Regularity as Regularization: Smooth and Strongly Convex Brenier Potentials in Optimal Transport

François-Pierre PATY *(ENSAE Paris)* Marco CUTURI *(Google Brain & ENSAE Paris)*

Goal: Estimate Wasserstein distances and Monge maps in high dimension, and regularize OT using regularity.

Approach: Enforce a Monge type optimal transport plan with prescribed regularity / distortion.

Results: Efficient algorithm in 1D, alternate convex minimization in higher dimension. New estimator for Wasserstein.

I. Regularity in Optimal Transport

2-Wasserstein Distance

$$W_2^2(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \int \|x-y\|^2 d\pi(x,y)$$

Brenier Theorem

If μ has a density wrt Lebesgue, then the optimal transport plan $\pi^\star$ is deterministic, there exists a (Monge) map $T^\star$ s.t.

$$d\pi^\star(x,y) = d\mu(x)\delta_{T^\star(x)}(y)$$

Moreover, there exists a **convex** function (Brenier potential) $f^\star$ s.t.

$$T^\star = \nabla f^\star$$

Monge-Ampère Equation

If μ and ν have densities p and q, the Brenier potential $f^\star$ is solution to

$$\det(\nabla^2 f) = \frac{p}{q\circ\nabla f}$$

Caffarelli Contraction Theorem

If V, W are convex, and $\mu = e^{V}\gamma$ and $\nu = e^{-W}\gamma$ where γ is the standard Gaussian, then any optimal Brenier potential $f^\star$ is 1-smooth.

More General Theorems: Caffarelli, De Philippis, Kim, Figalli, etc.

Under some assumptions on the measures (bounded support, convex support, bounded densities, etc.), we can get some (local, Hölder) regularity.

II. Regularity as Regularization

Smooth and Strongly convex Nearest Brenier Potentials (SSNB)

Given a partition $\mathcal{E}$ of $\mathbb{R}^d$, find a **potential** $f:\mathbb{R}^d\to\mathbb{R}$ s.t. for any cluster $E\in\mathcal{E}$, $f_{|E}$ is convex and any $x,y\in E$,

$$\ell\|x-y\| \le \|\nabla f(x) - \nabla f(y)\| \le L\|x-y\| \quad (\mathcal{F}_{\ell,L,\mathcal{E}})$$

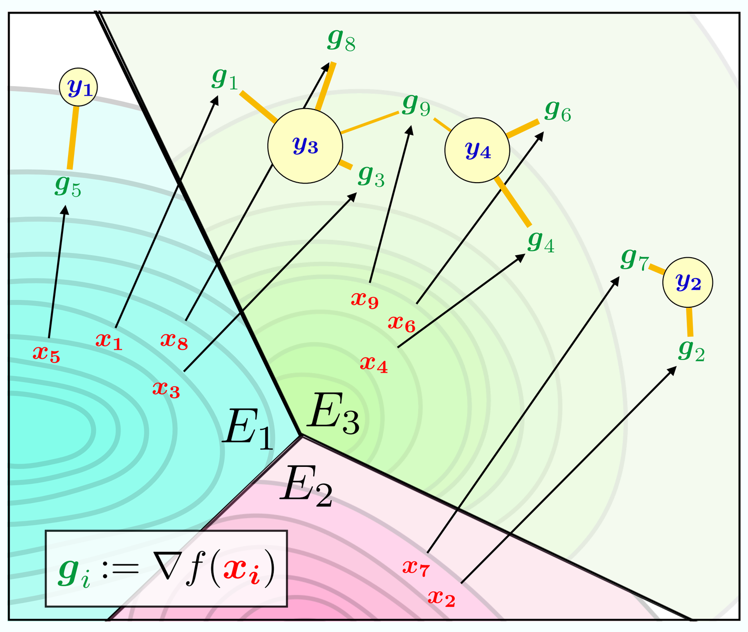
(Strong convexity and smoothness constraints)

and ∇f sends μ as close as possible to ν :

$$\min_{f\in\mathcal{F}_{\ell,L,\mathcal{E}}} W_2\left[\nabla f_\sharp\mu, \nu\right]$$

