Distributionally Robust Unit Commitment Models: Theory and Numerical Results

Mathis Azéma, Vincent Leclère, Wim van Ackooij

PSCC 2026

Outline

Outline

Unit Commitment Problem

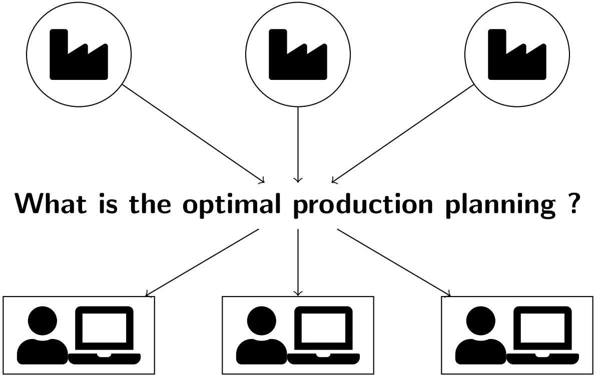
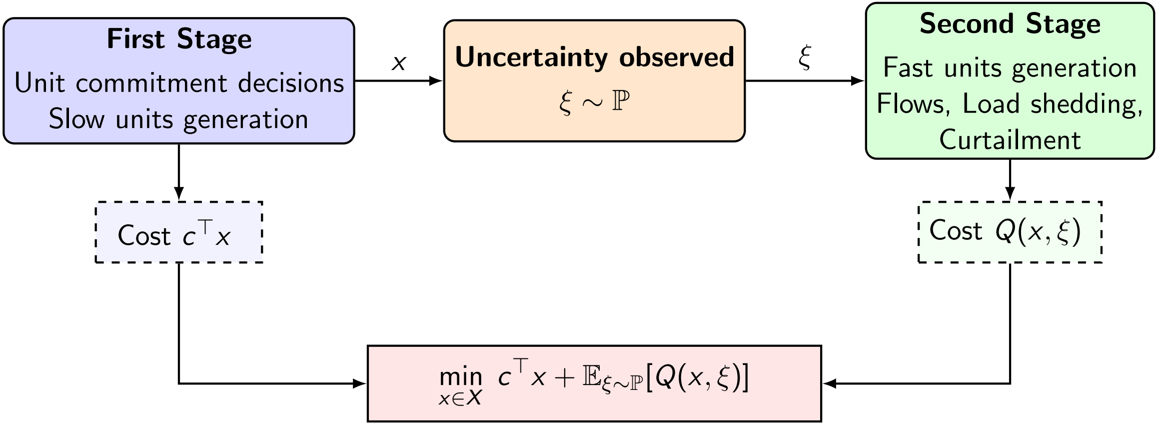


Figure: Unit Commitment

- Time horizon: 24 time steps for a day
- Decisions: on/off status of each unit at each time step, their production levels, and flows in the network.
- $\xi \in \mathbb{R}^T$: vector of demand

Two-Stage UC Formulation



- Problem: $\mathbb{P}$ is unknown, but samples are available.

Goal: Approximate the Recourse Function

Original stochastic problem

$$\min_{x \in X} \; c^\top x + \mathbb{E}_{\xi \sim \mathbb{P}}\left[Q(x,\xi)\right]$$

Approximate problem

$$\min_{x \in X} \; c^\top x + \widehat{Q}(x)$$

Goal

Find an approximation $\widehat{Q}$ such that the optimizer $\hat{x}$ of the approximate problem performs well for the original stochastic problem.

Out-of-sample evaluation

Using a large number M of scenarios, estimate:

$$c^\top \hat{x} + \mathbb{E}_{\xi \sim \mathbb{P}_M}\left[Q(\hat{x},\xi)\right].$$

Contributions

Method	Uses $\mathbb{P}_N$	Bound on the true problem	Tractable
Robust	×	✓	△
Risk-neutral	✓	×	✓
AVaR	✓	×	✓
Moments	△	✓	×
ϕ-divergence	✓	×	✓
Wasserstein	✓	✓	△

- Unified Benders decomposition framework.
- Efficient DCA algorithm for robust and Wasserstein DRO subproblems.
- First extensive numerical comparison on IEEE systems.
- Wasserstein-L2 achieves the best OOS performance.

