Multistage Stochastic Optimization: From Optimal Transport-Based Scenario Tree Reduction to Robust Optimization

Daniel Mimouni[1,2]

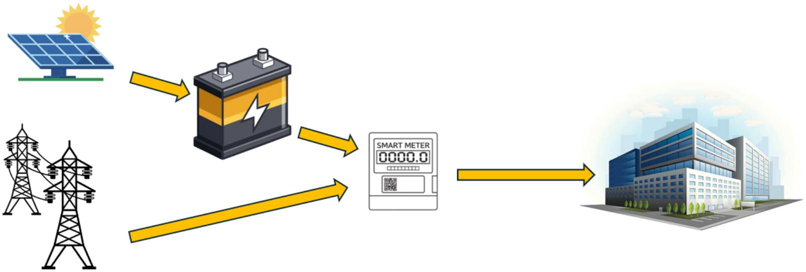
1 **Applied Mathematics Center - CMA, Mines Paris PSL, France**

2 **Applied Mathematics Department, IFP Energies nouvelles**

October 24, 2025

Alois Pichler	Professor, University of Technology, Chemnitz	Reviewer
Franck Iutzeler	Professor, Institut Mathématique de Toulouse	Reviewer
Michel de Lara	Professor, École Nationale des Ponts et Chaussées	Jury
Delphine Bresch-Pietri	Associate Professor, Mines Paris – PSL	Jury
Paul Malisani	Research Engineer, IFPEN	Supervisor
Welington de Oliveira	Associate Professor, Mines Paris – PSL	Thesis Director
Jiamin Zhu	Research Engineer, IFPEN	Co-Supervisor

Optimization System for Energy Management

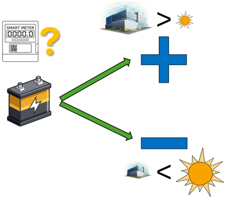


[1]

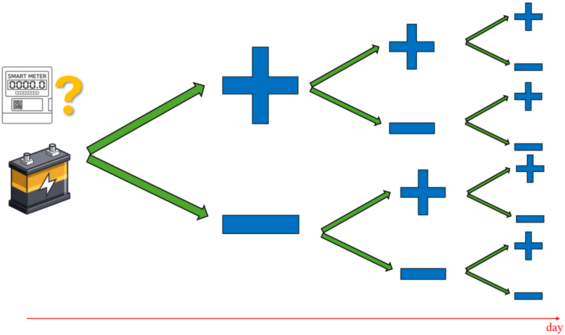
[1] J-M. Jancovici, C. Blain, *Le Monde sans fin - Miracle énergétique et dérive climatique*, 2021

Optimization Model for Energy Management

- two-stage:

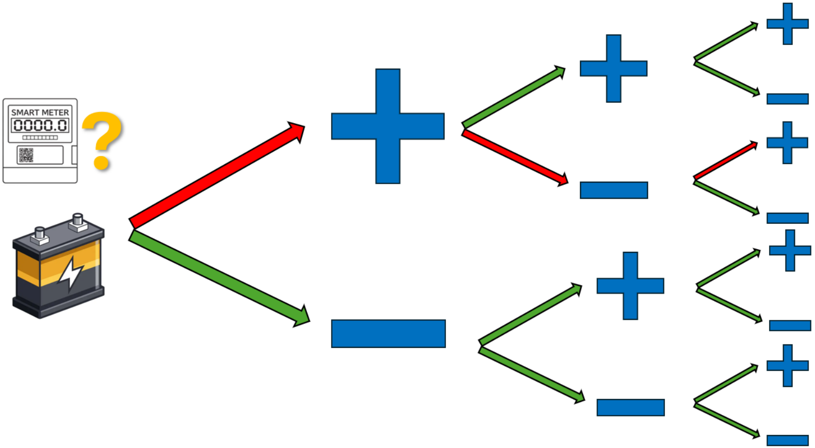


- multi-stage:



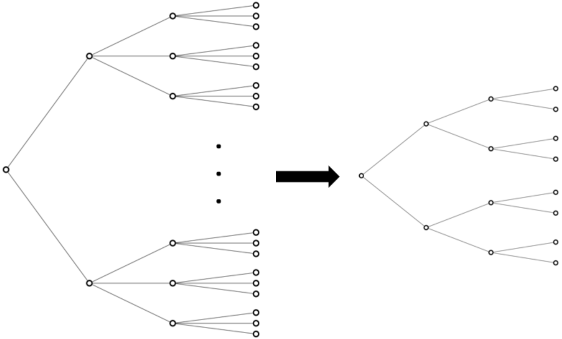
Optimization Model for Energy Management

- Decisions are inter-correlated
- Several methods exist to solve this difficult problem



Dealing with Trees

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



Stakes

- How to reduce trees while keeping good representativity?
  - How to compare trees with different size and structure?

