Computing Wasserstein Barycenter via Operator Splitting: the Method of Averaged Marginals

Daniel Mimouni[*,+]

Joint work with:
Paul Malisani*, Jiamin Zhu*, Welington de Oliveira[+]

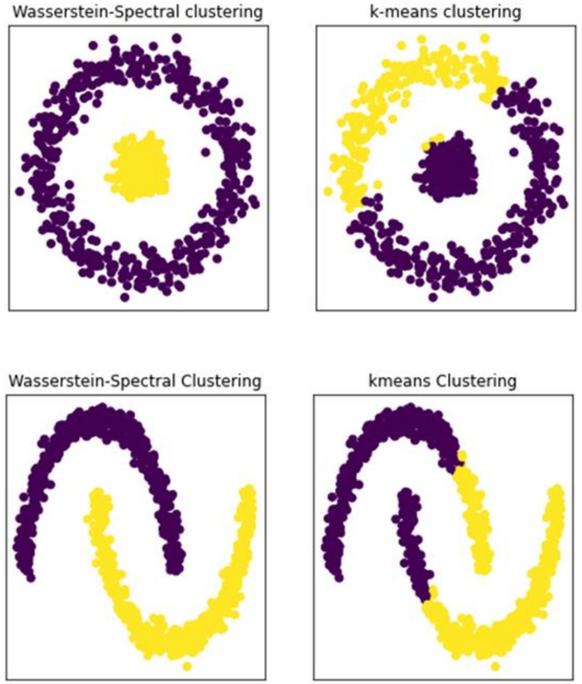
*IFP Energies nouvelles
[+]Mines Paris, PSL

Table of contents

I. Motivations

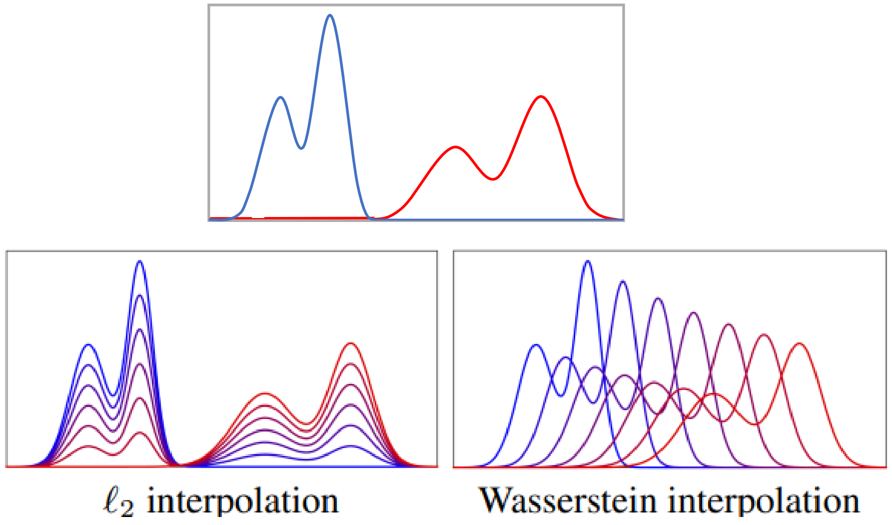
I.1. Applications

Clustering :



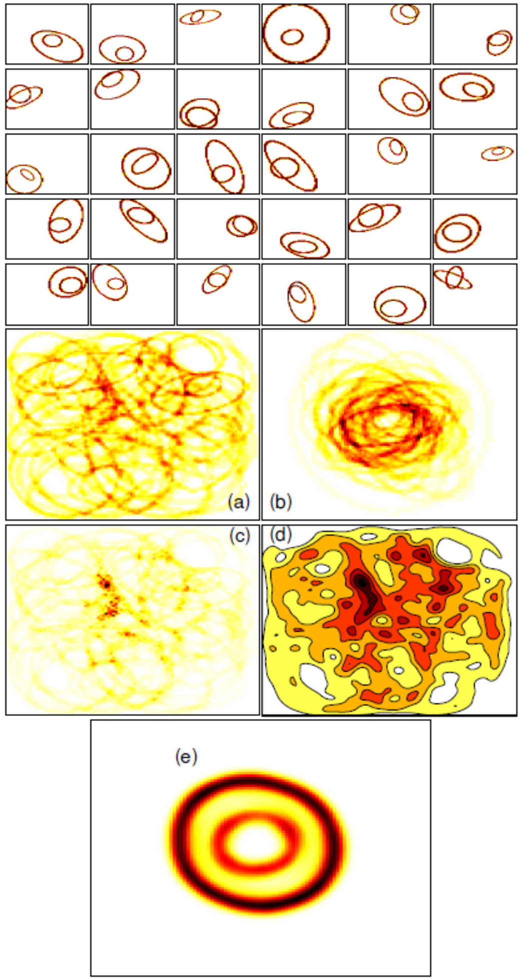
(a) Comparison of Wasserstein-Spectral clustering, spectral clustering, and k-means on Two-Circles dataset and Moons dataset [2]

