Statistical and algorithmic aspects of optimal quantization and sparse Wasserstein barycenters

Léo Portales

09/10/2025

Institut de Recherche
en Informatique de Toulouse
CNRS - INP - UT3 - UT1 - UT2J

Example 1: Optimal quantization

μ (Target Measure)

ν

$\frac{1}{N}\sum_{k=1}^{N}\delta_{y_k}$

$\sum_{k=1}^{N}\pi_k\delta_{y_k}$

Example 1: Optimal quantization

μ (Target Measure) $\quad\quad \nu$

$\frac{1}{N}\sum_{k=1}^{N}\delta_{y_k} \quad\quad \sum_{k=1}^{N}\pi_k\delta_{y_k}$

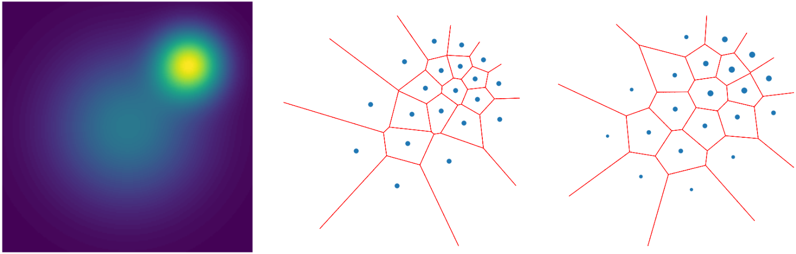


Figure: Optimal and uniform quantization of a Gaussian mixture

Simulations made using the library PyMongeAmpere by Quentin Mérigot.

Example 2: Empirical sparse Wasserstein barycenter

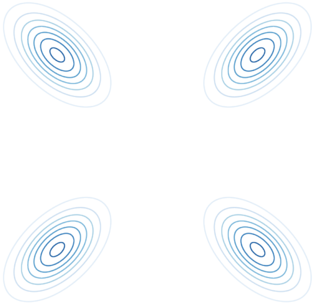


Figure: Wasserstein barycenter and its sample average approximation

Simulations made using the POT Library.

Example 2: Empirical sparse Wasserstein barycenter

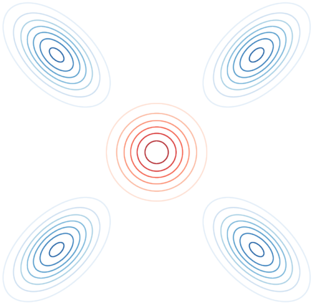


Figure: Wasserstein barycenter and its sample average approximation

Simulations made using the POT Library.

Example 2: Empirical sparse Wasserstein barycenter

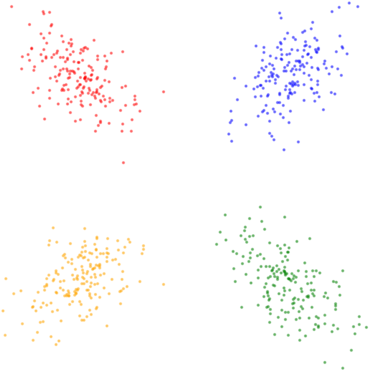


Figure: Wasserstein barycenter and its sample average approximation

Simulations made using the POT Library.

Example 2: Empirical sparse Wasserstein barycenter

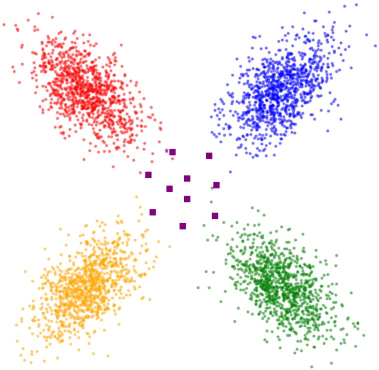
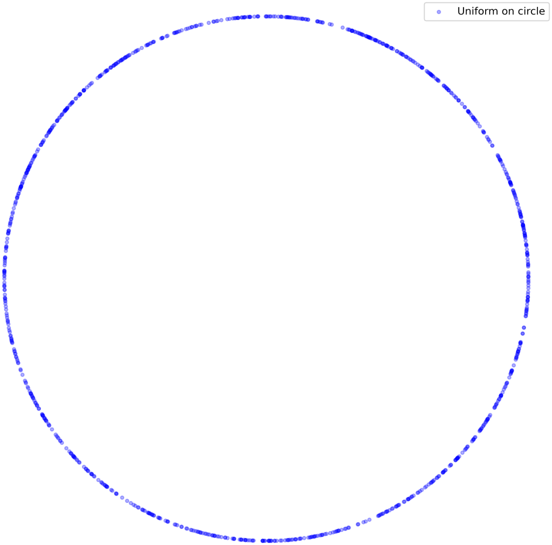


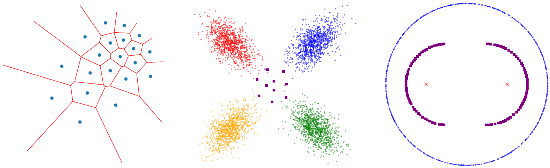
Figure: Wasserstein barycenter and its sample average approximation

Simulations made using the POT Library.

Example 3: Wasserstein geodesics

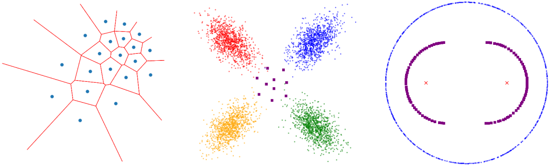


Connection between these examples



What do all these examples have in common?

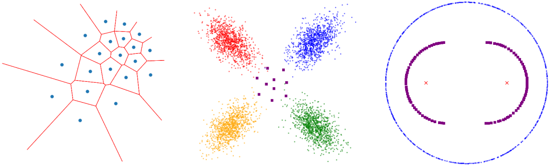
Connection between these examples



What do all these examples have in common?

They are all instances of problems of optimization of point clouds (seen as **probability measures**) and/or their probability weights.

Connection between these examples



What do all these examples have in common?

They are all instances of problems of optimization of point clouds (seen as **probability measures**) and/or their probability weights. More specifically, theses examples aim at solving

$$\min_{(Y,\pi)\in A} \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{D}\left(\mu^\ell, \sum_{i=1}^{N} \pi_i \delta_{y_i}\right)$$

where $\mu^1, \ldots, \mu^L$ are a collection of probability measures, A is a closed, non-empty subset of $(\mathbb{R}^d)^N \times \Delta_N$ and and $\mathbb{D}$ is a divergence between measures.

Objectives

- Define a good notion of metric between arbitrary measures (**discrete** or **continuous**)

Objectives

- Define a good notion of metric between arbitrary measures (**discrete** or **continuous**)
- Study interesting sub cases of

$$\min_{(Y,\pi)\in A} \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{D}\left(\mu^{\ell}, \sum_{i=1}^{N} \pi_i \delta_{y_i}\right)$$

In particular

1. Algorithmic properties
2. Statistical properties.

Optimal transport

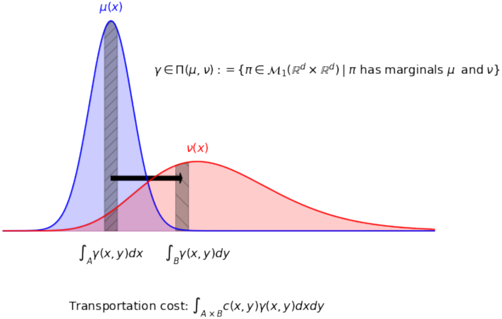
Optimal transport measures the minimal cost of transporting the target measure μ onto ν with couplings γ, according to a cost $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$.

$$\boxed{\mathrm{OT}_c(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma(x, y).}$$

Optimal transport

Optimal transport measures the minimal cost of transporting the target measure μ onto ν with couplings γ, according to a cost $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$.

$$\mathrm{OT}_c(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma(x, y).$$



Dual formulation

Kantorovich duality:

$$\begin{aligned}\mathrm{OT}_c(\mu,\nu) &= \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)d\gamma(x,y)\\ &= \sup_{\substack{\varphi,\psi\in\mathcal{C}(\mathbb{R}^d)\\ \varphi\oplus\psi\le c}} \int_{\mathbb{R}^d}\varphi d\mu + \int_{\mathbb{R}^d}\psi d\nu.\end{aligned}$$

Dual formulation

Kantorovich duality:

$$\begin{aligned}\mathrm{OT}_c(\mu,\nu) &= \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)d\gamma(x,y)\\ &= \sup_{\substack{\varphi,\psi\in\mathcal{C}(\mathbb{R}^d)\\ \varphi\oplus\psi\leq c}} \int_{\mathbb{R}^d}\varphi d\mu + \int_{\mathbb{R}^d}\psi d\nu.\end{aligned}$$

c-transform: $\varphi^c(x) := \inf_{y\in\mathbb{R}^d} c(x,y) - \psi(y)$.

Dual formulation

Kantorovich duality:

$$\mathrm{OT}_c(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)d\gamma(x,y)$$
$$= \sup_{\substack{\varphi,\psi\in\mathcal{C}(\mathbb{R}^d)\\ \varphi\oplus\psi\leq c}} \int_{\mathbb{R}^d} \varphi d\mu + \int_{\mathbb{R}^d} \psi d\nu.$$

c-transform: $\varphi^c(x) := \inf_{y\in\mathbb{R}^d} c(x,y) - \psi(y)$. Constraint $\varphi \oplus \psi \leq c$ is thus equivalent to $\psi \leq \varphi^c$.

$$\mathrm{OT}_c(\mu,\nu) = \sup_{\psi\in\mathcal{C}(\mathbb{R}^d)} \int_{\mathbb{R}^d} \varphi^c d\mu + \int_{\mathbb{R}^d} \varphi d\nu.$$

The quadratic cost

If $c(x,y) := \|x-y\|^p,\ p \geq 1$ then $W_p(\mu,\nu) := \mathrm{OT}_c(\mu,\nu)^{1/p}$ is a **distance** over the probability space $\mathcal{M}_1(\mathbb{R}^d)$ called the **Wasserstein distance**.

The quadratic cost

If $c(x, y) := \|x - y\|^p,\ p \geq 1$ then $W_p(\mu, \nu) := \mathrm{OT}_c(\mu, \nu)^{1/p}$ is a **distance** over the probability space $\mathcal{M}_1(\mathbb{R}^d)$ called the **Wasserstein distance**.

Moreover, if μ has density, then by Brenier's Theorem there exists a mapping $T : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ such that

$$\begin{aligned}\mathrm{OT}_c(\mu, \nu) &= W_p^p(\mu, \nu) \\ &= \int_{\mathbb{R}^d} \|x - T(x)\|^p d\mu(x)\end{aligned}$$

T is called the **optimal transport map**.

The semi discrete setting

Suppose $\nu := \sum_{i=1}^{N} \pi_i \delta_{y_i},\ Y \in (\mathbb{R}^d)^N,\ \pi \in \Delta_N$.

