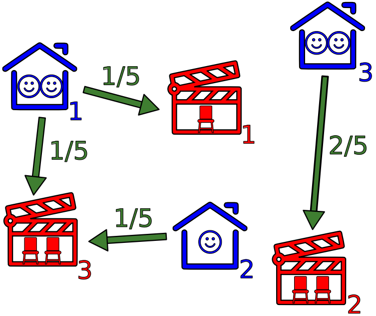
Computing Optimal Transport Barycentres

Eloi Tanguy, Julie Delon, Nathaël Gozlan

MAP5, Université Paris-Cité

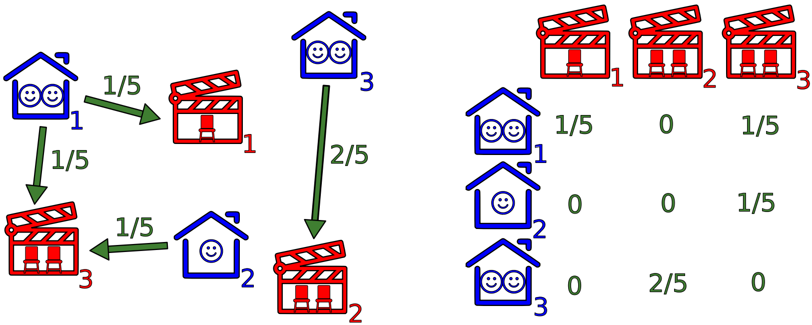
March 4, 2025

Discrete Optimal Transport



	1	2	3
1	1/5	0	1/5
2	0	0	1/5
3	0	2/5	0

Discrete Optimal Transport



Assignment Cost:

$$\frac{1}{5} \times c(x_1, y_1) + \frac{1}{5} \times c(x_1, y_3) + \frac{1}{5} \times c(x_2, y_3) + \frac{2}{5} \times c(x_3, y_2).$$

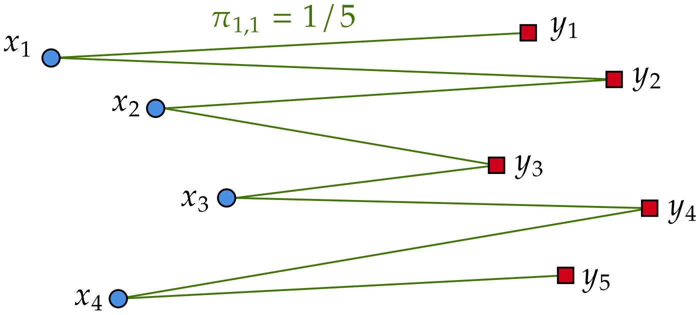
Constraints on $\pi \in \mathbb{R}_+^{3\times 3}$: $\pi \mathbf{1} = (2/5, 1/5, 2/5)$, $\pi^\top \mathbf{1} = (1/5, 2/5, 2/5)$.

$$\text{Optimal Transport Cost :} \ \min_{\pi} \sum_{i,j} c(x_i, y_j)\pi_{i,j}.$$

OT between discrete measures

$$\mu = \sum_{i=1}^{n} a_i \delta_{x_i}, \;\; \nu = \sum_{j=1}^{m} b_j \delta_{y_j}$$

$$\mathcal{T}_c(\mu, \nu) = \inf_{\pi \in \Pi(a,b)} \sum_{i,j} c(x_i, y_j) \pi_{i,j}.$$



OT Cost and 2-Wasserstein Distance

$$\mathcal{T}_c(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\mathrm{d}\pi(x,y) = \inf_{X\sim\mu, Y\sim\nu} \mathbb{E}[c(X,Y)].$$

$$\mathrm{W}_2^2(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} \|x-y\|_2^2\mathrm{d}\pi(x,y).$$

OT Cost and 2-Wasserstein Distance

$$\mathcal{T}_c(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\mathrm{d}\pi(x,y) = \inf_{X\sim\mu, Y\sim\nu} \mathbb{E}[c(X,Y)].$$

$$\mathrm{W}_2^2(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} \|x-y\|_2^2\mathrm{d}\pi(x,y).$$

Bures-Wasserstein

$$\mathrm{W}_2^2(\mathcal{N}(m_1,S_1),\mathcal{N}(m_2,S_2))$$
$$= \|m_1-m_2\|_2^2$$
$$+\mathrm{Tr}\left(S_1+S_2-2(S_1^{1/2}S_2S_1^{1/2})^{1/2}\right)$$

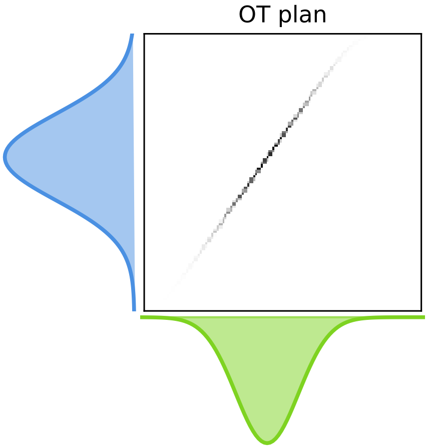
OT Cost and 2-Wasserstein Distance

$$\mathcal{T}_c(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\mathrm{d}\pi(x,y) = \inf_{X\sim\mu, Y\sim\nu} \mathbb{E}[c(X,Y)].$$

$$\mathrm{W}_2^2(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} \|x-y\|_2^2 \mathrm{d}\pi(x,y).$$

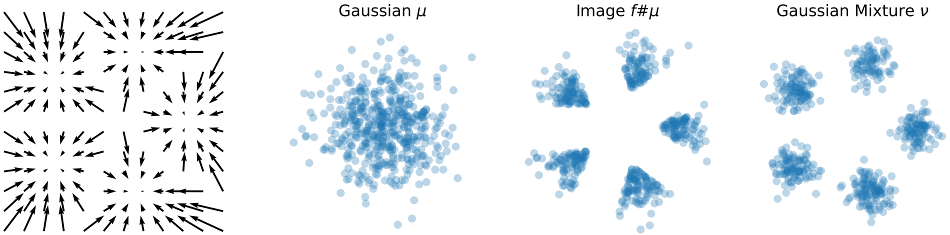
Bures-Wasserstein

$$\mathrm{W}_2^2(\mathcal{N}(m_1,S_1),\mathcal{N}(m_2,S_2))$$
$$= \|m_1 - m_2\|_2^2$$
$$+\mathrm{Tr}\left(S_1 + S_2 - 2(S_1^{1/2}S_2S_1^{1/2})^{1/2}\right)$$



