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

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

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

2 **IFP Energies nouvelles, France**

25th International Symposium on Mathematical Programming
Advances in Variational Analysis and Nonsmooth Optimization

Outline

I. The Wasserstein Barycenter Problem

III. The Method of Averaged Marginals

IV. Applications

V. Sparse (Nonconvex) Wasserstein Barycenter Problem

VI. Conclusion

- In applied probability, stochastic optimization, and data science, a crucial aspect is the ability to compare, summarize, and reduce the dimensionality of empirical (discrete) measures

- Since these tasks rely heavily on pairwise comparisons of measures, it is essential to use an appropriate metric for accurate data analysis

- Different metrics define different barycenters of a set of measures: a barycenter is a mean element that minimizes the (weighted) sum of all its distances to the set of target measures

- When the chosen metric is the optimal transport one, and there is mass equality between the measures, the underlying barycenter is denoted by Wasserstein Barycenter (WB)

Example extracted from [1]

30 artificial images

Barycenters using

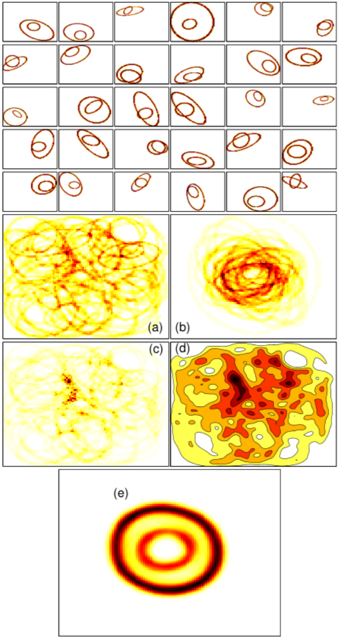
(a) Euclidean distance

(b) Euclidean + re-centering

(c) Jeffrey centroid

(d) RKHS distance

(e) 2-Wasserstein distance: Wasserstein barycenter



---

[1] M. Cuturi, A. Doucet. JMLR, 2014

THE (DISCRETE) WASSERSTEIN DISTANCE

Let $\xi, \zeta : \Omega \to \mathbb{R}^d$ be two random vectors having probability measures μ and ν:

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

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

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

DISCRETE WASSERSTEIN BARYCENTER

- Let $\alpha \in \mathbb{R}_+^M$ 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 $\nu^m \in \mathcal{P}(\Omega)$, $m = 1, \dots, M$, is a solution to the following optimization problem

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

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

DISCRETE WASSERSTEIN BARYCENTER

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

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

DISCRETE WASSERTEIN BARYCENTER - WB

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

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

- This LP scales exponentially in the number M of measures [2]
- If $M = 100$ $S^{(m)} = 3600$, $m = 1, \ldots, M$ (corresponding to figures with 60×60 pixels), the above LP has $1.2574 \cdot 10^{10}$ variables and $3.5288 \cdot 10^{6}$ constraints.

---

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

A vast body of the literature deals with inexact WBs

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 [3, 4, 5]
- Other approaches [6,7,8]

EXACT METHODS

- Methods for computing exact WBs are based on linear programming techniques and thus applicable to applications of moderate sizes [9,10]

---

[3] M. Cuturi, A. Doucet. JMLR, 2014

[4] J. Ye, J. Li. IEEE ICP (214)

[5] J. Ye et al. IEEE Transactions on Signal Processing (2017)

[6] G. Puccetti, L. Ruschendorf, S. Vanduffe. JMVA (2020)

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

[8] J. von Lindheim. COAP (2023)

[9] S. Borgwardt, S. Patterson (2020). INFORM J. Optimization

[10] J. Altschuler, E. Adsera. JMLR (2021)

Our Contribution: The Method of Averaged Marginals

OUR CONTRIBUTION

We provide an easy-to-implement, memory efficient and parallelizable algorithm based on the Douglas-Rachford splitting scheme to compute a solution to LPs of the form

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

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

Observe that we can drop the vector p (wlog)

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

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

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

$\{\pi^k\}$ converges to a solution to the above LP [11]

---

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

Given $\theta \in \mathbb{R}^{R\times\sum_{m=1}^{M} S^m}$, 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 (FIRST DR'S STEP)