Introduction

I. The scenario tree reduction problem

- D. Mimouni, P. Malisani, J. Zhu, and W. de Oliveira. *Scenario tree reduction via Wasserstein barycenters*. Submitted to **Annals of Operations Research**
  - opensource: `https://github.com/dan-mim/Nested_tree_reduction`

II. The Wasserstein barycenter problem

- D. Mimouni, P. Malisani, J. Zhu, and W. de Oliveira. *Computing Wasserstein barycenters via operator splitting: The method of averaged marginals*. Published in **SIAM Journal on Mathematics of Data Science** 2024
  - opensource: `https://github.com/dan-mim/Computing-Wasserstein-Barycenters-MAM`

III. Solving EMS problems

- D. Mimouni, J. Zhu, W. de Oliveira, and P. Malisani. *A comparative study of multi-stage stochastic optimization approaches for an energy management system*. Submitted to **IEEE Transactions on Control Systems Technology**
  - opensource: `https://github.com/dan-mim/EMS-RL-DRO`

Conclusion

- D. Mimouni, W. de Oliveira, and G. M. Sempere. *On the computation of constrained Wasserstein barycenters*. In press for **Pacific Journal of Optimization** 2025 for a special issue on Rockafellar
  - opensource: `https://github.com/dan-mim/Constrained-Optimal-Transport`

The scenario tree reduction problem

The (Discrete) Wasserstein Distance

We focus on discrete probability 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

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

The Wasserstein distance - discrete setting

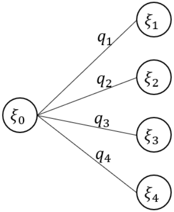
The 2-Wassestein distance between μ and ν is:

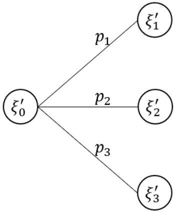
$$W(\mu,\nu) := \left(\min_{\pi \in U(\mu,\nu)} \sum_{r=1}^{R} \sum_{s=1}^{S} \|\xi_r - \zeta_s\|^2 \pi_{rs} \right)^{1/2}$$

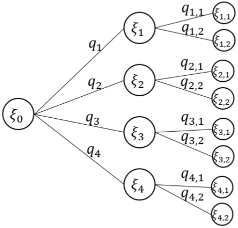
with

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

Distance between processes







- Two stage trees can be represented as discrete probability measures

- Multi stage trees have filtration

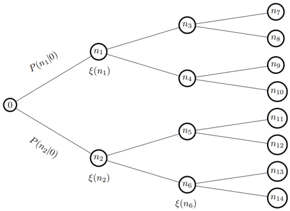
Challenge

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

Distance between processes

Let two T-period scenario trees with set of nodes $\mathcal{N}$, $\mathcal{N}'$:

- The ancestors of $n \in \mathcal{N}$ are $\mathcal{A}(n)$.
- The distance between two nodes at stage t, is $\mathtt{d}_{n_1,n_2}$.
- The transport mass between nodes at stage t, is noted $\pi_{i,j}$ or $\pi(i,j)$.



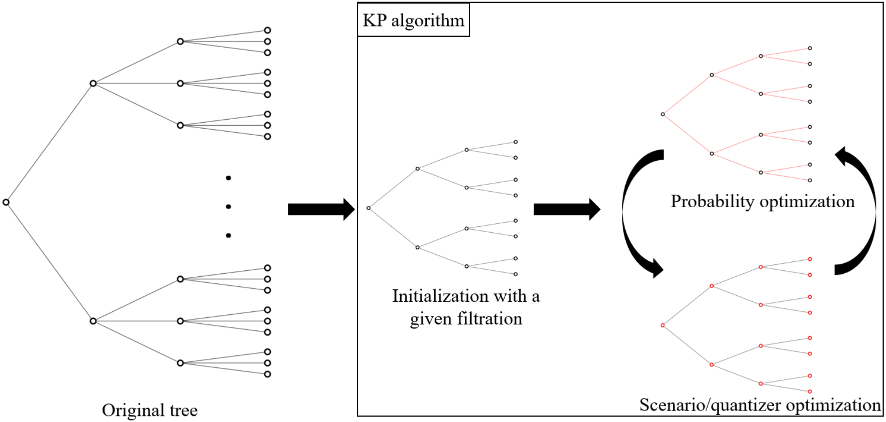
The Nested Distance[2]

The process distance of order 2 between $\mathbf{P}$ and $\mathbf{P}'$ is the square root of the optimal value of the following LP:

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

[2] GC Pflug, A Pichler 2012

Kovacevic and Pichler's algorithm (KP)



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 scenario values are alternatively optimized until convergence.

