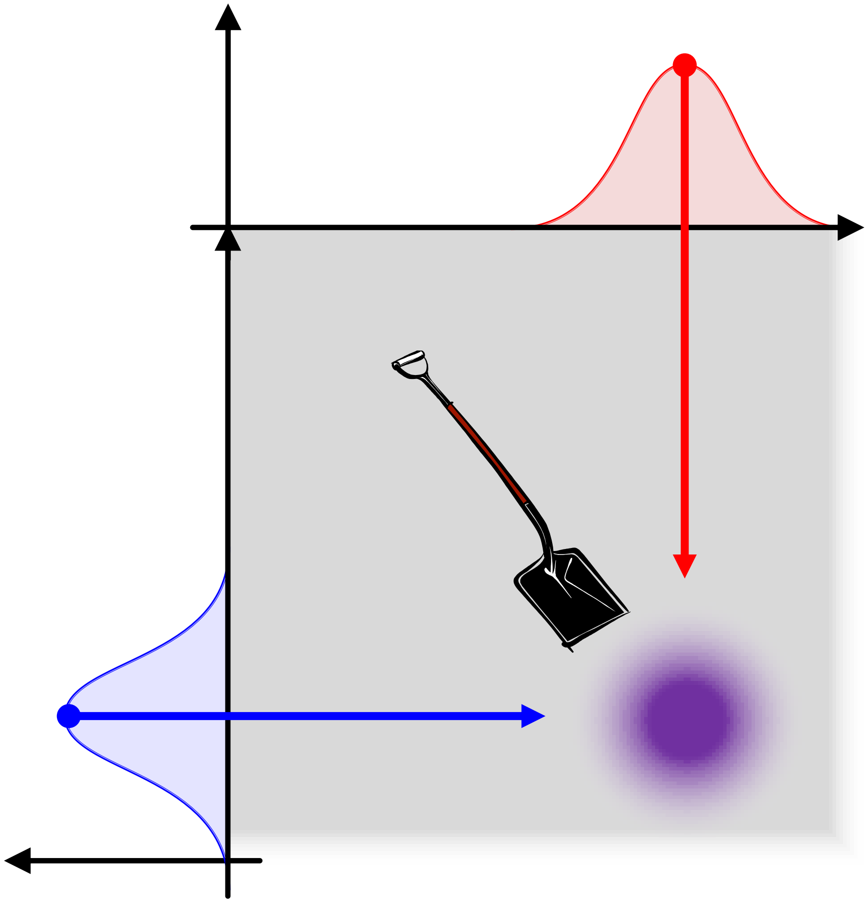
no effect

on these distances.

Alternative Idea

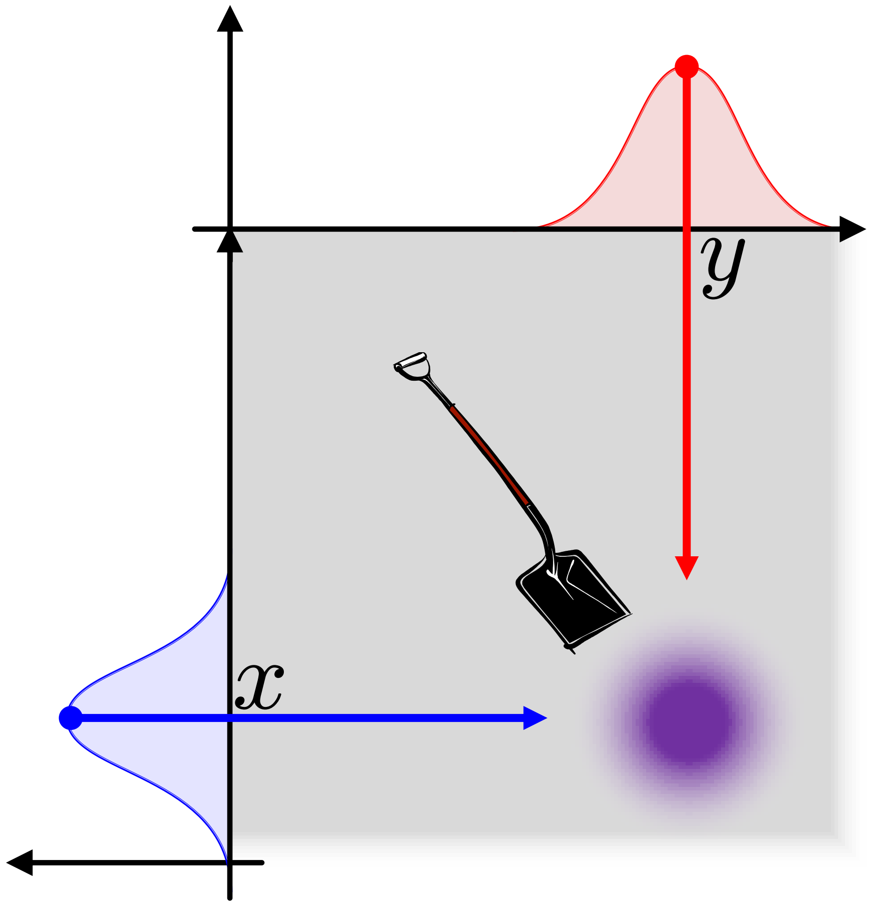


Alternative Idea



Match mass from the distributions

Earth Mover's Distance

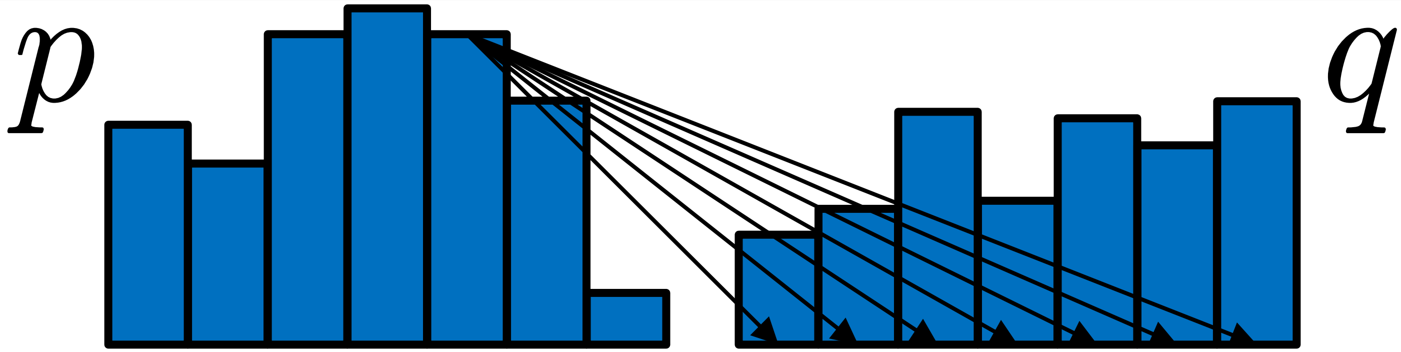


Cost to move mass m from x to y:

