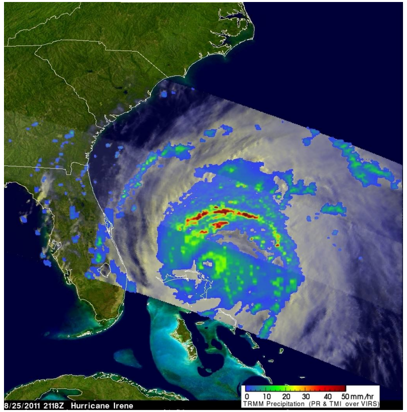
Stochastic Wasserstein Barycenters

Sebastian Claici, Edward Chien, Justin Solomon

Motivation

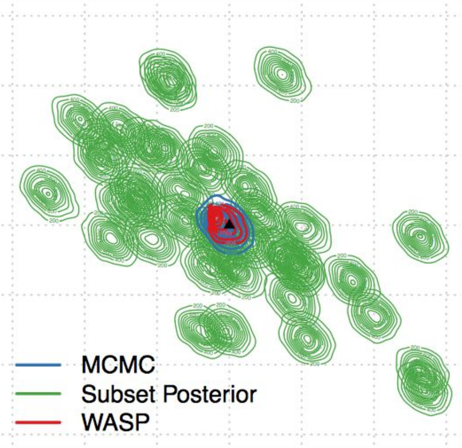
Summarizing data from multiple sources.

Weather Stations



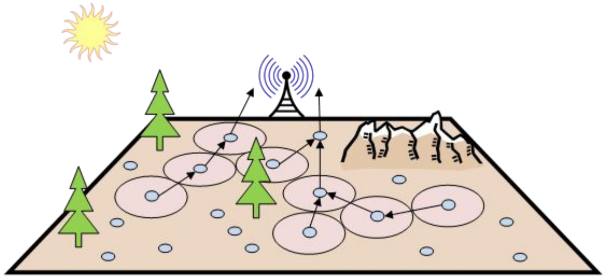
[NASA]

Subsets of a Dataset



[Srivastava et al., 2015]

Sensor Network

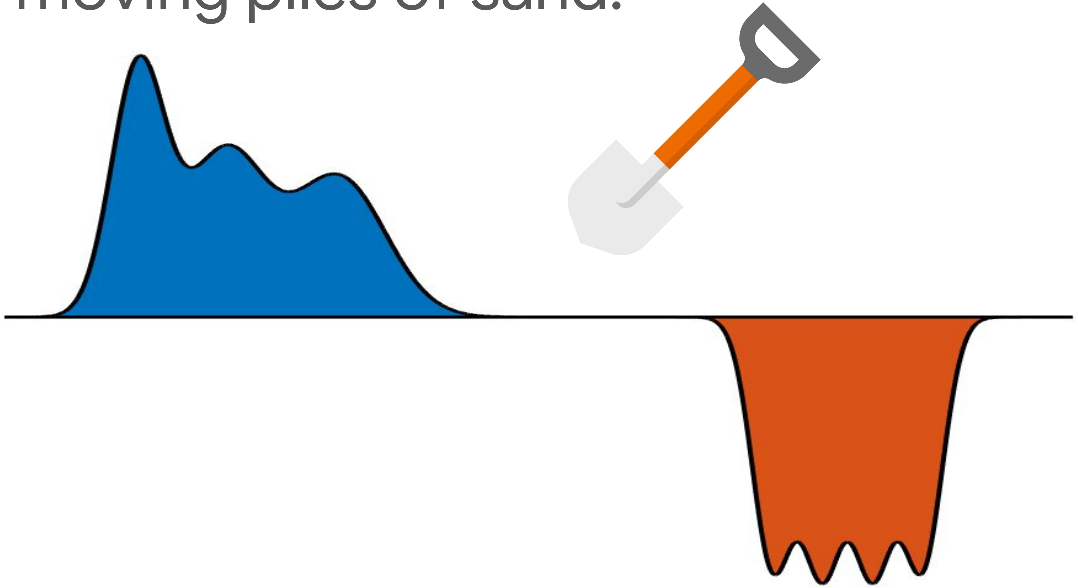


[Sarma Vrudhula]

Optimal Transport

A *geometry* for probability measures.