The semi discrete setting

Suppose $\nu := \sum_{i=1}^{N} \pi_i \delta_{y_i}$, $Y \in (\mathbb{R}^d)^N$, $\pi \in \Delta_N$. Then

$$
\begin{aligned}
\mathrm{OT}_c(\mu, \nu) &= \sup_{\varphi \in \mathcal{C}(\mathbb{R}^d)} \int_{\mathbb{R}^d} \varphi^c d\mu + \int_{\mathbb{R}^d} \varphi d\nu \\
&= \sup_{\varphi \in \mathbb{R}^N} \int_{\mathbb{R}^d} \min_{i=1,\dots,N} c(x, y_i) - \varphi_i \, d\mu(x) + \sum_{i=1}^{N} \pi_i \varphi_i \\
&= \sum_{i=1}^{N} \int_{L_i^c(Y, \varphi^*)} c(x, y_i) - \varphi_i^* \, d\mu(x) + \sum_{i=1}^{N} \pi_i \varphi_i^*
\end{aligned}
$$

The semi discrete setting

Suppose $\nu := \sum_{i=1}^{N} \pi_i \delta_{y_i}$, $Y \in (\mathbb{R}^d)^N$, $\pi \in \Delta_N$. Then

$$\begin{aligned}
\mathrm{OT}_c(\mu, \nu) &= \sup_{\varphi \in \mathcal{C}(\mathbb{R}^d)} \int_{\mathbb{R}^d} \varphi^c d\mu + \int_{\mathbb{R}^d} \varphi d\nu \\
&= \sup_{\varphi \in \mathbb{R}^N} \int_{\mathbb{R}^d} \min_{i=1,\dots,N} c(x, y_i) - \varphi_i \, d\mu(x) + \sum_{i=1}^{N} \pi_i \varphi_i \\
&= \sum_{i=1}^{N} \int_{L_i^c(Y, \varphi^*)} c(x, y_i) - \varphi_i^* \, d\mu(x) + \sum_{i=1}^{N} \pi_i \varphi_i^*
\end{aligned}$$

where

$$L_i^c(Y, \varphi^*) = \left\{ x \in \mathbb{R}^d \mid \, c(x, y_i) - \varphi_i^* \leq c(x, y_j) - \varphi_j^*, \, j \neq i \right\}.$$

The semi discrete setting

If $c(x, y) = \|x - y\|^p$, the power cells form a convex tesselation of $\mathbb{R}^d$ and define the OT map $T = \sum_{i=1}^N y_i 1_{L_i(Y,\varphi^*)}$.

The semi discrete setting

If $c(x,y) = \|x - y\|^p$, the power cells form a convex tesselation of $\mathbb{R}^d$ and define the OT map $T = \sum_{i=1}^{N} y_i \mathbb{1}_{L_i(Y,\varphi^*)}$.

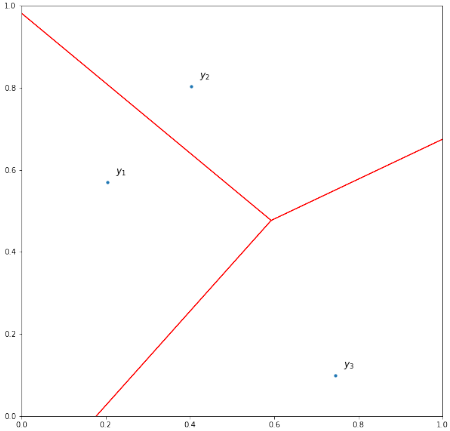


Figure: Example of semi-discrete OT between $\mu \sim \mathcal{U}_{[0,1]^2}$ and $\frac{1}{3}(\delta_{y_1} + \delta_{y_2} + \delta_{y_3})$.

Entropic Optimal transport

$$W_{\epsilon,p}(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \|x-y\|^p d\gamma(x,y) + \epsilon \mathrm{KL}(\gamma, \mu \bigotimes \nu)$$

Entropic Optimal transport

$W_{\epsilon,p}(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \|x-y\|^p d\gamma(x,y) + \epsilon \mathrm{KL}(\gamma, \mu \bigotimes \nu)$ (where $\mathrm{KL}(\alpha,\beta) = \int_{\mathbb{R}^d} \log\left(\frac{d\alpha(x)}{d\beta(x)}\right) d\alpha(x)$).

Entropic Optimal transport

$W_{\epsilon,p}(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \|x-y\|^p d\gamma(x,y) + \epsilon \mathrm{KL}(\gamma, \mu\bigotimes\nu)$ (where $\mathrm{KL}(\alpha,\beta) = \int_{\mathbb{R}^d} \log\left(\frac{d\alpha(x)}{d\beta(x)}\right) d\alpha(x)$).

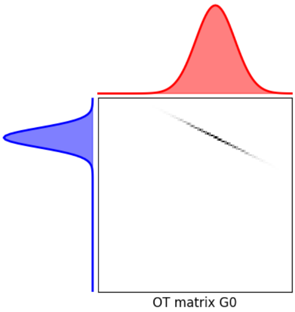


Figure: *POT library example of a $W_{\epsilon,2}$-transport plan*

Wider because it penalizes couplings too far from the product measure

Sliced optimal transport

1-D optimal transport: if $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$:

$$W_p^p(\mu,\nu) = \int_0^1 \mid F_\mu^{-1}(t) - F_\nu^{-1}(t) \mid^p dt$$

where F_α^{-1} denotes the generalized inverse of the c.d.f of α.

Sliced optimal transport

1-D optimal transport: if $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$:

$$W_p^p(\mu, \nu) = \int_0^1 \mid F_\mu^{-1}(t) - F_\nu^{-1}(t) \mid^p dt$$

where F_α^{-1} denotes the generalized inverse of the c.d.f of α.

Sliced optimal transport:
The sliced and max-sliced Wasserstein distances leverage the closed form of 1-d OT through projections on the line with directions θ.

Sliced optimal transport

1-D optimal transport: if $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$:

$$W_p^p(\mu, \nu) = \int_0^1 \mid F_\mu^{-1}(t) - F_\nu^{-1}(t) \mid^p dt$$

where F_α^{-1} denotes the generalized inverse of the c.d.f of α.

Sliced optimal transport:
The sliced and max-sliced Wasserstein distances leverage the closed form of 1-d OT through projections on the line with directions θ.

$$SW_p^p(\mu, \nu) = \int_{\mathbb{S}^{d-1}} W_p^p(P_\theta \sharp \mu, P_\theta \sharp \nu) d\sigma(\theta), \ \sigma \sim \mathcal{U}_{\mathbb{S}^{d-1}}$$

$$\text{max-}SW_p^p(\mu, \nu) = \max_{\theta \in \mathbb{S}^{d-1}} W_p^p(P_\theta \sharp \mu, P_\theta \sharp \nu).$$

Our object of study

Let $\mu^1, \ldots, \mu^L \in \mathcal{M}_1(\mathbb{R}^d)$, we study

$$\min_{(Y,\pi)\in A} \mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^{N} \pi_i \delta_{y_i}\right) := \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{D}\left(\mu^\ell, \sum_{i=1}^{N} \pi_i \delta_{y_i}\right).$$

Typically $\mathbb{D} = W_p^p$.

Our object of study

Let $\mu^1, \dots, \mu^L \in \mathcal{M}_1(\mathbb{R}^d)$, we study

$$\min_{(Y,\pi)\in A} \mathrm{F}_{\mathbb{D}} \left(\mu^1, \dots, \mu^L, \sum_{i=1}^{N} \pi_i \delta_{y_i} \right) := \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{D} \left(\mu^\ell, \sum_{i=1}^{N} \pi_i \delta_{y_i} \right).$$

Typically $\mathbb{D} = W_p^p$.

A	$L = 1$	$L > 1$
$(\mathbb{R}^d)^N \times \Delta_N$	Optimal quantization	OT barycenter with sparse support
$(\mathbb{R}^d)^N \times \{\bar{\pi}\}$	Constrained quantization	Free support OT barycenter
$\{\bar{Y}\} \times \Delta_N$	-	Fixed support OT barycenter

The objective function

Focus on $\mathbb{D} = W_p^p$ and $L = 1$

$$\boxed{(Y, \pi) \mapsto \sup_{w \in \mathbb{R}^N} \int_{\mathbb{R}^d} \|x - y_i\|^p - w_i d\mu(x) + \sum_{i=1}^{N} w_i \pi_i}$$

The objective function

Focus on $\mathbb{D} = W_p^p$ and $L = 1$

$$\boxed{(Y, \pi) \mapsto \sup_{w \in \mathbb{R}^N} \int_{\mathbb{R}^d} \|x - y_i\|^p - w_i d\mu(x) + \sum_{i=1}^{N} w_i \pi_i}$$

- Convex in π.
- Non convex in Y.
- GD works in practice but convergence guarantees are hard.

The objective function

Focus on $\mathbb{D} = W_p^p$ and $L = 1$

$$(Y, \pi) \mapsto \sup_{w \in \mathbb{R}^N} \int_{\mathbb{R}^d} \|x - y_i\|^p - w_i d\mu(x) + \sum_{i=1}^{N} w_i \pi_i$$

- Convex in π.
- Non convex in Y.
- GD works in practice but convergence guarantees are hard.

In this section we provide such guarantees.

Uniform and optimal quantization

$$\min_{Y:=(y_1,\dots,y_N)} \frac{1}{2} W_2^2 \left(\mu, \frac{1}{N} \sum_{i=1}^{N} \delta_{y_i} \right) = \min_{Y:=(y_1,\dots,y_N)} F_N(Y)$$

Uniform and optimal quantization

$$\min_{Y:=(y_1,\dots,y_N)} \frac{1}{2} W_2^2 \left(\mu, \frac{1}{N} \sum_{i=1}^{N} \delta_{y_i} \right) = \min_{Y:=(y_1,\dots,y_N)} F_N(Y)$$

where
$F_N(Y) = \max_{w \in \mathbb{R}^N} \frac{1}{2} \left\{ \int_{\mathbb{R}^d} \left(\min_{i=1,\dots,N} \|x - y_i\|^2 - w_i \right) f(x) dx + \frac{1}{N} \sum_{i=1}^{N} w_i \right\}$

Uniform and optimal quantization

$$\min_{Y:=(y_1,\dots,y_N)} \frac{1}{2} W_2^2 \left(\mu, \frac{1}{N}\sum_{i=1}^{N} \delta_{y_i}\right) = \min_{Y:=(y_1,\dots,y_N)} F_N(Y)$$

where

$$F_N(Y) = \max_{w\in\mathbb{R}^N} \tfrac{1}{2} \left\{ \int_{\mathbb{R}^d} \left(\min_{i=1,\dots,N} \|x - y_i\|^2 - w_i\right) f(x)dx + \tfrac{1}{N} \sum_{i=1}^{N} w_i \right\}$$

(Uniform Quantization)

Uniform and optimal quantization

$$\boxed{\min_{Y:=(y_1,\ldots,y_N)} \frac{1}{2} W_2^2 \left(\mu, \frac{1}{N} \sum_{i=1}^{N} \delta_{y_i} \right) = \min_{Y:=(y_1,\ldots,y_N)} F_N(Y)}$$

where

$$F_N(Y) = \max_{w \in \mathbb{R}^N} \tfrac{1}{2} \left\{ \int_{\mathbb{R}^d} \left(\min_{i=1,\ldots,N} \|x - y_i\|^2 - w_i \right) f(x) dx + \tfrac{1}{N} \sum_{i=1}^{N} w_i \right\}$$

(Uniform Quantization)

$$\boxed{\min_{Y, \pi > 0, \sum_{i=1}^{N} \pi_i = 1} \frac{1}{2} W_2^2 (\mu, \sum_{i=1}^{N} \pi_i \delta_{y_i}) = \min_{Y:=(y_1,\ldots,y_N)} G_N(Y)}$$

Uniform and optimal quantization

$$\min_{Y:=(y_1,\ldots,y_N)} \frac{1}{2} W_2^2\left(\mu, \frac{1}{N}\sum_{i=1}^{N} \delta_{y_i}\right) = \min_{Y:=(y_1,\ldots,y_N)} F_N(Y)$$

where
$F_N(Y) = \max_{w\in\mathbb{R}^N} \frac{1}{2} \left\{ \int_{\mathbb{R}^d} \left(\min_{i=1,\ldots,N} \|x - y_i\|^2 - w_i\right) f(x)dx + \frac{1}{N}\sum_{i=1}^N w_i \right\}$