Benders Decomposition

Approximate problem

$$\min_{x \in X} c^\top x + \widehat{Q}(x)$$

Polyhedral approximation

$$\widehat{Q}(x) = \max_{k \in [K]} \alpha_k^\top x + \beta_k$$

Master Problem

$$\begin{aligned} \min_{x \in X,\, \theta \in \mathbb{R}} \quad & c^\top x + \theta \\ \text{s.t.} \quad & \theta \geq \alpha_k^\top x + \beta_k, \\ & \qquad \forall k = 1, \ldots, j \end{aligned}$$

$\Longrightarrow$

$\Longleftarrow$

Oracle

Compute $\widehat{Q}(x_j^*)$ and the corresponding solution (α_j, β_j). Add the cut $\theta \geq \alpha_j^\top x + \beta_j$ to the master problem.

Heuristic to solve DC problems: DCA algorithm

Consider the DC problem

$$\min_{x \in X} \; f(x) = g(x) - h(x),$$

where g and h are convex functions.

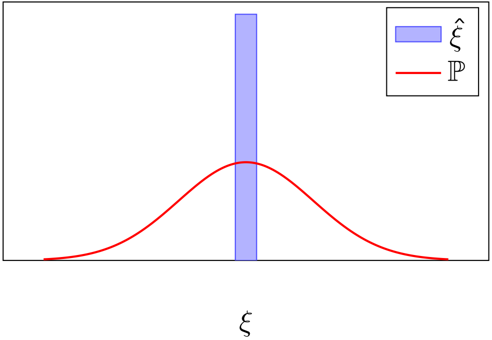
Difference of Convex Algorithm (DCA)

1. Choose an initial point $x^0 \in X$.
2. For $k = 0, 1, \ldots$
   1. Compute a subgradient
   $$y^k \in \partial h(x^k).$$
   2. Solve the convex optimization problem
   $$x^{k+1} \in \arg\min_{x \in X} \Big\{ g(x) - \langle y^k, x \rangle \Big\}.$$
   3. Stop if $\|x^{k+1} - x^k\| < \epsilon$.

Outline

Existing approaches: Deterministic model

$$\underbrace{\min_{x\in X} c^\top x + \mathbb{E}_{\xi\sim\mathbb{P}}[Q(x,\xi)]}_{\text{Real problem}} \quad\rightsquigarrow\quad \underbrace{\min_{x\in X} c^\top x + Q(x,\hat{\xi})}_{\text{Deterministic counterpart}}$$



Advantages

- Easy to understand.
- Fast to solve: Oracle is one LP.

Drawbacks

- Ignores uncertainty.

Existing approaches: Robust model

$$\underbrace{\min_{x \in X} c^\top x + \mathbb{E}_{\xi \sim \mathbb{P}}[Q(x,\xi)]}_{\text{Real problem}} \quad \rightsquigarrow \quad \underbrace{\min_{x \in X} c^\top x + \max_{\xi \in \Xi} Q(x,\xi)}_{\text{robust counterpart}}$$

Budget uncertainty set

$$\Xi = \left\{ \begin{array}{l} \xi \in \mathbb{R}^T : \xi_t = \hat{\xi}_t + \delta_t, \\ |\delta_t| \leq 1, \ \sum_{t=1}^{T} |\delta_t| \leq \Gamma \end{array} \right\}$$

Advantages

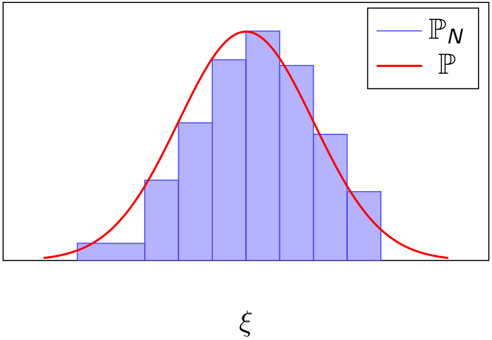
- Easy to understand.
- Robust against worst-case scenarios.

Drawbacks

- Overly conservative.
- Oracle is one DC problem equivalent to one MILP after linearization.

