Linearization of the Wasserstein space & quantitative stability of optimal transport maps

Frédéric Chazal   Alex Delalande   Quentin Mérigot

DGDVC Conference, CIRM 2021

1. Wasserstein distance

Wasserstein distance

- Let $\mathrm{Prob}_p(\mathbb{R}^d) = \{\mu \in \mathrm{Prob}(\mathbb{R}^d) \mid \int \|x\|^p \,\mathrm{d}\,\mu < +\infty\}$.

p-**Wasserstein distance** between $\mu, \nu \in \mathrm{Prob}_p(\mathbb{R}^d)$:

$$\mathrm{W}_p(\mu, \nu) = \left(\min_{\gamma \in \Gamma(\mu,\nu)} \|x - y\|^p \,\mathrm{d}\,\gamma(x, y)\right)^{1/p}.$$

where $\Gamma(\mu, \nu) =$ couplings between μ and $\nu \subseteq \mathrm{Prob}(\mathbb{R}^d \times \mathbb{R}^d)$.

Wasserstein distance

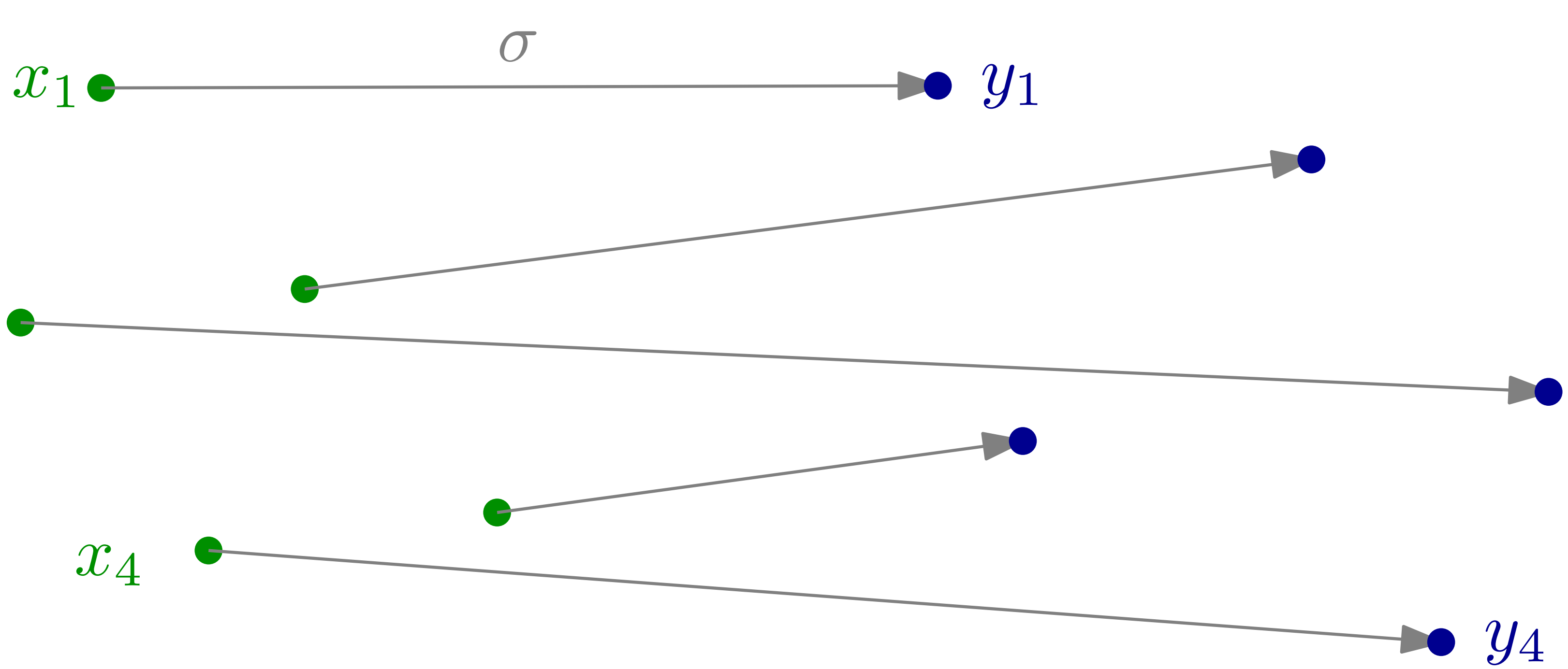
- Let $\mathrm{Prob}_p(\mathbb{R}^d) = \{\mu \in \mathrm{Prob}(\mathbb{R}^d) \mid \int \|x\|^p \,\mathrm{d}\,\mu < +\infty\}$.

p-**Wasserstein distance** between $\mu, \nu \in \mathrm{Prob}_p(\mathbb{R}^d)$:

$$\mathrm{W}_p(\mu, \nu) = \left(\min_{\gamma \in \Gamma(\mu,\nu)} \|x - y\|^p \,\mathrm{d}\,\gamma(x, y)\right)^{1/p}.$$

where $\Gamma(\mu, \nu)$ = couplings between μ and $\nu \subseteq \mathrm{Prob}(\mathbb{R}^d \times \mathbb{R}^d)$.

ex. $\mu = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}, \quad \nu = \frac{1}{N}\sum_{j=1}^{n} \delta_{y_j} \Longrightarrow \mathrm{W}_p(\mu, \nu) = \left(\min_{\sigma \in \mathfrak{S}_N} \|x_i - y_{\sigma(i)}\|^p\right)^{1/p}$



Wasserstein distance

- Let $\mathrm{Prob}_p(\mathbb{R}^d) = \{\mu \in \mathrm{Prob}(\mathbb{R}^d) \mid \int \|x\|^p \,\mathrm{d}\,\mu < +\infty\}$.

p-**Wasserstein distance** between $\mu, \nu \in \mathrm{Prob}_p(\mathbb{R}^d)$:

$$\mathrm{W}_p(\mu, \nu) = \left(\min_{\gamma \in \Gamma(\mu,\nu)} \|x - y\|^p \,\mathrm{d}\,\gamma(x, y)\right)^{1/p}.$$

where $\Gamma(\mu, \nu)$ = couplings between μ and $\nu \subseteq \mathrm{Prob}(\mathbb{R}^d \times \mathbb{R}^d)$.

ex. $\mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}, \quad \nu = \frac{1}{N} \sum_{j=1}^{n} \delta_{y_j} \Longrightarrow \mathrm{W}_p(\mu, \nu) = \left(\min_{\sigma \in \mathfrak{S}_N} \|x_i - y_{\sigma(i)}\|^p\right)^{1/p}$

- On $\mathrm{Prob}(X)$, with $X \subseteq \mathbb{R}^d$ compact, W_p metrizes the **weak$*$ convergence**

The Wasserstein distances makes sense for data (point clouds, gray images, meshes) which describe some distribution of mass, e.g. histograms.

Wasserstein distance

- Let $\mathrm{Prob}_p(\mathbb{R}^d) = \{\mu \in \mathrm{Prob}(\mathbb{R}^d) \mid \int \|x\|^p \,\mathrm{d}\,\mu < +\infty\}$.

> p-**Wasserstein distance** between $\mu, \nu \in \mathrm{Prob}_p(\mathbb{R}^d)$:
> $$\mathrm{W}_p(\mu,\nu) = \left(\min_{\gamma \in \Gamma(\mu,\nu)} \|x-y\|^p \,\mathrm{d}\,\gamma(x,y)\right)^{1/p}.$$
> where $\Gamma(\mu,\nu)$ = couplings between μ and $\nu \subseteq \mathrm{Prob}(\mathbb{R}^d \times \mathbb{R}^d)$.

ex. $\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}, \quad \nu = \frac{1}{N}\sum_{j=1}^n \delta_{y_j} \Longrightarrow \mathrm{W}_p(\mu,\nu) = \left(\min_{\sigma \in \mathfrak{S}_N} \|x_i - y_{\sigma(i)}\|^p\right)^{1/p}$

- On $\mathrm{Prob}(X)$, with $X \subseteq \mathbb{R}^d$ compact, W_p metrizes the **weak$*$ convergence**

> The Wasserstein distances makes sense for data (point clouds, gray images, meshes) which describe some distribution of mass, e.g. histograms.

- One can extend some usual statistical notions to the Wasserstein space:

ex. **Wasserstein barycenter** between $\mu_1, \ldots, \mu_N$ with coefficients $\alpha_1, \ldots, \alpha_N \geq 0$:

[Agueh, Carlier '10]

Wasserstein distance

- Let $\mathrm{Prob}_p(\mathbb{R}^d) = \{\mu \in \mathrm{Prob}(\mathbb{R}^d) \mid \int \|x\|^p \,\mathrm{d}\,\mu < +\infty\}$.

p-**Wasserstein distance** between $\mu, \nu \in \mathrm{Prob}_p(\mathbb{R}^d)$:

$$\mathrm{W}_p(\mu,\nu) = \left(\min_{\gamma\in\Gamma(\mu,\nu)} \|x-y\|^p \,\mathrm{d}\,\gamma(x,y)\right)^{1/p}.$$

where $\Gamma(\mu,\nu)$ = couplings between μ and $\nu \subseteq \mathrm{Prob}(\mathbb{R}^d \times \mathbb{R}^d)$.

ex. $\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}, \quad \nu = \frac{1}{N}\sum_{j=1}^n \delta_{y_j} \implies \mathrm{W}_p(\mu,\nu) = \left(\min_{\sigma\in\mathfrak{S}_N} \|x_i - y_{\sigma(i)}\|^p\right)^{1/p}$

- On $\mathrm{Prob}(X)$, with $X \subseteq \mathbb{R}^d$ compact, W_p metrizes the **weak$*$ convergence**

The Wasserstein distances makes sense for data (point clouds, gray images, meshes) which describe some distribution of mass, e.g. histograms.

- One can extend some usual statistical notions to the Wasserstein space:

ex. **Wasserstein barycenter** between $\mu_1, \dots, \mu_N$ with coefficients $\alpha_1, \dots, \alpha_N \geq 0$:

[Agueh, Carlier '10]

$$\mathrm{bary}((\mu_i),(\alpha_i)) := \arg\min_{\mu\in\mathrm{Prob}(X)} \sum_{1\leq i\leq N} \alpha_i \,\mathrm{W}_2^2(\mu,\mu_i).$$

Wasserstein distance

- Let $\mathrm{Prob}_p(\mathbb{R}^d) = \{\mu \in \mathrm{Prob}(\mathbb{R}^d) \mid \int \|x\|^p \,\mathrm{d}\,\mu < +\infty\}$.

p-**Wasserstein distance** between $\mu, \nu \in \mathrm{Prob}_p(\mathbb{R}^d)$:

$$\mathrm{W}_p(\mu,\nu) = \left(\min_{\gamma \in \Gamma(\mu,\nu)} \|x-y\|^p \,\mathrm{d}\,\gamma(x,y)\right)^{1/p}.$$

where $\Gamma(\mu,\nu)$ = couplings between μ and $\nu \subseteq \mathrm{Prob}(\mathbb{R}^d \times \mathbb{R}^d)$.

ex. $\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}, \quad \nu = \frac{1}{N}\sum_{j=1}^n \delta_{y_j} \implies \mathrm{W}_p(\mu,\nu) = \left(\min_{\sigma \in \mathfrak{S}_N} \|x_i - y_{\sigma(i)}\|^p\right)^{1/p}$

- On $\mathrm{Prob}(X)$, with $X \subseteq \mathbb{R}^d$ compact, W_p metrizes the **weak$*$ convergence**

The Wasserstein distances makes sense for data (point clouds, gray images, meshes) which describe some distribution of mass, e.g. histograms.

- One can extend some usual statistical notions to the Wasserstein space:

ex. **Wasserstein barycenter** between $\mu_1, \ldots, \mu_N$ with coefficients $\alpha_1, \ldots, \alpha_N \geq 0$:

[Agueh, Carlier '10]

$$\mathrm{bary}((\mu_i),(\alpha_i)) := \arg\min_{\mu \in \mathrm{Prob}(X)} \sum_{1 \leq i \leq N} \alpha_i \,\mathrm{W}_2^2(\mu,\mu_i).$$

but also: k-**means algorithm**, **principal component analysis**, etc.

1D Wasserstein space and Quantiles

- Given $\mu \in \mathrm{Prob}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\rho = \mu$, with $\rho =$ Lebesgue measure on $[0,1]$.

NB: $T_{\mu\#}\lambda = \mu \iff \forall B \subseteq \mathbb{R},\ \lambda(T_\mu^{-1}(B)) = \mu(B)$
$\iff \forall x \in \mathbb{R},\ T_\mu^{-1}(x) = \lambda([0, T_\mu^{-1}(x)]) = \mu((-\infty, x])$

1D Wasserstein space and Quantiles

- Given $\mu \in \mathrm{Prob}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\rho = \mu$, with $\rho =$ Lebesgue measure on $[0,1]$.

  NB: $T_{\mu\#}\lambda = \mu \iff \forall B \subseteq \mathbb{R},\ \lambda(T_\mu^{-1}(B)) = \mu(B)$
  $\iff \forall x \in \mathbb{R},\ T_\mu^{-1}(x) = \lambda([0, T_\mu^{-1}(x)]) = \mu((-\infty, x])$

- T_μ is the inverse cdf, also called *quantile function*.

  $\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ with $x_1 \leq \ldots, \leq x_N \Longrightarrow T_\mu = x_i$ on $\left[\frac{i-1}{N}, \frac{i}{N}\right]$.

1D Wasserstein space and Quantiles

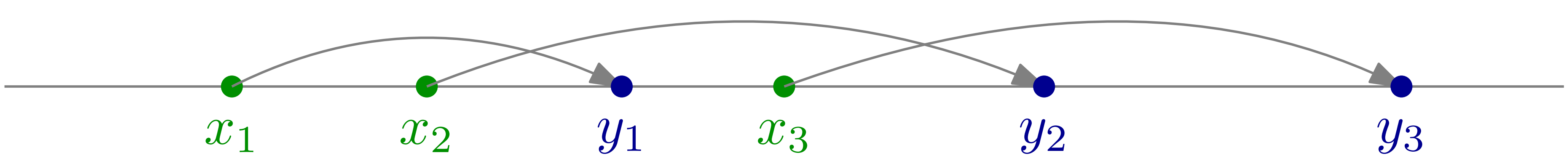
- Given $\mu \in \mathrm{Prob}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\rho = \mu$, with $\rho =$ Lebesgue measure on $[0,1]$.

NB: $T_{\mu\#}\lambda = \mu \iff \forall B \subseteq \mathbb{R},\ \lambda(T_\mu^{-1}(B)) = \mu(B)$
$\iff \forall x \in \mathbb{R},\ T_\mu^{-1}(x) = \lambda([0, T_\mu^{-1}(x)]) = \mu((-\infty, x])$

- T_μ is the inverse cdf, also called *quantile function*.

$\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ with $x_1 \leq \ldots, \leq x_N \implies T_\mu = x_i$ on $\left[\frac{i-1}{N}, \frac{i}{N}\right]$.
$\nu = \frac{1}{N}\sum_{i=1}^N \delta_{y_i}$ with $y_1 \leq \ldots, \leq y_N \implies T_\nu = y_i$ on $\left[\frac{i-1}{N}, \frac{i}{N}\right]$.

$$\mathrm{W}_p(\mu,\nu)^p = \frac{1}{N}\sum_{1\leq i\leq N} \|x_i - y_j\|^p = \|T_\mu - T_\nu\|^p_{\mathrm{L}^p([0,1])}$$



1D Wasserstein space and Quantiles

- Given $\mu \in \mathrm{Prob}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\rho = \mu$, with $\rho =$ Lebesgue measure on $[0,1]$.

  NB: $T_{\mu\#}\lambda = \mu \iff \forall B \subseteq \mathbb{R},\ \lambda(T_\mu^{-1}(B)) = \mu(B)$
  $\iff \forall x \in \mathbb{R},\ T_\mu^{-1}(x) = \lambda([0, T_\mu^{-1}(x)]) = \mu((-\infty, x])$

- T_μ is the inverse cdf, also called *quantile function*.

  $\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ with $x_1 \leq \ldots, \leq x_N \Longrightarrow T_\mu = x_i$ on $\left[\frac{i-1}{N}, \frac{i}{N}\right]$.
  $\nu = \frac{1}{N}\sum_{i=1}^N \delta_{y_i}$ with $y_1 \leq \ldots, \leq y_N \Longrightarrow T_\nu = y_i$ on $\left[\frac{i-1}{N}, \frac{i}{N}\right]$.

$$\mathrm{W}_p(\mu,\nu)^p = \frac{1}{N}\sum_{1\leq i\leq N} \|x_i - y_j\|^p = \|T_\mu - T_\nu\|^p_{\mathrm{L}^p([0,1])}$$

- This property remains true for any pair of probability measures $\mu, \nu \in \mathrm{Prob}(\mathbb{R})$:

$$\mathrm{W}_p(\mu,\nu) = \left(\int_{[0,1]} \|T_\mu(t) - T_\nu(t)\|^p \,\mathrm{d}\,t\right) = \|T_\mu - T_\nu\|_{\mathrm{L}^p([0,1])}$$

  In particular, $\mu \mapsto T_\mu$ embeds isometrically in $\mathrm{Prob}_p(\mathbb{R})$ into $\mathrm{L}^p([0,1])$!

1D Wasserstein space and Quantiles

- Given $\mu \in \mathrm{Prob}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\rho = \mu$, with $\rho =$ Lebesgue measure on $[0,1]$.

NB: $T_{\mu\#}\lambda = \mu \iff \forall B \subseteq \mathbb{R},\ \lambda(T_\mu^{-1}(B)) = \mu(B)$
$\iff \forall x \in \mathbb{R},\ T_\mu^{-1}(x) = \lambda([0, T_\mu^{-1}(x)]) = \mu((-\infty, x])$

- T_μ is the inverse cdf, also called *quantile function*.

$\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ with $x_1 \leq \ldots, \leq x_N \Longrightarrow T_\mu = x_i$ on $\left[\frac{i-1}{N}, \frac{i}{N}\right]$.
$\nu = \frac{1}{N}\sum_{i=1}^N \delta_{y_i}$ with $y_1 \leq \ldots, \leq y_N \Longrightarrow T_\nu = y_i$ on $\left[\frac{i-1}{N}, \frac{i}{N}\right]$.

$$\mathrm{W}_p(\mu,\nu)^p = \tfrac{1}{N}\textstyle\sum_{1\leq i\leq N} \|x_i - y_j\|^p = \|T_\mu - T_\nu\|^p_{\mathrm{L}^p([0,1])}$$

- This property remains true for any pair of probability measures $\mu, \nu \in \mathrm{Prob}(\mathbb{R})$:

$$\mathrm{W}_p(\mu,\nu) = \left(\int_{[0,1]} \|T_\mu(t) - T_\nu(t)\|^p \,\mathrm{d}\,t\right) = \|T_\mu - T_\nu\|_{\mathrm{L}^p([0,1])}$$

In particular, $\mu \mapsto T_\mu$ embeds isometrically in $\mathrm{Prob}_p(\mathbb{R})$ into $\mathrm{L}^p([0,1])$!

- No such isometric embedding in higher dimension: $(\mathrm{Prob}_p, \mathrm{W}_p)$ is *curved*.

