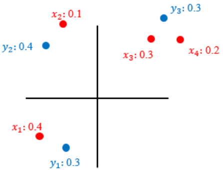
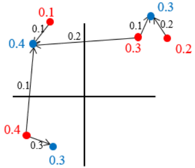
Annex

The (Discrete) Wasserstein Distance





		ξ_1 0.4	ξ_2 0.1	ξ_3 0.3	ξ_4 0.2
ζ_1	0.3	0.3	0	0	0
ζ_2	0.4	0.1	0.1	0.2	0
ζ_3	0.3	0	0	0.1	0.2

We focus on discrete probability measures based on

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

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

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

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

The Wasserstein distance - discrete setting

The 2-Wassestein distance between μ and ν is:

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

with

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

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 $\texttt{supp}(\mu) := \{\xi_1, \dots, \xi_R\}$

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

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

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

The Wasserstein distance - discrete setting

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

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

with

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

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:

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

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, $p^m := \sum_{s=1}^{S^m} \theta_{rs}^m$ 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:*

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

Proposition (Second DR's step)

The proximal mapping $\hat{\pi} = \arg\min_{\substack{\pi^m \in \Pi^m \\ m=1,\ldots,M}} \sum_{m=1}^{M} \langle d^m, \pi^m \rangle + \frac{\rho}{2} \|\pi - y\|^2$ *can be computed exactly, in parallel along the columns of each transport plan* y^m*, as follows: for all* $m \in \{1, \ldots, M\}$*,*

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

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

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)*

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 $\texttt{supp}(\nu^{(m)}) := \{\zeta_1, \dots, \zeta_S\}$, $m = 1, \dots, M$
- The measures are evenly weighted $\alpha_m = \frac{1}{M}$, $m = 1, \dots, M$
- Then $\texttt{supp}(\mu)$ has at most

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

points, as the finer grid only runs between the boundary points [1]

$\boxed{\text{This significantly reduces the LP's dimension}}$

[1] 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	$\lvert\mathtt{supp}(\mu)\rvert$ R	# variables $(MS+1)R$	# eq. constraints $(S+R)M$
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

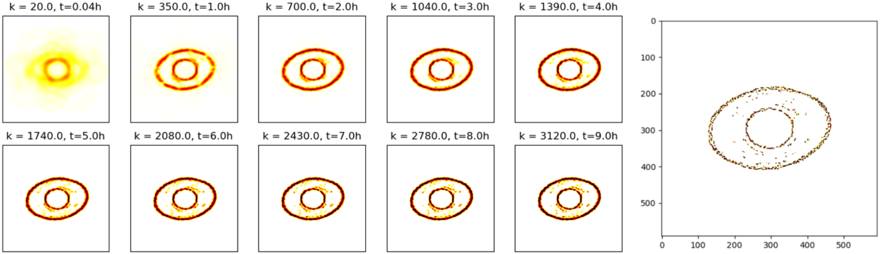
- In contrast to the worst-case, exponentially sized possible support set, there always exists a WB $\bar{\mu}$ with provably sparse support

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

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

Exact free-support resolution

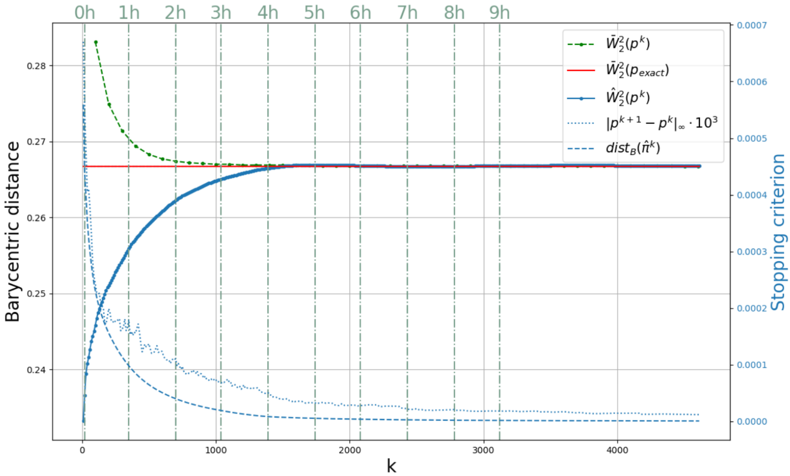
The dataset we use is the one from [2]: $M = 10$ images of 60×60 pixels
LP's dimension: $1.2574 \cdot 10^{10}$ variables and $3.5288 \cdot 10^{6}$ constraints
We compare with the dedicated solver of Altschuler and Boix-Adsera, available at [3]



