Quantitative Stability of Optimal Transport Maps under Variations of the Target Measure

Alex Delalande

Université Paris-Saclay & INRIA DataShape team

Based on joint works with Quentin Mérigot and Frédéric Chazal.

Orsay, June 2021

Introduction

Quadratic Optimal Transport

- Let $\Omega \subseteq \mathbb{R}^d$ and $\rho, \mu \in \mathcal{P}_2(\Omega)$.
  Quadratic Optimal Transport problem between ρ and μ:

  $$\boxed{\min_{\pi \in \Pi(\rho,\mu)} \int_{\Omega \times \Omega} ||x - y||^2 \mathrm{d}\pi(x, y),}$$

  where $\Pi(\rho, \mu) = \{\pi \in \mathcal{P}(\Omega \times \Omega) \mid (P_1)_{\#}\pi = \rho \text{ and } (P_2)_{\#}\pi = \mu\}$
  with $P_1(x, y) = x$ and $P_2(x, y) = y$.

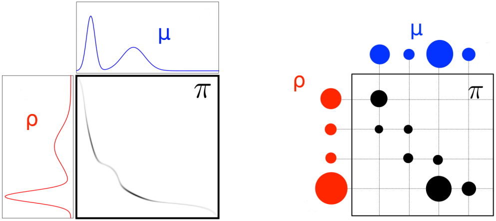
Introduction

Quadratic Optimal Transport

- Let $\Omega \subseteq \mathbb{R}^d$ and $\rho, \mu \in \mathcal{P}_2(\Omega)$.
  Quadratic Optimal Transport problem between ρ and μ:

$$\boxed{\min_{\pi \in \Pi(\rho,\mu)} \int_{\Omega \times \Omega} ||x - y||^2 \mathrm{d}\pi(x, y),}$$

where $\Pi(\rho, \mu) = \{\pi \in \mathcal{P}(\Omega \times \Omega) \mid (P_1)_{\#}\pi = \rho$ and $(P_2)_{\#}\pi = \mu\}$
with $P_1(x, y) = x$ and $P_2(x, y) = y$.



Examples of couplings π between ρ and μ (Figure from [Peyré and Cuturi (2019))]

Introduction

Quadratic Optimal Transport

- Let $\Omega \subseteq \mathbb{R}^d$ and $\rho, \mu \in \mathcal{P}_2(\Omega)$.
  Quadratic Optimal Transport problem between ρ and μ:

$$\boxed{\min_{\pi \in \Pi(\rho,\mu)} \int_{\Omega \times \Omega} ||x - y||^2 \mathrm{d}\pi(x, y),}$$

  where $\Pi(\rho, \mu) = \{\pi \in \mathcal{P}(\Omega \times \Omega) \mid (P_1)_{\#}\pi = \rho \text{ and } (P_2)_{\#}\pi = \mu\}$ with $P_1(x, y) = x$ and $P_2(x, y) = y$.
- **2-Wasserstein distance** between ρ and μ:

$$\boxed{\mathrm{W}_2(\rho, \mu) = \left(\min_{\pi \in \Pi(\rho,\mu)} \int_{\Omega \times \Omega} ||x - y||^2 \mathrm{d}\pi(x, y) \right)^{1/2}.}$$

- **2-Wasserstein space**: $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$.

Introduction

p-th Optimal Transport problem, $p \geq 1$

- Let $\Omega \subseteq \mathbb{R}^d$ and $\rho, \mu \in \mathcal{P}_2(\Omega)$.
  p-th Optimal Transport problem between ρ and μ:

$$\boxed{\min_{\pi \in \Pi(\rho,\mu)} \int_{\Omega \times \Omega} ||x - y||^p \mathrm{d}\pi(x, y),}$$

  where $\Pi(\rho, \mu) = \{\pi \in \mathcal{P}(\Omega \times \Omega) \mid (P_1)_\# \pi = \rho \text{ and } (P_2)_\# \pi = \mu\}$
  with $P_1(x, y) = x$ and $P_2(x, y) = y$.
- **p-Wasserstein distance** between ρ and μ:

$$\boxed{\mathrm{W}_p(\rho, \mu) = \left(\min_{\pi \in \Pi(\rho,\mu)} \int_{\Omega \times \Omega} ||x - y||^p \mathrm{d}\pi(x, y) \right)^{1/p}.}$$

- **p-Wasserstein space**: $(\mathcal{P}_p(\Omega), \mathrm{W}_p)$.

$$p \leq q \implies \mathrm{W}_p \leq \mathrm{W}_q$$

and if Ω compact, $\mathrm{W}_q \lesssim \mathrm{W}_p$

Introduction

Quadratic Optimal Transport

- Let $\Omega \subseteq \mathbb{R}^d$ and $\rho, \mu \in \mathcal{P}_2(\Omega)$.
  Quadratic Optimal Transport problem between ρ and μ:

$$\boxed{\min_{\pi \in \Pi(\rho,\mu)} \int_{\Omega \times \Omega} ||x - y||^2 \mathrm{d}\pi(x, y),}$$

  where $\Pi(\rho, \mu) = \{\pi \in \mathcal{P}(\Omega \times \Omega) \mid (P_1)_{\#}\pi = \rho \text{ and } (P_2)_{\#}\pi = \mu\}$
  with $P_1(x, y) = x$ and $P_2(x, y) = y$.

- **2-Wasserstein distance** between ρ and μ:

$$\boxed{\mathrm{W}_2(\rho, \mu) = \left(\min_{\pi \in \Pi(\rho,\mu)} \int_{\Omega \times \Omega} ||x - y||^2 \mathrm{d}\pi(x, y) \right)^{1/2}.}$$

- **2-Wasserstein space**: $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ is a curved metric space.

	Riemannian geometry	Optimal transport
Point	$x \in M$	$\mu \in \mathcal{P}_2(\Omega)$
Geodesic distance	$d_g(x, y)$	$\mathrm{W}_2(\mu, \nu)$

- Fix $\rho \in \mathcal{P}_{a.c.}(\Omega)$.