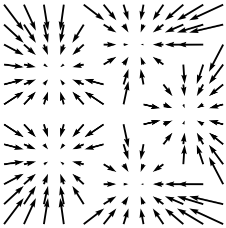
Push-forward measures and OT maps

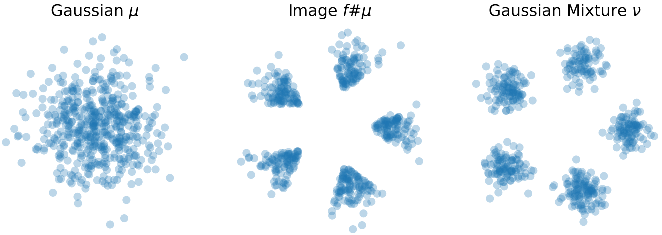
Image Measure: $f\#\mu := \mathrm{Law}_{X\sim\mu}[f(X)]$



Push-forward measures and OT maps

Image Measure: $f\#\mu := \mathrm{Law}_{X\sim\mu}[f(X)]$



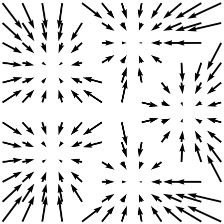


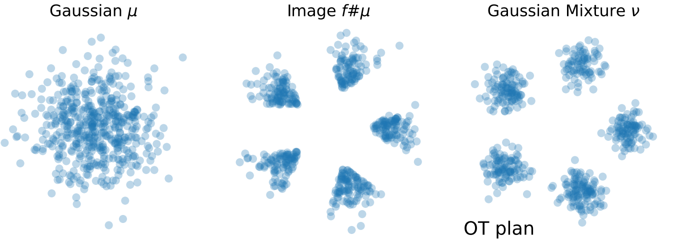
Brenier's Theorem

If $c(x,y) = \|x-y\|_2^2$, and $\mu \ll \mathscr{L}^d$, then there is a unique solution $\pi^* = (I, \nabla\varphi)\#\mu$, with φ convex.

Push-forward measures and OT maps

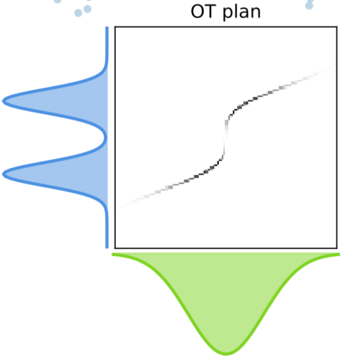
Image Measure: $f\#\mu := \mathrm{Law}_{X\sim\mu}[f(X)]$





Brenier's Theorem

If $c(x,y) = \|x-y\|_2^2$, and $\mu \ll \mathscr{L}^d$, then there is a unique solution $\pi^* = (I, \nabla\varphi)\#\mu$, with φ convex.

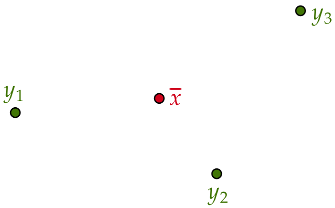


1. Optimal Transport
2. Wasserstein Barycentres
3. OT Barycentres
4. Discrete Case and Numerics
5. Application to GMMs

From Euclidean Combinations to Fréchet Means

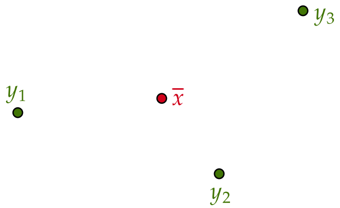
$$\overline{x} = \sum_{k=1}^{K} \lambda_k y_k$$

$$\overline{x} = \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{k=1}^{K} \lambda_k \|x - y_k\|_2^2$$



From Euclidean Combinations to Fréchet Means

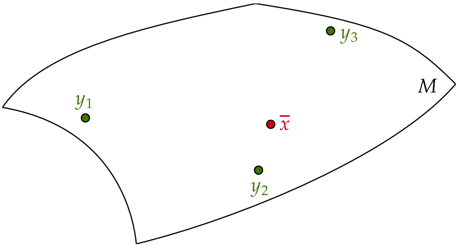
$$\overline{x} = \sum_{k=1}^{K} \lambda_k y_k$$



$$\overline{x} = \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{k=1}^{K} \lambda_k \|x - y_k\|_2^2$$

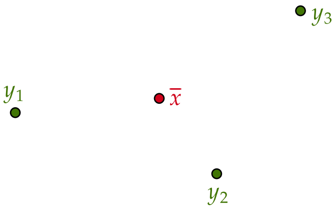
Fréchet mean:

$$\overline{x} \in \underset{x \in \mathcal{X}}{\operatorname{argmin}} \sum_{k=1}^{K} d(x, y_k)^2.$$



From Euclidean Combinations to Fréchet Means

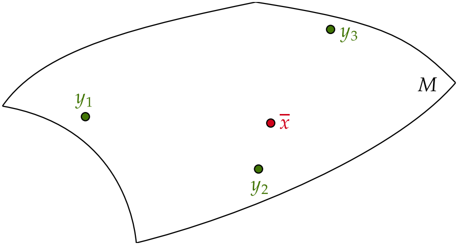
$$\overline{x} = \sum_{k=1}^{K} \lambda_k y_k$$



$$\overline{x} = \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{k=1}^{K} \lambda_k \|x - y_k\|_2^2$$