Intuition: moving piles of sand.



Wasserstein Distance

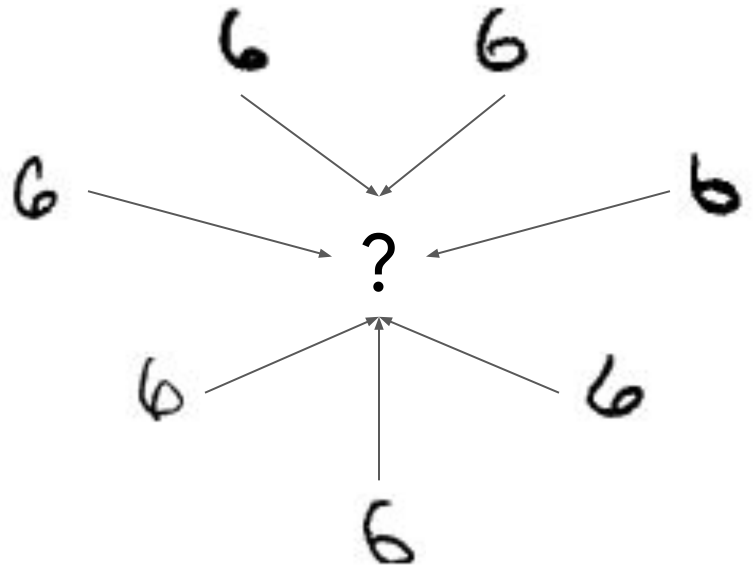
The 2-Wasserstein distance between μ_1 and μ_2 is

$$W_2^2(\mu_1, \mu_2) = \left(\underbrace{\inf_{\gamma \in \Gamma(\mu_1, \mu_2)}}_{\text{couplings}} \int_{X \times X} \underbrace{d(\mathbf{x}, \mathbf{y})^2}_{\text{metric}} \, \mathrm{d}\gamma(\mathbf{x}, \mathbf{y}) \right)$$

Intuition: coupling tells you where to shovel.

Barycenters in Probability Space

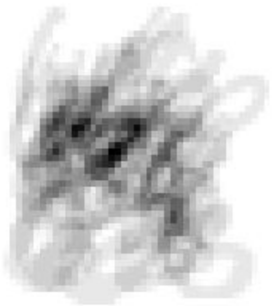
Often want to *average* distributions.



Barycenters in Probability Space

Euclidean averaging doesn’t contain geometric information.