Data preprocessing :



Comparison between Euclidean (left) and Optimal Transport (right) barycenters between two densities, one being a translated and scaled version of the other. Colors encode the progression of the interpolation. The Euclidean interpolation results in mixtures of the two initial densities, while Optimal Transport results in a progressive translation and scaling [3]

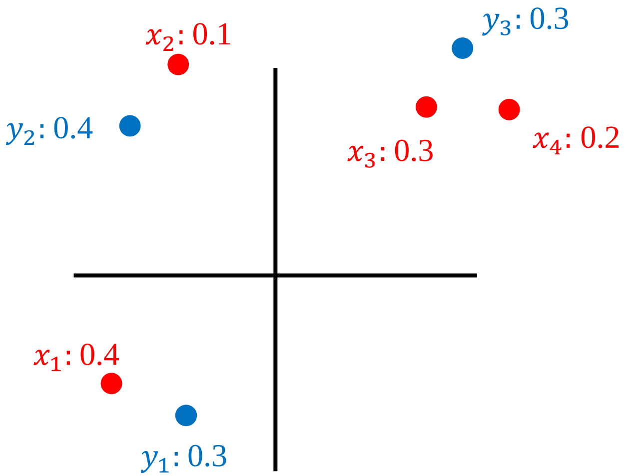
Visualization :



(Top) 30 artificial images of two nested random ellipses.
Mean measures using the (a) Euclidean distance (b) Euclidean after re-centering images (c) Jeffrey centroid (Nielsen, 2013) (d) RKHS distance (Gaussian kernel, σ= 0.002) (e) 2-Wasserstein distance. [1]

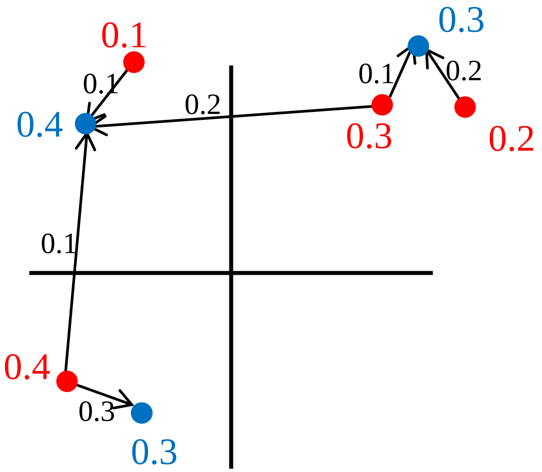
II. Background on Discrete Optimal Transport

II.1. Wasserstein distance



the ι-Wasserstein distance $W_\iota(\mu,\nu)$

$$\boxed{\mathtt{OT}_{\Xi,Z}(p,q) := \left\{ \begin{array}{lll} \min\limits_{\pi\geq 0} & \sum\limits_{r=1}^{R}\sum\limits_{s=1}^{S} \mathtt{d}(\xi_r,\zeta_s)^\iota \pi_{rs} & \\ \text{s.t.} & \sum_{r=1}^{R} \pi_{rs} = q_s, & s=1,\dots,S \\ & \sum_{s=1}^{S} \pi_{rs} = p_r, & r=1,\dots,R \end{array} \right.}$$



π	x_1 0.4	x_2 0.1	x_3 0.3	x_4 0.2
y_1 0.3	0.3	0	0	0
y_2 0.4	0.1	0.1	0.2	0
y_3 0.3	0	0	0.1	0.2

Recall:

$$\Delta_n(\tau) := \left\{ u \in \mathbb{R}^n_+ : \sum_{i=1}^n u_i = \tau \right\}$$

II. Background on Discrete Optimal Transport

II.1. Wasserstein distance

the ι-Wasserstein distance $W_\iota(\mu, \nu)$

finitely many R scenarios $\Xi := \{\xi_1, \dots, \xi_R\}$ for ξ and $S^{(m)}$ scenarios $Z^{(m)} := \{\zeta_1^{(m)}, \dots, \zeta_{S^{(m)}}^{(m)}\}$ for $\zeta^{(m)}$, $m = 1, \dots, M$, i.e., measures of the form

$$\mu = \textstyle\sum_{r=1}^R p_r \delta_{\xi_r} \quad \text{and} \quad \nu^{(m)} = \sum_{s=1}^{S^{(m)}} q_s^{(m)} \delta_{\zeta_s^{(m)}}, \quad m = 1, \dots, M,$$

with δ_u the Dirac unit mass on $u \in \Omega$, $p \in \Delta_R$, and $q^{(m)} \in \Delta_{S^{(m)}}$, $m = 1, \dots, M$.