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 [4]

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

[2] J. M. Altschuler and E. Boix-Adsera. JMLR (2021)
[3] `https://github.com/eboix/high_precision_barycenters`
[4] S. Borgwardt, S. Patterson (2020). INFORM J. Optimization

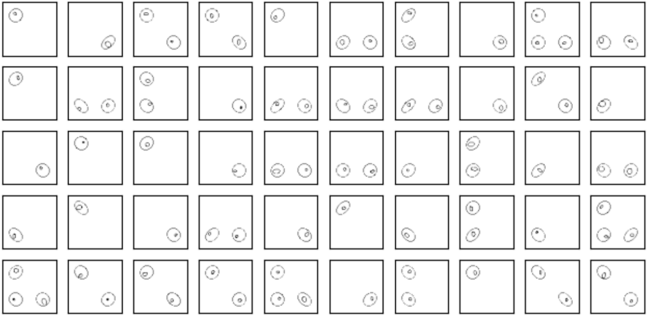


The optimal value of the WB problem is 0.2666

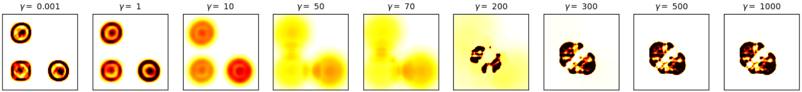
After 1 hour of processing, MAM had a barycenter distance of 0.2702, which improved to 0.2667 after 3.5 hours, when the solver of Altschuler and Boix-Adsera halts

Prospects - Unbalanced WB

- A new approach, very useful in the optimal transport community.



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



- Detailed in the SIAM publication.

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

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

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

Unbalanced WB ($\gamma > 0$)

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

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 $\mathcal{B}$

Algorithm

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

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

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

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

Constrained Wasserstein barycenters

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

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

- 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

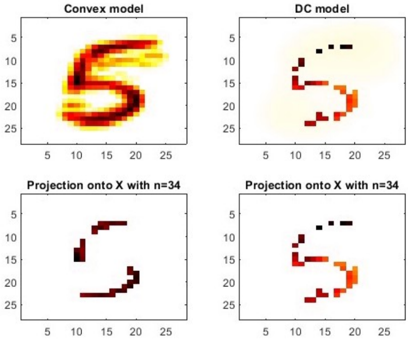
$$\begin{cases} \min_{p,\pi\geq 0} & \sum_{m=1}^{M} \langle d^m, \pi^m \rangle + \gamma\, \mathtt{dist}_X^2(p) \\ \text{s.t.} & \sum_{r=1}^{R} \pi_{rs}^m = q_s^m, \quad s = 1,\dots,S^m,\ m = 1,\dots,M \\ & \sum_{s=1}^{S^m} \pi_{rs}^m = p_r, \quad r = 1,\dots,R,\ m = 1,\dots,M \end{cases}$$

Constrained Wasserstein barycenters[5]

$$\begin{cases} \min_{p \geq 0, \pi \in \mathcal{B}} & \sum_{m=1}^{M} \langle d^m, \pi^m \rangle \\ \text{s.t.} & \pi^1 \in \Pi^1, \ldots, \pi^M \in \Pi^M \\ & p \in X \end{cases}$$

- If X is convex, MAM can be easily extended to compute constrained WB;
- If X is nonconvex, MAM is no longer convergent. We proposed a difference-of-convex extension.