Existing approaches: Risk-neutral model

$$\underbrace{\min_{x \in X} c^\top x + \mathbb{E}_{\xi \sim \mathbb{P}}[Q(x,\xi)]}_{\text{Real problem}} \quad \rightsquigarrow \quad \underbrace{\min_{x \in X} c^\top x + \mathbb{E}_{\xi \sim \mathbb{P}_N}[Q(x,\xi)]}_{\text{risk-neutral counterpart}}$$



Advantages

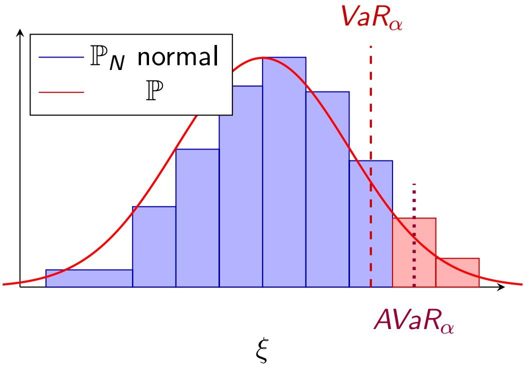
- Easy to understand.
- Fast to solve: Oracle is N LPs.

Drawbacks

- Overfitting.

Existing approaches: AVaR model

$$\underbrace{\min_{x \in X} c^\top x + \mathbb{E}_{\xi \sim \mathbb{P}}[Q(x,\xi)]}_{\text{Real problem}} \quad \rightsquigarrow \quad \underbrace{\min_{x \in X} c^\top x + \mathbb{E}_{\xi \sim \mathbb{P}_N}[Q(x,\xi) \mid Q(x,\xi) > VaR_\alpha(Q(x,\xi))]}_{\text{AVAR counterpart}}$$



Advantages

- Fast to solve: Oracle is N LPs.

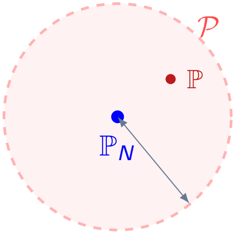
Drawbacks

- Overfitting.

Distributionally Robust Optimization

$$\underbrace{\min_{x \in X} c^\top x + \mathbb{E}_{\xi \sim \mathbb{P}}[Q(x,\xi)]}_{\text{Real problem}} \quad \rightsquigarrow \quad \underbrace{\min_{x \in X} c^\top x + \max_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\xi \sim \mathbb{Q}}[Q(x,\xi)]}_{\text{DRO counterpart}}$$

- Risk-neutral: $\mathcal{P} = \{\mathbb{P}_N\}$.
- Robust: $\mathcal{P} = \{\mathbb{Q} : \text{supp}(\mathbb{Q}) \subseteq \Xi\}$.



Core Challenge: How to construct $\mathcal{P}$ so that it is:

Statistically meaningful (i.e., $\mathbb{P} \in \mathcal{P}$)
$\Longleftrightarrow$ **Computationally tractable**

Existing approaches: Moments

$$\underbrace{\min_{x \in X} c^\top x + \mathbb{E}_{\xi \sim \mathbb{P}}[Q(x,\xi)]}_{\text{Real problem}} \quad \rightsquigarrow \quad \underbrace{\min_{x \in X} c^\top x + \max_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\xi \sim \mathbb{Q}}[Q(x,\xi)]}_{\text{DRO counterpart}}$$

Ambiguity set $\mathcal{P}$

$$\mathcal{P} = \left\{ \begin{array}{l} \mathbb{Q} \in \mathcal{B}(\mathcal{Z}) : \mathbb{E}_{\mathbb{Q}}[\xi] = \hat{\xi}, \\ \mathbb{E}_{\mathbb{Q}}\left[(\xi - \hat{\xi})(\xi - \hat{\xi})^\top\right] \preceq \gamma \hat{\Sigma} \end{array} \right\}$$

Advantages

- Easy to understand.
- $\mathbb{P} \in \mathcal{P}$

Drawbacks

- The master problem is a SDP problem.
- The oracle is non-convex.
- Does not fully leverage the information contained in $\mathbb{P}_N$.

Existing approaches: ϕ-divergence DRO

$$\underbrace{\min_{x\in X} c^\top x + \mathbb{E}_{\xi\sim\mathbb{P}}[Q(x,\xi)]}_{\text{Real problem}} \quad \rightsquigarrow \quad \underbrace{\min_{x\in X} c^\top x + \max_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}_{\xi\sim\mathbb{Q}}[Q(x,\xi)]}_{\text{DRO counterpart}}$$

Ambiguity set $\mathcal{P}$

$$\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) : D_\phi(\mathbb{Q},\mathbb{P}_N) \leq \rho\}, \quad \mathbb{P}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_i}, \quad D_\phi(\mathbb{Q},\mathbb{P}_N) = \mathbb{E}_{\mathbb{P}_N}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}_N}\right)\right]$$

Advantages

- Exploits information from $\mathbb{P}_N$.
- Master problem and oracle convex.

