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

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

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

- $T_\mu$ is the inverse cdf, also called quantile function.
Motivation 1: Monge-Kantorovich Quantiles

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

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

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

- $T_\mu$ is the inverse cdf, also called *quantile function*.

How to extend this notion to a multivariate setting?
Motivation 1: Monge-Kantorovich Quantiles

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

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

- $T_\mu$ is the inverse cdf, also called \textit{quantile function}.

How to extend this notion to a multivariate setting?

\textbf{Theorem (Brenier, McCann)} Given $\rho \in \text{Prob}^{ac}(\mathbb{R}^d)$ and $\mu \in \text{Prob}(\mathbb{R}^d)$, 

$\exists! \rho$-a.e. $T_\mu : \mathbb{R}^d \to \mathbb{R}^d$ such that $T_\mu \# \rho = \mu$ and $T_\mu = \nabla \phi$ with $\phi$ convex.
Motivation 1: Monge-Kantorovich Quantiles

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

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

- \( T_\mu \) is the inverse cdf, also called quantile function.

How to extend this notion to a multivariate setting?

**Theorem (Brenier, McCann)** Given \( \rho \in \text{Prob}^{ac}(\mathbb{R}^d) \) and \( \mu \in \text{Prob}(\mathbb{R}^d) \),

\[ \exists! \rho\text{-a.e. } T_\mu : \mathbb{R}^d \to \mathbb{R}^d \text{ such that } T_\mu \# \rho = \mu \text{ and } T_\mu = \nabla \phi \text{ with } \phi \text{ convex}. \]

- Monge-Kantorovich quantile := \( T_\mu \). Need of a reference probability density \( \rho \).

[Cherzounukov, Galichon, Hallin, Henry, '15]
Motivation 1: Monge-Kantorovich Quantiles

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

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

- \( T_\mu \) is the inverse cdf, also called quantile function.

How to extend this notion to a multivariate setting?

**Theorem (Brenier, McCann)** Given \( \rho \in \text{Prob}^{ac}(\mathbb{R}^d) \) and \( \mu \in \text{Prob}(\mathbb{R}^d) \),

\[ \exists! \rho\text{-a.e. } T_\mu : \mathbb{R}^d \to \mathbb{R}^d \text{ such that } T_\mu \# \rho = \mu \text{ and } T_\mu = \nabla \phi \text{ with } \phi \text{ convex.} \]

- Monge-Kantorovich quantile := \( T_\mu \). Need of a reference probability density \( \rho \).

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

- \( T_\mu \) is unique \( \rho \)-a.e. but the convex function \( \phi_\mu \) is not necessarily unique.
Motivation 1: Monge-Kantorovich Quantiles

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

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

- $T_\mu$ is the inverse cdf, also called quantile function.

How to extend this notion to a multivariate setting?

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

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

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

- $T_\mu$ is unique $\rho$-a.e. but the convex function $\phi_\mu$ is not necessarily unique.

- $T_\mu : \text{spt}(\rho) \to \mathbb{R}^d$ is monotone: $\langle T_\mu(x) - T_\mu(y)|x - y\rangle \geq 0$. 
Numerical Example: Monge-Kantorovich Depth

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

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

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

[Cherzonukov, Galichon, Hallin, Henry]
Numerical Example: Monge-Kantorovich Depth

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

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

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

[Cherzontukov, Galichon, Hallin, Henry]
Let $\text{Prob}_p(\mathbb{R}^d) = \{\mu \in \text{Prob}(\mathbb{R}^d) \mid \int \|x\|^p \, d\mu < +\infty\}$.

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

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

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

**Wasserstein space**
Let \( \text{Prob}_p(\mathbb{R}^d) = \{ \mu \in \text{Prob}(\mathbb{R}^d) \mid \int \|x\|^p \, d\mu < +\infty \} \).

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

\[
W_p(\mu, \nu) = \left( \min_{\gamma \in \Gamma(\mu, \nu)} \|x - y\|^p \, d\gamma(x, y) \right)^{1/p}.
\]

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

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

i.e. \( \lim_{n \to +\infty} W_p(\mu_n, \mu) = 0 \iff \forall \phi \in C^0(X), \lim_{n \to +\infty} \int \phi \, d\mu_n = \int \phi \, d\mu. \)
Let \( \text{Prob}_p(\mathbb{R}^d) = \{ \mu \in \text{Prob}(\mathbb{R}^d) \mid \int \|x\|^p \, d\mu < +\infty \} \).

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

\[
W_p(\mu, \nu) = \left( \min_{\gamma \in \Gamma(\mu, \nu)} \|x - y\|^p \, d\gamma(x, y) \right)^{1/p}.
\]

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

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

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

On \( \text{Prob}(\mathbb{R}) \), any **monotone** coupling \( \gamma \) between \( \mu, \nu \) is optimal in the def of \( W_p \).
Let \( \text{Prob}_p(\mathbb{R}^d) = \{ \mu \in \text{Prob}(\mathbb{R}^d) \mid \int \|x\|^p \, d\mu < +\infty \} \).

\[ p\text{-Wasserstein distance} \text{ between } \mu, \nu \in \text{Prob}_p(\mathbb{R}^d): \]
\[ W_p(\mu, \nu) = \left( \min_{\gamma \in \Gamma(\mu, \nu)} \int \|x - y\|^p \, d\gamma(x, y) \right)^{1/p}. \]

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

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

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

On \( \text{Prob}(\mathbb{R}) \), any monotone coupling \( \gamma \) between \( \mu, \nu \) is optimal in the def of \( W_p \).

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

\[ W_p(\mu, \nu) = \left( \int_{[0,1]} \|T_\mu(t) - T_\nu(t)\|^p \, dt \right) = \|T_\mu - T_\nu\|_{L^p([0,1])}. \]
Let \( \text{Prob}_p(\mathbb{R}^d) = \{ \mu \in \text{Prob}(\mathbb{R}^d) \mid \int \|x\|^p \, d\mu < +\infty \} \).

