Scenario Tree Reduction and Operator Splitting Method for Stochastic Optimization of Energy Systems

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

PhD. Mid-term Evaluation

Personal background

- Lycée Henri IV, et Louis Le Grand — Prépa MPSI/PSI* — 2015-2017
- École Centrale Lyon — Engineering school — 2017-2021
- Imperial College — Master of Science — 2019-2020
- Ile de France Mobilités — Data scientist — 2021-2022

Overview

Some context

Multistage stochastic optimization problem

$$\mathrm{val}(\xi) := \min_{u_1} c_1(x_1, u_1, \xi_1) + \min_{u_2} \mathbb{E}_{\xi_2}\left[c_2(x_2, u_2, \xi_2) + \cdots + \min_{u_T} \mathbb{E}_{\xi_T}\left[c_T(x_T, u_T, \xi_T)\right]\right]$$

under the following constraints

$$x_{t+1} = f_t(x_t, u_t, \xi_t) \tag{1a}$$

$$(u_t, x_t) \in K_t \subset \mathbb{R}^m \times \mathbb{R}^n \tag{1b}$$

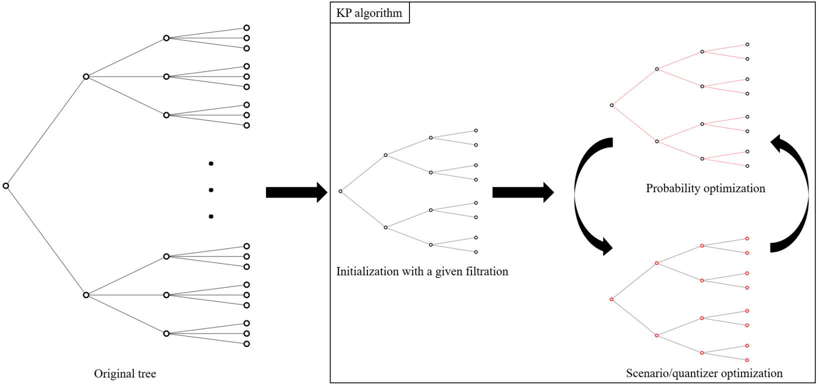
$$x_1 = x^0, \tag{1c}$$

- Decision management problems
- Multiple solving algorithms exist: MPC, Dynamic programming, Scenario based methods (Prog. Hedging, SDDP, etc.)
- In practice, the scenario process $\{\xi_t\}$ is approximated by a scenario tree

The Scenario Reduction Algorithm

Kovacevic and Pichler's algorithm (KP)

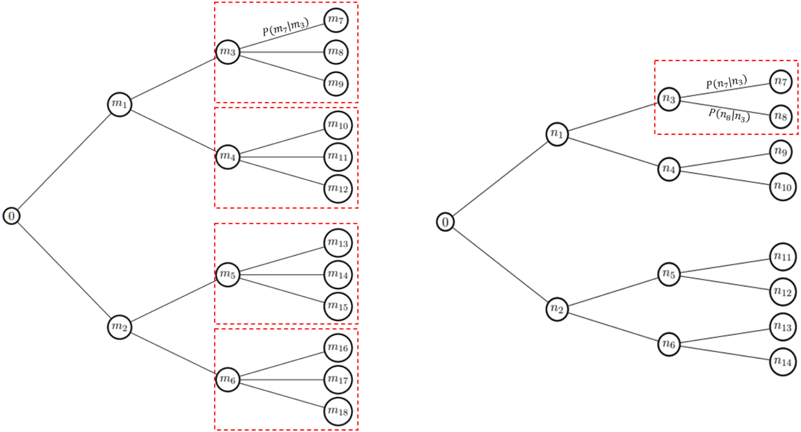
- For statistical representativity, the scenario tree should be large
- For computation tractability, the scenario tree should be small



KP algorithm: to approximate a tree, a smaller tree with a given filtration is improved in order to minimize the distance with the original tree. The probabilities and the scenaro values are alternatively optimized until convergence.

