A splitting method for computing Wasserstein Barycenters

Daniel Mimouni[1,2], Paul Malisani[2], Jiamin Zhu[2], Welington de Oliveira[1]

1 **Centre de Mathématiques Appliquées - CMA, Mines Paris PSL**

2 **IFP Energies nouvelles**

EUROPT 2024, Lund, Sweden
June 2024

Let ξ and ζ be two random vectors having probability measures μ and ν, that is,

$$\xi \sim \mu \quad \text{and} \quad \zeta \sim \nu$$

We focus on discrete measures based on

finitely many R atoms $\texttt{supp}(\mu) := \{\xi_1, \ldots, \xi_R\}$

finitely many S atoms $\texttt{supp}(\nu) := \{\zeta_1, \ldots, \zeta_S\}$,

i.e., the supports are finite and thus the measures are given by

$$\mu = \sum_{r=1}^{R} p_r \delta_{\xi_r} \quad \text{and} \quad \nu = \sum_{s=1}^{S} q_s \delta_{\zeta_s}$$

Quadratic Wasserstein distance - discrete setting

The 2-Wassestein distance between two discrete probability measures μ and ν is:

$$W_2(\mu, \nu) := \left(\min_{\pi \in U(\mu,\nu)} \sum_{r=1}^{R} \sum_{s=1}^{S} \|\xi_r - \zeta_s\|^2 \pi_{rs} \right)^{1/2}$$

with

$$U(\mu, \nu) := \left\{ \pi \geq 0 \;\middle|\; \begin{array}{ll} \sum_{r=1}^{R} \pi_{rs} = q_s, & s = 1, \ldots, S \\ \sum_{s=1}^{S} \pi_{rs} = p_r, & r = 1, \ldots, R \end{array} \right\}$$

- Let $\alpha \in \Re_+^M$ be a vector of weights: $\sum_{m=1}^M \alpha_m = 1$

Discrete Wassertein Barycenter - WB

A Wassertein barycenter of a set of M discrete probability measures ν^m, $m = 1, \dots, M$, is a solution to the following optimization problem

$$\min_{\mu \in P(\Omega)} \sum_{m=1}^M \alpha_m W_2^2(\mu, \nu^m)$$

- A WB of a set of M discrete probability measures is a discrete measure itself, supported on a subset of the of the finite set

$$\mathtt{supp}(\mu) := \left\{ \sum_{m=1}^M \alpha_m \zeta_s^m \, : \, \zeta_s^m \in \mathtt{supp}(\nu^m), \; m = 1, \dots, M \right\}$$

- This set has at most $\Pi_{m=1}^M S^m$ points, with $S^m = |\mathtt{supp}(\nu^m)|$
- If we enumerate all R points $\xi \in \mathtt{supp}(\mu)$, we get an LP formulation for the discrete WB

CMA Centre de Mathématiques Appliquées

$$\text{supp}(\mu) = \left\{ \sum_{m=1}^{M} \alpha_m \zeta_s^m \; : \; \zeta_s^m \in \text{supp}(\nu^m), \; m = 1, \ldots, M \right\}$$

Let $R = |\text{supp}(\mu)|$, $\xi \in \text{supp}(\mu)$ and $S^m = |\text{supp}(\nu^m)|$

Discrete Wassertein Barycenter - WB

A Wassertein barycenter of a set of M discrete probability measures ν^m, $m = 1, \ldots, M$, is a solution to the LP

$$\begin{cases} \min_{p,\pi \geq 0} & \sum_{m=1}^{M} \alpha_m \sum_{r=1}^{R} \sum_{s=1}^{S^m} \|\xi_r - \zeta_s^m\|^2 \pi_{rs}^m \\ \text{s.t.} & \sum_{r=1}^{R} \pi_{rs}^m = q_s^m, \quad s = 1, \ldots, S^m, \; m = 1, \ldots, M \\ & \sum_{s=1}^{S^m} \pi_{rs}^m = p_r, \quad r = 1, \ldots, R, \; m = 1, \ldots, M \end{cases}$$

- This LP scales exponentially in the number M of measures [[1]]