Sparse barycenter of 10 images 28×28 with $X := \{p \in \mathbb{R}^R : \|p\|_0 \leq \mathtt{n}\}$



[5] Joint work with Gregorio M. Sempere, Mines Paris PSL

MAM constrained

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**

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

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

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

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

$\quad\quad$ **end for**

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

$\quad$ **end for**

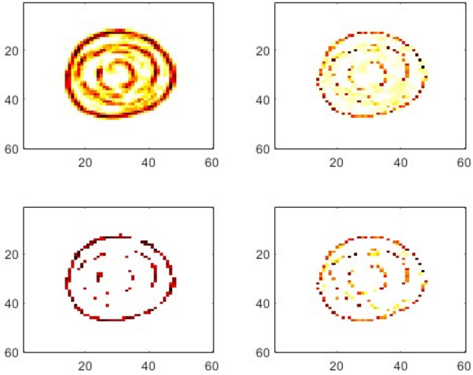
end while

Sparse Wasserstein barycenters

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

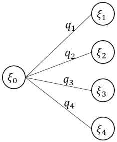
$$\begin{cases} \min\limits_{p \geq 0, \pi \in B} & \sum_{m=1}^{M} \langle d^m, \pi^m \rangle + \gamma \, \mathtt{dist}_X^2(p) \\ \text{s.t.} & \pi^1 \in \Pi^1, \ldots, \pi^M \in \Pi^M \end{cases}$$

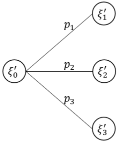
Barycenter of 10 images 28×28



Joint work with Gregorio M. Sempere, Mines Paris PSL

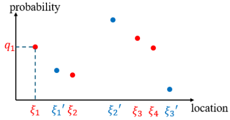
Distance between processes



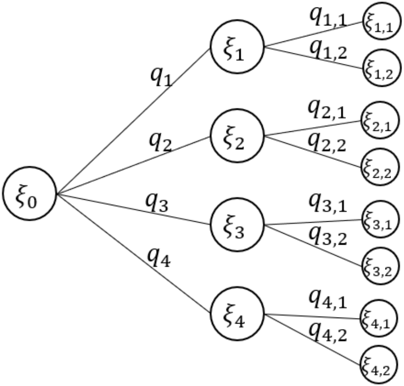


- Two stage trees can be represented as discrete probability measures.

For example if supports are 1D:



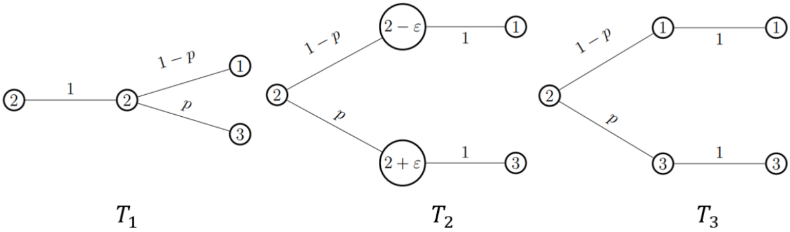
$$\mu = q_1\delta_{\xi_1} + q_2\delta_{\xi_2} + q_3\delta_{\xi_3} + q_4\delta_{\xi_4}$$
$$\nu = p_1\delta_{\xi'_1} + p_2\delta_{\xi'_2} + p_3\delta_{\xi'_3}$$



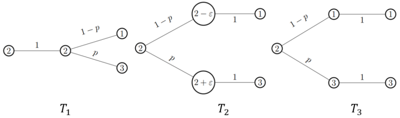
- Multi stage trees have filtration.

We need an extension of the Wasserstein distance to random processes.

Nested distance vs Wasserstein distance 1/2



Nested distance vs Wasserstein distance 2/2



- The Wasserstein distance: the trees are two scenarios. Here,

$$T_1: \quad \xi^1_{T_1} = (2,2,1), \quad \xi^2_{T_1} = (2,2,3),$$

and

$$T_2: \quad \xi^1_{T_2} = (2,2-\varepsilon,1), \quad \xi^2_{T_2} = (2,2+\varepsilon,3).$$

The squared Euclidean distance matrix between these scenarios is then

$$\begin{pmatrix} \varepsilon^2 & \varepsilon^2+4 \\ \varepsilon^2+4 & \varepsilon^2 \end{pmatrix}.$$

The optimal transport plan:

$$\begin{pmatrix} 1-p & 0 \\ 0 & p \end{pmatrix}.$$

Therefore, we obtain

$$W_2(T_1,T_2) = \varepsilon \xrightarrow[\varepsilon\to 0]{} 0.$$

- The nested distance : The d-distance at first stage is 0. At the middle stage is given by

$$\begin{pmatrix} \varepsilon^2 \\ \varepsilon^2 \end{pmatrix},$$

and at the final stage by

$$\begin{pmatrix} \varepsilon^2 & \varepsilon^2+4 \\ \varepsilon^2+4 & \varepsilon^2 \end{pmatrix}.$$