(Uniform Quantization)

$$\min_{Y,\pi>0,\, \sum_{i=1}^N \pi_i=1} \frac{1}{2} W_2^2(\mu, \sum_{i=1}^{N} \pi_i\delta_{y_i}) = \min_{Y:=(y_1,\ldots,y_N)} G_N(Y)$$

where $G_N(Y) = \frac{1}{2} \int_{\mathbb{R}^d} \min_{i=1,\ldots,N} \|x - y_i\|^2 f(x)dx$

Uniform and optimal quantization

$$\min_{Y:=(y_1,\dots,y_N)} \frac{1}{2} W_2^2 \left(\mu, \frac{1}{N} \sum_{i=1}^{N} \delta_{y_i} \right) = \min_{Y:=(y_1,\dots,y_N)} F_N(Y)$$

where
$F_N(Y) = \max_{w \in \mathbb{R}^N} \frac{1}{2} \left\{ \int_{\mathbb{R}^d} \left(\min_{i=1,\dots,N} \|x - y_i\|^2 - w_i \right) f(x)dx + \frac{1}{N} \sum_{i=1}^{N} w_i \right\}$

(Uniform Quantization)

$$\min_{Y,\pi>0,\, \sum_{i=1}^{N} \pi_i = 1} \frac{1}{2} W_2^2(\mu, \sum_{i=1}^{N} \pi_i \delta_{y_i}) = \min_{Y:=(y_1,\dots,y_N)} G_N(Y)$$

where $G_N(Y) = \frac{1}{2} \int_{\mathbb{R}^d} \min_{i=1,\dots,N} \|x - y_i\|^2 f(x)dx$

(Optimal Quantization)

Voronoï and Laguerre tesselation

- i'th Voronoï cell:
  $V_i(Y) = \{x \in \mathbb{R}^d \mid \|x - y_i\|^2 < \|x - y_j\|^2 \ \forall j = 1, ..., N\}$

Optimal Quantization

Uniform Quantization

Voronoï and Laguerre tesselation

- i'th Voronoï cell:
  $V_i(Y) = \{x \in \mathbb{R}^d \mid \|x - y_i\|^2 < \|x - y_j\|^2 \; \forall j = 1, ..., N\}$
- i'th Laguerre cell:
  $L_i(Y, w) = \{x \in \mathbb{R}^d \mid \|x - y_i\|^2 - w_i < \|x - y_j\|^2 - w_j \; \forall j = 1, ..., N\}$

Optimal Quantization

Uniform Quantization

Voronoï and Laguerre tesselation

- i'th Voronoï cell:
  $V_i(Y) = \{x \in \mathbb{R}^d \mid \|x - y_i\|^2 < \|x - y_j\|^2 \; \forall j = 1, ..., N\}$
- i'th Laguerre cell:
  $L_i(Y, w) = \{x \in \mathbb{R}^d \mid \|x - y_i\|^2 - w_i < \|x - y_j\|^2 - w_j \; \forall j = 1, ..., N\}$

Optimal Quantization

$$\begin{aligned} G_N(Y) &= \frac{1}{2} \int_{\mathbb{R}^d} \min_{i=1,\dots,N} \left\| x - y_i \right\|^2 f(x) dx \\ &= \frac{1}{2} \sum_{i=1}^{N} \int_{V_i(Y)} \left\| x - y_i \right\|^2 f(x) dx \end{aligned}$$

Discrete
Uniform Quantization

Voronoï and Laguerre tesselation

- i'th Voronoï cell:
  $V_i(Y) = \{x \in \mathbb{R}^d \mid \|x - y_i\|^2 < \|x - y_j\|^2 \; \forall j = 1, ..., N\}$
- i'th Laguerre cell:
  $L_i(Y, w) = \{x \in \mathbb{R}^d \mid \|x - y_i\|^2 - w_i < \|x - y_j\|^2 - w_j \; \forall j = 1, ..., N\}$

Optimal Quantization

$$G_N(Y) = \frac{1}{2} \int_{\mathbb{R}^d} \min_{i=1,\dots,N} \|x - y_i\|^2 f(x)dx$$
$$= \frac{1}{2} \sum_{i=1}^{N} \int_{V_i(Y)} \|x - y_i\|^2 f(x)dx$$

Discrete

Uniform Quantization

$$F_N(Y) = \max_{w \in \mathbb{R}^N} \frac{1}{2} \left\{ \int_{\mathbb{R}^d} \left(\min_{i=1,\dots,N} \|x - y_i\|^2 - w_i \right) f(x)dx + \frac{1}{N} \sum_{i=1}^{N} w_i \right\}$$
$$= \max_{w \in \mathbb{R}^N} \frac{1}{2} \left\{ \sum_{i=1}^{N} \int_{L_i(Y,w)} \left(\|x - y_i\|^2 - w_i \right) f(x)dx + \frac{1}{N} \sum_{i=1}^{N} w_i \right\}$$

Lloyd's algorithm[1]: two equivalent formulations

Can be described in two ways

1. Choose an initial points cloud Y.

[1] Lloyd, Stuart P. Least squares quantization in PCM, 1982.

Lloyd's algorithm[1]: two equivalent formulations

Can be described in two ways

1. Choose an initial points cloud Y.
2. Construct either a Voronoï (OQ) or a Laguerre tessellation (UQ) of the support of μ.

[1] Lloyd, Stuart P. Least squares quantization in PCM, 1982.

Lloyd's algorithm[1]: two equivalent formulations

Can be described in two ways

1. Choose an initial points cloud Y.
2. Construct either a Voronoï (OQ) or a Laguerre tessellation (UQ) of the support of μ.
3. Set the new points cloud as the barycenters of every cell.

[1] Lloyd, Stuart P. Least squares quantization in PCM, 1982.

Lloyd's algorithm[1]: two equivalent formulations

Can be described in two ways

1. Choose an initial points cloud Y.
2. Construct either a Voronoï (OQ) or a Laguerre tessellation (UQ) of the support of μ.
3. Set the new points cloud as the barycenters of every cell.

Which translates to

- UQ:

$$\begin{cases} Y_0 \in (\mathbb{R}^d)^N \setminus D_N \\ Y_{n+1} = Y_n - N.\nabla F_N(Y_n) \end{cases}$$

[1] Lloyd, Stuart P. Least squares quantization in PCM, 1982.

Lloyd's algorithm[1]: two equivalent formulations

Can be described in two ways

1. Choose an initial points cloud Y.
2. Construct either a Voronoï (OQ) or a Laguerre tessellation (UQ) of the support of μ.
3. Set the new points cloud as the barycenters of every cell.

Which translates to

- UQ:

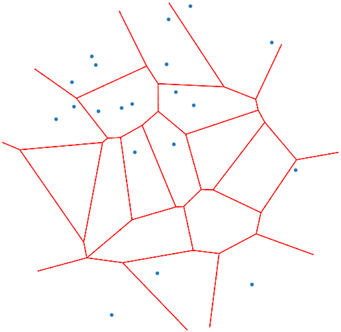
$$\begin{cases} Y_0 \in (\mathbb{R}^d)^N \setminus D_N \\ Y_{n+1} = Y_n - N.\nabla F_N(Y_n) \end{cases}$$

- OQ:

$$\begin{cases} Y_0 \in (\mathbb{R}^d)^N \setminus D_N \\ Y_{n+1} = Y_n - \frac{1}{\mu(V(Y_n))}.\nabla G_N(Y_n) \end{cases}$$

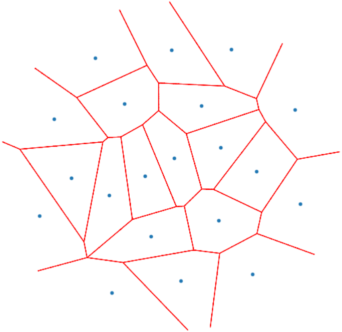
[1] Lloyd, Stuart P. Least squares quantization in PCM, 1982.

Lloyd's algorithm for (Uniform Quantization): an example with $\mu \sim \mathcal{N}_{\mathbb{S}^1}(0_2, \sigma^2 I_2)$



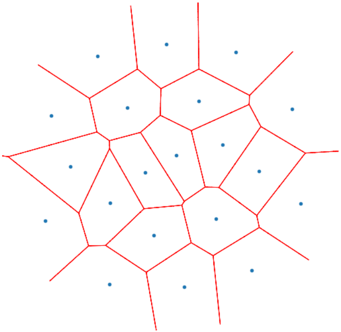
Simulations made using the library PyMongeAmpere by Quentin Mérigot.

Lloyd's algorithm for (Uniform Quantization): an example with $\mu \sim \mathcal{N}_{\mathbb{S}^1}(0_2, \sigma^2 I_2)$



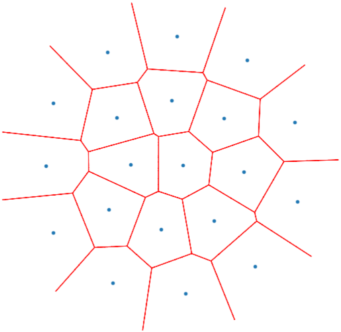
Simulations made using the library PyMongeAmpere by Quentin Mérigot.

Lloyd's algorithm for (Uniform Quantization): an example with $\mu \sim \mathcal{N}_{\mathbb{S}^1}(0_2, \sigma^2 I_2)$



Simulations made using the library PyMongeAmpere by Quentin Mérigot.

Lloyd's algorithm for (Uniform Quantization): an example with $\mu \sim \mathcal{N}_{\mathbb{S}^1}(0_2, \sigma^2 I_2)$



Simulations made using the library PyMongeAmpere by Quentin Mérigot.

Convergence of Lloyd's algorithm for OQ and UQ

Uniform Quantization

$F_N(Y) = \max_{w \in \mathbb{R}^N} \frac{1}{2} \left\{ \int_{\mathbb{R}^d} \left(\min_{i=1,\ldots,N} \|x - y_i\|^2 - w_i \right) f(x) dx + \frac{1}{N} \sum_{i=1}^{N} w_i \right\}$

Optimal Quantization

$G_N(Y) = \frac{1}{2} \int_{\mathbb{R}^d} \min_{i=1,\ldots,N} \|x - y_i\|^2 f(x) dx$

Convergence of Lloyd's algorithm for OQ and UQ

Uniform Quantization

$F_N(Y) = \max_{w\in\mathbb{R}^N} \frac{1}{2}\left\{\int_{\mathbb{R}^d}\left(\min_{i=1,\dots,N}\|x-y_i\|^2 - w_i\right) f(x)dx + \frac{1}{N}\sum_{i=1}^N w_i\right\}$

Let $(Y_n)_n$ be Lloyd iterates.

Optimal Quantization

$G_N(Y) = \frac{1}{2}\int_{\mathbb{R}^d} \min_{i=1,\dots,N}\|x-y_i\|^2 f(x)dx$

Convergence of Lloyd's algorithm for OQ and UQ

Uniform Quantization

$F_N(Y) = \max_{w\in\mathbb{R}^N} \frac{1}{2}\left\{\int_{\mathbb{R}^d}\left(\min_{i=1,\dots,N}\|x-y_i\|^2 - w_i\right) f(x)dx + \frac{1}{N}\sum_{i=1}^N w_i\right\}$

Let $(Y_n)_n$ be Lloyd iterates.

- Convergence to a set of critical points[2]: $\|\nabla F_N(Y_n)\| \underset{n\longrightarrow\infty}{\longrightarrow} 0$

Optimal Quantization

$G_N(Y) = \frac{1}{2}\int_{\mathbb{R}^d} \min_{i=1,\dots,N}\|x-y_i\|^2 f(x)dx$

[2] Quentin Merigot, Filippo Santambrogio, Clément Sarrazin. Non-asymptotic convergence bounds for Wasserstein approximation using point clouds. 2021

Convergence of Lloyd's algorithm for OQ and UQ

Uniform Quantization

$F_N(Y) = \max_{w \in \mathbb{R}^N} \frac{1}{2} \left\{ \int_{\mathbb{R}^d} \left(\min_{i=1,\dots,N} \|x - y_i\|^2 - w_i \right) f(x) dx + \frac{1}{N} \sum_{i=1}^N w_i \right\}$

Let $(Y_n)_n$ be Lloyd iterates.