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

Match mass from the distributions

Earth Mover's Distance



$$\min_T \sum_{ij} T_{ij} d(x_i, x_j)$$

$m \cdot d(x, y)$

$$\text{s.t.} \sum_j T_{ij} = p_i$$

Starts at p

$$\sum_i T_{ij} = q_j$$

Ends at q

$$T \geq 0$$

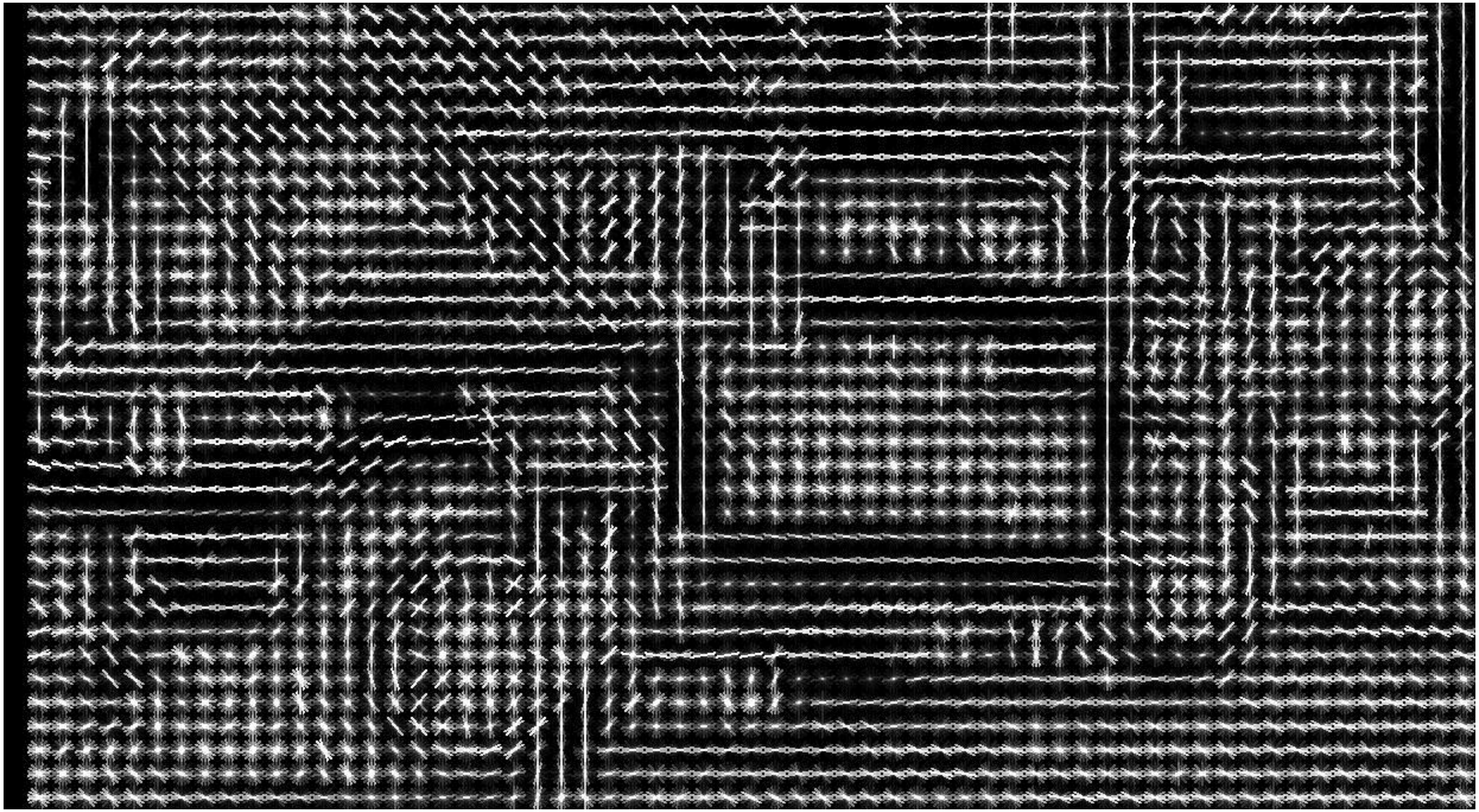
Positive mass

Important Theorem

EMD is a metric when d(x,y) satisfies the triangle inequality.

"The Earth Mover's Distance as a Metric for Image Retrieval"
Rubner, Tomasi, and Guibas
International Journal of Computer Vision 40.2 (2000): 99—121.