Drawbacks

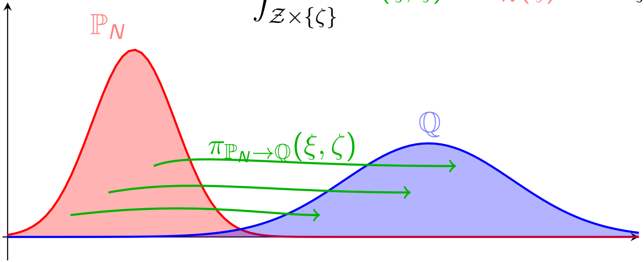
- $\forall \mathbb{Q} \in \mathcal{P}$, $\text{supp}(\mathbb{Q}) \subset \text{supp}(\mathbb{P}_N) = \{\xi_1, \ldots, \xi_N\}$
- $\mathbb{P} \notin \mathcal{P}$

Outline

Wasserstein ambiguity set: definition

$$\mathcal{P} = \{ \; \mathbb{Q} \in \mathcal{B}(\mathcal{Z}) : \mathcal{W}_p(\mathbb{Q}, \mathbb{P}_N) \le \rho^p \; \}$$

$$\begin{aligned} (\mathcal{W}_p(\mathbb{Q}, \mathbb{P}_N))^p = \inf_{\pi \in \mathcal{P}(\mathcal{Z},\mathcal{Z})} \; & \int_{\mathcal{Z}\times\mathcal{Z}} \|\xi - \zeta\|^p d\pi(\xi,\zeta) \\ s.t. \quad & \int_{\{\xi\}\times\mathcal{Z}} d\pi(\xi,\zeta) = \mathbb{Q}(\xi) \qquad \forall \xi \in \mathcal{Z} \\ & \int_{\mathcal{Z}\times\{\zeta\}} d\pi(\xi,\zeta) = \mathbb{P}_N(\zeta) \qquad \forall \zeta \in \mathcal{Z} \end{aligned}$$



Wasserstein ambiguity set: Benders decomposition

$$\underbrace{\min_{x \in X} c^\top x + \mathbb{E}_{\xi \sim \mathbb{P}}[Q(x,\xi)]}_{\text{Real problem}} \quad \rightsquigarrow \quad \underbrace{\min_{x \in X} c^\top x + \max_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\xi \sim \mathbb{Q}}[Q(x,\xi)]}_{\text{DRO counterpart}}$$

Ambiguity set $\mathcal{P}$

$$\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) : \mathcal{W}_p(\mathbb{Q}, \mathbb{P}_N) \leq \rho^p\}, \qquad \mathbb{P}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_i}.$$

Advantages

- Use information in $\mathbb{P}_N$.
- If $\rho = O(1/\sqrt{N})$, then with high probability the error is of order $O(1/N)$

Drawbacks

- Oracle: N DC problems.
- How to choose the norm and the support $\mathcal{Z}$?

Wasserstein ambiguity set: Which norm? Which support $\mathcal{Z}$?

Ambiguity set

$$\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) : \mathcal{W}_p(\mathbb{Q}, \mathbb{P}_N) \leq \rho^p\}, \qquad \mathbb{P}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_i}.$$

Choice	Consequence
$\mathcal{Z} = \mathbb{R}^T$, $\|\cdot\|_1$, $p = 1$	DRO is equivalent to $$\min_{x \in \mathcal{X}} c^\top x + \mathbb{E}_{\mathbb{P}_N}[Q(x, \xi)] + \text{constant}.$$

Wasserstein ambiguity set: Which norm? Which support $\mathcal{Z}$?

Ambiguity set

$$\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) : \mathcal{W}_p(\mathbb{Q}, \mathbb{P}_N) \leq \rho^p\}, \qquad \mathbb{P}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_i}.$$

Choice	Consequence
$\mathcal{Z} = [\underline{\xi}, \overline{\xi}]$, $\|\cdot\|_1$, $p = 1$	Worst-case distributions are supported on $$\{\underline{\xi}_t,\ \xi_{i,t},\ \overline{\xi}_t\},$$ which generally does not match the support of the true distribution.

Wasserstein ambiguity set: Which norm? Which support $\mathcal{Z}$?

Ambiguity set

$$\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) : \mathcal{W}_p(\mathbb{Q}, \mathbb{P}_N) \leq \rho^p\}, \qquad \mathbb{P}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_i}.$$

Choice	**Consequence**
$\mathcal{Z} = \mathbb{R}^T$, $\|\cdot\|_2$, $p = 2$	Leads to meaningful worst-case distributions.