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

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

Given $\theta \in \mathbb{R}^{R\times\sum_{m=1}^{M} S^m}$, let $a_m := \frac{\frac{1}{S^m}}{\sum_{j=1}^{M}\frac{1}{S^{(j)}}}$ be weights, $p^m := \sum_{s=1}^{S^m}\theta^m_{rs}$ the m^{th} marginal, $p := \sum_{m=1}^{M} a_m p^m$ the average of marginals

Proposition (First DR's step)

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

$$\pi^m_{rs} = \theta^m_{rs} + \frac{(p_r - p^m_r)}{S^m}, \quad s = 1,\dots,S^m,\ r = 1,\dots,R,\ m = 1,\dots,M$$

Proposition (Second DR's step)

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

THE METHOD OF AVERAGED MARGINALS (MAM)

MAM is a specialization of the DR algorithm applied to the WB problem

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

MAM ALGORITHM

```
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
    for $m = 1, \dots, M$ do
        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
        $p^m \leftarrow \sum_{s=1}^{S^m} \pi^m_{rs}$                                  ▷ Update the $m^{th}$ marginal
    end for
end while
```

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

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

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

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

Unbalanced WB ($\gamma > 0$)

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

MAM can be easily adapted to deal with both balanced and unbalanced WBs

Evaluating the proximal operator of $\mathtt{dist}_{\mathcal{B}}(\pi)$ amounts to projecting onto $\mathcal{B}$

CONVERGENCE ANALYSIS

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)*

Applications

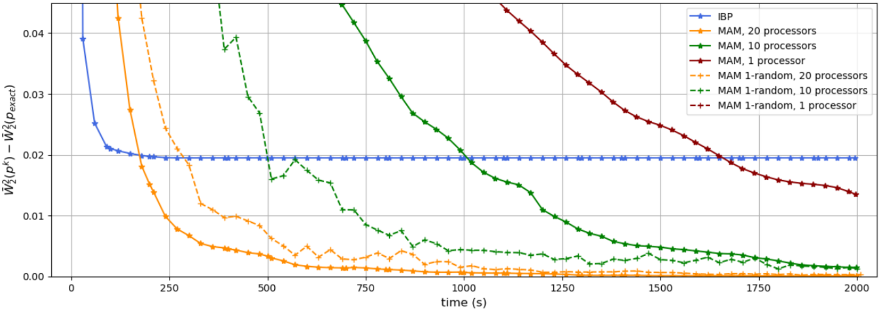
NUMERICAL EXPERIMENTS: FIXED SUPPORT $R = 1\,600$

We benchmark MAM, randomized MAM, and IBP (Iterative Bregman Projection of [12]) on the MNIST database with $M = 100$ images of 40×40 pixels. LP's dimension: 256 001 600 variables and 320 000 constraints

---

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

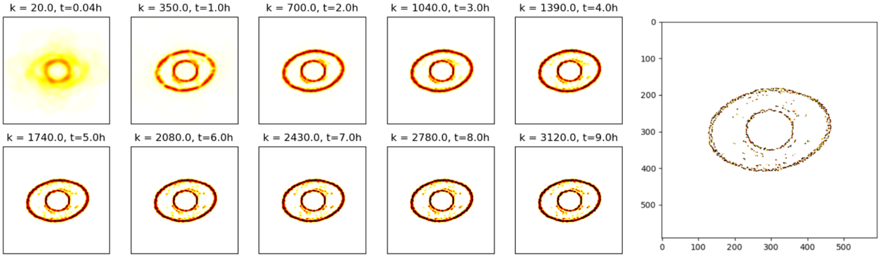
QUANTITATIVE COMPARISONS - FIXED SUPPORT $R = 1\,600$



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

Exact free-support resolution

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]



Evolution of the approximated MAM barycenter with time in regards with the exact barycenter of the Altschuler and Bois-Adsera algorithm computed in 3.5 hours [15]

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

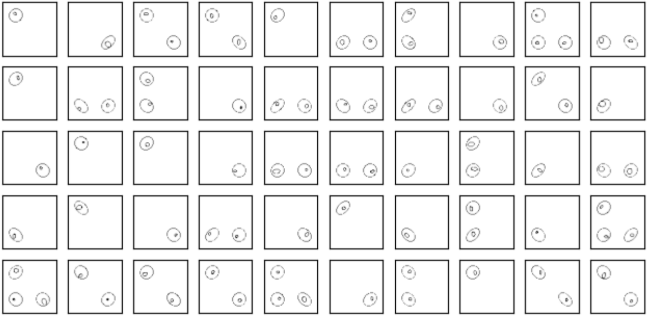
---

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

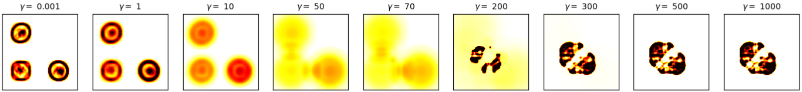
[14] `https://github.com/eboix/high_precision_barycenters`