Brenier's theorem [Brenier (1991)]:
For any $\mu \in \mathcal{P}_2(\Omega)$, $\exists!$ $\rho-$*a.e.* $T_\mu : \mathbb{R}^d \to \mathbb{R}^d$ *such that* $(T_\mu)_{\#}\rho = \mu$ *and* $T_\mu = \nabla \phi_\mu$ *with* ϕ_μ *convex.*

$T_{\#}\rho$ is the image measure of ρ under the map T:

$$\forall B \subset \mathbb{R}^d, \quad T_{\#}\rho(B) = \rho(T^{-1}(B)).$$

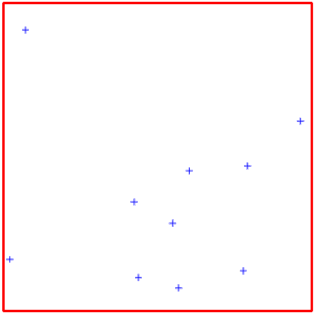
Introduction

Quadratic Optimal Transport - Brenier's theorem

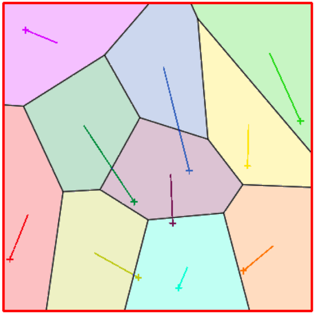
- Fix $\rho \in \mathcal{P}_{a.c.}(\Omega)$.

Brenier's theorem [Brenier (1991)]:
For any $\mu \in \mathcal{P}_2(\Omega)$, $\exists!$ $\rho-$a.e. $T_\mu : \mathbb{R}^d \to \mathbb{R}^d$ such that $(T_\mu)_\# \rho = \mu$ and $T_\mu = \nabla \phi_\mu$ with ϕ_μ convex.

Illustration:



$\rho = \lambda_{[0,1]^2}$, $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{y_i}$



T_μ

Introduction

Quadratic Optimal Transport - Brenier's theorem

- Fix $\rho \in \mathcal{P}_{a.c.}(\Omega)$.

Brenier's theorem [Brenier (1991)]:
For any $\mu \in \mathcal{P}_2(\Omega)$, $\exists!$ $\rho-$*a.e.* $T_\mu : \mathbb{R}^d \to \mathbb{R}^d$ *such that* $(T_\mu)_\# \rho = \mu$ *and* $T_\mu = \nabla \phi_\mu$ *with* ϕ_μ *convex.*

T_μ induces the optimal transport from ρ to μ ($\pi = (id, T_\mu)_\# \rho$):

$$\mathrm{W}_2(\rho, \mu) = \left(\int_\Omega ||x - T_\mu(x)||^2 \rho(x) \mathrm{d}x \right)^{1/2}.$$

Introduction

Quadratic Optimal Transport - Brenier's theorem

- Fix $\rho \in \mathcal{P}_{a.c.}(\Omega)$.

Brenier's theorem [Brenier (1991)]:
For any $\mu \in \mathcal{P}_2(\Omega)$, $\exists!$ $\rho-$a.e. $T_\mu : \mathbb{R}^d \to \mathbb{R}^d$ such that $(T_\mu)_{\#}\rho = \mu$ and $T_\mu = \nabla\phi_\mu$ with ϕ_μ convex.

Ambrosio et al. (2008):

	Riemannian geometry	Optimal transport
Point	$x \in M$	$\mu \in \mathcal{P}_2(\Omega)$
Geodesic distance	$d_g(x, y)$	$\mathrm{W}_2(\mu, \nu)$
Tangent space	$\mathcal{T}_\rho M$	$\mathcal{T}_\rho\mathcal{P}_2(\Omega) \subseteq \mathrm{L}^2(\rho, \mathbb{R}^d)$
Inverse exponential map	$\exp^{-1}(x) \in \mathcal{T}_\rho M$	$T_\mu - id \in \mathcal{T}_\rho\mathcal{P}_2(\Omega)$
Distance in tangent space	$\left\|\exp^{-1}(x) - \exp^{-1}(y)\right\|_{g(\rho)}$	$\\|T_\mu - T_\nu\\|_{\mathrm{L}^2(\rho)}$

Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$

- **"Linearized OT distance"**:

$$\boxed{\mathrm{W}_{2,\rho}(\mu,\nu) := \| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)}}$$

  - Embedding of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ into the Hilbert space $\mathrm{L}^2(\rho, \mathbb{R}^d)$.
  - Possibility to apply the "hilbertian" statistics toolbox to measures, consistently (?) with the geometry of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$.
  - Computational gains are possible.

Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 1: approximation of Wasserstein barycenters

- Wasserstein barycenters: $\mu_1, \ldots, \mu_K \in \mathcal{P}_2(\Omega), \alpha_1, \ldots, \alpha_K \geq 0$:

$$\bar{\mu} := \arg \min_{\mu \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{K} \alpha_i \mathrm{W}_2^2(\mu, \mu_i).$$

- "Linearized" Wasserstein barycenters:

$$\hat{\mu} := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$$

Remark: The image of the embedding $\{ T_\mu \quad | \quad \mu \in \mathcal{P}_2(\Omega) \}$ is convex!

Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 1: approximation of Wasserstein barycenters

- Wasserstein barycenters: $\mu_1, \ldots, \mu_K \in \mathcal{P}_2(\Omega), \alpha_1, \ldots, \alpha_K \geq 0$:

$$\bar{\mu} := \arg \min_{\mu \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{K} \alpha_i \mathrm{W}_2^2(\mu, \mu_i).$$

- "Linearized" Wasserstein barycenters:

$$\hat{\mu} := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$$

Illustration:

Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 1: approximation of Wasserstein barycenters

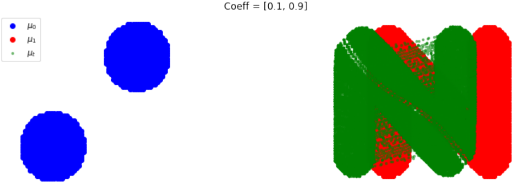
- Wasserstein barycenters: $\mu_1, \ldots, \mu_K \in \mathcal{P}_2(\Omega), \alpha_1, \ldots, \alpha_K \geq 0$:

$$\bar{\mu} := \arg\min_{\mu \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{K} \alpha_i \mathrm{W}_2^2(\mu, \mu_i).$$

- "Linearized" Wasserstein barycenters:

$$\hat{\mu} := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$$

Illustration:



Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 1: approximation of Wasserstein barycenters

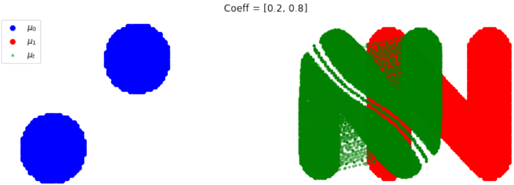
- Wasserstein barycenters: $\mu_1, \dots, \mu_K \in \mathcal{P}_2(\Omega), \alpha_1, \dots, \alpha_K \geq 0$:

$$\bar{\mu} := \arg \min_{\mu \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{K} \alpha_i \mathrm{W}_2^2(\mu, \mu_i).$$

