Convergence rates for discretized optimal transport

Quentin Mérigot

Université Paris-Sud 11

Based on joint work with F. Chazal and A. Delalande

Workshop on numerical solutions of HJB equations, Paris, January 2020

1. Motivations

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 $[0,1]$.

NB: $T_{\mu\#}\lambda = \mu \iff \forall B \subseteq \mathbb{R},\ \lambda(T_\mu^{-1}(B)) = \mu(B)$

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

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 $[0,1]$.

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

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

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 $[0,1]$.

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

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

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 $[0,1]$.

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

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

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.

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 $[0,1]$.

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

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

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 $[0,1]$.

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

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

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 $[0,1]$.

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

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

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]

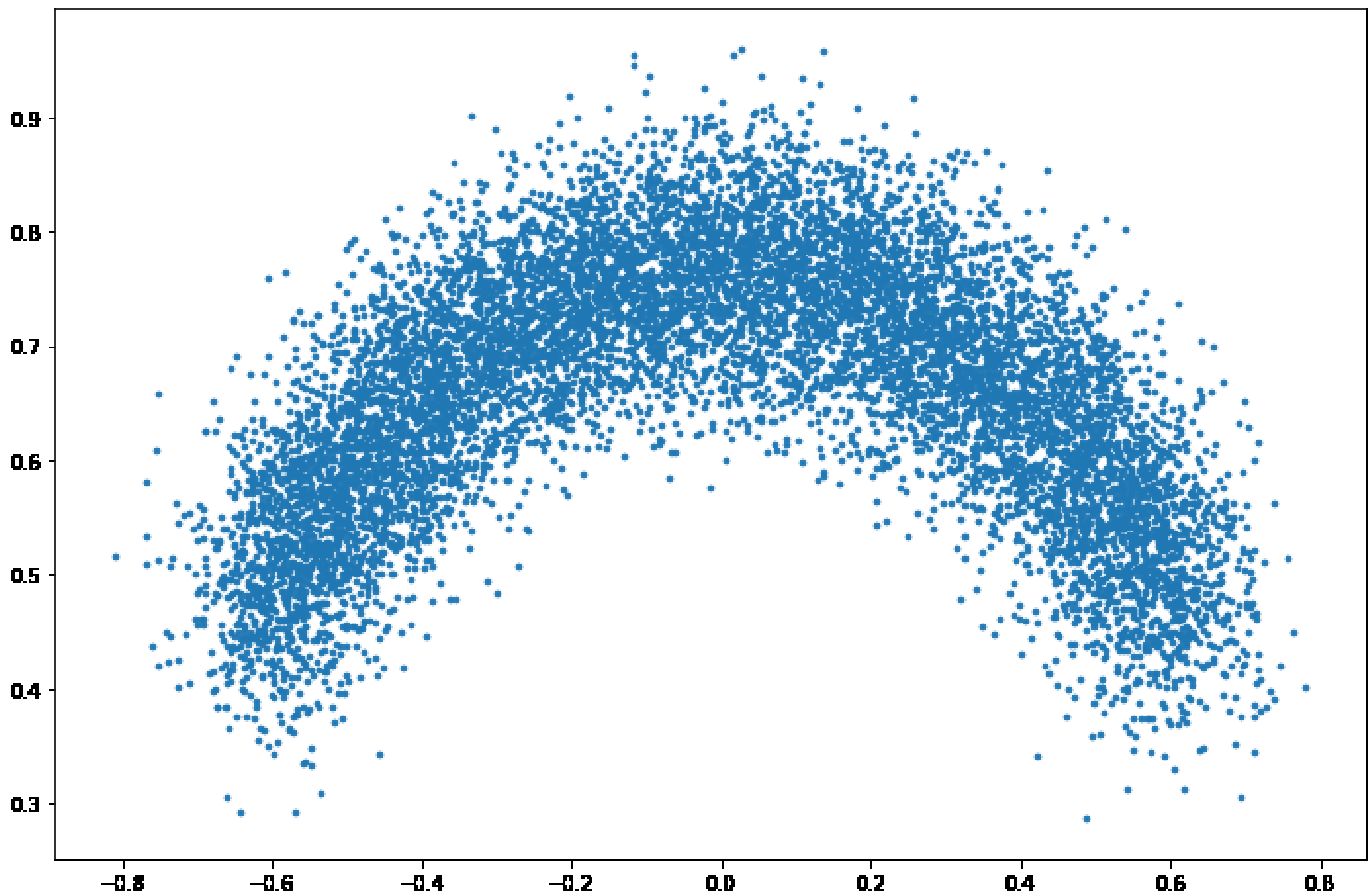
- T_μ is unique ρ-a.e. but the convex function ϕ_μ is not necessarily unique.

- $T_\mu : \mathrm{spt}(\rho) \to \mathbb{R}^d$ 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



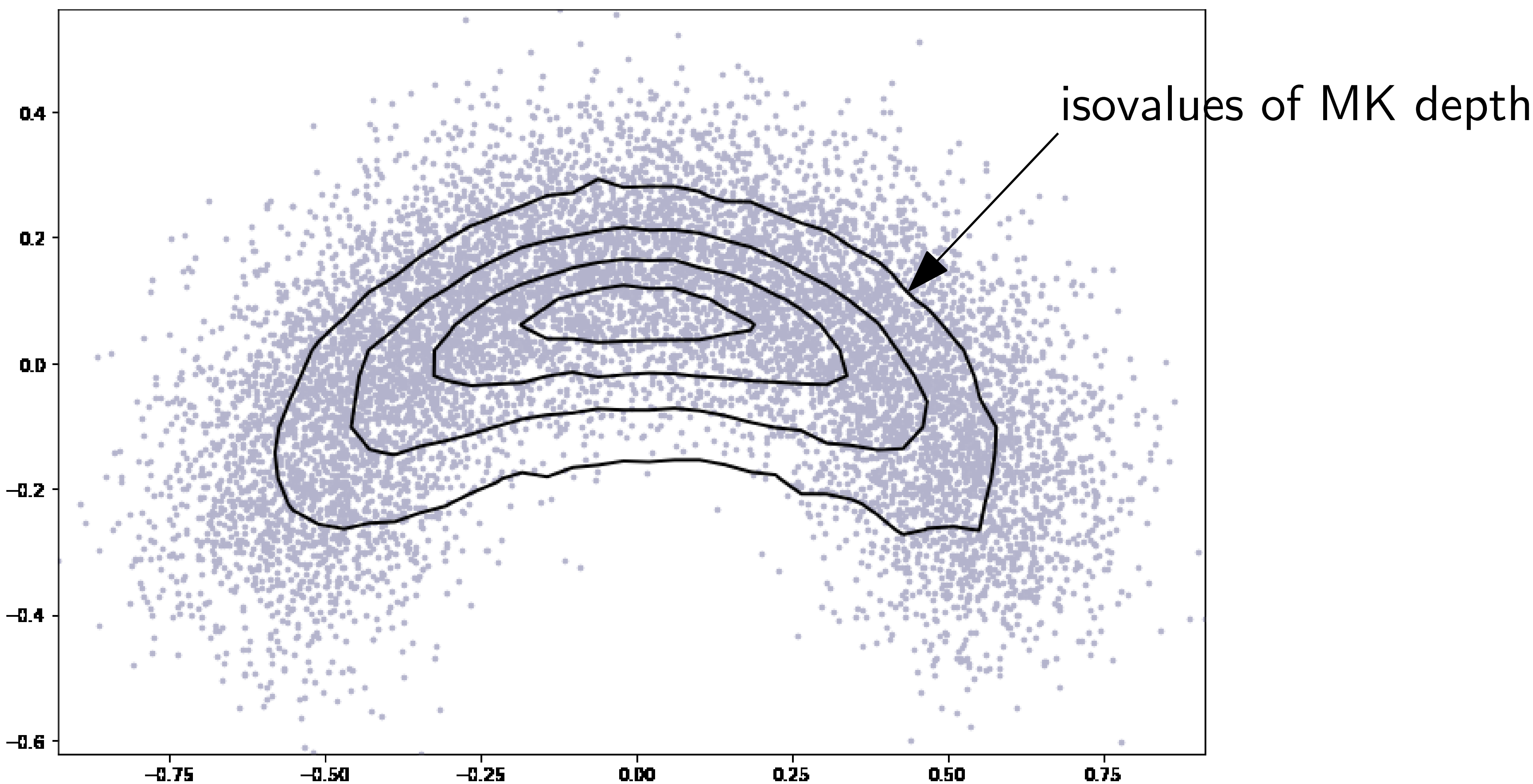
"Monge-Kantorovich depth of y_i" $\simeq \|T_\mu^{-1}(y_i)\|$.

[Cherzonukov, Galichon, Hallin, Henry]

Numerical Example: Monge-Kantorovich Depth

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

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



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

[Cherzonukov, Galichon, Hallin, Henry]

Wasserstein space

- 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 space

- 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 $\nu \subseteq \mathrm{Prob}(\mathbb{R}^d \times \mathbb{R}^d)$.

- On $\mathrm{Prob}(X)$, with $X \subseteq \mathbb{R}^d$ compact, W_p metrizes **narrow convergence**
  i.e. $\lim_{n\to+\infty} \mathrm{W}_p(\mu_n,\mu) = 0 \Longleftrightarrow \forall\phi \in \mathcal{C}^0(X), \lim_{n\to+\infty} \int \phi \,\mathrm{d}\,\mu_n = \int \phi \,\mathrm{d}\,\mu$.

Wasserstein space

- 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 $\nu \subseteq \mathrm{Prob}(\mathbb{R}^d \times \mathbb{R}^d)$.

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

  i.e. $\lim_{n\to+\infty} \mathrm{W}_p(\mu_n,\mu) = 0 \Longleftrightarrow \forall \phi \in \mathcal{C}^0(X), \lim_{n\to+\infty} \int \phi \,\mathrm{d}\,\mu_n = \int \phi \,\mathrm{d}\,\mu$.

- On $\mathrm{Prob}(\mathbb{R})$, any *monotone* coupling γ between μ, ν is optimal in the def of W_p.

Wasserstein space

- 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 $\nu \subseteq \mathrm{Prob}(\mathbb{R}^d \times \mathbb{R}^d)$.

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

  i.e. $\lim_{n\to+\infty} \mathrm{W}_p(\mu_n, \mu) = 0 \iff \forall \phi \in \mathcal{C}^0(X), \lim_{n\to+\infty} \int \phi \,\mathrm{d}\,\mu_n = \int \phi \,\mathrm{d}\,\mu$.

- On $\mathrm{Prob}(\mathbb{R})$, any *monotone* coupling γ between μ, ν is optimal in the def of W_p.

  For instance $\gamma := (T_\mu, T_\nu)_\# \rho$ with $\rho =$ Lebesgue on $[0, 1]$ is monotone, implying

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

Wasserstein space

- 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 $\nu \subseteq \mathrm{Prob}(\mathbb{R}^d \times \mathbb{R}^d)$.

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

  i.e. $\lim_{n\to+\infty} \mathrm{W}_p(\mu_n,\mu) = 0 \iff \forall \phi \in \mathcal{C}^0(X), \lim_{n\to+\infty} \int \phi \,\mathrm{d}\,\mu_n = \int \phi \,\mathrm{d}\,\mu$.

- On $\mathrm{Prob}(\mathbb{R})$, any *monotone* coupling γ between μ, ν is optimal in the def of W_p.

  For instance $\gamma := (T_\mu, T_\nu)_{\#}\rho$ with ρ = Lebesgue on $[0,1]$ is monotone, implying

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

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

Wasserstein space

- 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 $\nu \subseteq \mathrm{Prob}(\mathbb{R}^d \times \mathbb{R}^d)$.

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

  i.e. $\lim_{n\to+\infty} \mathrm{W}_p(\mu_n,\mu) = 0 \iff \forall \phi \in \mathcal{C}^0(X), \lim_{n\to+\infty} \int \phi \,\mathrm{d}\,\mu_n = \int \phi \,\mathrm{d}\,\mu$.

- On $\mathrm{Prob}(\mathbb{R})$, any *monotone* coupling γ between μ, ν is optimal in the def of W_p.

  For instance $\gamma := (T_\mu, T_\nu)_{\#}\rho$ with ρ = Lebesgue on $[0,1]$ is monotone, implying

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

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

The previous embedding is false in higher dimension: $(\mathrm{Prob}_p, \mathrm{W}_p)$ is *curved*.

Motivation 2: "Linearization" of W_2

Motivation 2: "Linearization" of W_2

- We fix a reference measure, $\rho = \mathrm{Leb}_X$ with $X \subseteq \mathbb{R}^d$ convex compact with $|X| = 1$.

Given $\mu \in \mathrm{Prob}_2(\mathbb{R}^d)$, we define T_μ as the unique map satisfying

(i) $T_\mu = \nabla\phi_\mu$ a.e. for some convex function $\phi_\mu : X \to \mathbb{R}$ and

(ii) $T_{\mu\#}\rho = \mu$.