Example: 1D Wasserstein barycenters

Proposition: The W_2 barycenter of $\mu_1, \ldots, \mu_N \in \mathrm{Prob}_2(\mathbb{R})$ and $\alpha_1, \ldots, \alpha_N \geq 0$ is

$$\mathrm{bary}((\mu_i), (\alpha_i)) = \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i}\right)_{\#} \rho$$

Example: 1D Wasserstein barycenters

Proposition: The W_2 barycenter of $\mu_1, \ldots, \mu_N \in \mathrm{Prob}_2(\mathbb{R})$ and $\alpha_1, \ldots, \alpha_N \geq 0$ is

$$\mathrm{bary}((\mu_i),(\alpha_i)) = \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i}\right)_{\#} \rho$$

Proof: $\mu \in \arg\min_\mu \sum_i \alpha_i \, \mathrm{W}_2^2(\mu_i, \mu) \iff \mu \in \arg\min_\mu \sum_i \alpha_i \|T_{\mu_i} - T_\mu\|^2$

Example: 1D Wasserstein barycenters

Proposition: The W_2 barycenter of $\mu_1, \ldots, \mu_N \in \mathrm{Prob}_2(\mathbb{R})$ and $\alpha_1, \ldots, \alpha_N \geq 0$ is

$$\mathrm{bary}((\mu_i), (\alpha_i)) = \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho$$

Proof: $\mu \in \arg\min_\mu \sum_i \alpha_i \, \mathrm{W}_2^2(\mu_i, \mu) \Longleftrightarrow \mu \in \arg\min_\mu \sum_i \alpha_i \|T_{\mu_i} - T_\mu\|^2$

$$\Longleftrightarrow T_\mu = \frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i}$$

Example: 1D Wasserstein barycenters

Proposition: The W_2 barycenter of $\mu_1, \ldots, \mu_N \in \mathrm{Prob}_2(\mathbb{R})$ and $\alpha_1, \ldots, \alpha_N \geq 0$ is

$$\mathrm{bary}((\mu_i), (\alpha_i)) = \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i}\right)_{\#} \rho$$

Proof: $\mu \in \arg\min_\mu \sum_i \alpha_i \mathrm{W}_2^2(\mu_i, \mu) \iff \mu \in \arg\min_\mu \sum_i \alpha_i \|T_{\mu_i} - T_\mu\|^2$

$$\iff T_\mu = \frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i}$$

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

Example: 1D Wasserstein barycenters

Proposition: The W_2 barycenter of $\mu_1, \ldots, \mu_N \in \mathrm{Prob}_2(\mathbb{R})$ and $\alpha_1, \ldots, \alpha_N \geq 0$ is

$$\mathrm{bary}((\mu_i), (\alpha_i)) = \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i}\right)_{\#} \rho$$

Proof: $\mu \in \arg\min_\mu \sum_i \alpha_i \mathrm{W}_2^2(\mu_i, \mu) \iff \mu \in \arg\min_\mu \sum_i \alpha_i \|T_{\mu_i} - T_\mu\|^2$

$$\iff T_\mu = \frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i}$$

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

Example: If $\mu_i = \frac{1}{N} \sum_j \delta_{x_i^j}$, with $x_i^1 \leq \ldots \leq x_i^N$ and $\alpha_i \geq 0$, the W_2 barycenter is

$$\mathrm{bary}((\mu_i), (\alpha_i)) = \frac{1}{N} \sum_j \delta_{\bar{x}^j} \text{ with } \bar{x}^j = \frac{1}{\sum_i \alpha_i} \sum_i \alpha_i x_i^j.$$

Example: 1D Wasserstein barycenters

Proposition: The W_2 barycenter of $\mu_1, \ldots, \mu_N \in \mathrm{Prob}_2(\mathbb{R})$ and $\alpha_1, \ldots, \alpha_N \geq 0$ is

$$\mathrm{bary}((\mu_i), (\alpha_i)) = \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho$$

Proof:

$$\begin{aligned} \mu \in \arg\min_\mu \textstyle\sum_i \alpha_i \, \mathrm{W}_2^2(\mu_i, \mu) &\Longleftrightarrow \mu \in \arg\min_\mu \textstyle\sum_i \alpha_i \|T_{\mu_i} - T_\mu\|^2 \\ &\Longleftrightarrow T_\mu = \frac{1}{\sum_i \alpha_i} \textstyle\sum_i \alpha_i T_{\mu_i} \\ &\Longleftrightarrow \mu = \left(\frac{1}{\sum_i \alpha_i} \textstyle\sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho \end{aligned}$$

Example: If $\mu_i = \frac{1}{N} \sum_j \delta_{x_i^j}$, with $x_i^1 \leq \ldots \leq x_i^N$ and $\alpha_i \geq 0$, the W_2 barycenter is

$$\mathrm{bary}((\mu_i), (\alpha_i)) = \frac{1}{N} \sum_j \delta_{\bar{x}^j} \text{ with } \bar{x}^j = \frac{1}{\sum_i \alpha_i} \sum_i \alpha_i x_i^j.$$

In 1D, computing a W_2 barycenter between empirical measures

$\Longleftrightarrow$ sorting the positions of the Dirac masses + averaging !

2. Linearized Wasserstein distance

[Wang, Slepcev, Basu, Ozolek, Rohde '13]

Motivation 1: Monge-Kantorovich Quantiles

- Given $\mu \in \mathrm{Prob}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\rho = \mu$, with $\rho =$ Lebesgue measure on $[0,1]$.
- The map $\mu \mapsto T_\mu$ embeds isometrically in $\mathrm{Prob}_p(\mathbb{R})$ into $\mathrm{L}^p([0,1])$.

Motivation 1: Monge-Kantorovich Quantiles

- Given $\mu \in \mathrm{Prob}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\rho = \mu$, with $\rho =$ Lebesgue measure on $[0,1]$.
- The map $\mu \mapsto T_\mu$ embeds isometrically in $\mathrm{Prob}_p(\mathbb{R})$ into $\mathrm{L}^p([0,1])$.

How to extend this notion to a multivariate setting ?

Motivation 1: Monge-Kantorovich Quantiles

- Given $\mu \in \mathrm{Prob}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\rho = \mu$, with $\rho =$ Lebesgue measure on $[0,1]$.
- The map $\mu \mapsto T_\mu$ embeds isometrically in $\mathrm{Prob}_p(\mathbb{R})$ into $\mathrm{L}^p([0,1])$.

How to extend this notion to a multivariate setting ?

Theorem (Brenier, McCann) Given $\rho \in \mathrm{Prob}^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mu \in \mathrm{Prob}(\mathbb{R}^d)$,
$\exists!$ ρ-a.e. $T_\mu : \mathbb{R}^d \to \mathbb{R}^d$ such that $T_{\mu\#}\rho = \mu$ and $T_\mu = \nabla\phi$ with ϕ convex.

- Monge-Kantorovich quantile $:= T_\mu$. Need of a reference probability density ρ.

[Cherzonukov, Galichon, Hallin, Henry, '15]

Motivation 1: Monge-Kantorovich Quantiles

- Given $\mu \in \mathrm{Prob}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\rho = \mu$, with $\rho =$ Lebesgue measure on $[0,1]$.
- The map $\mu \mapsto T_\mu$ embeds isometrically in $\mathrm{Prob}_p(\mathbb{R})$ into $\mathrm{L}^p([0,1])$.

How to extend this notion to a multivariate setting ?

Theorem (Brenier, McCann) Given $\rho \in \mathrm{Prob}^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mu \in \mathrm{Prob}(\mathbb{R}^d)$,
$\exists!$ ρ-a.e. $T_\mu : \mathbb{R}^d \to \mathbb{R}^d$ such that $T_{\mu\#}\rho = \mu$ and $T_\mu = \nabla\phi$ with ϕ convex.

- Monge-Kantorovich quantile $:= T_\mu$. Need of a reference probability density ρ.
  [Cherzonukov, Galichon, Hallin, Henry, '15]
- T_μ is unique ρ-a.e. but the convex function ϕ_μ is not necessarily unique.

Motivation 1: Monge-Kantorovich Quantiles

- Given $\mu \in \mathrm{Prob}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\rho = \mu$, with $\rho =$ Lebesgue measure on $[0,1]$.
- The map $\mu \mapsto T_\mu$ embeds isometrically in $\mathrm{Prob}_p(\mathbb{R})$ into $\mathrm{L}^p([0,1])$.

How to extend this notion to a multivariate setting ?

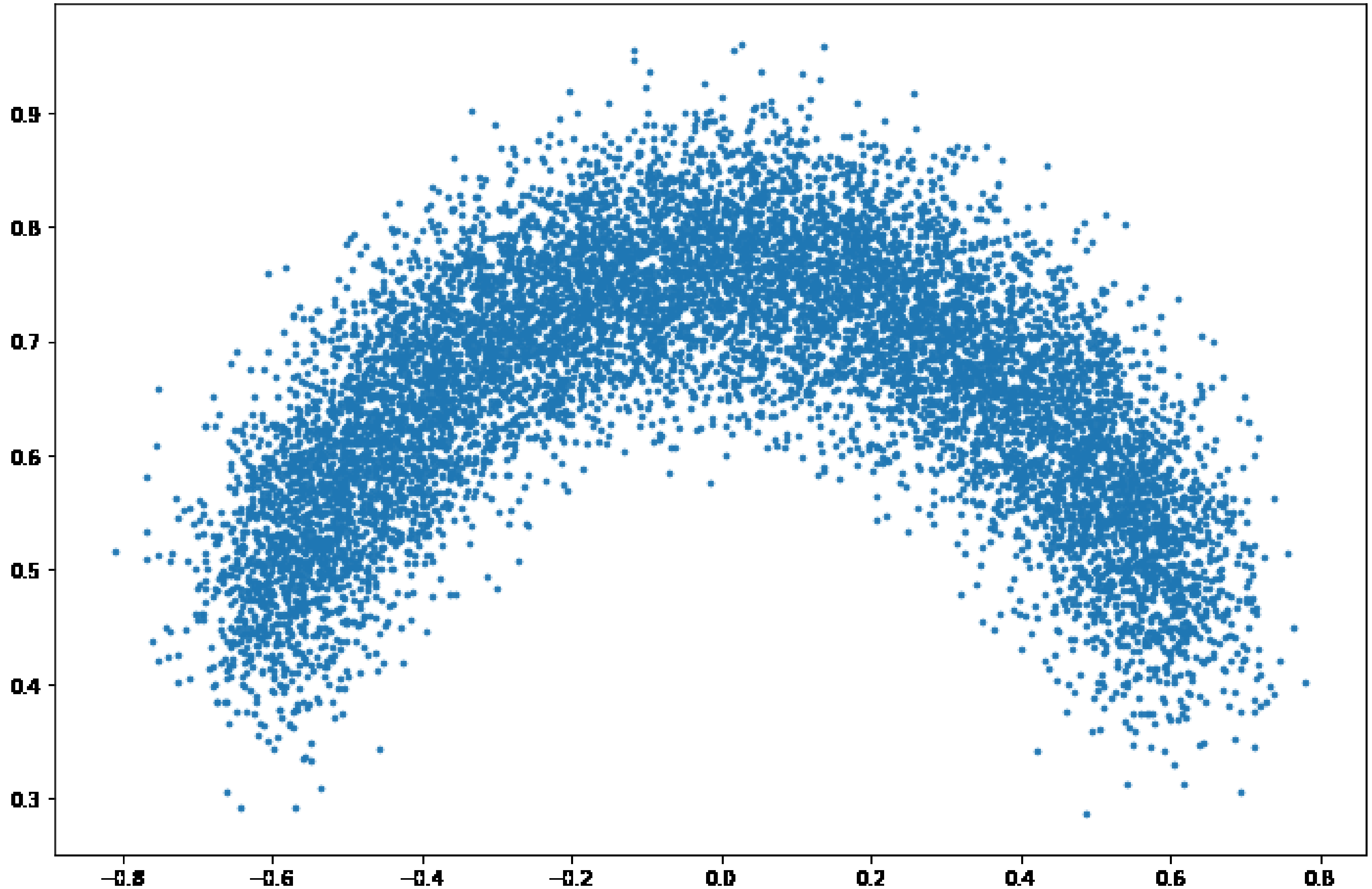
Theorem (Brenier, McCann) Given $\rho \in \mathrm{Prob}^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mu \in \mathrm{Prob}(\mathbb{R}^d)$,
$\exists!$ ρ-a.e. $T_\mu : \mathbb{R}^d \to \mathbb{R}^d$ such that $T_{\mu\#}\rho = \mu$ and $T_\mu = \nabla\phi$ with ϕ convex.

- Monge-Kantorovich quantile $:= T_\mu$. Need of a reference probability density ρ.
  [Cherzonukov, Galichon, Hallin, Henry, '15]
- T_μ is unique ρ-a.e. but the convex function ϕ_μ is not necessarily unique.
- $T_\mu : \mathrm{spt}(\rho) \to \mathbb{R}^d$ is *monotone*: $\langle T_\mu(x) - T_\mu(y) | x - y \rangle \geq 0$.

Numerical Example: Monge-Kantorovich Depth

Source: $\rho =$ uniform probability density on $B(0,1) \subseteq \mathbb{R}^2$

Target: $\mu = \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{y_i}$ with $N = 10^4$ points



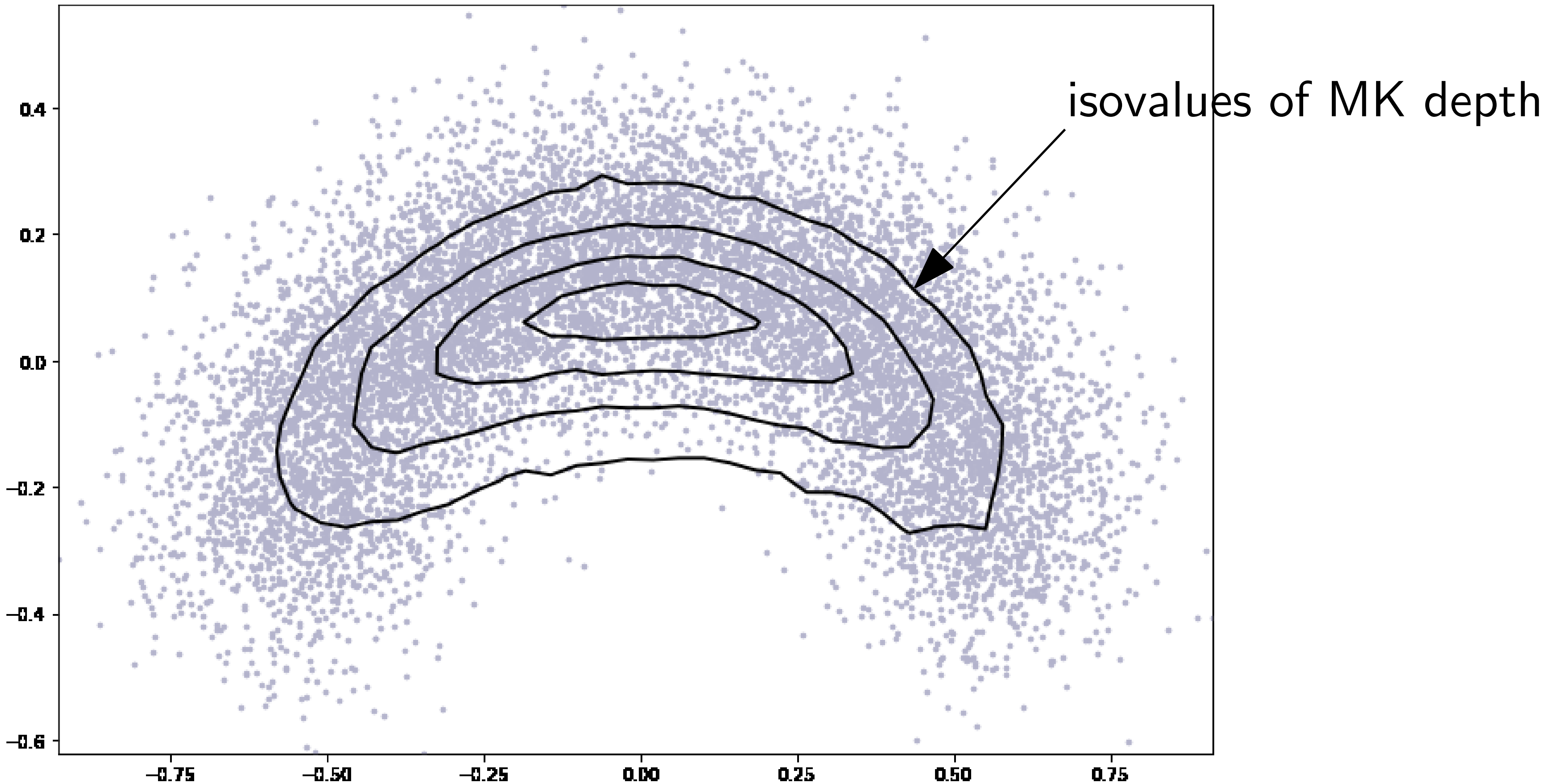
"Monge-Kantorovich depth of y_i" $\simeq \|T_\mu^{-1}(y_i)\|$.

[Cherzonukov, Galichon, Hallin, Henry]

Numerical Example: Monge-Kantorovich Depth

Source: $\rho =$ uniform probability density on $B(0,1) \subseteq \mathbb{R}^2$

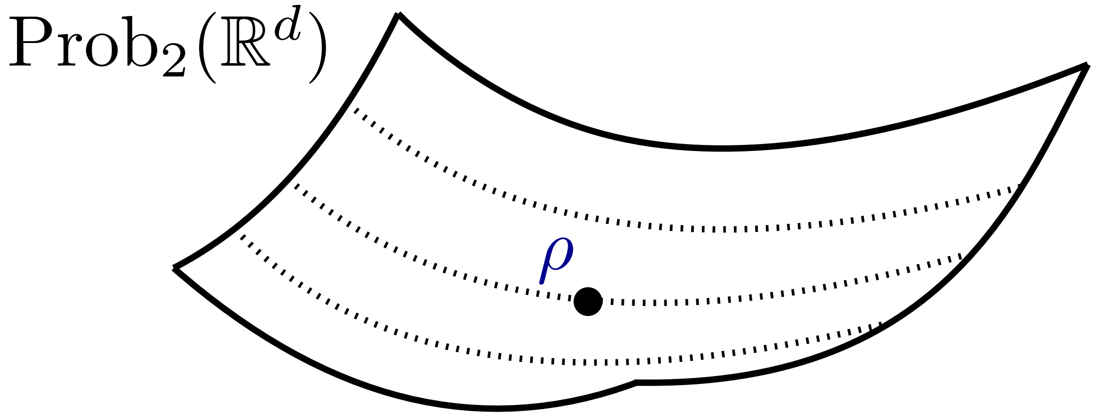
Target: $\mu = \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{y_i}$ with $N = 10^4$ points



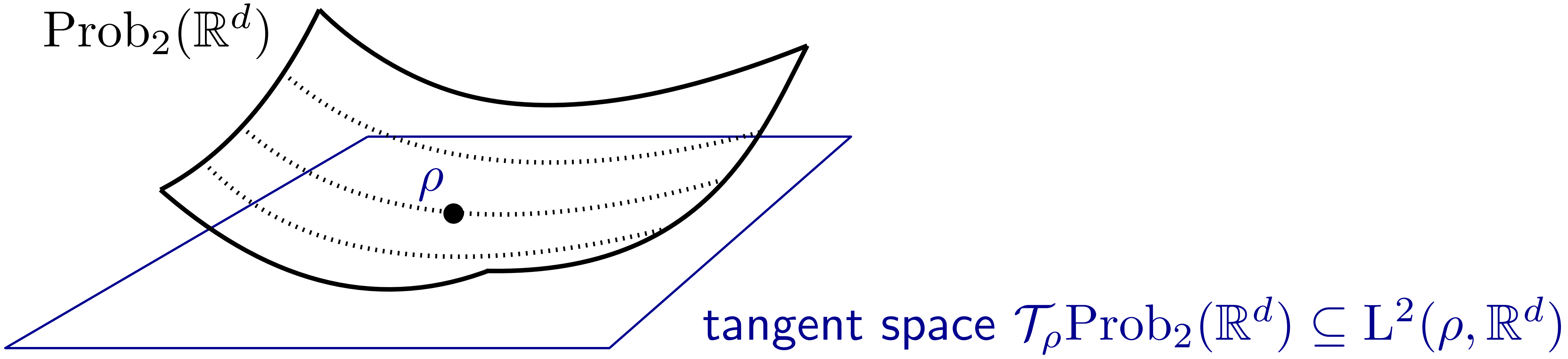
"Monge-Kantorovich depth of y_i" $\simeq \|T_\mu^{-1}(y_i)\|$.