Basic Application



http://web.mit.edu/vondrick/ihog/

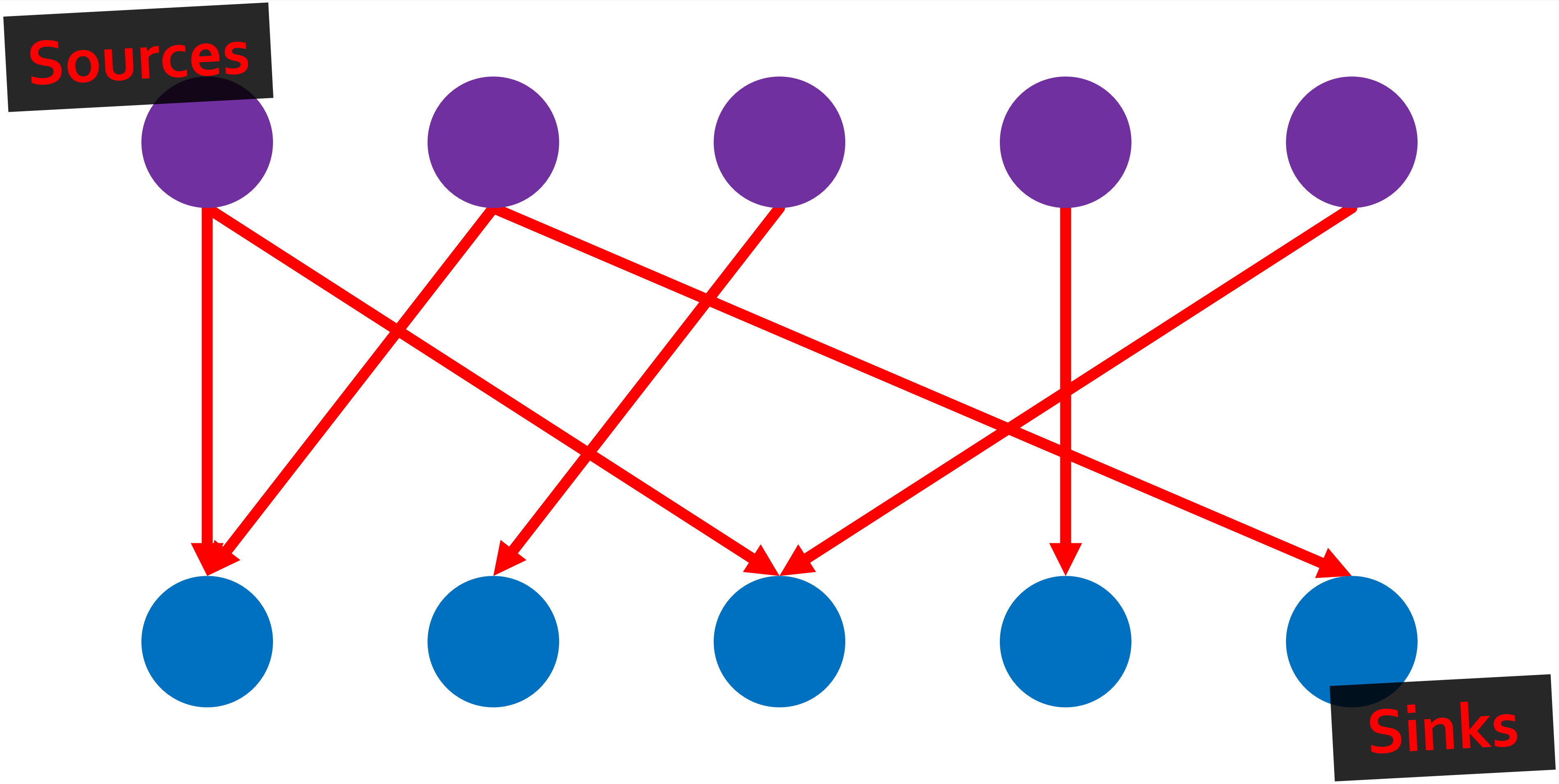
Comparing histogram descriptors

Computation of EMD

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

Quadratically-scaling LP

Discrete Perspective



Multi-Commodity Flow

Discrete Perspective

Useful conclusions:

1. Practical

Can do better than generic solvers.

Discrete Perspective

Useful conclusions:

1. Practical

Can do better than generic solvers.

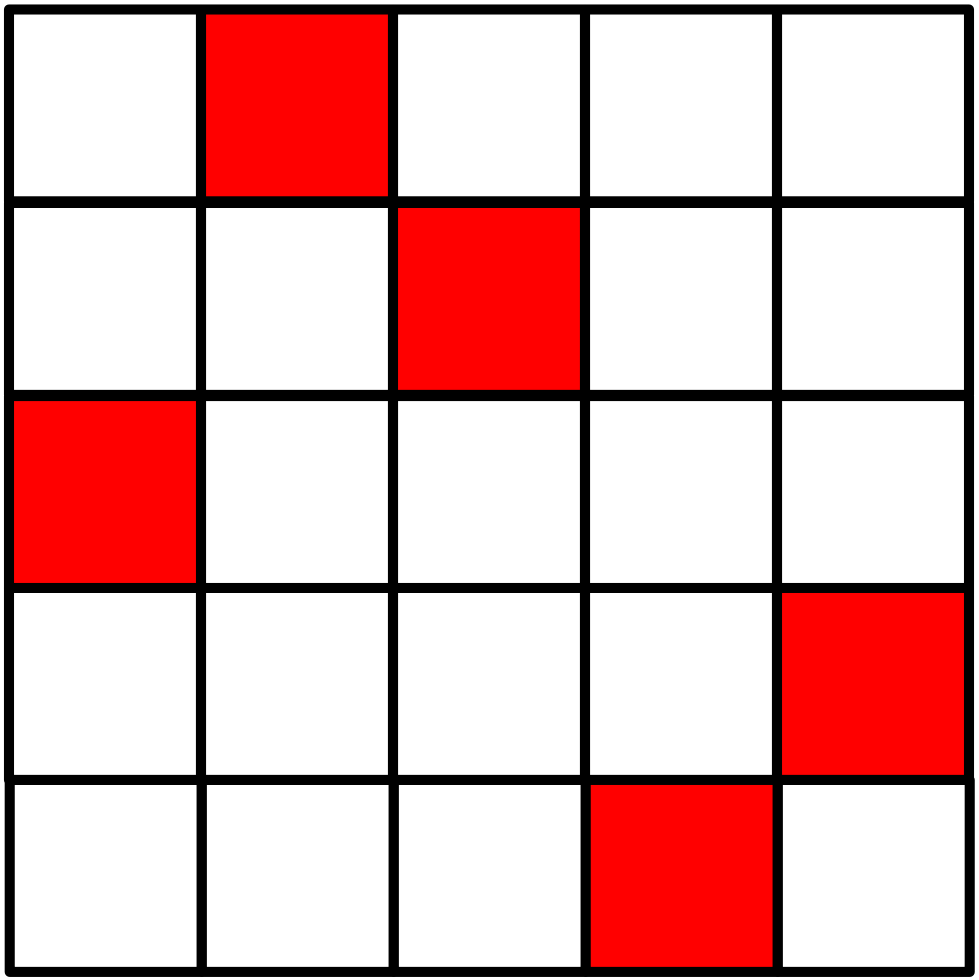
2. Theoretical

"Complementary slackness"