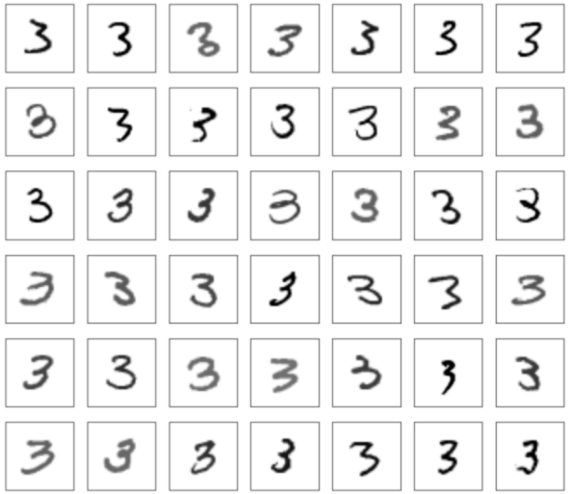
$$\mathtt{OT}_{\Xi,Z}(p,q) := \begin{cases} \min\limits_{\pi \ge 0} & \sum\limits_{r=1}^R \sum\limits_{s=1}^S \mathsf{d}(\xi_r, \zeta_s)^\iota \pi_{rs} \\ \text{s.t.} & \sum_{r=1}^R \pi_{rs} = q_s, \qquad s = 1, \dots, S \\ & \sum_{s=1}^S \pi_{rs} = p_r, \qquad r = 1, \dots, R \end{cases}$$

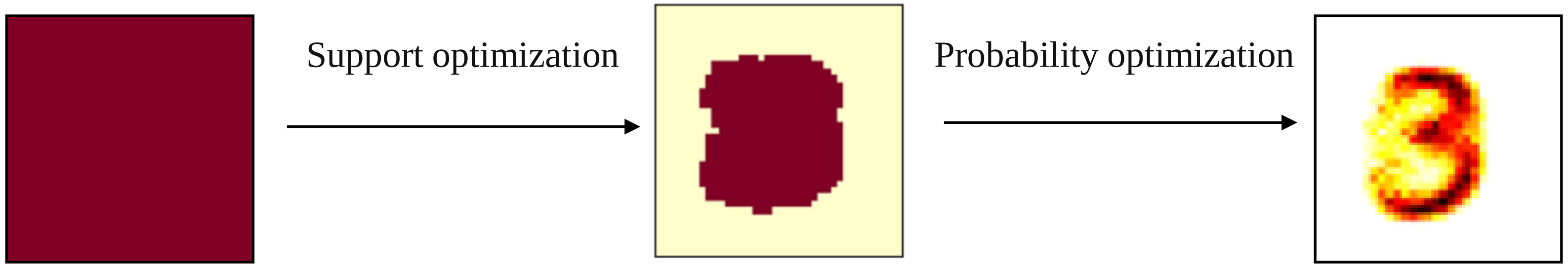
II. Background on Discrete Optimal Transport

II.2. Wasserstein Barycenter (WB)

General formulation of the WB problem:

$$\min_{\Xi, p \in \Delta_R} \sum_{m=1}^{M} \alpha_m \mathtt{OT}_{\Xi, Z^{(m)}}(p, q^{(m)})$$





General formulation of the WB problem:

$$\boxed{\min_{\Xi, p \in \Delta_R} \sum_{m=1}^{M} \alpha_m \mathtt{OT}_{\Xi, Z^{(m)}}(p, q^{(m)})}$$

Block coordinate optimization:

- Step 1: support optimization $\longrightarrow$ Straightforward solution exists if $\iota = 2$ (Euclidean norm) $\min_{\Xi} \sum_{m=1}^{M} \alpha_m \mathtt{OT}_{\Xi, Z^{(m)}}(p^k, q^{(m)})$
- Step 2: probability optimization $\longrightarrow$ $\min_{p \in \Delta_R} \sum_{m=1}^{M} \alpha_m \mathtt{OT}_{\Xi, Z^{(m)}}(p, q^{(m)})$

Repeat until convergence.

Recall:

$$\mathtt{OT}_{\Xi,Z}(p,q) := \begin{cases} \min_{\pi\geq 0} & \sum_{r=1}^{R}\sum_{s=1}^{S} \mathtt{d}(\xi_r,\zeta_s)^{\iota}\pi_{rs} \\ \text{s.t.} & \sum_{r=1}^{R}\pi_{rs} = q_s, \quad s=1,\ldots,S \\ & \sum_{s=1}^{S}\pi_{rs} = p_r, \quad r=1,\ldots,R. \end{cases}$$