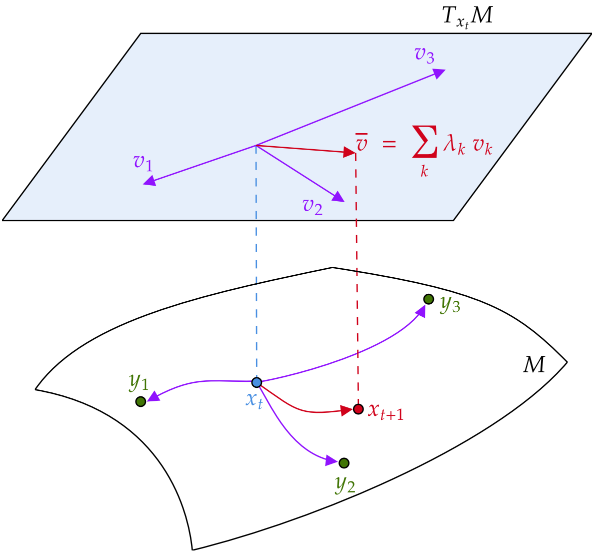
Fréchet mean:

$$\overline{x} \in \underset{x \in \mathcal{X}}{\operatorname{argmin}} \sum_{k=1}^{K} d(x, y_k)^2.$$

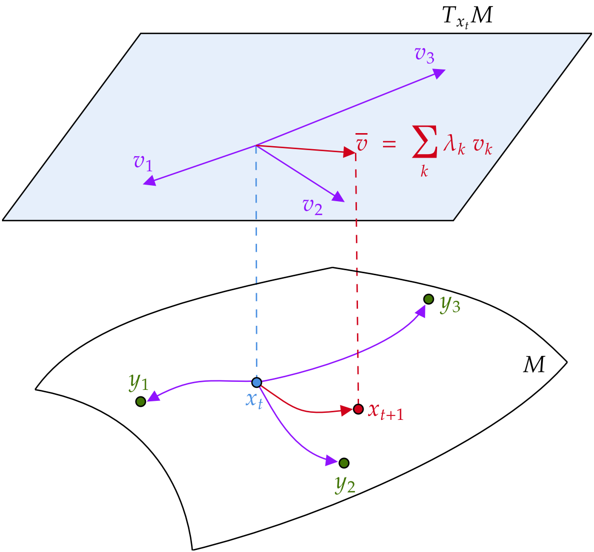


Generalisation: $\overline{x} \in \underset{x \in \mathcal{X}}{\operatorname{argmin}} \sum_{k=1}^{K} c_k(x, y_k)$.

Fixed-Point Algorithm for Fréchet Means on Manifolds



Fixed-Point Algorithm for Fréchet Means on Manifolds



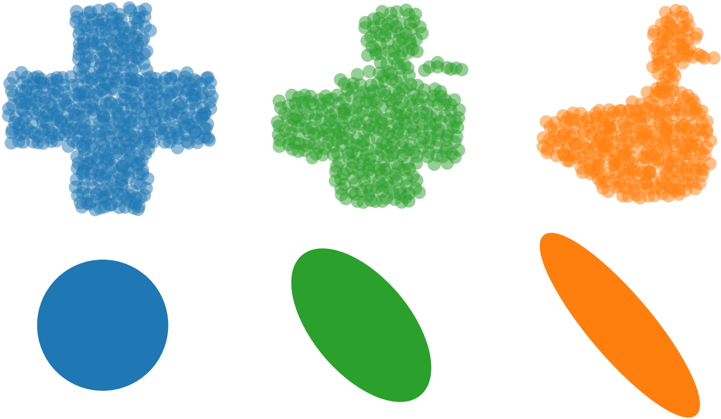
$$V(x) = \sum_{k=1}^{K} \lambda_k d(x, y_k)^2.$$

$$\nabla V(x) = -2 \sum_{k=1}^{K} \lambda_k \mathrm{Log}_x(y_k).$$

$$x_{t+1} = \mathrm{Exp}_{x_t}\left(-\frac{1}{2}\nabla V(x_t)\right).$$

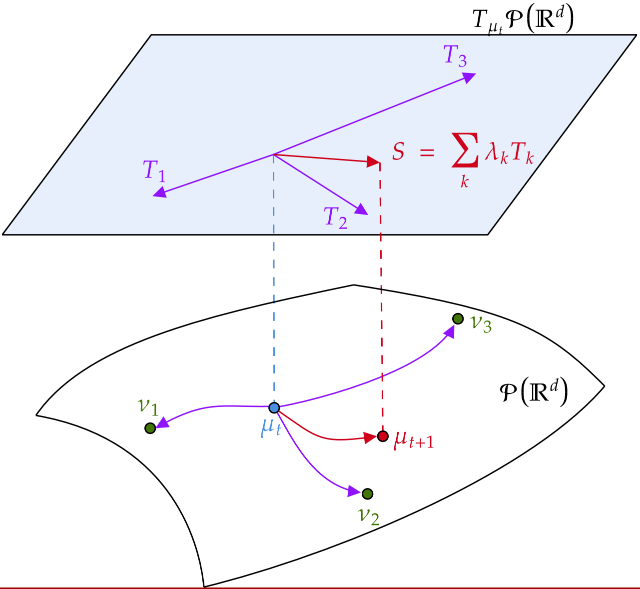
2-Wasserstein Barycentres (Agueh & Carlier 2011 [1])

$$\underset{\mu \in \mathcal{P}(\mathbb{R}^d)}{\operatorname{argmin}} \sum_{k=1}^{K} \lambda_k \mathrm{W}_2^2(\mu, \nu_k).$$



Fixed-Point Method (Alvarez-Esteban et al. 2016 [3])

Assumptions: $c(x, y) = \|x - y\|_2^2$, AC measures on $\mathbb{R}^d$.

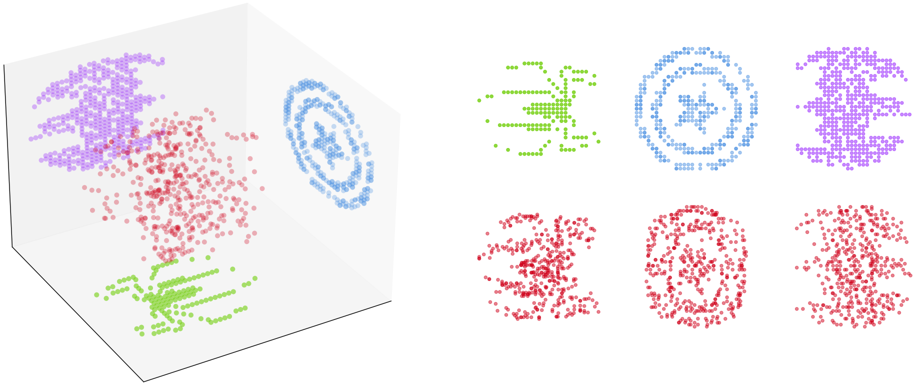


1. Optimal Transport
2. Wasserstein Barycentres
3. OT Barycentres
4. Discrete Case and Numerics
5. Application to GMMs

Motivation for OT barycenters with generic costs

$$\mathrm{W}_1(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \int \|x-y\|_2 \mathrm{d}\pi(x,y).$$

Find $\mu \in \mathcal{P}(\mathbb{R}^3)$ minimising $\sum_k \frac{1}{3}\mathrm{W}_1(P_k\#\mu,\nu_k)$ where $\nu_k \in \mathcal{P}(\mathbb{R}^2)$.