$T \in [0,1]^{n\times n}$ **usually contains** $O(n)$ **nonzeros.**

Multi-Commodity Flow

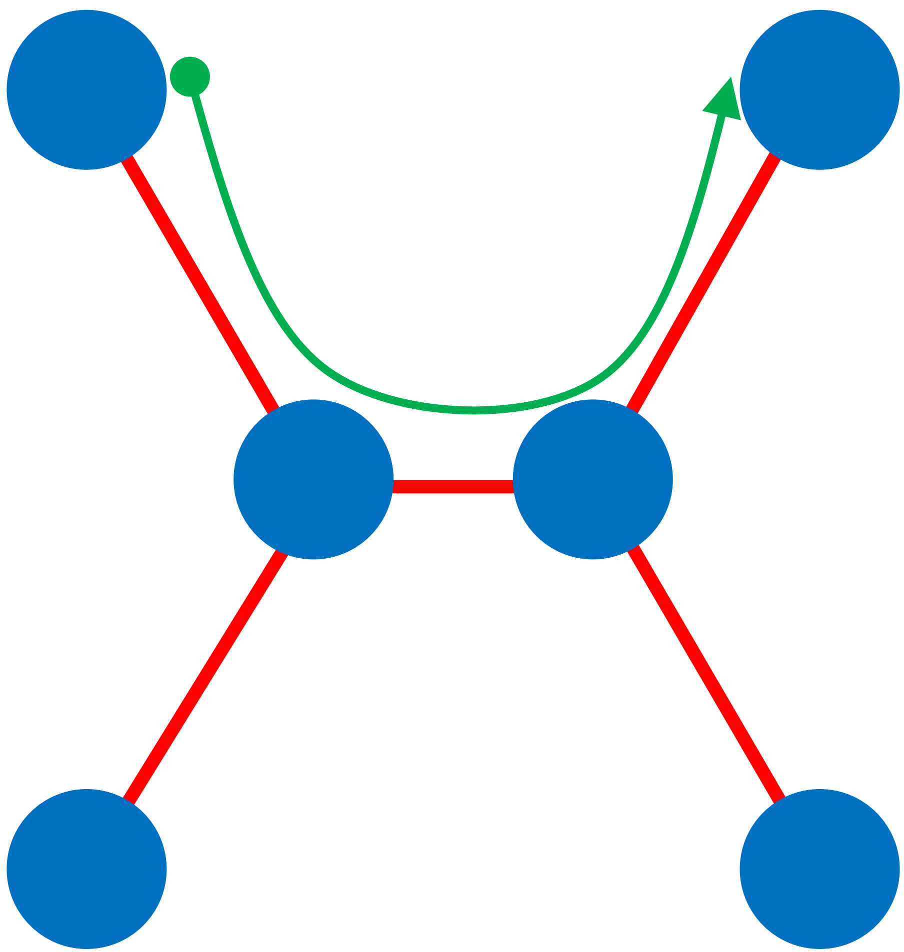
Transportation Matrix Structure



Matches bins

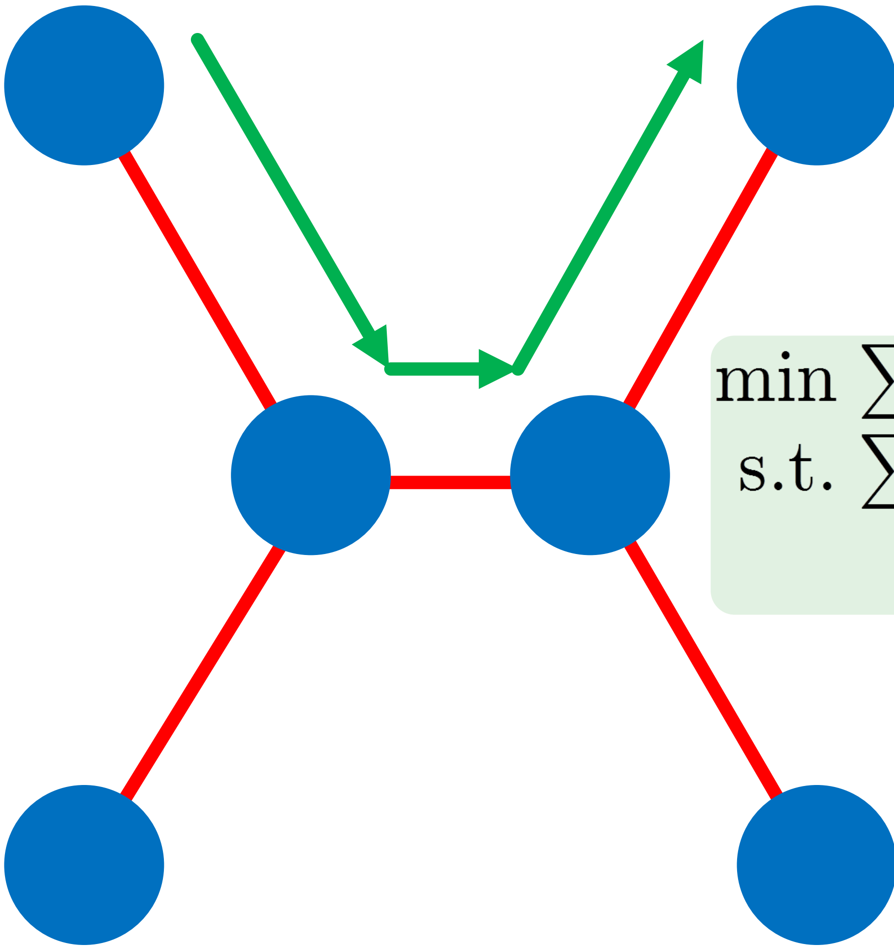
Underlying map!

Structured d(x,y)



$$
\begin{aligned}
\min_T \; & \sum_{ij} T_{ij} \\
\text{s.t.} \; & \sum_j T_{ij}
\end{aligned}
$$

Structured d(x,y)



$$\min \sum_{e \in E} |T(e)|$$
$$\text{s.t.} \sum_{e=(v,w)} T(e) - \sum_{e=(w,v)} T(e) = q_v - p_v \;\; \forall v \in V$$

Other Possibilities for Structure

- **Thresholded ground distance**
  Pele and Werman 2009
- **Linear/cyclic/grid domains**
  Assorted theory papers

Continuous Notation

<pause>

Beware:

Confusing notation!