WB problem written as a huge-scale LP:

$$\begin{cases} \min_{p,\pi} \sum_{r=1}^{R}\sum_{s=1}^{S^{(1)}} d_{rs}^{(1)}\pi_{rs}^{(1)} + \cdots + \sum_{r=1}^{R}\sum_{s=1}^{S^{(M)}} d_{rs}^{(M)}\pi_{rs}^{(M)} \\ \text{s.t.} \quad \sum_{r=1}^{R}\pi_{rs}^{(1)} = q_s^{(1)}, \quad s=1,\ldots,S^{(1)} \\ \qquad \ddots \qquad \vdots \\ \qquad \sum_{r=1}^{R}\pi_{rs}^{(M)} = q_s^{(M)}, \quad s=1,\ldots,S^{(M)} \\ \quad \sum_{s=1}^{S^{(1)}}\pi_{rs}^{(1)} = p_r, \quad r=1,\ldots,R \\ \qquad \ddots \qquad \vdots \\ \qquad \sum_{s=1}^{S^{(M)}}\pi_{rs}^{(M)} = p_r, \quad r=1,\ldots,R \\ \quad p\in\Delta_R, \pi^{(1)}\geq 0 \quad \cdots \quad \pi^{(M)}\geq 0 \end{cases}$$

Barycentric distance

Set of constraints on the $q^{(m)}$

$$\Pi^{(m)} := \left\{\pi^{(m)}\geq 0 : \sum_{r=1}^{R}\pi_{rs}^{(m)} = q_s^{(m)},\ s=1,\ldots,S^{(m)}\right\}$$

Set of constraints on p

$$\sum_{s=1}^{S(m)}\pi_{rs}^{(m)} = \sum_{s=1}^{S(m')}\pi_{rs}^{(m')}, \quad m,m'=1,\ldots,M$$

II.2. Wasserstein Barycenter (WB) - Previous works

How to compute ? $\xrightarrow{\text{Exact solution}}$ Solving the **huge LP** has limits : scalability, time consuming

Exact solution ?

Two alternatives:

approximation

i) Inexact methods via entropic regularization $\longrightarrow$
Iterative Bregman Projection (IBP) - 2013
Bregman ADMM (B-ADMM) - 2017
Iterative Swapping Algorithm (ISA) - 2020

Tends to an exact solution but low numerical efficiency

ii) Exact decomposition methods via solving smaller and simpler problems $\longrightarrow$
ADMM – 2010
ISA in certain configurations - 2020

II. Background on Discrete Optimal Transport

II.2. Wasserstein Barycenter (WB) - Reformulation

$$\mathcal{B} := \left\{ \pi = (\pi^{(1)}, \dots, \pi^{(S)}) \;\middle|\; \begin{array}{lcll} \sum_{s=1}^{S^{(1)}} \pi_{rs}^{(1)} & = & \sum_{s=1}^{S^{(2)}} \pi_{rs}^{(2)}, & r = 1, \dots, R \\ \sum_{s=1}^{S^{(2)}} \pi_{rs}^{(2)} & = & \sum_{s=1}^{S^{(3)}} \pi_{rs}^{(3)}, & r = 1, \dots, R \\ & \vdots & & \\ \sum_{s=1}^{S^{(M-1)}} \pi_{rs}^{(M-1)} & = & \sum_{s=1}^{S^{(M)}} \pi_{rs}^{(M)}, & r = 1, \dots, R \end{array} \right\}$$

$$\left\{ \begin{array}{llllll} \min\limits_{\pi} & \sum\limits_{r=1}^{R} \sum\limits_{s=1}^{S^{(1)}} d_{rs}^{(1)} \pi_{rs}^{(1)} & + \cdots + & \sum\limits_{r=1}^{R} \sum\limits_{s=1}^{S^{(M)}} d_{rs}^{(M)} \pi_{rs}^{(M)} & + & \mathbf{i}_{\mathcal{B}}(\pi) \\ \text{s.t.} & \sum_{r=1}^{R} \pi_{rs}^{(1)} & & & = q_s^{(1)}, & s = 1, \dots, S^{(1)} \\ & & \ddots & & \vdots & \\ & & & \sum_{r=1}^{R} \pi_{rs}^{(M)} & = q_s^{(M)}, & s = 1, \dots, S^{(M)} \\ & \pi^{(1)} \geq 0 & \cdots & \pi^{(M)} \geq 0 & & \end{array} \right.$$

III. The Method of Averaged Marginals (MAM)

III.1. Reformulation of the LP

Reformulation of the problem:

1. $$f^{(m)}(\pi^{(m)}) := \sum_{r=1}^{R} \sum_{s=1}^{S^{(m)}} d_{rs}^{(m)} \pi_{rs}^{(m)} + \mathbf{i}_{\Pi^{(m)}}(\pi^{(m)})$$

2. $$f(\pi) := \sum_{m=1}^{M} f^{(m)}(\pi^{(m)}) \quad \text{and} \quad g(x) := \mathbf{i}_{\mathcal{B}}(\pi)$$

3. $$\min_{\pi} \; f(\pi) + g(\pi)$$

$$\boxed{\text{find} \quad \pi \quad \text{such that} \quad 0 \in \partial f(\pi) + \partial g(\pi)}$$

New problem : finding the zero of the sum of two maximal monotone operators

→ Several methods exist

→ Douglas-Rachford operator splitting is the most popular one (see ADMM or progressive hedging methods)