Wasserstein Barycenter (WB) within the KP Algorithm

- In the Scenario Reduction problem with seek $\mathbf{G}$ (with given filtration $\mathcal{F}'_t$) that minimizes $\mathrm{dl}_2(\mathbf{H}, \mathbf{G})$
- Our first contribution is to notice than the steps of the KP algorithm is a Wasserstein Barycenter problems (WB)



(left) Original tree, *(right)* Approximated tree. The probabilities $(P(n_7|n_3), P(n_8|n_3))$ are computed as the Wasserstein Barycenter of the set of (known) probabilities associated to the boxed subtrees on the left.

Navigate the Wasserstein Barycenter Challenge

Extension to the WB

The Wasserstein distance

Let (Ω, P) a probabilty space and two measurable random vectors $\xi, \zeta : \Omega \mapsto \mathbb{R}^d$ such that $\mu := \xi_{\#}P$ and $\nu := \zeta_{\#}P$. Their (quadratic) 2-Wasserstein distance is

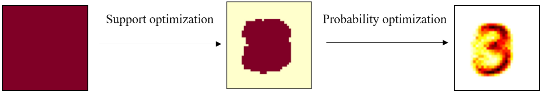
$$W_2(\xi, \zeta) := \left(\inf_{\pi \in U(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \mathrm{d}\pi(x, y) \right)^{1/2}, \qquad \text{(WD)}$$

Here $U(\mu, \nu)$ is the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ having marginals μ and ν.

Wasserstein Barycenter problem

Given M measures $\{\nu^{(1)}, \ldots, \nu^{(M)}\} \subset \mathcal{P}(\mathbb{R}^d)$ and $\alpha \in \Delta_M$, an 2-*Wasserstein barycenter* with weights α is a solution to the following optimization problem

$$\min_{\mu \in \mathcal{P}(\mathbb{R}^d)} \sum_{m=1}^{M} \alpha_m W_2^2(\mu, \nu^{(m)}) \,. \qquad (2)$$



Huge scale Linear Program (LP)

WB problem

It boils down to solve a huge LP with $MRS + R$ variables[1]:

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

[1] If $M = 100$ $R = S^{(m)} = 1600$, $m = 1,\ldots,M$ (corresponding to figures with 40×40 pixels), the above LP has 256 001 600 variables.

Reformulation of the LP

Separable sets

$$\Pi^{(m)} := \left\{ \pi^{(m)} \geq 0 : \sum_{r=1}^{R} \pi_{rs}^{(m)} = q_s^{(m)},\ s = 1, \ldots, S^{(m)} \right\},\ m = 1, \ldots, M \qquad (3)$$

and the linear subspace

$$\mathcal{B} := \left\{ \pi = (\pi^{(1)}, \ldots, \pi^{(M)}) : \begin{array}{lll} \sum_{s=1}^{S^{(1)}} \pi_{rs}^{(1)} & = \sum_{s=1}^{S^{(2)}} \pi_{rs}^{(2)}, & \forall r \\ \sum_{s=1}^{S^{(2)}} \pi_{rs}^{(2)} & = \sum_{s=1}^{S^{(3)}} \pi_{rs}^{(3)}, & \forall r \\ & \vdots & \\ \sum_{s=1}^{S^{(M-1)}} \pi_{rs}^{(M-1)} & = \sum_{s=1}^{S^{(M)}} \pi_{rs}^{(M)}, & \forall r \end{array} \right\} \qquad (4)$$

Douglas-Rachford theory

- Consider the indicator function $\mathbf{i}_C$ of a convex set C, and reformulate the problem:

$$f^{(m)}(\pi^{(m)}) := \sum_{r=1}^{R} \sum_{s=1}^{S^{(m)}} c_{rs}^{(m)} \pi_{rs}^{(m)} + \mathbf{i}_{\Pi^{(m)}}(\pi^{(m)}), \quad m = 1, \ldots, M, \tag{5}$$

- Recast problem:

$$f(\pi) := \sum_{m=1}^{M} f^{(m)}(\pi^{(m)}) \quad \text{and} \quad g(x) := \begin{cases} \mathbf{i}_{\mathcal{B}}(\pi) & \text{if balanced} \\ \gamma\, \mathtt{dist}_{\mathcal{B}}(\pi) & \text{if unbalanced.} \end{cases} \tag{6}$$