[Cherzonukov, Galichon, Hallin, Henry]

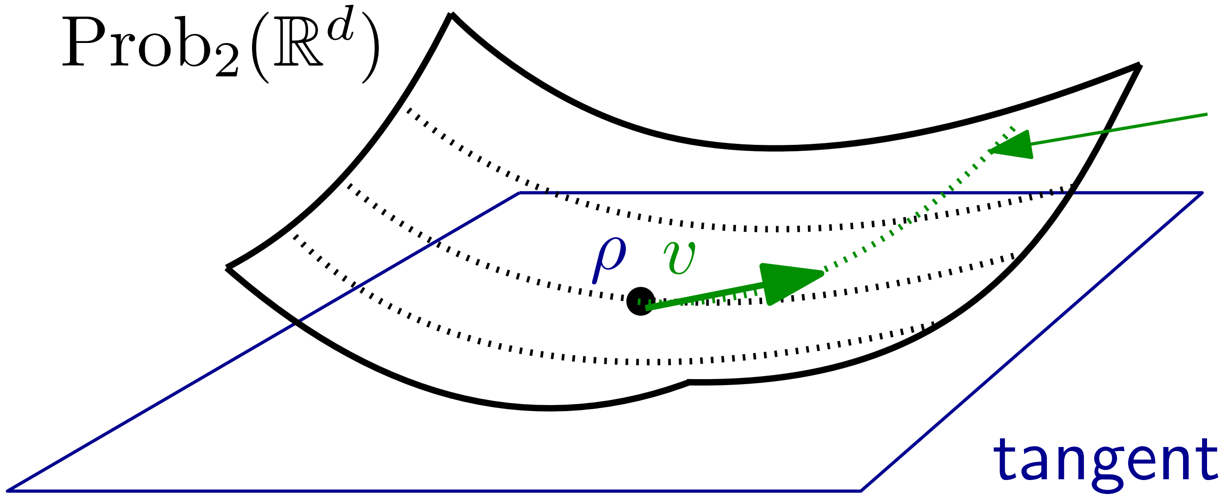
Motivation 2: $\mu \mapsto T_\mu - \mathrm{id}$ as a logarithm



Motivation 2: $\mu \mapsto T_\mu - \mathrm{id}$ as a logarithm



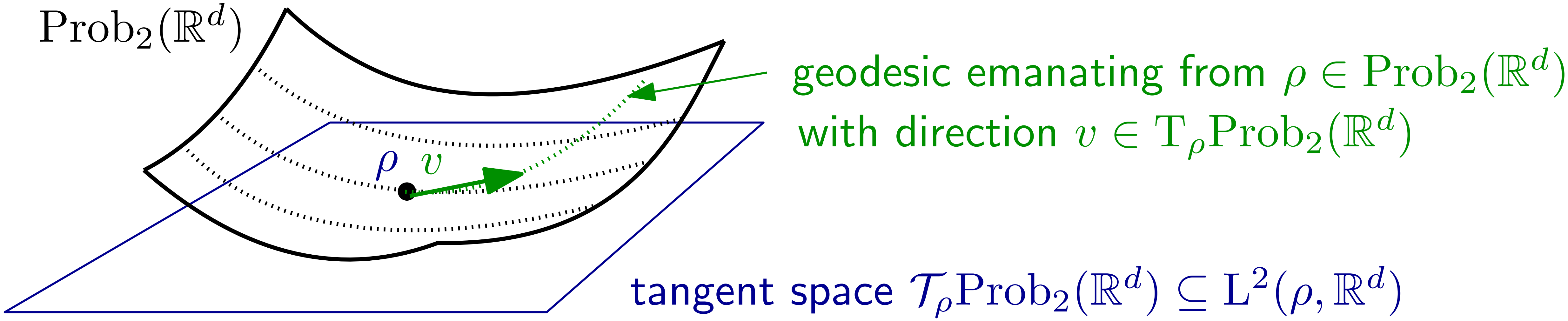
Motivation 2: $\mu \mapsto T_\mu - \mathrm{id}$ as a logarithm



geodesic emanating from $\rho \in \mathrm{Prob}_2(\mathbb{R}^d)$
with direction $v \in \mathrm{T}_\rho \mathrm{Prob}_2(\mathbb{R}^d)$

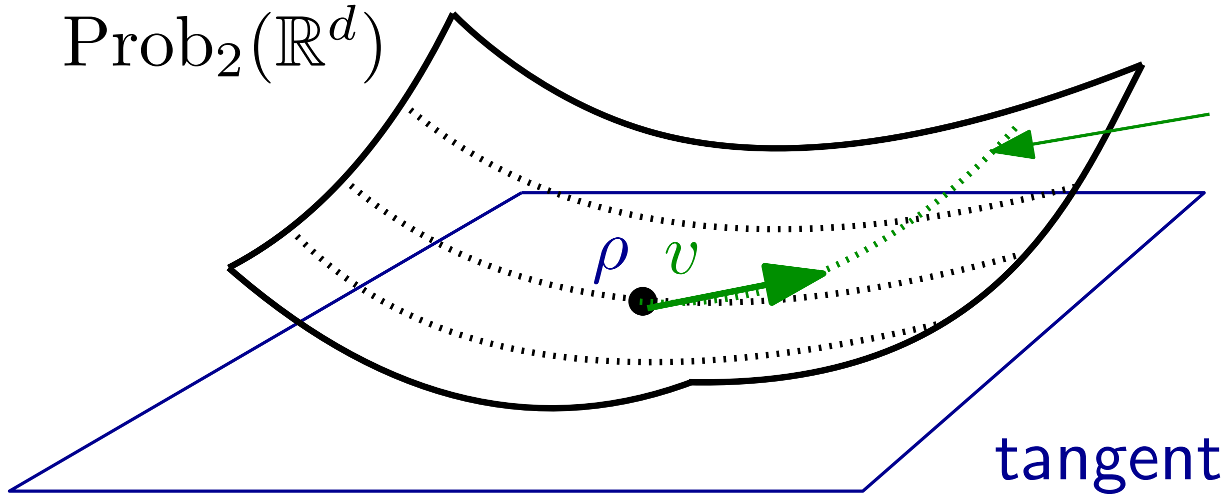
tangent space $\mathcal{T}_\rho \mathrm{Prob}_2(\mathbb{R}^d) \subseteq \mathrm{L}^2(\rho, \mathbb{R}^d)$

Motivation 2: $\mu \mapsto T_\mu - \mathrm{id}$ as a logarithm



	Riemannian geometry	Optimal transport
point	$x \in M$	$\mu \in \mathrm{Prob}_2(\mathbb{R}^d)$
geodesic distance	$\mathrm{d}_g(x,y)$	$\mathrm{W}_2(\mu,\nu)$
tangent space	$\mathrm{T}_\rho M$	$\mathrm{T}_\rho \mathrm{Prob}_2(\mathbb{R}^d) \subseteq \mathrm{L}^2(\rho, X)$
exponential map	$\exp_\rho : \mathrm{T}_\rho M \to M$	$v \in \mathrm{T}_\rho \mathrm{Prob}_2(\mathbb{R}^d) \mapsto (\mathrm{id}+v)_\#$
inverse exponential map	$\exp_\rho^{-1}(x) \in \mathrm{T}_\rho M$	$T_\mu - \mathrm{id} \in \mathrm{T}_\rho \mathrm{Prob}_2(X)$
distance in tangent space	$\\| \exp_\rho^{-1}(x) - \exp_\rho^{-1}(y) \\|_{g(x_0)}$	$\\|T_\mu - T_\nu\\|_{\mathrm{L}^2(\rho)}$

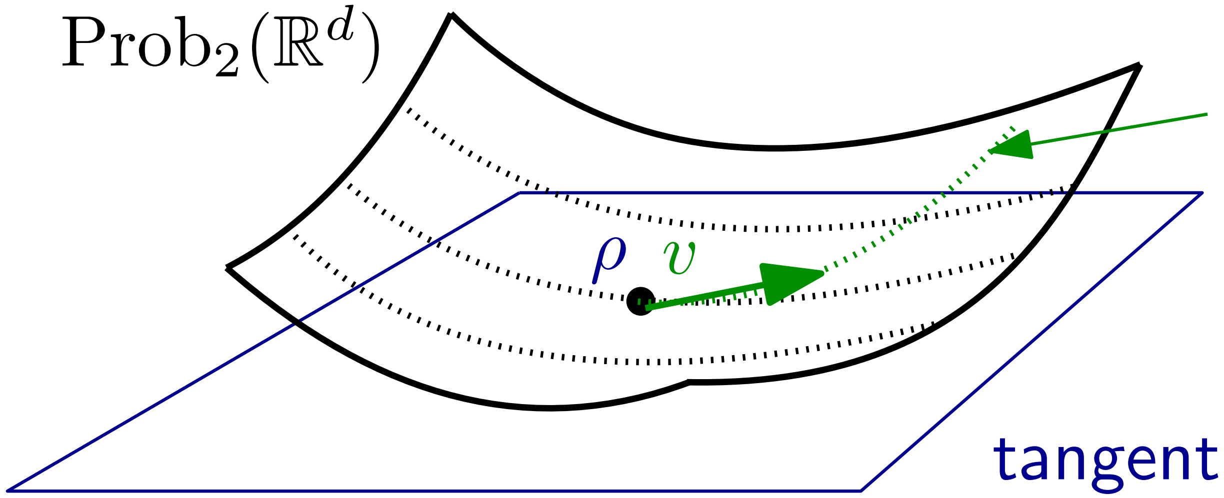
Motivation 2: $\mu \mapsto T_\mu - \mathrm{id}$ as a logarithm



	Riemannian geometry	Optimal transport
point	$x \in M$	$\mu \in \mathrm{Prob}_2(\mathbb{R}^d)$
geodesic distance	$\mathrm{d}_g(x,y)$	$\mathrm{W}_2(\mu,\nu)$
tangent space	$\mathrm{T}_\rho M$	$\mathrm{T}_\rho \mathrm{Prob}_2(\mathbb{R}^d) \subseteq \mathrm{L}^2(\rho, X)$
exponential map	$\exp_\rho : \mathrm{T}_\rho M \to M$	$v \in \mathrm{T}_\rho \mathrm{Prob}_2(\mathbb{R}^d) \mapsto (\mathrm{id}+v)_{\#}$
inverse exponential map	$\exp_\rho^{-1}(x) \in \mathrm{T}_\rho M$	$T_\mu - \mathrm{id} \in \mathrm{T}_\rho \mathrm{Prob}_2(X)$
distance in tangent space	$\Vert \exp_\rho^{-1}(x) - \exp_\rho^{-1}(y) \Vert_{g(x_0)}$	$\Vert T_\mu - T_\nu \Vert_{\mathrm{L}^2(\rho)}$

- The map $\mu \in \mathrm{Prob}_2(\mathbb{R}^d) \to T_\mu \in \mathrm{L}^2(X)$ is an injective map, with image the space of (square-integrable) gradients of convex functions on X.
  - Linearized OT framework $\longrightarrow$ [Wang, Slepcev, Basu, Ozolek, Rohde '13]

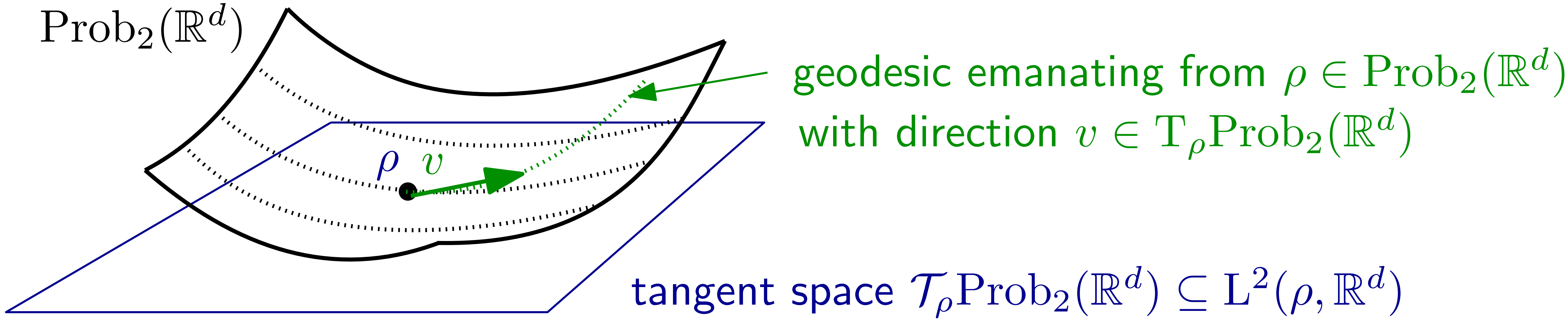
Motivation 2: $\mu \mapsto T_\mu - \mathrm{id}$ as a logarithm



	Riemannian geometry	Optimal transport
point	$x \in M$	$\mu \in \mathrm{Prob}_2(\mathbb{R}^d)$
geodesic distance	$\mathrm{d}_g(x, y)$	$\mathrm{W}_2(\mu, \nu)$
tangent space	$\mathrm{T}_\rho M$	$\mathrm{T}_\rho\mathrm{Prob}_2(\mathbb{R}^d) \subseteq \mathrm{L}^2(\rho, X)$
exponential map	$\exp_\rho : \mathrm{T}_\rho M \to M$	$v \in \mathrm{T}_\rho\mathrm{Prob}_2(\mathbb{R}^d) \mapsto (\mathrm{id} + v)_\#$
inverse exponential map	$\exp_\rho^{-1}(x) \in \mathrm{T}_\rho M$	$T_\mu - \mathrm{id} \in \mathrm{T}_\rho\mathrm{Prob}_2(X)$
distance in tangent space	$\\| \exp_\rho^{-1}(x) - \exp_\rho^{-1}(y) \\|_{g(x_0)}$	$\\|T_\mu - T_\nu\\|_{\mathrm{L}^2(\rho)}$

- The map $\mu \in \mathrm{Prob}_2(\mathbb{R}^d) \to T_\mu \in \mathrm{L}^2(X)$ is an injective map, with image the space of (square-integrable) gradients of convex functions on X.
  - Linearized OT framework $\longrightarrow$ [Wang, Slepcev, Basu, Ozolek, Rohde '13]
  - $\mathrm{W}_{2,\rho}(\mu, \nu) := \\|T_\mu - T_\nu\\|_{\mathrm{L}^2(\rho)} \longrightarrow$ [Ambrosio, Gigli, Savaré '04]

Motivation 2: $\mu \mapsto T_\mu - \mathrm{id}$ as a logarithm



	Riemannian geometry	Optimal transport
point	$x \in M$	$\mu \in \mathrm{Prob}_2(\mathbb{R}^d)$
geodesic distance	$\mathrm{d}_g(x, y)$	$\mathrm{W}_2(\mu, \nu)$
tangent space	$\mathrm{T}_\rho M$	$\mathrm{T}_\rho\mathrm{Prob}_2(\mathbb{R}^d) \subseteq \mathrm{L}^2(\rho, X)$
exponential map	$\exp_\rho : \mathrm{T}_\rho M \to M$	$v \in \mathrm{T}_\rho\mathrm{Prob}_2(\mathbb{R}^d) \mapsto (\mathrm{id} + v)_\#$
inverse exponential map	$\exp_\rho^{-1}(x) \in \mathrm{T}_\rho M$	$T_\mu - \mathrm{id} \in \mathrm{T}_\rho\mathrm{Prob}_2(X)$
distance in tangent space	$\\| \exp_\rho^{-1}(x) - \exp_\rho^{-1}(y)\\|_{g(x_0)}$	$\\|T_\mu - T_\nu\\|_{\mathrm{L}^2(\rho)}$

- The map $\mu \in \mathrm{Prob}_2(\mathbb{R}^d) \to T_\mu \in \mathrm{L}^2(X)$ is an injective map, with image the space of (square-integrable) gradients of convex functions on X.
  - Linearized OT framework $\longrightarrow$ [Wang, Slepcev, Basu, Ozolek, Rohde '13]
  - $\mathrm{W}_{2,\rho}(\mu, \nu) := \|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \longrightarrow$ [Ambrosio, Gigli, Savaré '04]

Geometric embedding of $\mathrm{Prob}_2(\mathbb{R}^d)$ into the Hilbert space $\mathrm{L}^2(\rho, \mathbb{R}^d)$.

Example: barycenter computation

- **Barycenter in Wasserstein space:** $\mu_1, \ldots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \ldots, \alpha_k \geq 0$:

$$\mu := \arg\min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i \, \mathrm{W}_2^2(\mu, \mu_i).$$

Example: barycenter computation

- **Barycenter in Wasserstein space:** $\mu_1, \ldots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \ldots, \alpha_k \geq 0$:

$$\mu := \arg\min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i \, \mathrm{W}_2^2(\mu, \mu_i).$$

$\longrightarrow$ Need to solve an optimisation problem every time the coefficients α_i are changed.

Example: barycenter computation

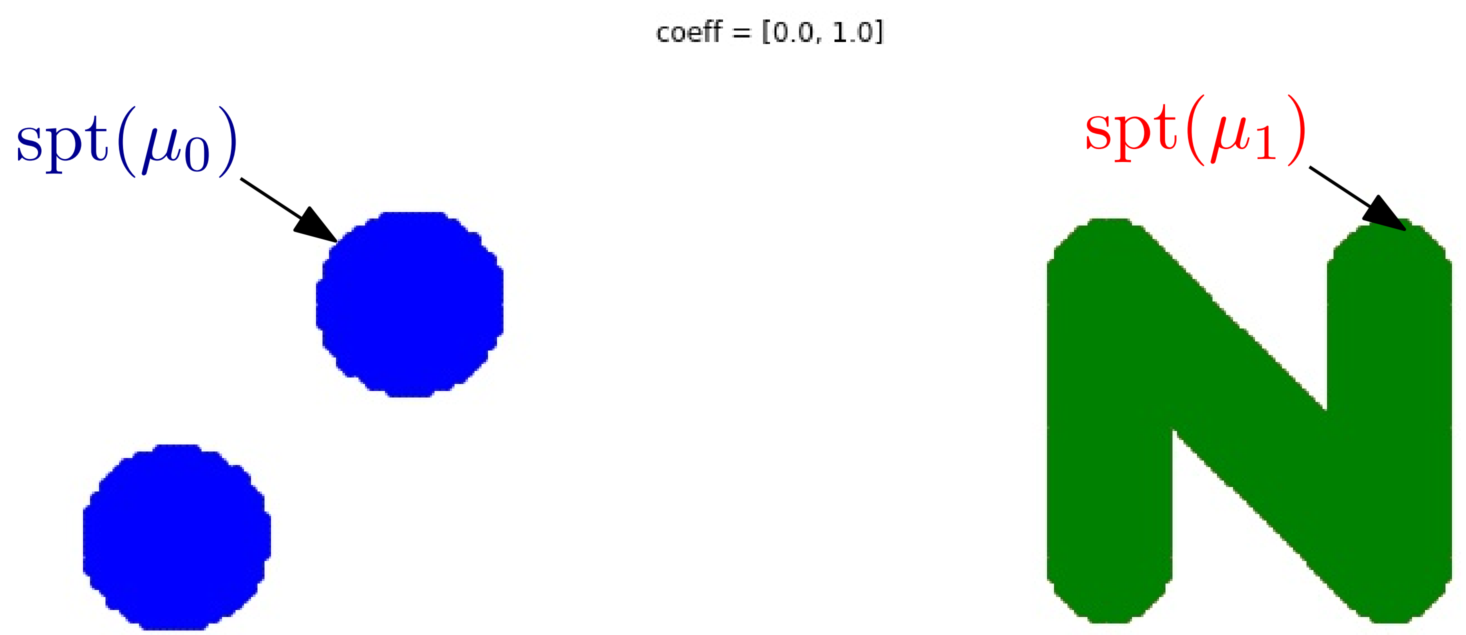
- **Barycenter in Wasserstein space:** $\mu_1, \ldots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \ldots, \alpha_k \geq 0$:

$$\mu := \arg\min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i \, \mathrm{W}_2^2(\mu, \mu_i).$$

$\longrightarrow$ Need to solve an optimisation problem every time the coefficients α_i are changed.