[15] S. Borgwardt, S. Patterson (2020). INFORM J. Optimization

UNBALANCED WB



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



Sparse (Nonconvex) Wasserstein Barycenter Problem

CONSTRAINED WASSERSTEIN BARYCENTERS

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

$$\begin{cases} \min\limits_{p,\pi\geq 0} & \sum\limits_{m=1}^{M} \langle d^m, \pi^m \rangle \\ \text{s.t.} & \sum\limits_{r=1}^{R} \pi_{rs}^m = q_s^m, \quad s = 1,\dots,S^m,\ m = 1,\dots,M \\ & \sum\limits_{s=1}^{S^m} \pi_{rs}^m = p_r, \quad r = 1,\dots,R,\ m = 1,\dots,M \\ & p \in X \end{cases}$$

- If X is convex, MAM can be easily extended to compute constrained WB
- If X is nonconvex, MAM is no longer convergent

OUR PROPOSAL: DIFFERENCE-OF-CONVEX (DC) MODEL

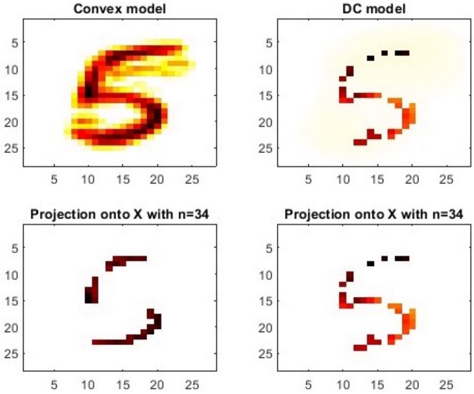
$$\begin{cases} \min\limits_{p,\pi\geq 0} & \sum\limits_{m=1}^{M} \langle d^m, \pi^m \rangle + \gamma\, \mathtt{dist}_X^2(p) \\ \text{s.t.} & \sum\limits_{r=1}^{R} \pi_{rs}^m = q_s^m, \quad s = 1,\dots,S^m,\ m = 1,\dots,M \\ & \sum\limits_{s=1}^{S^m} \pi_{rs}^m = p_r, \quad r = 1,\dots,R,\ m = 1,\dots,M \end{cases}$$

Sparse Wasserstein barycenters

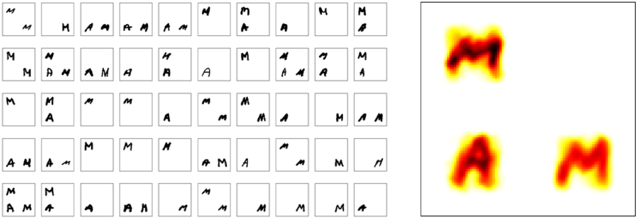
Let $X := \{p \in \mathbb{R}^R : \|p\|_0 \leq \mathtt{n}\}$

$$\left\{ \begin{array}{lll} \displaystyle\min_{p \geq 0, \pi \in B} & \displaystyle\sum_{m=1}^{M} \langle d^m, \pi^m \rangle + \gamma \, \mathtt{dist}_X^2(p) \\ \text{s.t.} & \pi^1 \in \Pi^1, \ldots, \pi^M \in \Pi^M \end{array} \right.$$

Barycenter of 10 images 28×28



Joint work with Gregorio M. Sempere, Mines Paris PSL



TAKE-AWAY MESSAGES

- New easy-to-implement and memory efficient 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
- It can be applied to the free or fixed-support settings
- It can handle convex constraints on the barycenter mass p
- For nonconvex constraints, an extension of MAM to the DC setting is under investigation

Thank you!

