Barycenter in Wasserstein spaces: existence and consistency

Thibaut Le Gouic and Jean-Michel Loubes*

Institut de Mathématiques de Marseille
École Centrale Marseille
Institut Mathématique de Toulouse*

October 29th 2015

The barycenter of a set $\{x_i\}_{1\leq i\leq J}$ of $\mathbb{R}^d$ for J points endowed with weights $(\lambda_i)_{1\leq i\leq J}$ is defined as

$$\sum_{1\leq i\leq J} \lambda_i x_i.$$

It is characterized by being the minimizer of

$$x \mapsto \sum_{1\leq i\leq J} \lambda_i \|x - x_i\|^2.$$

The barycenter of a set $\{x_i\}_{1\leq i\leq J}$ of $\mathbb{R}^d$ for J points endowed with weights $(\lambda_i)_{1\leq i\leq J}$ is defined as

$$\sum_{1\leq i\leq J} \lambda_i x_i.$$

It is characterized by being the minimizer of

$$x \mapsto \sum_{1\leq i\leq J} \lambda_i \|x - x_i\|^2.$$

Replace $(\mathbb{R}^d, \|.\|)$ by a metric space (E, d), and minimize

$$x \mapsto \sum_{1\leq i\leq J} \lambda_i d(x, x_i)^2.$$

Likewise, given a random variable/vector of law μ on $\mathbb{R}^d$, its expectation $\mathbb{E}X$ is characterized by being the minimizer of

$$x \mapsto \mathbb{E}\|X - x\|^2.$$

Likewise, given a random variable/vector of law μ on $\mathbb{R}^d$, its expectation $\mathbb{E}X$ is characterized by being the minimizer of

$$x \mapsto \mathbb{E}\|X - x\|^2.$$

$\rightarrow$ extension to a metric space (it summarizes the information staying in a geodesic space)

Definition (p-barycenter)

Given a probability measure μ on a geodesic space (E, d), the set

$$\arg\min \left\{ x \in E; \int d(x,y)^p d\mu(y) \right\},$$

is called the set of p-**barycenters** *of* μ.

Definition (p-barycenter)

Given a probability measure μ on a geodesic space (E, d), the set

$$\arg\min \left\{ x \in E; \int d(x,y)^p d\mu(y) \right\},$$

*is called the set of p-**barycenters** of μ.*

- Existence ?

1. Geodesic space
2. Wasserstein space
3. Applications

Definition (Geodesic space)

A complete metric space (E, d) is said to be geodesic if for all $x, y \in E$, there exists $z \in E$ such that

$$\frac{1}{2}d(x, y) = d(x, z) = d(z, y).$$

Definition (Geodesic space)

A complete metric space (E, d) is said to be geodesic if for all $x, y \in E$, there exists $z \in E$ such that

$$\frac{1}{2}d(x, y) = d(x, z) = d(z, y).$$

- Include many spaces (vectorial normed spaces, compact manifolds, ...),

Proposition (Existence)

The p-barycenter of any probability measure on a **locally compact** *geodesic space, with finite moments of order p, exists.*

Proposition (Existence)

The p-barycenter of any probability measure on a **locally compact** *geodesic space, with finite moments of order p, exists.*

- Not unique e.g. the sphere
- Non positively curved space $\rightarrow$ unique barycenter, 1-Lipschitz on 2-Wasserstein space.

Definition (Wasserstein metric)

Let μ *and* ν *be two probability measures on a metric space* (E, d) *and* $p \geq 1$*.*
The p-**Wasserstein distance** *between* μ *and* ν *is defined as*

$$W_p^p(\mu, \nu) = \inf_{\pi \in \Gamma(\mu,\nu)} \int d_E(x, y)^p d\pi(x, y),$$

where $\Gamma(\mu, \nu)$ *is the set of all probability measures on* $E \times E$ *with marginals* μ *and* ν*.*

Definition (Wasserstein metric)

Let μ and ν be two probability measures on a metric space (E, d) and $p \geq 1$.
*The p-**Wasserstein distance** between μ and ν is defined as*

$$W_p^p(\mu, \nu) = \inf_{\pi \in \Gamma(\mu,\nu)} \int d_E(x, y)^p d\pi(x, y),$$

where $\Gamma(\mu, \nu)$ is the set of all probability measures on $E \times E$ with marginals μ and ν.

- Defined for any measure for which moments of order p are finite : $Ed(X, x_0)^p < \infty$ (denote this set $\mathcal{P}_p(E)$),
- It is a metric on $\mathcal{P}_p(E)$; $(\mathcal{P}_p(E), W_p)$ is called the **Wasserstein space**,
- The topology of this metric is the weak convergence topology and convergence of moments of order p.

- The Wasserstein space of a complete geodesic space is a complete geodesic space.
- $(\mathcal{P}_p(E), W_p)$ is locally compact $\Leftrightarrow$ (E, d) is compact.
- $(E, d) \subset (\mathcal{P}_p(E), W_p)$ isometrically.
- Existence of the barycenter on $(\mathcal{P}_p(E), W_p)$?