- "Linearized" Wasserstein barycenters:

$$\hat{\mu} := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$$

Illustration:



Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 1: approximation of Wasserstein barycenters

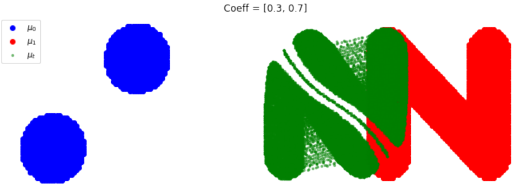
- Wasserstein barycenters: $\mu_1, \ldots, \mu_K \in \mathcal{P}_2(\Omega), \alpha_1, \ldots, \alpha_K \geq 0$:

$$\bar{\mu} := \arg \min_{\mu \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{K} \alpha_i \mathrm{W}_2^2(\mu, \mu_i).$$

- "Linearized" Wasserstein barycenters:

$$\hat{\mu} := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$$

Illustration:



Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 1: approximation of Wasserstein barycenters

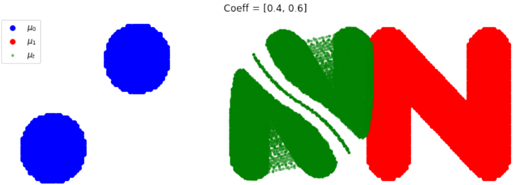
- Wasserstein barycenters: $\mu_1, \dots, \mu_K \in \mathcal{P}_2(\Omega), \alpha_1, \dots, \alpha_K \geq 0$:

$$\bar{\mu} := \arg \min_{\mu \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{K} \alpha_i \mathrm{W}_2^2(\mu, \mu_i).$$

- "Linearized" Wasserstein barycenters:

$$\hat{\mu} := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$$

Illustration:



Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 1: approximation of Wasserstein barycenters

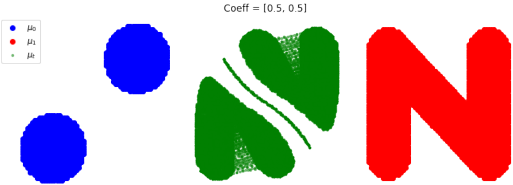
- Wasserstein barycenters: $\mu_1, \ldots, \mu_K \in \mathcal{P}_2(\Omega), \alpha_1, \ldots, \alpha_K \geq 0$:

$$\bar{\mu} := \arg \min_{\mu \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{K} \alpha_i \mathrm{W}_2^2(\mu, \mu_i).$$

- "Linearized" Wasserstein barycenters:

$$\hat{\mu} := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$$

Illustration:



Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 1: approximation of Wasserstein barycenters

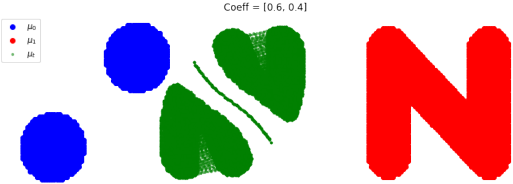
- Wasserstein barycenters: $\mu_1, \ldots, \mu_K \in \mathcal{P}_2(\Omega), \alpha_1, \ldots, \alpha_K \geq 0$:

$$\bar{\mu} := \arg \min_{\mu \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{K} \alpha_i \mathrm{W}_2^2(\mu, \mu_i).$$

- "Linearized" Wasserstein barycenters:

$$\hat{\mu} := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$$

Illustration:



Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 1: approximation of Wasserstein barycenters

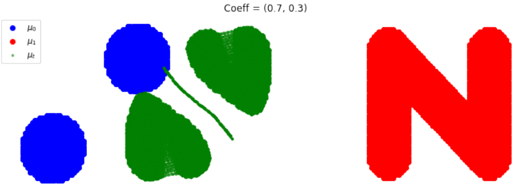
- Wasserstein barycenters: $\mu_1, \ldots, \mu_K \in \mathcal{P}_2(\Omega), \alpha_1, \ldots, \alpha_K \geq 0$:

$$\bar{\mu} := \arg \min_{\mu \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{K} \alpha_i \mathrm{W}_2^2(\mu, \mu_i).$$

- "Linearized" Wasserstein barycenters:

$$\hat{\mu} := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$$

Illustration:



Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 1: approximation of Wasserstein barycenters

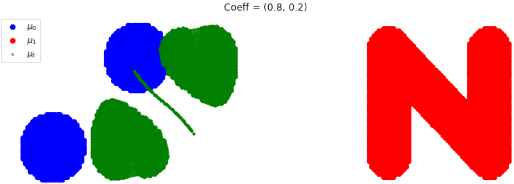
- Wasserstein barycenters: $\mu_1, \dots, \mu_K \in \mathcal{P}_2(\Omega), \alpha_1, \dots, \alpha_K \geq 0$:

$$\bar{\mu} := \arg \min_{\mu \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{K} \alpha_i \mathrm{W}_2^2(\mu, \mu_i).$$

- "Linearized" Wasserstein barycenters:

$$\hat{\mu} := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$$

Illustration:



Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 1: approximation of Wasserstein barycenters

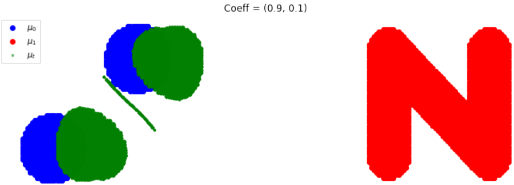
- Wasserstein barycenters: $\mu_1, \ldots, \mu_K \in \mathcal{P}_2(\Omega), \alpha_1, \ldots, \alpha_K \geq 0$:

$$\bar{\mu} := \arg \min_{\mu \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{K} \alpha_i \mathrm{W}_2^2(\mu, \mu_i).$$

- "Linearized" Wasserstein barycenters:

$$\hat{\mu} := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$$

Illustration:



Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 1: approximation of Wasserstein barycenters

- Wasserstein barycenters: $\mu_1, \ldots, \mu_K \in \mathcal{P}_2(\Omega), \alpha_1, \ldots, \alpha_K \geq 0$:

$$\bar{\mu} := \arg \min_{\mu \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{K} \alpha_i \mathrm{W}_2^2(\mu, \mu_i).$$

- "Linearized" Wasserstein barycenters:

$$\hat{\mu} := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$$

Illustration:

Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ - Example 2: clustering of the MNIST dataset

The MNIST training set is made of 60,000 images of handwritten digits (28×28 grayscale images).

Push-forwards of the 20 centroids after clustering of the Brenier maps of the MNIST training set.