- **"Linearized" Wasserstein barycenters:** $\mu := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$

$\longrightarrow$ Simple expression once the transport maps $T_{\mu_i} : \rho \to \mu_i$ have been computed.



Example: barycenter computation

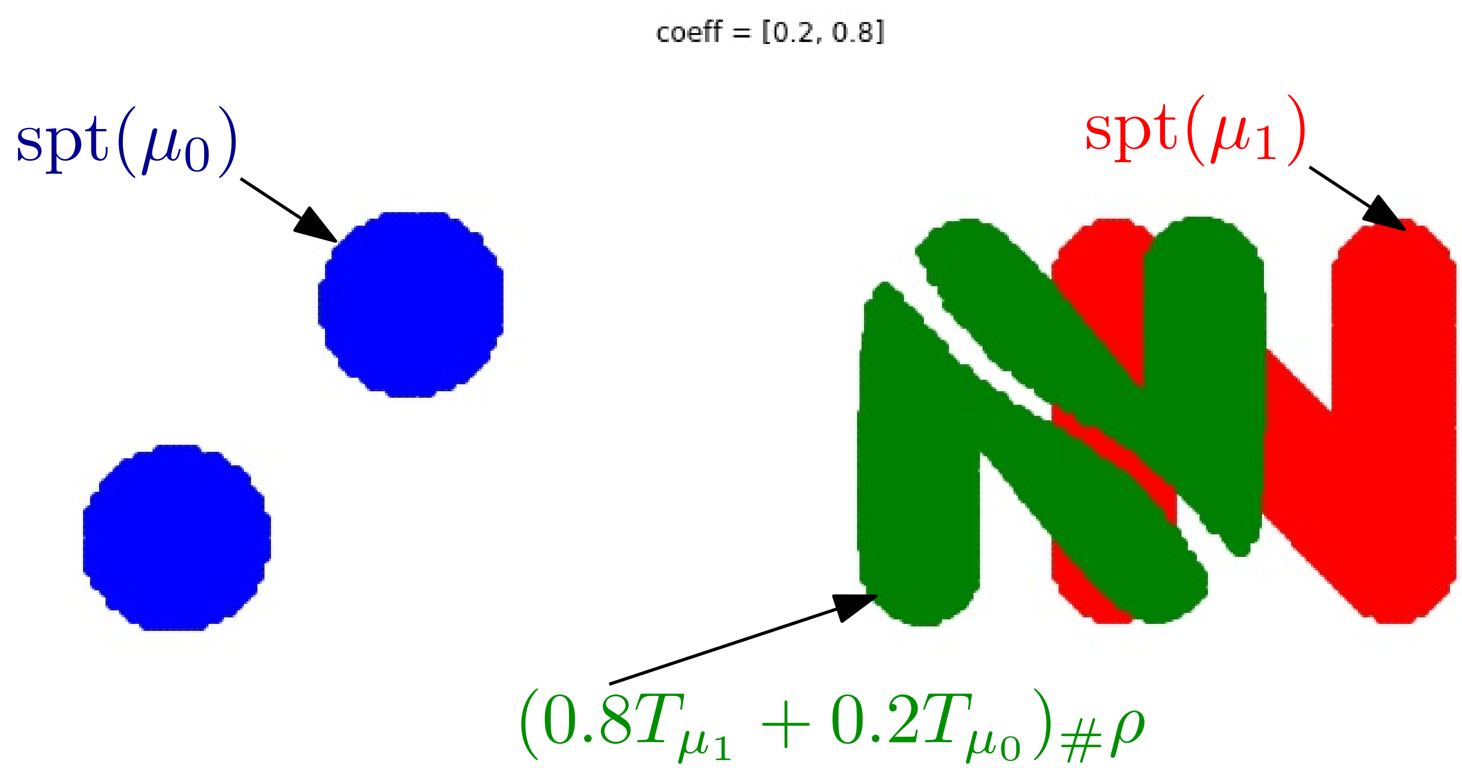
- **Barycenter in Wasserstein space:** $\mu_1, \ldots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \ldots, \alpha_k \geq 0$:

$$\mu := \arg\min_{1 \leq i \leq k} \textstyle\sum_{1 \leq i \leq k} \alpha_i \, \mathrm{W}_2^2(\mu, \mu_i).$$

$\longrightarrow$ Need to solve an optimisation problem every time the coefficients α_i are changed.

- **"Linearized" Wasserstein barycenters:** $\mu := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho$.

$\longrightarrow$ Simple expression once the transport maps $T_{\mu_i} : \rho \to \mu_i$ have been computed.



Example: barycenter computation

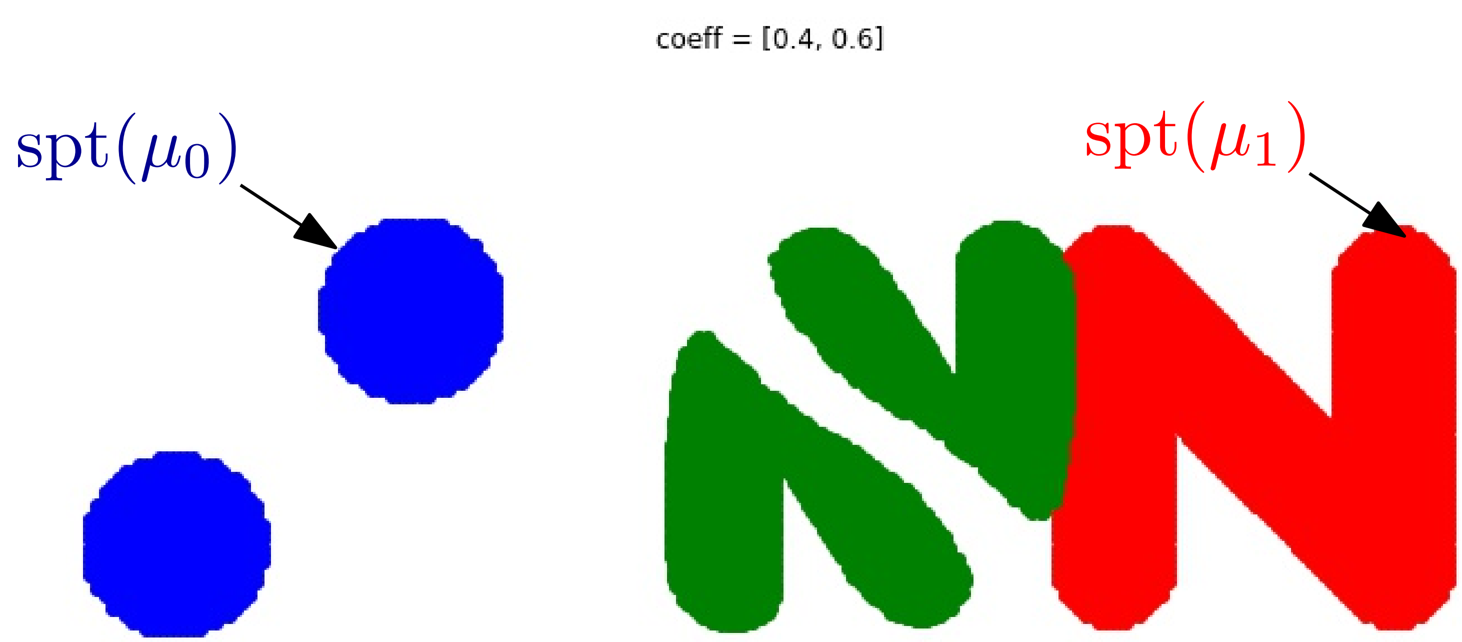
- **Barycenter in Wasserstein space:** $\mu_1, \ldots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \ldots, \alpha_k \geq 0$:

$$\mu := \arg\min_{1\leq i\leq k} \sum_{1\leq i\leq k} \alpha_i \, \mathrm{W}_2^2(\mu, \mu_i).$$

$\longrightarrow$ Need to solve an optimisation problem every time the coefficients α_i are changed.

- **"Linearized" Wasserstein barycenters:** $\mu := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i}\right)_{\#} \rho$.

$\longrightarrow$ Simple expression once the transport maps $T_{\mu_i} : \rho \to \mu_i$ have been computed.



Example: barycenter computation

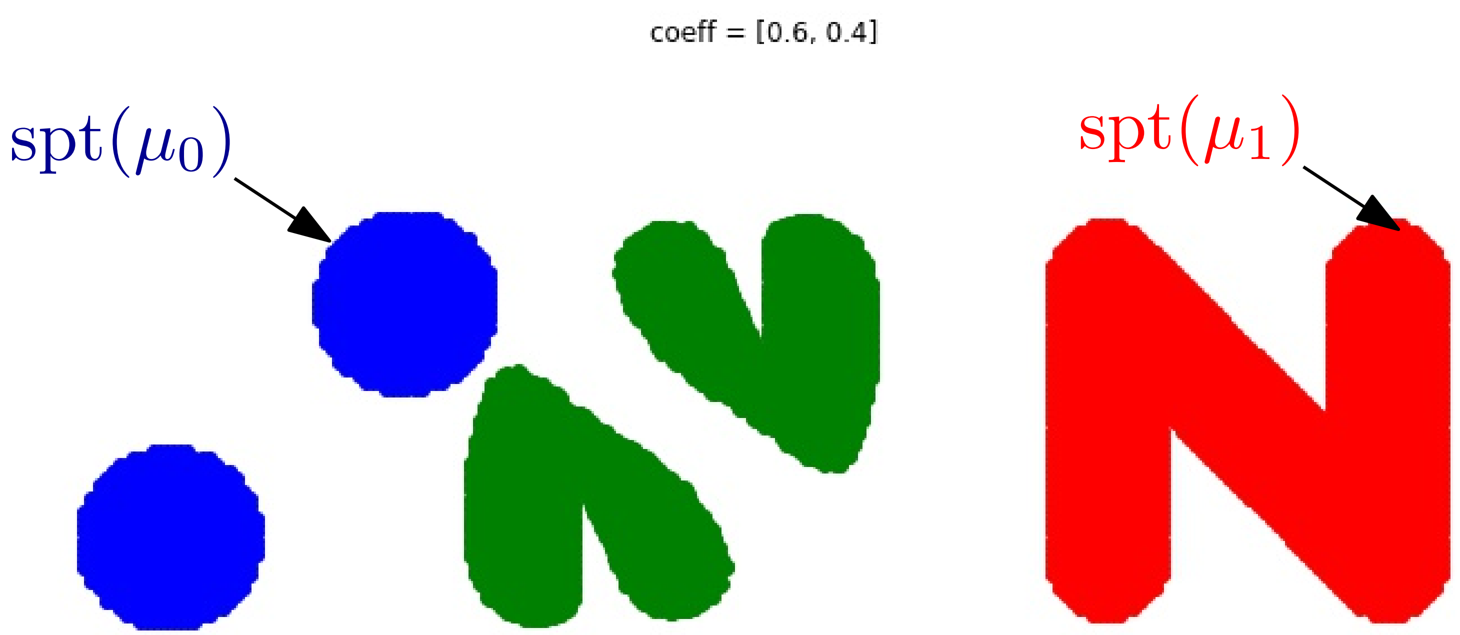
- **Barycenter in Wasserstein space:** $\mu_1, \ldots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \ldots, \alpha_k \geq 0$:

$$\mu := \arg\min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i \, \mathrm{W}_2^2(\mu, \mu_i).$$

$\longrightarrow$ Need to solve an optimisation problem every time the coefficients α_i are changed.

- **"Linearized" Wasserstein barycenters:** $\mu := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$

$\longrightarrow$ Simple expression once the transport maps $T_{\mu_i} : \rho \to \mu_i$ have been computed.



Example: barycenter computation

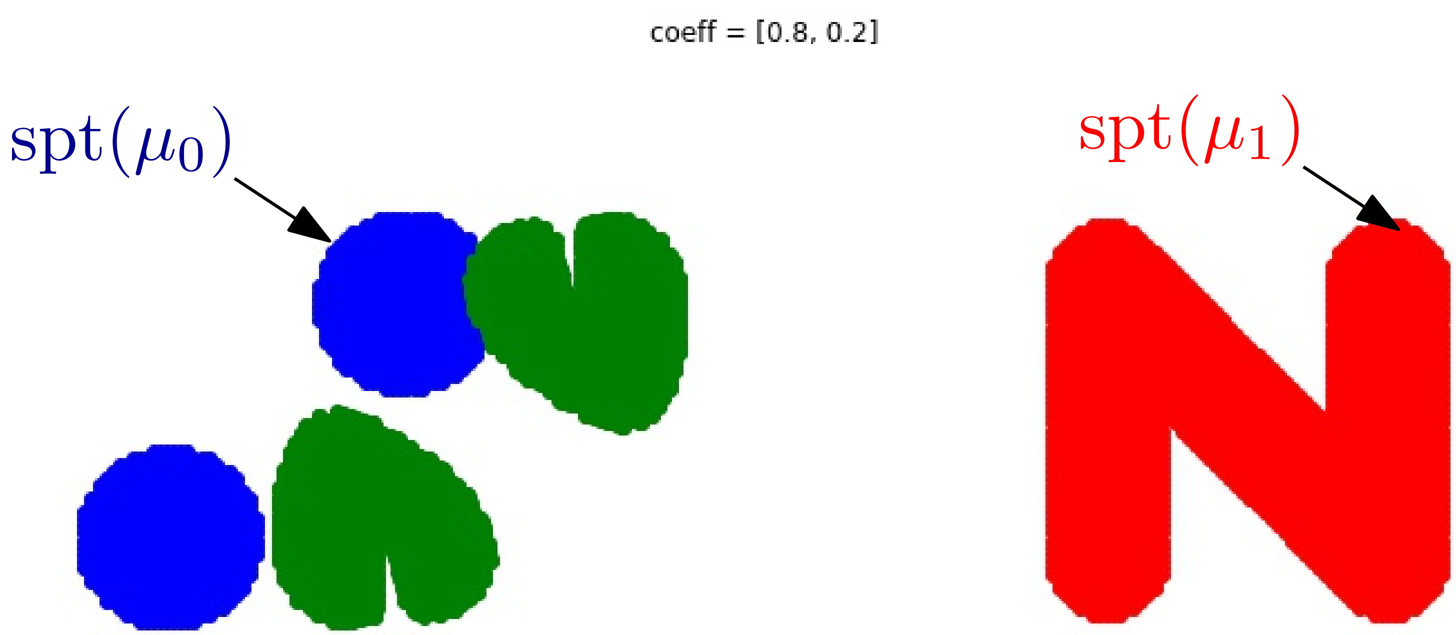
- **Barycenter in Wasserstein space:** $\mu_1, \ldots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \ldots, \alpha_k \geq 0$:

$$\mu := \arg\min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i \, \mathrm{W}_2^2(\mu, \mu_i).$$

$\longrightarrow$ Need to solve an optimisation problem every time the coefficients α_i are changed.

- **"Linearized" Wasserstein barycenters:** $\mu := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right)_{\#} \rho.$

$\longrightarrow$ Simple expression once the transport maps $T_{\mu_i} : \rho \to \mu_i$ have been computed.



Example: barycenter computation

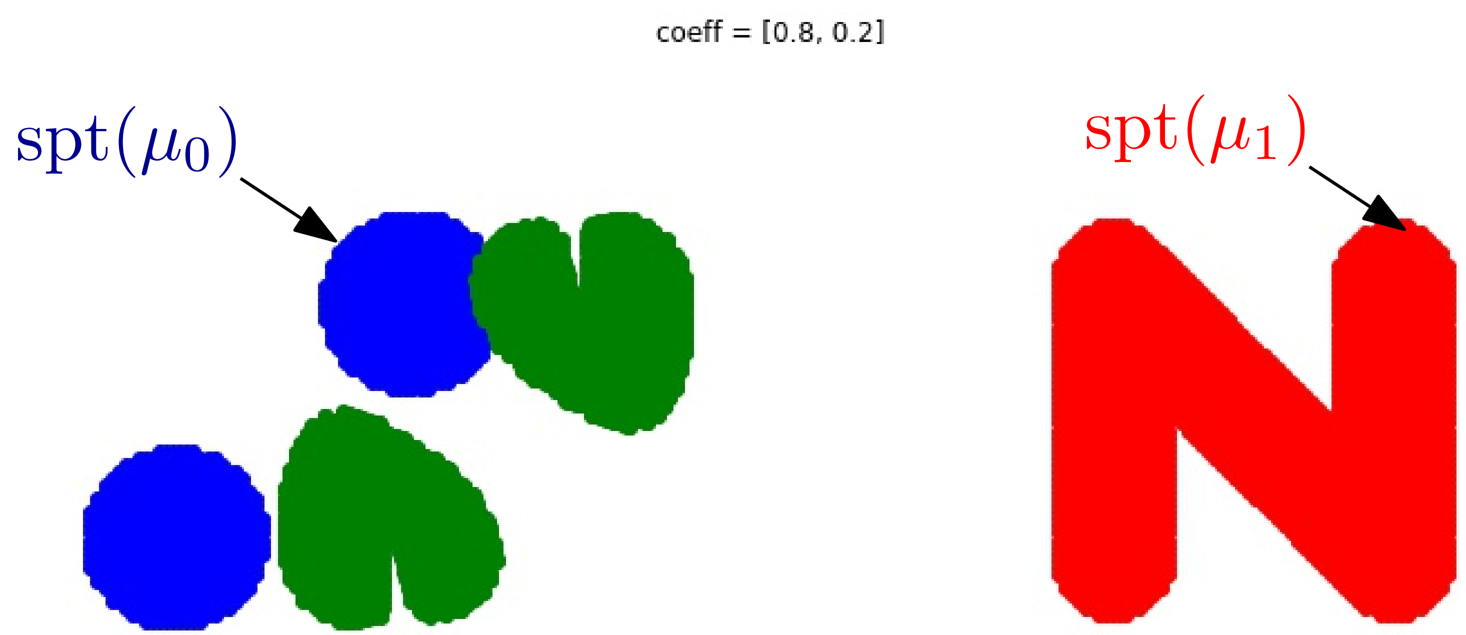
- **Barycenter in Wasserstein space:** $\mu_1, \dots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \dots, \alpha_k \geq 0$:

$$\mu := \arg\min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i \, \mathrm{W}_2^2(\mu, \mu_i).$$

$\longrightarrow$ Need to solve an optimisation problem every time the coefficients α_i are changed.