Generalises Delon et al. 2021 [5] where $c_k(x,y) = \|P_k(x)-y\|_2^2$.

Generalising Wasserstein Barycentres

Setting:

- $(\mathcal{X}, d_{\mathcal{X}})$ compact metric space for barycentres.
- $(\mathcal{Y}_k, d_{\mathcal{Y}_k})$ compact metric spaces for measures ν_k.
- $c_k : \mathcal{X} \times \mathcal{Y}_k \longrightarrow \mathbb{R}_+$ continuous cost functions.

Generalising Wasserstein Barycentres

Setting:

- $(\mathcal{X}, d_{\mathcal{X}})$ compact metric space for barycentres.
- $(\mathcal{Y}_k, d_{\mathcal{Y}_k})$ compact metric spaces for measures ν_k.
- $c_k : \mathcal{X} \times \mathcal{Y}_k \longrightarrow \mathbb{R}_+$ continuous cost functions.

$$\underset{\mu \in \mathcal{P}(\mathcal{X})}{\operatorname{argmin}}\ V(\mu), \quad V(\mu) := \sum_{k=1}^{K} \mathcal{T}_{c_k}(\mu, \nu_k).$$

Generalising Wasserstein Barycentres

Setting:

- $(\mathcal{X}, d_{\mathcal{X}})$ compact metric space for barycentres.
- $(\mathcal{Y}_k, d_{\mathcal{Y}_k})$ compact metric spaces for measures ν_k.
- $c_k : \mathcal{X} \times \mathcal{Y}_k \longrightarrow \mathbb{R}_+$ continuous cost functions.

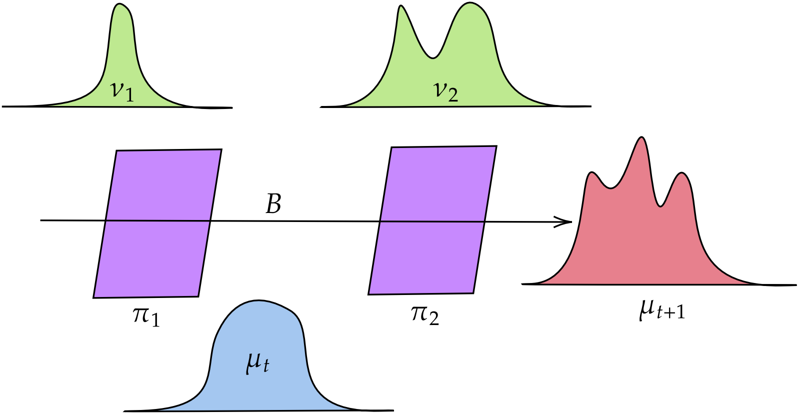
$$\underset{\mu \in \mathcal{P}(\mathcal{X})}{\operatorname{argmin}}\ V(\mu), \quad V(\mu) := \sum_{k=1}^{K} \mathcal{T}_{c_k}(\mu, \nu_k).$$

Assumption: The ground barycenter function

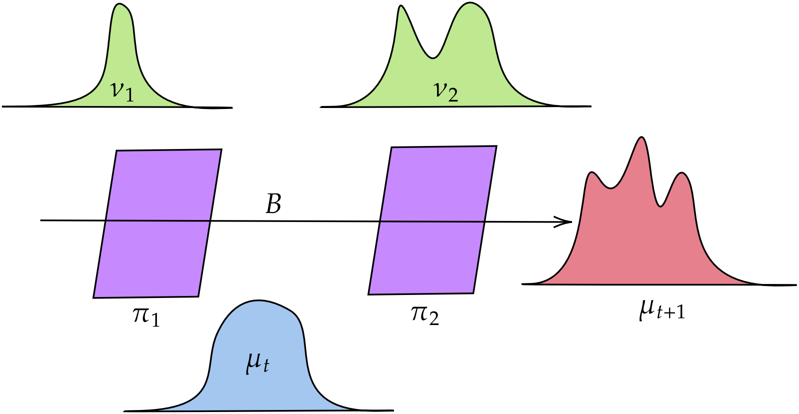
$$B(y_1, \cdots, y_K) := \underset{x \in \mathcal{X}}{\operatorname{argmin}} \sum_{k=1}^{K} c_k(x, y_k)$$

is well-defined.

Fixed-Point Algorithm: Intuition



Fixed-Point Algorithm: Intuition

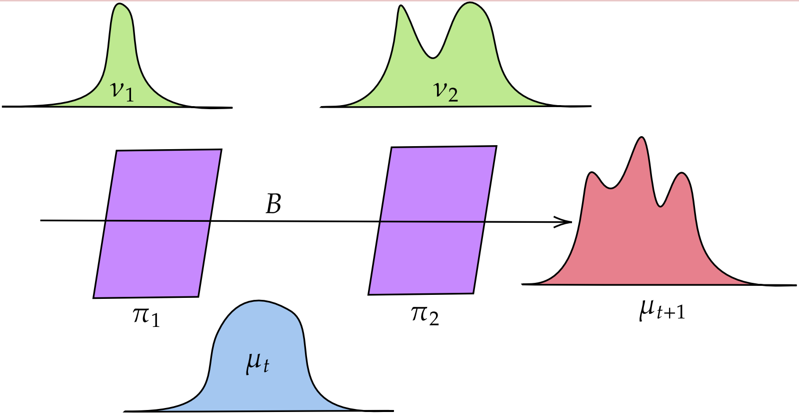


General Idea

Let $(X_t, Y_1, \cdots, Y_K)$ RVs such that $X_t \sim \mu_t$, $Y_k \sim \nu_k$ and $(X_t, Y_k) \sim \pi_k \in \Pi^*_{c_k}(\mu_t, \nu_k)$. Take $X_{t+1} = B(Y_1, \cdots, Y_K)$.