Expectation



Euclidean Average

Wasserstein Barycenters

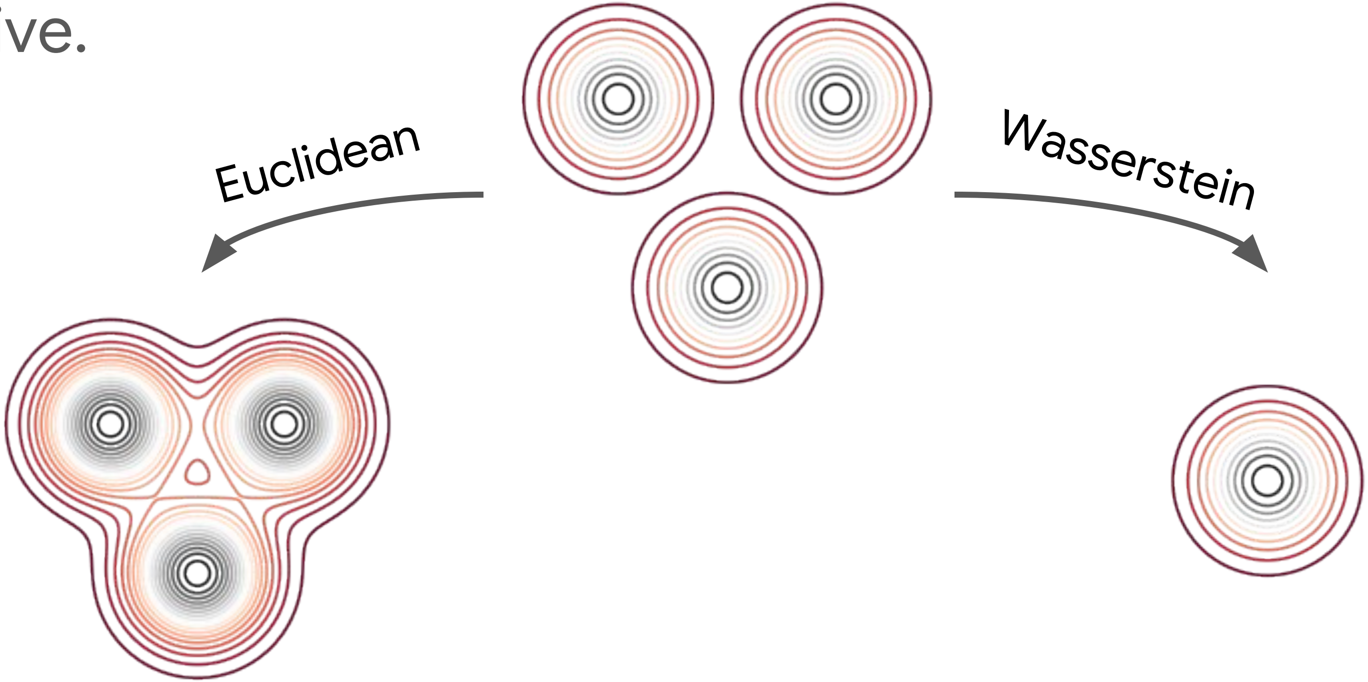
The Wasserstein barycenter of measures μ_j is the measure ν that minimizes

$$F[\nu] = \frac{1}{N} \sum_{j=1}^{N} W_2^2(\nu, \mu_j)$$

The standard Euclidean mean is the analog of this.

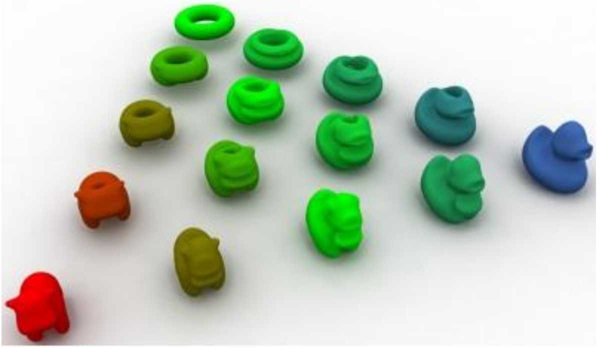
Why Wasserstein Barycenters?

Barycenters under the Wasserstein distance are more intuitive.

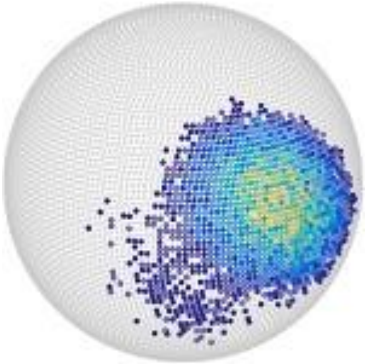


Related Work

[Cuturi & Doucet, 2014]



[Solomon et al., 2015]



[Staib et al., 2017]

Pipeline

Approximate barycenter by a discrete uniform measure:

$$\nu = \frac{1}{m} \sum_{i=1}^{m} \delta_{x_i}$$

Solve for *positions* of the points x_i.

Make use of duality and an alternating optimization scheme.

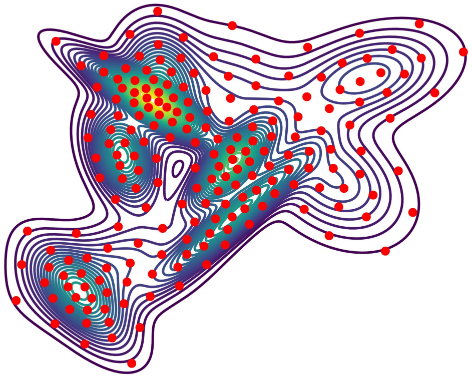
Only assume sample access to input measures.