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

\[
W_p(\mu, \nu) = \left( \min_{\gamma \in \Gamma(\mu, \nu)} \|x - y\|^p \, d\gamma(x, y) \right)^{1/p}.
\]

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

On \( \text{Prob}(X) \), with \( X \subseteq \mathbb{R}^d \) compact, \( W_p \) metrizes \textbf{narrow convergence}

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

On \( \text{Prob}(\mathbb{R}) \), any \textbf{monotone} coupling \( \gamma \) between \( \mu, \nu \) is optimal in the def of \( W_p \).

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

\[
W_p(\mu, \nu) = \left( \int_{[0,1]} \|T_\mu(t) - T_\nu(t)\|^p \, dt \right) = \|T_\mu - T_\nu\|_{L^p([0,1])}.
\]

In particular, \( (\text{Prob}_p(\mathbb{R}), W_p) \) embeds isometrically in \( L^p([0, 1]) \)!
Let $\text{Prob}_p(\mathbb{R}^d) = \{\mu \in \text{Prob}(\mathbb{R}^d) \mid \int \|x\|^p \, d\mu < +\infty\}$.

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

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

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

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

On $\text{Prob}(\mathbb{R})$, any **monotone** coupling $\gamma$ between $\mu, \nu$ is optimal in the def of $W_p$. For instance $\gamma := (T\mu, T\nu) \# \rho$ with $\rho = \text{Lebesgue on } [0, 1]$ is monotone, implying

$$W_p(\mu, \nu) = \left( \int_{[0,1]} \|T\mu(t) - T\nu(t)\|^p \, dt \right) = \|T\mu - T\nu\|_{L^p([0,1])}.$$  

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

The previous embedding is false in higher dimension: $\left(\text{Prob}_p, W_p\right)$ is **curved**.
Motivation 2: "Linearization" of $W_2$
Motivation 2: "Linearization" of $\mathcal{W}_2$

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

Given $\mu \in \text{Prob}_2(\mathbb{R}^d)$, we define $T_\mu$ 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, $\rho = \text{Leb}_X$ with $X \subseteq \mathbb{R}^d$ convex compact with $|X| = 1$.

Given $\mu \in \text{Prob}_2(\mathbb{R}^d)$, we define $T_\mu$ 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 $\mu \in \text{Prob}_2(\mathbb{R}^d) \to T_\mu \in L^2(X)$ is an injective map, with image the space of (square-integrable) gradients of convex functions on $X$. 

6 - 3
Motivation 2: "Linearization" of \( W_2 \)

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

\[
\text{Given } \mu \in \text{Prob}_2(\mathbb{R}^d), \text{ we define } T_\mu \text{ as the unique map satisfying }
\]

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

(ii) \( T_\mu \# \rho = \mu \).

- The map \( \mu \in \text{Prob}_2(\mathbb{R}^d) \rightarrow T_\mu \in L^2(X) \) is an injective map, with image the space of (square-integrable) gradients of convex functions on \( X \).

\[
W_{2,\rho}(\mu, \nu) := \|T_\mu - T_\nu\|_{L^2(\rho)} \rightarrow [\text{Ambrosio, Gigli, Savaré '04}]
\]

<table>
<thead>
<tr>
<th>Riemannian geometry</th>
<th>Optimal transport</th>
</tr>
</thead>
<tbody>
<tr>
<td>point ( x \in M )</td>
<td>( \mu \in \text{Prob}_2(\mathbb{R}^d) )</td>
</tr>
<tr>
<td>geodesic distance ( d_g(x, y) )</td>
<td>( W_2(\mu, \nu) )</td>
</tr>
<tr>
<td>tangent space ( T_\rho M )</td>
<td>( T_\rho \text{Prob}_2(\mathbb{R}^d) \subseteq L^2(\rho, X) )</td>
</tr>
<tr>
<td>inverse exponential map ( \exp^{-1}<em>\rho(x) \in T</em>\rho M )</td>
<td>( T_\mu \in T_\rho \text{Prob}_2(X) )</td>
</tr>
<tr>
<td>distance in tangent space ( | \exp^{-1}<em>\rho(x) - \exp^{-1}</em>\rho(y) |_{g(x_0)} )</td>
<td>( |T_\mu - T_\nu|_{L^2(\rho)} )</td>
</tr>
</tbody>
</table>
Motivation 2: ”Linearization” of $W_2$

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

Given $\mu \in \text{Prob}_2(\mathbb{R}^d)$, we define $T_\mu$ 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 $\mu \in \text{Prob}_2(\mathbb{R}^d) \rightarrow T_\mu \in L^2(X)$ is an injective map, with image
the space of (square-integrable) gradients of convex functions on $X$.

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

<table>
<thead>
<tr>
<th>Riemannian geometry</th>
<th>Optimal transport</th>
</tr>
</thead>
<tbody>
<tr>
<td>point</td>
<td>$x \in M$</td>
</tr>
<tr>
<td>geodesic distance</td>
<td>$d_g(x, y)$</td>
</tr>
<tr>
<td>tangent space</td>
<td>$T_\rho M$</td>
</tr>
<tr>
<td>inverse exponential</td>
<td>$\exp^{-1}<em>\rho (x) \in T</em>\rho M$</td>
</tr>
<tr>
<td>map in tangent space</td>
<td>$|\exp^{-1}<em>\rho (x) - \exp^{-1}</em>\rho (y)|_{g(x_0)}$</td>
</tr>
<tr>
<td>distance in tangent</td>
<td>$T_\rho \text{Prob}_2(\mathbb{R}^d) \subseteq L^2(\rho, X)$</td>
</tr>
<tr>
<td>space</td>
<td>$T_\mu \in T_\rho \text{Prob}_2(X)$</td>
</tr>
<tr>
<td></td>
<td>$|T_\mu - T_\nu|_{L^2(\rho)}$</td>
</tr>
</tbody>
</table>

- Used in image analysis $\rightarrow$ [Wang, Slepcev, Basu, Ozolek, Rohde '13]
Motivation 2: "Linearization" of $W_2$

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

  Given $\mu \in \text{Prob}_2(\mathbb{R}^d)$, we define $T_\mu$ 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 $\mu \in \text{Prob}_2(\mathbb{R}^d) \to T_\mu \in L^2(X)$ is an injective map, with image the space of (square-integrable) gradients of convex functions on $X$.