- Convergence to a set of critical points[2]: $\|\nabla F_N(Y_n)\| \underset{n \longrightarrow \infty}{\longrightarrow} 0$

Optimal Quantization

$G_N(Y) = \frac{1}{2} \int_{\mathbb{R}^d} \min_{i=1,\dots,N} \|x - y_i\|^2 f(x) dx$

Let (Z_n) be Lloyd iterates.

[2] Quentin Merigot, Filippo Santambrogio, Clément Sarrazin. Non-asymptotic convergence bounds for Wasserstein approximation using point clouds. 2021

Convergence of Lloyd's algorithm for OQ and UQ

Uniform Quantization

$F_N(Y) = \max_{w \in \mathbb{R}^N} \frac{1}{2} \left\{ \int_{\mathbb{R}^d} \left(\min_{i=1,\ldots,N} \|x - y_i\|^2 - w_i \right) f(x) dx + \frac{1}{N} \sum_{i=1}^N w_i \right\}$

Let $(Y_n)_n$ be Lloyd iterates.

- Convergence to a set of critical points[2]: $\|\nabla F_N(Y_n)\| \underset{n \longrightarrow \infty}{\longrightarrow} 0$

Optimal Quantization

$G_N(Y) = \frac{1}{2} \int_{\mathbb{R}^d} \min_{i=1,\ldots,N} \|x - y_i\|^2 f(x) dx$

Let (Z_n) be Lloyd iterates.

- Convergence to a set of critical points[3]: $\|\nabla G_N(Z_n)\| \underset{n \longrightarrow \infty}{\longrightarrow} 0$

[2] Quentin Merigot, Filippo Santambrogio, Clément Sarrazin. Non-asymptotic convergence bounds for Wasserstein approximation using point clouds. 2021

[3] Maria Emelianenko, Lili Ju and Alexander Rand. Nondegeneracy and Weak Global Convergence of the Lloyd Algorithm in $\mathbb{R}^d$. 2008

Convergence of Lloyd's algorithm for OQ and UQ

Uniform Quantization

$$F_N(Y) = \max_{w \in \mathbb{R}^N} \frac{1}{2} \left\{ \int_{\mathbb{R}^d} \left(\min_{i=1,\dots,N} \|x - y_i\|^2 - w_i \right) f(x) dx + \frac{1}{N} \sum_{i=1}^N w_i \right\}$$

Let $(Y_n)_n$ be Lloyd iterates.

- Convergence to a set of critical points[2]: $\|\nabla F_N(Y_n)\| \underset{n \longrightarrow \infty}{\longrightarrow} 0$

Optimal Quantization

$$G_N(Y) = \frac{1}{2} \int_{\mathbb{R}^d} \min_{i=1,\dots,N} \|x - y_i\|^2 f(x) dx$$

Let (Z_n) be Lloyd iterates.

- Convergence to a set of critical points[3]: $\|\nabla G_N(Z_n)\| \underset{n \longrightarrow \infty}{\longrightarrow} 0$

Question:

What about $(Y_n)_{n \geq 1}$ and $(Z_n)_{n \geq 1}$?

---

[2] Quentin Merigot, Filippo Santambrogio, Clément Sarrazin. Non-asymptotic convergence bounds for Wasserstein approximation using point clouds. 2021

[3] Maria Emelianenko, Lili Ju and Alexander Rand. Nondegeneracy and Weak Global Convergence of the Lloyd Algorithm in $\mathbb{R}^d$. 2008

Main result

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is ???. Then the iterates of Lloyd for both uniform (UQ) and optimal quantization (OQ) converge.

Main result

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is ???. Then the iterates of Lloyd for both uniform (UQ) and optimal quantization (OQ) converge.

Remark

Convexity of Supp(μ) is not necessary for uniform quantization.

Convergence of gradient schemes in non-convex setting

- F_N and G_N both non convex.

[4] H. Attouch, J. Bolte, and B.F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. 2013.

Convergence of gradient schemes in non-convex setting

- F_N and G_N both non convex.

We follow the geometric approach in optimization pioneered by J.Bolte.

[4] H. Attouch, J. Bolte, and B.F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. 2013.

Convergence of gradient schemes in non-convex setting

- F_N and G_N both non convex.

We follow the geometric approach in optimization pioneered by J.Bolte.

- The iterates of gradient sequences converge in this setting under **Lojasiewicz/Kurdyka-Lojasiewicz** inequality. [4]

[4] H. Attouch, J. Bolte, and B.F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. 2013.

Convergence of gradient schemes in non-convex setting

- F_N and G_N both non convex.

We follow the geometric approach in optimization pioneered by J.Bolte.

- The iterates of gradient sequences converge in this setting under **Lojasiewicz/Kurdyka-Lojasiewicz** inequality. [4]
- Such properties are true whenever the underlying objective function's graph belongs to an o-minimal structure.

[4] H. Attouch, J. Bolte, and B.F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. 2013.

Convergence of gradient schemes in non-convex setting

- F_N and G_N both non convex.

We follow the geometric approach in optimization pioneered by J.Bolte.

- The iterates of gradient sequences converge in this setting under **Lojasiewicz/Kurdyka-Lojasiewicz** inequality. [4]
- Such properties are true whenever the underlying objective function's graph belongs to an o-minimal structure.
- F_N and G_N do so under an analyticity assumption on the target density.

[4] H. Attouch, J. Bolte, and B.F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. 2013.

KL inequality

Kurdyka-Lojasiewicz Inequality

f verifies a KL inequality at x^* if there exists a neighborhood U of x^*, $c > 0$, $\eta > 0$ and a strictly increasing positive function $\Psi : [0, \eta[\longrightarrow \mathbb{R}$ such that:

$$\|\nabla(\Psi \circ f)(x)\| \geq c, \ \forall x \in U \text{ such that } \ 0 < f(x) < \eta. \tag{1}$$

KL inequality

<u>Kurdyka-Lojasiewicz Inequality</u>

f verifies a KL inequality at x^* if there exists a neighborhood U of x^*, $c > 0$, $\eta > 0$ and a strictly increasing positive function $\Psi : [0, \eta[\longrightarrow \mathbb{R}$ such that:

$$\|\nabla(\Psi \circ f)(x)\| \geq c, \; \forall x \in U \text{ such that } 0 < f(x) < \eta. \tag{1}$$

<u>Main idea of KL:</u> Composing f with Ψ makes it sharper around its critical points.

KL inequality

Kurdyka-Lojasiewicz Inequality
f verifies a KL inequality at x^* if there exists a neighborhood U of x^*, $c > 0$, $\eta > 0$ and a strictly increasing positive function $\Psi : [0, \eta[\longrightarrow \mathbb{R}$ such that:

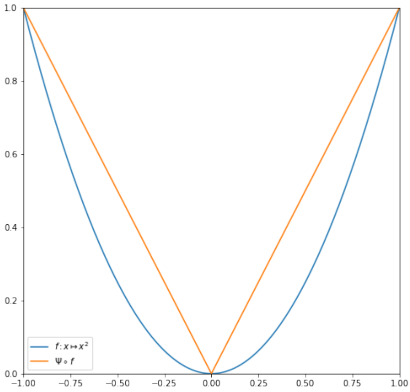
$$\|\nabla(\Psi \circ f)(x)\| \geq c, \ \forall x \in U \text{ such that } 0 < f(x) < \eta. \tag{1}$$

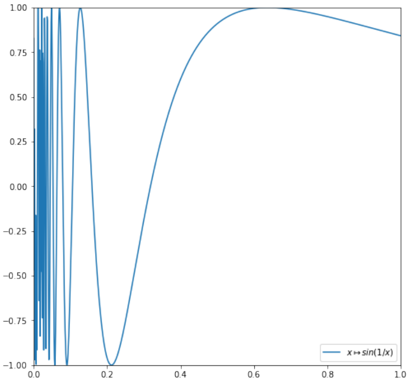
Main idea of KL: Composing f with Ψ makes it sharper around its critical points.

A function whose differential go to zero can have infinitely many oscillations as it goes to 0. The KL property prevents that from happening.

KL inequality

A case where the function is KL... and another one where it is not (at 0).





f:x$\mapsto x^2$, $\Psi : x \mapsto \sqrt{x}$ $\quad$ $g : x \mapsto \sin 1/x$.

Strong descent conditions

Strong descent conditions: Let $(Y_k)_{k\geq 1}$ be some sequence and F be some differentiable cost function.

Strong descent conditions

Strong descent conditions: Let $(Y_k)_{k\geq 1}$ be some sequence and F be some differentiable cost function.

- $F(Y_k) - F(Y_{k+1}) \geq \sigma \|\nabla F(Y_k)\|.\|Y_{k+1} - Y_k\|$

Strong descent conditions

Strong descent conditions: Let $(Y_k)_{k\geq 1}$ be some sequence and F be some differentiable cost function.

- $F(Y_k) - F(Y_{k+1}) \geq \sigma \|\nabla F(Y_k)\|.\|Y_{k+1} - Y_k\|$
- $F(Y_{k+1}) = F(Y_k) \implies Y_{k+1} = Y_k$.

Strong descent conditions

Strong descent conditions: Let $(Y_k)_{k\geq 1}$ be some sequence and F be some differentiable cost function.

- $F(Y_k) - F(Y_{k+1}) \geq \sigma \|\nabla F(Y_k)\|.\|Y_{k+1} - Y_k\|$
- $F(Y_{k+1}) = F(Y_k) \implies Y_{k+1} = Y_k$.

(We say that the sequence $(Y_k)_{k\geq 1}$ is SDC)

Strong descent conditions

Strong descent conditions: Let $(Y_k)_{k\geq 1}$ be some sequence and F be some differentiable cost function.

- $F(Y_k) - F(Y_{k+1}) \geq \sigma \|\nabla F(Y_k)\|.\|Y_{k+1} - Y_k\|$
- $F(Y_{k+1}) = F(Y_k) \implies Y_{k+1} = Y_k$.

(We say that the sequence $(Y_k)_{k\geq 1}$ is SDC)

- We show that G_N is SDC (largely thanks to Emelianenko, Ju, Rand [5]).

[5] M. Emelianenko, L. Ju, and A. Rand. Nondegeneracy and weak global convergence of the Lloyd algorithm in $\mathbb{R}^d$. 2008

Strong descent conditions

Strong descent conditions: Let $(Y_k)_{k\geq 1}$ be some sequence and F be some differentiable cost function.

- $F(Y_k) - F(Y_{k+1}) \geq \sigma \|\nabla F(Y_k)\|.\|Y_{k+1} - Y_k\|$
- $F(Y_{k+1}) = F(Y_k) \implies Y_{k+1} = Y_k$.

(We say that the sequence $(Y_k)_{k\geq 1}$ is SDC)

- We show that G_N is SDC (largely thanks to Emelianenko, Ju, Rand [5]).
- F_N SDC (can be deduced from Merigot, Santambrogio, Sarrazin [6]).

[5] M. Emelianenko, L. Ju, and A. Rand. Nondegeneracy and weak global convergence of the Lloyd algorithm in $\mathbb{R}^d$. 2008

[6] Quentin Merigot, Filippo Santambrogio, Clément Sarrazin. Non-asymptotic convergence bounds for Wasserstein approximation using point clouds. 2021

Idea of proof

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is ???. Then the iterates of Lloyd for both uniform (UQ) and optimal quantization (OQ) converge.

Sketch of proof

Idea of proof

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is ???. Then the iterates of Lloyd for both uniform (UQ) and optimal quantization (OQ) converge.