III. The Method of Averaged Marginals (MAM)

III.2. Douglas-Rachford (DR) theory

Douglas-Rachford steps:

given initial point $\theta^0 = (\theta^{(1),0}, \dots, \theta^{(M),0})$ and prox-parameter $\rho > 0$:

$$\begin{cases} \pi^{k+1} &= \mathrm{prox}_{g/\rho}(\theta^k) \\ \hat{\pi}^{k+1} &= \mathrm{prox}_{f/\rho}(2\pi^{k+1} - \theta^k) \\ \theta^{k+1} &= \theta^k + \hat{\pi}^{k+1} - \pi^{k+1} \end{cases}$$

Projection onto $\mathcal{B}$ is explicit

$$\pi^{k+1} = \mathtt{Proj}_{\mathcal{B}}(\theta^k)$$

Projections onto the simplex

$$\Delta_n(\tau) := \left\{ u \in \mathbb{R}^n_+ : \sum_{i=1}^n u_i = \tau \right\}$$

$$\begin{pmatrix} \hat{\pi}^{(m)}_{1s} \\ \vdots \\ \hat{\pi}^{(m)}_{Rs} \end{pmatrix} = \mathtt{Proj}_{\Delta_R(q_s^{(m)})} \begin{pmatrix} y_{1s} - \frac{1}{\rho} d^{(m)}_{1s} \\ \vdots \\ y_{Rs} - \frac{1}{\rho} d^{(m)}_{Rs} \end{pmatrix}, \quad s = 1, \dots, S^{(m)}$$

III. The Method of Averaged Marginals (MAM)

III.3. Algorithm - Main steps

Step 1: Given a multi-transportation plan θ^k

- Marginals $p^{(m),k} = \theta^{(m),k}\mathbb{1}, m = 1, \ldots, M$
- p^k is a weighted average of $\{p^{(1),k}, \ldots, p^{(M),k}\}$

Step 2: Given θ^k, p^k and distance matrices

- Compute a multi-transportation plan π^k by performing $\sum_{m=1}^{M} S^{(m)}$ independant projections onto the simplex Δ_R

Step 3: Givent θ^k, p^k and π^k

- Compute θ^{k+1} by a straightforward operation
- Set $k = k + 1$ and repeat

III.3. Algorithm - Feelings and philosophy

Algorithm 5.1 METHOD OF AVERAGED MARGINALS - MAM

▷ Step 0: input

Given $\rho > 0$, the distance matrix and initial point $d, \theta^0 \in \mathbb{R}^{R \times \sum_{m=1}^{M} S^{(m)}}$, and $a \in \Delta_M$ as in (5.3a), set $k \leftarrow 0$ and $p_r^{(m)} \leftarrow \sum_{s=1}^{S^{(m)}} \theta_{rs}^{(m),0}$, $r = 1, \dots, R$, $m = 1, \dots, M$

Set $\gamma \leftarrow \infty$ if $q^{(m)} \in \mathbb{R}_+^{S^{(m)}}$, $m = 1, \dots, M$, are balanced; otherwise, choose $\gamma \in (0, \infty)$

while not converged **do**

▷ Step 1: average the marginals

  Compute $p^k \leftarrow \sum_{m=1}^{M} a_m p^{(m)}$

Unbalanced formulation

  Set $t^k = 1$ if $\rho \sqrt{\sum_{m=1}^{M} \frac{\|p^k - p^{(m)}\|^2}{S^{(m)}}} \leq \gamma$; otherwise, $t^k \leftarrow \gamma / \left(\rho \sqrt{\sum_{m=1}^{M} \frac{\|p^k - p^{(m)}\|^2}{S^{(m)}}}\right)$

  Choose an index set $\emptyset \neq \mathcal{M}^k \subseteq \{1, \dots, M\}$

Can be executed in parallel or randomized

  for $m \in \mathcal{M}^k$ **do**

▷ Step 2: update the m^{th} plan

    for $s = 1, \dots, S^{(m)}$ **do**

      Define $w_r \leftarrow \theta_{rs}^{(m),k} + 2\, t^k \frac{p_r^k - p_r^{(m)}}{S^{(m)}} - \frac{1}{\rho} d_{rs}^{(m)}$, $r = 1, \dots, R$

      Compute $(\hat{\pi}_{1s}^{(m)}, \dots, \hat{\pi}_{Rs}^{(m)}) \leftarrow \mathtt{Proj}_{\Delta_R(q_s^{(m)})}(w)$

Projection onto the simplex performed exactly by using efficient methods

      Update $\theta_{rs}^{(m),k+1} \leftarrow \hat{\pi}_{rs}^{(m)} - t^k \frac{p_r^k - p_r^{(m)}}{S^{(m)}}$, $r = 1, \dots, R$