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

<table>
<thead>
<tr>
<th>Riemannian geometry</th>
<th>Optimal transport</th>
</tr>
</thead>
<tbody>
<tr>
<td>point</td>
<td>$\mu \in \text{Prob}_2(\mathbb{R}^d)$</td>
</tr>
<tr>
<td>geodesic distance</td>
<td>$W_2(\mu, \nu)$</td>
</tr>
<tr>
<td>tangent space</td>
<td>$T_\rho M$</td>
</tr>
<tr>
<td>inverse exponential map</td>
<td>$T_\rho \text{Prob}_2(\mathbb{R}^d) \subseteq L^2(\rho, X)$</td>
</tr>
<tr>
<td>distance in tangent space</td>
<td>$T_\mu \in T_\rho \text{Prob}_2(X)$</td>
</tr>
</tbody>
</table>

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

- Representing family of probability measures by family of functions in $L^2(\rho)$.
Example: barycenter computation

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

\[
\mu := \arg \min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i W_2^2(\mu, \mu_i).
\]
Example: barycenter computation

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

\[
\mu := \arg\min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i W_2^2(\mu, \mu_i).
\]

→ Need to solve an optimisation problem every time the coefficients \( \alpha_i \) are changed.
Example: barycenter computation

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

\[
\mu := \arg \min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i \ W_2^2(\mu, \mu_i).
\]

\[\rightarrow\] Need to solve an optimisation problem every time the coefficients \( \alpha_i \) are changed.

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

\[\rightarrow\] Simple expression once the transport maps \( T_{\mu_i} : \rho \rightarrow \mu_i \) have been computed.
Example: barycenter computation

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

  \[
  \mu := \arg\min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i \ W_2^2(\mu, \mu_i).
  \]

  → Need to solve an optimisation problem every time the coefficients \( \alpha_i \) are changed.

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

  → Simple expression once the transport maps \( T_{\mu_i} : \rho \to \mu_i \) have been computed.

coeff = [0.2, 0.8]

\( \text{spt}(\mu_0) \)

\( \text{spt}(\mu_1) \)

\((0.8T_{\mu_1} + 0.2T_{\mu_0}) \# \rho\)
Example: barycenter computation

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

$$\mu := \arg\min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i W_2^2(\mu, \mu_i).$$

$\rightarrow$ Need to solve an optimisation problem every time the coefficients $\alpha_i$ are changed.

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

$\rightarrow$ Simple expression once the transport maps $T_{\mu_i} : \rho \rightarrow \mu_i$ have been computed.
Example: barycenter computation

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

\[
\mu := \arg\min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i W_2^2(\mu, \mu_i).
\]

\( \rightarrow \) Need to solve an optimisation problem every time the coefficients \( \alpha_i \) are changed.

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

\( \rightarrow \) Simple expression once the transport maps \( T_{\mu_i} : \rho \rightarrow \mu_i \) have been computed.
Example: barycenter computation

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

  \[
  \mu := \arg \min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i W_2^2(\mu, \mu_i).
  \]

  \( \rightarrow \) Need to solve an optimisation problem every time the coefficients \( \alpha_i \) are changed.

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

  \( \rightarrow \) Simple expression once the transport maps \( T_{\mu_i} : \rho \rightarrow \mu_i \) have been computed.
Example: barycenter computation

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

\[
\mu := \arg\min_{1 \leq i \leq k} \sum_{1 \leq i \leq k} \alpha_i W_2^2(\mu, \mu_i).
\]

\( \rightarrow \) Need to solve an optimisation problem every time the coefficients \( \alpha_i \) are changed.

”Linearized” Wasserstein barycenters: \( \mu := \left( \frac{1}{\sum_i \alpha_i} \sum_i \alpha_i T_{\mu_i} \right) \# \rho. \)

\( \rightarrow \) Simple expression once the transport maps \( T_{\mu_i} : \rho \rightarrow \mu_i \) have been computed.

What amount of the Wasserstein geometry is preserved by the embedding \( \mu \mapsto T_\mu? \)
Motivation 3: numerical analysis of optimal transport

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

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

- Approximate $\mu$ 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 $[(i_1 - 1)/k, i_1/k] \times \ldots \times [(i_d - 1)/k, i_d/k]$. 


Motivation 3: numerical analysis of optimal transport

**Theorem (Brenier, McCann)** Given $\rho \in \text{Prob}^{ac}(\mathbb{R}^d)$ and $\mu \in \text{Prob}(\mathbb{R}^d)$,

$\exists! \rho$-a.e. $T_\mu : \mathbb{R}^d \to \mathbb{R}^d$ such that $T_\mu # \rho = \mu$ and $T_\mu = \nabla \phi$ with $\phi$ convex.

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

- Approximate $\mu$ 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 $[(i_1 - 1)/k, i_1/k] \times \ldots \times [(i_d - 1)/k, i_d/k]$

(Then, $W_p(\mu_k, \mu) \lesssim \frac{1}{k}$.)
Motivation 3: numerical analysis of optimal transport

**Theorem (Brenier, McCann)** Given $\rho \in \text{Prob}^{ac}(\mathbb{R}^d)$ and $\mu \in \text{Prob}(\mathbb{R}^d)$, there exists a unique $\rho$-a.e. $T_\mu : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $T_\mu \# \rho = \mu$ and $T_\mu = \nabla \phi$ with $\phi$ convex.

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

- Approximate $\mu$ 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, $W_p(\mu_k, \mu) \lesssim 1/k$.)

- Compute *exactly* the optimal transport plan $T_{\mu_k}$ between $\rho$ and $\mu_k$, (using a semi-discrete optimal transport solver).
Motivation 3: numerical analysis of optimal transport

**Theorem (Brenier, McCann)** Given \( \rho \in \text{Prob}^{ac}(\mathbb{R}^d) \) and \( \mu \in \text{Prob}(\mathbb{R}^d) \),

\[ \exists! \, \rho \text{-a.e. } T_\mu : \mathbb{R}^d \to \mathbb{R}^d \text{ such that } T_\mu \# \rho = \mu \text{ and } T_\mu = \nabla \phi \text{ with } \phi \text{ convex}. \]

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

- **Approximate** \( \mu \) by a discrete measure, for instance
  
  \[ \mu_k = \sum_{i_1 \leq \ldots \leq i_k} \mu(B_{i_1, \ldots, i_k}) \delta(\frac{i_1}{k}, \ldots, \frac{i_k}{k}) \]
  
  where \( B_{i_1, \ldots, i_k} \) is the cube \([\frac{i_1 - 1}{k}, \frac{i_1}{k}] \times \ldots \times [\frac{i_d - 1}{k}, \frac{i_d}{k}]\]

  (Then, \( W_p(\mu_k, \mu) \lesssim \frac{1}{k} \).)