Continuous Notation

$$\min_{\pi \in \Pi(\mu,\nu)} \iint_{X \times X} c(x,y)\, d\pi(x,y)$$

Monge-Kantorovich Problem

Continuous Notation

$$\min_{\pi \in \Pi(\mu,\nu)} \iint_{X \times X} c(x,y)\, d\pi(x,y)$$

Measure coupling

$$\forall U, V \subseteq X$$

$$\mu(U) = \pi(U \times X)$$

$$\nu(V) = \pi(X \times V)$$

Monge-Kantorovich Problem

p-Wasserstein Distance

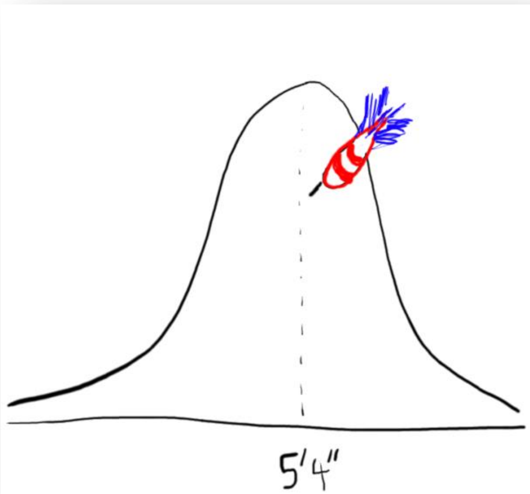
$$\mathcal{W}_p(\mu,\nu) \equiv \min_{\pi \in \Pi(\mu,\nu)} \left(\iint_{X \times X} d(x,y)^p \, d\pi(x,y) \right)^{1/p}$$