Projection onto $\mathcal{B}$

    end for

▷ Step 3: update the m^{th} marginal

    Update $p_r^{(m)} \leftarrow \sum_{s=1}^{S^{(m)}} \theta_{rs}^{(m),k+1}$, $r = 1, \dots, R$

  end for

end while

Return $\bar{p} \leftarrow p^k$

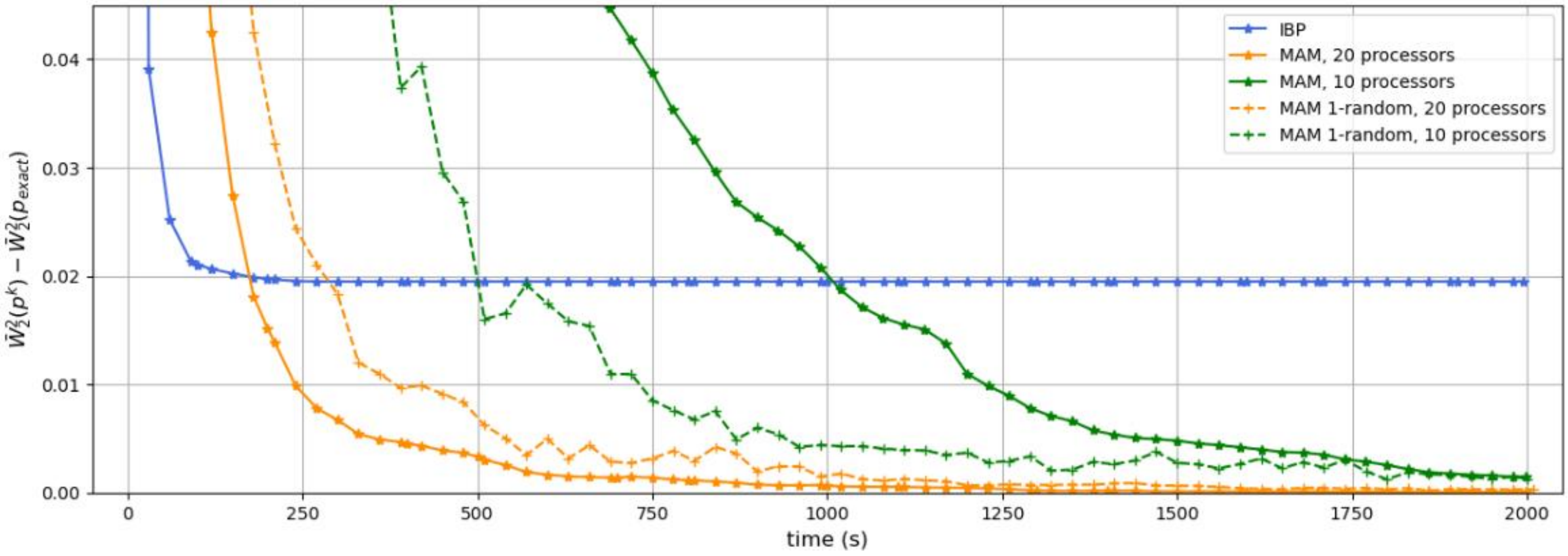
IV. Applications

IV.1. Qualitative comparison

MAM	IBP
Exact algorithm	Iterative Bregman Projection • state-of-the-art algorithm for WB • based on an entropic regularization of the problem thus computes inexact WB