- Since f is polyhedral, it follows that computing one of its solutions is equivalent to

$$\text{find} \quad \pi \quad \text{such that} \quad 0 \in \partial f(\pi) + \partial g(\pi). \tag{7}$$

Douglas-Rachford algorithm

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

$$\begin{cases} \pi^{k+1} = \mathrm{prox}_{g/\rho}(\theta^k) \\ \hat{\pi}^{k+1} = \mathrm{prox}_{f/\rho}(2\pi^{k+1} - \theta^k) \\ \theta^{k+1} = \theta^k + \hat{\pi}^{k+1} - \pi^{k+1} \end{cases} \tag{8}$$

The Method of Averaged Marginals (MAM)

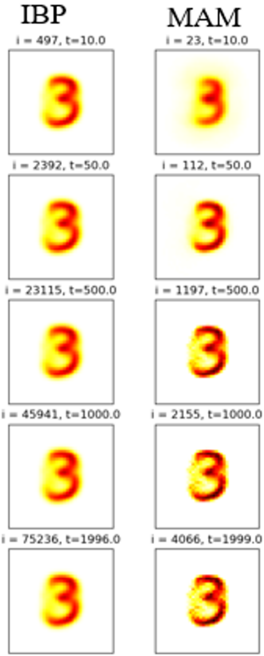
The Douglas Rachford and the MAM Algorithm

$$\begin{cases} \pi^{k+1} & = & \texttt{Proj}_{\mathcal{B}}(\theta^k) \xrightarrow{\text{explicit}} \pi_{rs}^{(m)} := \theta_{rs}^{(m)} + \frac{(p_r^k - p_r^{(m)})}{S^{(m)}} \\ \hat{\pi}^{k+1} & := & \begin{pmatrix} \hat{\pi}_{1s}^{(m)} \\ \vdots \\ \hat{\pi}_{Rs}^{(m)} \end{pmatrix} = \texttt{Proj}_{\Delta_R(q_s^{(m)})} \begin{pmatrix} y_{1s} - \frac{1}{\rho} c_{1s}^{(m)} \\ \vdots \\ y_{Rs} - \frac{1}{\rho} c_{Rs}^{(m)} \end{pmatrix}, \quad s = 1, \dots, S^{(m)} \\ \theta^{k+1} & = & \theta^k + \hat{\pi}^{k+1} - \pi^{k+1} \end{cases}$$

- At every iteration the barycenter approximation p^k is a weighted average of the M marginals $p^{(m)} := \sum_{s=1}^{S^{(m)}} \theta_{rs}^{(m)}$.
- The projection onto the simplex can be accomplished in parallel, exactly and efficiently by specialized algorithms[2].
- The whole sequence converges to an exact barycenter.