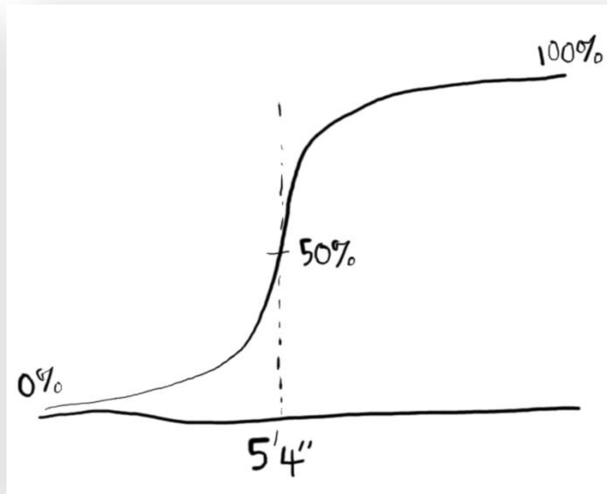
Shortest path distance

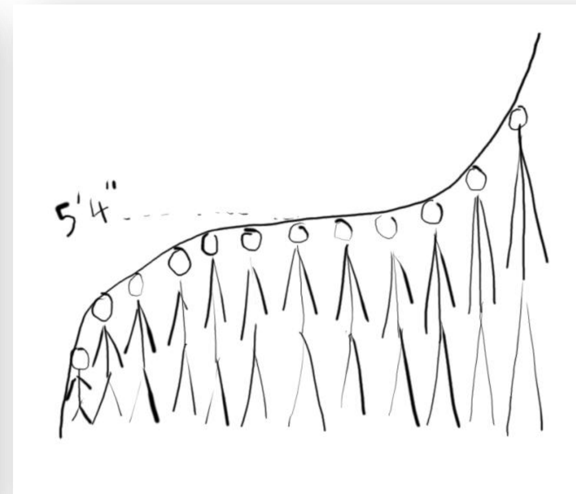
Expectation

Ground distance from shortest path

In One Dimension







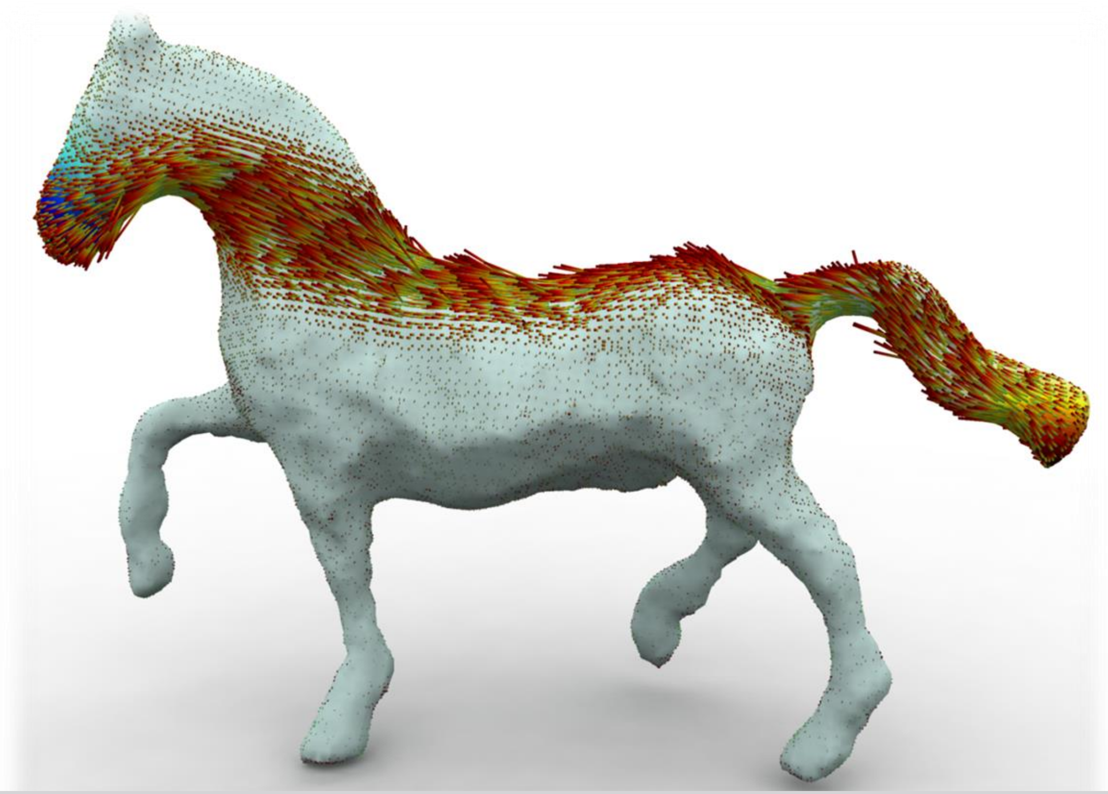
PDF ⋯⋯▶ **[CDF]** ⋯⋯▶ **CDF^{-1}**

$$\mathcal{W}_1(\mu,\nu) = \|\text{CDF}(\mu) - \text{CDF}(\nu)\|_1$$

$$\mathcal{W}_2(\mu,\nu) = \|\text{CDF}^{-1}(\mu) - \text{CDF}^{-1}(\nu)\|_2$$

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

Connection to Fluid Dynamics



\omit{equations}

Benamou & Brenier

Advect distributions using minimal work.

Similar Alternative for W1

"Beckmann problem"

$$\overbrace{\inf_J \int_M \|J(x)\|\,dx}^{\text{Total work}}$$

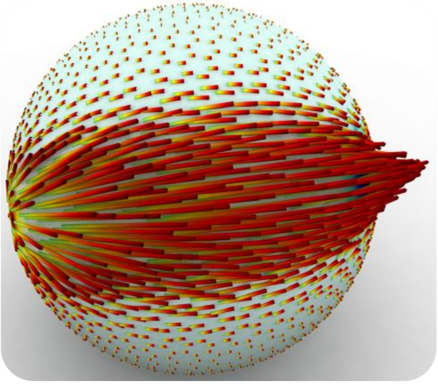
$$\text{s.t. } \nabla \cdot J(x) = \rho_1(x) - \rho_0(x)$$

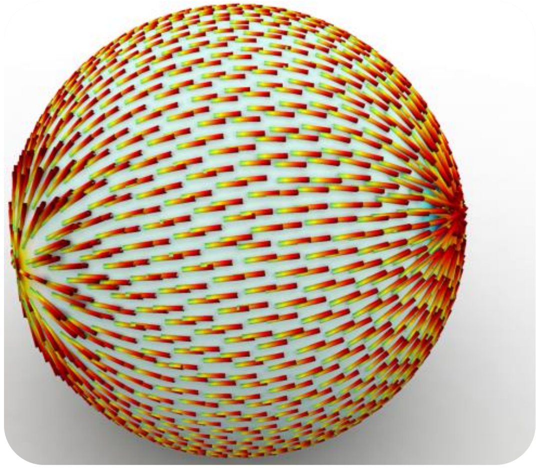
$$\underbrace{J(x) \cdot n(x) = 0 \;\forall x \in \partial M}_{\text{Advects from } \rho_0 \text{ to } \rho_1}$$

Similar to graph problem

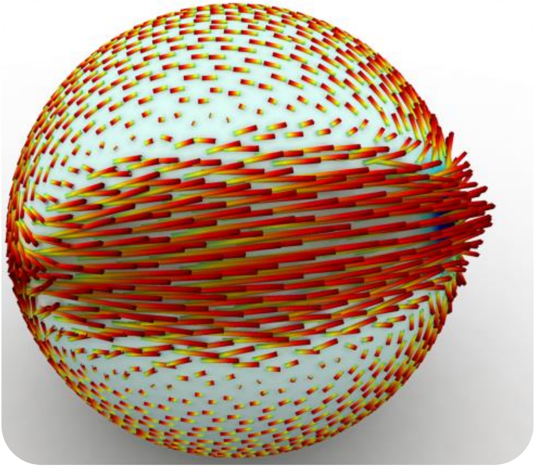
Hodge Decomposition of *J*

$$J(x) = \nabla f(x) + \mathcal{R} \cdot \nabla g(x)$$





Curl-free

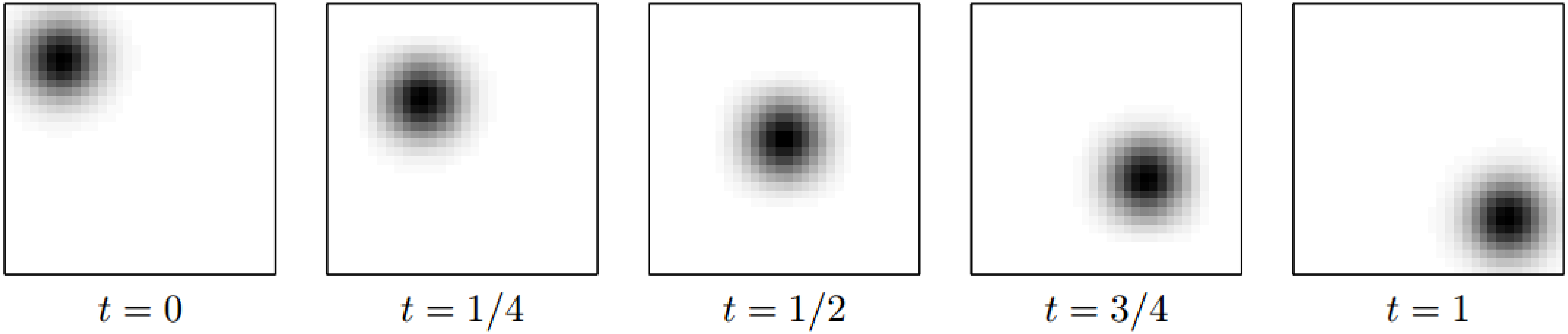


Div-free

$$\nabla \cdot J = \Delta f = \rho_1 - \rho_0$$

Displacement Interpolation

$$\mathcal{W}_2$$



"Explains" shortest path.