Simplifying

Barycenter problem is additive.

Assume that we only have one input measure.

What is the best discrete approximation to a measure?



Kantorovich Dual Problem

The transport problem admits a dual formulation:

$$\sup_f \int_X f(x) \mathrm{d}\mu_1(x) + \int_X \inf_x \left\{ d(x,y)^2 - f(x) \right\} \mathrm{d}\mu_2(y)$$

Dual problem has a supply/demand interpretation.

Kantorovich Dual Problem

Can simplify when one measure is discrete:

$$F[f, x_1, \ldots, x_m] = \frac{1}{m}\sum_{i=1}^{m} f_i + \sum_{i=1}^{m} \int_{V_{x_i}} \left(d(x_i, y)^2 - f_i\right) \mathrm{d}\mu_2(y)$$

weights

Voronoi regions

Optimization -- Weights

The dual problem is concave in the weights.

Can apply gradient ascent with gradient:

$$\frac{\partial F}{\partial f_i} = \frac{1}{m} - \underbrace{\int_{V_{x_i}} \mathrm{d}\mu_2(y)}_{\text{density of cell}}$$

Optimization -- Points

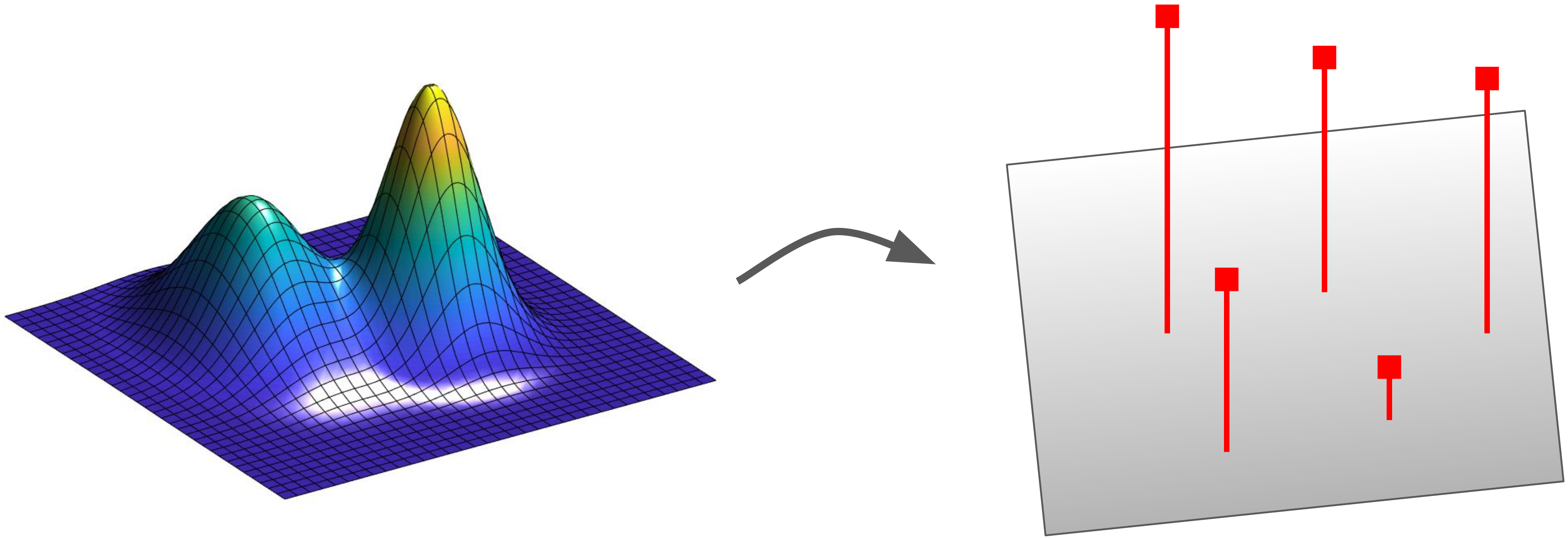
The dual is non-convex in the point positions.