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

It is know that \( T_{\mu_k} \) converges to \( T_\mu \) but convergence rates are unknown in general...
Motivation 3: numerical analysis of optimal transport

Theorem (Brenier, McCann) Given $\rho \in \text{Prob}^{ac}(\mathbb{R}^d)$ and $\mu \in \text{Prob}(\mathbb{R}^d)$, 

$\exists! \ \rho$-a.e. $T_\mu : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $T_\mu \# \rho = \mu$ and $T_\mu = \nabla \phi$ with $\phi$ convex.

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

- Approximate $\mu$ 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 $[(i_1 - 1)/k, i_1/k] \times \ldots \times [(i_d - 1)/k, i_d/k]$

(Then, $W_p(\mu_k, \mu) \lesssim \frac{1}{k}$.)

- Compute exactly the optimal transport plan $T_{\mu_k}$ between $\rho$ and $\mu_k$, (using a \textbf{semi-discrete} optimal transport solver).

It is known that $T_{\mu_k}$ converges to $T_\mu$ 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 \( \rho \in \text{Prob}^{ac}(\mathbb{R}^d) \) and \( \mu \in \text{Prob}(\mathbb{R}^d) \),

\[ \exists! \rho \text{-a.e. } T_\mu : \mathbb{R}^d \to \mathbb{R}^d \text{ such that } T_\mu \# \rho = \mu \text{ and } T_\mu = \nabla \phi \text{ with } \phi \text{ convex.} \]

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

- Approximate \( \mu \) 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, \( W_p(\mu_k, \mu) \lesssim \frac{1}{k} \).)

- Compute exactly the optimal transport plan \( T_{\mu_k} \) between \( \rho \) and \( \mu_k \),

(\text{using a } \text{semi-discrete } \text{optimal transport solver}).

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

Indeed: since \( T_\mu \# \rho = \mu \) and \( T_\nu \# \rho = \nu \), one has \( \gamma := (T_\mu, T_\nu) \# \rho \in \Gamma(\mu, \nu) \).
The map $\mu \mapsto T_\mu$ is reverse-Lipschitz, i.e. $\|T_\mu - T_\nu\|_{L^2(\rho)} \geq W_2(\mu, \nu)$.

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

Thus, $W_2^2(\mu, \nu) \leq \int \|x - y\|^2 \, d\gamma(x, y) = \int \|T_\mu(x) - T_\nu(x)\|^2 \, d\rho(x)$. 

\[ \text{Elementary remarks} \]
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: since $T_\mu \# \rho = \mu$ and $T_\nu \# \rho = \nu$, one has $\gamma := (T_\mu, T_\nu) \# \rho \in \Gamma(\mu, \nu)$.

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

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

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

Thus, $W_2^2(\mu, \nu) \leq \int \|x - y\|^2 \, \mathrm{d} \gamma(x, y) = \int \|T_\mu(x) - T_\nu(x)\|^2 \, \mathrm{d} \rho(x)$.

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. $\|T_\mu - T_\nu\|_{L^2(\rho)} \geq W_2(\mu, \nu)$.

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

  Thus, $W_2^2(\mu, \nu) \leq \int \|x - y\|^2 \, d \gamma(x, y) = \int \|T_\mu(x) - T_\nu(x)\|^2 \, d \rho(x)$.

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

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

  Take $\rho = \frac{1}{\pi} \text{Leb}_{B(0,1)}$ 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. $\|T_\mu - T_\nu\|_{L^2(\rho)} \geq W_2(\mu, \nu)$.

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

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

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

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

Take $\rho = \frac{1}{\pi} \text{Leb}_{B(0,1)}$ 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 $\|T_{\mu_\theta} - T_{\mu_{\theta+\delta}}\|_{L^2(\rho)}^2 \geq C\delta$

Since on the other hand, $W_2(\mu_\theta, \mu_{\theta+\delta}) \leq C\delta$,

$\|T_{\mu_\theta} - T_{\mu_{\theta+\delta}}\|_{L^2(\rho)} \geq C W_2(\mu_\theta, \mu_{\theta+\delta})^{1/2}$
**Local \( \frac{1}{2} \)-Hölder continuity**

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

If \( T_\mu \) is \( L \)-Lipschitz, then

\[
\|T_\mu - T_\nu\|_2^2 \leq C W_1(\mu, \nu) \quad \text{with} \quad C = 4L \text{diam}(X).
\]
Local $\frac{1}{2}$-Hölder continuity

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

If $T_\mu$ is $L$-Lipschitz, then $\|T_\mu - T_\nu\|_2^2 \leq C W_1(\mu, \nu)$ with $C = 4L \text{diam}(X)$.

$\Rightarrow [\text{Ambrosio, Gigli '09}]$ with slightly better upper bound. See also [Berman '18].
Local $\frac{1}{2}$-Hölder continuity

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

If $T_\mu$ is $L$-Lipschitz, then $\|T_\mu - T_\nu\|_2^2 \leq C W_1(\mu, \nu)$ with $C = 4L \text{diam}(X)$.

- $\simeq$ [Ambrosio,Gigli ’09] with slightly better upper bound. See also [Berman ’18].
- No regularity assumption on $\nu$ $\longrightarrow$ consequences in statistics and numerical analysis.
Thm: Assume $\rho \in \text{Prob}^{\text{ac}}(X)$ and $\mu, \nu \in \text{Prob}(Y)$ with $X, Y \subseteq \mathbb{R}^d$ compact.