If $\Pi^*_{c_k}(\mu_t, \nu_k) = \{(I, T_k)\#\mu_t\}$ then $\mu_{t+1} = B(T_1, \cdots, T_K)\#\mu_t$.

Fixed-point Algorithm: (more) formal definition



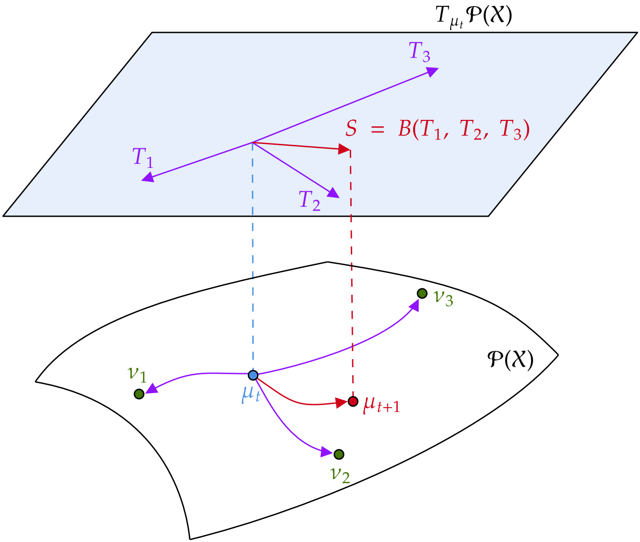
$$\Gamma(\mu) := \left\{\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}_1 \cdots \times \mathcal{Y}_K) : \ \forall k \in [\![1, K]\!],\ \gamma_{0,k} \in \Pi^*_{c_k}(\mu, \nu_k)\right\},$$

$$G := \left\{ \begin{array}{ccc} \mathcal{P}(\mathcal{X}) & \rightrightarrows & \mathcal{P}(\mathcal{X}) \\ \mu & \mapsto & B\#\Gamma(\mu) \end{array} \right., \quad \mu_{t+1} \in G(\mu_t).$$

$$B\#\Gamma(\mu) := \{B\#\gamma,\ \gamma \in \Gamma(\mu)\},\ B\#\gamma = \mathrm{Law}_{(X,Y_1,\cdots,Y_K)\sim\gamma} B(Y_1, \cdots, Y_K).$$

Relation to Alvarez-Esteban et al. 2016 [3]

Dream case: $\mathcal{X} = \mathcal{Y}_1 = \cdots = \mathcal{Y}_K$ and maps exist.



Reality:

$$\gamma : \gamma_{0,k} \in \Pi^*_{c_k}(\mu_t, \nu_k),$$

$$\mu_{t+1} = B\#\gamma.$$

Algorithm Convergence

Ground Barycentre Lemma

$$\sum_k c_k(x, y_k) \geq \sum_k c_k(B(y_1, \cdots, y_K), y_k) + \delta(x, B(y_1, \cdots, y_K)).$$

Case $\|x - y\|_2^2$: simply $\sum_k \lambda_k \|x - y_k\|_2^2 = \sum_k \|\overline{x} - y_k\|_2^2 + \|x - \overline{x}\|_2^2$.

Algorithm Convergence

Ground Barycentre Lemma

$$\sum_k c_k(x, y_k) \geq \sum_k c_k(B(y_1, \cdots, y_K), y_k) + \delta(x, B(y_1, \cdots, y_K)).$$

Case $\|x - y\|_2^2$: simply $\sum_k \lambda_k \|x - y_k\|_2^2 = \sum_k \|\overline{x} - y_k\|_2^2 + \|x - \overline{x}\|_2^2$.

Decrease Property

$$\forall \overline{\mu} \in G(\mu),\ V(\mu) \geq V(\overline{\mu}) + \mathcal{T}_\delta(\mu, \overline{\mu}).$$

If μ^* is a barycentre then $G(\mu^*) = \{\mu^*\}$.

Algorithm Convergence

Ground Barycentre Lemma

$$\sum_k c_k(x, y_k) \geq \sum_k c_k(B(y_1, \cdots, y_K), y_k) + \delta(x, B(y_1, \cdots, y_K)).$$

Case $\|x-y\|_2^2$: simply $\sum_k \lambda_k \|x - y_k\|_2^2 = \sum_k \|\overline{x} - y_k\|_2^2 + \|x - \overline{x}\|_2^2$.

Decrease Property

$$\forall \overline{\mu} \in G(\mu),\ V(\mu) \geq V(\overline{\mu}) + \mathcal{T}_\delta(\mu, \overline{\mu}).$$

If μ^* is a barycentre then $G(\mu^*) = \{\mu^*\}$.

Using arcane magic about the regularity of the multimap G:

Convergence

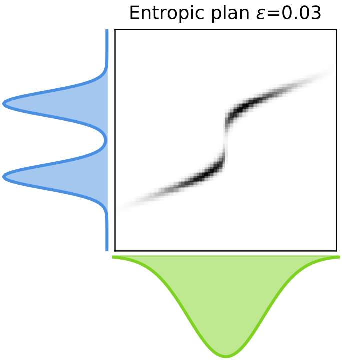
If μ is a subsequential limit of (μ_t) then $\mu \in G(\mu)$.