[1] S. Borgwardt. Operational Research (2022)

Inexact approaches

- Mostly based on reformulations via an entropic regularization: several papers by M. Cuturi, G. Peyré, G. Carlier and others
- Block-coordinate approach: fix the support and optimize the probability, then fix the probability and optimize the support [[2], [3], [4]]
- Other approaches [[5],[6],[7]]

Exact methods

- Methods for computing exact WBs are based on linear programming techniques [[8],[9]]

[2] M. Cuturi, A. Doucet. JMLR, 2014
[3] J. Ye, J. Li. IEEE ICP (214)
[4] J. Ye et al. IEEE Transactions on Signal Processing (2017)
[5] G. Puccetti, L. Ruschendorf, S. Vanduffe. JMVA (2020)
[6] S. Borgwardt. Operational Research (2022)
[7] J. von Lindheim. COAP (2023)
[8] S. Borgwardt, S. Patterson (2020). INFORM J. Optimization
[9] J. Altschuler, E. Adsera. JMLR (2021)

We provide an embarrassingly parallelizable algorithm based on the Douglas-Rachford splitting scheme to compute a solution to LPs of the form

$$\left\{\begin{array}{lll} \min\limits_{p,\pi\geq 0} & \sum\limits_{m=1}^{M}\sum\limits_{r=1}^{R}\sum\limits_{s=1}^{S^m} d_{rs}^m \pi_{rs}^m & \\ \text{s.t.} & \sum\limits_{r=1}^{R} \pi_{rs}^m = q_s^m, & s=1,\ldots,S^m,\ m=1,\ldots,M \\ & \sum\limits_{s=1}^{S^m} \pi_{rs}^m = p_r, & r=1,\ldots,R,\ m=1,\ldots,M \end{array}\right.$$

with given $d^m \in \Re^{R\times S^m}$ (e.g. $d_{rs}^m := \alpha_m \|\xi_r - \zeta_s^m\|^2$)

Observe that we can drop the vector p (wlog)

$$
\left\{
\begin{array}{ll}
\min\limits_{\pi \geq 0} & \sum_{m=1}^{M} \sum_{r=1}^{R} \sum_{s=1}^{S^m} d_{rs}^m \pi_{rs}^m \\
\text{s.t.} & \sum_{r=1}^{R} \pi_{rs}^1 = q_s^1, \quad s = 1, \ldots, S^1 \\
& \vdots \\
& \sum_{r=1}^{R} \pi_{rs}^M = q_s^M, \quad s = 1, \ldots, S^M \\
& \sum_{s=1}^{S^1} \pi_{rs}^1 = \cdots = \sum_{s=1}^{S^M} \pi_{rs}^M, \quad r = 1, \ldots, R
\end{array}
\right.
\equiv
\left\{
\begin{array}{ll}
\min\limits_{\pi} & \sum_{m=1}^{M} \langle d^m, \pi^m \rangle \\
\text{s.t.} & \pi^1 \in \Pi^m \\
& \vdots \\
& \pi^M \in \Pi^M \\
& \pi \in \mathcal{B}
\end{array}
\right.
$$

This LP can be solved by the Douglas-Rachford splitting (DR) method

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

DR ALGORITHM

$$
\left\{
\begin{array}{lll}
\pi^{k+1} & = & \texttt{Proj}_{\mathcal{B}}(\theta^k) \\
\hat{\pi}^{k+1} & = & \arg \min\limits_{\substack{\pi^m \in \Pi^m \\ m=1,\ldots,M}} \sum_{m=1}^{M} \langle d^m, \pi^m \rangle + \frac{\rho}{2} \| \pi - (2\pi^{k+1} - \theta^k) \|^2 \\
\theta^{k+1} & = & \theta^k + \hat{\pi}^{k+1} - \pi^{k+1}
\end{array}
\right.
$$

$\{\pi^k\}$ converges to a solution to the above LP [10]

[10] H.H. Bauschke, P.L. Combettes. Chapter 25. (2017)