Motivation 2: "Linearization" of W_2

- We fix a reference measure, $\rho = \mathrm{Leb}_X$ with $X \subseteq \mathbb{R}^d$ convex compact with $|X| = 1$.

Given $\mu \in \mathrm{Prob}_2(\mathbb{R}^d)$, we define T_μ as the unique map satisfying

(i) $T_\mu = \nabla \phi_\mu$ a.e. for some convex function $\phi_\mu : X \to \mathbb{R}$ and

(ii) $T_{\mu\#}\rho = \mu$.

- 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.

Motivation 2: "Linearization" of W_2

- We fix a reference measure, $\rho = \mathrm{Leb}_X$ with $X \subseteq \mathbb{R}^d$ convex compact with $|X| = 1$.

Given $\mu \in \mathrm{Prob}_2(\mathbb{R}^d)$, we define T_μ as the unique map satisfying

(i) $T_\mu = \nabla\phi_\mu$ a.e. for some convex function $\phi_\mu : X \to \mathbb{R}$ and

(ii) $T_{\mu\#}\rho = \mu$.

- 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.
  - $\mathrm{W}_{2,\rho}(\mu,\nu) := \|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \longrightarrow$ [Ambrosio, Gigli, Savaré '04]

	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	$\mathrm{T}_\rho M$	$\mathrm{T}_\rho\mathrm{Prob}_2(\mathbb{R}^d) \subseteq \mathrm{L}^2(\rho, X)$
inverse exponential map	$\exp_\rho^{-1}(x) \in \mathrm{T}_\rho M$	$T_\mu \in \mathrm{T}_\rho\mathrm{Prob}_2(X)$
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: "Linearization" of W_2

- We fix a reference measure, $\rho = \mathrm{Leb}_X$ with $X \subseteq \mathbb{R}^d$ convex compact with $|X| = 1$.

Given $\mu \in \mathrm{Prob}_2(\mathbb{R}^d)$, we define T_μ as the unique map satisfying

(i) $T_\mu = \nabla \phi_\mu$ a.e. for some convex function $\phi_\mu : X \to \mathbb{R}$ and

(ii) $T_{\mu\#}\rho = \mu$.

- 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.
  - $\mathrm{W}_{2,\rho}(\mu, \nu) := \|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \longrightarrow$ [Ambrosio, Gigli, Savaré '04]

	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	$\mathrm{T}_\rho M$	$\mathrm{T}_\rho \mathrm{Prob}_2(\mathbb{R}^d) \subseteq \mathrm{L}^2(\rho, X)$
inverse exponential map	$\exp_\rho^{-1}(x) \in \mathrm{T}_\rho M$	$T_\mu \in \mathrm{T}_\rho \mathrm{Prob}_2(X)$
distance in tangent space	$\\| \exp_\rho^{-1}(x) - \exp_\rho^{-1}(y) \\|_{g(x_0)}$	$\\|T_\mu - T_\nu\\|_{\mathrm{L}^2(\rho)}$

  - Used in image analysis $\longrightarrow$ [Wang, Slepcev, Basu, Ozolek, Rohde '13]

Motivation 2: "Linearization" of W_2

- We fix a reference measure, $\rho = \mathrm{Leb}_X$ with $X \subseteq \mathbb{R}^d$ convex compact with $|X| = 1$.

Given $\mu \in \mathrm{Prob}_2(\mathbb{R}^d)$, we define T_μ as the unique map satisfying

(i) $T_\mu = \nabla\phi_\mu$ a.e. for some convex function $\phi_\mu : X \to \mathbb{R}$ and

(ii) $T_{\mu\#}\rho = \mu$.

- 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.
  - $\mathrm{W}_{2,\rho}(\mu,\nu) := \|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \longrightarrow$ [Ambrosio, Gigli, Savaré '04]

	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	$\mathrm{T}_\rho M$	$\mathrm{T}_\rho \mathrm{Prob}_2(\mathbb{R}^d) \subseteq \mathrm{L}^2(\rho, X)$
inverse exponential map	$\exp_\rho^{-1}(x) \in \mathrm{T}_\rho M$	$T_\mu \in \mathrm{T}_\rho \mathrm{Prob}_2(X)$
distance in tangent space	$\|\exp_\rho^{-1}(x) - \exp_\rho^{-1}(y)\|_{g(x_0)}$	$\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)}$

  - Used in image analysis $\longrightarrow$ [Wang, Slepcev, Basu, Ozolek, Rohde '13]

$\longrightarrow$ Representing family of probability measures by family of functions in $\mathrm{L}^2(\rho)$.

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

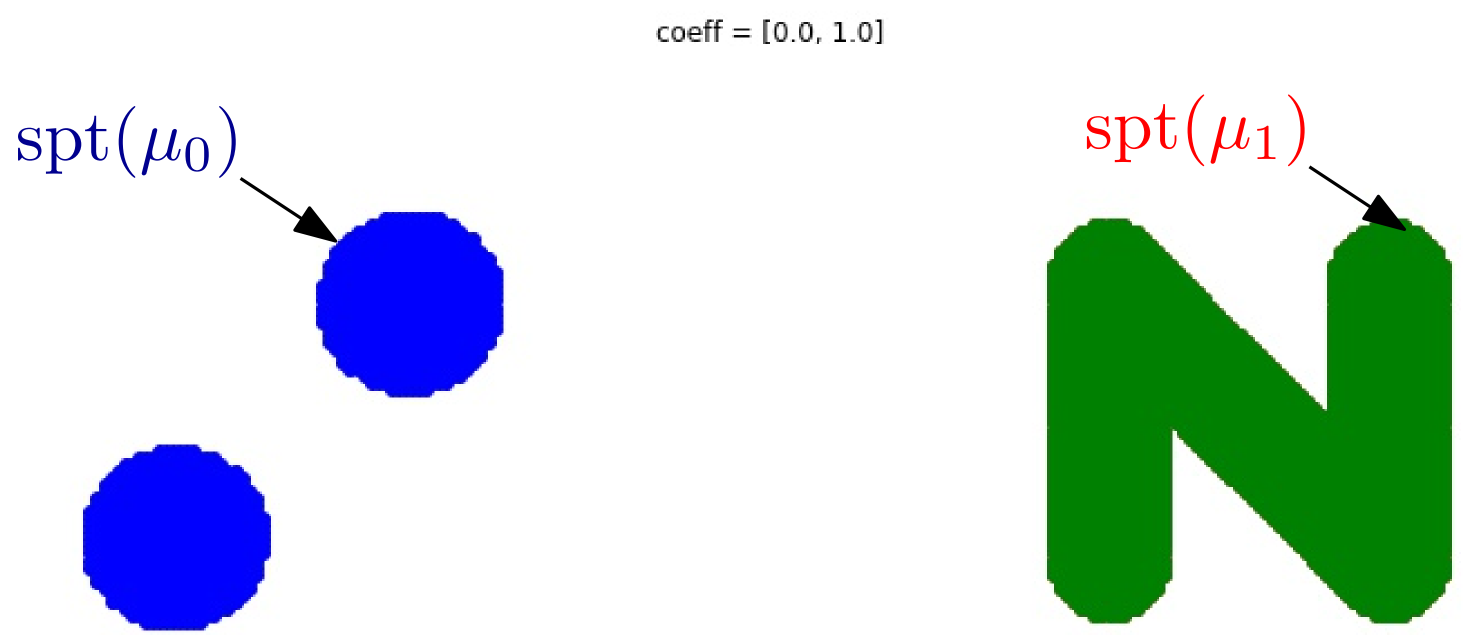
- **Barycenter in Wasserstein space:** $\mu_1, \ldots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \ldots, \alpha_k \geq 0$:

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

$\longrightarrow$ 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

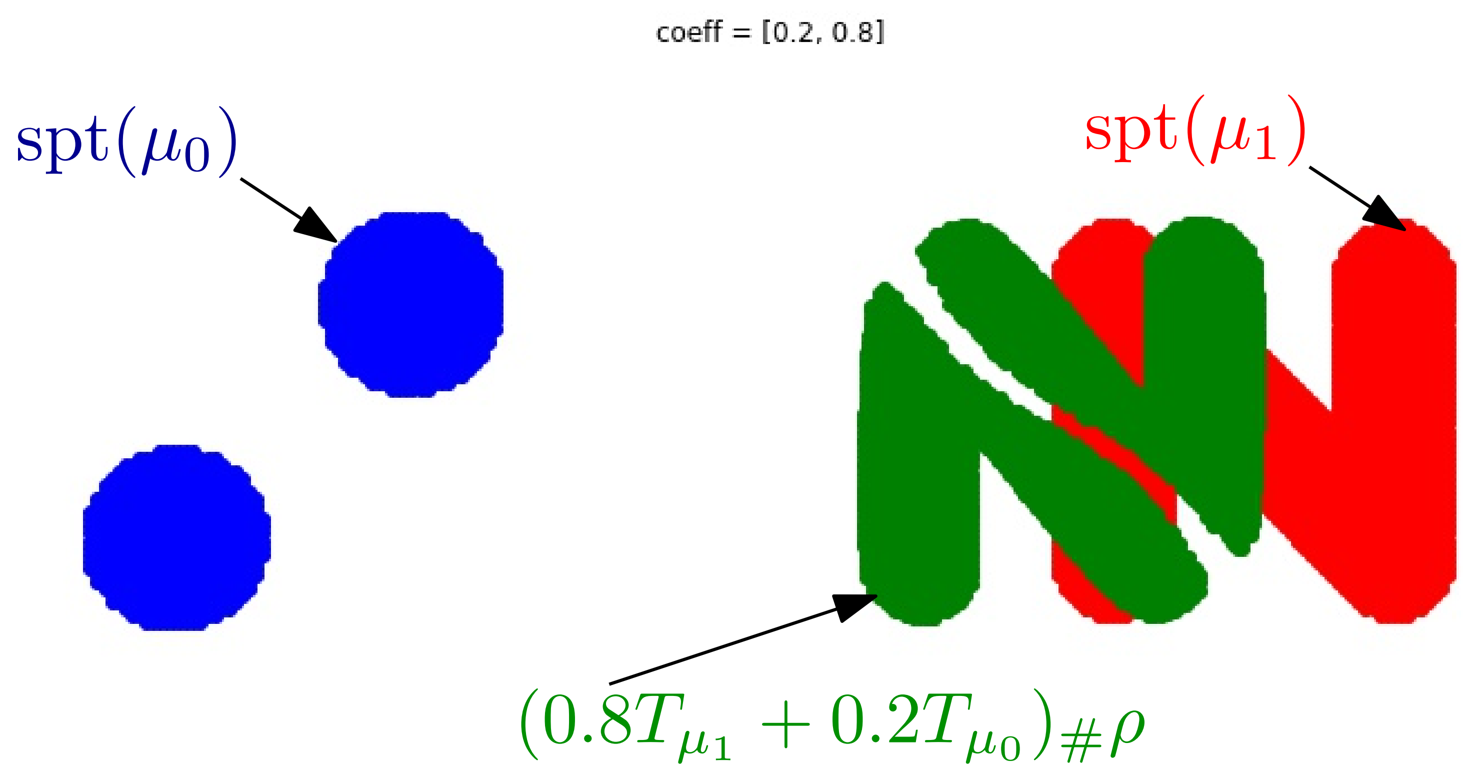
- **Barycenter in Wasserstein space:** $\mu_1, \dots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \dots, \alpha_k \geq 0$:

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

$\longrightarrow$ 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.



Example: barycenter computation

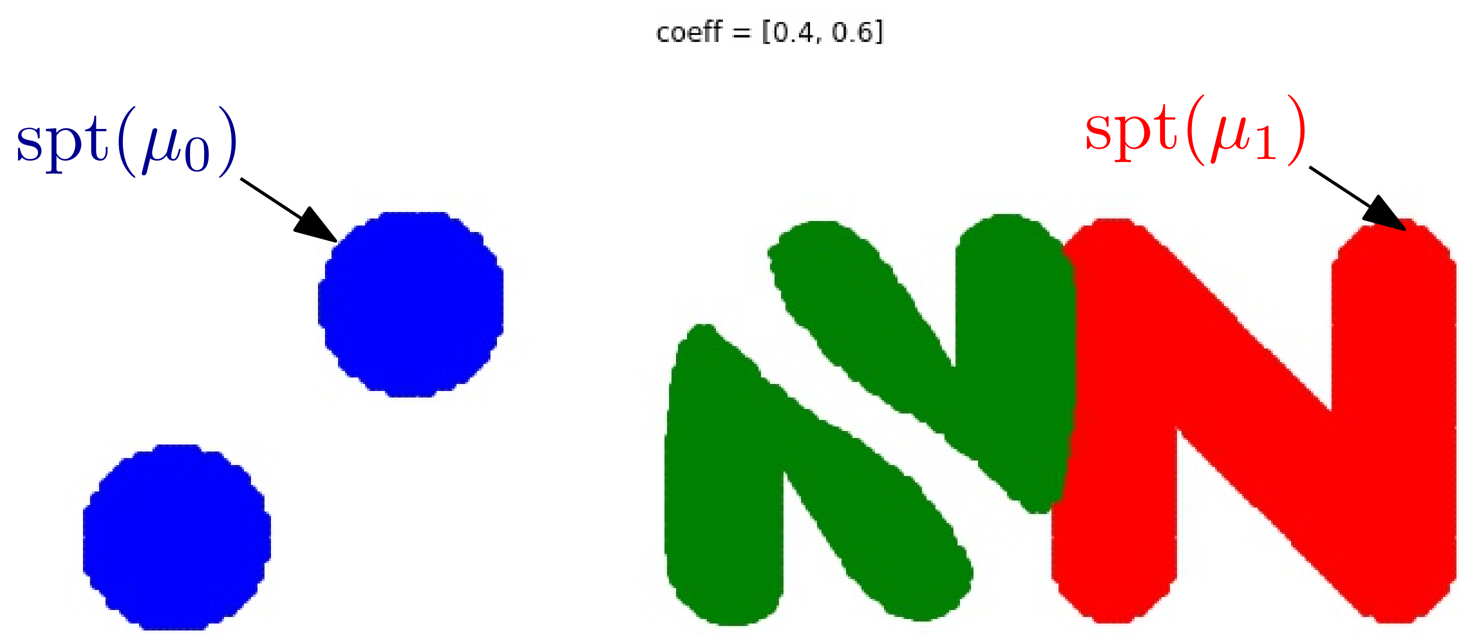
- **Barycenter in Wasserstein space:** $\mu_1, \dots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \dots, \alpha_k \geq 0$:

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

$\longrightarrow$ 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.