Introduction

"Linearization" of $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$

- **2-Wasserstein distance** between ρ and μ:

$$\mathrm{W}_2(\rho, \mu) = \left(\min_{\pi \in \Pi(\rho,\mu)} \int_{\Omega \times \Omega} ||x - y||^2 \mathrm{d}\pi(x, y) \right)^{1/2}.$$

- **2-Wasserstein space**: $(\mathcal{P}_2(\Omega), \mathrm{W}_2)$ is a curved metric space.
- **"Linearized OT distance"**:

$$\mathrm{W}_{2,\rho}(\mu, \nu) := \| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)}$$

How good is the approximation $\mathrm{W}_2(\mu, \nu) \approx \mathrm{W}_{2,\rho}(\mu, \nu)$?

Setting

- Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact and convex set. Let $\rho \in \mathcal{P}_{a.c.}(\mathcal{X})$ with a density s.t.
$$0 < m_\rho \leq \rho \leq M_\rho < +\infty.$$

- For $\Omega \subseteq \mathbb{R}^d$, we study the stability of the mapping
$$\boxed{\mu \in (\mathcal{P}_2(\Omega), \mathrm{W}_2) \quad \mapsto \quad T_\mu \in \mathrm{L}^2(\rho; \Omega).}$$

Elementary properties

- $\mu \mapsto T_\mu$ is reverse-Lipschitz: $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \geq \mathrm{W}_2(\mu, \nu)$

Elementary properties

- $\mu \mapsto T_\mu$ is reverse-Lipschitz: $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \geq \mathrm{W}_2(\mu, \nu)$
  Indeed one has $\pi := (T_\mu, T_\nu)_\# \rho \in \Pi(\mu, \nu)$.
  $\implies \mathrm{W}_2^2(\mu, \nu) \leq \int_{\Omega \times \Omega} \|x - y\|^2 \, \mathrm{d}\pi(x, y) = \int_\Omega \|T_\mu(x) - T_\nu(x)\|^2 \, \mathrm{d}\rho(x)$.

Elementary properties

- $\mu \mapsto T_\mu$ is reverse-Lipschitz: $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \geq \mathrm{W}_2(\mu, \nu)$

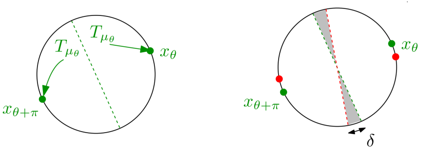
- $\mu \mapsto T_\mu$ is continuous (Corollary 5.23 in Villani (2008))

Elementary properties

- $\mu \mapsto T_\mu$ is reverse-Lipschitz: $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \geq \mathrm{W}_2(\mu, \nu)$
- $\mu \mapsto T_\mu$ is continuous (Corollary 5.23 in Villani (2008))
- $\mu \mapsto T_\mu$ is not better than $\frac{1}{2}$-Hölder

Elementary properties

- $\mu \mapsto T_\mu$ is reverse-Lipschitz: $\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \geq \mathrm{W}_2(\mu, \nu)$
- $\mu \mapsto T_\mu$ is continuous (Corollary 5.23 in Villani (2008))
- $\mu \mapsto T_\mu$ is not better than $\frac{1}{2}$-Hölder
  Take $\rho := \frac{1}{\pi}\mathrm{Leb}_{B(0,1)}$ on $\mathbb{R}^2$ and $\mu_\theta := \frac{\delta_{x_\theta} + \delta_{x_{\theta+\pi}}}{2}$ with $x_\theta = (\cos(\theta), \sin(\theta))$. Then $|| T_{\mu_\theta} - T_{\mu_{\theta+\delta}} ||^2_{\mathrm{L}^2(\rho)} \geq C\delta$ while $\mathrm{W}_2(\mu_\theta, \mu_{\theta+\delta}) \leq C\delta$.



Elementary properties

- $\mu \mapsto T_\mu$ is reverse-Lipschitz: $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \geq \mathrm{W}_2(\mu, \nu)$
- $\mu \mapsto T_\mu$ is continuous (Corollary 5.23 in Villani (2008))
- $\mu \mapsto T_\mu$ is not better than $\frac{1}{2}$-Hölder

Quantitative continuity of $\mu \mapsto T_\mu$?

Known stability results

A negative result

Theorem(Andoni, Naor, Neiman):
$(\mathcal{P}_2(\mathbb{R}^3), \mathrm{W}_2)$ *does not admit a uniform, coarse or quasisymmetric embedding into any* L^p *space.*

Andoni, A., Naor, A., and Neiman, O. (2018). Snowflake universality of Wasserstein spaces.
Annales Scientifiques de l'Ecole Normale Superieure, 51(3):657–700.

Known stability results

A negative result

Theorem(Andoni, Naor, Neiman):
$(\mathcal{P}_2(\mathbb{R}^3), \mathrm{W}_2)$ *does not admit a uniform, coarse or quasisymmetric embedding into any* L^p *space.*

Andoni, A., Naor, A., and Neiman, O. (2018). Snowflake universality of Wasserstein spaces.
Annales Scientifiques de l'Ecole Normale Superieure, 51(3):657–700.

Definition (Metric space embeddings): For $(X, d_X), (Y, d_Y)$ two metric spaces and $f : X \to Y$, the embedding f is said to be:

- *uniform* if both f and f^{-1} are uniformly continuous.
- *coarse* if there exists non-decreasing functions $\alpha, \beta : \mathbb{R}_+ \to \mathbb{R}_+$ with $\lim_{t\to\infty} \alpha(t) = \infty$ s.t. for every $x, y \in X$,

$$\alpha(d_X(x, y)) \leq d_Y(f(x), f(y)) \leq \beta(d_X(x, y)).$$

Known stability results

A negative result

Theorem(Andoni, Naor, Neiman):
$(\mathcal{P}_2(\mathbb{R}^3), \mathrm{W}_2)$ *does not admit a uniform, coarse or quasisymmetric embedding into any* L^p *space.*

Andoni, A., Naor, A., and Neiman, O. (2018). Snowflake universality of Wasserstein spaces.
Annales Scientifiques de l'Ecole Normale Superieure, 51(3):657–700.

$\implies$ $\mu \mapsto T_\mu$ **is not bi-Hölder on the whole set** $\mathcal{P}_2(\mathbb{R}^d)$

Known stability results

$\frac{1}{2}$-Hölder continuity near a regular measure (compact support)

Theorem(Ambrosio, reported in Gigli (2011)):
Let $\Omega \subset \mathbb{R}^d$ be a compact set and $\mu, \nu \in \mathcal{P}(\Omega)$. Assume that the Brenier map T_μ from ρ to μ is L-Lipschitz. Then,

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq 2\sqrt{\mathrm{diam}(\mathcal{X}) L}\mathrm{W}_1(\mu, \nu)^{1/2}.$$

Gigli, N. (2011). On hölder continuity-in-time of the optimal transport map towards measures along a curve.
Proceedings of the Edinburgh Mathematical Society, 54(2):401–409.