Entropic Barycentres

$$\mathcal{T}_{c,\varepsilon}(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c\mathrm{d}\pi + \varepsilon\,\mathrm{KL}(\pi|\mu\otimes\nu).$$

$$V_\varepsilon(\mu) := \sum_{k=1}^{K} \mathcal{T}_{c,\varepsilon}(\mu,\nu_k).$$

$$G_\varepsilon(\mu) := B\#\gamma, \text{ with } \gamma_{0,k} = \Pi^*_{c_k,\varepsilon}(\mu,\nu_k).$$

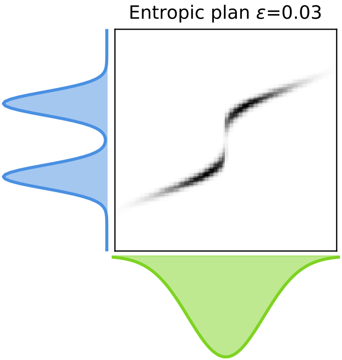


Entropic Barycentres

$$\mathcal{T}_{c,\varepsilon}(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c\mathrm{d}\pi + \varepsilon\,\mathrm{KL}(\pi|\mu\otimes\nu).$$

$$V_\varepsilon(\mu) := \sum_{k=1}^K \mathcal{T}_{c,\varepsilon}(\mu,\nu_k).$$

$$G_\varepsilon(\mu) := B\#\gamma, \text{ with } \gamma_{0,k} = \Pi^*_{c_k,\varepsilon}(\mu,\nu_k).$$



Decrease Property

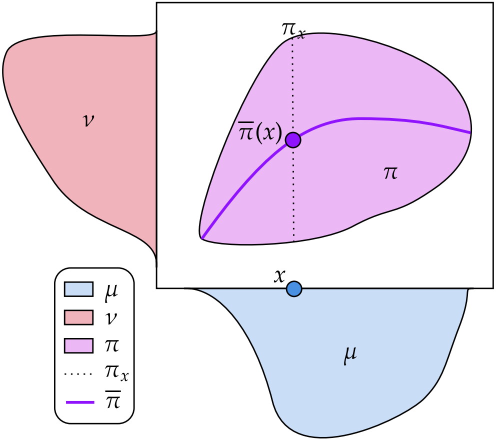
$V_\varepsilon(\mu) \geq V_\varepsilon(G_\varepsilon(\mu)) + \mathcal{T}_\delta(\mu, G_\varepsilon(\mu))$. If μ^* barycentre, $G_\varepsilon(\mu^*) = \mu^*$.

Convergence

If μ is a subsequential limit of (μ_t) then $\mu = G_\varepsilon(\mu)$.

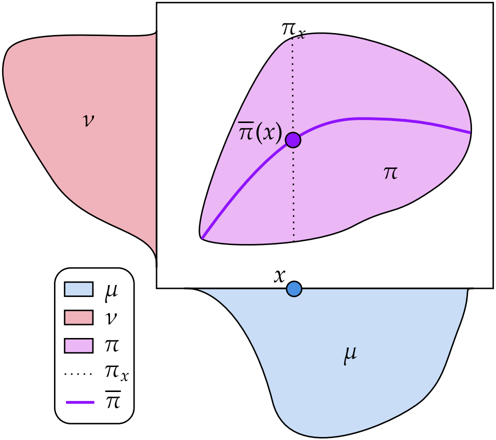
Barycentric Projections

Replace a coupling π with a map $\overline{\pi}$.



Barycentric Projections

Replace a coupling π with a map $\overline{\pi}$.



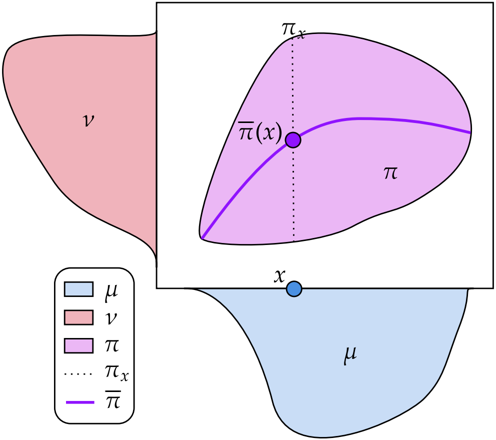
$$\overline{\pi}(x) = \int y \mathrm{d}\pi_x(y).$$

$$\overline{\pi}(x) = \mathbb{E}_{(X,Y)\sim\pi}[Y|X = x].$$

$$\overline{\pi} = \underset{f\in L^2(\mu)}{\mathrm{argmin}} \int \|f(x)-y\|_2^2 \mathrm{d}\pi(x,y).$$

Barycentric Projections

Replace a coupling π with a map $\overline{\pi}$.



$$\overline{\pi}(x) = \int y \mathrm{d}\pi_x(y).$$

$$\overline{\pi}(x) = \mathbb{E}_{(X,Y)\sim\pi}[Y|X = x].$$

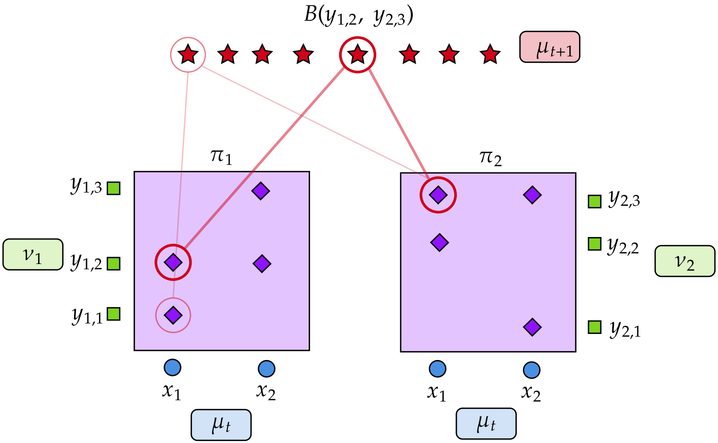
$$\overline{\pi} = \underset{f\in L^2(\mu)}{\operatorname{argmin}} \int \|f(x)-y\|_2^2 \mathrm{d}\pi(x,y).$$

$$H(\mu) = \left\{B(\overline{\pi_1}, \cdots, \overline{\pi_K})\#\mu,\ \pi_k \in \Pi^*_{c_k}(\mu, \nu_k)\right\}.$$

No guarantees.