Example: barycenter computation

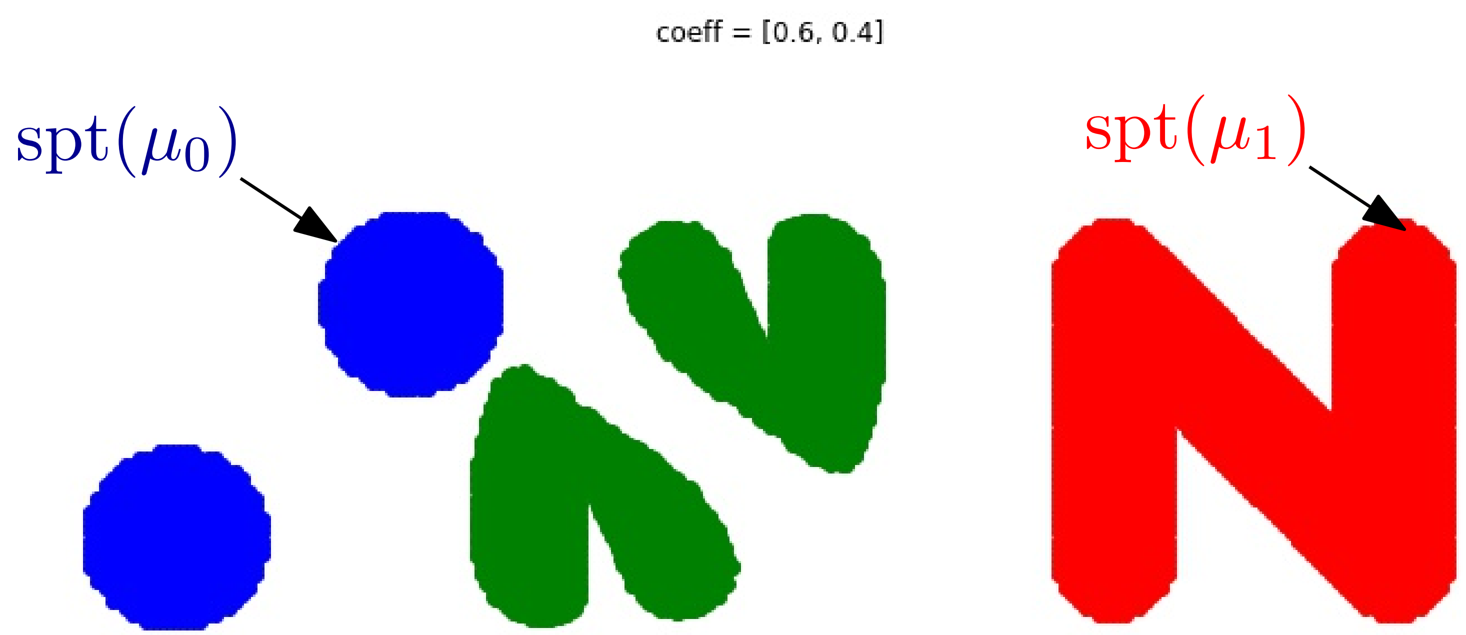
- **Barycenter in Wasserstein space:** $\mu_1, \ldots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \ldots, \alpha_k \geq 0$:

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

$\longrightarrow$ 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

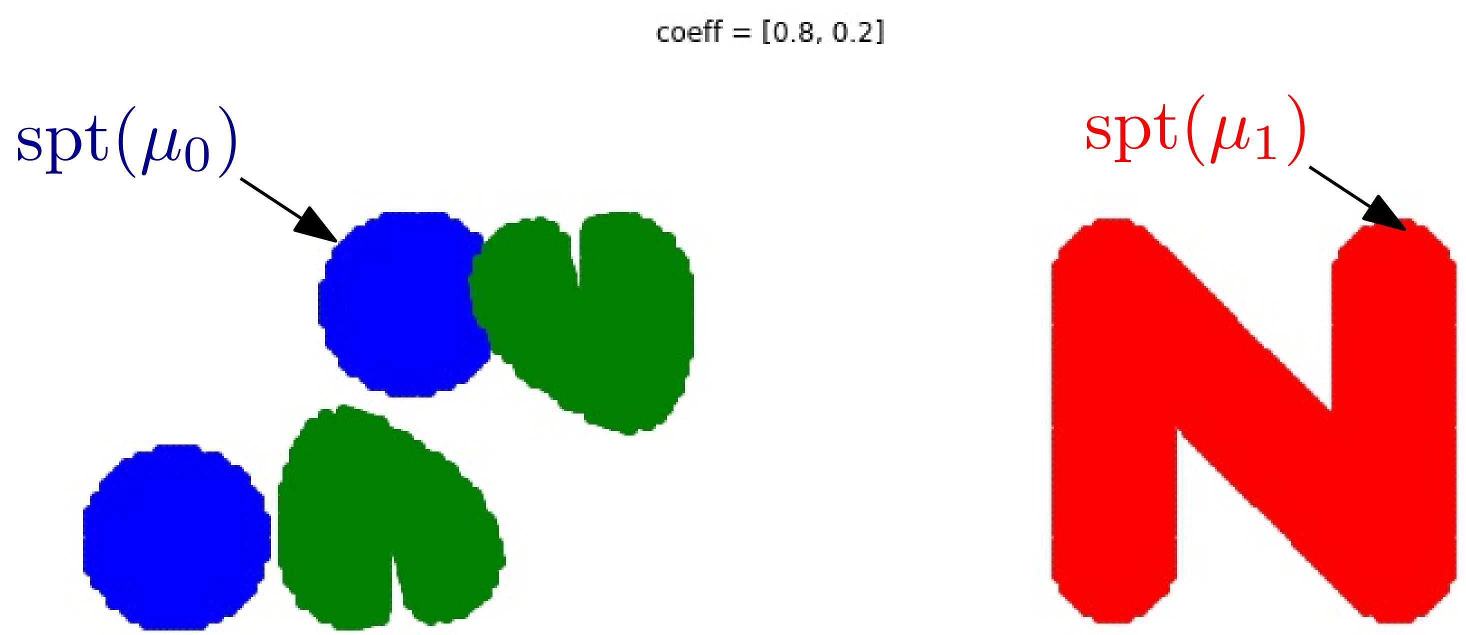
- **Barycenter in Wasserstein space:** $\mu_1, \dots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \dots, \alpha_k \geq 0$:

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

$\longrightarrow$ 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.



Example: barycenter computation

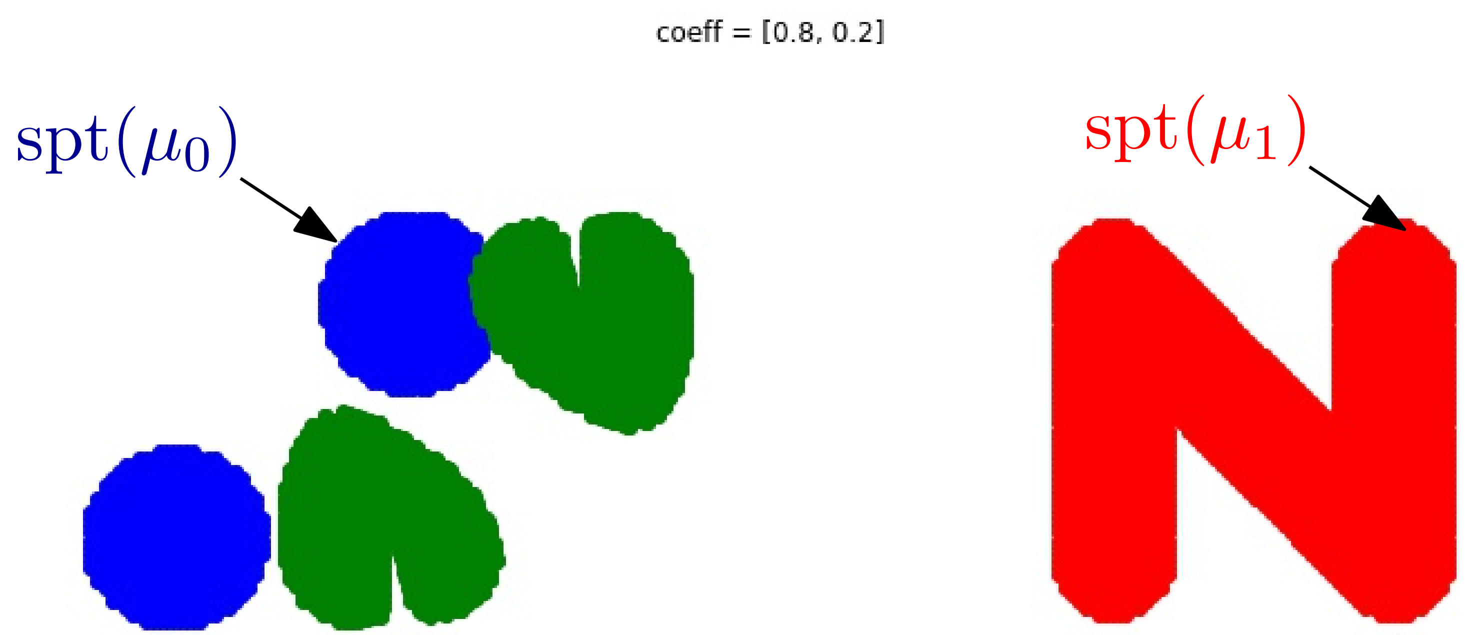
- **Barycenter in Wasserstein space:** $\mu_1, \ldots, \mu_k \in \mathrm{Prob}_2(\mathbb{R}^d)$, $\alpha_1, \ldots, \alpha_k \geq 0$:

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

$\longrightarrow$ 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.



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

Motivation 3: numerical analysis of optimal transport

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.

To solve numerically an OT problem between $\rho \in \mathrm{Prob}^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mu \in \mathrm{Prob}([0,1]^d)$:

- Approximate μ by a discrete measure, for instance

  $$\mu_k = \sum_{i_1 \leq \ldots \leq i_k} \mu(B_{i_1,\ldots,i_k})\delta_{(i_1/k,\ldots,i_k/k)}$$

  where $B_{i_1,\ldots,i_k}$ is the cube $[(i_1-1)/k, i_1/k] \times \ldots [(i_d-1)/k, i_d/k]$

Motivation 3: numerical analysis of optimal transport

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.

To solve numerically an OT problem between $\rho \in \mathrm{Prob}^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mu \in \mathrm{Prob}([0,1]^d)$:

- Approximate μ by a discrete measure, for instance

$$\mu_k = \sum_{i_1 \leq \ldots \leq i_k} \mu(B_{i_1,\ldots,i_k})\delta_{(i_1/k,\ldots,i_k/k)}$$

where $B_{i_1,\ldots,i_k}$ is the cube $[(i_1-1)/k, i_1/k] \times \ldots [(i_d-1)/k, i_d/k]$

(Then, $\mathrm{W}_p(\mu_k,\mu) \lesssim \frac{1}{k}$.)

Motivation 3: numerical analysis of optimal transport

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.

To solve numerically an OT problem between $\rho \in \mathrm{Prob}^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mu \in \mathrm{Prob}([0,1]^d)$:

- Approximate μ by a discrete measure, for instance

$$\mu_k = \sum_{i_1 \leq \ldots \leq i_k} \mu(B_{i_1,\ldots,i_k})\delta_{(i_1/k,\ldots,i_k/k)}$$

where $B_{i_1,\ldots,i_k}$ is the cube $[(i_1-1)/k, i_1/k] \times \ldots [(i_d-1)/k, i_d/k]$

(Then, $\mathrm{W}_p(\mu_k,\mu) \lesssim \frac{1}{k}$.)