Centre de Mathématiques Appliquées

First DR's step

Projecting onto the subspace of balanced plans

Given $\theta \in \Re^{R \times \sum_{m=1}^{M} S^m}$, let

- $a_m := \frac{\frac{1}{S^m}}{\sum_{j=1}^{M} \frac{1}{S(j)}}$ be weights
- $p^m := \sum_{s=1}^{S^m} \theta_{rs}^m$ the m^{th} marginal
- $p := \sum_{m=1}^{M} a_m p^m$ the average of marginals

Proposition

The projection $\pi = \texttt{Proj}_{\mathcal{B}}(\theta)$ *has the explicit form:*

$$\pi_{rs}^m = \theta_{rs}^m + \frac{(p_r - p_r^m)}{S^m}, \quad s = 1, \ldots, S^m, \; r = 1, \ldots, R, \; m = 1, \ldots, M$$

This projection can be computed in parallel

Proposition

The proximal mapping

$$\hat{\pi} = \arg \min_{\substack{\pi^m \in \Pi^m \\ m=1,\dots,M}} \sum_{m=1}^{M} \langle d^m, \pi^m \rangle + \frac{\rho}{2} \|\pi - y\|^2$$

can be computed exactly, in parallel along the columns of each transport plan y^m, as follows: for all $m \in \{1, \dots, M\}$,

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

Here, $\Delta_R(\tau) = \left\{ x \in \Re^R_+ : \sum_{r=1}^R x_r = \tau \right\}$

Every projection onto $\Delta_R(q^m_s)$ can be carried out (in parallel) efficiently and exactly [11]

[11] L. Condat. Math.Prog. (2016)

Centre de Mathématiques Appliquées

The Method of Averaged Marginals - MAM

Easy-to-implement and memory efficient algorithm to compute WBs

Algorithm

Input: initial plan $\pi = (\pi^1, \ldots, \pi^m)$ and parameter $\rho > 0$

Define $a_m \leftarrow (\frac{1}{S^m})/(\sum_{j=1}^M \frac{1}{S^j})$ and set $p^m \leftarrow \sum_{s=1}^{S^m} \pi^m_{rs}$, $m = 1, \ldots, M$

while not converged **do**

  $p \leftarrow \sum_{m=1}^M a_m p^m$ ▷ Average the marginals

  for $m = 1, \ldots, M$ **do**

    for $s = 1, \ldots, S^m$ **do**

      $\pi^m_{:s} \leftarrow \mathtt{Proj}_{\Delta(q^m_s)}\left(\pi^m_{:s} + 2\frac{p - p^m}{S^m} - \frac{1}{\rho} d^m_{:s}\right) - \frac{p - p^m}{S^m}$

    end for

    $p^m \leftarrow \sum_{s=1}^{S^m} \theta^m_{rs}$ ▷ Update the m^{th} marginal

  end for

end while

This algorithm is embarrassingly parallelizable and can run in a randomized manner...

Algorithm (randomized)

Input: initial plan $\pi = (\pi^1, \dots, \pi^m)$ and parameter $\rho > 0$

Define $a_m \leftarrow (\frac{1}{S^m})/(\sum_{j=1}^{M} \frac{1}{S^j})$ and set $p^m \leftarrow \sum_{s=1}^{S^m} \pi^m_{rs}$, $m = 1, \dots, M$

while not converged **do**

  $p \leftarrow \sum_{m=1}^{M} a_m p^m$ ▷ Average the marginals

  Draw randomly $m \in \{1, 2, \dots, M\}$ with probability $\alpha_m > 0$

  for $s = 1, \dots, S^m$ **do**

    $\pi^m_{:s} \leftarrow \texttt{Proj}_{\Delta(q^m_s)}\left(\pi^m_{:s} + 2\frac{p - p^m}{S^m} - \frac{1}{\rho} d^m_{:s}\right) - \frac{p - p^m}{S^m}$

  end for

  $p^m \leftarrow \sum_{s=1}^{S^m} \theta^m_{rs}$ ▷ Update the m^{th} marginal

end while

Centre de Mathématiques Appliquées

Suppose the probability vector p is constrained to a closed convex set $X \subset \Re^R$:

$$\left\{\begin{array}{lll} \min\limits_{p,\pi\geq 0} & \sum\limits_{m=1}^{M} \langle d^m, \pi^m \rangle & \\ \text{s.t.} & \sum\limits_{r=1}^{R} \pi^m_{rs} = q^m_s, & s = 1,\ldots,S^m,\ m = 1,\ldots,M \\ & \sum\limits_{s=1}^{S^m} \pi^m_{rs} = p_r, & r = 1,\ldots,R,\ m = 1,\ldots,M \\ & p \in X & \end{array}\right.$$

How MAM can be modified to compute constrained WBs?