- **"Linearized" Wasserstein barycenters:** $\mu := \left(\frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i}\right)_{\#} \rho$.

$\longrightarrow$ Simple expression once the transport maps $T_{\mu_i} : \rho \to \mu_i$ have been computed.



What amount of the Wasserstein geometry is preserved by the embedding $\mu \mapsto T_\mu$?

Example: k-Means for MNIST digits

MNIST has $M = 60\,000$ images grayscale images (64×64 pixels) representing digits.

Example: k-Means for MNIST digits

MNIST has $M = 60\,000$ images grayscale images (64×64 pixels) representing digits.

Each image $\alpha^\ell \in \mathcal{M}_{64}(\mathbb{R})$ is transformed into a probability measure on $[0,1]^2$ via

$$\mu^\ell = \frac{1}{\sum_{i,j} \alpha^\ell_{ij}} \sum_{i,j} \alpha^\ell_{i,j} \delta_{(x_i, x_j)}, \quad \text{with } x_i = \frac{i}{63}$$

Example: k-Means for MNIST digits

MNIST has $M = 60\,000$ images grayscale images (64×64 pixels) representing digits.

Each image $\alpha^\ell \in \mathcal{M}_{64}(\mathbb{R})$ is transformed into a probability measure on $[0,1]^2$ via

$$\mu^\ell = \frac{1}{\sum_{i,j} \alpha^\ell_{ij}} \sum_{i,j} \alpha^\ell_{i,j} \delta_{(x_i, x_j)}, \quad \text{with } x_i = \frac{i}{63}$$

$$T^\ell = T_{\mu^\ell} \in \mathrm{L}^2([0,1], \mathbb{R}^2) \quad \text{[OT map from } \rho = \mathrm{Leb}_{[0,1]^2} \text{ to } \mu^\ell\text{]}$$

Example: k-Means for MNIST digits

MNIST has $M = 60\,000$ images grayscale images (64×64 pixels) representing digits.

Each image $\alpha^\ell \in \mathcal{M}_{64}(\mathbb{R})$ is transformed into a probability measure on $[0,1]^2$ via

$$\mu^\ell = \frac{1}{\sum_{i,j} \alpha^\ell_{ij}} \sum_{i,j} \alpha^\ell_{i,j} \delta_{(x_i, x_j)}, \quad \text{with } x_i = \frac{i}{63}$$

$$T^\ell = T_{\mu^\ell} \in \mathrm{L}^2([0,1], \mathbb{R}^2) \quad \text{[OT map from } \rho = \mathrm{Leb}_{[0,1]^2} \text{ to } \mu^\ell]$$

We run the K-Means method on the transport plans, with $K = 20$.

Each cluster $X^k \subseteq \{0, \dots, M\}$ yields an *average transport plan* $S^k = \frac{1}{|X^k|} \sum_{\ell \in X} T^\ell$,

Example: k-Means for MNIST digits

MNIST has $M = 60\,000$ images grayscale images (64×64 pixels) representing digits.

Each image $\alpha^\ell \in \mathcal{M}_{64}(\mathbb{R})$ is transformed into a probability measure on $[0,1]^2$ via

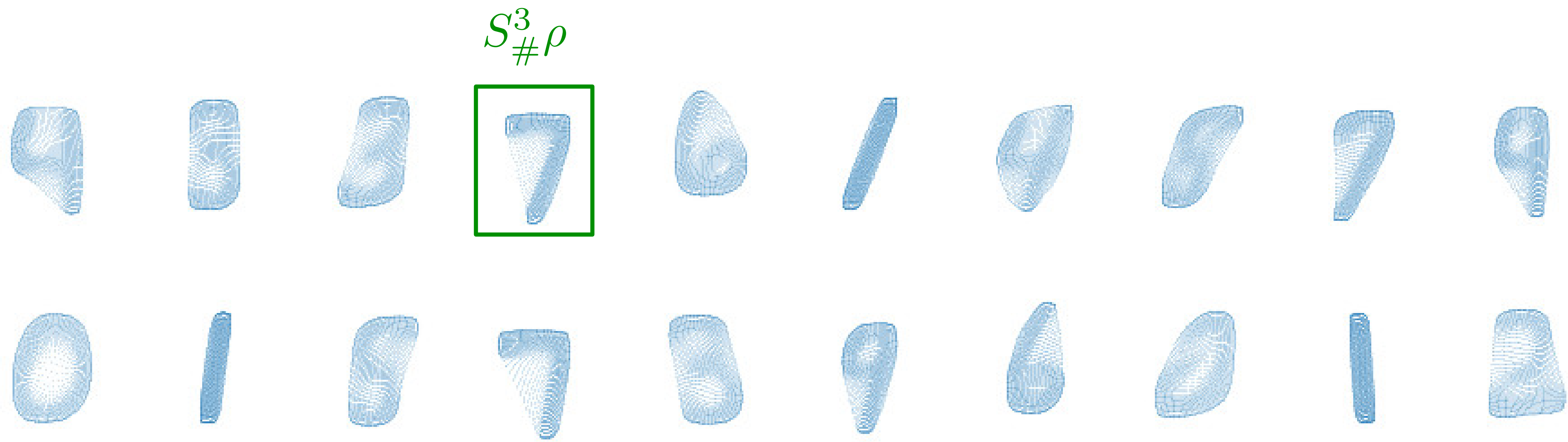
$$\mu^\ell = \frac{1}{\sum_{i,j} \alpha^\ell_{ij}} \sum_{i,j} \alpha^\ell_{i,j} \delta_{(x_i, x_j)}, \quad \text{with } x_i = \frac{i}{63}$$

$$T^\ell = T_{\mu^\ell} \in \mathrm{L}^2([0,1], \mathbb{R}^2) \quad \text{[OT map from } \rho = \mathrm{Leb}_{[0,1]^2} \text{ to } \mu^\ell]$$

We run the K-Means method on the transport plans, with $K = 20$.

Each cluster $X^k \subseteq \{0, \dots, M\}$ yields an *average transport plan* $S^k = \frac{1}{|X^k|} \sum_{\ell \in X} T^\ell$,

and $S^k_\# \rho$ is the "reconstructed measure".



2. Known properties of $\mu \mapsto T_\mu$.

Elementary remarks

- **The map** $\mu \mapsto T_\mu$ **is reverse-Lipschitz**, i.e. $\|T_\mu - T_\nu\|_{L^2(\rho)} \geq W_2(\mu, \nu)$.

Elementary remarks

- **The map $\mu \mapsto T_\mu$ is reverse-Lipschitz**, i.e. $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \geq \mathrm{W}_2(\mu, \nu)$.

  Indeed: $(T_\mu, T_\nu)_{\#}\rho$ is a coupling between μ and ν, with cost $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)}$.

Elementary remarks

- **The map $\mu \mapsto T_\mu$ is reverse-Lipschitz**, i.e. $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \geq \mathrm{W}_2(\mu, \nu)$.

  Indeed: $(T_\mu, T_\nu)_{\#}\rho$ is a coupling between μ and ν, with cost $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)}$.

- **The map $\mu \mapsto T_\mu$ is continuous.**

Elementary remarks

- **The map $\mu \mapsto T_\mu$ is reverse-Lipschitz**, i.e. $\|T_\mu - T_\nu\|_{L^2(\rho)} \geq W_2(\mu, \nu)$.

  Indeed: $(T_\mu, T_\nu)_\# \rho$ is a coupling between μ and ν, with cost $\|T_\mu - T_\nu\|_{L^2(\rho)}$.

- **The map $\mu \mapsto T_\mu$ is continuous.**

- **The map $\mu \mapsto T_\mu$ is not better than $\frac{1}{2}$-Hölder.**

Elementary remarks

- **The map $\mu \mapsto T_\mu$ is reverse-Lipschitz**, i.e. $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \geq \mathrm{W}_2(\mu,\nu)$.

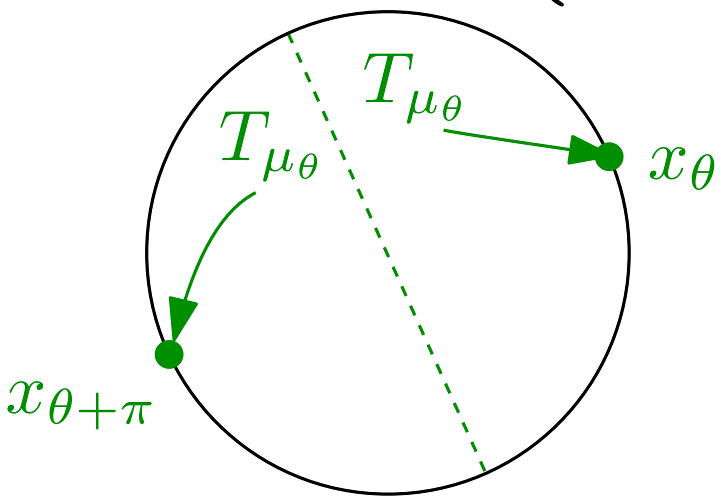
  Indeed: $(T_\mu, T_\nu)_\# \rho$ is a coupling between μ and ν, with cost $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)}$.

- **The map $\mu \mapsto T_\mu$ is continuous.**

- **The map $\mu \mapsto T_\mu$ is not better than $\frac{1}{2}$-Hölder.**

  Take $\rho = \frac{1}{\pi}\mathrm{Leb}_{\mathrm{B}(0,1)}$ on $\mathbb{R}^2$, and let $\mu_\theta = \frac{1}{2}(\delta_{x_\theta} + \delta_{x_{\theta+\pi}})$, with $x_\theta = (\cos(\theta), \sin(\theta))$.

  Then $T_{\mu_\theta}(x) = \begin{cases} x_\theta & \langle x_\theta | x\rangle \geq 0 \\ x_{\theta+\pi} & \text{if not} \end{cases}$,



Elementary remarks

- **The map $\mu \mapsto T_\mu$ is reverse-Lipschitz**, i.e. $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \geq \mathrm{W}_2(\mu,\nu)$.

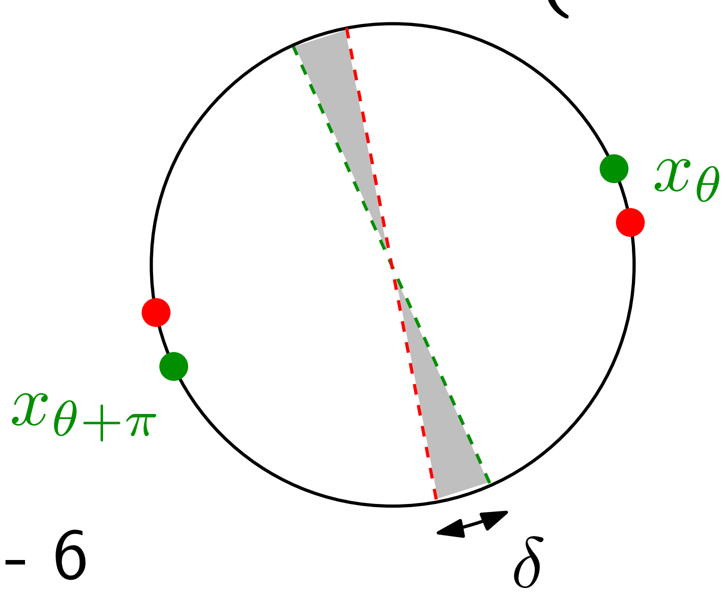
  Indeed: $(T_\mu, T_\nu)_\# \rho$ is a coupling between μ and ν, with cost $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)}$.

- **The map $\mu \mapsto T_\mu$ is continuous.**

- **The map $\mu \mapsto T_\mu$ is not better than $\frac{1}{2}$-Hölder.**

  Take $\rho = \frac{1}{\pi}\mathrm{Leb}_{\mathrm{B}(0,1)}$ on $\mathbb{R}^2$, and let $\mu_\theta = \frac{1}{2}(\delta_{x_\theta} + \delta_{x_{\theta+\pi}})$, with $x_\theta = (\cos(\theta), \sin(\theta))$.

  Then $T_{\mu_\theta}(x) = \begin{cases} x_\theta & \langle x_\theta | x \rangle \geq 0 \\ x_{\theta+\pi} & \text{if not} \end{cases}$, so that $\|T_{\mu_\theta} - T_{\mu_{\theta+\delta}}\|^2_{\mathrm{L}^2(\rho)} \geq C\delta$



Since on the other hand, $\mathrm{W}_2(\mu_\theta, \mu_{\theta+\delta}) \leq C\delta$,

$$\|T_{\mu_\theta} - T_{\mu_{\theta+\delta}}\|_{\mathrm{L}^2(\rho)} \geq C\,\mathrm{W}_2(\mu_\theta, \mu_{\theta+\delta})^{1/2}$$

Local $\frac{1}{2}$-Hölder continuity

Thm: Assume $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ and $\mu, \nu \in \mathrm{Prob}(Y)$ with $X, Y \subseteq \mathbb{R}^d$ compact

If T_μ is L-Lipschitz, then $\|T_\mu - T_\nu\|_2^2 \leq C\, \mathrm{W}_1(\mu, \nu)$ with $C = 4L\, \mathrm{diam}(X)$.

Local $\frac{1}{2}$-Hölder continuity

Thm: Assume $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ and $\mu, \nu \in \mathrm{Prob}(Y)$ with $X, Y \subseteq \mathbb{R}^d$ compact
If T_μ is L-Lipschitz, then $\|T_\mu - T_\nu\|_2^2 \leq C\,\mathrm{W}_1(\mu, \nu)$ with $C = 4L\,\mathrm{diam}(X)$.

[Ambrosio,Gigli '09], see also [Berman '18].

Local $\frac{1}{2}$-Hölder continuity

Thm: Assume $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ and $\mu, \nu \in \mathrm{Prob}(Y)$ with $X, Y \subseteq \mathbb{R}^d$ compact
If T_μ is L-Lipschitz, then $\|T_\mu - T_\nu\|_2^2 \leq C\,\mathrm{W}_1(\mu, \nu)$ with $C = 4L\,\mathrm{diam}(X)$.

[Ambrosio,Gigli '09], see also [Berman '18].

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

Local $\frac{1}{2}$-Hölder continuity

Thm: Assume $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ and $\mu, \nu \in \mathrm{Prob}(Y)$ with $X, Y \subseteq \mathbb{R}^d$ compact
If T_μ is L-Lipschitz, then $\|T_\mu - T_\nu\|_2^2 \leq C\,\mathrm{W}_1(\mu, \nu)$ with $C = 4L\,\mathrm{diam}(X)$.

[Ambrosio,Gigli '09], see also [Berman '18].

- No regularity assumption on $\nu \longrightarrow$ applicable in statistics and numerical analysis.
- The hypothesis that T_μ is Lipschitz is practically restricting:
  1) it implies that $\mathrm{spt}(\mu)$ is connected.

Local $\frac{1}{2}$-Hölder continuity

Thm: Assume $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ and $\mu, \nu \in \mathrm{Prob}(Y)$ with $X, Y \subseteq \mathbb{R}^d$ compact
If T_μ is L-Lipschitz, then $\|T_\mu - T_\nu\|_2^2 \leq C\,\mathrm{W}_1(\mu, \nu)$ with $C = 4L\,\mathrm{diam}(X)$.

[Ambrosio,Gigli '09], see also [Berman '18].

- No regularity assumption on $\nu \longrightarrow$ applicable in statistics and numerical analysis.
- The hypothesis that T_μ is Lipschitz is practically restricting:
  1) it implies that $\mathrm{spt}(\mu)$ is connected.
  2) 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 $\mathcal{C}^\alpha$ densities bounded from above and below, then T_μ is $\mathcal{C}^{1,\alpha}$.

Global Hölder continuity

Theorem (Berman '18): Assume that ρ is the Lebesgue measure on X, and $\mu, \nu \in \mathrm{Prob}(Y)$ with X convex compact and Y compact. Then,

$$\|T_\mu - T_\nu\|^2_{\mathrm{L}^2(\rho)} \leq C\,\mathrm{W}_1(\mu,\nu)^\alpha \text{ with } \alpha = \frac{1}{2^{d-1}(d+2)}$$

Global Hölder continuity

Theorem (Berman '18): Assume that ρ is the Lebesgue measure on X, and $\mu, \nu \in \mathrm{Prob}(Y)$ with X convex compact and Y compact. Then,

$$\|T_\mu - T_\nu\|^2_{\mathrm{L}^2(\rho)} \leq C\,\mathrm{W}_1(\mu,\nu)^\alpha \text{ with } \alpha = \frac{1}{2^{d-1}(d+2)}$$

- The Hölder exponent is not tight, but the inequality holds without regularity assumption on μ, ν!

Global Hölder continuity

Theorem (Berman '18): Assume that ρ is the Lebesgue measure on X, and $\mu, \nu \in \mathrm{Prob}(Y)$ with X convex compact and Y compact. Then,

$$\|T_\mu - T_\nu\|^2_{\mathrm{L}^2(\rho)} \leq C\,\mathrm{W}_1(\mu,\nu)^\alpha \text{ with } \alpha = \frac{1}{2^{d-1}(d+2)}$$

- The Hölder exponent is not tight, but the inequality holds without regularity assumption on μ, ν!
- Proof of Berman's theorem relies on techniques from complex geometry, and in particular an inequality due to Blocki.

Global Hölder continuity

Theorem (Berman '18): Assume that ρ is the Lebesgue measure on X, and $\mu, \nu \in \mathrm{Prob}(Y)$ with X convex compact and Y compact. Then,

$$\|T_\mu - T_\nu\|^2_{\mathrm{L}^2(\rho)} \leq C\,\mathrm{W}_1(\mu,\nu)^\alpha \text{ with } \alpha = \frac{1}{2^{d-1}(d+2)}$$

- The Hölder exponent is not tight, but the inequality holds without regularity assumption on μ, ν!
- Proof of Berman's theorem relies on techniques from complex geometry, and in particular an inequality due to Blocki.

- By [Andoni, Naor, Neiman '18], the space $(\mathrm{Prob}_2(\mathbb{R}^d), W_2)$ does not admit a bi-Hölder embedding into any L^p space when $d \geq 3$.

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

Main theorem

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu,\nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

Main theorem

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu,\nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

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

Main theorem

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu,\nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

- 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 exponent $\frac{1}{6}$ is probably not optimal...

Main theorem

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu,\nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

- 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 exponent $\frac{1}{6}$ is probably not optimal...
- The constant $C(X,Y) \lesssim \mathrm{diam}(X)^{d+1}\,\mathrm{diam}(Y)$.

Main theorem

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu,\nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

- 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 exponent $\frac{1}{6}$ is probably not optimal...
- The constant $C(X,Y) \lesssim \mathrm{diam}(X)^{d+1}\,\mathrm{diam}(Y)$.
- Proof relies on the semidiscrete setting, i.e. the bound is established in the case

$$\mu = \textstyle\sum_i \mu_i \delta_{y_i},\ \nu = \sum_i \nu_i \delta_{y_i}.$$

  and one concludes using a density argument.