- Compute *exactly* the optimal transport plan T_{μ_k} between ρ and μ_k,
(using a **semi-discrete** optimal transport solver).

Motivation 3: numerical analysis of optimal transport

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.

To solve numerically an OT problem between $\rho \in \mathrm{Prob}^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mu \in \mathrm{Prob}([0,1]^d)$:

- Approximate μ by a discrete measure, for instance

  $$\mu_k = \sum_{i_1 \leq \ldots \leq i_k} \mu(B_{i_1,\ldots,i_k})\delta_{(i_1/k,\ldots,i_k/k)}$$

  where $B_{i_1,\ldots,i_k}$ is the cube $[(i_1-1)/k, i_1/k] \times \ldots [(i_d-1)/k, i_d/k]$

  (Then, $\mathrm{W}_p(\mu_k,\mu) \lesssim \frac{1}{k}$.)

- Compute *exactly* the optimal transport plan T_{μ_k} between ρ and μ_k, (using a **semi-discrete** optimal transport solver).

It is know that T_{μ_k} converges to T_μ but convergence rates are unknown in general...

Motivation 3: numerical analysis of optimal transport

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.

To solve numerically an OT problem between $\rho \in \mathrm{Prob}^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mu \in \mathrm{Prob}([0,1]^d)$:

- Approximate μ by a discrete measure, for instance

  $$\mu_k = \sum_{i_1 \leq \ldots \leq i_k} \mu(B_{i_1,\ldots,i_k})\delta_{(i_1/k,\ldots,i_k/k)}$$

  where $B_{i_1,\ldots,i_k}$ is the cube $[(i_1-1)/k, i_1/k] \times \ldots [(i_d-1)/k, i_d/k]$

  (Then, $\mathrm{W}_p(\mu_k,\mu) \lesssim \frac{1}{k}$.)

- Compute *exactly* the optimal transport plan T_{μ_k} between ρ and μ_k, (using a **semi-discrete** optimal transport solver).

It is know that T_{μ_k} converges to T_μ but convergence rates are unknown in general...

In general, the numerical analysis for optimal transport is virtually inexistent, whatever the discretization method.

Motivation 3: numerical analysis of optimal transport

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.

To solve numerically an OT problem between $\rho \in \mathrm{Prob}^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mu \in \mathrm{Prob}([0,1]^d)$:

- Approximate μ by a discrete measure, for instance

  $$\mu_k = \sum_{i_1 \leq \ldots \leq i_k} \mu(B_{i_1,\ldots,i_k})\delta_{(i_1/k,\ldots,i_k/k)}$$

  where $B_{i_1,\ldots,i_k}$ is the cube $[(i_1-1)/k, i_1/k] \times \ldots [(i_d-1)/k, i_d/k]$

  (Then, $\mathrm{W}_p(\mu_k,\mu) \lesssim \frac{1}{k}$.)

- Compute *exactly* the optimal transport plan T_{μ_k} between ρ and μ_k, (using a **semi-discrete** optimal transport solver).

It is know that T_{μ_k} converges to T_μ but convergence rates are unknown in general...

In general, the numerical analysis for optimal transport is virtually inexistent, whatever the discretization method.

2. Continuity of $\mu \mapsto T_\mu$.

Elementary remarks

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

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: since $T_{\mu\#}\rho = \mu$ and $T_{\nu\#}\rho = \nu$, one has $\gamma := (T_\mu, T_\nu)_{\#}\rho \in \Gamma(\mu,\nu)$.

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: since $T_{\mu\#}\rho = \mu$ and $T_{\nu\#}\rho = \nu$, one has $\gamma := (T_\mu, T_\nu)_{\#}\rho \in \Gamma(\mu,\nu)$.
>
> Thus, $\mathrm{W}_2^2(\mu,\nu) \leq \int \|x-y\|^2 \,\mathrm{d}\,\gamma(x,y) = \int \|T_\mu(x) - T_\nu(x)\|^2 \,\mathrm{d}\,\rho(x)$.

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: since $T_{\mu\#}\rho = \mu$ and $T_{\nu\#}\rho = \nu$, one has $\gamma := (T_\mu, T_\nu)_{\#}\rho \in \Gamma(\mu,\nu)$.
  >
  > Thus, $\mathrm{W}_2^2(\mu,\nu) \leq \int \|x-y\|^2 \,\mathrm{d}\,\gamma(x,y) = \int \|T_\mu(x) - T_\nu(x)\|^2 \,\mathrm{d}\,\rho(x)$.

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

Elementary remarks

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

> Indeed: since $T_{\mu\#}\rho = \mu$ and $T_{\nu\#}\rho = \nu$, one has $\gamma := (T_\mu, T_\nu)_{\#}\rho \in \Gamma(\mu,\nu)$.
> Thus, $\mathrm{W}_2^2(\mu,\nu) \leq \int \|x-y\|^2 \,\mathrm{d}\,\gamma(x,y) = \int \|T_\mu(x) - T_\nu(x)\|^2 \,\mathrm{d}\,\rho(x)$.

- **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. $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} \geq \mathrm{W}_2(\mu,\nu)$.

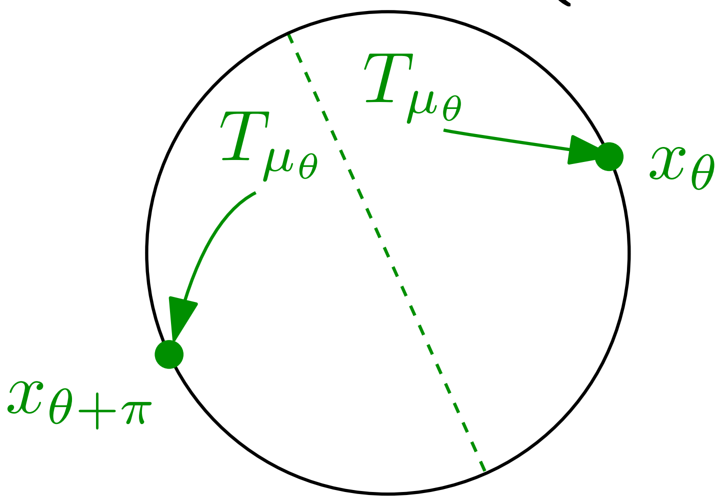
  > Indeed: since $T_{\mu\#}\rho = \mu$ and $T_{\nu\#}\rho = \nu$, one has $\gamma := (T_\mu, T_\nu)_{\#}\rho \in \Gamma(\mu,\nu)$.
  >
  > Thus, $\mathrm{W}_2^2(\mu,\nu) \leq \int \|x-y\|^2 \,\mathrm{d}\,\gamma(x,y) = \int \|T_\mu(x) - T_\nu(x)\|^2 \,\mathrm{d}\,\rho(x)$.

- **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 define $\mu_\theta = \frac{\delta_{x_\theta} + \delta_{x_{\theta+\pi}}}{2}$, 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: since $T_{\mu\#}\rho = \mu$ and $T_{\nu\#}\rho = \nu$, one has $\gamma := (T_\mu, T_\nu)_{\#}\rho \in \Gamma(\mu,\nu)$.

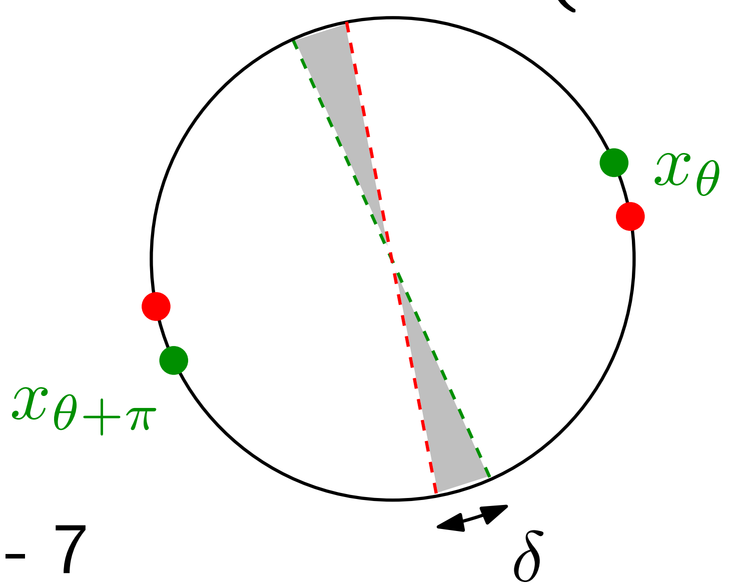
Thus, $\mathrm{W}_2^2(\mu,\nu) \leq \int \|x-y\|^2 \,\mathrm{d}\,\gamma(x,y) = \int \|T_\mu(x) - T_\nu(x)\|^2 \,\mathrm{d}\,\rho(x)$.

- **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 define $\mu_\theta = \frac{\delta_{x_\theta} + \delta_{x_{\theta+\pi}}}{2}$, 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, $\mathrm{W}_2(\mu_\theta, \mu_{\theta+\delta}) \leq C\delta$,

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

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

- $\simeq$ [Ambrosio,Gigli '09] with slightly better upper bound. 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 $\boxed{T_\mu \text{ is } L\text{-Lipschitz,}}$ then $\|T_\mu - T_\nu\|_2^2 \leq C\,\mathrm{W}_1(\mu, \nu)$ with $C = 4L\,\mathrm{diam}(X)$.

- $\simeq$ [Ambrosio,Gigli '09] with slightly better upper bound. See also [Berman '18].
- No regularity assumption on $\nu \longrightarrow$ consequences 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 $\boxed{T_\mu \text{ is } L\text{-Lipschitz,}}$ then $\|T_\mu - T_\nu\|_2^2 \leq C\,\mathrm{W}_1(\mu,\nu)$ with $C = 4L\,\mathrm{diam}(X)$.

- $\simeq$ [Ambrosio,Gigli '09] with slightly better upper bound. See also [Berman '18].
- No regularity assumption on $\nu \longrightarrow$ consequences in statistics and numerical analysis.
- Let $\phi_\mu : X \to \mathbb{R}$ convex s.t. $T_\mu = \nabla\phi_\mu$.
  $\psi_\mu : Y \to \mathbb{R}$ its Legendre transform: $\boxed{\psi_\mu(y) = \max_{x\in X}\langle x|y\rangle - \phi_\mu(x)}$

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

- $\simeq$ [Ambrosio,Gigli '09] with slightly better upper bound. See also [Berman '18].
- No regularity assumption on $\nu \longrightarrow$ consequences in statistics and numerical analysis.
- Let $\phi_\mu : X \to \mathbb{R}$ convex s.t. $T_\mu = \nabla \phi_\mu$.
  $\psi_\mu : Y \to \mathbb{R}$ its Legendre transform: $\boxed{\psi_\mu(y) = \max_{x \in X} \langle x | y \rangle - \phi_\mu(x)}$

Prop: If T_μ is L-Lipschitz, then $\|T_\mu - T_\nu\|^2_{\mathrm{L}^2(\rho)} \leq -2L \int (\psi_\mu - \psi_\nu)\,\mathrm{d}(\mu - \nu)$.

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

- $\simeq$ [Ambrosio,Gigli '09] with slightly better upper bound. See also [Berman '18].
- No regularity assumption on $\nu \longrightarrow$ consequences in statistics and numerical analysis.
- Let $\phi_\mu : X \to \mathbb{R}$ convex s.t. $T_\mu = \nabla \phi_\mu$.
  $\psi_\mu : Y \to \mathbb{R}$ its Legendre transform: $\boxed{\psi_\mu(y) = \max_{x \in X} \langle x | y \rangle - \phi_\mu(x)}$

Prop: If T_μ is L-Lipschitz, then $\|T_\mu - T_\nu\|^2_{\mathrm{L}^2(\rho)} \leq -2L \int (\psi_\mu - \psi_\nu)\, \mathrm{d}(\mu - \nu)$.

- **Prop** $\Longrightarrow$ **Thm:** Kantorovich-Rubinstein theorem

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

- $\simeq$ [Ambrosio,Gigli '09] with slightly better upper bound. See also [Berman '18].
- No regularity assumption on $\nu \longrightarrow$ consequences in statistics and numerical analysis.
- Let $\phi_\mu : X \to \mathbb{R}$ convex s.t. $T_\mu = \nabla\phi_\mu$.
  $\psi_\mu : Y \to \mathbb{R}$ its Legendre transform: $\boxed{\psi_\mu(y) = \max_{x\in X}\langle x|y\rangle - \phi_\mu(x)}$

Prop: If T_μ is L-Lipschitz, then $\|T_\mu - T_\nu\|^2_{\mathrm{L}^2(\rho)} \leq -2L\int(\psi_\mu - \psi_\nu)\,\mathrm{d}(\mu - \nu)$.