Centre de Mathématiques Appliquées

The Method of Averaged Marginals - MAM

Constrained setting

Algorithm

Input: initial plan $\pi = (\pi^1, \ldots, \pi^m)$ and parameter $\rho > 0$

Define $a_m \leftarrow (\frac{1}{S^m})/(\sum_{j=1}^{M} \frac{1}{S^j})$ and set $p^m \leftarrow \sum_{s=1}^{S^m} \pi^m_{rs}$, $m = 1, \ldots, M$

while not converged **do**

  $p \leftarrow \texttt{Proj}_X \left(\sum_{m=1}^{M} a_m p^m \right)$ ▷ Average the marginals

  for $m = 1, \ldots, M$ **do**

    for $s = 1, \ldots, S^m$ **do**

      $\pi^m_{:s} \leftarrow \texttt{Proj}_{\Delta(q^m_s)} \left(\pi^m_{:s} + 2\frac{p - p^m}{S^m} - \frac{1}{\rho} d^m_{:s} \right) - \frac{p - p^m}{S^m}$

    end for

    $p^m \leftarrow \sum_{s=1}^{S^m} \theta^m_{rs}$ ▷ Update the m^{th} marginal

  end for

end while

Unbalanced Wasserstein barycenters

Linear subspace of balanced plans:

$$\mathcal{B} = \left\{ \pi : \sum_{s=1}^{S^1} \pi^1_{rs} = \cdots = \sum_{s=1}^{S^M} \pi^M_{rs}, \quad r = 1, \ldots, R \right\}$$

ν^m, $m = 1, \ldots, M$, have equal masses

Balanced WB

$$\begin{cases} \min_{\pi \in \mathcal{B}} & \sum_{m=1}^{M} \langle d^m, \pi^m \rangle \\ \text{s.t.} & \pi^1 \in \Pi^m \\ & \vdots \\ & \pi^M \in \Pi^M \end{cases}$$

ν^m, $m = 1, \ldots, M$, have different masses

Unbalanced WB ($\gamma > 0$)

$$\begin{cases} \min_{\pi} & \sum_{m=1}^{M} \langle d^m, \pi^m \rangle + \gamma\, \mathtt{dist}_{\mathcal{B}}(\pi) \\ \text{s.t.} & \pi^1 \in \Pi^m \\ & \vdots \\ & \pi^M \in \Pi^M \end{cases}$$

How MAM can be modified to compute unbalanced WBs?

CMA
Centre de Mathématiques Appliquées

The Method of Averaged Marginals - MAM

Algorithm

Input: initial plan $\pi = (\pi^1, \dots, \pi^m)$ and parameters $\rho, \gamma > 0$

Define $a_m \leftarrow (\frac{1}{S^m})/(\sum_{j=1}^{M} \frac{1}{S^j})$ and set $p^m \leftarrow \sum_{s=1}^{S^m} \pi^m_{rs}$, $m = 1, \dots, M$

while not converged **do**

  $p \leftarrow \sum_{m=1}^{M} a_m p^m$ ▷ Average the marginals

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

  for $m = 1, \dots, M$ **do**

    for $s = 1, \dots, S^m$ **do**

      $\pi^m_{:s} \leftarrow \mathtt{Proj}_{\Delta(q^m_s)} \left(\pi^m_{:s} + 2t \frac{p - p^m}{S^m} - \frac{1}{\rho} d^m_{:s} \right) - t \frac{p - p^m}{S^m}$

    end for

    $p^m \leftarrow \sum_{s=1}^{S^m} \theta^m_{rs}$ ▷ Update the m^{th} marginal

  end for

end while

Set $\gamma = \infty$ to compute balanced WB (if the measures are balanced)
Otherwise, choose $\gamma \in (0, \infty)$ to compute unbalanced WB