Sketch of proof

- F_N and G_N verify a KL inequality at all points of their domain.

Idea of proof

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is ???. Then the iterates of Lloyd for both uniform (UQ) and optimal quantization (OQ) converge.

Sketch of proof

- F_N and G_N verify a KL inequality at all points of their domain.
- SDC gradient methods on KL functions converge *(Bolte et.al, Absil et.al.)*

Idea of proof

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is ???. Then the iterates of Lloyd for both uniform (UQ) and optimal quantization (OQ) converge.

Sketch of proof

- F_N and G_N verify a KL inequality at all points of their domain.
- SDC gradient methods on KL functions converge *(Bolte et.al, Absil et.al.)*
- Lloyd's algorithm for OQ and UQ are SDC gradient methods on KL functions and thus the iterates converge.

Idea of proof

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is ???. Then the iterates of Lloyd for both uniform (UQ) and optimal quantization (OQ) converge.

Sketch of proof

- F_N and G_N verify a KL inequality at all points of their domain.
- SDC gradient methods on KL functions converge *(Bolte et.al, Absil et.al.)*
- Lloyd's algorithm for OQ and UQ are SDC gradient methods on KL functions and thus the iterates converge.

Question:

Why are F_N and G_N KL?

Idea of proof

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is ???. Then the iterates of Lloyd for both uniform (UQ) and optimal quantization (OQ) converge.

Sketch of proof

- F_N and G_N verify a KL inequality at all points of their domain.
- SDC gradient methods on KL functions converge *(Bolte et.al, Absil et.al.)*
- Lloyd's algorithm for OQ and UQ are SDC gradient methods on KL functions and thus the iterates converge.

Question:

Why are F_N and G_N KL? Because they are definable in an o minimal structure.

A (very quick) overview on o-minimality

$$\mathcal{SA} \subset \mathbb{R}_{\mathsf{an}} \subset \mathbb{R}_{\mathsf{an,exp}} \subset \mathcal{P}(\mathbb{R})$$

[7] H. Hironaka. Introduction to Real-analytic Sets and Real-analytic Maps. 1973

[8] L.van den Dries and C.Miller, On the real exponential field with restricted analytic functions. 1994

A (very quick) overview on o-minimality

$$\mathcal{SA} \subset \mathbb{R}_{\mathsf{an}} \subset \mathbb{R}_{\mathsf{an,exp}} \subset \mathcal{P}(\mathbb{R})$$

- $\mathcal{SA}$:

[7] H. Hironaka. Introduction to Real-analytic Sets and Real-analytic Maps. 1973

[8] L.van den Dries and C.Miller, On the real exponential field with restricted analytic functions. 1994

A (very quick) overview on o-minimality

$$\mathcal{SA} \subset \mathbb{R}_{\text{an}} \subset \mathbb{R}_{\text{an,exp}} \subset \mathcal{P}(\mathbb{R})$$

- $\mathcal{SA}$: $\bigcup_{i=1}^{N} \{x \in \mathbb{R}^d \mid P_{i_1}(x) \overset{>}{\geq} 0, ..., P_{i_l} \overset{>}{\geq} 0, (P_{i_k})_{k=1,\ldots,l} \subset \mathbb{R}_d[X]\}$

[7] H. Hironaka. Introduction to Real-analytic Sets and Real-analytic Maps. 1973

[8] L.van den Dries and C.Miller, On the real exponential field with restricted analytic functions. 1994

A (very quick) overview on o-minimality

$$\mathcal{SA} \subset \mathbb{R}_{\text{an}} \subset \mathbb{R}_{\text{an,exp}} \subset \mathcal{P}(\mathbb{R})$$

- $\mathcal{SA}$: $\bigcup_{i=1}^{N} \{x \in \mathbb{R}^d \mid P_{i_1}(x) \overset{>}{\geq} 0, ..., P_{i_l} \overset{>}{\geq} 0, (P_{i_k})_{k=1,\dots,l} \subset \mathbb{R}_d[X]\}$
  Definability stable by $+$, $\times$, $(.)^{-1}$, min, ...

[7] H. Hironaka. Introduction to Real-analytic Sets and Real-analytic Maps. 1973

[8] L.van den Dries and C.Miller, On the real exponential field with restricted analytic functions. 1994

A (very quick) overview on o-minimality

$$\mathcal{SA} \subset \mathbb{R}_{\text{an}} \subset \mathbb{R}_{\text{an,exp}} \subset \mathcal{P}(\mathbb{R})$$

- $\mathcal{SA}$: $\bigcup_{i=1}^{N} \{x \in \mathbb{R}^d \mid P_{i_1}(x) \overset{>}{\geq} 0, ..., P_{i_l} \overset{>}{\geq} 0, (P_{i_k})_{k=1,...,l} \subset \mathbb{R}_d[X]\}$
  Definability stable by $+$, $\times$, $(.)^{-1}$, min, ...
- $\mathbb{R}_{\text{an}}$: Same with the analytic maps restricted to a compact within their domain of definition [7] .

[7] H. Hironaka. Introduction to Real-analytic Sets and Real-analytic Maps. 1973

[8] L.van den Dries and C.Miller, On the real exponential field with restricted analytic functions. 1994

A (very quick) overview on o-minimality

$$\mathcal{SA} \subset \mathbb{R}_{\text{an}} \subset \mathbb{R}_{\text{an,exp}} \subset \mathcal{P}(\mathbb{R})$$

- $\mathcal{SA}$: $\bigcup_{i=1}^{N}\{x \in \mathbb{R}^d \mid P_{i_1}(x) \overset{>}{\geq} 0, ..., P_{i_l} \overset{>}{\geq} 0, (P_{i_k})_{k=1,...,l} \subset \mathbb{R}_d[X]\}$
  Definability stable by $+$, $\times$, $(.)^{-1}$, min, ...
- $\mathbb{R}_{\text{an}}$: Same with the analytic maps restricted to a compact within their domain of definition [7] example: $\exp|_{[-1,1]}$.

[7] H. Hironaka. Introduction to Real-analytic Sets and Real-analytic Maps. 1973
[8] L.van den Dries and C.Miller, On the real exponential field with restricted analytic functions. 1994

A (very quick) overview on o-minimality

$$\mathcal{SA} \subset \mathbb{R}_{\mathrm{an}} \subset \mathbb{R}_{\mathrm{an,exp}} \subset \mathcal{P}(\mathbb{R})$$

- $\mathcal{SA}$: $\bigcup_{i=1}^{N} \{x \in \mathbb{R}^d \mid P_{i_1}(x) \overset{>}{\geq} 0, \dots, P_{i_l} \overset{>}{\geq} 0, (P_{i_k})_{k=1,\dots,l} \subset \mathbb{R}_d[X]\}$
  Definability stable by $+$, $\times$, $(.)^{-1}$, min, ...
- $\mathbb{R}_{\mathrm{an}}$: Same with the analytic maps restricted to a compact within their domain of definition [7] example: $\exp|_{[-1,1]}$.
- $\mathbb{R}_{\mathrm{an,exp}}$: Same with the exponential map.[8]

[7] H. Hironaka. Introduction to Real-analytic Sets and Real-analytic Maps. 1973

[8] L.van den Dries and C.Miller, On the real exponential field with restricted analytic functions. 1994

[9] Raf Cluckers, Daniel J. Miller. Stability under integration of sums of products of real globally subanalytic functions and their logarithms. 2009

A (very quick) overview on o-minimality

$$\mathcal{SA} \subset \mathbb{R}_{\text{an}} \subset \mathbb{R}_{\text{an,exp}} \subset \mathcal{P}(\mathbb{R})$$

- $\mathcal{SA}$: $\bigcup_{i=1}^{N} \{x \in \mathbb{R}^d \mid P_{i_1}(x) \overset{>}{\geq} 0, \ldots, P_{i_l} \overset{>}{\geq} 0, (P_{i_k})_{k=1,\ldots,l} \subset \mathbb{R}_d[X]\}$
  Definability stable by $+$, $\times$, $(.)^{-1}$, min, ...
- $\mathbb{R}_{\text{an}}$: Same with the analytic maps restricted to a compact within their domain of definition [7] example: $\exp|_{[-1,1]}$.
- $\mathbb{R}_{\text{an,exp}}$: Same with the exponential map.[8]
- If f is ??? then $Y \mapsto \int_{\mathcal{X}} f(x, Y)dx$ is definable in an o-minimal structure.[9]

---

[7] H. Hironaka. Introduction to Real-analytic Sets and Real-analytic Maps. 1973

[8] L.van den Dries and C.Miller, On the real exponential field with restricted analytic functions. 1994

[9] Raf Cluckers, Daniel J. Miller. Stability under integration of sums of products of real globally subanalytic functions and their logarithms. 2009

A (very quick) overview on o-minimality

$$\mathcal{SA} \subset \mathbb{R}_{\text{an}} \subset \mathbb{R}_{\text{an,exp}} \subset \mathcal{P}(\mathbb{R})$$

- $\mathcal{SA}$: $\bigcup_{i=1}^{N} \{x \in \mathbb{R}^d \mid P_{i_1}(x) \overset{>}{\geq} 0, \ldots, P_{i_l} \overset{>}{\geq} 0, (P_{i_k})_{k=1,\ldots,l} \subset \mathbb{R}_d[X]\}$
  Definability stable by $+$, $\times$, $(.)^{-1}$, min, ...
- $\mathbb{R}_{\text{an}}$: Same with the analytic maps restricted to a compact within their domain of definition [7] example: $\exp|_{[-1,1]}$.
- $\mathbb{R}_{\text{an,exp}}$: Same with the exponential map.[8]
- If f is GSA then $Y \mapsto \int_{\mathcal{X}} f(x, Y)dx$ is definable in an o-minimal structure.[9]
- Functions definable in an o minimal structure are KL.

[7] H. Hironaka. Introduction to Real-analytic Sets and Real-analytic Maps. 1973

[8] L.van den Dries and C.Miller, On the real exponential field with restricted analytic functions. 1994

[9] Raf Cluckers, Daniel J. Miller. Stability under integration of sums of products of real globally subanalytic functions and their logarithms. 2009

Examples of globally subanalytic (GSA) functions

- An image (regarded as a probability density.)

Examples of globally subanalytic (GSA) functions

- An image (regarded as a probability density.)
- Semi-algebraic densities (Uniform law on a semi algebraic set, the triangular distribution,...)

Examples of globally subanalytic (GSA) functions

- An image (regarded as a probability density.)
- Semi-algebraic densities (Uniform law on a semi algebraic set, the triangular distribution,...)
- Any probability density which is analytic and which is truncated on a compact semi algebraic subset (ex: Gaussians truncated on a sphere).

Examples of globally subanalytic (GSA) functions

- An image (regarded as a probability density.)
- Semi-algebraic densities (Uniform law on a semi algebraic set, the triangular distribution,...)
- Any probability density which is analytic and which is truncated on a compact semi algebraic subset (ex: Gaussians truncated on a sphere).
- Any mix of the previous ones.

Examples of globally subanalytic (GSA) functions

- An image (regarded as a probability density.)
- Semi-algebraic densities (Uniform law on a semi algebraic set, the triangular distribution,...)
- Any probability density which is analytic and which is truncated on a compact semi algebraic subset (ex: Gaussians truncated on a sphere).
- Any mix of the previous ones.