Known stability results

$\frac{1}{2}$-Hölder continuity near a regular measure (compact support)

Theorem(Ambrosio, reported in Gigli (2011)):
Let $\Omega \subset \mathbb{R}^d$ be a compact set and $\mu, \nu \in \mathcal{P}(\Omega)$. Assume that the Brenier map T_μ from ρ to μ is L-Lipschitz. Then,

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq 2\sqrt{\mathrm{diam}(\mathcal{X}) L}\,\mathrm{W}_1(\mu, \nu)^{1/2}.$$

- No regularity assumption on $\nu \implies$ applicable in statistics and numerical analysis.

Known stability results

$\frac{1}{2}$-Hölder continuity near a regular measure (compact support)

Theorem(Ambrosio, reported in Gigli (2011)):
Let $\Omega \subset \mathbb{R}^d$ be a compact set and $\mu, \nu \in \mathcal{P}(\Omega)$. Assume that the Brenier map T_μ from ρ to μ is L-Lipschitz. Then,

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq 2\sqrt{\mathrm{diam}(\mathcal{X}) L}\mathrm{W}_1(\mu, \nu)^{1/2}.$$

- No regularity assumption on $\nu \implies$ applicable in statistics and numerical analysis.
- The hypothesis that $x \mapsto T_\mu(x)$ is Lipschitz is very restricting in practice:
  - it implies that $\mathrm{spt}(\mu)$ must be at least connected

Known stability results

$\frac{1}{2}$-Hölder continuity near a regular measure (compact support)

Theorem(Ambrosio, reported in Gigli (2011)):
Let $\Omega \subset \mathbb{R}^d$ be a compact set and $\mu, \nu \in \mathcal{P}(\Omega)$. Assume that the Brenier map T_μ from ρ to μ is L-Lipschitz. Then,

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq 2\sqrt{\mathrm{diam}(\mathcal{X}) L}\, \mathrm{W}_1(\mu, \nu)^{1/2}.$$

- No regularity assumption on $\nu \implies$ applicable in statistics and numerical analysis.
- The hypothesis that $x \mapsto T_\mu(x)$ is Lipschitz is very restricting in practice:
  - it implies that $\mathrm{spt}(\mu)$ must be at least connected
  - it can be proven only under very strong conditions on the data: e.g. if ρ, μ are absolutely continuous on smooth uniformly convex sets, with $\mathcal{C}^{0,\alpha}$ densities bounded from above and below, then T_μ is $\mathcal{C}^{1,\alpha}$.

Known stability results

A global Hölder continuity result (compact support)

Theorem(Berman (2020)):
Let $\Omega \subset \mathbb{R}^d$ *be a compact set and* $\mu, \nu \in \mathcal{P}(\Omega)$*. If* $\rho \equiv 1$ *on* $\mathcal{X}$*, then,*

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq C_{\mathcal{X},\Omega} \mathrm{W}_1(\mu, \nu)^{\frac{1}{2^{(d-1)}(d+2)}}.$$

Berman, R. J. (2020). Convergence rates for discretized monge–ampère equations and quantitative stability of optimal transport. *Foundations of Computational Mathematics*.

Known stability results

A global Hölder continuity result (compact support)

Theorem(Berman (2018)):
Let $\Omega \subset \mathbb{R}^d$ *be a compact set and* $\mu, \nu \in \mathcal{P}(\Omega)$*. If* $\rho \equiv 1$ *on* $\mathcal{X}$*, then,*

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \leq C_{\mathcal{X},\Omega} \mathrm{W}_1(\mu,\nu)^{\frac{1}{2^{(d-1)}(d+2)}}.$$

- No regularity assumption on μ and ν, only a compact support.

Known stability results

A global Hölder continuity result (compact support)

Theorem(Berman (2018)):
Let $\Omega \subset \mathbb{R}^d$ be a compact set and $\mu, \nu \in \mathcal{P}(\Omega)$. If $\rho \equiv 1$ on $\mathcal{X}$, then,

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq C_{\mathcal{X},\Omega} \mathrm{W}_1(\mu, \nu)^{\frac{1}{2^{(d-1)}(d+2)}}.$$

- No regularity assumption on μ and ν, only a compact support.
- The Hölder exponent is not tight.

Global, dimension-independent Hölder-continuity of $\mu \mapsto T_\mu$

New stability results

The compact case

Theorem(Chazal, Mérigot, D.):
Let $\Omega \subset \mathbb{R}^d$ *be a compact set and* $\mu, \nu \in \mathcal{P}(\Omega)$. *Then,*

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \leq C_{d,\mathcal{X},\Omega} \mathrm{W}_1(\mu,\nu)^{1/6}.$$

New stability results

The compact case

Theorem(Chazal, Mérigot, D.):
Let $\Omega \subset \mathbb{R}^d$ *be a compact set and* $\mu, \nu \in \mathcal{P}(\Omega)$*. Then,*

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq C_{d,\mathcal{X},\Omega} \mathrm{W}_1(\mu, \nu)^{1/6}.$$

- First global and dimension-independent stability result for optimal transport maps.

New stability results

The compact case

Theorem(Chazal, Mérigot, D.):
Let $\Omega \subset \mathbb{R}^d$ *be a compact set and* $\mu, \nu \in \mathcal{P}(\Omega)$. *Then,*

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq C_{d,\mathcal{X},\Omega} \mathrm{W}_1(\mu, \nu)^{1/6}.$$

- First global and dimension-independent stability result for optimal transport maps.
- Gap between lower-bound and upper bound for Hölder exponent: $\frac{1}{6} < \frac{1}{2}$.

New stability results

The compact case

Theorem(Chazal, Mérigot, D.):
Let $\Omega \subset \mathbb{R}^d$ be a compact set and $\mu, \nu \in \mathcal{P}(\Omega)$. Then,

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq C_{d,\mathcal{X},\Omega} \mathrm{W}_1(\mu,\nu)^{1/6}.$$

- First global and dimension-independent stability result for optimal transport maps.
- Gap between lower-bound and upper bound for Hölder exponent: $\frac{1}{6} < \frac{1}{2}$.
- The constant is explicit, $C_{d,\mathcal{X},\Omega} \approx \mathrm{diam}(\mathcal{X})^{d+1} \mathrm{diam}(\Omega)$.