1. Optimal Transport
2. Wasserstein Barycentres
3. OT Barycentres
4. Discrete Case and Numerics
5. Application to GMMs

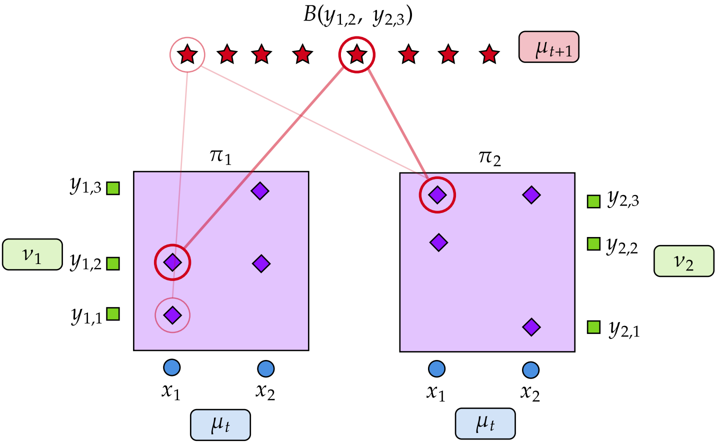
Discrete G



$$\mu = \sum_{i=1}^{n} a_i \delta_{x_i}$$

$$\nu_k = \sum_{j=1}^{n_k} b_{k,j} \delta_{y_{k,j}}$$

Discrete G

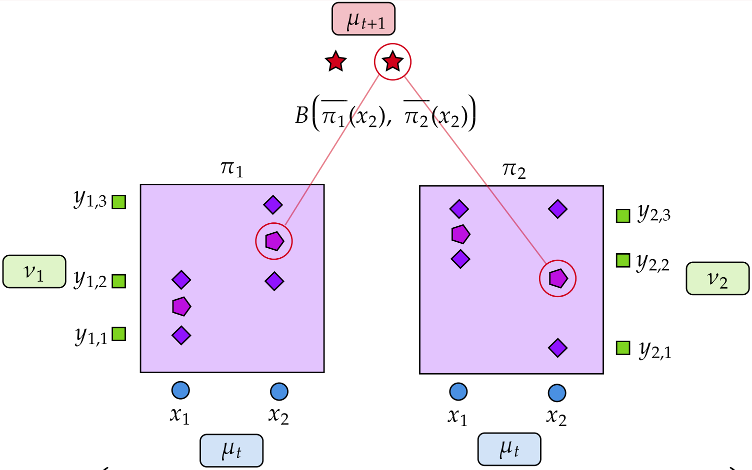


$$\mu = \sum_{i=1}^{n} a_i \delta_{x_i}$$

$$\nu_k = \sum_{j=1}^{n_k} b_{k,j} \delta_{y_{k,j}}$$

$$G(\mu) = \left\{ \sum_{j_1, \cdots, j_K} \left(\sum_{i=1}^{n} \frac{1}{a_i^{K-1}} \pi^{(1)}_{i,j_1} \times \cdots \times \pi^{(K)}_{i,j_K} \right) \delta\left(B(y_{1,j_1}, \cdots, y_{K,j_K})\right), \right.$$
$$\left. \pi^{(k)} \in \Pi^*_{c_k}(\mu, \nu_k) \right\}.$$

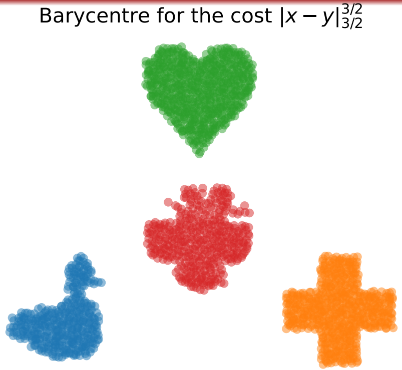
Discrete H (Generalises Cuturi & Doucet 2014 [4])

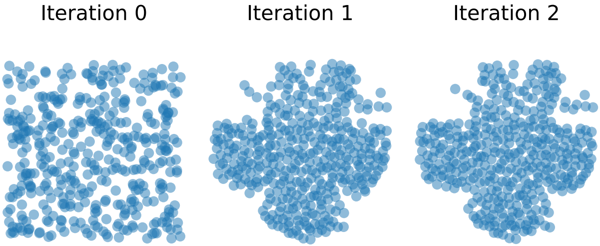


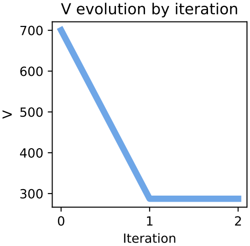
$$H(\mu) = \left\{ \sum_{i=1}^{n} a_i \delta\left(B\left(\overline{\pi}_1(x_i), \cdots, \overline{\pi}_K(x_i)\right)\right),\ \pi_k \in \Pi^*_{c_k}(\mu, \nu_k) \right\},$$

$$\overline{\pi}_k(x_i) = (1/a_i) \sum_{j=1}^{n_1} \pi^{(k)}_{i,j} y_{1,j}.$$

Illustration for $c(x,y) = \|x-y\|_{1.5}^{1.5}$

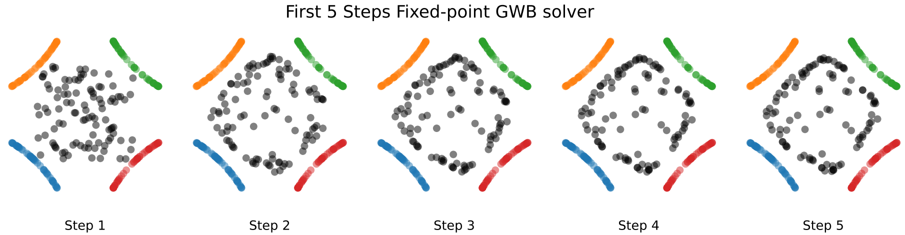


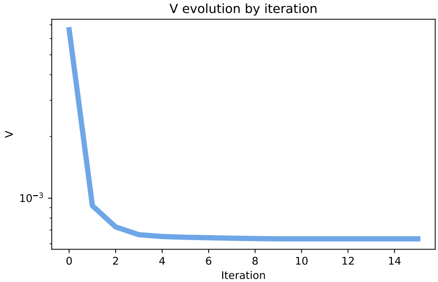




Non-linear Generalised Wasserstein Barycentre

$\mathrm{argmin}_\mu \sum_{k=1}^4 \frac{1}{4} \mathrm{W}_2^2(P_k\#\mu, \nu_k)$ where P_k is the projection onto circle k.