Theorem (MAM's convergence analysis)

- *(Deterministic.) MAM asymptotically computes a balanced (unbalanced) Wasserstein barycenter should the measures be balanced (unbalanced)*
- *(Randomized.) MAM computes almost surely a balanced (unbalanced) Wasserstein barycenter should the measures be balanced (unbalanced)*

Centre de Mathématiques Appliquées

Numerical experiments: fixed support $R = 1\,600$

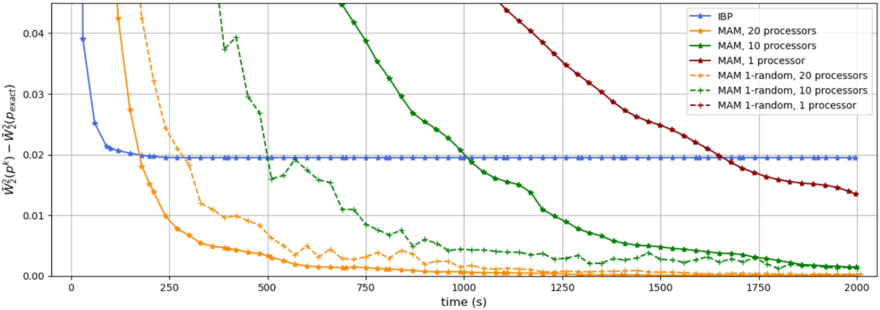
We benchmark MAM, randomized MAM, and IBP (Iterative Bregman Projection of [12]) on the MNIST database with $M = 60$ images of 40×40 pixels. LP's dimension: 153 601 600 variables and 192 000 constraints

[12] [J.-D. Benamou et al. SIAM Journal on Scientific Computing. (2015)]

MAM versus IBP

Wasserstein barycentric distance:

$$\bar{W}_2^2(\mu) := \sum_{m=1}^{M} \alpha_m W_2^2(\mu, \nu^m)$$



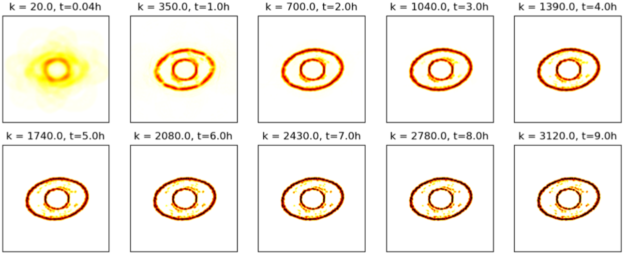
Evolution with respect to time of the difference between the Wasserstein barycenter distance of an approximation, $\bar{W}_2^2(p^k)$, and the Wasserstein barycentric distance of the exact solution $\bar{W}_2^2(p_{exact})$ given by the LP. The time step between two points is 30 seconds

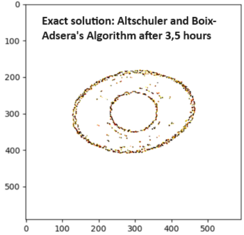
Numerical experiments: free support $R = 34\,9281$

The dataset we use is the one from [13]: $M = 10$ images of 60×60 pixels

LP's dimension: $1.2574 \cdot 10^{10}$ variables and $3.5288 \cdot 10^{6}$ constraints

We compare with the dedicated solver of Altschuler and Boix-Adsera, available at [14]