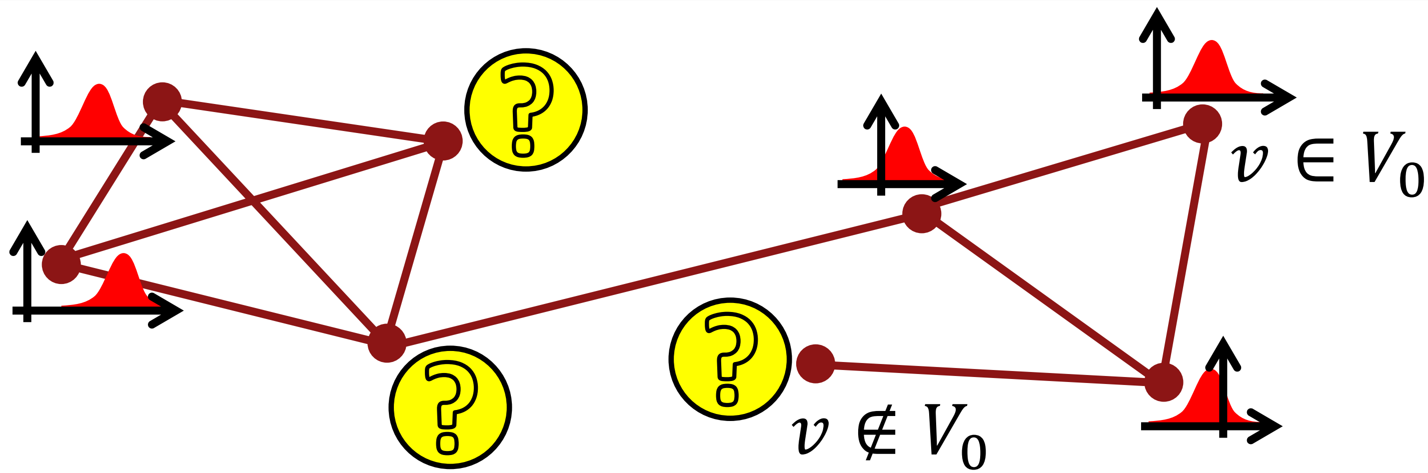
Image from "Optimal Transport with Proximal Splitting" (Papadakis, Peyré, and Oudet)

Mass moves along shortest paths

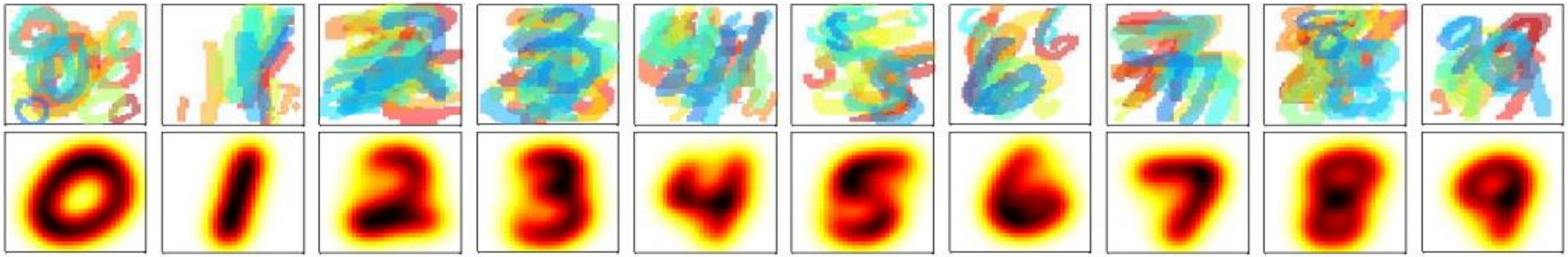
Parallel to Information Geometry

- Consider set of distributions as a **manifold**
- **Tangent spaces** from advection
- **Geodesics** from displacement interpolation

Computational Applications



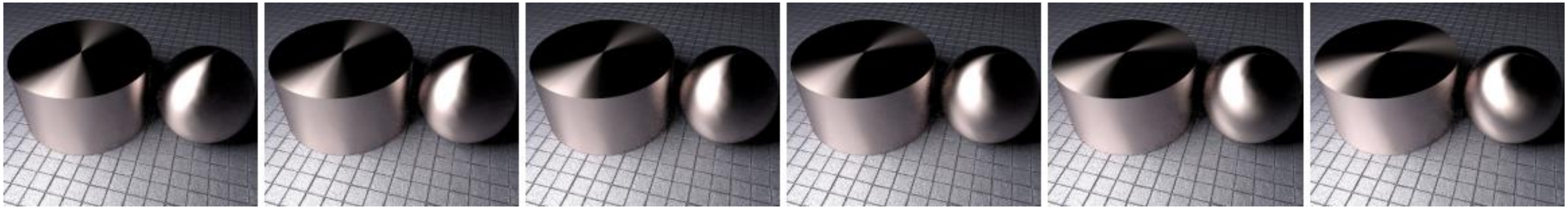
"Wasserstein Propagation for Semi-Supervised Learning" (Solomon et al.)



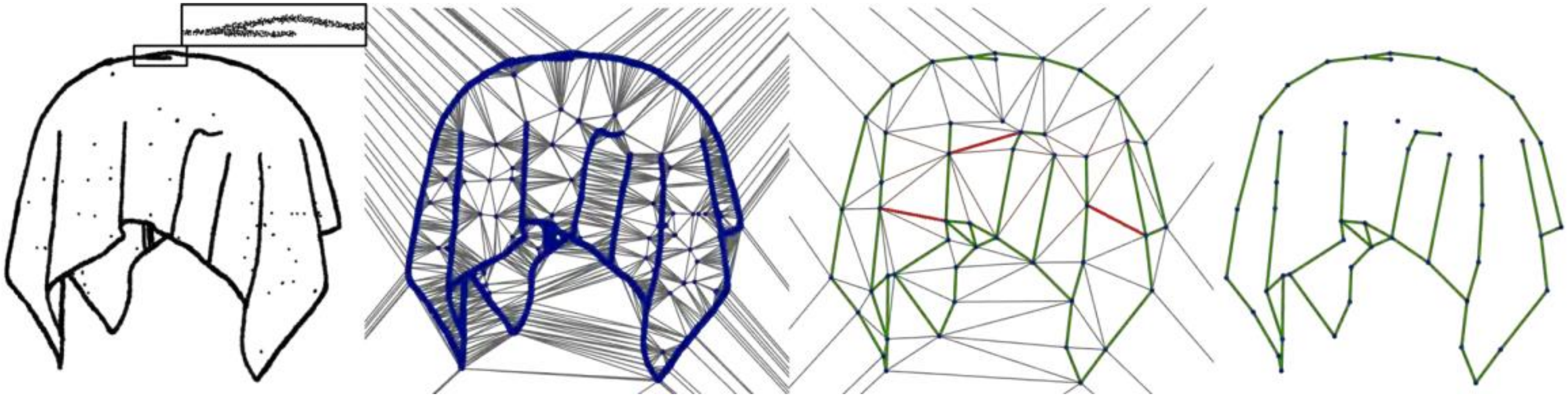
"Fast Computation of Wasserstein Barycenters" (Cuturi and Doucet)