Gradient yields a simple fixed point iteration scheme.

$$\frac{\partial F}{\partial x_i} = x_i \int_{V_{x_i}} \mathrm{d}\mu_2(y) - \underbrace{\int_{V_{x_i}} y \mathrm{d}\mu_2(y)}_{\text{barycenter of cell}}$$

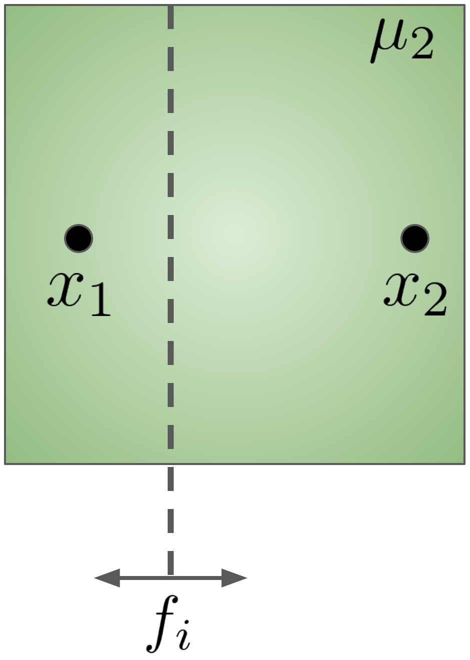
Semidiscrete Problem

Map a continuous distribution to a discrete one.