Think of it as analytic functions restricted to a semi-algebraic compact subset

Proving the definability of G_N and F_N

$$G_N(Y) = \frac{1}{2} \int_{\mathbb{R}^d} \min_{i=1,\dots,N} \|x - y_i\|^2 \underbrace{f(x)}_{\text{GSA}} \, dx$$

Proving the definability of G_N and F_N

$$G_N(Y) = \frac{1}{2} \int_{\mathbb{R}^d} \underbrace{\min_{i=1,\dots,N} \|x - y_i\|^2 \underbrace{f(x)}_{\text{GSA}}}_{\text{GSA}} \, dx$$

Proving the definability of G_N and F_N

$$G_N(Y) = \underbrace{\frac{1}{2}\int_{\mathbb{R}^d} \underbrace{\min_{i=1,\dots,N} \|x - y_i\|^2 \underbrace{f(x)}_{\text{GSA}}\, dx}_{\text{GSA}}}_{\text{definable}}$$

Proving the definability of G_N and F_N

$$G_N(Y) = \underbrace{\frac{1}{2}\int_{\mathbb{R}^d} \underbrace{\min_{i=1,\dots,N} \|x - y_i\|^2 \underbrace{f(x)}_{\text{GSA}} \, dx}_{\text{GSA}}}_{\text{definable}}$$

<u>Conclusion:</u> G_N is a KL function.

Proving the definability of G_N and F_N

<u>Conclusion:</u> G_N is a KL function.

$$F_N(Y) = \underbrace{\max_{w\in\mathbb{R}^N} \frac{1}{2} \int_{\mathbb{R}^d} \underbrace{\min_{i=1,\dots,N} (\|x-y_i\|^2 - w_i) \underbrace{f(x)}_{\text{GSA}}}_{\text{GSA}} dx + \frac{1}{N}\sum_{i=1}^{N} w_i}_{\text{definable}}$$

Proving the definability of G_N and F_N

Conclusion: G_N is a KL function.

$$F_N(Y) = \underbrace{\max_{w\in\mathbb{R}^N} \underbrace{\frac{1}{2}\int_{\mathbb{R}^d} \underbrace{\min_{i=1,\dots,N}(\|x-y_i\|^2 - w_i)\underbrace{f(x)}_{\text{GSA}}\, dx}_{\text{GSA}} + \frac{1}{N}\sum_{i=1}^N w_i}_{\text{definable}}}_{\text{definable}}\ .$$

Conclusion: F_N is a KL function.

Results

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is globally subanalytic. Then the iterates of Lloyd for both uniform and optimal quantization converge.

Results

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is globally subanalytic. Then the iterates of Lloyd for both uniform and optimal quantization converge.

Additional results:

We showed along the way the definability of $Y \mapsto \mathbb{D}(\mu, \frac{1}{N}\sum_{k=1}^{N} \delta_{y_k})$ for

Results

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is globally subanalytic. Then the iterates of Lloyd for both uniform and optimal quantization converge.

Additional results:

We showed along the way the definability of $Y \mapsto \mathbb{D}(\mu, \frac{1}{N}\sum_{k=1}^{N} \delta_{y_k})$ for

- $\mathbb{D} = W_p^p$ (general Wasserstein)

Results

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is globally subanalytic. Then the iterates of Lloyd for both uniform and optimal quantization converge.

Additional results:

We showed along the way the definability of $Y \mapsto \mathbb{D}(\mu, \frac{1}{N}\sum_{k=1}^{N} \delta_{y_k})$ for

- $\mathbb{D} = W_p^p$ (general Wasserstein)
- $\mathbb{D} = \text{max-}SW_p^p$ (max Sliced Wasserstein)

Results

Theorem

Let μ be a continuous probability measure supported on a compact and convex subset of $\mathbb{R}^d$. Suppose its density is globally subanalytic. Then the iterates of Lloyd for both uniform and optimal quantization converge.

Additional results:

We showed along the way the definability of $Y \mapsto \mathbb{D}(\mu, \frac{1}{N}\sum_{k=1}^{N} \delta_{y_k})$ for

- $\mathbb{D} = W_p^p$ (general Wasserstein)
- $\mathbb{D} = \text{max-}SW_p^p$ (max Sliced Wasserstein)
- $\mathbb{D} = W_{\epsilon,p}$ (Entropic regularization of the semi-discrete OT problem)

From optimization to statistics

We investigated the **deterministic** case where target measures were known.

From optimization to statistics

We investigated the **deterministic** case where target measures were known. However in most ML context μ is unknown, or known but has to be **discretized** (if continuous).

From optimization to statistics

We investigated the **deterministic** case where target measures were known. However in most ML context μ is unknown, or known but has to be **discretized** (if continuous).
We go back to the more general study of

$$\min_{(Y,\pi)\in A} \mathrm{F}_{\mathbb{D}} \left(\mu^1, \ldots, \mu^L, \sum_{i=1}^{N} \pi_i \delta_{y_i} \right) := \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{D} \left(\mu^\ell, \sum_{i=1}^{N} \pi_i \delta_{y_i} \right).$$

for $\mathbb{D} = W_p^p,\ W_{\epsilon,p},\ SW_p^p,\ \text{max-}SW_p^p$.

From optimization to statistics

We investigated the **deterministic** case where target measures were known. However in most ML context μ is unknown, or known but has to be **discretized** (if continuous).
We go back to the more general study of

$$\min_{(Y,\pi)\in A} \mathsf{F}_{\mathbb{D}}\left(\mu^1,\dots,\mu^L,\sum_{i=1}^N \pi_i\delta_{y_i}\right) := \frac{1}{L}\sum_{\ell=1}^L \mathbb{D}\left(\mu^\ell,\sum_{i=1}^N \pi_i\delta_{y_i}\right).$$

for $\mathbb{D} = W_p^p,\ W_{\epsilon,p},\ SW_p^p,\ \text{max-}SW_p^p$. More specifically we investigate the **sample average approximation** of $\mathsf{F}_{\mathbb{D}}$.

Sparse optimal transport barycenters

This problem is called the **empirical** **sparse optimal transport barycenter** problem.

Sparse optimal transport barycenters

This problem is called the **empirical** **sparse optimal transport barycenter** problem.

We want to compute how much information is lost in average by computing

$$\min_{(Y,\pi)\in A} \mathrm{F}_{\mathbb{D}}\left(\mu_n^1, \ldots, \mu^L, \sum_{i=1}^{N} \pi_i \delta_{y_i}\right)$$

Sparse optimal transport barycenters

This problem is called the **empirical sparse optimal transport barycenter** problem.

We want to compute how much information is lost in average by computing

$$\min_{(Y,\pi)\in A} \mathrm{F}_{\mathbb{D}}\left(\mu_n^1, \dots, \mu^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right)$$

In stead of

$$\min_{(Y,\pi)\in A} \mathrm{F}_{\mathbb{D}}\left(\mu^1, \dots, \mu^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right).$$

Empirical risk minimization:

For a general loss function L

$$\mathbb{E}\left[L(f_n(X)) - \min_{f\in\mathcal{F}} \mathbb{E}[L(f(X))]\right] \leq 2\mathbb{E}\left[\sup_{f\in\mathcal{F}} \mid \mathbb{P}_n L(f) - \mathbb{E}[L(f(X))] \mid\right]$$
$$\lesssim \frac{\mathcal{C}(n)}{\sqrt{n}}$$

where f_n is a minimizer of $\mathbb{P}_n L(f)$ and $\mathcal{C}(n)$ is a measure of complexity (Rademacher, VC dimension, log-entropy, ...)

The K-means as an empirical risk minimization scheme

It was shown[10] that empirical optimal quantization (or K-means) behaves as

$$\mathbb{E}\left[D_2(\mu, Y_n) - \min_{Y \in \overline{\mathrm{B}_d(0,R)}^N} D_2(\mu, Y)\right] \lesssim \sqrt{\frac{N\log(N)}{n}}$$

where $D_2(\mu, Y) = \int_{\mathbb{R}^d} \min_{i=1,\ldots,N} \|x - y_i\|^2 d\mu(x)$.

[10] Principles of nonparametric learning, Section 4.4, L.Gyorfi. 2002.

The K-means as an empirical risk minimization scheme

It was shown[10] that empirical optimal quantization (or K-means) behaves as

$$\mathbb{E}\left[D_2(\mu, Y_n) - \min_{Y \in \overline{B_d(0,R)}^N} D_2(\mu, Y)\right] \lesssim \sqrt{\frac{N\log(N)}{n}}$$

where $D_2(\mu, Y) = \int_{\mathbb{R}^d} \min_{i=1,\dots,N} \|x - y_i\|^2 d\mu(x)$.

(Here $\mathcal{F} = \left\{ x \mapsto \min_{i=1,\dots,N} \|x - y_i\|^2 \mid Y \in \overline{B_d(0,R)}^N \right\}$).

[10] Principles of nonparametric learning, Section 4.4, L.Gyorfi. 2002.

The K-means as an empirical risk minimization scheme

It was shown[10] that empirical optimal quantization (or K-means) behaves as

$$\mathbb{E}\left[D_2(\mu, Y_n) - \min_{Y \in \overline{B_d(0,R)}^N} D_2(\mu, Y)\right] \lesssim \sqrt{\frac{N\log(N)}{n}}$$

where $D_2(\mu, Y) = \int_{\mathbb{R}^d} \min_{i=1,\dots,N} \|x - y_i\|^2 d\mu(x)$.

(Here $\mathcal{F} = \left\{x \mapsto \min_{i=1,\dots,N} \|x - y_i\|^2 \mid Y \in \overline{B_d(0,R)}^N\right\}$).

Question:

Does this result also hold for the more general problem

$$\min_{(Y,\pi)\in A} \frac{1}{L}\sum_{l=1}^{L} \mathbb{D}\left(\mu^l, \sum_{i=1}^{N} \pi_i \delta_{y_i}\right) \quad A \subset (\mathbb{R}^d)^N \times \Delta_N \text{ ?}$$

[10] Principles of nonparametric learning, Section 4.4, L.Gyorfi. 2002.

Main result

Theorem

- $\mu^1, ..., \mu^L \in \mathcal{M}_1\left(\overline{B_d(0, R)}\right)$.

Main result

Theorem

- $\mu^1, ..., \mu^L \in \mathcal{M}_1\left(\overline{B_d(0,R)}\right)$.
- $\mu_n^1, \ldots, \mu_n^L$ *empirical measures over i.i.d r.v of respective law* $\mu^1, \ldots \mu^L$.

Main result

Theorem

- $\mu^1, ..., \mu^L \in \mathcal{M}_1\left(\overline{B_d(0,R)}\right)$.
- $\mu_n^1, \ldots, \mu_n^L$ *empirical measures over i.i.d r.v of respective law* $\mu^1, \ldots \mu^L$.
- $\mathbb{D} = W_p^p,\ W_{p,\epsilon},\ SW_p^p$ *or max-*SW_p^p.

Main result

Theorem

- $\mu^1, ..., \mu^L \in \mathcal{M}_1\left(\overline{B_d(0,R)}\right)$.
- $\mu_n^1, \ldots, \mu_n^L$ *empirical measures over i.i.d r.v of respective law* $\mu^1, \ldots \mu^L$.
- $\mathbb{D} = W_p^p,\ W_{p,\epsilon},\ SW_p^p$ *or max-*SW_p^p.

Let $\nu_n^* := \sum_{i=1}^N \pi_i^n \delta_{y_i^n}$ *be a minimizer of* $\mathsf{F}_{\mathbb{D}}(\mu_n^1, \ldots, \mu_n^L, \nu)$ *over* $A \subset (\mathbb{R}^d)^N \times \Delta_N$. *Then*

Main result

Theorem

- $\mu^1, ..., \mu^L \in \mathcal{M}_1\left(\overline{B_d(0,R)}\right)$.
- $\mu_n^1, \ldots, \mu_n^L$ *empirical measures over i.i.d r.v of respective law* $\mu^1, \ldots \mu^L$.
- $\mathbb{D} = W_p^p,\ W_{p,\epsilon},\ SW_p^p$ *or max-*SW_p^p.