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

- $\simeq$ [Ambrosio, Gigli '09] with slightly better upper bound. See also [Berman '18].
- No regularity assumption on $\nu \rightarrow$ consequences in statistics and numerical analysis.
- Let $\phi_\mu : X \rightarrow \mathbb{R}$ convex s.t. $T_\mu = \nabla \phi_\mu$.
  \begin{align*}
  \psi_\mu : Y \rightarrow \mathbb{R} \text{ its Legendre transform: } \psi_\mu(y) &= \max_{x \in X} \langle x | y \rangle - \phi_\mu(x)
  \end{align*}
Local $\frac{1}{2}$-Hölder continuity

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

If $T_\mu$ is $L$-Lipschitz, then $\|T_\mu - T_\nu\|^2_{L^2(\rho)} \leq C W_1(\mu, \nu)$ with $C = 4L \text{diam}(X)$.

$\simeq$ [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 $\phi_\mu : X \rightarrow \mathbb{R}$ convex s.t. $T_\mu = \nabla \phi_\mu$.

$\psi_\mu : Y \rightarrow \mathbb{R}$ its Legendre transform:

$\psi_\mu(y) = \max_{x \in X} \langle x | y \rangle - \phi_\mu(x)$

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

$\simeq$ [Ambrosio,Gigli '09] with slightly better upper bound. See also [Berman '18].

$\longrightarrow$ consequences in statistics and numerical analysis.
Local $\frac{1}{2}$-Hölder continuity

**Thm:** Assume $\rho \in \text{Prob}^{\text{ac}}(X)$ and $\mu, \nu \in \text{Prob}(Y)$ with $X, Y \subseteq \mathbb{R}^d$ compact. If $T_\mu$ is $L$-Lipschitz, then $\|T_\mu - T_\nu\|_2^2 \leq C W_1(\mu, \nu)$ with $C = 4L \text{diam}(X)$.

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

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

- **Prop**$\implies$ **Thm:** Kantorovich-Rubinstein theorem
Local $\frac{1}{2}$-Hölder continuity

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

If $T_\mu$ is $L$-Lipschitz, then $\|T_\mu - T_\nu\|_2^2 \leq C W_1(\mu, \nu)$ with $C = 4L \text{diam}(X)$.

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

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

- $\int \psi_\nu d(\mu - \nu) = \int \psi_\nu d(\nabla \phi_\mu # \rho - \nabla \phi_\nu # \rho) = \int \psi_\nu (\nabla \phi_\mu) - \psi_\nu (\nabla \phi_\nu) d \rho$
Local $\frac{1}{2}$-Hölder continuity

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

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

\[\int \psi_\nu \, d(\mu - \nu) = \int \psi_\nu \, d(\nabla \phi_\mu \# \rho - \nabla \phi_\nu \# \rho) = \int \psi_\nu(\nabla \phi_\mu) - \psi_\nu(\nabla \phi_\nu) \, d\rho\]

(convexity: $\psi_\nu(y) - \psi_\nu(x) \geq \langle y - x | \nabla \psi_\nu(x) \rangle$)

\[\geq \int \langle \nabla \psi_\mu - \nabla \psi_\nu | \nabla \psi_\nu(\nabla \phi_\nu) \rangle \, d\rho\]
Local $\frac{1}{2}$-Hölder continuity

**Thm:** Assume $\rho \in \text{Prob}^{ac}(X)$ and $\mu, \nu \in \text{Prob}(Y)$ with $X, Y \subseteq \mathbb{R}^d$ compact. If $T_\mu$ is $L$-Lipschitz, then $\|T_\mu - T_\nu\|_{L^2(\rho)}^2 \leq C W_1(\mu, \nu)$ with $C = 4L \text{diam}(X)$.

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

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

\[
\int \psi_\nu \, d(\mu - \nu) = \int \psi_\nu \, d(\nabla \phi_\mu \# \rho - \nabla \phi_\nu \# \rho) = \int \psi_\nu (\nabla \phi_\mu) - \psi_\nu (\nabla \phi_\nu) \, d \rho
\]

(convexity: $\psi_\nu(y) - \psi_\nu(x) \geq \langle y - x | \nabla \psi_\nu(x) \rangle$) \[\geq \int \langle \nabla \psi_\mu - \nabla \psi_\nu | \nabla \psi_\nu (\nabla \phi_\nu) \rangle \, d \rho \]

\[= \int \langle \nabla \psi_\mu - \nabla \psi_\nu | \text{id} \rangle \, d \rho\]
Local $\frac{1}{2}$-Hölder continuity

**Thm:** Assume $\rho \in \text{Prob}^{\text{ac}}(X)$ and $\mu, \nu \in \text{Prob}(Y)$ with $X, Y \subseteq \mathbb{R}^d$ compact. If $T_\mu$ is $L$-Lipschitz, then $\|T_\mu - T_\nu\|_2^2 \leq C W_1(\mu, \nu)$ with $C = 4L \text{diam}(X)$.

$\simeq$ [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: $\psi_\mu(y) = \max_{x \in X} \langle x | y \rangle - \phi_\mu(x)$

**Prop:** If $T_\mu$ is $L$-Lipschitz, then $\|T_\mu - T_\nu\|_2^2 \leq -2L \int (\psi_\mu - \psi_\nu) d(\mu - \nu)$.

\[ \int \psi_\nu d(\mu - \nu) = \int \psi_\nu d(\nabla \phi_\mu \# \rho - \nabla \phi_\nu \# \rho) = \int \psi_\nu (\nabla \phi_\mu) - \psi_\nu (\nabla \phi_\nu) d \rho \]

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

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

\[ \int \psi_\mu d(\nu - \mu) \geq \int \langle \nabla \psi_\nu - \nabla \psi_\mu | \text{id} \rangle d \rho + \frac{L}{2} \| \nabla \phi_\mu - \nabla \phi_\nu \|_{L^2(\rho)}^2 \]

($T_\mu = \nabla \phi_\mu$ $L$-Lipschitz $\iff$ $\psi_\mu = \phi^*_\mu$ is $L$-strongly convex)
Thm (Berman, ’18): Let $\rho \in \text{Prob}^{ac}(X)$ and $\mu, \nu \in \text{Prob}(Y)$ with $X,Y$ compact. Then, $\|\nabla \psi_\mu - \nabla \psi_\nu\|_{L^2(Y)}^2 \leq C W_1(\mu, \nu)^\alpha$ with $\alpha = \frac{1}{2^{d-1}}$. 
**Global Hölder continuity**