- $\int \psi_\nu\,\mathrm{d}(\mu - \nu) = \int \psi_\nu\,\mathrm{d}(\nabla\phi_{\mu\#}\rho - \nabla\phi_{\nu\#}\rho) = \int \psi_\nu(\nabla\phi_\mu) - \psi_\nu(\nabla\phi_\nu)\,\mathrm{d}\,\rho$

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

- $\simeq$ [Ambrosio,Gigli '09] with slightly better upper bound. See also [Berman '18].
- No regularity assumption on $\nu \longrightarrow$ consequences in statistics and numerical analysis.
- Let $\phi_\mu : X \to \mathbb{R}$ convex s.t. $T_\mu = \nabla \phi_\mu$.
  $\psi_\mu : Y \to \mathbb{R}$ its Legendre transform: $\boxed{\psi_\mu(y) = \max_{x \in X} \langle x | y \rangle - \phi_\mu(x)}$

Prop: If T_μ is L-Lipschitz, then $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)}^2 \leq -2L \int (\psi_\mu - \psi_\nu)\, \mathrm{d}(\mu - \nu)$.

- $\int \psi_\nu \, \mathrm{d}(\mu - \nu) = \int \psi_\nu \, \mathrm{d}(\nabla \phi_{\mu\#}\rho - \nabla \phi_{\nu\#}\rho) = \int \psi_\nu(\nabla \phi_\mu) - \psi_\nu(\nabla \phi_\nu) \, \mathrm{d}\, \rho$

(convexity: $\psi_\nu(y) - \psi_\nu(x) \geq \langle y - x | \nabla \psi_\nu(x) \rangle$) $\quad \geq \int \langle \nabla \psi_\mu - \nabla \psi_\nu | \nabla \psi_\nu(\nabla \phi_\nu) \rangle \, \mathrm{d}\, \rho$

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

- $\simeq$ [Ambrosio,Gigli '09] with slightly better upper bound. See also [Berman '18].
- No regularity assumption on $\nu \longrightarrow$ consequences in statistics and numerical analysis.
- Let $\phi_\mu : X \to \mathbb{R}$ convex s.t. $T_\mu = \nabla\phi_\mu$.
  $\psi_\mu : Y \to \mathbb{R}$ its Legendre transform: $\boxed{\psi_\mu(y) = \max_{x\in X} \langle x|y\rangle - \phi_\mu(x)}$

Prop: If T_μ is L-Lipschitz, then $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)}^2 \leq -2L \int (\psi_\mu - \psi_\nu)\,\mathrm{d}(\mu - \nu)$.

- $\int \psi_\nu \,\mathrm{d}(\mu-\nu) = \int \psi_\nu \,\mathrm{d}(\nabla\phi_{\mu\#}\rho - \nabla\phi_{\nu\#}\rho) = \int \psi_\nu(\nabla\phi_\mu) - \psi_\nu(\nabla\phi_\nu)\,\mathrm{d}\,\rho$

(convexity: $\psi_\nu(y) - \psi_\nu(x) \geq \langle y - x|\nabla\psi_\nu(x)\rangle$) $\quad \geq \int \langle \nabla\psi_\mu - \nabla\psi_\nu | \nabla\psi_\nu(\nabla\phi_\nu)\rangle\,\mathrm{d}\,\rho$

$$= \int \langle \nabla\psi_\mu - \nabla\psi_\nu | \mathrm{id}\rangle\,\mathrm{d}\,\rho$$

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

- $\simeq$ [Ambrosio,Gigli '09] with slightly better upper bound. See also [Berman '18].
- No regularity assumption on $\nu \longrightarrow$ consequences in statistics and numerical analysis.
- Let $\phi_\mu : X \to \mathbb{R}$ convex s.t. $T_\mu = \nabla\phi_\mu$.
  $\psi_\mu : Y \to \mathbb{R}$ its Legendre transform: $\boxed{\psi_\mu(y) = \max_{x\in X}\langle x|y\rangle - \phi_\mu(x)}$

Prop: If T_μ is L-Lipschitz, then $\|T_\mu - T_\nu\|^2_{\mathrm{L}^2(\rho)} \leq -2L\int(\psi_\mu - \psi_\nu)\,\mathrm{d}(\mu - \nu)$.

- $\int \psi_\nu\,\mathrm{d}(\mu-\nu) = \int \psi_\nu\,\mathrm{d}(\nabla\phi_{\mu\#}\rho - \nabla\phi_{\nu\#}\rho) = \int \psi_\nu(\nabla\phi_\mu) - \psi_\nu(\nabla\phi_\nu)\,\mathrm{d}\,\rho$

(convexity: $\psi_\nu(y) - \psi_\nu(x) \geq \langle y - x|\nabla\psi_\nu(x)\rangle$) $\quad \geq \int\langle\nabla\psi_\mu - \nabla\psi_\nu|\nabla\psi_\nu(\nabla\phi_\nu)\rangle\,\mathrm{d}\,\rho$

$$= \int\langle\nabla\psi_\mu - \nabla\psi_\nu|\mathrm{id}\rangle\,\mathrm{d}\,\rho$$

- $\int \psi_\mu\,\mathrm{d}(\nu-\mu) \geq \int\langle\nabla\psi_\nu - \nabla\psi_\mu|\mathrm{id}\rangle\,\mathrm{d}\,\rho + \frac{L}{2}\|\nabla\phi_\mu - \nabla\phi_\nu\|_{L^2(\rho)}$
  ($T_\mu = \nabla\phi_\mu$ L-Lipschitz $\iff \psi_\mu = \phi_\mu^*$ is L-strongly convex)

Global Hölder continuity

Thm (Berman, '18): Let $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ and $\mu, \nu \in \mathrm{Prob}(Y)$ with X, Y compact.
Then, $\|\nabla\psi_\mu - \nabla\psi_\nu\|^2_{\mathrm{L}^2(Y)} \leq C\,\mathrm{W}_1(\mu,\nu)^\alpha$ with $\alpha = \frac{1}{2^{d-1}}$

Global Hölder continuity

Thm (Berman, '18): Let $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ and $\mu, \nu \in \mathrm{Prob}(Y)$ with X, Y compact.
Then, $\|\nabla\psi_\mu - \nabla\psi_\nu\|^2_{\mathrm{L}^2(Y)} \leq C\,\mathrm{W}_1(\mu,\nu)^\alpha$ with $\alpha = \frac{1}{2^{d-1}}$

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

Global Hölder continuity

Thm (Berman, '18): Let $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ and $\mu, \nu \in \mathrm{Prob}(Y)$ with X, Y compact.
Then, $\|\nabla\psi_\mu - \nabla\psi_\nu\|^2_{\mathrm{L}^2(Y)} \leq C\,\mathrm{W}_1(\mu,\nu)^\alpha$ with $\alpha = \frac{1}{2^{d-1}}$

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

- The Hölder exponent is terrible, but inequality holds without assumptions on μ, ν!

Global Hölder continuity

Thm (Berman, '18): Let $\rho \in \mathrm{Prob}^{\mathrm{ac}}(X)$ and $\mu, \nu \in \mathrm{Prob}(Y)$ with X, Y compact.
Then, $\|\nabla\psi_\mu - \nabla\psi_\nu\|^2_{\mathrm{L}^2(Y)} \leq C\, \mathrm{W}_1(\mu,\nu)^\alpha$ with $\alpha = \frac{1}{2^{d-1}}$

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

- The Hölder exponent is terrible, but inequality holds without assumptions on μ, ν!
- Proof of Berman's theorem relies on techniques from complex geometry.

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

Main theorem

Thm (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu,\nu)^{1/5}.$$

Main theorem

Thm (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu, \nu)^{1/5}.$$

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

Main theorem

Thm (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu, \nu)^{1/5}.$$

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

- Gap between lower-bound and upper bound for Hölder exponent: $\frac{1}{5} < \frac{1}{2}$.

  The exponent $\frac{1}{5}$ is certainly not optimal...

Main theorem

Thm (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu, \nu)^{1/5}.$$

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

- Gap between lower-bound and upper bound for Hölder exponent: $\frac{1}{5} < \frac{1}{2}$.

  The exponent $\frac{1}{5}$ is certainly not optimal...

- The constant C depend polynomially on $\mathrm{diam}(X), \mathrm{diam}(Y)$.

Main theorem

Thm (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu,\nu)^{1/5}.$$

- First global and dimension-independent stability result for optimal transport maps.
- Gap between lower-bound and upper bound for Hölder exponent: $\frac{1}{5} < \frac{1}{2}$.

  The exponent $\frac{1}{5}$ is certainly not optimal...
- The constant C depend polynomially on $\mathrm{diam}(X), \mathrm{diam}(Y)$.
- Proof relies on the semidiscrete setting, i.e. the bound is established in the case

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

  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,

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

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

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,

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

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

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,

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

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

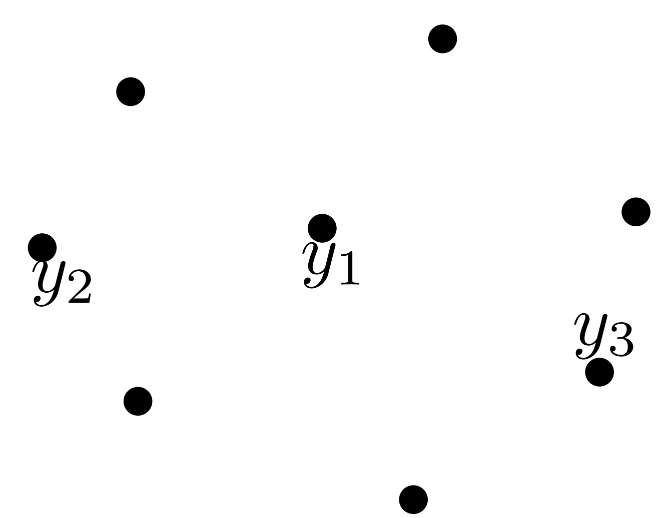
Kantorovich duality

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

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,

$$\mathcal{T}(\rho,\mu) = \max_{\gamma\in\Gamma(\rho,\mu)} \int \langle x|y\rangle \,\mathrm{d}\,\gamma(x,y)$$
$$\boxed{= \min_{\phi\oplus\psi\geq\langle\cdot|\cdot\rangle} \int \phi \,\mathrm{d}\,\rho + \int \psi \,\mathrm{d}\,\mu}$$
$$= \min_{\psi} \int \psi^* \,\mathrm{d}\,\rho + \int \psi \,\mathrm{d}\,\mu$$

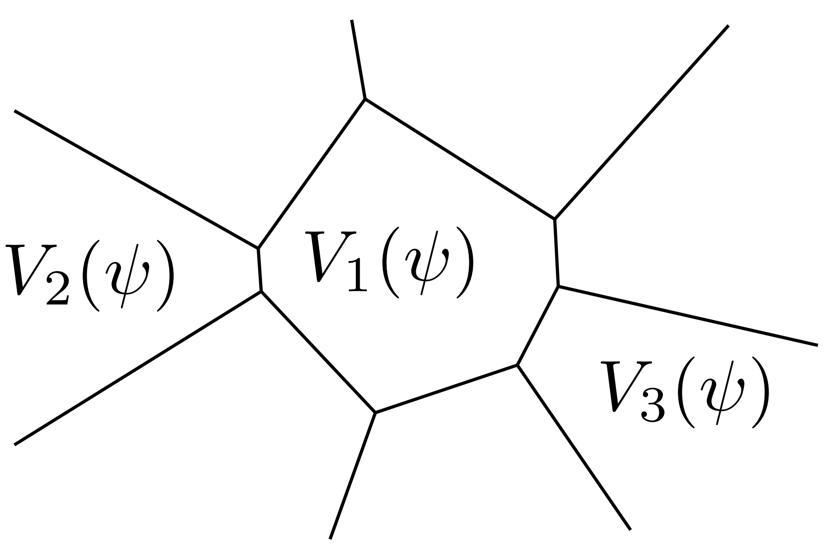
Kantorovich duality

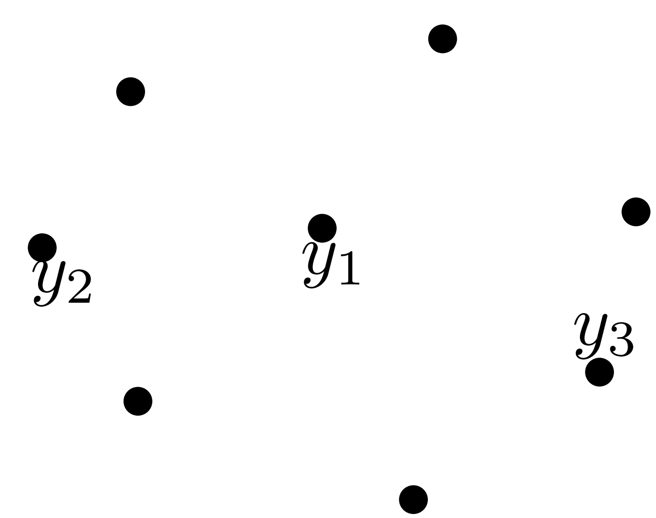
Legendre-Fenchel transform:

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

- 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,

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

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

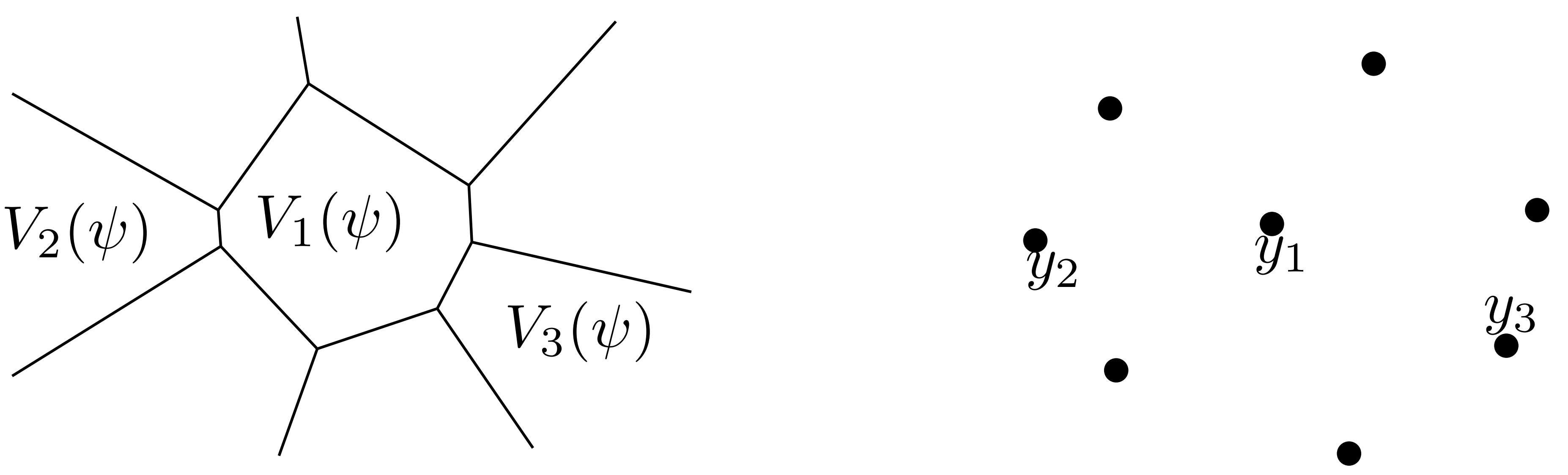
$$= \min_{\psi} \int \psi^* \,\mathrm{d}\,\rho + \int \psi \,\mathrm{d}\,\mu$$

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

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



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

Optimality condition and economic interpretation

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

Optimality condition and economic interpretation

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) - \sum_i \mu_i\psi_i$, 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 $G_i(\psi) = \rho(V_i(\psi))$.

Optimality condition and economic interpretation

$\mathcal{T}(\rho, \mu) = \min_{\psi \in \mathbb{R}^N} \Phi(\psi) - \sum_i \mu_i \psi_i$, 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 $G_i(\psi) = \rho(V_i(\psi))$.

$\psi \in \mathbb{R}^N$ is a minimizer of dual pb $\Longleftrightarrow \forall i, \rho(V_i(\psi)) = \mu_i$

Optimality condition and economic interpretation

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) - \sum_i \mu_i\psi_i$, 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 $G_i(\psi) = \rho(V_i(\psi))$.

$\psi \in \mathbb{R}^N$ is a minimizer of dual pb $\iff \forall i, \rho(V_i(\psi)) = \mu_i$

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

Optimality condition and economic interpretation

$\mathcal{T}(\rho, \mu) = \min_{\psi \in \mathbb{R}^N} \Phi(\psi) - \sum_i \mu_i \psi_i$, 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 $G_i(\psi) = \rho(V_i(\psi))$.

$$\begin{aligned}
\psi \in \mathbb{R}^N \text{ is a minimizer of dual pb} &\Longleftrightarrow \forall i, \rho(V_i(\psi)) = \mu_i \\
&\Longleftrightarrow G(\psi) = \mu \text{ with } G = (G_1, \ldots, G_N),\ \mu \in \mathbb{R}^N \\
&\Longleftrightarrow T = \nabla \psi^* \text{ transports } \rho \text{ onto } \textstyle\sum_i \mu_i \delta_{y_i}
\end{aligned}$$

Optimality condition and economic interpretation

$\mathcal{T}(\rho, \mu) = \min_{\psi \in \mathbb{R}^N} \Phi(\psi) - \sum_i \mu_i \psi_i$, 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 $G_i(\psi) = \rho(V_i(\psi))$.

$\psi \in \mathbb{R}^N$ is a minimizer of dual pb $\iff \forall i, \rho(V_i(\psi)) = \mu_i$

$\iff G(\psi) = \mu$ 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

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) - \sum_i \mu_i\psi_i$, 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 $G_i(\psi) = \rho(V_i(\psi))$.

$\psi \in \mathbb{R}^N$ is a minimizer of dual pb $\Longleftrightarrow \forall i, \rho(V_i(\psi)) = \mu_i$

$\Longleftrightarrow G(\psi) = \mu$ 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.

Optimality condition and economic interpretation

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) - \sum_i \mu_i\psi_i$, 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 $G_i(\psi) = \rho(V_i(\psi))$.

$\psi \in \mathbb{R}^N$ is a minimizer of dual pb $\iff \forall i, \rho(V_i(\psi)) = \mu_i$

$\iff G(\psi) = \mu$ 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.

Optimality condition and economic interpretation

$\mathcal{T}(\rho,\mu) = \min_{\psi\in\mathbb{R}^N} \Phi(\psi) - \sum_i \mu_i\psi_i$, 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 $G_i(\psi) = \rho(V_i(\psi))$.

$\psi \in \mathbb{R}^N$ is a minimizer of dual pb $\Longleftrightarrow \forall i, \rho(V_i(\psi)) = \mu_i$

$\Longleftrightarrow G(\psi) = \mu$ 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.

$\longrightarrow \rho(V_i) =$ amount of customers for product y_i.

Optimality condition and economic interpretation

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

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

$\psi\in\mathbb{R}^N$ is a minimizer of dual pb $\iff \forall i, \rho(V_i(\psi)) = \mu_i$

$\iff G(\psi) = \mu$ 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, $\{y_i\}_{1\le i\le N} =$ 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.

$\longrightarrow \rho(V_i) =$ amount of customers for product y_i.

$\boxed{\text{Optimal transport} = \text{finding prices satisfying capacity constraints } \rho(V_i(\psi)) = \mu_i.}$

Optimality condition and economic interpretation

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

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

$\psi \in \mathbb{R}^N$ is a minimizer of dual pb $\iff \forall i, \rho(V_i(\psi)) = \mu_i$

$\iff G(\psi) = \mu$ 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, $\{y_i\}_{1\le i\le N} =$ 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.

$\longrightarrow \rho(V_i) =$ amount of customers for product y_i.

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

- **Algorithm (Oliker–Prussner):** coordinate-wise increment. Complexity: $O(N^3)$.

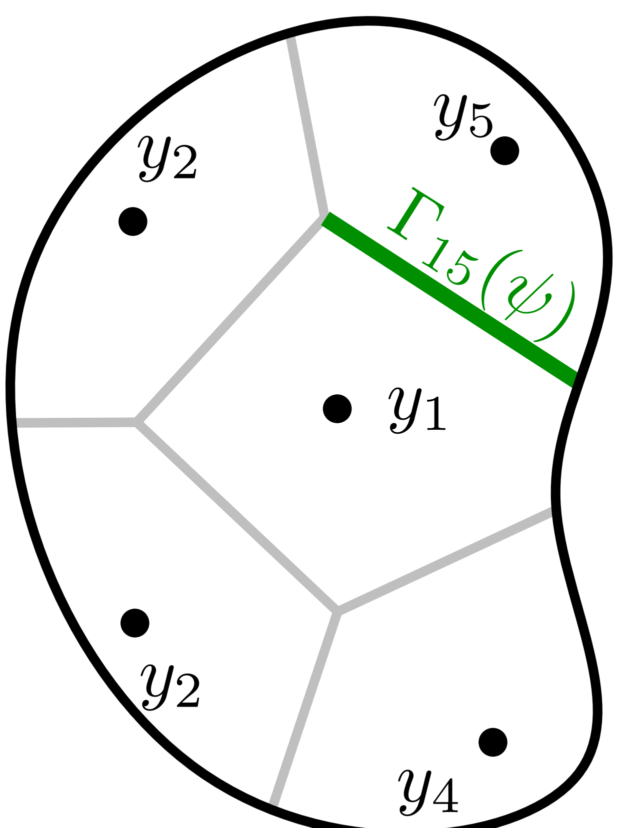
Hessian on Φ and Newton's Algorithm

(Recall that $G_i(\psi) = \rho(V_i(\psi))$ and $\nabla\Phi = -(G_1, \ldots, G_N)$)

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

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

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



Hessian on Φ and Newton's Algorithm

(Recall that $G_i(\psi) = \rho(V_i(\psi))$ and $\nabla\Phi = -(G_1, \ldots, G_N)$)

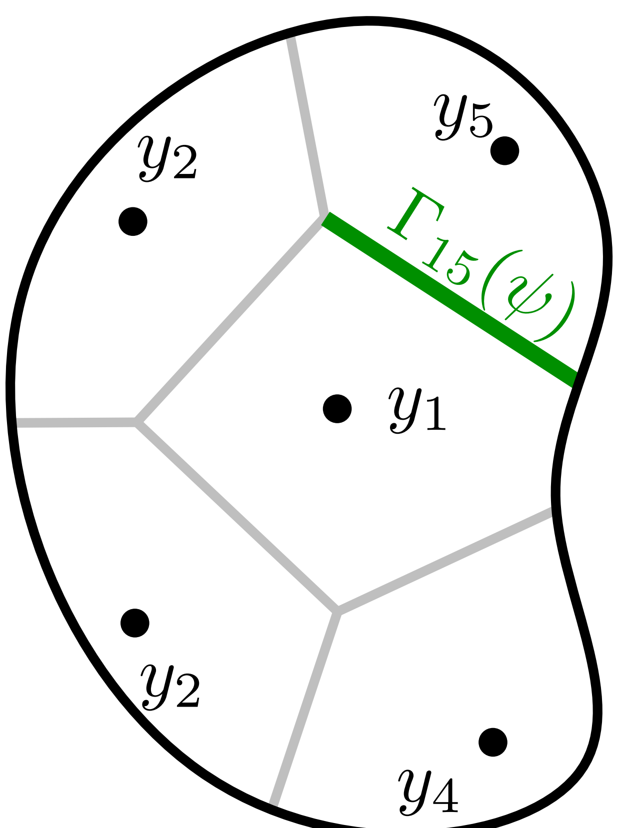
Proposition: ▶ If $\rho \in \mathcal{C}^0(X)$ and $(y_i)_{1 \leq i \leq N}$ is generic, then $\Phi \in \mathcal{C}^2(\mathbb{R}^N)$ and

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

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

Let $E = \{\psi \in \mathbb{R}^N \mid \forall i, G_i(\psi) > 0\}$

▶ If $\Omega = \{\rho > 0\}$ is connected and $\psi \in E$, then $\mathrm{Ker} \mathrm{D}^2 \Phi(\psi) = \mathbb{R}(1, \ldots, 1)$.



Hessian on Φ and Newton's Algorithm

(Recall that $G_i(\psi) = \rho(V_i(\psi))$ and $\nabla\Phi = -(G_1, \dots, G_N)$)

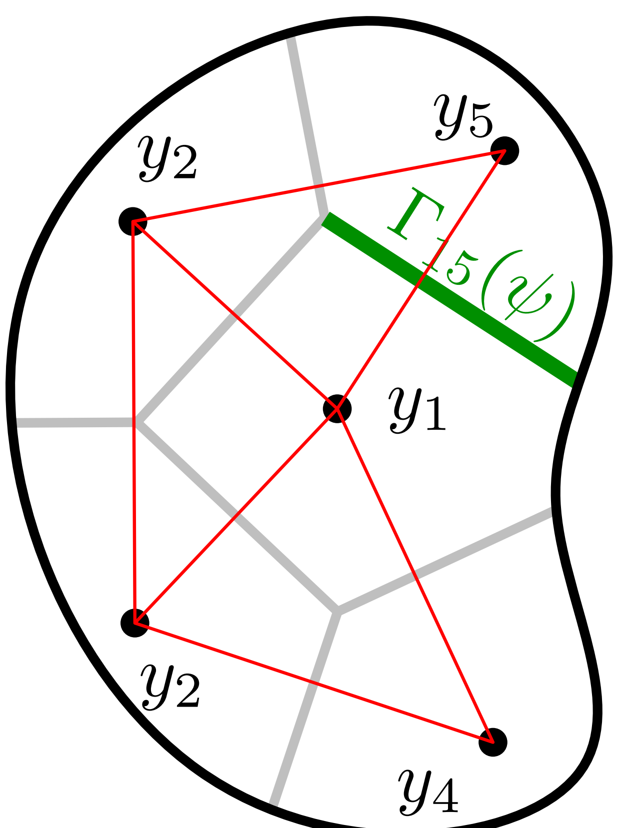
Proposition: ▶ If $\rho \in \mathcal{C}^0(X)$ and $(y_i)_{1 \le i \le N}$ is generic, then $\Phi \in \mathcal{C}^2(\mathbb{R}^N)$ and

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

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

Let $E = \{\psi \in \mathbb{R}^N \mid \forall i, G_i(\psi) > 0\}$

▶ If $\Omega = \{\rho > 0\}$ is connected and $\psi \in E$, then $\mathrm{Ker} \mathrm{D}^2 \Phi(\psi) = \mathbb{R}(1, \dots, 1)$.