Probability Optimization

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.

Recursive Problem

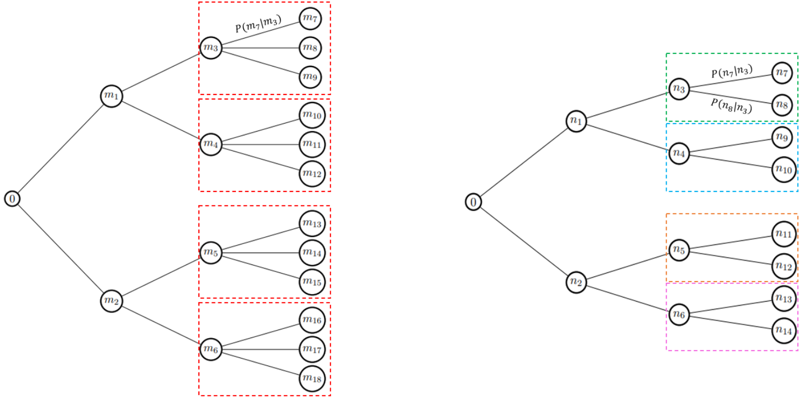
$$\begin{cases} \min\limits_{\pi} & \sum_{m \in \mathcal{N}_t} \pi(m,n) \sum_{i \in m+, j \in n+} \pi(i,j|m,n) \delta_\iota(i,j) \\ \text{s.t.} & \sum_{j \in n+} \pi(i,j|m,n) = P(i|m), \qquad (i \in m+) \\ & \sum_{i \in m+} \pi(i,j|m,n) = \sum_{i \in \tilde{m}+} \pi(i,j|\tilde{m},n), \qquad (j \in n+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

- Computationally expensive due to the solving of potentially large-scale LPs repeatedly. Can be untractable for large-scale scenario trees.

Wasserstein Barycenter (WB) within the KP Algorithm

- In the scenario reduction problem we seek $\mathbf{P}'$ (with given filtration $\mathcal{F}'_t$) that minimizes $\mathrm{ND}(\mathbf{P}, \mathbf{P}')$

The steps of the KP algorithm are Wasserstein Barycenter (WB) problems



(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) red probabilities associated to the boxed subtrees on the left.

The Wasserstein barycenter problem

Motivations

30 artificial images

Barycenters using

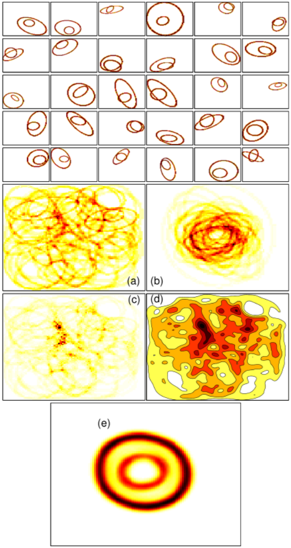
(a) Euclidean distance

(b) Euclidean + re-centering

(c) Jeffrey centroid

(d) RKHS distance

(e) 2-Wasserstein distance: Wasserstein barycenter



[a]

[a] M. Cuturi, A. Doucet, 2014

Discrete Wasserstein Barycenter

Discrete Wassertein Barycenter - WB

A Wassertein barycenter of a set of M discrete probability measures $\nu^m \in \mathcal{P}(\Omega)$, $m = 1, \ldots, M$, is a solution to the following optimization problem

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

where $\sum_{m=1}^{M} \alpha_m = 1$, $\alpha_m > 0$ for $m = 1, \ldots, M$

- The evaluation of the barycenter distance already requires to solve M Wasserstein distance optimization problems!

Discrete Wasserstein Barycenter

Recall $\mu = \sum_{r=1}^{R} p_r \delta_{\xi_r}$

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

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

- This LP scales exponentially in the number M of measures [3]
- 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.

[3] 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 [[4], [5], [6]]
- Other approaches [[7],[8],[9]]

Exact methods

- Methods for computing exact WBs are based on linear programming techniques and thus applicable to applications of moderate sizes [[10],[11]]

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

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

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

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

Observe that we can drop the vector p.