In the discrete setting $\mu = \sum_{i=1}^n a_i\delta_{x_i}$, this can be recast as an alternate convex QCQP/OT minimization:

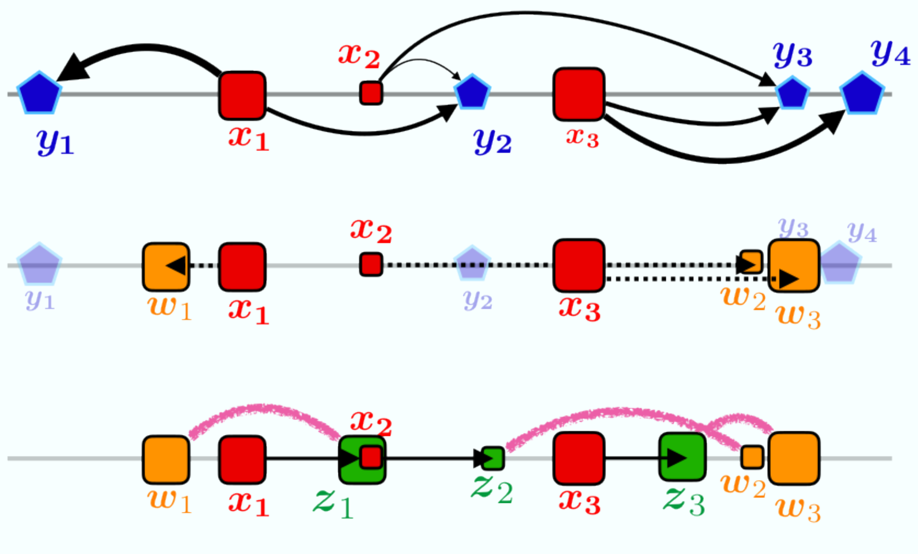
$$\min_{\substack{z_1,\dots,z_n\in\mathbb{R}^d\\ u\in\mathbb{R}^n}} W_2^2\left(\sum_{i=1}^n a_i\delta_{z_i}, \nu\right)$$



$$u_i \ge u_j + \langle z_j, x_i - x_j\rangle + \frac{1}{2(1-\ell/L)}\left(\frac{1}{L}\|z_i - z_j\|^2 + \ell\|x_i - x_j\|^2 - 2\frac{\ell}{L}\langle z_j - z_i, x_j - x_i\rangle\right)$$

III. One-Dimensional Case

1. Compute classical OT plan
2. Compute the Barycentric Projection
3. Enforce regularity via strong-monotonic and Lipschitz isotonic regression



Schiele *Van Gogh* *Classical OT* *SSNB*

IV. Estimation of Monge Map and Wasserstein Distance

Monge Map

Once a SSNB potential $f^\star$ has been computed, we can easily compute the map on any new point $x\in\mathbb{R}^d$ by solving a cheap QCQP:

$$\min_{v\in\mathbb{R},\, g\in\mathbb{R}^d} v$$
$$\text{s.t. } \forall i,\ v \ge u_i + \langle z_i, x - x_i\rangle + \frac{1}{2(1-\ell/L)}\left(\frac{1}{L}\|g - z_i\|^2 + \ell\|x - x_i\|^2 - 2\frac{\ell}{L}\langle z_i - g, x_i - x\rangle\right)$$

Wasserstein Distance

Draw n iid samples $x_1,\dots,x_n\sim\mu$ and $y_1,\dots,y_n\sim\nu$, and consider the empirical measures $\hat\mu_n, \hat\nu_n$ on those points.

The classical OT estimator of $W_2(\mu,\nu)$ is $W_2(\hat\mu_n,\hat\nu_n)$.