**Thm (Berman, ’18):** Let $\rho \in \text{Prob}^{ac}(X)$ and $\mu, \nu \in \text{Prob}(Y)$ with $X, Y$ compact. Then, $\|\nabla \psi_\mu - \nabla \psi_\nu\|_{L^2(Y)}^2 \leq C W_1(\mu, \nu)\alpha$ with $\alpha = \frac{1}{2d-1}$

**Corollary:** $\|T_\mu - T_\nu\|_{L^2(\rho)}^2 \leq C W_1(\mu, \nu)\alpha$ with $\alpha = \frac{1}{2d-1(d+2)}$
**Global Hölder continuity**

**Thm (Berman, ’18):** Let $\rho \in \text{Prob}^{ac}(X)$ and $\mu, \nu \in \text{Prob}(Y)$ with $X, Y$ compact. Then, $\|\nabla \psi_\mu - \nabla \psi_\nu\|_{L^2(Y)}^2 \leq C W_1(\mu, \nu)^{\alpha}$ with $\alpha = \frac{1}{2d-1}$

**Corollary:** $\|T_\mu - T_\nu\|_{L^2(\rho)}^2 \leq C W_1(\mu, \nu)^{\alpha}$ with $\alpha = \frac{1}{2d-1(d+2)}$

- The Hölder exponent is terrible, but inequality holds without assumptions on $\mu, \nu$!
Global Hölder continuity

Thm (Berman, ’18): Let $\rho \in \text{Prob}^{ac}(X)$ and $\mu, \nu \in \text{Prob}(Y)$ with $X, Y$ compact.

Then, $\|\nabla \psi_\mu - \nabla \psi_\nu\|_{L^2(Y)}^2 \leq C W_1(\mu, \nu)^\alpha$ with $\alpha = \frac{1}{2d-1}$

Corollary: $\|T_\mu - T_\nu\|_{L^2(\rho)}^2 \leq C W_1(\mu, \nu)^\alpha$ with $\alpha = \frac{1}{2d-1(d+2)}$

- The Hölder exponent is terrible, but inequality holds without assumptions on $\mu, \nu$!
- 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 $\rho = \text{Leb}_X$, and let $Y$ be compact. Then, there exists $C$ s.t. for all $\mu, \nu \in \text{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$
Main theorem

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

$$||T_\mu - T_\nu||_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$ 

- 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 $\rho = \text{Leb}_X$, and let $Y$ be compact. Then, there exists $C$ s.t. for all $\mu, \nu \in \text{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$  

- 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 $\rho = \text{Leb}_X$, and let $Y$ be compact. Then, there exists $C$ s.t. for all $\mu, \nu \in \text{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$

- 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 $\text{diam}(X), \text{diam}(Y)$. 

Main theorem

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

$$\|T_{\mu} - T_{\nu}\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$  

- 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 $\text{diam}(X), \text{diam}(Y)$.
- Proof relies on the semidiscrete setting, i.e. the bound is established in the case
  $$\mu = \sum \mu_i \delta_{y_i}, \nu = \sum \nu_i \delta_{y_i}.$$  
  and one concludes using a density argument.
Semidiscrete OT for $c(x, y) = -\langle x|y \rangle$

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

$$\mathcal{T}(\rho, \mu) = \max_{\gamma \in \Gamma(\rho, \mu)} \int \langle x|y \rangle \, d\gamma(x, y)$$
Semidiscete OT for $c(x, y) = -\langle x | y \rangle$

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

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

Kantorovich duality

$$= \min_{\phi \oplus \psi \succeq \langle \cdot | \cdot \rangle} \int \phi \, d\rho + \int \psi \, d\mu$$
Semidiscrete OT for $c(x, y) = -\langle x | y \rangle$

Let $\rho, \nu \in \text{Prob}_{1}^{ac}(\mathbb{R}^d)$ and $\Gamma(\rho, \mu) =$ couplings between $\rho, \mu$,

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

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

Legendre-Fenchel transform:

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

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

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

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

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

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

Kantorovich duality

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

Legendre-Fenchel transform:

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

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

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

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

\[
V_i(\psi) = \{ x \mid \forall j, \langle x | y_i \rangle - \psi_i \geq \langle x | y_j \rangle - \psi_j \}
\]
Semidiscrete OT for $c(x, y) = -\langle x | y \rangle$

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

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

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

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

Kantorovich duality

- 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

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

Legendre-Fenchel transform:

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

Thus, $\mathcal{T}(\rho, \mu) = \min_{\psi \in \mathbb{R}^N} \sum_i \int_{V_i(\psi)} \langle x | y_i \rangle - \psi_i \, d \rho(x) + \sum_i \mu_i \psi_i$
\[ T(\rho, \mu) = \min_{\psi \in \mathbb{R}^N} \Phi(\psi) - \sum_{i} \mu_i \psi_i, \text{ where:} \]
\[ \Phi(\psi) := \sum_{i} \int_{V_i(\psi)} \langle x | y_i \rangle - \psi_i \, d\rho(x) \]
Optimality condition and economic interpretation

\[ \mathcal{T}(\rho, \mu) = \min_{\psi \in \mathbb{R}^N} \Phi(\psi) - \sum_i \mu_i \psi_i, \text{ where:} \]

\[ \Phi(\psi) := \sum_i \int V_i(\psi) \langle x | y_i \rangle - \psi_i \, d\rho(x) \]

- Gradient: \[ \nabla \Phi(\psi) = -(G_i(\psi))_{1 \leq i \leq N} \text{ where } G_i(\psi) = \rho(V_i(\psi)). \]
Optimality condition and economic interpretation

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

\[ \nabla \Phi(\psi) = -(G_i(\psi))_{1 \leq i \leq N} \quad \text{where} \quad G_i(\psi) = \rho(V_i(\psi)). \]

\[ \psi \in \mathbb{R}^N \text{ is a minimizer of dual pb} \iff \forall i, \rho(V_i(\psi)) = \mu_i \]
Optimality condition and economic interpretation

\[ T(\rho, \mu) = \min_{\psi \in \mathbb{R}^N} \Phi(\psi) - \sum_i \mu_i \psi_i, \text{ where:} \]

\[ \Phi(\psi) := \sum_i \int V_i(\psi) \langle x | y_i \rangle - \psi_i \, d\rho(x) \]