Learning

Computational Applications



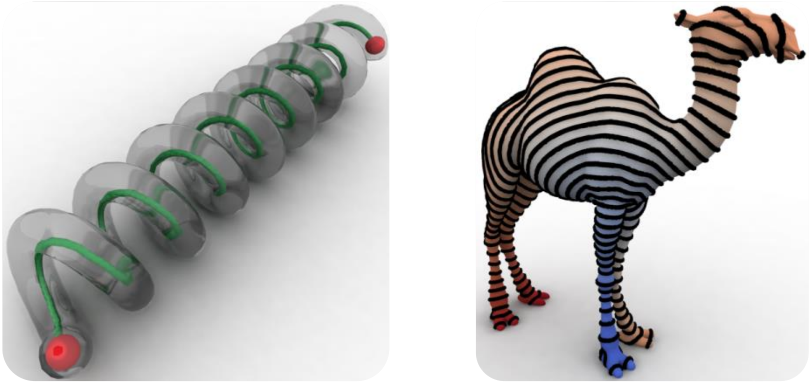
“Displacement Interpolation Using Lagrangian Mass Transport” (Bonneel et al.)



“An Optimal Transport Approach to Robust Reconstruction and Simplification of 2D Shapes” (de Goes et al.)

Morphing and registration

Computational Applications

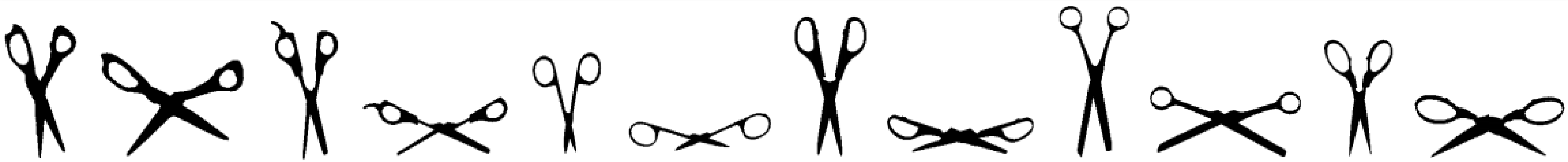


"Earth Mover's Distances on Discrete Surfaces" (Solomon et al.)

"Blue Noise Through Optimal Transport" (de Goes et al.)

Graphics

Computational Applications



"Geodesic Shape Retrieval via Optimal Mass Transport" (Rabin, Peyré, and Cohen)

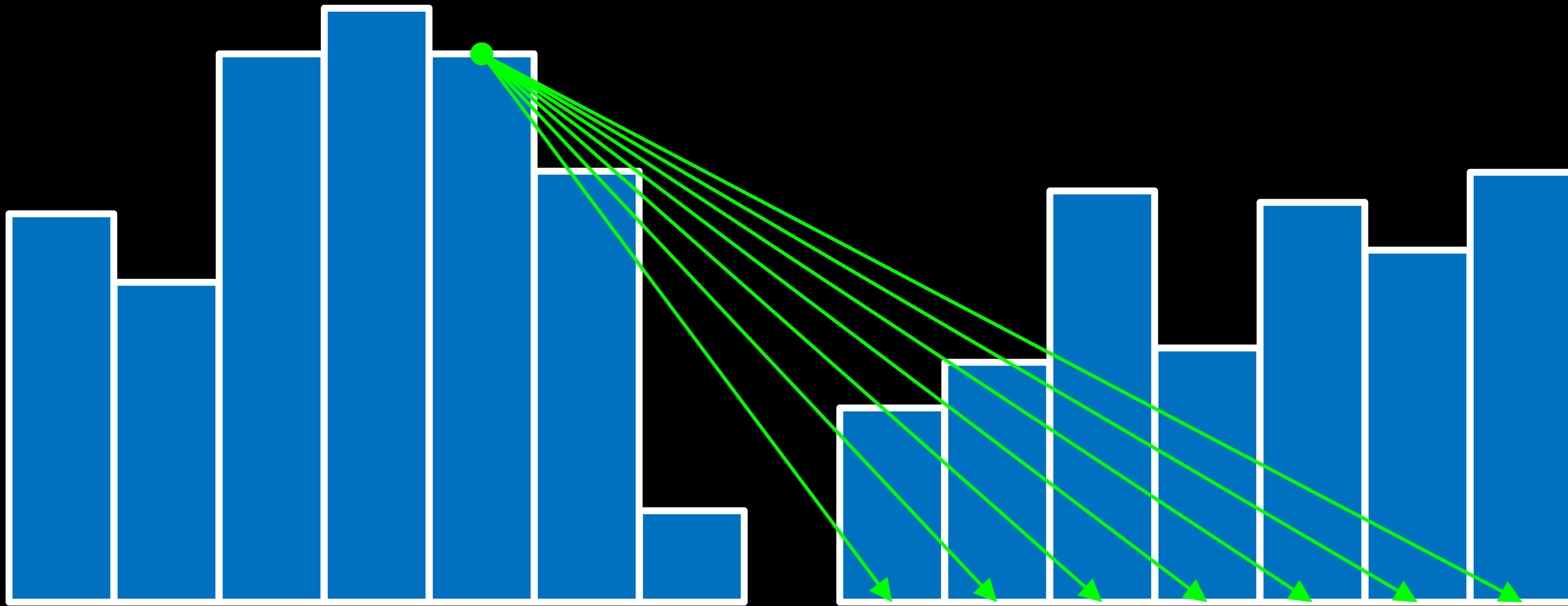
"Adaptive Color Transfer with Relaxed Optimal Transport" (Rabin, Ferradans, and Papadakis)

Vision and image processing

What's Left?

- **Learning applications**
  Variational methods, metric learning, …

- **Efficient computation in L_2 case**



Transportation Distances

An Informal Tutorial

Questions?