Then the transport plan at the middle stage:

$$\begin{pmatrix} 1-p \\ p \end{pmatrix},$$

and at the final stage:

$$\begin{pmatrix} (1-p)^2 & (1-p)p \\ (1-p)p & p^2 \end{pmatrix}.$$

Therefore,

$$\mathrm{ND}_2(T_1,T_2) = 2\varepsilon^2\big(1+p(1-p)\big)+8p(1-p) \xrightarrow[\varepsilon\to 0]{} 8p(1-p).$$

- $W_2(T_1,T_3) = 1$ and $W_2(T_2,T_3) = 1-\varepsilon$, means that, as $\varepsilon \to 0$, the trees T_2 and T_3 become equidistant from T_1.
- $\mathrm{ND}_2(T_1,T_3) = 2+10p(1-p)$, and $\mathrm{ND}_2(T_2,T_3) = 2(1-\varepsilon)^2$.

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[6]:

$$|\text{val}(\mathbf{H}) - \text{val}(\mathbf{G})| \leq L_2 \cdot \text{dl}_2(\mathbf{H}, \mathbf{G})^2 \qquad (1)$$

- It is not the case when using the WD.

[6] See Pflug and Pichler 2012

Probability Optimization in the KP algorithm

Given the stochastic process quantizers $\{\xi'(n) \in \Xi : n \in \mathcal{N}'\}$ and structure of $(\mathcal{N}', A')$, we are looking for the optimal probability measure P' to approximate $\mathbf{P} :=(\Xi^{T+1}, \mathcal{F}, P)$, regarding the nested distance.

Large non-convex optimization problem

$$\left\{\begin{array}{lll} \min\limits_{\pi, P'} & \sum\limits_{i \in \mathcal{N}_T, j \in \mathcal{N}'_T} \pi(i,j) \mathtt{d}^{\iota}_{i,j} & \\ & & \\ \text{s.t.} & \sum_{j \in n+} \frac{\pi(i,j)}{\pi(m,n)} = P(i|m), & (\forall m \in \mathcal{A}(i), n) \\ & \sum_{i \in m+} \frac{\pi(i,j)}{\pi(m,n)} = P'(j|n), & (\forall n \in \mathcal{A}(j), m) \\ & \pi_{i,j} \geq 0 \text{ and } \sum_{i,j} \pi_{i,j} = 1 & \\ & P'(j|j-) \geq 0. & \end{array}\right. \tag{2}$$

- This is a bilinear problem
- There is a large number of decision variables and bilinear constraints.

From Bilinear to Recursive Problem

- $\pi(i,j) = \pi(i,j|m,n) \times \pi(m,n)$,
- $\delta_\iota(m,n) := \sum_{i\in m+, j\in n+} \pi(i,j|m,n)\delta_\iota(i,j)$ for $m \in \mathcal{N}_t, n \in \mathcal{N}'_t$,
- $\delta_\iota(i,j) = \mathtt{d}_\iota(\xi_i, \xi'_j)^\iota =: \mathtt{d}^\iota_{i,j}$ for the leaves i, j of the trees.

$$\sum_{i\in\mathcal{N}_T, j\in\mathcal{N}'_T} \pi(i,j)\mathtt{d}^\iota_{i,j} = \sum_{i\in\mathcal{N}_T, j\in\mathcal{N}'_T} \pi(i,j)\delta_\iota(i,j) \tag{3a}$$

$$= \sum_{i\in\mathcal{N}_T, j\in\mathcal{N}'_T} \sum_{m\in i-, n\in j-} \pi(i,j|m,n)\pi(m,n)\delta_\iota(i,j) \tag{3b}$$

$$= \sum_{n\in\mathcal{N}'_{T-1}} \sum_{m\in\mathcal{N}_{T-1}} \pi(m,n) \underbrace{\sum_{i\in m+, j\in n+} \pi(i,j|m,n)\delta_\iota(i,j)}_{\delta_\iota(m,n)} \tag{3c}$$

$$= \sum_{n\in\mathcal{N}'_{T-1}} \sum_{m\in\mathcal{N}_{T-1}} \pi(m,n)\delta_\iota(m,n). \tag{3d}$$

Evaluate the ND recursively

$\delta_\iota(0,0) = \mathrm{ND}(\mathbf{P}, \mathbf{P}')$

From a large LP to an Optimal Transport Problem

The recursive problem (RP) is a Wasserstein Barycenter problem.