New stability results

The compact case - Sketch of proof

0. Introduction of the Kantorovich functional:

$$\min_{\pi\in\Pi(\rho,\mu)} \int_{\Omega\times\Omega} \|x-y\|^2 \,\mathrm{d}\pi(x,y) = M_2(\rho) + M_2(\mu) - 2 \max_{\pi\in\Pi(\rho,\mu)} \int_{\Omega\times\Omega} \langle x|y\rangle \mathrm{d}\pi(x,y).$$

New stability results

The compact case - Sketch of proof

0. Introduction of the Kantorovich functional:

$$\min_{\pi\in\Pi(\rho,\mu)} \int_{\mathcal{X}\times\Omega} \|x-y\|^2 \,\mathrm{d}\pi(x,y) = \quad M_2(\rho) + M_2(\mu) - 2 \underbrace{\max_{\pi\in\Pi(\rho,\mu)} \int_{\mathcal{X}\times\Omega} \langle x|y\rangle \mathrm{d}\pi(x,y).}_{=\min_{\psi\in\mathcal{C}_b(\Omega)} \int_{\mathcal{X}} \psi^* \mathrm{d}\rho + \int_\Omega \psi \mathrm{d}\mu}$$

Legendre transform of ψ:

$$\psi^*(x) = \max_{y\in\Omega} \langle x|y\rangle - \psi(y)$$

New stability results

The compact case - Sketch of proof

0. Introduction of the Kantorovich functional:

$$\min_{\psi \in \mathcal{C}_b(\Omega)} \int_{\mathcal{X}} \psi^* \mathrm{d}\rho + \int_{\Omega} \psi \mathrm{d}\mu = \min_{\psi \in \mathcal{C}_b(\Omega)} \mathcal{K}(\psi) + \int_{\Omega} \psi \mathrm{d}\mu,$$

where $\mathcal{K}(\psi) = \int_{\mathcal{X}} \psi^* \mathrm{d}\rho$.

New stability results

The compact case - Sketch of proof

0. Introduction of the Kantorovich functional:

$$\min_{\psi \in \mathcal{C}_b(\Omega)} \int_{\mathcal{X}} \psi^* \mathrm{d}\rho + \int_\Omega \psi \mathrm{d}\mu = \min_{\psi \in \mathcal{C}_b(\Omega)} \mathcal{K}(\psi) + \int_\Omega \psi \mathrm{d}\mu,$$

where $\mathcal{K}(\psi) = \int_{\mathcal{X}} \psi^* \mathrm{d}\rho$.

- $\mathcal{K}$ is convex and $\nabla \mathcal{K}(\psi) = -\left(\nabla \psi^*\right)_{\#} \rho$.

New stability results

The compact case - Sketch of proof

0. Introduction of the Kantorovich functional:

$$\min_{\psi \in \mathcal{C}_b(\Omega)} \int_{\mathcal{X}} \psi^* \mathrm{d}\rho + \int_{\Omega} \psi \mathrm{d}\mu = \min_{\psi \in \mathcal{C}_b(\Omega)} \mathcal{K}(\psi) + \int_{\Omega} \psi \mathrm{d}\mu,$$

where $\mathcal{K}(\psi) = \int_{\mathcal{X}} \psi^* \mathrm{d}\rho$.

- $\mathcal{K}$ is convex and $\nabla \mathcal{K}(\psi) = -\left(\nabla \psi^*\right)_{\#} \rho$.
- $\implies$ For a minimzer ψ_μ, $\left(\nabla \psi_\mu^*\right)_{\#} \rho = \mu$.

New stability results

The compact case - Sketch of proof

0. Introduction of the Kantorovich functional:

$$\min_{\psi \in \mathcal{C}_b(\Omega)} \int_{\mathcal{X}} \psi^* \mathrm{d}\rho + \int_\Omega \psi \mathrm{d}\mu = \min_{\psi \in \mathcal{C}_b(\Omega)} \mathcal{K}(\psi) + \int_\Omega \psi \mathrm{d}\mu,$$

where $\mathcal{K}(\psi) = \int_{\mathcal{X}} \psi^* \mathrm{d}\rho$.

- $\mathcal{K}$ is convex and $\nabla \mathcal{K}(\psi) = -\left(\nabla \psi^*\right)_{\#} \rho$.
- $\implies$ For a minimzer ψ_μ, $\left(\nabla \psi_\mu^*\right)_{\#} \rho = \mu$.
- Relation to the Brenier map: $\nabla \psi_\mu^* = T_\mu$.

New stability results

The compact case - Sketch of proof

0. Introduction of the Kantorovich functional:

$$\min_{\psi \in \mathcal{C}_b(\Omega)} \int_{\mathcal{X}} \psi^* \mathrm{d}\rho + \int_\Omega \psi \mathrm{d}\mu = \min_{\psi \in \mathcal{C}_b(\Omega)} \mathcal{K}(\psi) + \int_\Omega \psi \mathrm{d}\mu,$$

where $\mathcal{K}(\psi) = \int_{\mathcal{X}} \psi^* \mathrm{d}\rho$.

- $\mathcal{K}$ is convex and $\nabla \mathcal{K}(\psi) = -\left(\nabla \psi^*\right)_{\#} \rho$.
- $\implies$ For a minimzer ψ_μ, $\left(\nabla \psi_\mu^*\right)_{\#} \rho = \mu$.
- Relation to the Brenier map: $\nabla \psi_\mu^* = T_\mu$.
- Stability of $\mu \mapsto T_\mu$ can be deduced from stability of $\mu \mapsto \psi_\mu$.

New stability results

The compact case - Sketch of proof

0. Introduction of the Kantorovich functional:

$$\min_{\psi \in \mathcal{C}_b(\Omega)} \int_{\mathcal{X}} \psi^* \mathrm{d}\rho + \int_{\Omega} \psi \mathrm{d}\mu = \min_{\psi \in \mathcal{C}_b(\Omega)} \mathcal{K}(\psi) + \int_{\Omega} \psi \mathrm{d}\mu,$$

where $\mathcal{K}(\psi) = \int_{\mathcal{X}} \psi^* \mathrm{d}\rho$.

- $\mathcal{K}$ is convex and $\nabla \mathcal{K}(\psi) = -\left(\nabla \psi^*\right)_{\#} \rho$.
- $\implies$ For a minimzer ψ_μ, $\left(\nabla \psi_\mu^*\right)_{\#} \rho = \mu$.
- Relation to the Brenier map: $\nabla \psi_\mu^* = T_\mu$.
- Stability of $\mu \mapsto T_\mu$ can be deduced from stability of $\mu \mapsto \psi_\mu$.
- Optimality condition: $\nabla \mathcal{K}(\psi_\mu) = -\mu \iff \psi_\mu = (\nabla \mathcal{K})^{-1}(-\mu)$.