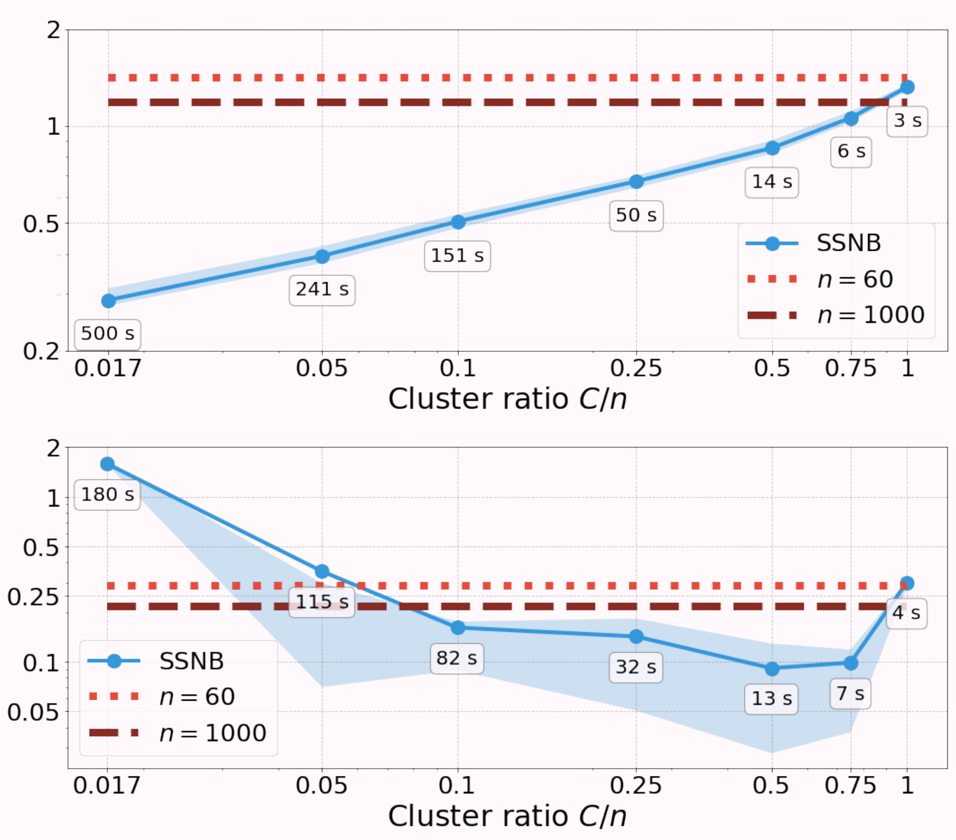
The SSNB estimator is $W_2(\mu, f^\star{}_\sharp\mu)$. If $\mathcal{E} = \{\mathbb{R}^d\}$ *(global regularity)*,

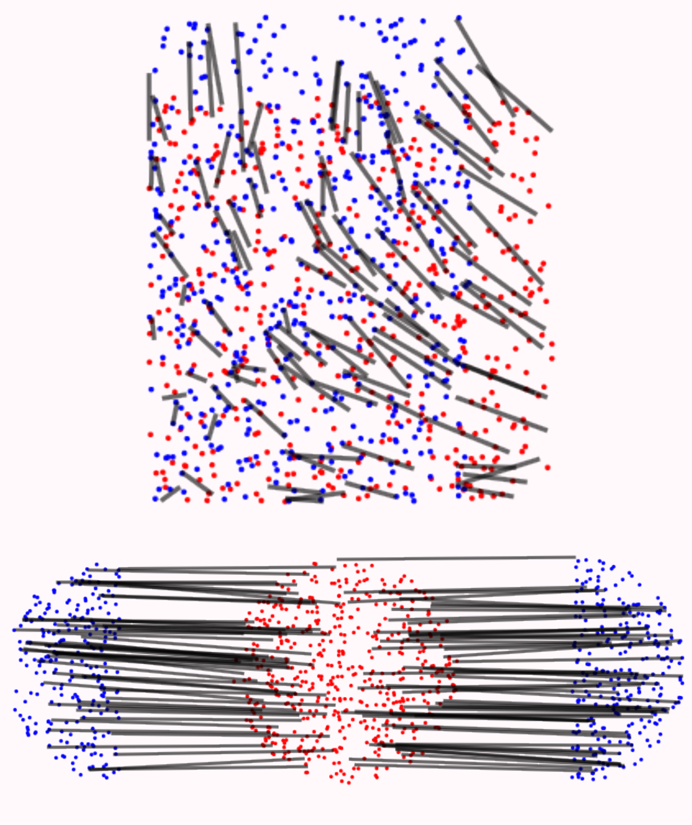
$$W_2(\mu, f^\star{}_\sharp\mu) = \left(\int \|x - \nabla f^\star(x)\|^2 d\mu(x)\right)^{1/2}$$