Given $t \in \{1, \dots, T\}$ and $n \in \mathcal{N}'_t$, problem (RP) reads as:

$$
\left\{
\begin{array}{lll}
\min\limits_{\pi} & \sum_{m \in \mathcal{N}_t} \pi(m,n) \sum_{i \in m+, j \in n+} \pi(i,j|m,n) \delta_\iota(i,j) & \\
 & & \\
\text{s.t.} & \sum_{j \in n+} \pi(i,j|m,n) = P(i|m), & (i \in m+) \\
 & \sum_{i \in m+} \pi(i,j|m,n) = \sum_{i \in \tilde{m}+} \pi(i,j|\tilde{m},n), & (j \in n+ \text{ and } m, \tilde{m} \in \mathcal{N}_t) \\
 & \pi(i,j|m,n) \geq 0. &
\end{array}
\right.
\tag{RP}
$$

- $\sum_{i \in m+} \pi(i,j|m,n) = P'(j|n)$ for $j \in n+$, for all $m = m_1, \dots, m_M$.

From a large LP to an Optimal Transport Problem

The recursive problem (RP) is a Wasserstein Barycenter problem.

Given $t \in \{1, \dots, T\}$ and $n \in \mathcal{N}'_t$, problem (RP) reads as:

$$
\begin{cases}
\min\limits_{P',\pi} & \pi(m_1,n) \sum\limits_{i \in m_1+, j \in n+} \pi(i,j|m_1,n)\delta_\iota(i,j) \quad + \cdots + \quad \pi(m_M,n) \sum\limits_{i \in m_M+, j \in n+} \pi(i,j|m_M,n) \\
\text{s.t.} & \sum_{j \in n+} \pi(i,j|m_1,n) \qquad\qquad = P(i|m_1), \qquad (i \in m_1+) \\
& \qquad \ddots \qquad\qquad\qquad \vdots \\
& \qquad\qquad \sum_{j \in n+} \pi(i,j|m_M,n) = P(i|m_M), \qquad (i \in m_M+) \\
& \sum_{i \in m_1+} \pi(i,j|m_1,n) \qquad\qquad = P'(j|n), \qquad (j \in n+) \\
& \qquad \ddots \qquad\qquad\qquad \vdots \\
& \qquad\qquad \sum_{i \in m_M+} \pi(i,j|m_M,n) = P'(j|n), \qquad (j \in n+) \\
& \sum_{j \in n+} P'(j|n) = 1, \pi(i,j|m_1,n) \geq 0 \quad \cdots \quad \pi(i,j|m_M,n) \geq 0,
\end{cases}
\tag{WB}
$$

- $\sum_{i=1}^{M} \pi(m_i,n) = 1$ per definition. Kovacevic and Pichler fix $\pi(m,n)$ with the values computed from the previous iteration.

From a large LP to an Optimal Transport Problem

The recursive problem (RP) is a Wasserstein Barycenter problem.

Given $t \in \{1, \dots, T\}$ and $n \in \mathcal{N}'_t$, problem (RP) reads as:

$$\left\{\begin{array}{lllll}
\min\limits_{P',\pi} & \alpha_1^n \sum\limits_{i \in m_1+, j \in n+} \pi(i,j|m_1,n)\delta_\iota(i,j) & + \cdots + & \alpha_M^n \sum\limits_{i \in m_M+, j \in n+} \pi(i,j|m_M,n)\delta_\iota(i,j & \\
\text{s.t.} & \sum_{j \in n+} \pi(i,j|m_1,n) & & = P(i|m_1), & (i \in m_1+) \\
& \ddots & \vdots & & \\
& \quad \sum_{j \in n+} \pi(i,j|m_M,n) & & = P(i|m_M), & (i \in m_M+) \\
& \sum_{i \in m_1+} \pi(i,j|m_1,n) & & = P'(j|n), & (j \in n+) \\
& \ddots & \vdots & & \\
& \quad \sum_{i \in m_M+} \pi(i,j|m_M,n) & & = P'(j|n), & (j \in n+) \\
& \sum_{j \in n+} P'(j|n) = 1, \pi(i,j|m_1,n) \geq 0 & \cdots & \pi(i,j|m_M,n) \geq 0, &
\end{array}\right. \tag{WB}$$