Let $\nu_n^* := \sum_{i=1}^N \pi_i^n \delta_{y_i^n}$ *be a minimizer of* $\mathsf{F}_{\mathbb{D}}(\mu_n^1, \ldots, \mu_n^L, \nu)$ *over* $A \subset (\mathbb{R}^d)^N \times \Delta_N$. *Then*

$$\mathbb{E}\left[\mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^N \pi_i^n \delta_{y_i^n}\right) - \min_{(Y,\pi)\in A} \mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right)\right]$$
$$\lesssim \sqrt{\frac{N}{n}}.$$

Tightness of the bound

Lemma (Bartlett et.al, 2002)

Let $\mathbb{D} = W_2^2$*,* $n \in \mathbb{N}^*$ *and* $R > 0$*. For any* $N \in \mathbb{N}^*$ *there exists* $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ *and* $\pi \in \Delta_N$ *such that*

$$\mathbb{E}\left[\mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^N \pi_i \delta_{y_i^n}\right) - \min_{Y \in \mathbb{B}_R^N} \mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^N \pi_i \delta_{y_i^*}\right)\right] \gtrsim \sqrt{\frac{N^{1-\frac{4}{d}}}{n}}$$

where $(y_1^n, \ldots, y_N^n)$ *denotes a minimizer of* $(y_1, \ldots, y_N) \mapsto \mathsf{F}_{W_2^2}\left(\mu_n^1, \ldots, \mu_n^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right)$.

Tightness of the bound

Lemma (Bartlett et.al, 2002)

Let $\mathbb{D} = W_2^2$, $n \in \mathbb{N}^*$ *and* $R > 0$. *For any* $N \in \mathbb{N}^*$ *there exists* $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ *and* $\pi \in \Delta_N$ *such that*

$$\mathbb{E}\left[\mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^{N} \pi_i \delta_{y_i^n}\right) - \min_{Y \in \mathbb{B}_R^N} \mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^{N} \pi_i \delta_{y_i^*}\right)\right] \gtrsim \sqrt{\frac{N^{1-\frac{4}{d}}}{n}}$$

where $(y_1^n, \ldots, y_N^n)$ *denotes a minimizer of*
$(y_1, \ldots, y_N) \mapsto \mathsf{F}_{W_2^2}\left(\mu_n^1, \ldots, \mu_n^L, \sum_{i=1}^{N} \pi_i \delta_{y_i}\right)$.

- Our bound is min-max optimal in n.

Tightness of the bound

Lemma (Bartlett et.al, 2002)

Let $\mathbb{D} = W_2^2$, $n \in \mathbb{N}^*$ *and* $R > 0$. *For any* $N \in \mathbb{N}^*$ *there exists* $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ *and* $\pi \in \Delta_N$ *such that*

$$\mathbb{E}\left[\mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^{N} \pi_i \delta_{y_i^n}\right) - \min_{Y \in \mathbb{B}_R^N} \mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^{N} \pi_i \delta_{y_i^*}\right)\right] \gtrsim \sqrt{\frac{N^{1-\frac{4}{d}}}{n}}$$

where $(y_1^n, \ldots, y_N^n)$ *denotes a minimizer of* $(y_1, \ldots, y_N) \mapsto \mathsf{F}_{W_2^2}\left(\mu_n^1, \ldots, \mu_n^L, \sum_{i=1}^{N} \pi_i \delta_{y_i}\right)$.

- Our bound is min-max optimal in n.
- But not optimal in N (depends heavily on the target measures).

Illustration of the convergence

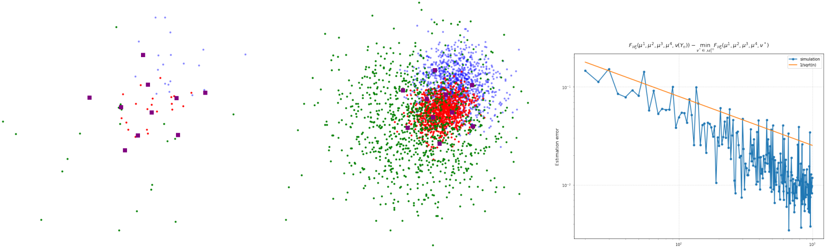


Figure: Empirical barycenter with increasing sample size.

Illustration of the convergence

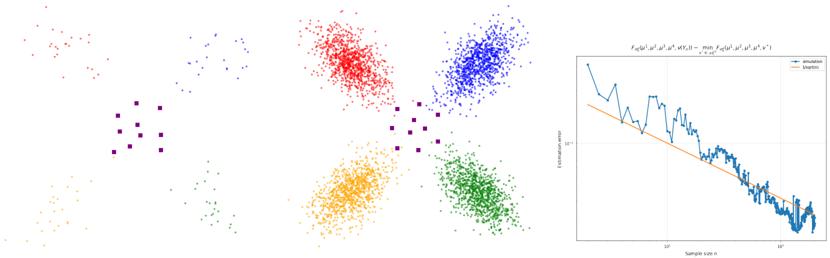


Figure: Empirical barycenter with increasing sample size.

Illustration of the convergence

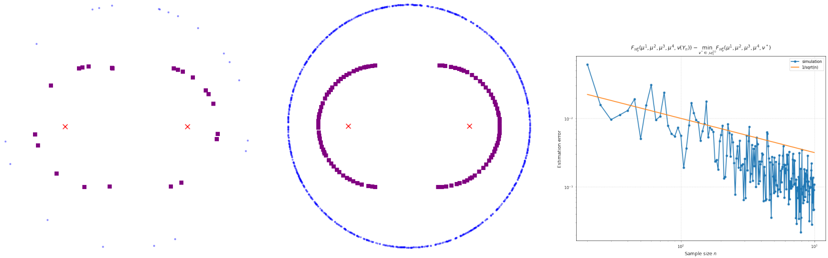


Figure: Empirical barycenter with increasing sample size.

Sketch of proof: generalization error

- Let $\nu_n^* := \sum_{i=1}^N \pi_i^n \delta_{y_i^n}$ be a minimizer of $\mathsf{F}_{\mathbb{D}}(\mu_n^1, \ldots, \mu_n^L, \nu(Y, \pi))$ over $A \subset (\mathbb{R}^d)^N \times \Delta_N$.
- Let $\nu^* := \sum_{i=1}^N \pi_i^* \delta_{y_i^*}$ be a minimizer of $\mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu(Y, \pi))$ over $A \subset (\mathbb{R}^d)^N \times \Delta_N$.

Then

$$\mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu_n^*) - \mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu^*)$$

Sketch of proof: generalization error

- Let $\nu_n^* := \sum_{i=1}^N \pi_i^n \delta_{y_i^n}$ be a minimizer of $\mathsf{F}_{\mathbb{D}}(\mu_n^1, \ldots, \mu_n^L, \nu(Y, \pi))$ over $A \subset (\mathbb{R}^d)^N \times \Delta_N$.
- Let $\nu^* := \sum_{i=1}^N \pi_i^* \delta_{y_i^*}$ be a minimizer of $\mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu(Y, \pi))$ over $A \subset (\mathbb{R}^d)^N \times \Delta_N$.

Then

$$\begin{aligned}
&\mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu_n^*) - \mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu^*) \\
&\leq \mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu_n^*)) - \mathsf{F}_{\mathbb{D}}(\mu_n^1, \ldots, \mu_n^L, \nu_n^*) \\
&+ \mathsf{F}_{\mathbb{D}}(\mu_n^1, \ldots, \mu_n^L, \nu^*) - \mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu^*)
\end{aligned}$$

Sketch of proof: generalization error

- Let $\nu_n^* := \sum_{i=1}^N \pi_i^n \delta_{y_i^n}$ be a minimizer of $\mathsf{F}_{\mathbb{D}}(\mu_n^1, \ldots, \mu_n^L, \nu(Y,\pi))$ over $A \subset (\mathbb{R}^d)^N \times \Delta_N$.
- Let $\nu^* := \sum_{i=1}^N \pi_i^* \delta_{y_i^*}$ be a minimizer of $\mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu(Y,\pi))$ over $A \subset (\mathbb{R}^d)^N \times \Delta_N$.

Then

$$
\begin{aligned}
&\mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu_n^*) - \mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu^*) \\
&\leq \mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu_n^*)) - \mathsf{F}_{\mathbb{D}}(\mu_n^1, \ldots, \mu_n^L, \nu_n^*) \\
&+ \mathsf{F}_{\mathbb{D}}(\mu_n^1, \ldots, \mu_n^L, \nu^*) - \mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu^*) \\
&\leq 2 \sup_{(Y,\pi)\in A} \left| \mathsf{F}_{\mathbb{D}}\left(\mu_n^1, \ldots, \mu_n^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right) - \mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right) \right|.
\end{aligned}
$$

Sketch of proof: generalization error

- Let $\nu_n^* := \sum_{i=1}^N \pi_i^n \delta_{y_i^n}$ be a minimizer of $\mathsf{F}_{\mathbb{D}}(\mu_n^1, \ldots, \mu_n^L, \nu(Y,\pi))$ over $A \subset (\mathbb{R}^d)^N \times \Delta_N$.
- Let $\nu^* := \sum_{i=1}^N \pi_i^* \delta_{y_i^*}$ be a minimizer of $\mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu(Y,\pi))$ over $A \subset (\mathbb{R}^d)^N \times \Delta_N$.

Then

$$
\begin{aligned}
&\mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu_n^*) - \mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu^*) \\
&\leq \mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu_n^*)) - \mathsf{F}_{\mathbb{D}}(\mu_n^1, \ldots, \mu_n^L, \nu_n^*) \\
&+ \mathsf{F}_{\mathbb{D}}(\mu_n^1, \ldots, \mu_n^L, \nu^*) - \mathsf{F}_{\mathbb{D}}(\mu^1, \ldots, \mu^L, \nu^*) \\
&\leq 2 \sup_{(Y,\pi)\in A} \left| \mathsf{F}_{\mathbb{D}}\left(\mu_n^1, \ldots, \mu_n^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right) - \mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right) \right|.
\end{aligned}
$$

(Generalization error)

Sketch of proof: control of the generalization error

For any fixed $(Y, \pi) \in (\mathbb{R}^d)^N \times \Delta_N$:

$$\left| \mathsf{F}_{\mathbb{D}}\left(\mu_n^1, \dots, \mu_n^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right) - \mathsf{F}_{\mathbb{D}}\left(\mu^1, \dots, \mu^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right) \right|$$

Sketch of proof: control of the generalization error

For any fixed $(Y, \pi) \in (\mathbb{R}^d)^N \times \Delta_N$:

$$\left| \mathrm{F}_{\mathbb{D}}\left(\mu_n^1, \dots, \mu_n^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right) - \mathrm{F}_{\mathbb{D}}\left(\mu^1, \dots, \mu^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right) \right|$$
$$\leq \frac{1}{L} \sum_{\ell=1}^L \left| W_p^p\left(\mu_n^\ell, \sum_{i=1}^N \pi_i \delta_{y_i}\right) - W_p^p\left(\mu^\ell, \sum_{i=1}^N \pi_i \delta_{y_i}\right) \right|$$

Sketch of proof: control of the generalization error

For any fixed $(Y, \pi) \in (\mathbb{R}^d)^N \times \Delta_N$:

$$\left| \mathsf{F}_{\mathbb{D}}\left(\mu_n^1, \ldots, \mu_n^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right) - \mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^N \pi_i \delta_{y_i}\right) \right|$$

$$\leq \frac{1}{L} \sum_{\ell=1}^L \left| W_p^p\left(\mu_n^\ell, \sum_{i=1}^N \pi_i \delta_{y_i}\right) - W_p^p\left(\mu^\ell, \sum_{i=1}^N \pi_i \delta_{y_i}\right) \right|$$

$$\leq \frac{1}{L} \sum_{\ell=1}^L \sup_{\|w\| \leq K} \left| \mathbb{P}_n^\ell f_{Y,w} - \mathbb{E}_{X^\ell \sim \mu^\ell}[f_{Y,w}(X^\ell)] \right|$$