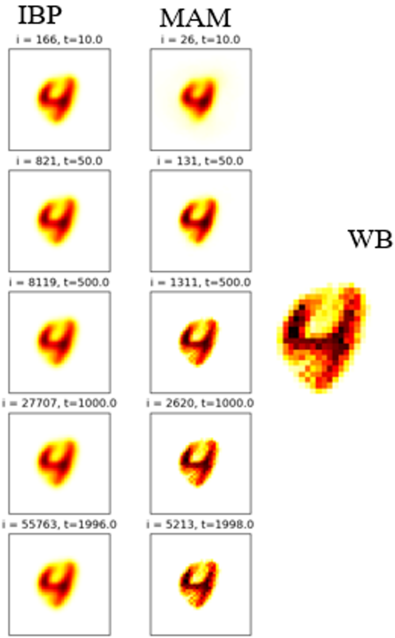
[2] see L. Condat 2016

Qualitative comparisons

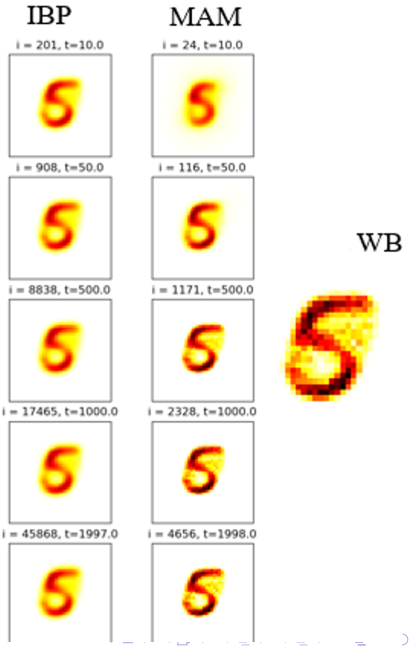


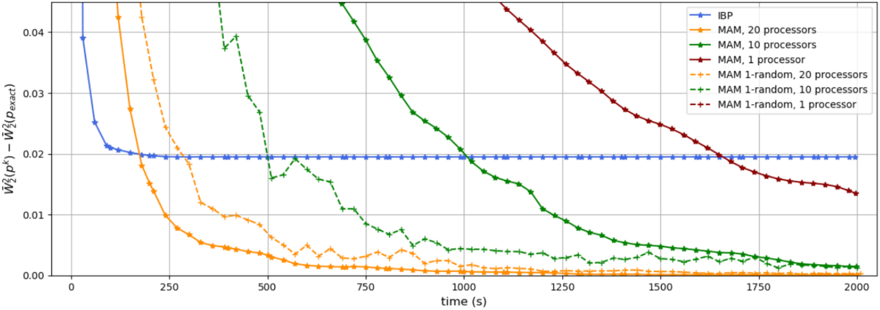
WB

Qualitative comparisons

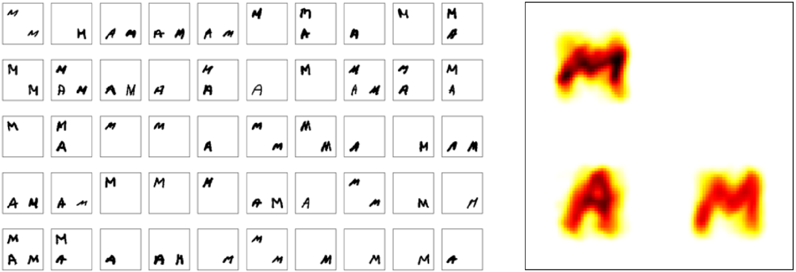


Qualitative comparisons





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



Take-away messages

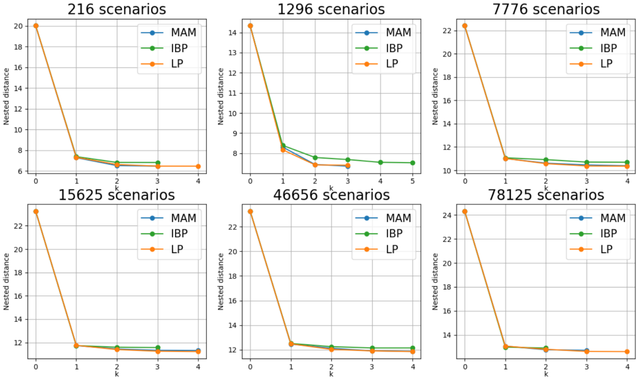
- New algorithm for computing WBs which is parallelizable and can run in a randomized manner if necessary
- It can be applied to both balanced WB and unbalanced WB problems upon setting a single parameter
- It can be applied to the free or fixed-support settings
- The method is submitted (after review) in SIAM Journal on Mathematics and Data Science `https://arxiv.org/pdf/2309.05315.pdf`
- Our Python code is freely available at `https://ifpen-gitlab.appcollaboratif.fr/detocs/mam_wb`

Exploring Utilizations of the Enhanced Scenario Reduction Algorithm

Reduction scenario applications

Scenario tree reduction employing different solvers to compute the WBs:

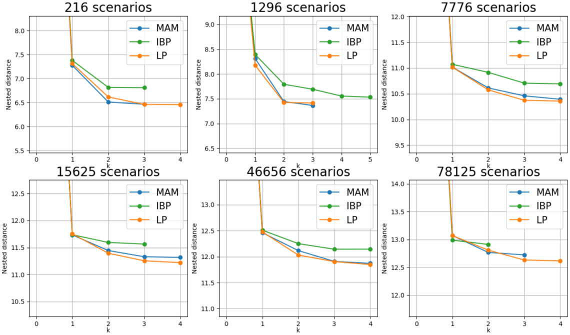
- A classic LP (KP setting),
- Iterative Bregmann Projection algorithm [3],
- MAM.