New stability results

The compact case - Sketch of proof

0. Introduction of the Kantorovich functional:

$$\min_{\psi \in \mathcal{C}_b(\Omega)} \int_{\mathcal{X}} \psi^* \mathrm{d}\rho + \int_{\Omega} \psi \mathrm{d}\mu = \min_{\psi \in \mathcal{C}_b(\Omega)} \mathcal{K}(\psi) + \int_{\Omega} \psi \mathrm{d}\mu,$$

where $\mathcal{K}(\psi) = \int_{\mathcal{X}} \psi^* \mathrm{d}\rho$.

- $\mathcal{K}$ is convex and $\nabla \mathcal{K}(\psi) = -\left(\nabla \psi^*\right)_{\#} \rho$.
- $\implies$ For a minimzer ψ_μ, $\left(\nabla \psi_\mu^*\right)_{\#} \rho = \mu$.
- Relation to the Brenier map: $\nabla \psi_\mu^* = T_\mu$.
- Stability of $\mu \mapsto T_\mu$ can be deduced from stability of $\mu \mapsto \psi_\mu$.
- Optimality condition: $\nabla \mathcal{K}(\psi_\mu) = -\mu \iff \psi_\mu = (\nabla \mathcal{K})^{-1}(-\mu)$.
- $\implies$ Need for a strong convexity estimate of $\mathcal{K}$. In the absence of regularization, $\mathcal{K}$ is not globally strongly convex.

New stability results

The compact case - Sketch of proof

1. Local strong convexity of the Kantorovich functional:

Let $\mu^0, \mu^1 \in \mathcal{P}(\Omega)$ and for $k \in \{0, 1\}$,

$$\psi^k \in \arg\min_{\psi \in \mathcal{C}_b(\Omega)} \mathcal{K}(\psi) + \int \psi \mathrm{d}\mu^k.$$

New stability results

The compact case - Sketch of proof

1. Local strong convexity of the Kantorovich functional:

Let $\mu^0, \mu^1 \in \mathcal{P}(\Omega)$ and for $k \in \{0,1\}$,

$$\psi^k \in \arg\min_{\psi \in \mathcal{C}_b(\Omega)} \mathcal{K}(\psi) + \int \psi \mathrm{d}\mu^k.$$

For $t \in [0,1]$ denote $\psi^t = (1-t)\psi^0 + t\psi^1 = \psi^0 + tv$, and notice that:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}(\psi^t) = -\mathbb{E}_\rho v(\nabla\psi^{t*}),$$

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{K}(\psi^t) = \mathbb{E}_\rho \langle \nabla v(\nabla\psi^{t*}) | \left(\mathrm{D}^2\psi^t\right)^{-1} \nabla v(\nabla\psi^{t*})\rangle.$$

New stability results

The compact case - Sketch of proof

1. Local strong convexity of the Kantorovich functional:

Brascamp-Lieb inequality:

$$\boxed{\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{K}(\psi^t) = \mathbb{E}_\rho\langle \nabla v(\nabla\psi^{t*}) | \left(\mathrm{D}^2\psi^t\right)^{-1} \nabla v(\nabla\psi^{t*})\rangle \gtrsim \mathsf{Var}_\rho(v(\nabla\psi^{t*})).}$$

New stability results

The compact case - Sketch of proof

1. Local strong convexity of the Kantorovich functional:

Brascamp-Lieb inequality:

$$\boxed{\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{K}(\psi^t) = \mathbb{E}_\rho\langle \nabla v(\nabla\psi^{t*}) | \left(\mathrm{D}^2\psi^t\right)^{-1} \nabla v(\nabla\psi^{t*})\rangle \gtrsim \mathsf{Var}_\rho(v(\nabla\psi^{t*})).}$$

2. "Global" strong convexity of the Kantorovich functional:

$\int_0^1 \dots \mathrm{d}t +$ concavity of $A \mapsto \det(A)^{1/d}$:

$$\boxed{\langle \nabla\mathcal{K}(\psi^1) - \nabla\mathcal{K}(\psi^0) | v\rangle \gtrsim \mathsf{Var}_{\mu^0+\mu^1}(v).}$$

New stability results

The compact case - Sketch of proof

2. "Global" strong convexity of the Kantorovich functional:

$$\begin{aligned} &\langle \nabla \mathcal{K}(\psi^1) - \nabla \mathcal{K}(\psi^0) | v \rangle \gtrsim \mathrm{Var}_{\mu^0+\mu^1}(v) \\ \iff &\langle \mu^0 - \mu^1 | \psi^1 - \psi^0 \rangle \gtrsim \mathrm{Var}_{\mu^0+\mu^1}(\psi^1 - \psi^0). \end{aligned}$$

New stability results

The compact case - Sketch of proof

2. "Global" strong convexity of the Kantorovich functional:

$$\begin{aligned} &\langle \nabla\mathcal{K}(\psi^1) - \nabla\mathcal{K}(\psi^0) | v\rangle \gtrsim \mathsf{Var}_{\mu^0+\mu^1}(v) \\ \iff &\langle \mu^0 - \mu^1 | \psi^1 - \psi^0\rangle \gtrsim \mathsf{Var}_{\mu^0+\mu^1}(\psi^1 - \psi^0). \end{aligned}$$

3. Stability of dual potentials:

Kantorovich-Rubinstein duality:

$$\begin{aligned} \mathsf{Var}_{\mu^0+\mu^1}(\psi^1 - \psi^0) &\lesssim \mathrm{Lip}(\psi^1 - \psi^0)\mathrm{W}_1(\mu^0, \mu^1) \\ &\lesssim \mathrm{W}_1(\mu^0, \mu^1). \end{aligned}$$

New stability results

The compact case - Sketch of proof

3. Stability of dual potentials:

$$\boxed{\mathsf{Var}_{\mu^0+\mu^1}(\psi^1 - \psi^0) \lesssim \mathrm{W}_1(\mu^0, \mu^1).}$$

4. Stability of primal potentials:

Fenchel-Young (in)equality:

$$\boxed{\mathsf{Var}_{\rho}(\psi^{1*} - \psi^{0*}) \leq \mathsf{Var}_{\mu^0+\mu^1}(\psi^1 - \psi^0) \lesssim \mathrm{W}_1(\mu^0, \mu^1).}$$