Semidiscrete OT for $c(x,y) = -\langle x|y\rangle$

- Let $\rho, \nu \in \mathrm{Prob}_1^{\mathrm{ac}}(\mathbb{R}^d)$ and $\Gamma(\rho,\mu) =$ couplings between ρ, μ,

  $\mathcal{T}(\rho,\mu) = \max_{\gamma\in\Gamma(\rho,\mu)} \int \langle x|y\rangle \,\mathrm{d}\,\gamma(x,y)$

Semidiscrete OT for $c(x, y) = -\langle x|y\rangle$

- Let $\rho, \nu \in \mathrm{Prob}_1^{\mathrm{ac}}(\mathbb{R}^d)$ and $\Gamma(\rho, \mu) =$ couplings between ρ, μ,

$$\mathcal{T}(\rho, \mu) = \max_{\gamma \in \Gamma(\rho,\mu)} \int \langle x|y\rangle \,\mathrm{d}\,\gamma(x, y)$$

$$= \min_{\phi \oplus \psi \geq \langle \cdot|\cdot\rangle} \int \phi \,\mathrm{d}\,\rho + \int \psi \,\mathrm{d}\,\mu$$

Kantorovich duality

Semidiscrete OT for $c(x,y) = -\langle x|y\rangle$

- Let $\rho, \nu \in \mathrm{Prob}_1^{\mathrm{ac}}(\mathbb{R}^d)$ and $\Gamma(\rho,\mu)$ = couplings between ρ, μ,

$$\mathcal{T}(\rho,\mu) = \max_{\gamma\in\Gamma(\rho,\mu)} \int \langle x|y\rangle \,\mathrm{d}\,\gamma(x,y)$$

$$= \min_{\phi\oplus\psi\geq\langle\cdot|\cdot\rangle} \int \phi \,\mathrm{d}\,\rho + \int \psi \,\mathrm{d}\,\mu$$

Kantorovich duality

$$= \min_{\psi} \int \psi^* \,\mathrm{d}\,\rho + \int \psi \,\mathrm{d}\,\mu$$

Legendre-Fenchel transform:

$$\psi^*(x) = \max_y \langle x|y\rangle - \psi(y)$$

Semidiscrete OT for $c(x,y) = -\langle x|y\rangle$

- Let $\rho, \nu \in \mathrm{Prob}_1^{\mathrm{ac}}(\mathbb{R}^d)$ and $\Gamma(\rho,\mu)$ = couplings between ρ, μ,

$$\mathcal{T}(\rho,\mu) = \max_{\gamma\in\Gamma(\rho,\mu)} \int \langle x|y\rangle \,\mathrm{d}\,\gamma(x,y)$$

$$= \min_{\phi\oplus\psi\geq\langle\cdot|\cdot\rangle} \int \phi \,\mathrm{d}\,\rho + \int \psi \,\mathrm{d}\,\mu$$

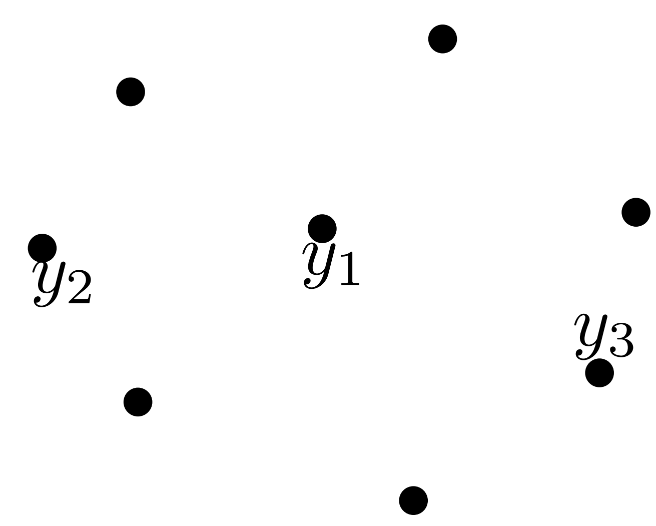
Kantorovich duality

$$= \min_{\psi} \int \psi^* \,\mathrm{d}\,\rho + \int \psi \,\mathrm{d}\,\mu$$

Legendre-Fenchel transform:

$$\psi^*(x) = \max_y \langle x|y\rangle - \psi(y)$$

- Let $\mu = \sum_{1\leq i\leq N} \mu_i \delta_{y_i}$ and $\psi_i = \psi(y_i)$.



Semidiscrete OT for $c(x, y) = -\langle x|y\rangle$

- Let $\rho, \nu \in \mathrm{Prob}_1^{\mathrm{ac}}(\mathbb{R}^d)$ and $\Gamma(\rho, \mu)$ = couplings between ρ, μ,

$$\mathcal{T}(\rho, \mu) = \max_{\gamma \in \Gamma(\rho,\mu)} \int \langle x|y\rangle \,\mathrm{d}\,\gamma(x, y)$$

$$\boxed{= \min_{\phi \oplus \psi \geq \langle \cdot|\cdot\rangle} \int \phi \,\mathrm{d}\,\rho + \int \psi \,\mathrm{d}\,\mu}$$

$$= \min_\psi \int \psi^* \,\mathrm{d}\,\rho + \int \psi \,\mathrm{d}\,\mu$$

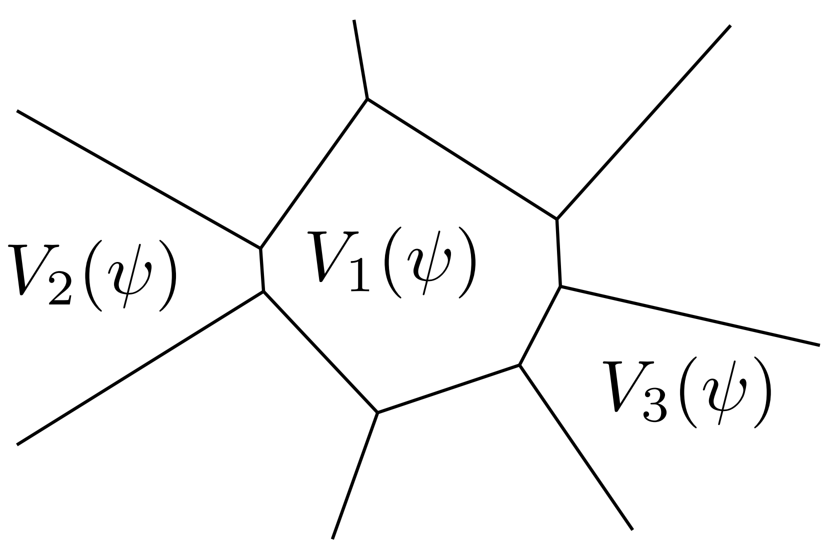
Kantorovich duality

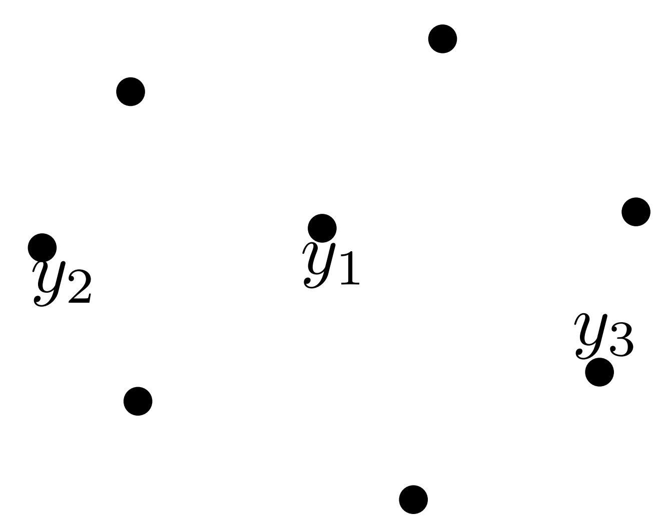
Legendre-Fenchel transform:

$$\psi^*(x) = \max_y \langle x|y\rangle - \psi(y)$$

- Let $\mu = \sum_{1 \leq i \leq N} \mu_i \delta_{y_i}$ and $\psi_i = \psi(y_i)$. Then, $\psi^*|_{V_i(\psi)} := \langle \cdot|y_i\rangle - \psi_i$ where

$$\boxed{V_i(\psi) = \{x \mid \forall j,\ \langle x|y_i\rangle - \psi_i \geq \langle x|y_j\rangle - \psi_j\}}$$





Semidiscrete OT for $c(x, y) = -\langle x|y \rangle$

- Let $\rho, \nu \in \mathrm{Prob}_1^{\mathrm{ac}}(\mathbb{R}^d)$ and $\Gamma(\rho, \mu)$ = couplings between ρ, μ,

$$\mathcal{T}(\rho, \mu) = \max_{\gamma \in \Gamma(\rho,\mu)} \int \langle x|y \rangle \,\mathrm{d}\,\gamma(x, y)$$

$$= \min_{\phi \oplus \psi \geq \langle \cdot|\cdot \rangle} \int \phi \,\mathrm{d}\,\rho + \int \psi \,\mathrm{d}\,\mu$$

Kantorovich duality

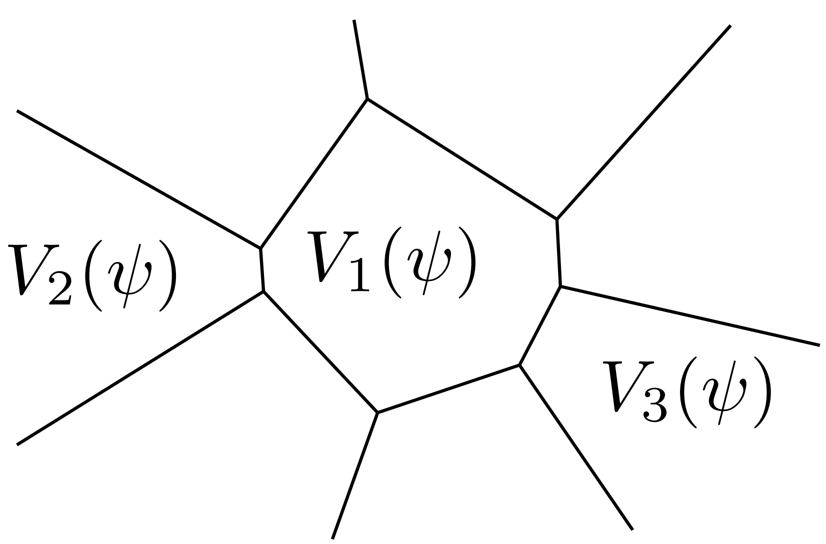
$$= \min_{\psi} \int \psi^* \,\mathrm{d}\,\rho + \int \psi \,\mathrm{d}\,\mu$$

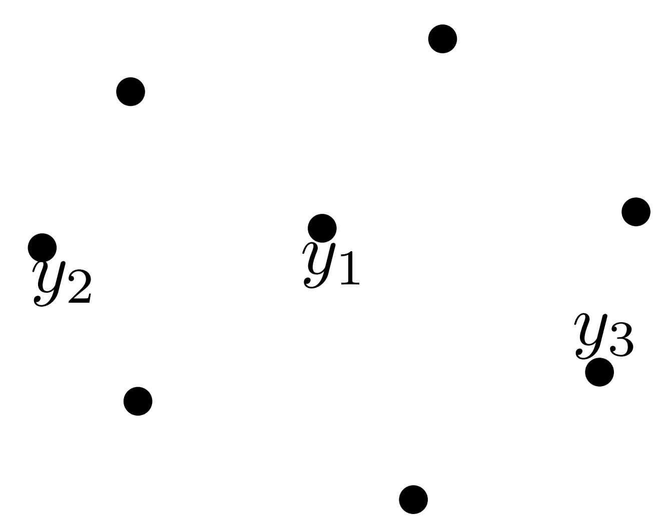
Legendre-Fenchel transform:

$$\psi^*(x) = \max_y \langle x|y \rangle - \psi(y)$$

- Let $\mu = \sum_{1 \leq i \leq N} \mu_i \delta_{y_i}$ and $\psi_i = \psi(y_i)$. Then, $\psi^*|_{V_i(\psi)} := \langle \cdot|y_i \rangle - \psi_i$ where

$$V_i(\psi) = \{x \mid \forall j,\ \langle x|y_i \rangle - \psi_i \geq \langle x|y_j \rangle - \psi_j\}$$





Thus, $\mathcal{T}(\rho, \mu) = \min_{\psi \in \mathbb{R}^N} \sum_i \int_{V_i(\psi)} \langle x|y_i \rangle - \psi_i \,\mathrm{d}\,\rho(x) + \sum_i \mu_i \psi_i$

Optimality condition and economic interpretation

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) + \langle \mu|\psi\rangle$, where:

$$\Phi(\psi) := \sum_i \int_{V_i(\psi)} \langle x|y_i\rangle - \psi_i \,\mathrm{d}\,\rho(x)$$

Optimality condition and economic interpretation

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) + \langle \mu|\psi\rangle$, where:

$$\Phi(\psi) := \sum_i \int_{V_i(\psi)} \langle x|y_i\rangle - \psi_i \,\mathrm{d}\,\rho(x)$$

- **Gradient:** $\nabla\Phi(\psi) = -(G_i(\psi))_{1\leq i\leq N}$ where $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$.

Optimality condition and economic interpretation

$\mathcal{T}(\rho, \mu) = \min_{\psi \in \mathbb{R}^N} \Phi(\psi) + \langle \mu | \psi \rangle$, where:

$$\boxed{\Phi(\psi) := \sum_i \int_{V_i(\psi)} \langle x | y_i \rangle - \psi_i \,\mathrm{d}\,\rho(x)}$$

- **Gradient:** $\nabla \Phi(\psi) = -(G_i(\psi))_{1 \leq i \leq N}$ where $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$.

$\psi \in \mathbb{R}^N$ minimizes $\Phi + \langle \mu | \cdot \rangle \Longleftrightarrow \nabla \Phi(\psi) = -\mu$

Optimality condition and economic interpretation

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) + \langle \mu|\psi\rangle$, where:

$$\boxed{\Phi(\psi) := \sum_i \int_{V_i(\psi)} \langle x|y_i\rangle - \psi_i \,\mathrm{d}\,\rho(x)}$$

- **Gradient:** $\nabla\Phi(\psi) = -(G_i(\psi))_{1\leq i\leq N}$ where $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$.

$\psi \in \mathbb{R}^N$ minimizes $\Phi + \langle\mu|\cdot\rangle \Longleftrightarrow \nabla\Phi(\psi) = -\mu$

$\Longleftrightarrow G(\psi) = \mu$ with $G = (G_1,\dots,G_N)$, $\mu\in\mathbb{R}^N$

Optimality condition and economic interpretation

$\mathcal{T}(\rho, \mu) = \min_{\psi \in \mathbb{R}^N} \Phi(\psi) + \langle \mu | \psi \rangle$, where:

$$\boxed{\Phi(\psi) := \sum_i \int_{V_i(\psi)} \langle x | y_i \rangle - \psi_i \, \mathrm{d}\, \rho(x)}$$

- **Gradient:** $\nabla \Phi(\psi) = -(G_i(\psi))_{1 \leq i \leq N}$ where $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\, \rho$.

$$\begin{aligned}
\psi \in \mathbb{R}^N \text{ minimizes } \Phi + \langle \mu | \cdot \rangle &\Longleftrightarrow \nabla \Phi(\psi) = -\mu \\
&\Longleftrightarrow G(\psi) = \mu \text{ with } G = (G_1, \dots, G_N),\ \mu \in \mathbb{R}^N \\
&\Longleftrightarrow T = \nabla \psi^* \text{ transports } \rho \text{ onto } \textstyle\sum_i \mu_i \delta_{y_i}
\end{aligned}$$

Optimality condition and economic interpretation

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) + \langle \mu|\psi\rangle$, where:

$$\boxed{\Phi(\psi) := \sum_i \int_{V_i(\psi)} \langle x|y_i\rangle - \psi_i \,\mathrm{d}\,\rho(x)}$$

- **Gradient:** $\nabla\Phi(\psi) = -(G_i(\psi))_{1\leq i\leq N}$ where $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$.

$\psi\in\mathbb{R}^N$ minimizes $\Phi + \langle\mu|\cdot\rangle \iff \nabla\Phi(\psi) = -\mu$

$\iff G(\psi) = \mu$ with $G = (G_1,\dots,G_N)$, $\mu\in\mathbb{R}^N$

$\iff T = \nabla\psi^*$ transports ρ onto $\sum_i \mu_i\delta_{y_i}$

- **Economic interpretation:** $\rho =$ density of customers, $\{y_i\}_{1\leq i\leq N} =$ product types

Optimality condition and economic interpretation

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) + \langle \mu|\psi\rangle$, where:

$$\boxed{\Phi(\psi) := \sum_i \int_{V_i(\psi)} \langle x|y_i\rangle - \psi_i \,\mathrm{d}\,\rho(x)}$$

- **Gradient:** $\nabla\Phi(\psi) = -(G_i(\psi))_{1\leq i\leq N}$ where $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$.

$\psi\in\mathbb{R}^N$ minimizes $\Phi + \langle\mu|\cdot\rangle \iff \nabla\Phi(\psi) = -\mu$

$\iff G(\psi) = \mu$ with $G = (G_1,\dots,G_N)$, $\mu\in\mathbb{R}^N$

$\iff T = \nabla\psi^*$ transports ρ onto $\sum_i \mu_i\delta_{y_i}$

- **Economic interpretation:** $\rho =$ density of customers, $\{y_i\}_{1\leq i\leq N} =$ product types

$\longrightarrow$ given prices $\psi\in\mathbb{R}^N$, a customer x maximizes $\langle x|y_i\rangle - \psi_i$ over all products.

Optimality condition and economic interpretation

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) + \langle\mu|\psi\rangle$, where: $\boxed{\Phi(\psi) := \sum_i \int_{V_i(\psi)} \langle x|y_i\rangle - \psi_i \,\mathrm{d}\,\rho(x)}$

- **Gradient:** $\nabla\Phi(\psi) = -(G_i(\psi))_{1\leq i\leq N}$ where $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$.

$\psi \in \mathbb{R}^N$ minimizes $\Phi + \langle\mu|\cdot\rangle \Longleftrightarrow \nabla\Phi(\psi) = -\mu$

$\Longleftrightarrow G(\psi) = \mu$ with $G = (G_1,\dots,G_N)$, $\mu\in\mathbb{R}^N$

$\Longleftrightarrow T = \nabla\psi^*$ transports ρ onto $\sum_i \mu_i\delta_{y_i}$

- **Economic interpretation:** $\rho =$ density of customers, $\{y_i\}_{1\leq i\leq N} =$ product types

$\longrightarrow$ given prices $\psi\in\mathbb{R}^N$, a customer x maximizes $\langle x|y_i\rangle - \psi_i$ over all products.

$\longrightarrow V_i(\psi) = \{x \mid i \in \arg\max_j \langle x|y_j\rangle - \psi_j\} =$ customers choosing product y_i.

Optimality condition and economic interpretation

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) + \langle \mu|\psi\rangle$, where: $\boxed{\Phi(\psi) := \sum_i \int_{V_i(\psi)} \langle x|y_i\rangle - \psi_i \,\mathrm{d}\,\rho(x)}$

- **Gradient:** $\nabla\Phi(\psi) = -(G_i(\psi))_{1\leq i\leq N}$ where $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$.

$\psi \in \mathbb{R}^N$ minimizes $\Phi + \langle\mu|\cdot\rangle \iff \nabla\Phi(\psi) = -\mu$

$\iff G(\psi) = \mu$ with $G = (G_1,\dots,G_N)$, $\mu\in\mathbb{R}^N$

$\iff T = \nabla\psi^*$ transports ρ onto $\sum_i \mu_i\delta_{y_i}$

- **Economic interpretation:** $\rho =$ density of customers, $\{y_i\}_{1\leq i\leq N} =$ product types

$\longrightarrow$ given prices $\psi\in\mathbb{R}^N$, a customer x maximizes $\langle x|y_i\rangle - \psi_i$ over all products.