Evolution of the Nested Distance along the reduction iterations for different initial trees.

[3] see the work of D. Bennammou and G. Peyré

Reduction scenario applications



Evolution of the Nested Distance along the reduction iterations for different initial trees with a zoom.

	LP	IBP	MAM	MAM 4 processors
T=4, cpn=6	0.17	0.49	2.21	0.56
T=5, cpn=6	1.54	14.83	18.23	6.28
T=6, cpn=6	74.25	161.19	344.83	124.44
T=7, cpn=5	487.58	323.76	816.46	341.62
T=7, cpn=6	4905	2136	2541	1256
T=8, cpn=5	13797	4334	3458	1635

Table: Total time (in seconds) per method for the studied trees.

Conclusion

Take-away messages

- We developed a new approach to reduce scenario tree, that can deployed even for very large trees
- Our Python code is freely available at `https://ifpen-gitlab.appcollaboratif.fr/detocs/tree_reduction`

Valorization of the Research

- The MAM algorithm is the subject of a paper submitted to SIAM Journal on Mathematics and Data Sciences (SIMODS) [4],
- Presentation during the PGMO days, the annual conference of the Optimization, OR, and Data Science program of the FMJH [5],
- Presentation of a poster at the Consortium in Applied Mathematics (CIROQUO)[6],
- Presentation of a poster during the DATA IA days of CentraleSupelec[7],
- Presentation of extensions of MAM at the international conference ISMP 2024 (International Symposium on Mathematical Programming[8].

---

[4] https://www.siam.org/publications/journals/siam-journal-on-mathematics-of-data-science-simods

[5] Fondation Mathématiques Jacques Hadamard: https://smf.emath.fr/evenements-smf/pgmo-days-2023

[6] https://ciroquo.ec-lyon.fr/evenements.html

[7] https://www.dataia.eu/

[8] https://ismp2024.gerad.ca/

Annexes

Distance between processes

- How to evaluate such a distance?

The Wasserstein distance

Let (Ω, P) a probabilty space and two measurable random vectors $\xi, \zeta : \Omega \mapsto \mathbb{R}^d$ such that $\mu := \xi_{\#}P$ and $\nu := \zeta_{\#}P$. Their (quadratic) 2-Wasserstein distance is

$$W_2(\xi, \zeta) := \left(\inf_{\pi \in U(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \mathrm{d}\pi(x, y) \right)^{1/2}, \qquad \text{(WD)}$$

Here $U(\mu, \nu)$ is the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ having marginals μ and ν.

The Nested Distance

Let $\mathbf{H} := (\Xi, (\mathcal{F}_t)_t, P)$ and $\mathbf{G} := (Z, (\mathcal{F}'{}_t)_t, P')$ be two filtered probability spaces, $\Xi := \{\xi^{(1)}, \dots, \xi^{(R)}\}$ and $Z := \{\zeta^{(1)}, \dots, \zeta^{(S)}\}$ are the scenario values. The process distance of order 2, between the trees $\mathbf{H}$ and $\mathbf{G}$ is the root square of the optimal value of the following LP,

$$\mathrm{dl}_2(\mathbf{H}, \mathbf{G})^2 := \begin{cases} \min\limits_{\pi} & \sum\limits_{i=1}^{R} \sum\limits_{j=1}^{S} \|\xi^{(i)} - \zeta^{(j)}\|_{\ell^2}^2 \, \pi(\xi^{(i)}, \zeta^{(j)}) \\ \text{s.t.} & \pi(M \times Z | \mathcal{F}_t \otimes \mathcal{F}'_t) = P(M | \mathcal{F}_t), \qquad (M \in \mathcal{F}_T, t = 0, \dots, T) \\ & \pi(\Xi \times N | \mathcal{F}_t \otimes \mathcal{F}'_t) = P'(N | \mathcal{F}'_t), \qquad (N \in \mathcal{F}'_T, t = 0, \dots, T) \\ & \pi \geq 0. \end{cases} \qquad \text{(ND)}$$

Stability Results for ND

Stability result for the ND

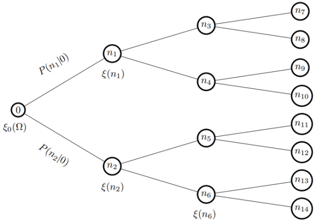
Consider the value function val($\mathbf{H}$) of stochastic optimization problem seen earlier so that $\text{val}(\mathbf{H}) := \text{val}(\xi^H)$, and L_2 a constant, then it holds[9]:

$$|\text{val}(\mathbf{H}) - \text{val}(\mathbf{G})| \leq L_2 \cdot \text{dl}_2(\mathbf{H}, \mathbf{G})^2 \tag{9}$$

- It is not the case when using the WD.