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

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

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

DR algorithm

$$
\begin{cases}
\pi^{k+1} & = & \texttt{Proj}_{\mathcal{B}}(\theta^k) \longrightarrow & \text{explicit} \\
\hat{\pi}^{k+1} & = & \arg \min\limits_{\substack{\pi^m \in \Pi^m \\ m=1,\dots,M}} \sum\limits_{m=1}^{M} \langle d^m, \pi^m \rangle + \frac{\rho}{2} \| \pi - (2\pi^{k+1} - \theta^k) \|^2 \longrightarrow & \text{proj. simplex} \\
\theta^{k+1} & = & \theta^k + \hat{\pi}^{k+1} - \pi^{k+1}
\end{cases}
$$

$\{\pi^k\}$ converges to a solution to the above LP [12]

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

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

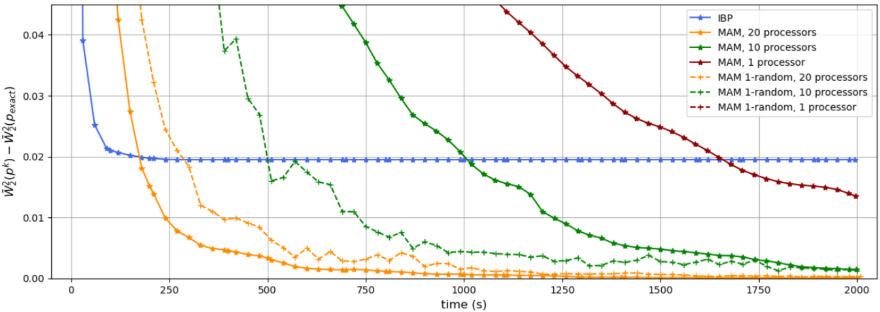
- This algorithm is parallelizable and can run in a randomized manner...
- MAM asymptotically computes Wasserstein barycenter.
- MAM randomized computes almost surely Wasserstein barycenter.

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

We benchmark MAM, randomized MAM, and IBP (Iterative Bregman Projection of [13]) on the MNIST database with $M = 100$ images of 40×40 pixels. LP's dimension: $256\,001\,600$ variables and $320\,000$ constraints

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

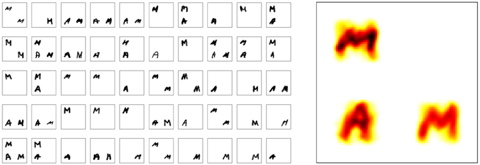
Quantitative comparisons - Fixed support $R = 1\,600$



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 $\bar{W}_2^2(p_{\text{exact}})$ given by the LP. The time step between two points is 30 seconds

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 fixed or free-support settings (optimization on probability and support)
- It can handle convex constraints on the barycenter mass p
- For nonconvex constraints, we extended MAM using the DC setting.



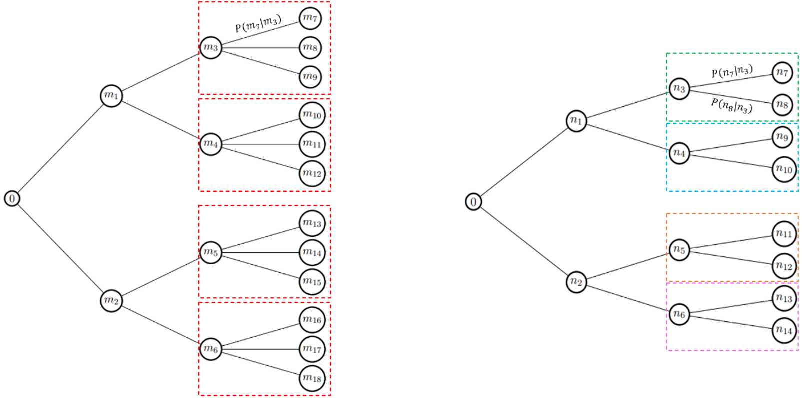
Contributions

- **SIAM Journal on Mathematics of Data Science** 2024 (published): D. Mimouni, P. Malisani, J. Zhu, and W. de Oliveira. *Computing Wasserstein barycenters via operator splitting: The method of averaged marginals.*
    - opensource code: `https://github.com/dan-mim/Computing-Wasserstein-Barycenters-MAM`
- International conferences: EUROPT 2024 (with W. de Oliveira), ISMP 2024
- National events: PGMO 2023, CIROQUO 2023

Wasserstein Barycenter (WB) within the KP Algorithm

- In the scenario reduction problem with seek $\mathbf{P}'$ (with given filtration $\mathcal{F}'_t$) that minimizes $\mathrm{ND}(\mathbf{P}, \mathbf{P}')$

The steps of the KP algorithm are WB problems



(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) red probabilities associated to the boxed subtrees on the left.

Boosted KP Algorithm

Scenario tree reduction via nested distance and Wasserstein barycenters

▷ Step 0: input

Let the original scenario tree $\mathbf{P} = (\Xi^{T+1}, \mathcal{F}, P)$ and a smaller scenario tree $\mathbf{P}'^{0} = (\Xi^{T+1}, \mathcal{F}', P'^{0})$ be given.

Choose a tolerance $\texttt{Tol} > 0$

for $k = 0, 1, 2, \ldots$ **do**

▷ **Step 1:** Improve the scenario values (quantizers)

If $\iota = 2$ use an analytic solution otherwise do a gradiant descent.