Definition (Measurable barycenter application)

Let (E, d) be a geodesic space. (E, d) is said to admit **measurable barycenter applications** *if for any $J \geq 1$ and any weights $(\lambda_j)_{1 \leq j \leq J}$, there exists a measurable application T from E^J to E such that for all $(x_1, ..., x_J) \in E^J$,*

$$\min_{x \in E} \sum_{j=1}^{J} \lambda_j d(x, x_j)^p = \sum_{j=1}^{J} \lambda_j d(T(x_1, ..., x_J), x_j)^p.$$

Definition (Measurable barycenter application)

Let (E, d) be a geodesic space. (E, d) is said to admit **measurable barycenter applications** *if for any $J \geq 1$ and any weights $(\lambda_j)_{1 \leq j \leq J}$, there exists a measurable application T from E^J to E such that for all $(x_1, ..., x_J) \in E^J$,*

$$\min_{x \in E} \sum_{j=1}^{J} \lambda_j d(x, x_j)^p = \sum_{j=1}^{J} \lambda_j d(T(x_1, ..., x_J), x_j)^p.$$

- Locally compact geodesic spaces admit measurable barycenter applications.

Theorem (Existence of barycenter)

Let (E, d) be a geodesic space that admits measurable barycenter applications. Then any probability measure $\mathbb{P}$ on $(\mathcal{P}_p(E), W_p)$ has a barycenter.

Theorem (Existence of barycenter)

Let (E, d) be a geodesic space that admits measurable barycenter applications. Then any probability measure $\mathbb{P}$ on $(\mathcal{P}_p(E), W_p)$ has a barycenter.

- Barycenter is not unique e.g. :
  $E = \mathbb{R}^d$ with $\mathbb{P} = \frac{1}{2}\delta_{\mu_1} + \frac{1}{2}\delta_{\mu_2}$,
  $\mu_1 = \frac{1}{2}\delta_{(-1,-1)} + \frac{1}{2}\delta_{(1,1)}$ and $\mu_2 = \frac{1}{2}\delta_{(1,-1)} + \delta_{(-1,1)}$

Theorem (Existence of barycenter)

Let (E, d) be a geodesic space that admits measurable barycenter applications. Then any probability measure $\mathbb{P}$ on $(\mathcal{P}_p(E), W_p)$ has a barycenter.

- Barycenter is not unique e.g. :
  $E = \mathbb{R}^d$ with $\mathbb{P} = \frac{1}{2}\delta_{\mu_1} + \frac{1}{2}\delta_{\mu_2}$,
  $\mu_1 = \frac{1}{2}\delta_{(-1,-1)} + \frac{1}{2}\delta_{(1,1)}$ and $\mu_2 = \frac{1}{2}\delta_{(1,-1)} + \delta_{(-1,1)}$
- Consistency of the barycenter ?

1. Multimarginal problem
2. Weak consistency
3. Approximation by finitely supported measures

Definition (Push forward)

Given a measure ν on E and an measurable application $T : E \to (F, \mathcal{F})$, the push forward of ν by T is given by

$$T_{\#\nu}(A) = \nu\left(T^{-1}(A)\right), \forall A \in \mathcal{F}.$$

- Probabilist version : X is a r.v. on $(\Omega, \mathcal{A}, \mathbb{P})$, then $\mathbb{P}_X = X_{\#\mathbb{P}}$.

Theorem (Barycenter and multi-marginal problem [Agueh and Carlier, 2011])

Let (E, d) be a complete separable geodesic space, $p \geq 1$ and $J \in \mathbb{N}^*$. Given $(\mu_i)_{1 \leq i \leq J} \in \mathcal{P}_p(E)^J$ and weights $(\lambda_i)_{1 \leq i \leq J}$, there exists a measure $\gamma \in \Gamma(\mu_1, ..., \mu_J)$ minimizing

$$\hat{\gamma} \mapsto \int \inf_{x \in E} \sum_{1 \leq i \leq J} \lambda_i d(x_i, x)^p d\hat{\gamma}(x_1, ..., x_J).$$

- If (E, d) admits a measurable barycenter application $T : E^J \to E$ then the measure $\nu = T_{\#}\gamma$ is a barycenter of $(\mu_i)_{1 \leq i \leq J}$
- If T is unique, ν is of the form $\nu = T_{\#}\gamma$.

Theorem (Weak consistency of the barycenter)

Let (E, d) be a geodesic space that admits measurable barycenter. Take $(\mathbb{P}_j)_{j\geq 1} \subset \mathcal{P}_p(E)$ converging to $\mathbb{P} \in \mathcal{P}_p(E)$. Take any barycenter μ_j of $\mathbb{P}_j$.
Then the sequence $(\mu^j)_{j\geq 1}$ is (weakly) tight and any limit point is a barycenter of $\mathbb{P}$.