- **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 \( G = (G_1, \ldots, G_N), \mu \in \mathbb{R}^N \)
Optimality condition and economic interpretation

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

▶ 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 \( \rho \) onto \( \sum_i \mu_i \delta_{y_i} \)
Optimality condition and economic interpretation

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

- **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 \( \rho \) onto \( \sum_i \mu_i \delta_{y_i} \)

- **Economic interpretation**: \( \rho = \text{density of customers}, \{y_i\}_{1 \leq i \leq N} = \text{product types} \)
Optimality condition and economic interpretation

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

- **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 \( \rho \) onto \( \sum_i \mu_i \delta_{y_i} \)

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

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

- **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 \( \forall i, \rho(V_i(\psi)) = \mu_i \)

\[ \iff G(\psi) = \mu \text{ with } G = (G_1, \ldots, G_N), \mu \in \mathbb{R}^N \]
\[ \iff T = \nabla \psi^* \text{ transports } \rho \text{ onto } \sum_i \mu_i \delta_{y_i} \]

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

\( \rightarrow \) given prices \( \psi \in \mathbb{R}^N \), a customer \( x \) maximizes \( \langle x|y_i \rangle - \psi_i \) over all products.
\( \rightarrow \) \( V_i(\psi) = \{x \mid i \in \arg \max_j \langle x|y_j \rangle - \psi_j\} = \) customers choosing product \( y_i \).
Optimality condition and economic interpretation

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

**Gradient:** \[ \nabla \Phi(\psi) = -(G_i(\psi))_{1 \leq i \leq N} \text{ where } G_i(\psi) = \rho(V_i(\psi)). \]

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

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

\( \iff T = \nabla \psi^* \text{ transports } \rho \text{ onto } \sum_i \mu_i \delta_{y_i} \)

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

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

\( \rightarrow V_i(\psi) = \{x | i \in \text{arg max}_j \langle x|y_j \rangle - \psi_j \} = \) customers choosing product \( y_i \).

\( \rightarrow \rho(V_i) = \) amount of customers for product \( y_i \).
Optimality condition and economic interpretation

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

> **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 \( \iff \forall i, \rho(V_i(\psi)) = \mu_i \)
\[ \iff G(\psi) = \mu \text{ with } G = (G_1, \ldots, G_N), \mu \in \mathbb{R}^N \]
\[ \iff T = \nabla \psi^* \text{ transports } \rho \text{ onto } \sum_i \mu_i \delta_{y_i} \]

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

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

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

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

Optimal transport = finding prices satisfying capacity constraints \( \rho(V_i(\psi)) = \mu_i \).
Optimality condition and economic interpretation

\[ T(\rho, \mu) = \min_{\psi \in \mathbb{R}^N} \Phi(\psi) - \sum_i \mu_i \psi_i, \text{ where:} \]

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

\[ \textbf{Gradient:} \quad \nabla \Phi(\psi) = -(G_i(\psi))_{1 \leq i \leq N} \text{ where } G_i(\psi) = \rho(V_i(\psi)). \]

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

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

\[ \iff T = \nabla \psi^* \text{ transports } \rho \text{ onto } \sum_i \mu_i \delta_{y_i} \]

\[ \textbf{Economic interpretation:} \quad \rho = \text{density of customers}, \{y_i\}_{1 \leq i \leq N} = \text{product types} \]

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

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

\[ \rightarrow \rho(V_i) = \text{amount of customers for product } y_i. \]

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

\[ \textbf{Algorithm (Oliker–Prussner):} \quad \text{coordinate-wise increment. Complexity: } O(N^3). \]
Hessian on $\Phi$ and Newton’s Algorithm

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

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

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x) \, dx$$

where $\Gamma_{ij} = V_i(\psi) \cap V_j(\psi)$.

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

(Recall that $G_i(\psi) = \rho(V_i(\psi))$ and $\nabla \Phi = -(G_1, \ldots, G_N)$)
Hessian on $\Phi$ and Newton’s Algorithm

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

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

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

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

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 $\text{Ker}D^2\Phi(\psi) = \mathbb{R}(1, \ldots, 1)$. 

\[\Gamma_{15}(\psi)\]
Proposition: If $\rho \in C^0(X)$ and $(y_i)_{1 \leq i \leq N}$ is generic, then $\Phi \in C^2(\mathbb{R}^N)$ and

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x) \, dx$$

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

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

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

$$(i, j) \in H \iff L_{ij} > 0$$
Hessian on $\Phi$ and Newton’s Algorithm

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

**Proposition:**

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

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

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

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 $\text{Ker} D^2 \Phi(\psi) = \mathbb{R}(1, \ldots, 1)$.

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

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

- If $\Omega$ is connected and $\psi \in E$, then $H$ is connected.
Hessian on $\Phi$ and Newton’s Algorithm

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

**Proposition:**

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

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x) \, dx$$

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

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

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

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

If $\Omega$ is connected and $\psi \in E$, then $H$ is connected.

$L$ is the Laplacian of a connected graph $\implies \text{Ker} L = \mathbb{R} \cdot \text{cst}$
Hessian on $\Phi$ and Newton’s Algorithm

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

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

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x) \, dx$$

where $\Gamma_{ij} = V_i(\psi) \cap V_j(\psi)$.

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

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

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

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

- If $\Omega$ is connected and $\psi \in E$, then $H$ is connected

- $L$ is the Laplacian of a connected graph $\implies \text{Ker}L = \mathbb{R} \cdot \text{cst}$

**Corollary:** Global convergence of a damped Newton algorithm.

[Kitagawa, M., Thibert 16]
Numerical example

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

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: $\rho = \text{uniform on } [0, 1]^2,$

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

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

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

NB: The points do not move.
Numerical example

Source: \( \rho = \text{uniform on } [0, 1]^2, \)

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

\( \psi_0 = \frac{1}{2} \| \cdot \|^2 \)

\( \psi_1 = \text{Newt}(\psi_0) \)

\( \psi_2 = \text{Newt}(\psi_1) \)

NB: The points do not move.
Numerical example

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

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

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

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

$\psi_2 = \text{Newt}(\psi_1)$

NB: The points do not move.