D. Mimouni, P. Malisani, J. Zhu, W. de Oliveira. Computing Wasserstein barycenter via operator splitting: the method of averaged marginals.
To appear in SIAM Mathematics of Data Science, 2024

- Preprint available at `https://arxiv.org/pdf/2309.05315.pdf`
- Python code is freely available at `https://ifpen-gitlab.appcollaboratif.fr/detocs/mam_wb`

CONTACT:

✉ daniel.mimouni@ifpen.fr

⌨ https://dan-mim.github.io

Annexes

SPECIAL SETTING: GRID-STRUCTURED DATA

- All measures share the same finite support: suppose that all measures $\nu^{(m)}$ are supported on a d-dimensional regular grid of integer step sizes in each direction, each coordinate going from 1 to K: $S^{(m)} = S = K^d$, and $\mathtt{supp}(\nu^{(m)}) := \{\zeta_1, \dots, \zeta_S\}$, $m = 1, \dots, M$
- The measures are evenly weighted $\alpha_m = \frac{1}{M}$, $m = 1, \dots, M$
- Then $\mathtt{supp}(\mu)$ has at most

$$R \leq ((K-1)M+1)^d$$

points, as the finer grid only runs between the boundary points [16]

This significantly reduces the LP's dimension

---

[16] S. Borgwardt, S. Patterson (2020). INFORM J. Optimization

2-WASSERSTEIN DISTANCE SETTING

EXAMPLE (LP'S DIMENSIONS)

Consider the case: $M = 10$, $d = 2$, $K = 40 \Rightarrow S = 1600$

data	$\lvert\mathtt{supp}(\mu)\rvert$	# variables	# eq. constraints
	R	$(MS+1)R$	$(S+R)M$
general	$1.0995 \cdot 10^{32}$	$1.7593 \cdot 10^{36}$	$1.0995 \cdot 10^{33}$
grid-structured	152881	$2.4462 \cdot 10^{9}$	1544810

- In contrast to the worst-case, exponentially sized possible support set, there always exists a WB $\bar{\mu}$ with provably sparse support

$$|\mathtt{supp}(\bar{\mu})| \leq \sum_{m=1}^{M} S^{(m)} - M + 1$$

- For the above example $|\mathtt{supp}(\bar{\mu})| \leq 15991$

- This fact motivates heuristics for computing inexact WBs: fixed-support approaches, which generally fix R to $\sum_{m=1}^{M} S^{(m)} - M + 1$ (or fewer) points

THE METHOD OF AVERAGED MARGINALS - MAM

UNBALANCED WASSERSTEIN BARYCENTER

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
        $p^m \leftarrow \sum_{s=1}^{S^m} \pi^m_{rs}$                   ▷ Update the $m^{th}$ marginal
    end for
end while
```

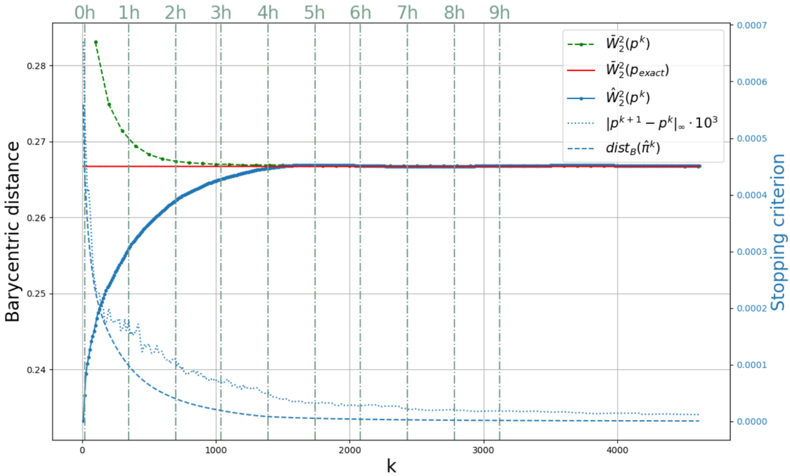
Set $\gamma = \infty$ to compute balanced WB (if the measures are balanced)
Otherwise, choose $\gamma \in (0, \infty)$ to compute unbalanced WB

THE METHOD OF AVERAGED MARGINALS - MAM

CONSTRAINED SETTING

ALGORITHM

```
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
    $p \leftarrow \texttt{Proj}_X \left( \sum_{m=1}^{M} a_m p^m \right)$        ▷ Average the marginals
    for $m = 1, \dots, M$ do
        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
        $p^m \leftarrow \sum_{s=1}^{S^m} \pi^m_{rs}$        ▷ Update the $m^{th}$ marginal
    end for
end while
```



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