Outline

Experimental Setup

Benchmarks:

- IEEE-6: 3 units, 6 buses, 24 time periods
- IEEE-118: 54 units, 118 buses, 24 time periods

Uncertainty generation:

- ARMA(1,1) process:

$$\xi_t = \hat{\xi}_t(1 + \varepsilon_t),$$
$$\varepsilon_t = 0.67\varepsilon_{t-1} + \eta_t + 0.117\eta_{t-1},\ \eta_t \sim \mathcal{N}(0, 2.9 \times 10^{-4}\sigma),$$

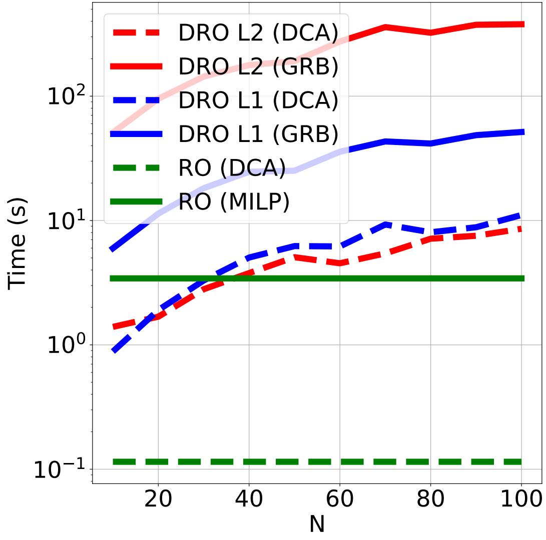
- Three uncertainty levels:

$$\sigma = 1, 2, 3$$

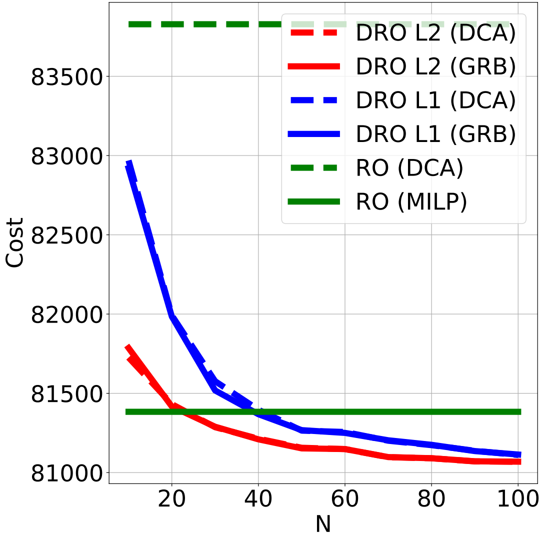
Evaluation:

- Out-of-sample (OOS) cost: expectation on 1000 scenarios.
- 10 independent runs.

Impact of DCA: IEEE-6

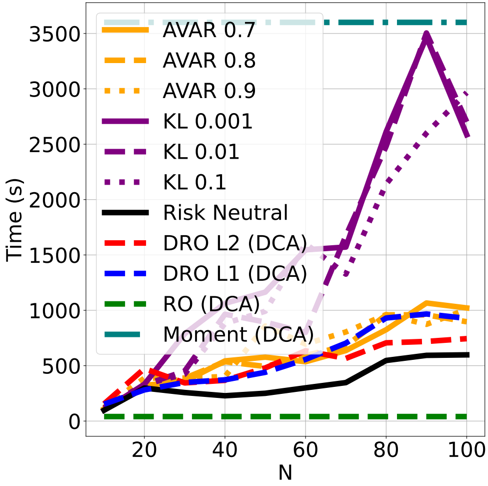


(a) Computational times

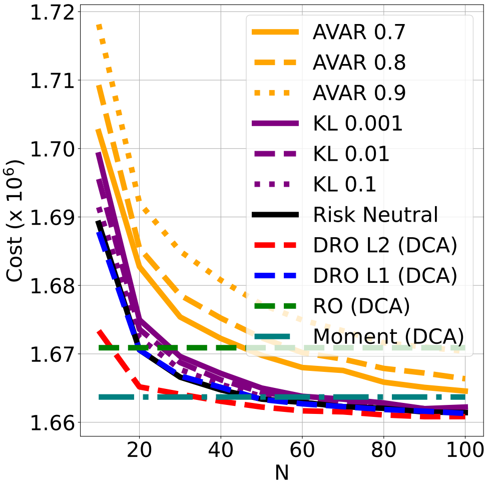


(b) OOS costs

Performance Comparison ($\sigma = 1$)