▷ **Step 2:** Improve the probabilities

for $t = T-1, \ldots, 0$ **do** ▷ Recursivity

for all $n \in \mathcal{N}_t'$ **do** ▷ Wasserstein barycenters

Set $\alpha_m^n \leftarrow \pi^k(m, n)$, $m \in \mathcal{N}_t$

Use IBP, or MAM to compute $\pi^{k+1}(\cdot, \cdot | \cdot, n)$

end for

end for

▷ **Step 3:** Stopping test

if $\delta_\iota^k(0,0) - \delta_\iota^{k+1}(0,0) \leq \texttt{Tol}$ **then**

Define $P'(n_T) = \sum_{m_T \in \mathcal{N}_T} \pi^{k+1}(m_T, n_T)$ for all $n_T \in \mathcal{N}_T'$ then $P'(n) = \sum_{j \in n+} P'(j)$ for all $n \in \mathcal{N}_t', t \neq T$

Set $\mathrm{ND}_\iota(\mathbf{P}, \mathbf{P}') \leftarrow \delta_\iota^{k+1}(0,0)$

Stop and return with the reduced tree $\mathbf{P}' = (\Xi^{T+1}, \mathcal{F}', P')$ and nested distance $\mathrm{ND}_\iota(\mathbf{P}, \mathbf{P}')$

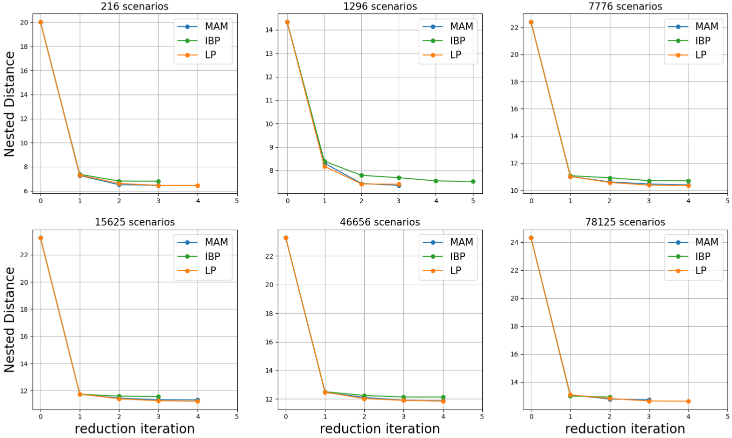
end if

end for

Reduction scenario applications

Scenario tree reduction employing different solvers to compute the WBs:

- A classic LP : KP algorithm + LP,
- Iterative Bregmann Projection (IBP) [14] : KP algorithm + IBP,
- Method of Averaged Marginals (MAM)[15] : KP algorithm + MAM.

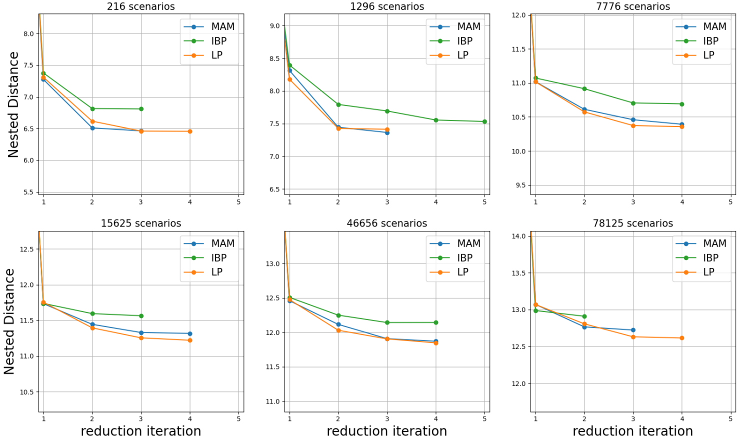


Evolution of the Nested Distance along the reduction iterations for different initial trees.

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

[15] see the work of Mimouni, Malisani, Zhu, de Oliveira

Reduction scenario applications



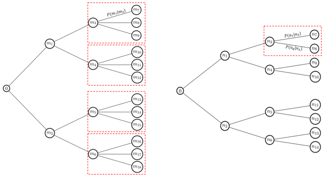
Evolution of the Nested Distance along the reduction iterations for different initial trees with a zoom.