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

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

Hessian on Φ and Newton's Algorithm

(Recall that $G_i(\psi) = \rho(V_i(\psi))$ and $\nabla\Phi = -(G_1, \dots, G_N)$)

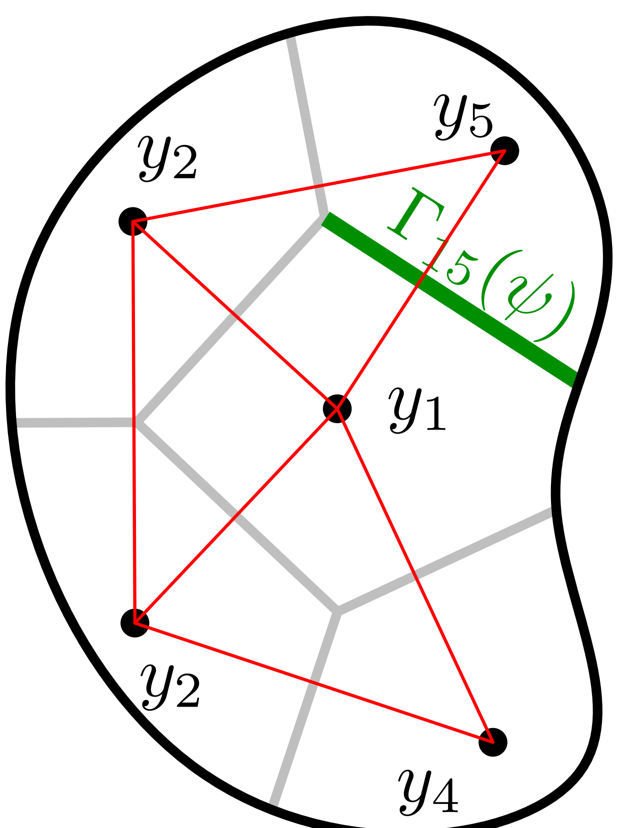
Proposition: ▶ If $\rho \in \mathcal{C}^0(X)$ and $(y_i)_{1\leq i\leq N}$ is generic, then $\Phi \in \mathcal{C}^2(\mathbb{R}^N)$ and

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

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

Let $E = \{\psi \in \mathbb{R}^N \mid \forall i, G_i(\psi) > 0\}$

▶ If $\Omega = \{\rho > 0\}$ is connected and $\psi \in E$, then $\mathrm{Ker}\mathrm{D}^2\Phi(\psi) = \mathbb{R}(1, \dots, 1)$.



▶ 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 on Φ and Newton's Algorithm

(Recall that $G_i(\psi) = \rho(V_i(\psi))$ and $\nabla\Phi = -(G_1, \ldots, G_N)$)

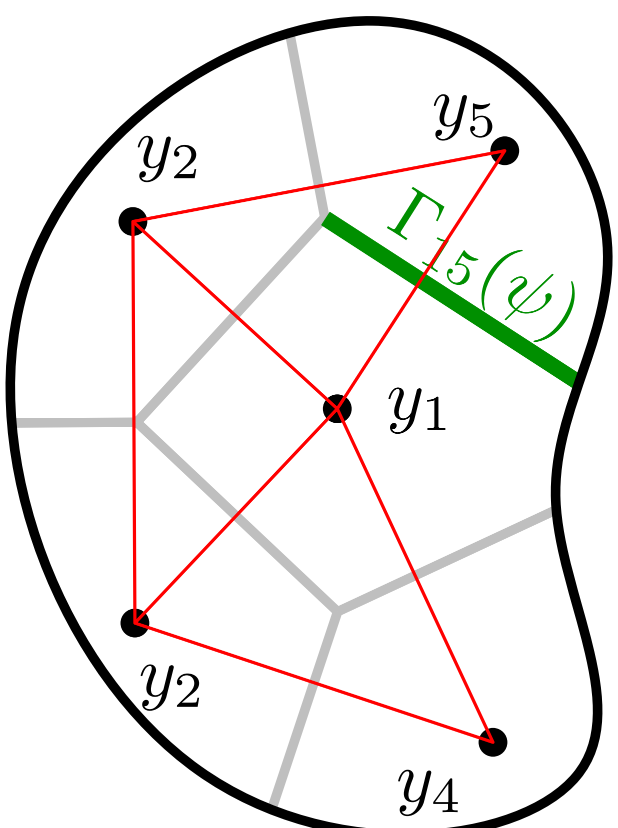
Proposition: ▶ If $\rho \in \mathcal{C}^0(X)$ and $(y_i)_{1 \leq i \leq N}$ is generic, then $\Phi \in \mathcal{C}^2(\mathbb{R}^N)$ and

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

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

Let $E = \{\psi \in \mathbb{R}^N \mid \forall i, G_i(\psi) > 0\}$

▶ If $\Omega = \{\rho > 0\}$ is connected and $\psi \in E$, then $\mathrm{KerD}^2\Phi(\psi) = \mathbb{R}(1, \ldots, 1)$.



- ▶ 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 $\Longrightarrow \mathrm{Ker}L = \mathbb{R} \cdot \mathrm{cst}$

Hessian on Φ and Newton's Algorithm

(Recall that $G_i(\psi) = \rho(V_i(\psi))$ and $\nabla\Phi = -(G_1, \ldots, G_N)$)

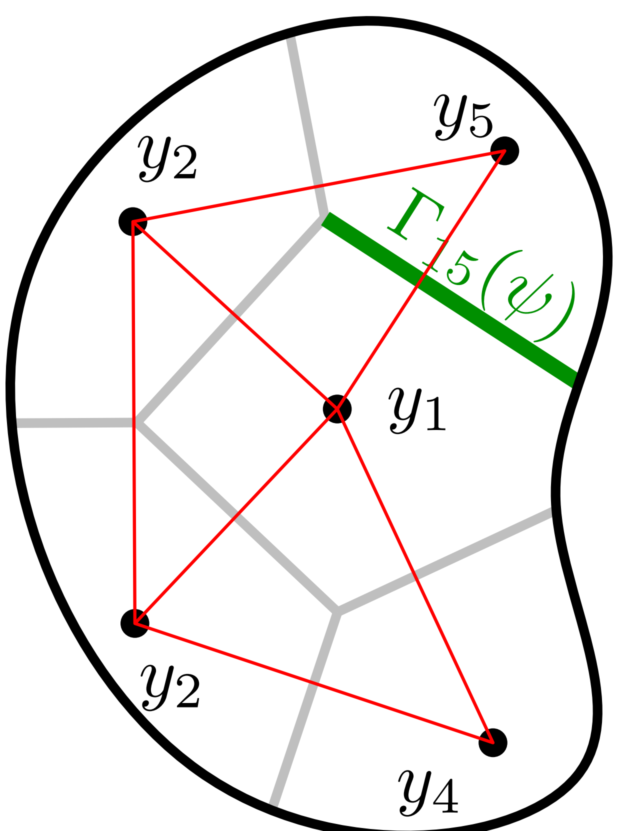
Proposition: ▶ If $\rho \in \mathcal{C}^0(X)$ and $(y_i)_{1\leq i\leq N}$ is generic, then $\Phi \in \mathcal{C}^2(\mathbb{R}^N)$ and

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

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

Let $E = \{\psi \in \mathbb{R}^N \mid \forall i, G_i(\psi) > 0\}$

▶ If $\Omega = \{\rho > 0\}$ is connected and $\psi \in E$, then $\mathrm{Ker}\mathrm{D}^2\Phi(\psi) = \mathbb{R}(1, \ldots, 1)$.



- 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 $\implies \mathrm{Ker}L = \mathbb{R}\cdot\mathrm{cst}$

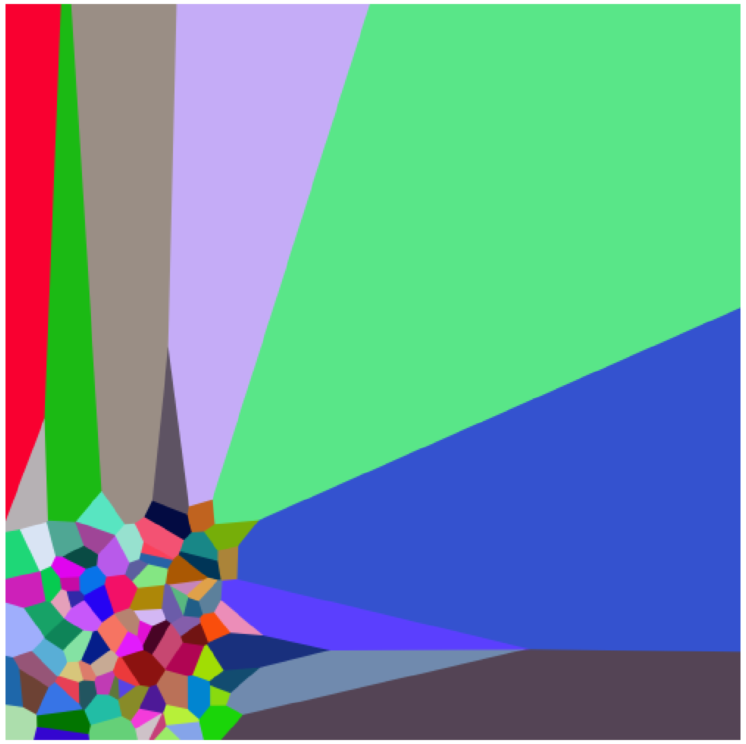
Corollary: Global convergence of a damped Newton algorithm.

[Kitagawa, M., Thibert 16]

Numerical example

Source: $\rho =$ uniform on $[0,1]^2$,

Target: $\mu = \frac{1}{N}\sum_{1\leq i\leq N}\delta_{y_i}$ with y_i uniform i.i.d. in $[0,\frac{1}{3}]^2$

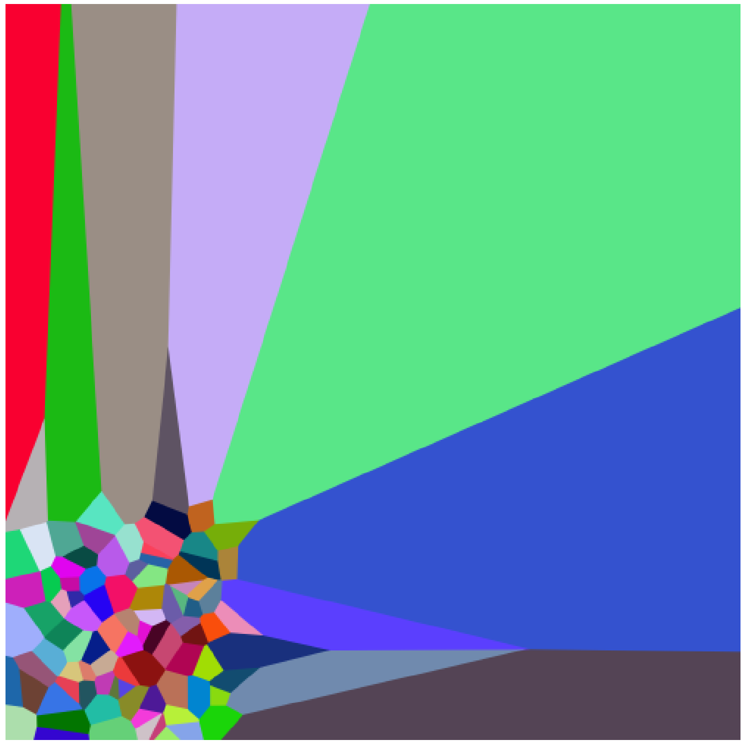


$\psi_0 = \frac{1}{2}\|\cdot\|^2$

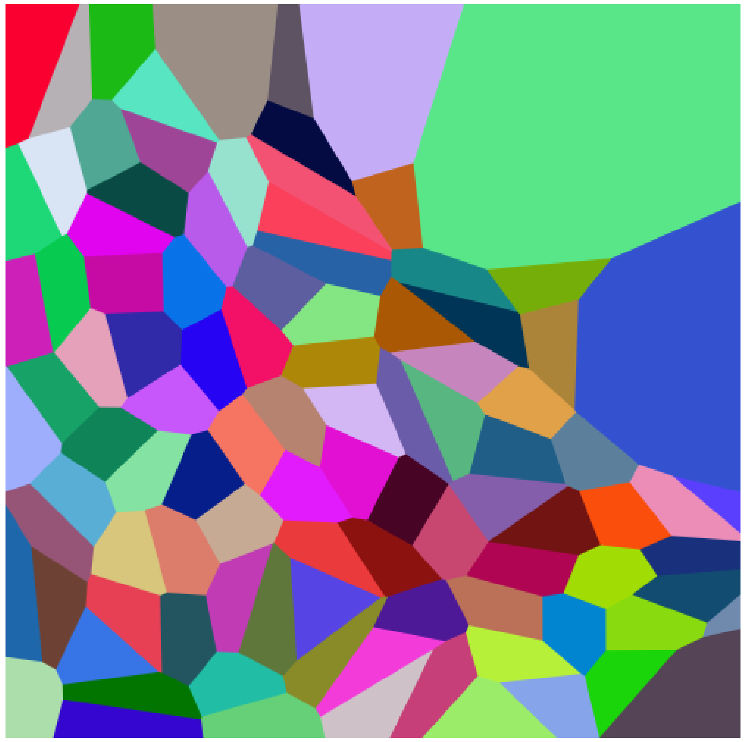
Numerical example

Source: $\rho =$ uniform on $[0,1]^2$,

Target: $\mu = \frac{1}{N}\sum_{1\leq i\leq N} \delta_{y_i}$ with y_i uniform i.i.d. in $[0, \frac{1}{3}]^2$