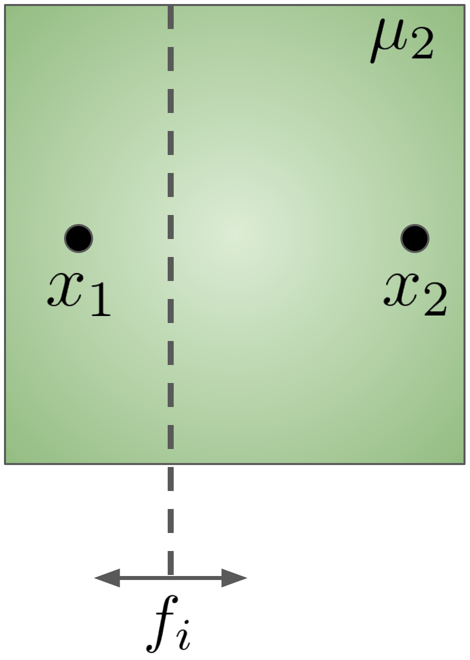
Optimization -- Weights

Adjust weights

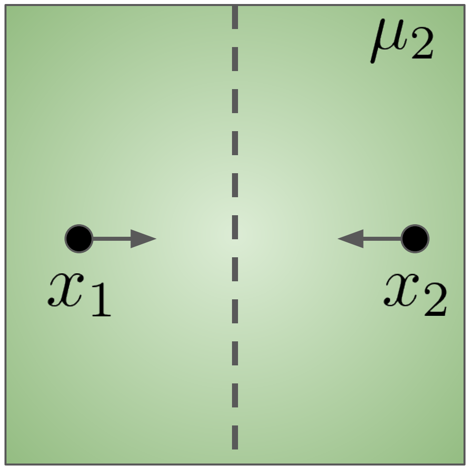


Optimization -- Points

Adjust weights



Move points



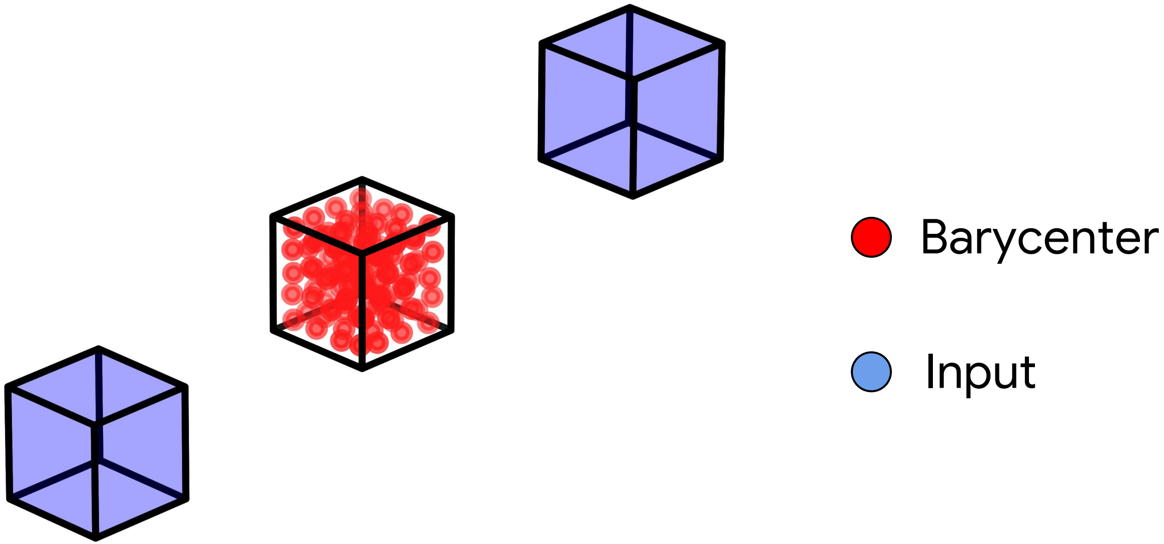
Algorithm

Extension to barycenter as simple as summing gradients.

- Sample seed points for the barycenter.
- Repeat until converged:
    1. Solve for weights via gradient ascent.
    2. Shift points via fixed point iteration.

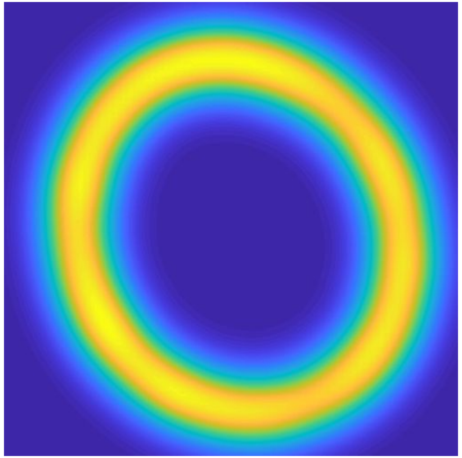
Results -- Known Barycenter

A few cases where we know what the true barycenter is

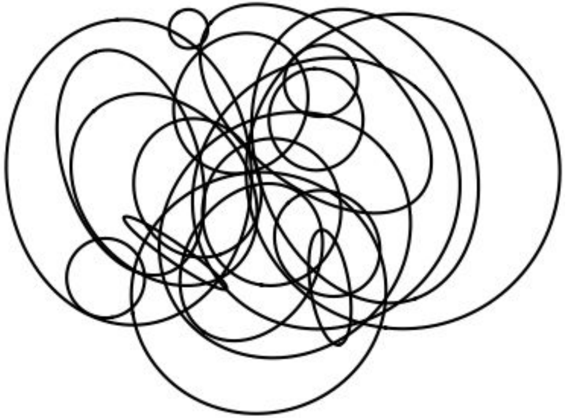


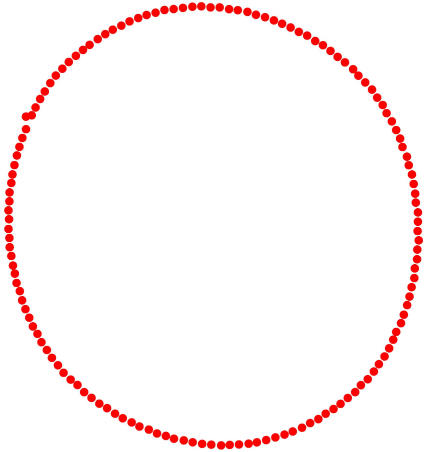
Results -- Sharp Features

Our method is inherently suited to distributions with sharp features.