which can be computed using Monte-Carlo integration. If $\mathcal{E} \ne \{\mathbb{R}^d\}$ *(local regularity)*, this is only an upper bound on $W_2(\mu, f^\star{}_\sharp\mu)$.

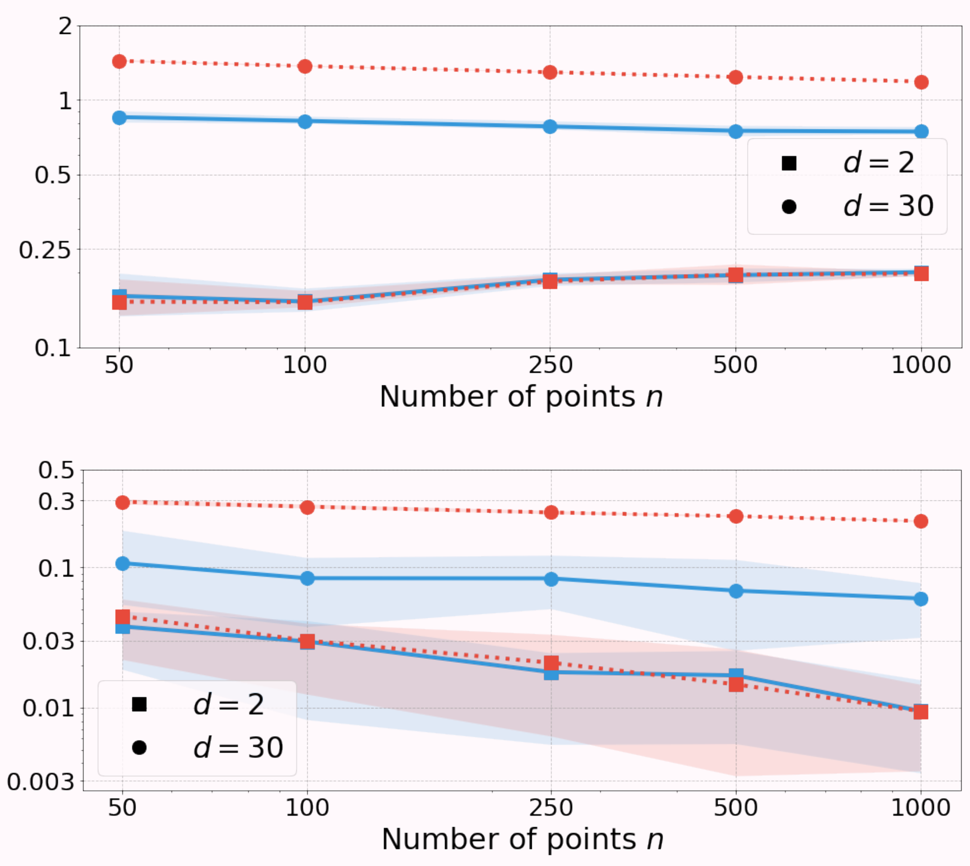
V. Experiments

Estimation Error Depending on the Number of Clusters





Estimation Error Depending on the Number of Points



Global regularity: trade-off between accuracy and computation time.

Local regularity: many small clusters are better in terms of both accuracy and computation time.

General case: SSNB estimator seem to have classical OT rate, but with a much better constant.