Proposition (Approximation by finitely supported measure)

For any measure $\mathbb{P}$ on $\mathcal{P}_p(E)$ there exists a sequence of finitely supported measures $(\mathbb{P}_j)_{j\geq 1} \subset \mathcal{P}_p(E)$ such that

$$W_p(\mathbb{P}_j, \mathbb{P}) \to 0 \text{ as } j \to \infty.$$

1. Multimarginal problem

2. Weak consistency

3. Approximation by finitely supported measures

1. Multimarginal problem
   $\rightarrow$ *existence of barycenter for* $\mathbb{P}$ *finitely supported.*
2. Weak consistency
3. Approximation by finitely supported measures

1. Multimarginal problem
   → *existence of barycenter for* $\mathbb{P}$ *finitely supported.*
2. Weak consistency
   → *existence of barycenter for probabilities that can be approximated by measures with barycenters.*
3. Approximation by finitely supported measures

1. Multimarginal problem
   $\rightarrow$ *existence of barycenter for* $\mathbb{P}$ *finitely supported.*
2. Weak consistency
   $\rightarrow$ *existence of barycenter for probabilities that can be approximated by measures with barycenters.*
3. Approximation by finitely supported measures
   $\rightarrow$ *any probability can be approximated by a finitely supported probability measure.*

Theorem (Consistency of the barycenter)

Let (E, d) be a geodesic space that admits measurable barycenter. Take $(\mathbb{P}_j)_{j\geq 1} \subset \mathcal{P}_p(E)$ and $\mathbb{P} \in \mathcal{P}_p(E)$. Take any barycenter μ_j of $\mathbb{P}_j$.
Then the sequence $(\mu_j)_{j\geq 1}$ is totally bounded in $(\mathcal{P}_p(E), W_p)$ and any limit point is a barycenter of $\mathbb{P}$.

Theorem (Consistency of the barycenter)

Let (E, d) be a geodesic space that admits measurable barycenter. Take $(\mathbb{P}_j)_{j\geq 1} \subset \mathcal{P}_p(E)$ and $\mathbb{P} \in \mathcal{P}_p(E)$. Take any barycenter μ_j of $\mathbb{P}_j$.
Then the sequence $(\mu_j)_{j\geq 1}$ is totally bounded in $(\mathcal{P}_p(E), W_p)$ and any limit point is a barycenter of $\mathbb{P}$.

- Imply continuity of barycenter when barycenter are unique.
- No rate of convergence (barycenter Lipschitz on $(E, d) \not\Rightarrow$ Lipschitz on $\mathcal{P}_p(E)$).
- Imply compactness of the set of barycenters.

Take $(\mu_i^n)_{1\leq j\leq J} \to \mu_j$ when $n \to \infty$ and weights $(\lambda_j)_{1\leq j\leq J}$.
Set μ_B^n the barycenter of $(\mu_i^n)_{1\leq j\leq J}$.
Then, as $n \to \infty$,

$$\mu_B^n \to \mu_B.$$

Take $(\mu_i^n)_{1\leq j\leq J} \to \mu_j$ when $n \to \infty$ and weights $(\lambda_j)_{1\leq j\leq J}$.
Set μ_B^n the barycenter of $(\mu_i^n)_{1\leq j\leq J}$.
Then, as $n \to \infty$,

$$\mu_B^n \to \mu_B.$$

- Texture mixing [Rabin et al., 2011]

Take $(\mu_n)_{n\geq 1}$ such that

$$\frac{1}{n}\sum_{i=1}^{n}\mu_i \to \mathbb{P}.$$

Set μ_B^n the barycenter of

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{\mu_i}.$$

Then, as $n\to\infty$,

$$\mu_B^n \to \mu_B$$

Take $(\mu_n)_{n\geq 1}$ such that

$$\frac{1}{n}\sum_{i=1}^{n}\mu_i \to \mathbb{P}.$$

Set μ_B^n the barycenter of

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{\mu_i}.$$

Then, as $n \to \infty$,

$$\mu_B^n \to \mu_B$$

- Average of template deformation
  [Bigot and Klein, 2012],[Agulló-Antolín et al., 2015]

Agueh, M. and Carlier, G. (2011).
Barycenters in the wasserstein space.
SIAM Journal on Mathematical Analysis, 43(2) :904–924.

Agulló-Antolín, M., Cuesta-Albertos, J. A., Lescornel, H., and Loubes, J.-M. (2015).
A parametric registration model for warped distributions with Wasserstein's distance.
J. Multivariate Anal., 135 :117–130.

Bigot, J. and Klein, T. (2012).
Consistent estimation of a population barycenter in the Wasserstein space.
ArXiv e-prints.

Rabin, J., Peyré, G., Delon, J., and Bernot, M. (2011).
Wasserstein Barycenter and its Application to Texture Mixing.
SSVM'11, pages 435–446.