New stability results

The compact case - Sketch of proof

4. Stability of primal potentials:

$$\boxed{\mathsf{Var}_\rho(\psi^{1*}-\psi^{0*}) \leq \mathsf{Var}_{\mu^0+\mu^1}(\psi^1-\psi^0) \lesssim \mathrm{W}_1(\mu^0,\mu^1).}$$

5. Stability of Brenier maps:

A Gagliardo-Niremberg type inequatliy for convex functions u, v:

$$\|\nabla u - \nabla v\|^2_{\mathrm{L}^2(\mathcal{X})} \leq C_d \mathcal{H}^{d-1}(\partial\mathcal{X})^{2/3}(\|\nabla u\|_{\mathrm{L}^\infty(\mathcal{X})}+\|\nabla v\|_{\mathrm{L}^\infty(\mathcal{X})})^{4/3}\|u-v\|^{2/3}_{\mathrm{L}^2(\mathcal{X})}.$$

$$\begin{aligned}\|T_{\mu^1}-T_{\mu^0}\|_{\mathrm{L}^2(\rho)} = \|\nabla\psi^{1*}-\nabla\psi^{0*}\|_{\mathrm{L}^2(\rho)} &\lesssim \mathsf{Var}_\rho(\psi^{1*}-\psi^{0*})^{1/6} \\ &\lesssim \mathrm{W}_1(\mu^0,\mu^1)^{1/6}.\end{aligned}$$

New stability results

General case

Theorem(Mérigot, D.):
Let $p > d$ and $p \geq 4$. Assume that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ have bounded p-th moment, i.e. $\max(M_p(\mu), M_p(\nu)) \leq M_p < +\infty$. *Then*

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq C_{d,p,\mathcal{X},\rho,M_p} \mathrm{W}_1(\mu, \nu)^{\frac{p}{6p+16d}}.$$

New stability results

General case

Theorem(Mérigot, D.):
Let $p > d$ and $p \geq 4$. Assume that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ have bounded p-th moment, i.e. $\max(M_p(\mu), M_p(\nu)) \leq M_p < +\infty$. *Then*

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq C_{d,p,\mathcal{X},\rho,M_p} \mathrm{W}_1(\mu,\nu)^{\frac{p}{6p+16d}}.$$

- When $p \to +\infty$, we recover the compact case: $\frac{p}{6p+16d} \to \frac{1}{6}$.

New stability results

General case

Theorem(Mérigot, D.):
Let $p > d$ and $p \geq 4$. Assume that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ have bounded p-th moment, i.e. $\max(M_p(\mu), M_p(\nu)) \leq M_p < +\infty$. *Then*

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq C_{d,p,\mathcal{X},\rho,M_p} \mathrm{W}_1(\mu, \nu)^{\frac{p}{6p+16d}}.$$

- When $p \to +\infty$, we recover the compact case: $\frac{p}{6p+16d} \to \frac{1}{6}$.
- The moment assumption is satisfied by a large family of probability measures, including sub-Gaussian and sub-exponential measures.

New stability results

General case

Theorem(Mérigot, D.):
Let $p > d$ and $p \geq 4$. Assume that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ have bounded p-th moment, i.e. $\max(M_p(\mu), M_p(\nu)) \leq M_p < +\infty$. *Then*

$$\| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq C_{d,p,\mathcal{X},\rho,M_p} \mathrm{W}_1(\mu,\nu)^{\frac{p}{6p+16d}}.$$

- When $p \to +\infty$, we recover the compact case: $\frac{p}{6p+16d} \to \frac{1}{6}$.
- The moment assumption is satisfied by a large family of probability measures, including sub-Gaussian and sub-exponential measures.
- Remind the by Andoni et al. (2018), the space $(\mathcal{P}_2(\mathbb{R}^d), \mathrm{W}_2)$ does not admit a bi-Hölder embedding into any L^p space when $d \geq 3$.

Summary

- Optimal transport maps can be used to embed $(\mathcal{P}_2(\mathbb{R}^d), \mathrm{W}_2)$ into $\mathrm{L}^2(\rho; \mathbb{R}^d)$, with a controlled distorsion on subsets of $(\mathcal{P}_2(\mathbb{R}^d), \mathrm{W}_2)$:

For $p > d$ and $p \geq 4$, take $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ s.t. $\max(M_p(\mu), M_p(\nu)) \leq M_p < +\infty$, then

$$\mathrm{W}_2(\mu, \nu) \leq \| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq C_{d,p,\mathcal{X},\rho,M_p} \mathrm{W}_2(\mu, \nu)^{\frac{p}{6p+16d}}.$$

Summary

- Optimal transport maps can be used to embed $(\mathcal{P}_2(\mathbb{R}^d), \mathrm{W}_2)$ into $\mathrm{L}^2(\rho; \mathbb{R}^d)$, with a controlled distorsion on subsets of $(\mathcal{P}_2(\mathbb{R}^d), \mathrm{W}_2)$:

  For $p > d$ and $p \geq 4$, take $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ s.t. $\max(M_p(\mu), M_p(\nu)) \leq M_p < +\infty$, then

  $$\mathrm{W}_2(\mu, \nu) \leq \| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)} \leq C_{d,p,\mathcal{X},\rho,M_p} \mathrm{W}_2(\mu, \nu)^{\frac{p}{6p+16d}}.$$

- **Open questions:**
  - Optimal exponents?
  - What happens for other cost functions?
  - Generalizations to regularized and unbalanced optimal transport?
  - Is there an embedding into $\mathrm{L}^2(\rho, \mathbb{R}^d)$ of the "Wasserstein balls"

$$B_{\mathrm{W}_2}(\delta_0, R) = \{\mu \in \mathcal{P}_2(\mathbb{R}^d) \quad | \quad M_2(\mu) \leq R\}?$$

Thank you for your attention!

References I

Ambrosio, L., Gigli, N., and Savaré, G. (2008). *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media.

Andoni, A., Naor, A., and Neiman, O. (2018). Snowflake universality of Wasserstein spaces. *Annales Scientifiques de l'Ecole Normale Superieure*, 51(3):657–700.

Berman, R. J. (2018). Convergence rates for discretized monge-ampère equations and quantitative stability of optimal transport. arXiv preprint 1803.00785.

Berman, R. J. (2020). Convergence rates for discretized monge–ampère equations and quantitative stability of optimal transport. *Foundations of Computational Mathematics*.

Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417.

References II

Gigli, N. (2011). On hölder continuity-in-time of the optimal transport map towards measures along a curve. *Proceedings of the Edinburgh Mathematical Society*, 54(2):401–409.

Peyré, G. and Cuturi, M. (2019). Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607.

Villani, C. (2008). *Optimal transport: old and new*, volume 338. Springer Science & Business Media.