$\longrightarrow V_i(\psi) = \{x \mid i \in \arg\max_j \langle x|y_j\rangle - \psi_j\} =$ customers choosing product y_i.

$\longrightarrow G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho =$ amount of customers for product y_i.

Optimality condition and economic interpretation

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) + \langle\mu|\psi\rangle$, where:

$$\Phi(\psi) := \sum_i \int_{V_i(\psi)} \langle x|y_i\rangle - \psi_i \,\mathrm{d}\,\rho(x)$$

- **Gradient:** $\nabla\Phi(\psi) = -(G_i(\psi))_{1\leq i\leq N}$ where $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$.

$\psi \in \mathbb{R}^N$ minimizes $\Phi + \langle\mu|\cdot\rangle \Longleftrightarrow \nabla\Phi(\psi) = -\mu$

$\Longleftrightarrow G(\psi) = \mu$ with $G = (G_1,\ldots,G_N)$, $\mu\in\mathbb{R}^N$

$\Longleftrightarrow T = \nabla\psi^*$ transports ρ onto $\sum_i \mu_i\delta_{y_i}$

- **Economic interpretation:** $\rho =$ density of customers, $\{y_i\}_{1\leq i\leq N} =$ product types

$\longrightarrow$ given prices $\psi\in\mathbb{R}^N$, a customer x maximizes $\langle x|y_i\rangle - \psi_i$ over all products.

$\longrightarrow V_i(\psi) = \{x \mid i \in \arg\max_j \langle x|y_j\rangle - \psi_j\} =$ customers choosing product y_i.

$\longrightarrow G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho =$ amount of customers for product y_i.

Optimal transport $=$ finding prices satisfying capacity constraints $G_i(\psi) = \mu_i$.

Optimality condition and economic interpretation

$\mathcal{T}(\rho, \mu) = \min_{\psi \in \mathbb{R}^N} \Phi(\psi) + \langle \mu | \psi \rangle$, where: $\boxed{\Phi(\psi) := \sum_i \int_{V_i(\psi)} \langle x | y_i \rangle - \psi_i \, \mathrm{d}\rho(x)}$

- **Gradient:** $\nabla \Phi(\psi) = -(G_i(\psi))_{1 \leq i \leq N}$ where $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\rho$.

$\psi \in \mathbb{R}^N$ minimizes $\Phi + \langle \mu | \cdot \rangle \iff \nabla \Phi(\psi) = -\mu$

$\iff G(\psi) = \mu$ with $G = (G_1, \ldots, G_N)$, $\mu \in \mathbb{R}^N$

$\iff T = \nabla \psi^*$ transports ρ onto $\sum_i \mu_i \delta_{y_i}$

- **Economic interpretation:** $\rho =$ density of customers, $\{y_i\}_{1 \leq i \leq N} =$ product types

$\longrightarrow$ given prices $\psi \in \mathbb{R}^N$, a customer x maximizes $\langle x | y_i \rangle - \psi_i$ over all products.

$\longrightarrow$ $V_i(\psi) = \{x \mid i \in \arg\max_j \langle x | y_j \rangle - \psi_j\} =$ customers choosing product y_i.

$\longrightarrow$ $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\rho =$ amount of customers for product y_i.

Optimal transport $=$ finding prices satisfying capacity constraints $G_i(\psi) = \mu_i$.

- Hölder-stability of optimal transport maps $\simeq$ strong concavity of Φ.

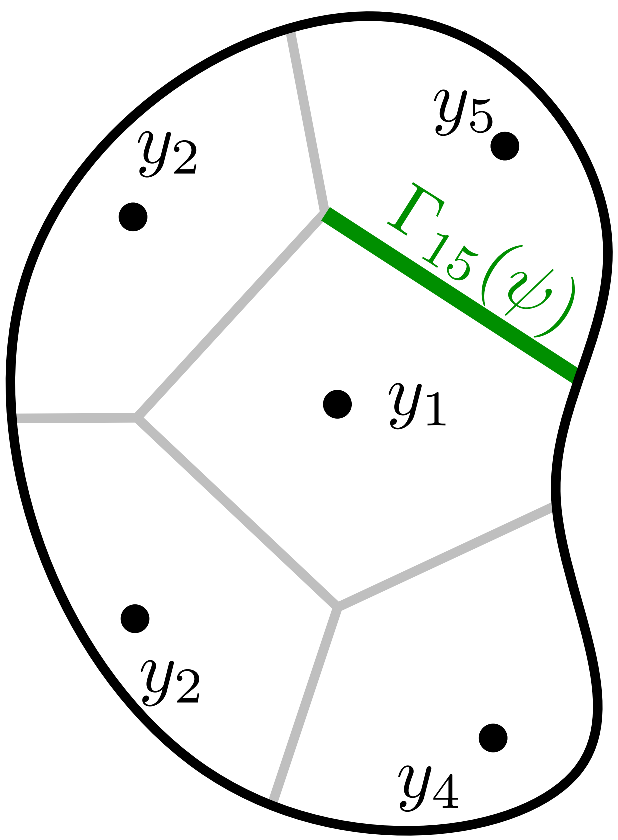
Hessian of Φ and strong convexity

(Recall that $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$, $\nabla\Phi = -(G_1, \dots, G_N)$, $\mathrm{D}G = -\mathrm{D}^2\Phi$)

Proposition: ▶ If $\rho \in \mathcal{C}^0(X)$ and $(y_i)_{1 \le i \le N}$ is generic, then $\Phi \in \mathcal{C}^2(\mathbb{R}^N)$ and

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x)\,\mathrm{d}\,x \text{ where } \Gamma_{ij} = V_i(\psi) \cap V_j(\psi).$$

$$\forall i, \quad \frac{\partial G_i}{\partial \psi_i}(\psi) = -\sum_{j \neq i} \frac{\partial G_i}{\partial \psi_j}(\psi)$$



Hessian of Φ and strong convexity

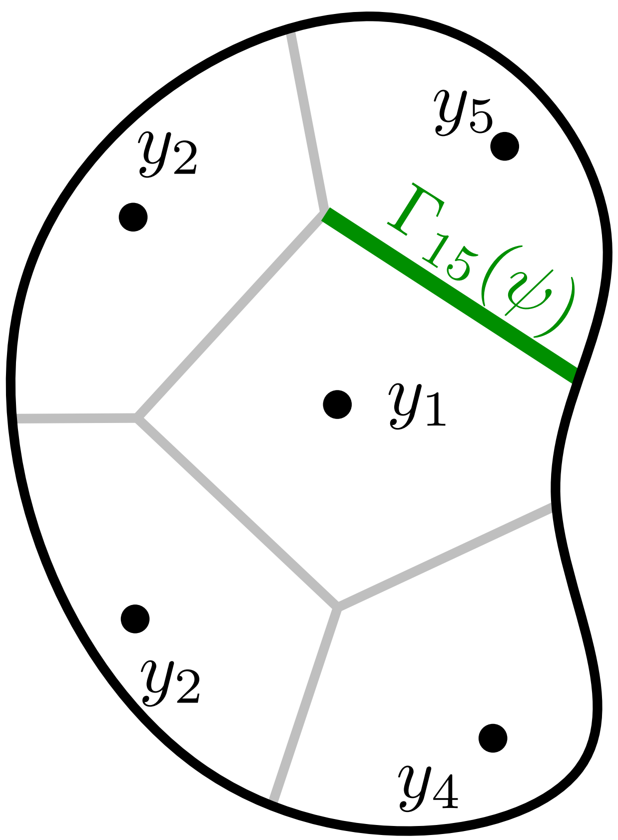
(Recall that $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$, $\nabla\Phi = -(G_1, \dots, G_N)$, $\mathrm{D}G = -\mathrm{D}^2\Phi$)

Proposition: ▶ If $\rho \in \mathcal{C}^0(X)$ and $(y_i)_{1\leq i\leq N}$ is generic, then $\Phi \in \mathcal{C}^2(\mathbb{R}^N)$ and

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x)\,\mathrm{d}\,x \text{ where } \Gamma_{ij} = V_i(\psi) \cap V_j(\psi).$$

$$\forall i, \quad \frac{\partial G_i}{\partial \psi_i}(\psi) = -\sum_{j\neq i} \frac{\partial G_i}{\partial \psi_j}(\psi)$$

▶ If $\Omega = \{\rho > 0\}$ is connected and $\forall i, G_i(\psi) > 0$, then $\mathrm{Ker}(\mathrm{D}G(\psi)) = \mathbb{R}(1, \dots, 1)$.



Hessian of Φ and strong convexity

(Recall that $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$, $\nabla\Phi = -(G_1, \dots, G_N)$, $\mathrm{D}G = -\mathrm{D}^2\Phi$)

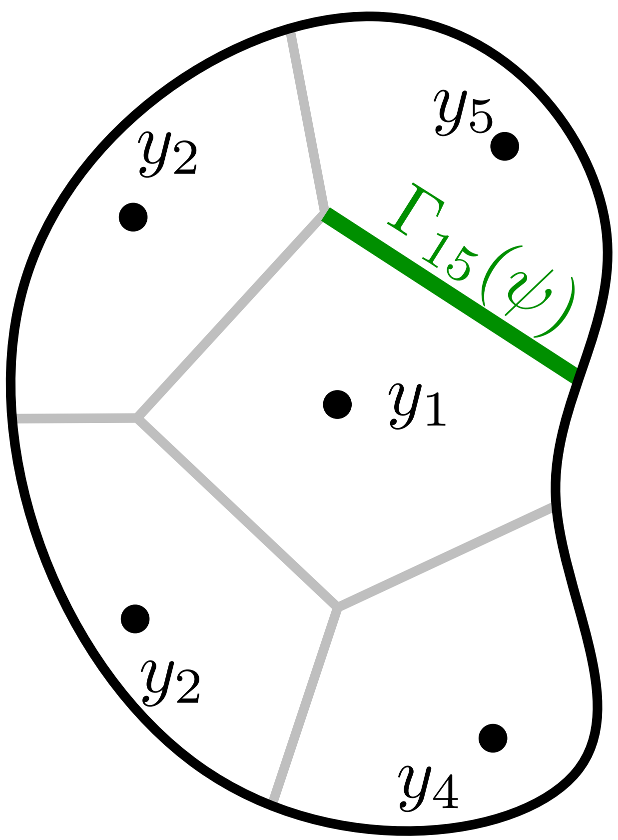
Proposition: ▶ If $\rho \in \mathcal{C}^0(X)$ and $(y_i)_{1\leq i\leq N}$ is generic, then $\Phi \in \mathcal{C}^2(\mathbb{R}^N)$ and

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x)\,\mathrm{d}\,x \text{ where } \Gamma_{ij} = V_i(\psi) \cap V_j(\psi).$$

$$\forall i, \quad \frac{\partial G_i}{\partial \psi_i}(\psi) = -\sum_{j\neq i} \frac{\partial G_i}{\partial \psi_j}(\psi)$$

▶ If $\Omega = \{\rho > 0\}$ is connected and $\forall i, G_i(\psi) > 0$, then $\mathrm{Ker}(\mathrm{D}G(\psi)) = \mathbb{R}(1, \dots, 1)$.

NB: if $V_i(\psi) = \emptyset$, then $\mathbf{1}_{\{y_i\}} \in \mathrm{Ker}(\mathrm{D}^2\Phi(\psi))$



Hessian of Φ and strong convexity

(Recall that $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$, $\nabla\Phi = -(G_1, \dots, G_N)$, $\mathrm{D}G = -\mathrm{D}^2\Phi$)

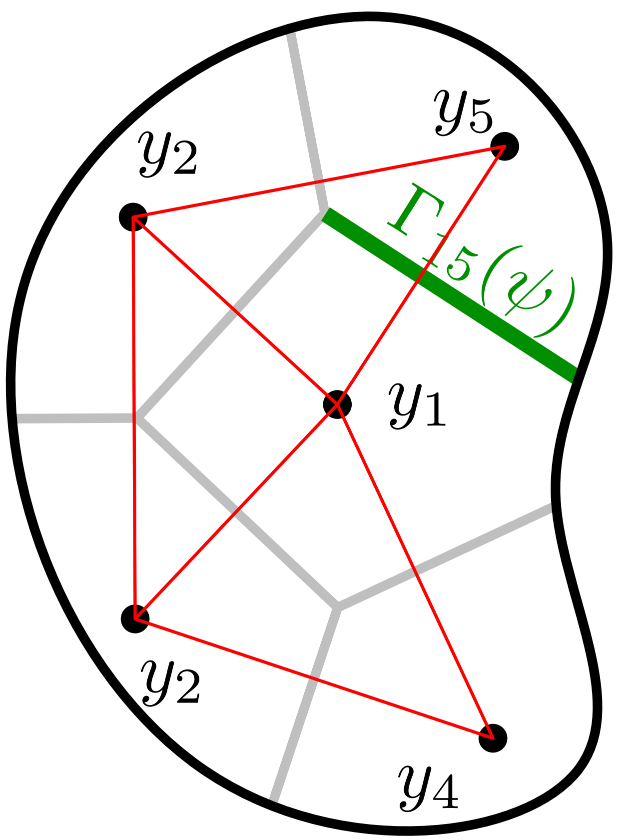
Proposition: ▶ If $\rho \in \mathcal{C}^0(X)$ and $(y_i)_{1\le i\le N}$ is generic, then $\Phi \in \mathcal{C}^2(\mathbb{R}^N)$ and

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x)\,\mathrm{d}\,x \text{ where } \Gamma_{ij} = V_i(\psi) \cap V_j(\psi).$$

$$\forall i, \quad \frac{\partial G_i}{\partial \psi_i}(\psi) = -\sum_{j\neq i} \frac{\partial G_i}{\partial \psi_j}(\psi)$$

▶ If $\Omega = \{\rho > 0\}$ is connected and $\forall i, G_i(\psi) > 0$, then $\mathrm{Ker}(\mathrm{D}G(\psi)) = \mathbb{R}(1, \dots, 1)$.

NB: if $V_i(\psi) = \emptyset$, then $\mathbf{1}_{\{y_i\}} \in \mathrm{Ker}(\mathrm{D}^2\Phi(\psi))$



Proof:

▶ Consider the matrix $L = DG(\psi)$ and the graph H:

$$(i, j) \in H \iff L_{ij} > 0$$

Hessian of Φ and strong convexity

(Recall that $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$, $\nabla\Phi = -(G_1, \dots, G_N)$, $\mathrm{D}G = -\mathrm{D}^2\Phi$)

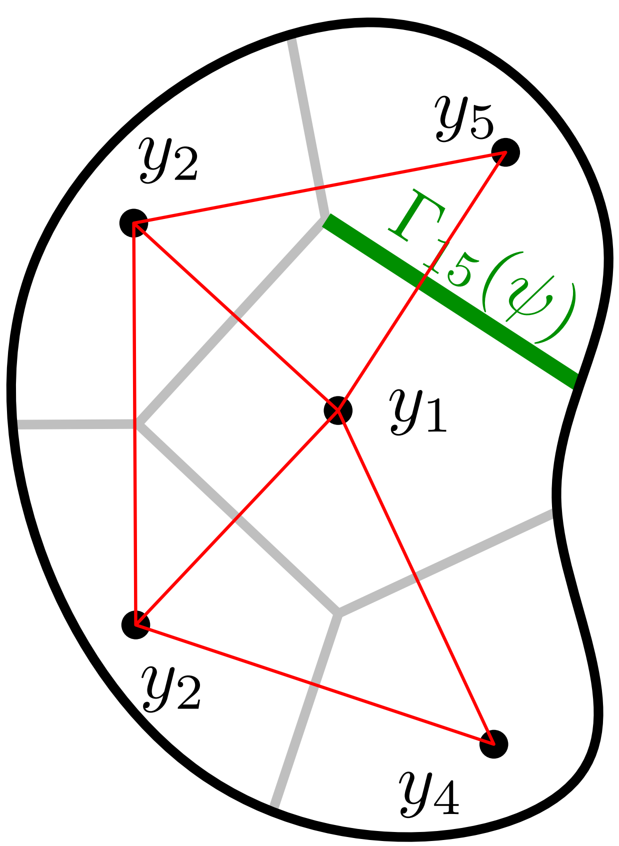
Proposition: ▶ If $\rho \in \mathcal{C}^0(X)$ and $(y_i)_{1\leq i \leq N}$ is generic, then $\Phi \in \mathcal{C}^2(\mathbb{R}^N)$ and

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x)\,\mathrm{d}\,x \text{ where } \Gamma_{ij} = V_i(\psi) \cap V_j(\psi).$$

$$\forall i, \quad \frac{\partial G_i}{\partial \psi_i}(\psi) = -\sum_{j\neq i} \frac{\partial G_i}{\partial \psi_j}(\psi)$$

▶ If $\Omega = \{\rho > 0\}$ is connected and $\forall i, G_i(\psi) > 0$, then $\mathrm{Ker}(\mathrm{D}G(\psi)) = \mathbb{R}(1, \dots, 1)$.

NB: if $V_i(\psi) = \emptyset$, then $\mathbf{1}_{\{y_i\}} \in \mathrm{Ker}(\mathrm{D}^2\Phi(\psi))$



Proof:

▶ Consider the matrix $L = DG(\psi)$ and the graph H:

$$(i,j) \in H \iff L_{ij} > 0$$

▶ If Ω is connected and $\psi \in E$, then H is connected

Hessian of Φ and strong convexity

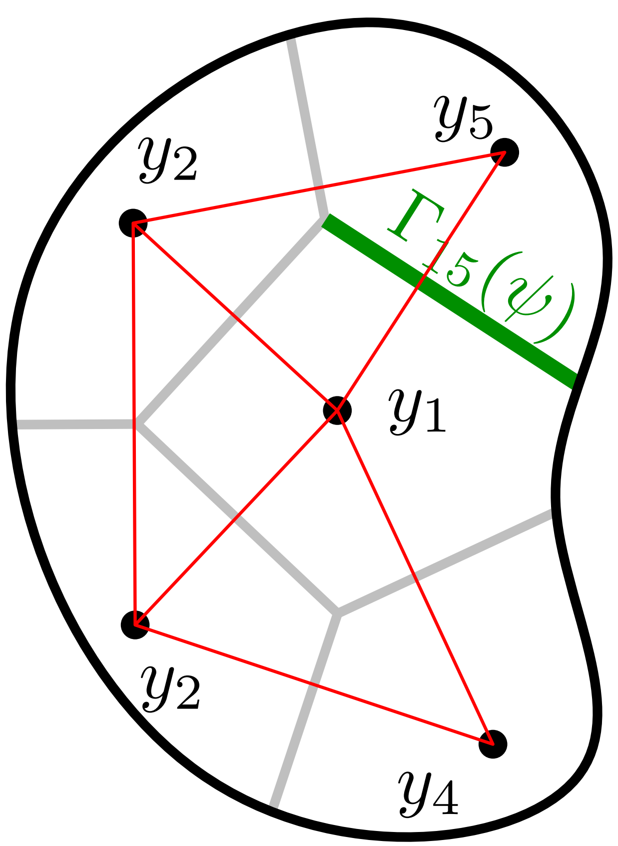
(Recall that $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$, $\nabla\Phi = -(G_1, \dots, G_N)$, $\mathrm{D}G = -\mathrm{D}^2\Phi$)

Proposition: ▶ If $\rho \in \mathcal{C}^0(X)$ and $(y_i)_{1\leq i\leq N}$ is generic, then $\Phi \in \mathcal{C}^2(\mathbb{R}^N)$ and

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x)\,\mathrm{d}\,x \text{ where } \Gamma_{ij} = V_i(\psi) \cap V_j(\psi).$$

$$\forall i, \quad \frac{\partial G_i}{\partial \psi_i}(\psi) = -\sum_{j\neq i} \frac{\partial G_i}{\partial \psi_j}(\psi)$$

▶ If $\Omega = \{\rho > 0\}$ is connected and $\forall i, G_i(\psi) > 0$, then $\mathrm{Ker}(\mathrm{D}G(\psi)) = \mathbb{R}(1, \dots, 1)$.