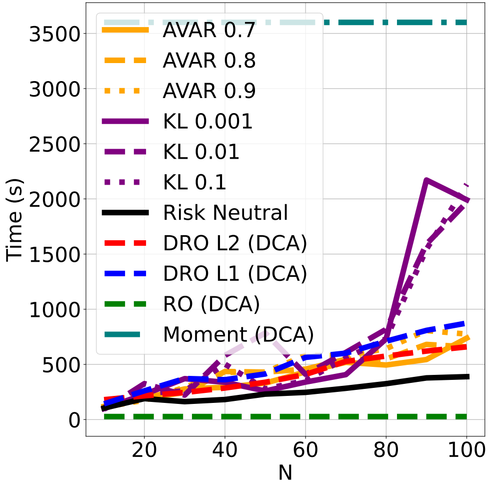


(a) Computational times ($\sigma = 1$)

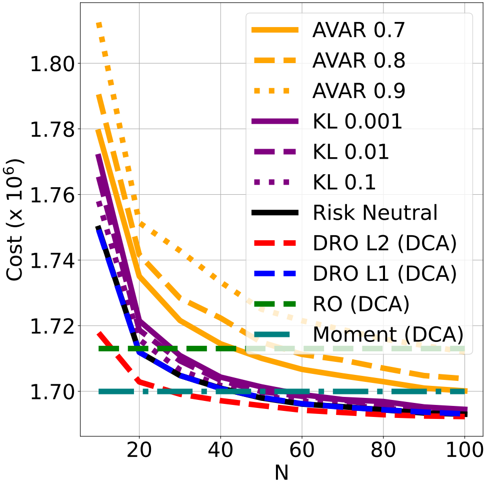


(b) OOS costs ($\sigma = 1$)

Performance Comparison ($\sigma = 2$)

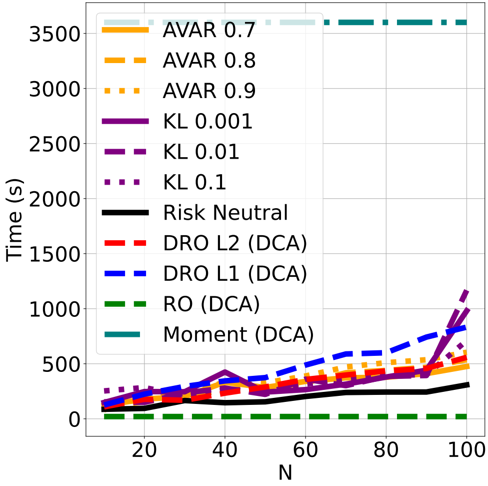


(a) Computational times ($\sigma = 2$)

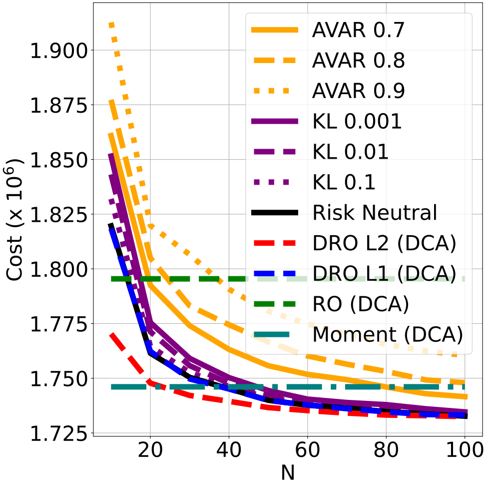


(b) OOS costs ($\sigma = 2$)

Performance Comparison ($\sigma = 3$)



(a) Computational times ($\sigma = 3$)



(b) OOS costs ($\sigma = 3$)

Conclusion

1. Unified Benders Decomposition framework for:
   - Risk-neutral UC
   - AVaR
   - Robust UC
   - DRO with moments
   - DRO with ϕ-divergence
   - DRO with Wasserstein distance
2. First extensive comparison of these approaches on IEEE systems.
3. Efficient solution of non-convex subproblems using DCA algorithm.
4. Wasserstein DRO with L_2 distance provides:
   - Best out-of-sample performance
   - Reasonable computational effort
5. Data and code: `https://github.com/MathisAzema/PSCC-SUC`