Scenarios	LP	IBP	MAM	MAM 4 processors
216→16	0.17	0.49	2.21	0.56
1296→32	1.54	14.83	18.23	6.28
7776→ 64	74.25	161.19	344.83	124.44
15625→ 128	487.58	323.76	816.46	341.62
46656→ 128	4905	2136	2541	1256
78125→ 256	13797	4334	3458	1635

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

Take-away messages

- New approach to tackle scenario tree reduction
- New easy-to-implement and memory efficient algorithm for reducing scenario trees
- Can leverage parallelization of transport optimal techniques
- Makes more accessible (because more efficient) a technique that keeps maximal information from the initial modelization

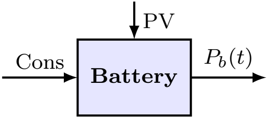


Contribution

- **Annals of Operations Research** 2025 (submitted): D. Mimouni, P. Malisani, J. Zhu, and W. de Oliveira. *Scenario tree reduction via Wasserstein barycenters.*
  - opensource code: `https://github.com/dan-mim/Nested_tree_reduction`
- International conference: ICSP 2025
- National conference: PGMO 2024

Solving EMS problems

Optimization Model for Multistage Energy Management



The effective power demand:

$$\boldsymbol{P_m}(t) = \mathbf{Cons}(t) - \mathbf{PV}(t) + \frac{1}{\rho_c}\max\{P_b(t), 0\} + \rho_d \min\{P_b(t), 0\}$$

Bill

The electricity bill to minimize:

$$J_{t_1:t_2}(P_b, \mathbf{Cons}, \mathbf{PV}) := \int_{t_1}^{t_2} p_r^c(t)\max\{\boldsymbol{P_m}(t), 0\} + p_r^d(t)\min\{\boldsymbol{P_m}(t), 0\}\mathrm{d}t$$

which, in our setting, is a convex function of P_b.

The battery's dynamics are governed by the constraints C:

$$\dot{E}(t) = P_b(t)$$

and the operational constraints of the battery's charging/discharging process are

$$0 < E < 13 \text{ kWh},$$

$$-8 < P_b < 8\rho_c \text{ kW}$$

$$E(t_1) = E(t_2) = E^0$$

Model Predictive Control (MPC)

MPC solves a deterministic optimization problem

- Currently used

MPC

One of the most used model.

- MPC predicts a realization ($\widehat{\text{Cons}}$ and $\widehat{\text{PV}}$) of the random variables **Cons** and **PV**

Solves:

$$\min_{P_b \in C} J_{t:t+\Delta T}(P_b, \widehat{\text{Cons}}, \widehat{\text{PV}}) \tag{MPC}$$

- Easy to implement
- Solution can be of poor quality depending on the accuracy of $\widehat{\text{Cons}}$ and $\widehat{\text{PV}}$

Scenario based models are studied to manage the multistage EMS problem.

SCENARIOS

The following models rely on a finite set of scenarios:

$$\Xi^S_{t_1:t_2} := \left\{ \xi^s := \mathrm{Cons}^s - \mathrm{PV}^s : s = 1, \ldots, S \right\} \quad (2)$$

- Each scenario, ξ^s, is associated to a probability $p_s > 0$, satisfying $\sum_{s=1}^{S} p_s = 1$
- The more scenarios the better representativity but the harder the problem

Deterministic Control for SP

DSP computes a single control policy that minimizes the expected cost across all scenarios

- We look for an optimal control P_b for all scenarios

DSP

$$\inf_{P_b \in \mathrm{C}} \left[\sum_{s=1}^{S} p_s J_{t_1:t_2}(P_b, \mathrm{Cons}^s, \mathrm{PV}^s) \right], \tag{DSP}$$

- Can be solved using standard control methods such as [16]
- DSP is numerically tractable even when using a large number of scenarios

[16] Malisani, P. (2024). Interior point methods in optimal control. ESAIM: Control, Optimisation and Calculus of Variations, 30, 59.

Stochastic Programming (SP)

SP optimizes expected cost over known scenarios

- We look for an optimal control P_b^s for each scenario.
- the controls undergo a non-anticipativity constraint $\mathcal{N}$

SP

$$\inf_{\substack{\boldsymbol{P_b} \in \mathbb{C} \\ \boldsymbol{P_b} \in \mathcal{N}}} \left[\sum_{s=1}^{S} p_s J(\boldsymbol{P_b}(s), \mathrm{Cons}^s, \mathrm{PV}^s) \right] \tag{SP}$$

- p_s is the probability of the scenario s

- Requires estimation of probability distribution.
- A possible algorithm: Progressive Hedging Algorithm (PHA).

Robust Optimization (RO)

RO assumes the worst-case scenario among all the possible outcomes.

RO

Minimization of the cost under worst-case scenario:

$$\inf_{\substack{\boldsymbol{P_b} \in \mathbb{C} \\ \boldsymbol{P_b} \in \mathcal{N}}} \max_{s \in \{1,\ldots,S\}} J(\boldsymbol{P_b}(s), \mathrm{Cons}^s, \mathrm{PV}^s) \tag{RO}$$

- Overly conservative in general.