NB: if $V_i(\psi) = \emptyset$, then $\mathbf{1}_{\{y_i\}} \in \mathrm{Ker}(\mathrm{D}^2\Phi(\psi))$

Proof:

- ▶ Consider the matrix $L = DG(\psi)$ and the graph H:
  $$(i,j) \in H \iff L_{ij} > 0$$
- ▶ If Ω is connected and $\psi \in E$, then H is connected
- ▶ L is the Laplacian of a connected graph $\implies \mathrm{Ker} L = \mathbb{R} \cdot \mathrm{cst}$

Hessian of Φ and strong convexity

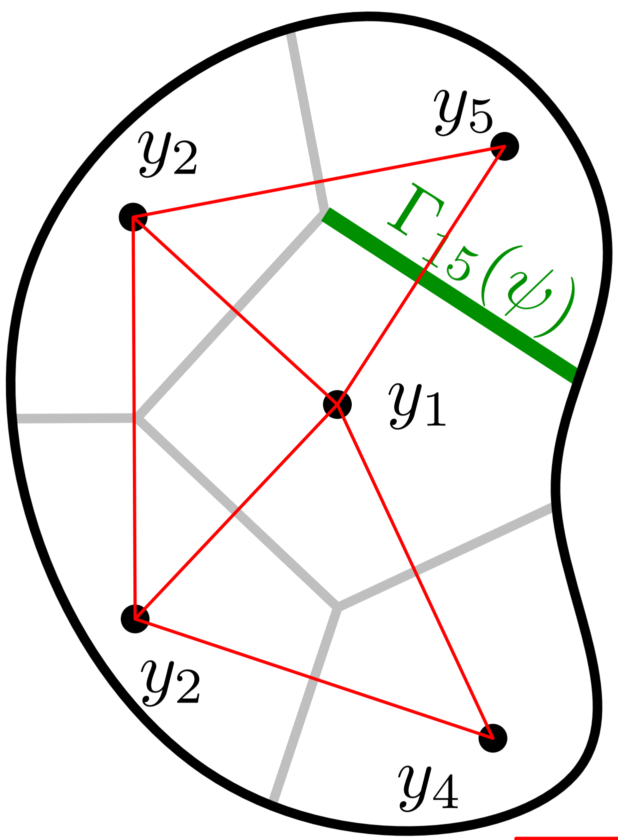
(Recall that $G_i(\psi) = \int_{V_i(\psi)} \mathrm{d}\,\rho$, $\nabla\Phi = -(G_1, \dots, G_N)$, $\mathrm{D}G = -\mathrm{D}^2\Phi$)

Proposition: ▶ If $\rho \in \mathcal{C}^0(X)$ and $(y_i)_{1\leq i\leq N}$ is generic, then $\Phi \in \mathcal{C}^2(\mathbb{R}^N)$ and

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\psi) = \frac{1}{\|y_i - y_j\|}\int_{\Gamma_{ij}(\psi)} \rho(x)\,\mathrm{d}\,x \text{ where } \Gamma_{ij} = V_i(\psi) \cap V_j(\psi).$$

$$\forall i, \quad \frac{\partial G_i}{\partial \psi_i}(\psi) = -\sum_{j\neq i} \frac{\partial G_i}{\partial \psi_j}(\psi)$$

▶ If $\Omega = \{\rho > 0\}$ is connected and $\forall i, G_i(\psi) > 0$, then $\mathrm{Ker}(\mathrm{D}G(\psi)) = \mathbb{R}(1, \dots, 1)$.



NB: if $V_i(\psi) = \emptyset$, then $\mathbf{1}_{\{y_i\}} \in \mathrm{Ker}(\mathrm{D}^2\Phi(\psi))$

Proof:

- ▶ Consider the matrix $L = DG(\psi)$ and the graph H:
  $$(i,j) \in H \iff L_{ij} > 0$$
- ▶ If Ω is connected and $\psi \in E$, then H is connected
- ▶ L is the Laplacian of a connected graph $\Longrightarrow \mathrm{Ker}L = \mathbb{R}\cdot\mathrm{cst}$

Proposition $\Longrightarrow$ local strong convexity of Φ, albeit **non-quantitative**.

Proof ingredients

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu,\nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

Proof ingredients

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu, \nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

- **Strategy of proof**: let $\mu^k = \sum_i \mu_i^k \delta_{y_i}$ for $k \in \{0, 1\}$, assume all $\mu_i^k > 0$.

Proof ingredients

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu,\nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu,\nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

- **Strategy of proof**: let $\mu^k = \sum_i \mu_i^k \delta_{y_i}$ for $k \in \{0,1\}$, assume all $\mu_i^k > 0$.
  Consider $\psi^k \in \mathbb{R}^Y$ s.t. $G(\psi^k) = \mu^k$, and $\psi^t = \psi^0 + tv$ with $v = \psi^1 - \psi^0$. Then,

Proof ingredients

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu,\nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

- **Strategy of proof**: let $\mu^k = \sum_i \mu_i^k \delta_{y_i}$ for $k \in \{0,1\}$, assume all $\mu_i^k > 0$.
  Consider $\psi^k \in \mathbb{R}^Y$ s.t. $G(\psi^k) = \mu^k$, and $\psi^t = \psi^0 + tv$ with $v = \psi^1 - \psi^0$. Then,
  $$\langle \mu^1 - \mu^0 | v\rangle = \langle G(\psi^1) - G(\psi^0) | v\rangle = \int_0^1 \langle \mathrm{D}G(\psi^t)v|v\rangle \,\mathrm{d}\, t$$

Proof ingredients

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu,\nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

- **Strategy of proof**: let $\mu^k = \sum_i \mu_i^k \delta_{y_i}$ for $k \in \{0,1\}$, assume all $\mu_i^k > 0$.
  Consider $\psi^k \in \mathbb{R}^Y$ s.t. $G(\psi^k) = \mu^k$, and $\psi^t = \psi^0 + tv$ with $v = \psi^1 - \psi^0$. Then,

$$\langle \mu^1 - \mu^0 | v\rangle = \langle G(\psi^1) - G(\psi^0)|v\rangle = \int_0^1 \langle \mathrm{D}G(\psi^t)v|v\rangle \,\mathrm{d}\,t$$

a) **Control of the eigengap**: $\langle \mathrm{D}G(\psi^t)v|v\rangle \leq -C(X)\|v\|^2_{\mathrm{L}^2(\mu_t)}$ if $\int v\,\mathrm{d}\,\mu_t = 0$.
with $\mu^t = G(\psi^t) \longrightarrow$ [Eymard, Gallouët, Herbin '00].

Proof ingredients

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu,\nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

- **Strategy of proof**: let $\mu^k = \sum_i \mu_i^k \delta_{y_i}$ for $k \in \{0,1\}$, assume all $\mu_i^k > 0$.
  Consider $\psi^k \in \mathbb{R}^Y$ s.t. $G(\psi^k) = \mu^k$, and $\psi^t = \psi^0 + tv$ with $v = \psi^1 - \psi^0$. Then,

$$\langle \mu^1 - \mu^0 | v\rangle = \langle G(\psi^1) - G(\psi^0)|v\rangle = \int_0^1 \langle \mathrm{D}G(\psi^t)v|v\rangle \,\mathrm{d}\,t$$

  a) **Control of the eigengap**: $\langle \mathrm{D}G(\psi^t)v|v\rangle \leq -C(X)\|v\|^2_{\mathrm{L}^2(\mu_t)}$ if $\int v \,\mathrm{d}\,\mu_t = 0$.
  with $\mu^t = G(\psi^t) \longrightarrow$ [Eymard, Gallouët, Herbin '00].

  b) **Control of μ_t**: Brunn-Minkowski's inequality implies $\mu^t \geq (1-t)^d \mu^0$.

Proof ingredients

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu,\nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

- **Strategy of proof**: let $\mu^k = \sum_i \mu_i^k \delta_{y_i}$ for $k \in \{0,1\}$, assume all $\mu_i^k > 0$.
  Consider $\psi^k \in \mathbb{R}^Y$ s.t. $G(\psi^k) = \mu^k$, and $\psi^t = \psi^0 + tv$ with $v = \psi^1 - \psi^0$. Then,

$$\langle \mu^1 - \mu^0 | v\rangle = \langle G(\psi^1) - G(\psi^0) | v \rangle = \int_0^1 \langle \mathrm{D}G(\psi^t) v | v\rangle \,\mathrm{d}\, t$$

  a) **Control of the eigengap**: $\langle \mathrm{D}G(\psi^t) v | v\rangle \leq -C(X) \|v\|^2_{\mathrm{L}^2(\mu_t)}$ if $\int v \,\mathrm{d}\, \mu_t = 0$.
  with $\mu^t = G(\psi^t) \longrightarrow$ [Eymard, Gallouët, Herbin '00].

  b) **Control of μ_t**: Brunn-Minkowski's inequality implies $\mu^t \geq (1-t)^d \mu^0$.

Combining a) and b) we get $\|\psi^1 - \psi^0\|^2_{\mathrm{L}^2(\mu^0)} \lesssim |\langle \mu^1 - \mu^0 | \psi^1 - \psi^0 \rangle|$

Proof ingredients

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu,\nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

- **Strategy of proof**: let $\mu^k = \sum_i \mu_i^k \delta_{y_i}$ for $k \in \{0,1\}$, assume all $\mu_i^k > 0$.
  Consider $\psi^k \in \mathbb{R}^Y$ s.t. $G(\psi^k) = \mu^k$, and $\psi^t = \psi^0 + tv$ with $v = \psi^1 - \psi^0$. Then,

$$\langle \mu^1 - \mu^0 | v\rangle = \langle G(\psi^1) - G(\psi^0) | v\rangle = \int_0^1 \langle \mathrm{D}G(\psi^t) v | v\rangle \,\mathrm{d}\, t$$

  a) **Control of the eigengap**: $\langle \mathrm{D}G(\psi^t) v | v\rangle \leq -C(X)\|v\|^2_{\mathrm{L}^2(\mu_t)}$ if $\int v \,\mathrm{d}\,\mu_t = 0$.
  with $\mu^t = G(\psi^t) \longrightarrow$ [Eymard, Gallouët, Herbin '00].

  b) **Control of μ_t**: Brunn-Minkowski's inequality implies $\mu^t \geq (1-t)^d \mu^0$.

Combining a) and b) we get $\|\psi^1 - \psi^0\|^2_{\mathrm{L}^2(\mu^0)} \lesssim |\langle \mu^1 - \mu^0 | \psi^1 - \psi^0\rangle|$

(Kantorovich-Rubinstein) $\leq \mathrm{Lip}(\psi^1 - \psi^0)\, \mathrm{W}_1(\mu^0, \mu_1)$

Proof ingredients

Thm: Let X be convex compact and ρ a density on X with $\|\log(\rho)\|_\infty < +\infty$
Let Y be compact. Then, $\forall \mu, \nu \in \mathrm{Prob}(Y)$, $\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu, \nu)^{1/6}$.

[M., Delalande, Chazal '19; Delalande, M. '21]

- **Strategy of proof**: let $\mu^k = \sum_i \mu_i^k \delta_{y_i}$ for $k \in \{0, 1\}$, assume all $\mu_i^k > 0$.
  Consider $\psi^k \in \mathbb{R}^Y$ s.t. $G(\psi^k) = \mu^k$, and $\psi^t = \psi^0 + tv$ with $v = \psi^1 - \psi^0$. Then,

$$\langle \mu^1 - \mu^0 | v \rangle = \langle G(\psi^1) - G(\psi^0) | v \rangle = \int_0^1 \langle \mathrm{D}G(\psi^t) v | v \rangle \, \mathrm{d}\, t$$

a) **Control of the eigengap**: $\langle \mathrm{D}G(\psi^t) v | v \rangle \leq -C(X) \|v\|^2_{\mathrm{L}^2(\mu_t)}$ if $\int v \, \mathrm{d}\, \mu_t = 0$.
with $\mu^t = G(\psi^t) \longrightarrow$ [Eymard, Gallouët, Herbin '00].

b) **Control of μ_t**: Brunn-Minkowski's inequality implies $\mu^t \geq (1-t)^d \mu^0$.

Combining a) and b) we get $\|\psi^1 - \psi^0\|^2_{\mathrm{L}^2(\mu^0)} \lesssim |\langle \mu^1 - \mu^0 | \psi^1 - \psi^0 \rangle|$

(Kantorovich-Rubinstein) $\leq \mathrm{Lip}(\psi^1 - \psi^0)\, \mathrm{W}_1(\mu^0, \mu_1)$

$\lesssim \mathrm{W}_2(\mu^0, \mu^1)$

- We lose a little in the exponent to control the difference between OT maps...

The non-compact space

Theorem Let $\mu^0, \mu^1 \in \mathrm{Prob}_2(\mathbb{R}^d)$, ϕ^k the Brenier potential from $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ to μ^k, where X is convex and ρ is bounded from above and below. Assume that

(i) $\forall k \in \{0, 1\}, \forall x, y \in X, |\phi^k(x) - \phi^k(y)| \leq C_H \|x - y\|^\alpha$

(ii) $M_4(\mu^k) \leq M$.

Then, $\mathrm{W}_2(\mu^0, \mu^1) \leq \|T_{\mu^1} - T_{\mu^0}\|_{\mathrm{L}^2(\rho)} \leq C(d, X, \rho, C_H, M)\, \mathrm{W}_1(\mu^0, \mu^1)^{\frac{1}{2(11-8\alpha)}}$.

[Delalande, M. 2021]

The non-compact space

Theorem Let $\mu^0, \mu^1 \in \mathrm{Prob}_2(\mathbb{R}^d)$, ϕ^k the Brenier potential from $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ to μ^k, where X is convex and ρ is bounded from above and below. Assume that

(i) $\forall k \in \{0,1\}, \forall x, y \in X, |\phi^k(x) - \phi^k(y)| \leq C_H \|x - y\|^\alpha$

(ii) $M_4(\mu^k) \leq M$.

Then, $\mathrm{W}_2(\mu^0, \mu^1) \leq \|T_{\mu^1} - T_{\mu^0}\|_{\mathrm{L}^2(\rho)} \leq C(d, X, \rho, C_H, M)\, \mathrm{W}_1(\mu^0, \mu^1)^{\frac{1}{2(11-8\alpha)}}$.

[Delalande, M. 2021]

- When $\alpha = 1$, we recover the exponent of the compact case: $2(11 - 8\alpha) = 6$.
- By Morrey's inequality, (i) holds when $M_p(\mu^k) < +\infty$ for $p > d$. Indeed,

$$M_p(\mu^k) = \int \|y\|^p \,\mathrm{d}\,\mu^k(y) = \int \|\nabla\phi^k\|^p \,\mathrm{d}\,\rho < +\infty \Longrightarrow \phi^k \in W^{1,p}(X) \subseteq \mathcal{C}^{1-\frac{d}{p}}(X)$$

The non-compact space

Theorem Let $\mu^0, \mu^1 \in \mathrm{Prob}_2(\mathbb{R}^d)$, ϕ^k the Brenier potential from $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ to μ^k, where X is convex and ρ is bounded from above and below. Assume that

(i) $\forall k \in \{0,1\}, \forall x, y \in X, |\phi^k(x) - \phi^k(y)| \leq C_H \|x - y\|^\alpha$

(ii) $M_4(\mu^k) \leq M$.

Then, $\mathrm{W}_2(\mu^0, \mu^1) \leq \|T_{\mu^1} - T_{\mu^0}\|_{\mathrm{L}^2(\rho)} \leq C(d, X, \rho, C_H, M)\, \mathrm{W}_1(\mu^0, \mu^1)^{\frac{1}{2(11-8\alpha)}}$.

[Delalande, M. 2021]

- When $\alpha = 1$, we recover the exponent of the compact case: $2(11 - 8\alpha) = 6$.
- By Morrey's inequality, (i) holds when $M_p(\mu^k) < +\infty$ for $p > d$. Indeed,

$$M_p(\mu^k) = \int \|y\|^p \,\mathrm{d}\,\mu^k(y) = \int \|\nabla \phi^k\|^p \,\mathrm{d}\,\rho < +\infty \Longrightarrow \phi^k \in W^{1,p}(X) \subseteq \mathcal{C}^{1-\frac{d}{p}}(X)$$

In particular, this result applies to sub-exponential or sub-Gaussian measures.

The non-compact space

Theorem Let $\mu^0, \mu^1 \in \mathrm{Prob}_2(\mathbb{R}^d)$, ϕ^k the Brenier potential from $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ to μ^k, where X is convex and ρ is bounded from above and below. Assume that

(i) $\forall k \in \{0,1\}, \forall x, y \in X, |\phi^k(x) - \phi^k(y)| \leq C_H \|x - y\|^\alpha$

(ii) $M_4(\mu^k) \leq M$.

Then, $\mathrm{W}_2(\mu^0, \mu^1) \leq \|T_{\mu^1} - T_{\mu^0}\|_{\mathrm{L}^2(\rho)} \leq C(d, X, \rho, C_H, M)\, \mathrm{W}_1(\mu^0, \mu^1)^{\frac{1}{2(11-8\alpha)}}$.

[Delalande, M. 2021]

- When $\alpha = 1$, we recover the exponent of the compact case: $2(11 - 8\alpha) = 6$.
- By Morrey's inequality, (i) holds when $M_p(\mu^k) < +\infty$ for $p > d$. Indeed,

$$M_p(\mu^k) = \int \|y\|^p \,\mathrm{d}\,\mu^k(y) = \int \|\nabla \phi^k\|^p \,\mathrm{d}\,\rho < +\infty \Longrightarrow \phi^k \in W^{1,p}(X) \subseteq \mathcal{C}^{1-\frac{d}{p}}(X)$$

  In particular, this result applies to sub-exponential or sub-Gaussian measures.

- By [Andoni, Naor, Neiman '18], the space $(\mathrm{Prob}_2(\mathbb{R}^d), W_2)$ does not admit a bi-Hölder embedding into any L^p space when $d \geq 3$.

Summary

Optimal transport plans can be used to embed of $\mathrm{Prob}_2(\mathbb{R}^d)$ into $\mathrm{L}^2(\rho, \mathbb{R}^d)$, while preserving some of its metric geometry, with applications in data analysis.

`https://github.com/sd-ot`

Summary

Optimal transport plans can be used to embed of $\mathrm{Prob}_2(\mathbb{R}^d)$ into $\mathrm{L}^2(\rho, \mathbb{R}^d)$, while preserving some of its metric geometry, with applications in data analysis.

```
https://github.com/sd-ot
```

Open questions/current work:

- optimal Hölder exponent for $\mu \mapsto T_\mu$ in the compact case?
- what happens for other cost functions?
- is there a bi-Hölder embedding of $\{\mu \in \mathrm{Prob}_2(\mathbb{R}^d) \mid M_2(\mu) \leq R\}$ into $\mathrm{L}^2(\rho)$?

Thank you for your attention!