This is a Wasserstein barycenter problem, with:

- The right constraints on mass conservation,
- The left constraints on unicity of the barycenter.

The Method of Averaged Marginals

MAM algorithm

Input: Initial plan $\pi = (\pi^1, \ldots, \pi^m)$ and parameter $\rho > 0$
Set $S^m \leftarrow\mid \texttt{supp}(q^m) \mid$, for $m = 1, \ldots, M$
Define $a_m \leftarrow (\frac{1}{S^m})/(\sum_{j=1}^M \frac{1}{S^j})$ and set $p^m \leftarrow \sum_{s=1}^{S^m} \pi^m_{rs}$, $m = 1, \ldots, M$
Set $D^m \leftarrow \alpha_m \, (\delta_\iota(i,j))_{(i,j)\in m+\times n+}$ and set $q^m = (P(i|m))_{i\in m+}$

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**
      $\pi^m_{:s} \leftarrow \texttt{Proj}_{\Delta(q^m_s)}\left(\pi^m_{:s} + 2\frac{p-p^m}{S^m} - \frac{1}{\rho} D^m_{:s}\right) - \frac{p-p^m}{S^m}$
    end for
    $p^m \leftarrow \sum_{s=1}^{S^m} \pi^m_{rs}$ ▷ Update the m^{th} marginal
  end for

end while

The Iterative Bregman Projection

IBP algorithm

Input: Given α_m for $m = 1, \dots, M$, $\lambda > 0$, initialize v^0 and u^0 with an arbitrary positive vector, for example $\mathbb{1}_S$ Initialize p^0, for example $\mathbb{1}_R/R$
Set $D^m \leftarrow \alpha_m \left(\delta_\iota(i,j)\right)_{(i,j)\in m+ \times n+}$ and set $q^m = (P(i|m))_{i \in m+}$
Define $K^m = e^{-\lambda D^m}$ for all $m = 1, \dots, M$
while not converged **do**

▷ Projections onto the constraints

  for m=1,...,M **do**

  $v^{m,k+1} = \frac{q^m}{(K^m)^T u^{m,k}}$

  $u^{m,k+1} = \frac{p^{k+1}}{K^m v^{m,k+1}}$

  end for

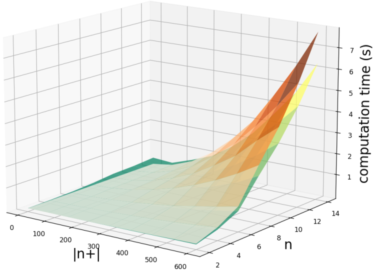
▷ Approximation of the barycenter

  $p^{k+1} = \prod_{m=1}^{M} (K^m v^{m,k+1})^{\alpha_m}$

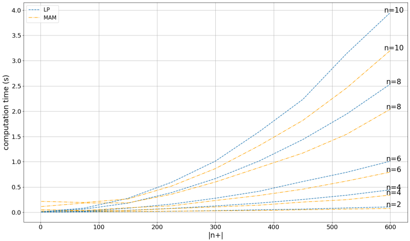
end while

return $\pi^m = \mathrm{diag}(u^m) K^m \mathrm{diag}(v^m)$ for all $m = 1, \dots, M$

Impact of the tree structure

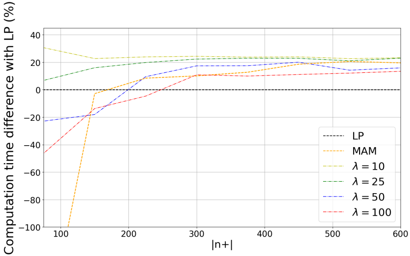


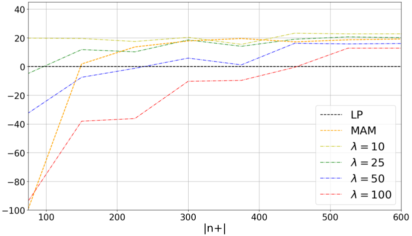
Influence of the tree structure on the computation time of a stage, depending on the method in use: MAM in green and LP in orange.