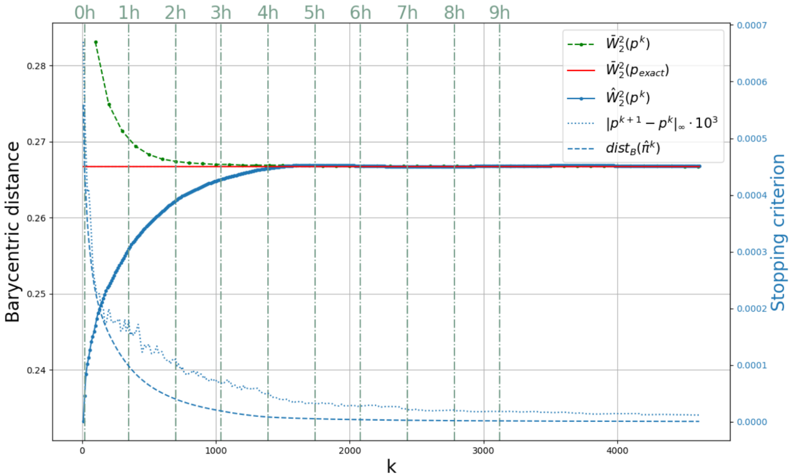




However, MAM can solve larger problems than the method Altschuler and Boix-Adsera

[13] J. M. Altschuler and E. Boix-Adsera. JMLR (2021)

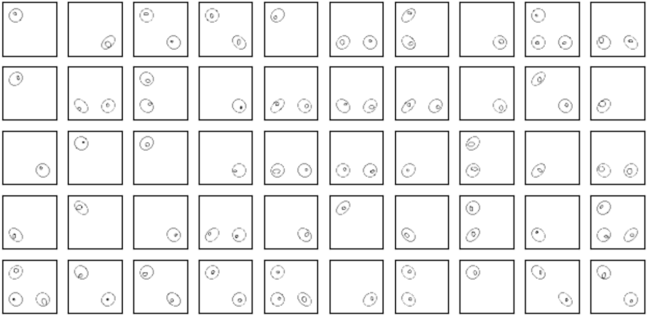
[14] https://github.com/eboix/high_precision_barycenters



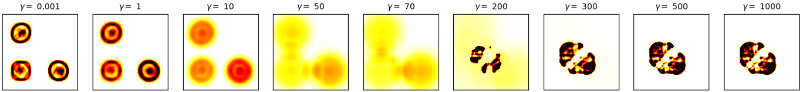
The optimal value of the WB problem is 0.2666

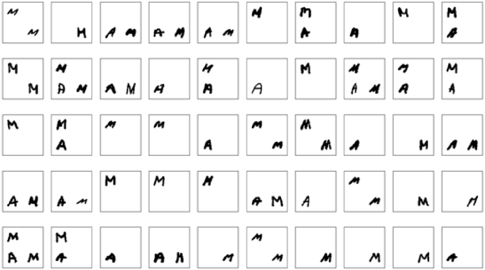
After 1 hour of processing, MAM had a barycenter distance of 0.2702, which improved to 0.2667 after 3.5 hours, when the solver of Altschuler and Boix-Adsera halts

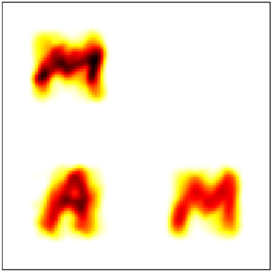
UNBALANCED WB



$$\begin{cases} \min\limits_{\pi} & \sum_{m=1}^{M} \langle d^m, \pi^m \rangle + \gamma \mathtt{dist}_{\mathcal{B}}(\pi) \\ \text{s.t.} & \pi^1 \in \Pi^m, \ldots, \pi^M \in \Pi^M \end{cases}$$







Take-away messages

- New algorithm for computing WBs which is parallelizable and can run in a randomized manner if necessary
- It can be applied to both balanced WB and unbalanced WB problems upon setting a single parameter
- Can handle additional constraints on the barycenter mass p
- It can be applied to the free or fixed-support settings
- Our Python code is freely available at `https://ifpen-gitlab.appcollaboratif.fr/detocs/mamwb`

Thank you!

D. Mimouni, P. Malisani, J. Zhu, W. de Oliveira. Computing Wasserstein barycenter via operator splitting: the method of averaged marginals, https://arxiv.org/pdf/2309.05315.pdf, 2023

Contact:

✉ welington.oliveira@minesparis.psl.eu

www.oliveira.mat.br

MINES PARIS | PSL