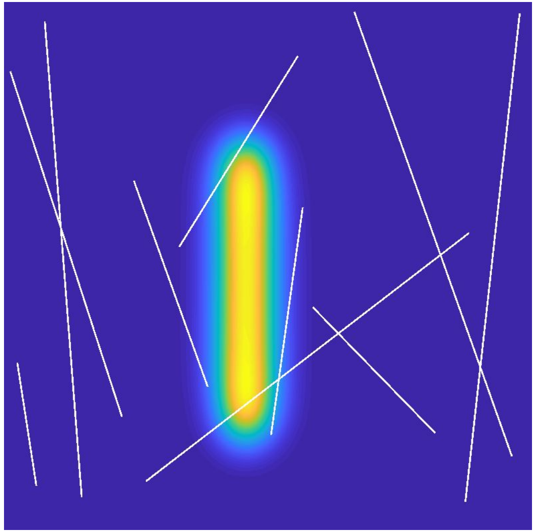
[Solomon et al., 2015]



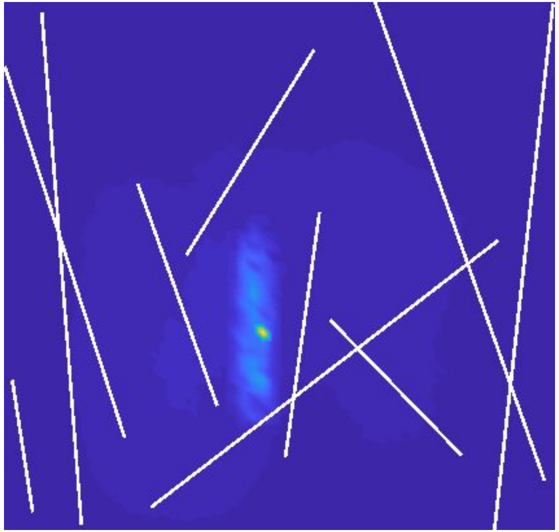


Ours -- 100 samples

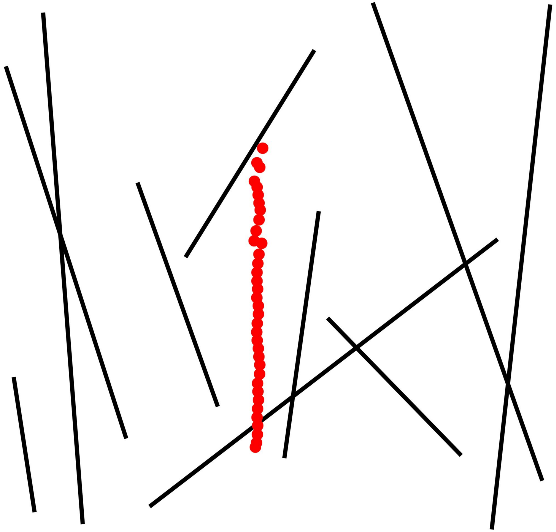
Results -- Sharp Features



[Solomon et al., 2015]



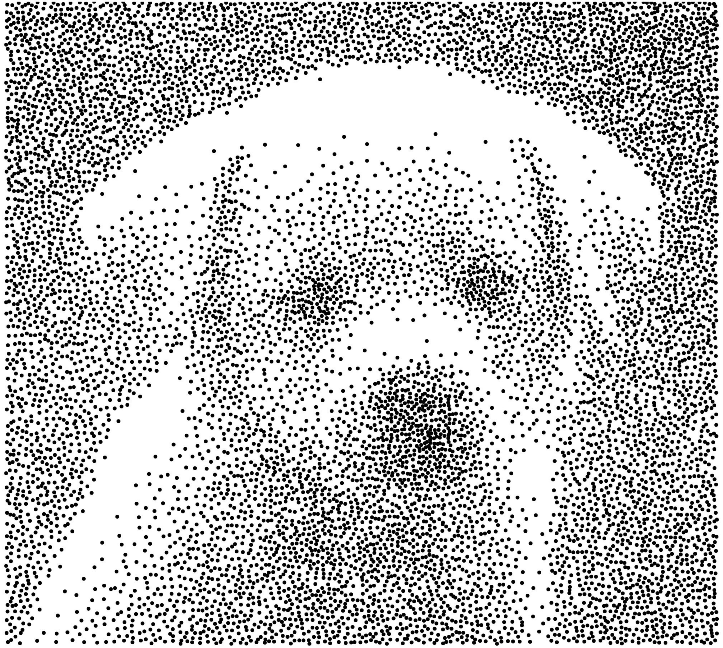
[Staib et al., 2017]