Distributional Robust Optimization (DRO)

DRO optimizes against worst-case distribution in an ambiguity set.

DRO

$$\inf_{\substack{\boldsymbol{P_b}\in\mathbb{C}\\ \boldsymbol{P_b}\in\mathcal{N}}} \sup_{q\in\mathcal{P}_\theta} \left[\sum_{s=1}^{S} q_s J(\boldsymbol{P_b}(s), \mathrm{Cons}^s, \mathrm{PV}^s)\right] \qquad \text{(DRO)}$$

- The scenarios are fixed but DRO optimize over the worst distribution of weights of the scenarios within an ambiguity set.

Wasserstein-based ambiguity sets.

$$\mathcal{P}_\theta := \left\{ q \in \mathbb{R}_+^S : \sum_s q_s = 1, \mathrm{W}_2\left(\sum_{s=1}^{S} q_s \delta_{\xi^s}, \sum_{l=1}^{L} p_l \delta_{\xi^l}\right) \le \theta \right\}$$

where W_2 is the 2-Wasserstein distance: a distance between probability measures

- θ controls the size of $\mathcal{P}_\theta$,
- large values of θ can lead to RO
- small values of θ can lead to SP

- A possible algorithm for multistage DRO: SDAP[17].

[17] van Ackooij, W. S., and de Oliveira, W. L. (2025). Scenario Decomposition with Alternating Projections. In Methods of Nonsmooth Optimization in Stochastic Programming

Stochastic Programming with Variance Penalization (VSP)

VSP introduces a variance regularization into SP.

VSP

$$\inf_{\substack{\boldsymbol{P_b} \in \mathbb{C} \\ \boldsymbol{P_b} \in \mathcal{N}}} \Bigg[\sum_{s=1}^{S} p_s J_{t_1:t_2}(\boldsymbol{P_b}(s), \mathrm{Cons}^s, \mathrm{PV}^s) + \frac{\alpha}{2} \sum_{s=1}^{S} p_s \|\boldsymbol{P_b}(s) - \textstyle\sum_{s'=1}^{S} p_{s'} \boldsymbol{P_b}(s')\|_{\mathrm{L}^2}^2 \Bigg] \quad \text{(VSP)}$$

- $\alpha = 0$: VSP is equivalent to SP
- $\alpha = \infty$: VSP is equivalent to DSP

- Trade-off parameter α controls robustness.
- Solved via Regularized Progressive Hedging (RPHA)[18].

[18] Malisani, P., Spagnol, A., and Smis-Michel, V. (2024). Robust stochastic optimization via regularized PHA: application to Energy Management Systems.

Reinforcement Learning (RL)

RL learns an optimal behavior by interacting with an environment and receiving costs from these interactions.

- We define a Markov Decision Process as $(\mathcal{T}, \mathcal{S}, \mathcal{A}, \mathbb{P}, c)$

- $\mathcal{T} = \{t_1, t_1 + \Delta, \ldots, t_2\}$ the finite time horizon,
- $\mathcal{S}$ is the state space:
  - $E \in \{0, dE, \cdots, 13 \text{ kWh}\}$ is the discretized energy state of the battery,
  - $\bar{\xi} = \text{Cons} - \text{PV} \in \{-8 \times 2, -16 + d\bar{\xi}, \cdots, 16\rho_c \text{ kW}\}$ is the difference between the electricity demand and the solar production
- $\mathcal{A}$ is the action space where the action P_b is the battery charging power,
- $\mathbb{P}$ is the transition probability of passing from state s to state s' given action a,
- c is the cost function : $c^\tau(s, P_b) = p_r^c(\tau) \max\{P_m^\tau, 0\} + p_r^d(\tau) \min\{P_m^\tau, 0\}$.

Bellman Equation

RL[19] minimizes the Q-function with respect to π

- π is the politic : $P_b^\tau = \pi(\tau, s)$

$$Q^\pi(t, s, P_b) = \mathbb{E}\left[\sum_{\tau=t}^{t_2} c^\tau(\boldsymbol{s}^\tau, \boldsymbol{P}_b^\tau) \mid \boldsymbol{s}^t = s, \boldsymbol{P}_b^t = P_b\right]. \quad (3)$$

- Learns control policy through experience.
- Scenario-free method.