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

Proof gives a better Hölder exponent ($1/3$ Hölder) for $\mu \mapsto \nu$ (no upper bound).
Thm (M., Delalande, Chazal ’19): Let $X$ convex compact with $|X| = 1$ and $ho = \text{Leb}_X$, and let $Y$ be compact. Then, there exists $C$ s.t. for all $\mu, \nu \in \text{Prob}(Y)$, 

$$\|T_\mu - T_\nu\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$
Proof ingredients

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

$$\|T_\mu - T_\nu\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$ 

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 $\rho = \text{Leb}_X$, and let $Y$ be compact. Then, there exists $C$ s.t. for all $\mu, \nu \in \text{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}. $$

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

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 $\rho = \text{Leb}_X$, and let $Y$ be compact. Then, there exists $C$ s.t. for all $\mu, \nu \in \text{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$

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

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,

$$\langle \mu^1 - \mu^0 | v \rangle = \langle G(\psi^1) - G(\psi^0) | v \rangle = \int_0^1 \langle DG(\psi^t) v | v \rangle \, dt$$
Proof ingredients

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

$$\|T_\mu - T_\nu\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$ 

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,

$$\langle \mu^1 - \mu^0 | v \rangle = \langle G(\psi^1) - G(\psi^0) | v \rangle = \int_0^1 \langle DG(\psi^t) v | v \rangle \, dt$$

a) Control of the eigengap: $\langle DG(\psi^t) v | v \rangle \leq -C(X) \|v\|_{L^2(\mu^t)}^2$ if $\int v \, d\mu_t = 0$.

with $\mu^t = G(\psi^t) \rightarrow$ [Eymard, Gallouët, Herbin ’00].
Proof ingredients

Thm (M., Delalande, Chazal ’19): Let $X$ convex compact with $|X| = 1$ and $\rho = \text{Leb}_X$, and let $Y$ be compact. Then, there exists $C$ s.t. for all $\mu, \nu \in \text{Prob}(Y)$,
$$\|T_\mu - T_\nu\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$  

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,
$$\langle \mu^1 - \mu^0 | v \rangle = \langle G(\psi^1) - G(\psi^0) | v \rangle = \int_0^1 \langle DG(\psi^t) v | v \rangle \, dt$$

a) Control of the eigengap: $\langle DG(\psi^t) v | v \rangle \leq -C(X)\|v\|_{L^2(\mu^t)}^2$ if $\int v \, d\mu_t = 0$.

with $\mu^t = G(\psi^t)$ $\longrightarrow$ [Eymard, Gallouët, Herbin ’00].

b) Control of $\mu_t$: 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 $\rho = \text{Leb}_X$, and let $Y$ be compact. Then, there exists $C$ s.t. for all $\mu, \nu \in \text{Prob}(Y)$,

$$\|T_\mu - T_\nu\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$  

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

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,

$$\langle \mu^1 - \mu^0 | v \rangle = \langle G(\psi^1) - G(\psi^0) | v \rangle = \int_0^1 \langle DG(\psi^t)v | v \rangle \, dt$$

a) **Control of the eigengap:** $\langle DG(\psi^t)v | v \rangle \leq -C(X)\|v\|_{L^2(\mu^t)}^2$ if $\int v \, d\mu^t = 0$.

with $\mu^t = G(\psi^t) \longrightarrow [\text{Eymard, Gallouët, Herbin '00}].$

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

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

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

$$\|T_\mu - T_\nu\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$  

\textbf{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,

$$\langle \mu^1 - \mu^0 | v \rangle = \langle G(\psi^1) - G(\psi^0) | v \rangle = \int_0^1 \langle DG(\psi^t)v | v \rangle \, dt$$

\begin{itemize}
  \item[a)] Control of the eigengap: $\langle DG(\psi^t)v | v \rangle \leq -C(X)\|v\|_{L^2(\mu^t)}^2$ if $\int v \, d\mu_t = 0$.
  \item[b)] Control of $\mu_t$: Brunn-Minkowski’s inequality implies $\mu^t \geq (1 - t)^d \mu^0$.
\end{itemize}

Combining a) and b) we get

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

Then, by Kantorovich-Rubinstein,

$$\leq \text{Lip}(\psi^1 - \psi^0) W_1(\mu^0, \mu_1)$$
Proof ingredients

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

$$\|T_\mu - T_\nu\|_{L^2(X)} \leq C W_2(\mu, \nu)^{1/5}.$$  

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

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,

$$\langle \mu^1 - \mu^0 | v \rangle = \langle G(\psi^1) - G(\psi^0) | v \rangle = \int_0^1 \langle DG(\psi^t)v | v \rangle \, dt$$

a) **Control of the eigengap:** $\langle DG(\psi^t)v | v \rangle \leq -C(X)\|v\|^2_{L^2(\mu^t)}$ if $\int v \, d\mu_t = 0$.

with $\mu^t = G(\psi^t) \longrightarrow$ [Eymard, Gallouët, Herbin ’00].

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

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

Then, by Kantorovich-Rubinstein,

$$\lesssim \text{Lip}(\psi^1 - \psi^0) W_1(\mu^0, \mu_1)$$

$$\lesssim W_2(\mu^0, \mu^1)$$

**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 ($64 \times 64$ pixels) representing digits.
Example: $\kappa$-Means for MNIST digits

MNIST has $M = 60,000$ images grayscale images ($64 \times 64$ 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: $\kappa$-Means for MNIST digits

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

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

$$T^\ell = T_{\mu^\ell} \in L^2([0, 1], \mathbb{R}^2) \quad \text{[OT map from } \rho = \text{Leb}_{[0,1]^2} \text{ to } \mu^\ell]$$
Example: $k$-Means for MNIST digits

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

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

$$T^\ell = T_{\mu^\ell} \in L^2([0, 1], \mathbb{R}^2) \quad \text{[OT map from } \rho = \text{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 ($64 \times 64$ pixels) representing digits. Each image $\alpha^\ell \in \mathcal{M}_{64}(\mathbb{R})$ is transformed into a probability measure on $[0,1]^2$ via

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

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

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

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 \( \text{Prob}(\mathbb{R}^d) \) into \( L^2(\rho, \mathbb{R}^d) \), with possible applications in data analysis. Computations can be easily performed using

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!