[9] See Pflug and Pichler 2012

The Nested Distance



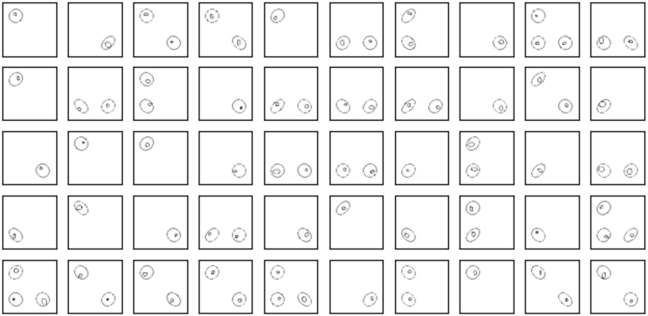
The Nested distance for trees

the process distance of order 2, between **H** and **G** is the square root of the optimal value of the following LP,

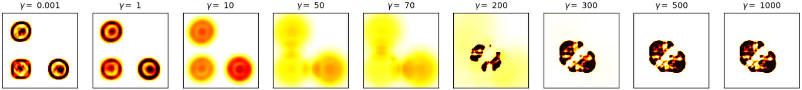
$$\mathrm{dl}_2(\mathbf{H},\mathbf{G}) := \begin{cases} \min\limits_{\pi} & \sum\limits_{i\in\mathcal{N}_T, j\in\mathcal{N}'_T} \pi_{i,j}\mathtt{d}^2_{i,j} \\ \text{s.t.} & \sum_{\{j:n\in\mathcal{A}(j)\}} \pi(i,j|m,n) = P(i|m), \quad (m\in\mathcal{A}(i), n) \\ & \sum_{\{i:m\in\mathcal{A}(i)\}} \pi(i,j|m,n) = P'(j|n), \quad (n\in\mathcal{A}(j), m) \\ & \pi_{i,j}\geq 0 \text{ and } \sum_{i,j}\pi_{i,j} = 1. \end{cases} \tag{NDT}$$

This LP can be decomposed into several Optimal Transport problems (OT).

Unbalanced WB



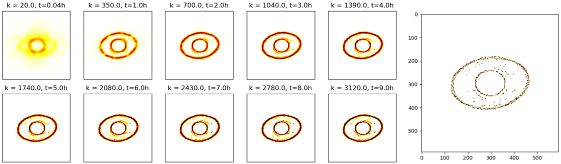
$$\begin{cases} \min\limits_{\pi} & \sum_{m=1}^{M} \langle c^{(m)}, \pi^{(m)} \rangle + \gamma \mathtt{dist}_{\mathcal{B}}(\pi) \\ \text{s.t.} & \pi^{(1)} \in \Pi^{(m)}, \ldots, \pi^{(M)} \in \Pi^{(M)} \end{cases}$$



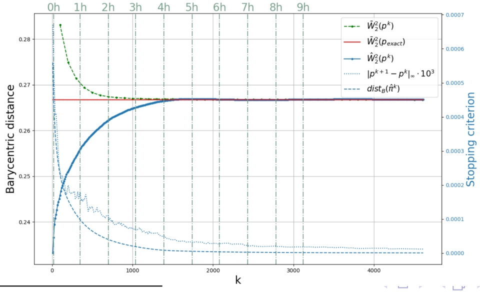
Exact free-support resolution

- 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: $S^{(m)} = S = K^d$.

- The measures are evenly weighted $\alpha_m = \frac{1}{M}$, $m = 1, \ldots, M$



Evolution of the approximated MAM barycenter with time in regards with the exact barycenter of the Altschuler and Bois-Adsera algorithm computed in 4 hours [10]



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