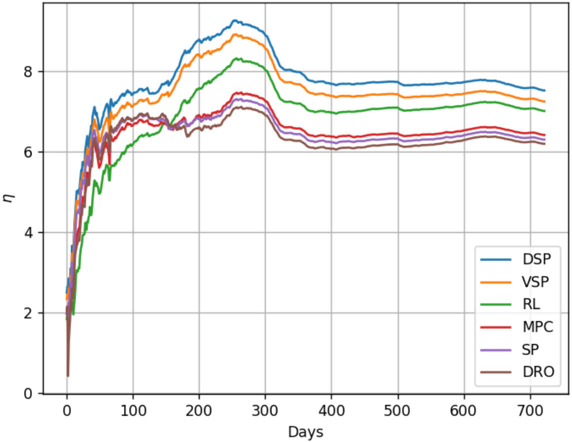
[19] Weber, L., Bušić, A., and Zhu, J. (2023). Reinforcement learning based demand charge minimization using energy storage. In 2023 62nd IEEE CDC

Models to Methods

- MPC
  One prediction
- DSP
  One control for all scenarios
- SP $\rightarrow$ PHA
  Assumed distribution
- DRO $\rightarrow$ SDAP
  Ambiguity set
- VSP $\rightarrow$ RPHA
  Variance penalization
- RL
  Scenario free model

Performance Comparison

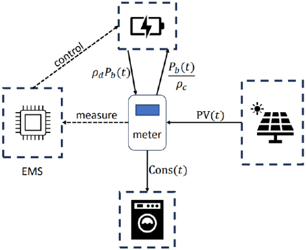
- After cross-validation for θ in DRO and α in VSP, the methods are tested over a 2 year period: 2022-01-22 to 2024-01-22.
- $\eta(\text{Day}) := 100 \times \frac{B-\text{Bill}}{B}$, where $B := \text{Bill}_{\text{no battery}}$



- DSP performs best overall.
- RL is competitive with less tuning.
- DRO probably limited by the number of scenarios it can deal with.

Take-away messages

- All models have been tested in real conditions over a 2 year period,
- DSP is the best model for our EMS problem. DSP is fast to compute and easy to implement.



Contribution

- **IEEE Transactions on Control Systems Technology** (submitted):
  D. Mimouni, J. Zhu, W. de Oliveira, and P. Malisani. *A comparative study of multi-stage stochastic optimization approaches for an energy management system.*
  - *public:* `https://github.com/dan-mim/EMS-RL-DRO`,
  - *industrial:* `https://gitlab.ifpen.fr/R1150/malisanp`
- International conference: ICCOPT 2025

Conclusion

Contributions

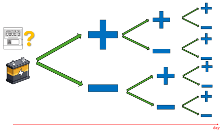
Scenario tree reduction

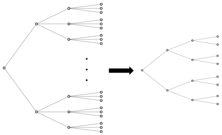
- **Annals of Operations Research** (submitted): D. Mimouni, P. Malisani, J. Zhu, and W. de Oliveira. *Scenario tree reduction via Wasserstein barycenters.*
  - opensource code: `https://github.com/dan-mim/Nested_tree_reduction`
- PGMO 2024 – The Gaspard Monge Program for Optimization, Operations Research and their Interactions with Data Science
- ICSP 2025 – Int. Conf. on Stochastic Programming

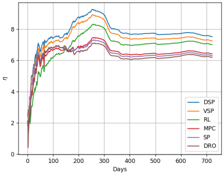
MAM

- **SIAM Journal on Mathematics of Data Science** 2024 (published): D. Mimouni, P. Malisani, J. Zhu, and W. de Oliveira. *Computing Wasserstein barycenters via operator splitting: The method of averaged marginals.*
  - opensource code: `https://github.com/dan-mim/Computing-Wasserstein-Barycenters-MAM`
- CIROQUO 2023
- PGMO 2023
- EUROPT 2024 - with W. de Oliveira
- ISMP 2024 – Int. Symposium on Mathematical Programming

Contributions







EMS

- **IEEE Transactions on Control Systems Technology** (submitted): D. Mimouni, J. Zhu, W. de Oliveira, and P. Malisani. *A comparative study of multi-stage stochastic optimization approaches for an energy management system.*
  - opensource code: `https://github.com/dan-mim/EMS-RL-DRO`
- International conference: ICCOPT 2025 – Int. Conf. on Continuous Optimization

Can we go further ?

- MAM could be applied in the context of reduced-basis methods for PDEs at IFPEN
- Use the reduced tree for tree-based exploration in reinforcement learning
- Include more commonly used models in our comparison for the IFPEN problem

Other published works: Constrained Barycenter

- **Pacific Journal of Optimization**, special issue for Rockafellar (in press): D. Mimouni, W. de Oliveira, and G. M. Sempere. *On the computation of constrained Wasserstein barycenters.*
  - opensource code `https://github.com/dan-mim/Constrained-Optimal-Transport`
- International conference: ICSP 2025 - Int. Conference on Stochastic Programming; with G. Sempere

Thank you!

Contact:

daniel.mimouni1@gmail.com

https://dan-mim.github.io