OT between GMMs

$$\mathrm{W}_2^2(\mathcal{N}(m_1, S_1), \mathcal{N}(m_2, S_2)) = \|m_1 - m_2\|_2^2 + \underbrace{\mathrm{Tr}\left(S_1 + S_2 - 2(S_1^{1/2} S_2 S_1^{1/2})^{1/2}\right)}_{d_{\mathrm{BW}}^2(S_1, S_2) :=}.$$

OT between GMMs

$$\mathrm{W}_2^2(\mathcal{N}(m_1,S_1),\mathcal{N}(m_2,S_2)) = \|m_1-m_2\|_2^2+\underbrace{\mathrm{Tr}\left(S_1+S_2-2(S_1^{1/2}S_2S_1^{1/2})^{1/2}\right)}_{d_{\mathrm{BW}}^2(S_1,S_2):=}.$$

Ground space: $(\mathcal{X},d)=(\mathcal{Y}_k,d_{\mathcal{Y}_k})=(\mathcal{N},\mathrm{W}_2)$ with ground cost $c=\mathrm{W}_2^2$.

$$\mu=\sum_{i=1}^n a_i\delta_{\mathcal{N}(m_i,S_i)},\ \nu=\sum_{j=1}^m b_j\delta_{\mathcal{N}(m_j',S_j')}\in\mathcal{P}(\mathcal{N});$$

$$\mathcal{T}_{\mathrm{W}_2^2}(\mu,\nu)=\min_{\pi\in\Pi(a,b)}\sum_{i,j}(\|m_i-m_j'\|_2^2+d_{\mathrm{BW}}^2(S_i,S_j'))\pi_{i,j}.$$

Ground Barycentre Between Gaussians

Gaussian barycentres (Agueh & Carlier 2011 [1]).

$$B(\mathcal{N}(m_1, S_1), \cdots, \mathcal{N}(m_K, S_K)) = \mathcal{N}(\overline{m}, \overline{S}),$$

$$\overline{m} := \sum_{k=1}^{K} \lambda_k m_k, \ \overline{S} := \underset{S \in S_d^{++}(\mathbb{R})}{\operatorname{argmin}} \sum_{k=1}^{K} \lambda_k d_{\mathrm{BW}}^2(S, S_k).$$

Fixed-point computation for $\overline{S}$:

$$G_{\mathcal{N}}(S) = S^{-1/2} \left(\sum_{k=1}^{K} \lambda_k (S^{1/2} S_k S^{1/2})^{1/2} \right)^2 S^{-1/2}.$$

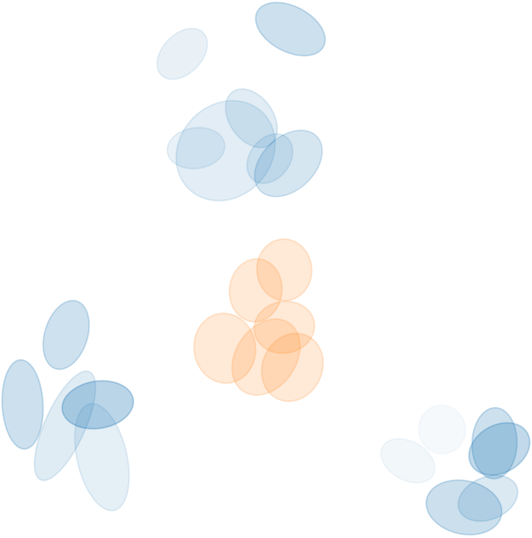
Riemannian gradient descent interpretation by Altschuler et al. 2021 [2].

GMM Barycentre

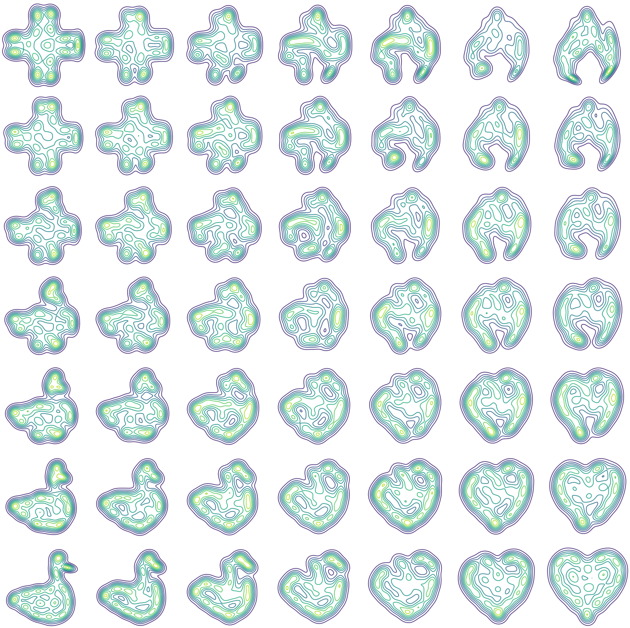
$$\mu = \sum_{i=1}^{n} a_i \delta_{\mathcal{N}(m_i, S_i)},$$

$$\nu_k = \sum_{j=1}^{n_k} b_k \delta_{\mathcal{N}(m_{k,j}, S_{k,j})},$$

$$V(\mu) = \sum_{k=1}^{K} \lambda_k \mathcal{T}_{\mathrm{W}_2^2}(\mu, \nu_k).$$



GMM Barycentre Example



- Talk based on *ET, Julie Delon and Nathaël Gozlan (2024):* Computing Barycentres of Measures for Generic Transport Costs. arXiv preprint 2501.04016.
- All code at `https://github.com/eloitanguy/ot_bar`
- Functions (soon) released on `https://pythonot.github.io/`
- Slides at `https://eloitanguy.github.io/publications/`

Thanks!

[1] Martial Agueh and Guillaume Carlier.
Barycenters in the Wasserstein space.
SIAM Journal on Mathematical Analysis, 43(2):904–924, 2011.

[2] Jason Altschuler, Sinho Chewi, Patrik R Gerber, and Austin Stromme.
Averaging on the bures-wasserstein manifold: dimension-free convergence of gradient descent.
Advances in Neural Information Processing Systems, 34:22132–22145, 2021.

[3] Pedro C Álvarez-Esteban, E Del Barrio, JA Cuesta-Albertos, and C Matrán.
A fixed-point approach to barycenters in Wasserstein space.
Journal of Mathematical Analysis and Applications, 441(2):744–762, 2016.

[4] Marco Cuturi and Arnaud Doucet.
Fast computation of Wasserstein barycenters.
In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 685–693, Bejing, China, 22–24 Jun 2014. PMLR.

[5] Julie Delon, Nathaël Gozlan, and Alexandre Saint-Dizier.
Generalized Wasserstein barycenters between probability measures living on different subspaces, 2021.