Sketch of proof: Log-entropy

Taking the expectation on both sides

$$\mathbb{E}\left[\sup_{(Y,\pi)} \left| \mathsf{F}_{\mathbb{D}}\left(\mu_n^1, \ldots, \mu_n^L, \sum_{i=1}^{N} \pi_i \delta_{y_i}\right) - \mathsf{F}_{\mathbb{D}}\left(\mu^1, \ldots, \mu^L, \sum_{i=1}^{N} \pi_i \delta_{y_i}\right) \right|\right]$$

Sketch of proof: Log-entropy

Taking the expectation on both sides

$$\mathbb{E}\left[\sup_{(Y,\pi)}\left|\mathsf{F}_{\mathbb{D}}\left(\mu_n^1,\ldots,\mu_n^L,\sum_{i=1}^N \pi_i\delta_{y_i}\right)-\mathsf{F}_{\mathbb{D}}\left(\mu^1,\ldots,\mu^L,\sum_{i=1}^N \pi_i\delta_{y_i}\right)\right|\right]$$

$$\leq \frac{1}{L}\sum_{\ell=1}^L \mathbb{E}\left[\sup_{(Y,\pi,w)}\left|\frac{1}{n}\sum_{i=1}^n f_{Y,w}(X_i^\ell)-\mathbb{E}_{X^\ell\sim\mu^\ell}[f_{Y,w}(X^\ell)]\right|\right]$$

Sketch of proof: Log-entropy

Taking the expectation on both sides

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{(Y,\pi)}\left| \mathsf{F}_{\mathbb{D}}\left(\mu_n^1,\dots,\mu_n^L,\sum_{i=1}^{N}\pi_i\delta_{y_i}\right) - \mathsf{F}_{\mathbb{D}}\left(\mu^1,\dots,\mu^L,\sum_{i=1}^{N}\pi_i\delta_{y_i}\right)\right|\right] \\
&\leq \frac{1}{L}\sum_{\ell=1}^{L}\mathbb{E}\left[\sup_{(Y,\pi,w)}\left|\frac{1}{n}\sum_{i=1}^{n} f_{Y,w}(X_i^\ell) - \mathbb{E}_{X^\ell\sim\mu^\ell}[f_{Y,w}(X^\ell)]\right|\right] \\
&= \frac{1}{L}\sum_{\ell=1}^{L}\sup_{f\in\mathcal{F}_p}\left|\mathbb{P}_n^\ell(f) - \mathbb{E}_{X\sim\mu^\ell}[f(X)]\right|
\end{aligned}
$$

Sketch of proof: Log-entropy

Taking the expectation on both sides

$$\mathbb{E}\left[\sup_{(Y,\pi)}\left|\mathsf{F}_{\mathbb{D}}\left(\mu_n^1,\dots,\mu_n^L,\sum_{i=1}^N \pi_i\delta_{y_i}\right)-\mathsf{F}_{\mathbb{D}}\left(\mu^1,\dots,\mu^L,\sum_{i=1}^N \pi_i\delta_{y_i}\right)\right|\right]$$

$$\leq \frac{1}{L}\sum_{\ell=1}^L \mathbb{E}\left[\sup_{(Y,\pi,w)}\left|\frac{1}{n}\sum_{i=1}^n f_{Y,w}(X_i^\ell)-\mathbb{E}_{X^\ell\sim\mu^\ell}[f_{Y,w}(X^\ell)]\right|\right]$$

$$=\frac{1}{L}\sum_{\ell=1}^L \sup_{f\in\mathcal{F}_p}\left|\mathbb{P}_n^\ell(f)-\mathbb{E}_{X\sim\mu^\ell}[f(X)]\right|$$

$$\leq \frac{C_1}{\sqrt{n}}\mathbb{E}\left[\int_0^{C_2}\sqrt{\log\Big(2\mathcal{N}(\tau/2,\mathcal{F}_p,\|\cdot\|_{\mathbb{L}^2(\mu_n^\ell)})\Big)}\,d\tau\right]$$

Sketch of proof: Log-entropy

Taking the expectation on both sides

$$\mathbb{E}\left[\sup_{(Y,\pi)}\left| \mathrm{F}_{\mathbb{D}}\left(\mu_n^1,\ldots,\mu_n^L,\sum_{i=1}^N \pi_i\delta_{y_i}\right) - \mathrm{F}_{\mathbb{D}}\left(\mu^1,\ldots,\mu^L,\sum_{i=1}^N \pi_i\delta_{y_i}\right)\right|\right]$$

$$\leq \frac{1}{L}\sum_{\ell=1}^L \mathbb{E}\left[\sup_{(Y,\pi,w)}\left|\frac{1}{n}\sum_{i=1}^n f_{Y,w}(X_i^\ell) - \mathbb{E}_{X^\ell\sim\mu^\ell}[f_{Y,w}(X^\ell)]\right|\right]$$

$$= \frac{1}{L}\sum_{\ell=1}^L \sup_{f\in\mathcal{F}_p}\left|\mathbb{P}_n^\ell(f) - \mathbb{E}_{X\sim\mu^\ell}[f(X)]\right|$$

$$\leq \frac{C_1}{\sqrt{n}}\mathbb{E}\left[\int_0^{C_2}\sqrt{\log\left(2\mathcal{N}(\tau/2,\mathcal{F}_p,\|\cdot\|_{\mathbb{L}^2(\mu_n^\ell)})\right)}\,d\tau\right]$$

$$\lesssim \sqrt{\frac{N}{n}}$$

Conclusion

Although non-convex, problems of points cloud optimization with **OT** losses have good rigidity properties under definability assumption on the target measures. This allows for good behavior of gradient sequences (such as **Lloyd's algorithms**).

Furthermore these problems have sample complexity $O(n^{-1/2})$, independent (exponentially) of the dimension. This rate is optimal in general.

Thank you for your attention!

Statistical aspect of optimal transport

Study of asymptotic behavior of $\mathbb{D}(\mu, \mu_n)$ and $\mathbb{D}(\mu, \nu_n)$ $\nu \neq \mu$.

[11] On the rate of convergence in Wasserstein distance of the empirical measure, N.Fournier and A.Guillin. 2013

Statistical aspect of optimal transport

Study of asymptotic behavior of $\mathbb{D}(\mu, \mu_n)$ and $\mathbb{D}(\mu, \nu_n)$ $\nu \neq \mu$. In all generality it depends on dimensionality. For instance if μ's support is compact[11]:

$$\mathbb{E}\left[W_p(\mu_n, \mu)\right] \lesssim \begin{cases} n^{-1/d} & \text{if } d > 2p \\ n^{-1/2p} \log(n)^{1/p} & \text{if } d = 2p \\ n^{-1/2p} & \text{if } d < 2p \end{cases}$$

[11] On the rate of convergence in Wasserstein distance of the empirical measure, N.Fournier and A.Guillin. 2013

Statistical aspect of optimal transport

Study of asymptotic behavior of $\mathbb{D}(\mu, \mu_n)$ and $\mathbb{D}(\mu, \nu_n)$ $\nu \neq \mu$. In all generality it depends on dimensionality. For instance if μ's support is compact[11]:

$$\mathbb{E}\left[W_p(\mu_n, \mu)\right] \lesssim \begin{cases} n^{-1/d} & \text{if } d > 2p \\ n^{-1/2p} \log(n)^{1/p} & \text{if } d = 2p \\ n^{-1/2p} & \text{if } d < 2p \end{cases}$$

and by triangle inequality this is also true for for $\mid W_p(\mu_n, \nu) - W_p(\mu, \nu) \mid$.

[11] On the rate of convergence in Wasserstein distance of the empirical measure, N.Fournier and A.Guillin. 2013

Semi-discrete OT: fundamental $1/\sqrt{n}$ rate of convergence

Sharper results by investigating $\mathbb{D}(\mu_n, \nu)$ ($\mu \neq \nu$) insted. This is because the sample complexity of OT adapts to the "less complex" measure[12].

[12] Empirical optimal transport between different measures adapts to lower complexity, S.Hundrieser, T.Staudt and A.Munk.2022.

Semi-discrete OT: fundamental $1/\sqrt{n}$ rate of convergence

Sharper results by investigating $\mathbb{D}(\mu_n, \nu)$ $(\mu \neq \nu)$ insted. This is because the sample complexity of OT adapts to the "less complex" measure[12]. In particular if ν is discrete then

$$\mathbb{E}[\mid W_p^p(\mu_n, \nu) - W_p^p(\mu, \nu) \mid] = O(1/\sqrt{n}).$$

[12] Empirical optimal transport between different measures adapts to lower complexity, S.Hundrieser, T.Staudt and A.Munk.2022.

Statistical aspect of barycenters

Empirical Wasserstein barycenters are solutions of $\min_{\mu\in\mathcal{M}_1(\mathbb{R}^d)} \frac{1}{L}\sum_{l=1}^{L} W_p^p(\mu_n^l,\mu)$ with for all $l=1,\dots,L$ μ_n^l empirical measures over i.i.d realizations of μ^l.

[13] Quantitative stability of barycenters in the Wasserstein space, G.Carlier, A.Delalande and Q.Mérigot. 2022.

[14] Randomized Wasserstein barycenter computation: Resampling with statistical guarantees, F.Heinemann, A.Munk and Y.Zemel. 2023.

Statistical aspect of barycenters

Empirical Wasserstein barycenters are solutions of $\min_{\mu\in\mathcal{M}_1(\mathbb{R}^d)} \frac{1}{L}\sum_{l=1}^{L} W_p^p(\mu_n^l,\mu)$ with for all $l = 1,\dots,L$ μ_n^l empirical measures over i.i.d realizations of μ^l. They converge to the true barycenter, furthermore we have[13]

$$\mathbb{E}\left[W_2^2\left(\mu_n^*,\mu^*\right)\right] \lesssim \begin{cases} n^{-1/12} \text{ if } d<4 \\ n^{-1/12}\log(n)^{1/6} \text{ if } d=4 \\ n^{-1/3d} \text{ if } d>4. \end{cases}$$

[13] Quantitative stability of barycenters in the Wasserstein space, G.Carlier, A.Delalande and Q.Mérigot. 2022.

[14] Randomized Wasserstein barycenter computation: Resampling with statistical guarantees, F.Heinemann, A.Munk and Y.Zemel. 2023.

Statistical aspect of barycenters

Empirical Wasserstein barycenters are solutions of $\min_{\mu \in \mathcal{M}_1(\mathbb{R}^d)} \frac{1}{L} \sum_{l=1}^{L} W_p^p(\mu_n^l, \mu)$ with for all $l = 1, \ldots, L$ μ_n^l empirical measures over i.i.d realizations of μ^l. They converge to the true barycenter, furthermore we have[13]

$$\mathbb{E}\left[W_2^2\left(\mu_n^*, \mu^*\right)\right] \lesssim \begin{cases} n^{-1/12} \text{ if } d < 4 \\ n^{-1/12} \log(n)^{1/6} \text{ if } d = 4 \\ n^{-1/3d} \text{ if } d > 4. \end{cases}$$

In the discrete setting we have[14]

$$\mathbb{E}\left[\frac{1}{L}\sum_{l=1}^{L} W_p^p(\mu^l, \nu_n) - \min_{\nu \in \mathcal{M}_1(\mathbb{R}^d)} \frac{1}{L}\sum_{l=1}^{L} W_p^p(\mu^l, \nu)\right] \lesssim \frac{1}{\sqrt{n}}.$$

Where ν_n is an empirical barycenter.

[13] Quantitative stability of barycenters in the Wasserstein space, G.Carlier, A.Delalande and Q.Mérigot. 2022.

[14] Randomized Wasserstein barycenter computation: Resampling with statistical guarantees, F.Heinemann, A.Munk and Y.Zemel. 2023.