Ours -- 50 samples

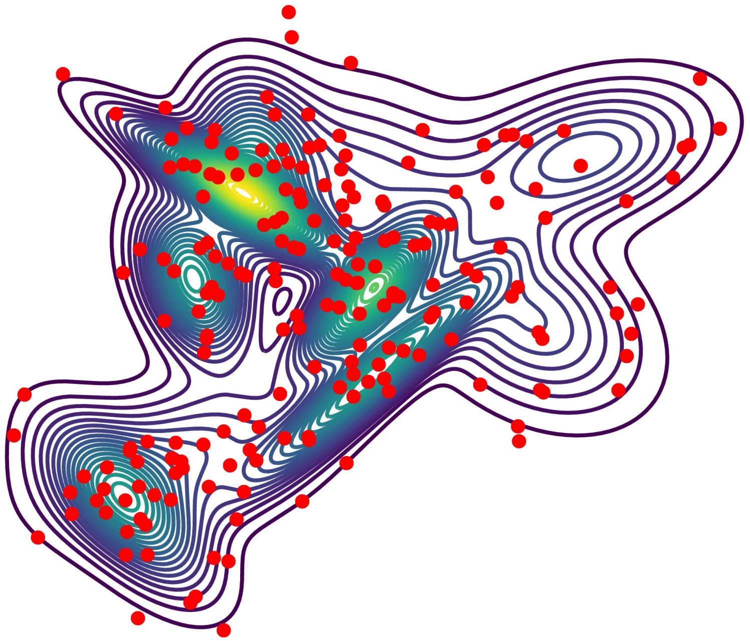
Applications -- Blue Noise Approximation

Original image

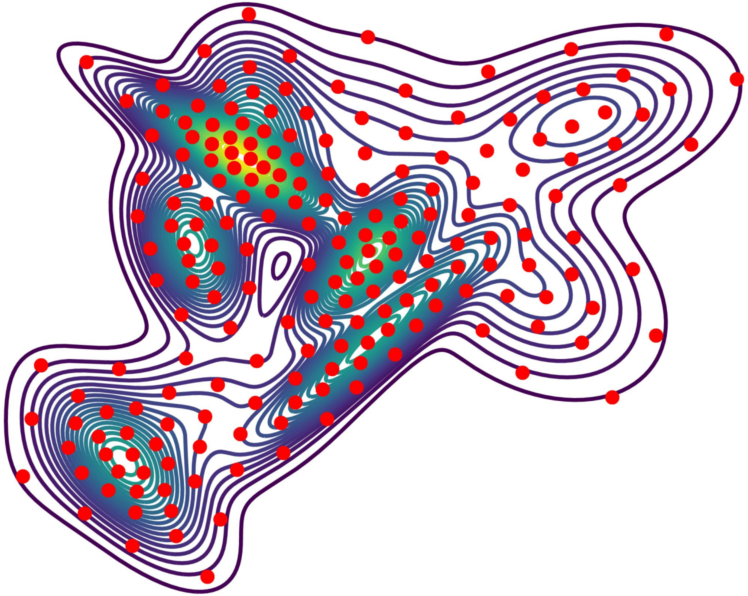


Blue noise approximation

Applications -- Super Samples



Uniform samples



Ours

Conclusions

Compute Wasserstein barycenters *without* regularization.

Works with *continuous* input distributions.

Easy to implement and parallelizable.

Come see our poster at stand #69!