$\psi_0 = \frac{1}{2}\|\cdot\|^2$



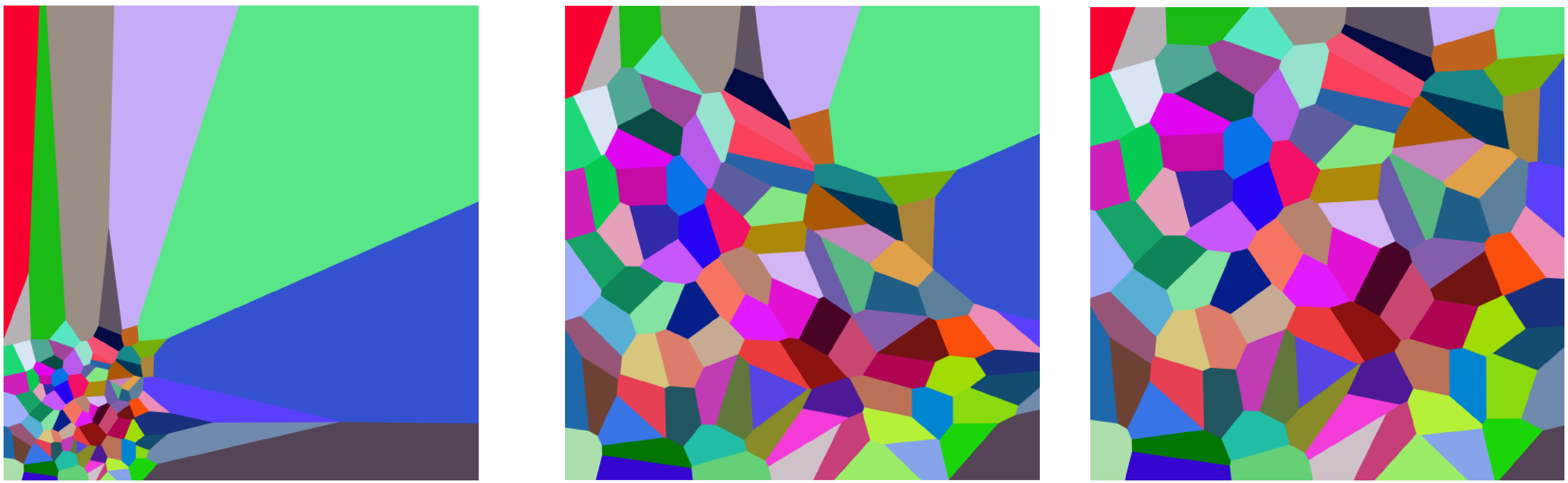
$\psi_1 = \mathrm{Newt}(\psi_0)$

NB: The points do **not** move.

Numerical example

Source: $\rho =$ uniform on $[0,1]^2$,

Target: $\mu = \frac{1}{N}\sum_{1\leq i\leq N}\delta_{y_i}$ with y_i uniform i.i.d. in $[0, \frac{1}{3}]^2$



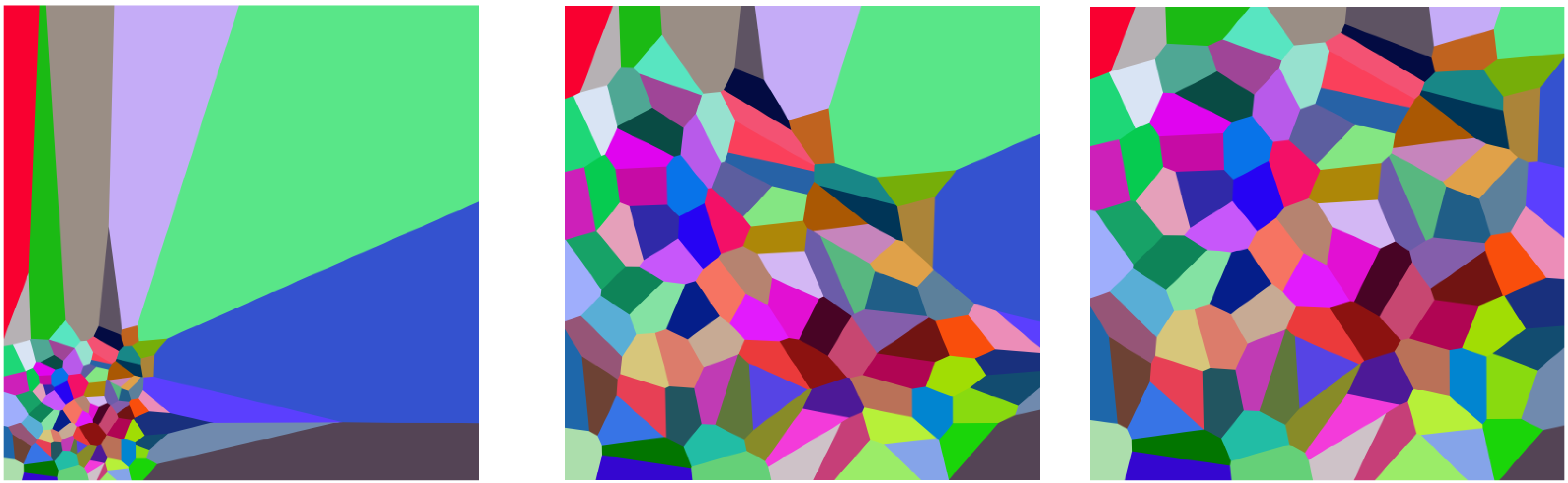
$\psi_0 = \frac{1}{2}\|\cdot\|^2$ $\quad$ $\psi_1 = \mathrm{Newt}(\psi_0)$ $\quad$ $\psi_2 = \mathrm{Newt}(\psi_1)$

NB: The points do **not** move.

Numerical example

Source: $\rho =$ uniform on $[0,1]^2$,

Target: $\mu = \frac{1}{N}\sum_{1\leq i\leq N}\delta_{y_i}$ with y_i uniform i.i.d. in $[0, \frac{1}{3}]^2$



$\psi_0 = \frac{1}{2}\|\cdot\|^2$ $\quad$ $\psi_1 = \mathrm{Newt}(\psi_0)$ $\quad$ $\psi_2 = \mathrm{Newt}(\psi_1)$

NB: The points do **not** move.

Convergence is *very* fast when $\mathrm{spt}(\rho)$ convex: 17 Newton iterations for $N \geq 10^7$ in 3D.

Proof ingredients

Proof ingredients

Thm (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu, \nu)^{1/5}.$$

Proof ingredients

Thm (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu, \nu)^{1/5}.$$

- **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 (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu,\nu)^{1/5}.$$

- **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 (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu,\nu)^{1/5}.$$

- **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 (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu,\nu)^{1/5}.$$

- **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 (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu,\nu)^{1/5}.$$

- **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 (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu,\nu)^{1/5}.$$

- **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 $\boxed{\|\psi^1 - \psi^0\|^2_{\mathrm{L}^2(\mu^0)} \lesssim |\langle \mu^1 - \mu^0 | \psi^1 - \psi^0\rangle|}$

Proof ingredients

Thm (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\, \mathrm{W}_2(\mu,\nu)^{1/5}.$$

- **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 $\boxed{\|\psi^1 - \psi^0\|^2_{\mathrm{L}^2(\mu^0)} \lesssim |\langle \mu^1 - \mu^0 | \psi^1 - \psi^0 \rangle|}$

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

Proof ingredients

Thm (M., Delalande, Chazal '19): Let X convex compact with $|X| = 1$ and $\rho = \mathrm{Leb}_X$, and let Y be compact. Then, there exists C s.t. for all $\mu, \nu \in \mathrm{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{\mathrm{L}^2(X)} \leq C\,\mathrm{W}_2(\mu,\nu)^{1/5}.$$

- **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 $\boxed{\|\psi^1 - \psi^0\|^2_{\mathrm{L}^2(\mu^0)} \lesssim |\langle \mu^1 - \mu^0 | \psi^1 - \psi^0\rangle|}$

  Then, by 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...

A toy application

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

Example: k-Means for MNIST digits

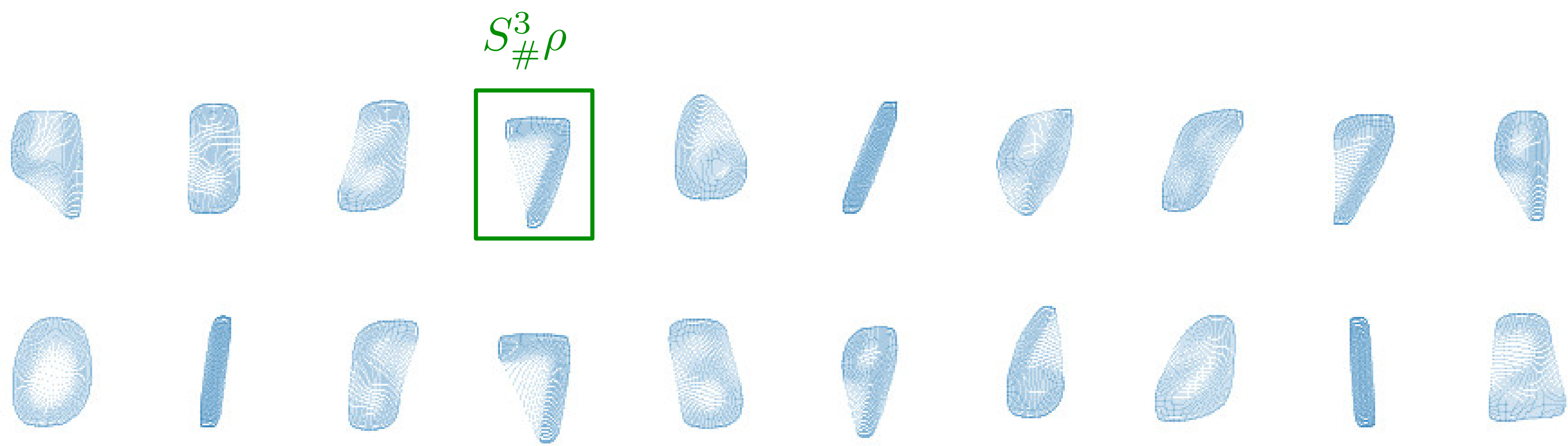
MNIST has $M = 60\,000$ images grayscale images (64×64 pixels) representing digits.
Each image $\alpha^\ell \in \mathcal{M}_{64}(\mathbb{R})$ is transformed into a probability measure on $[0,1]^2$ via

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

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

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

Each cluster $X^k \subseteq \{0,\ldots,M\}$ 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".



Summary

Optimal transport can be used to embed of $\mathrm{Prob}(\mathbb{R}^d)$ into $\mathrm{L}^2(\rho, \mathbb{R}^d)$, with possible applications in data analysis. Computations can be easily performed using

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

Summary

Optimal transport can be used to embed of $\mathrm{Prob}(\mathbb{R}^d)$ into $\mathrm{L}^2(\rho, \mathbb{R}^d)$, with possible applications in data analysis. Computations can be easily performed using

$\boxed{\texttt{https://github.com/sd-ot}}$

The analysis of this approach relies on the stability theory for $\mu \mapsto T_\mu$, both with respect to W_2, which has many open questions.

Summary

Optimal transport can be used to embed of $\mathrm{Prob}(\mathbb{R}^d)$ into $\mathrm{L}^2(\rho, \mathbb{R}^d)$, with possible applications in data analysis. Computations can be easily performed using

$\boxed{\texttt{https://github.com/sd-ot}}$

The analysis of this approach relies on the stability theory for $\mu \mapsto T_\mu$, both with respect to W_2, which has many open questions.

Thank you for your attention!