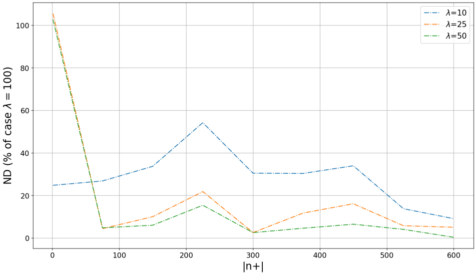
Influence of the tree structure on the computation time for small n.

Impact of the tree structure





Speed comparison with IBP for different λ: A positive time difference means the method is faster than LP. Each curve is obtained by averaging the ND accuracy over $n \in \{2, 4, 6, 8, 10, 12, 14, 16\}$.



Average influence of λ in the precision. Each curve is obtained by averaging the ND accuracy over $n \in \{2, 4, 6, 8, 10, 12, 14, 16\}$.

Impact of the initialization in the tree reduction

- *Kmeans method*, starting from 100 scenarios it creates 25 clusters using the Euclidean norm, and then computes the 25 corresponding barycenters;
- *Fast Forward Selection (FFS) method*, introduced by Heitsch and Römisch[7]. The method iteratively selects scenarios that minimize the Wasserstein distance to the remaining scenarios. At each step, the scenario that best approximates the distribution is added to the reduced set until the desired number of 25 scenarios is reached, ensuring an efficient yet effective reduction.

Scenario set	Filtrations	initial ND	reduced ND
1	**Kmeans**	2757	1219
	FFS	1384	658
2	**Kmeans**	1699	1092
	FFS	1936	896
3	**Kmeans**	1653	832
	FFS	1499	716
4	**Kmeans**	1858	963
	FFS	1161	566
5	**Kmeans**	1968	1054
	FFS	917	540

Comparison of the ND to the original tree before and after tree reduction using different initialization techniques.

- Detailed in the Annals of Operations Research publication.

[7] Computational Optimization and Applications (2003)

Decision Process for Scenario Based Methods

Process 1 Decision process for scenario-based methods

Every day do

From $\text{Cons}^{\text{meas}}$ and PV^{meas} on past ΔT, generate L scenarios

↓

- Select S scenarios with FFS
- Improve the reduced set with KP (optional)

↓

Compute $\boldsymbol{P_b}$ with a scenario-based algorithm (DSP, SDAP, RPHA...)

↓

for $t \in [0{:}00,\ 24{:}00]$

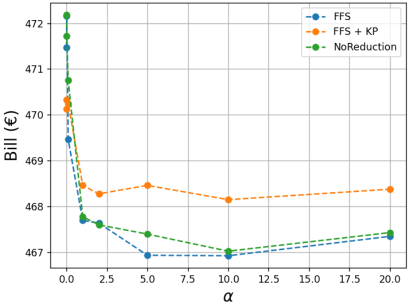
Measure $\text{Cons}^{\text{meas}}(t)$ and $\text{PV}^{\text{meas}}(t)$

↓

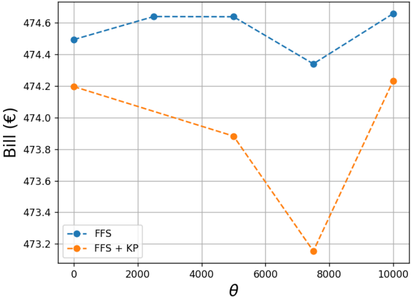
Using the filtration $\mathcal{N}_\Delta$, $\text{Cons}^{\text{meas}}(t)$, and $\text{PV}^{\text{meas}}(t)$ to select the optimal $P_b^{s^*}$ and apply decision $P_b^{s^*}(t+\Delta)$

The control algorithm, where $\text{Cons}^{\text{meas}}$, PV^{meas} denote, respectively, the electric consumption and photovoltaic production measured at the meter.