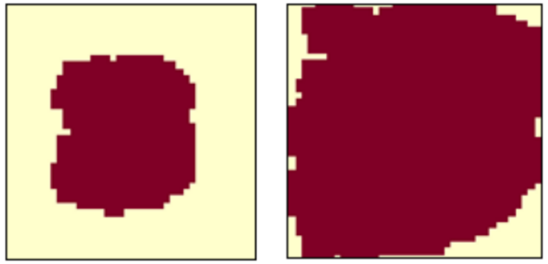
IV. Applications

IV.2. Quantitative comparison

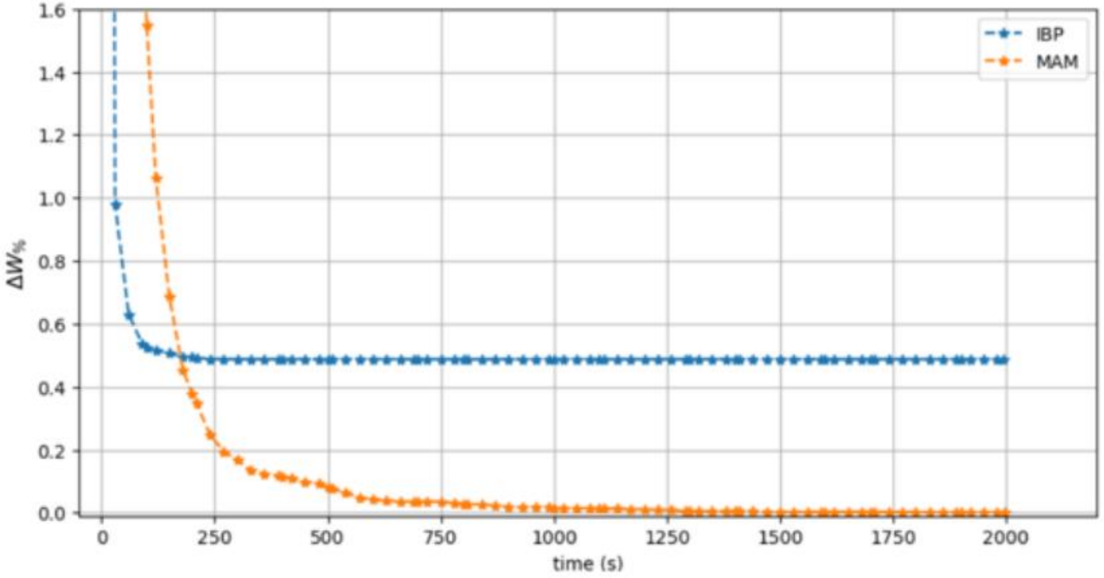


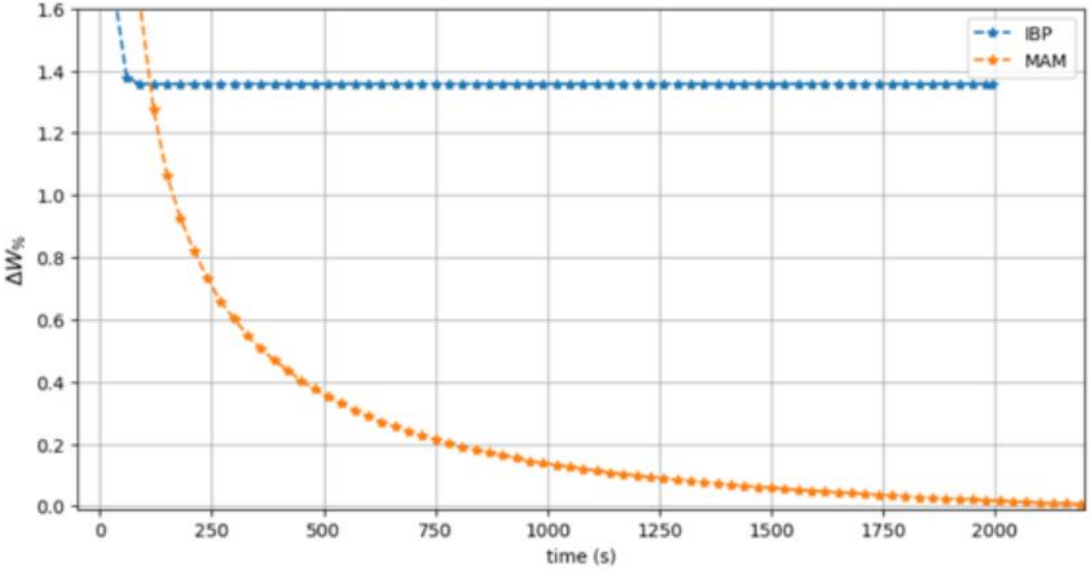
IV. Applications

IV.3. Influence of the support



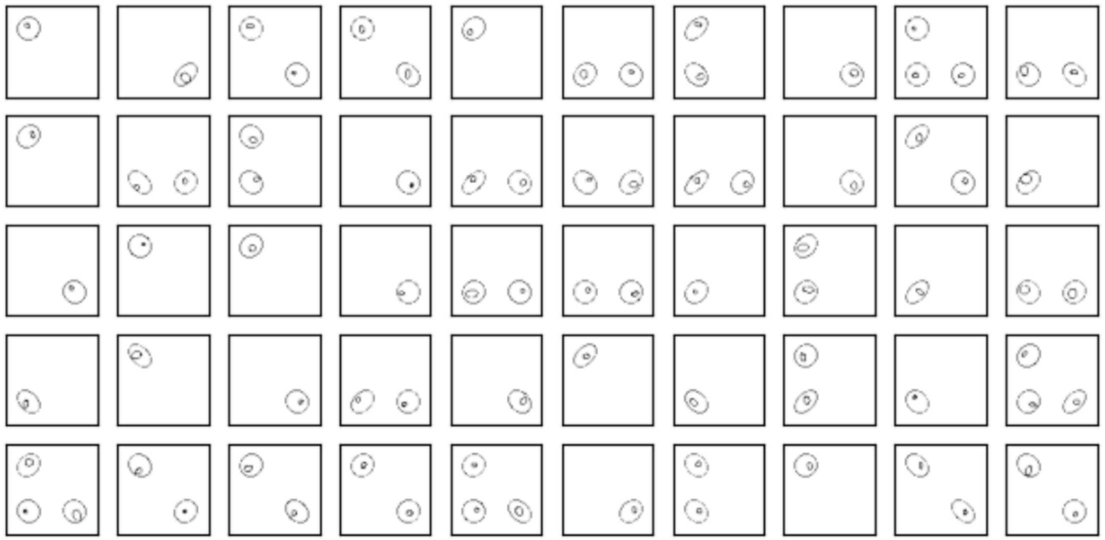
Union of the dataset support





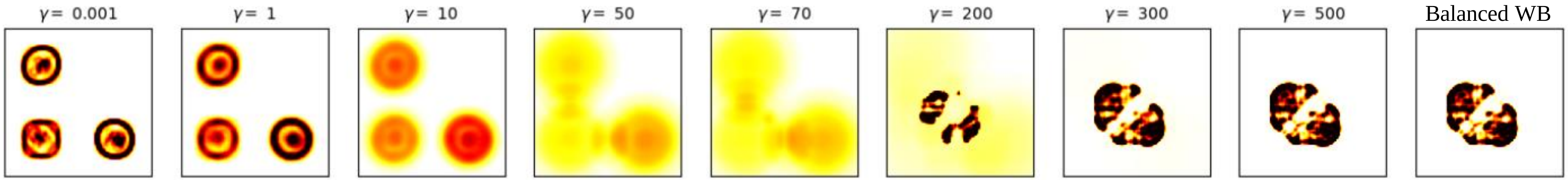
IV. Applications

IV.4. Unbalanced Wasserstein Barycenter

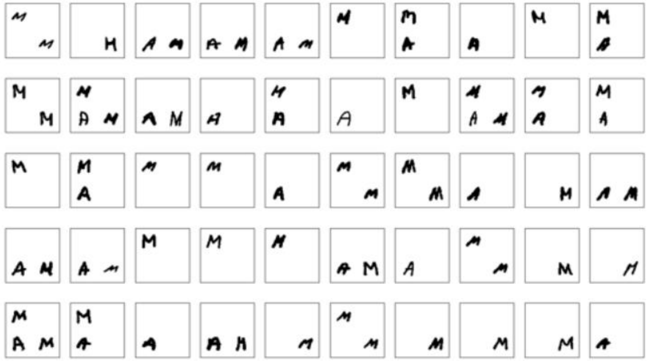


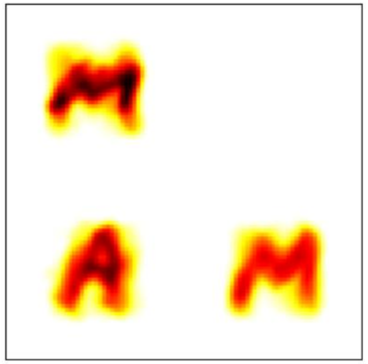
Dataset composed by 50 pictures with nested ellipses randomly positionned in the top left, bottom right and left corners :

The standard (balanced) WB is not always the best tool for summarizing:



Conclusion and future works





<u>What have been introduced?</u>

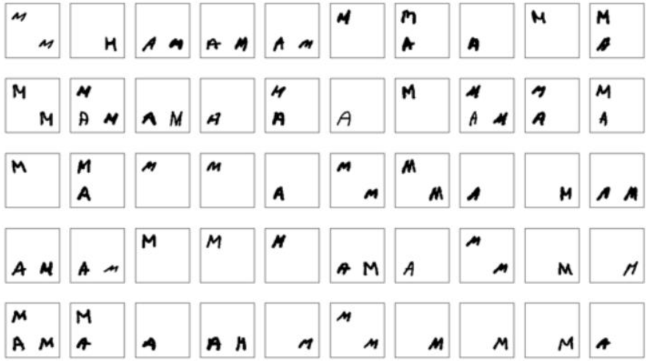
A novel approach for computing Wasserstein barycenters of discrete measures, that:

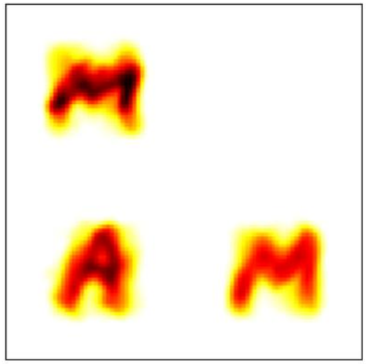
→Asymptotically exact
→Embarrassingly parallelizable and can be used on a randomized manner (almost surely convergence)
→Can tackle both the balanced and unbalanced case!

<u>What can be done now?</u>

→Adapt MAM to tackle scenario trees reduction problem in stochastic optimization
→Real life examples

References





[1] Cuturi, M., & Doucet, A. (2014, June). Fast computation of Wasserstein barycenters. In *International conference on machine learning* (pp. 685-693). PMLR

[2] Bonneel, N., Peyré, G., & Cuturi, M. (2016). Wasserstein barycentric coordinates: histogram regression using optimal transport. *ACM Trans. Graph.*, *35*(4), 71-1.

[3] El Hamri, M., Bennani, Y., & Falih, I. (2022). Hierarchical optimal transport for unsupervised domain adaptation. *Machine Learning*, *111*(11), 4159-4182.

[4] Mimouni, D., Malisani, P., Zhu, J., & de Oliveira, W. (2023). Computing Wasserstein Barycenter via operator splitting: the method of averaged marginals. *arXiv preprint arXiv:2309.05315*.