Cross validations on a 60 days period



Cross validation VSP



Cross validation DRO

Reduction tree and stoch. optim.

- For this specific application the reduction tree via KP was not the most effective approach;
- For other applications it can!
- Also in more general cases, the reduction tree would be more efficient in these cases:
  - When we trust the scenario generator enough, for example when predicting temperature (for a building...);
  - When uncertainties are present in the dynamics of the system.

Performance Ratio

We define four criteria to help explain the differences in model performance, where $P_d := -\min(\rho_d P_b^{\text{meas}}, 0)$ and $P_c := \max(P_b^{\text{meas}}/\rho_c, 0)$:

- **Autoproduction gain ratio:**

$$\texttt{PG} = 100 \times \frac{\int \texttt{min}\{P_d(t), \text{Cons}^{\text{meas}}(t) - \text{PV}^{\text{meas}}(t)\}\mathbb{1}_{\mathcal{C}_1}(t)\text{d}t}{\int \text{Cons}^{\text{meas}}(t)dt}. \quad (4)$$

Where, $\mathcal{C}_1 : \text{PV}^{\text{meas}}(t) < \text{Cons}^{\text{meas}}(t)$.

- **Autoconsumption gain ratio:**

$$\texttt{CG} = 100 \times \frac{\int \texttt{max}\{P_c(t), \text{PV}^{\text{meas}}(t) - \text{Cons}\}\mathbb{1}_{\mathcal{C}_2}(t)dt}{\int \text{PV}^{\text{meas}}(t)dt}.$$

Where, $\mathcal{C}_2 : \text{PV}^{\text{meas}}(t) > \text{Cons}^{\text{meas}}(t)$.

Performance Ratio

- **Discharging error ratio:**

$\texttt{DE} =$

$$100 \times \frac{\int \left(P_d(t) - \left(\mathrm{Cons}^{\mathrm{meas}}(t) - \mathrm{PV}^{\mathrm{meas}}(t)\right)\right) \mathbb{1}_{\mathcal{C}_3}(t) dt}{\int \mathrm{Cons}^{\mathrm{meas}}(t) dt}.$$

Where, $\mathcal{C}_3 : \mathrm{PV}^{\mathrm{meas}}(t) < \mathrm{Cons}^{\mathrm{meas}}(t)$ and $\mathrm{Cons}^{\mathrm{meas}}(t) - \mathrm{PV}^{\mathrm{meas}}(t) < P_d(t)$.

- **Grid charging ratio:**

$\texttt{GC} =$

$$100 \times \frac{\int (P_c(t) - (\mathrm{PV}^{\mathrm{meas}}(t) - \mathrm{Cons}^{\mathrm{meas}}(t))) \mathbb{1}_{\mathcal{C}_4}(t) dt}{\int \mathrm{PV}^{\mathrm{meas}}(t) dt}.$$

Where, $\mathcal{C}_4 : \mathrm{PV}^{\mathrm{meas}}(t) > \mathrm{Cons}^{\mathrm{meas}}(t)$ and $\mathrm{PV}^{\mathrm{meas}}(t) - \mathrm{Cons}^{\mathrm{meas}}(t) < P_c(t)$.

Performance Ratio

Table: Evaluation of models according to four performance criteria. Values in parentheses indicate the percentage difference relative to MPC.

Method	`CG`	`PG`
DSP	1.2900 (49.7%)	1.4437 (12.2%)
VSP	0.9248 (7.3%)	1.3086 (1.7%)
RL	0.9189 (6.6%)	1.1824 (-8.1%)
MPC	0.8620 (0.0%)	1.2867 (0.0%)
DRO	0.8850 (2.7%)	1.2715 (-1.2%)
SP	0.8811 (2.1%)	1.2883 (0.1%)

Method	`DE`	`GC`
DSP	0.0378 (-65.3%)	2.1109 (-9.8%)
VSP	0.0783 (-28.1%)	2.2602 (-3.4%)
RL	0.0649 (-40.5%)	2.0080 (-14.2%)
MPC	0.1090 (0.0%)	2.3400 (0.0%)
DRO	0.1135 (4.1%)	2.2927 (-2.0%)
SP	0.1223 (12.2%)	2.3593 (+0.8%)