Solving a two-stage distributionally robust unit commitment model using Wasserstein distance

Mathis Azéma (ENPC, Paris, France)

Supervisors: Vincent Leclère, Wim Van Ackooij (EDF R&D)

ICSP

July 29th, 2025

CERMICS
École nationale des ponts et chaussées

EDF R&D

Contents

Contents

Contents

Two-stage Unit Commitment problem

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

$$Q(x,\xi) = \max_{\alpha,\beta,\gamma} \alpha^\top \xi + \beta^\top x + \gamma^\top d = \max_{k \in [K]} \alpha_k^\top \xi + \beta_k^\top x + \gamma_k^\top d,$$
$$\text{s.t. } (\alpha, \beta, \gamma) \in \Lambda$$

- $\mathcal{M}$: set of units
- T: time horizon
- $\xi \in \mathbb{R}^T$: demand vector
- x_i: commitment variables.
- y_i: production variables.

$$Q(x,\xi) = \min_{y} C^\top y$$
$$\text{s.t. } T_i x_i + W_i y_i = h_i, \quad \forall i$$
$$\sum_{i \in \mathcal{M}} y_{i,t} = \xi_t, \quad \forall t$$

Distributionally Robust Optimization

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

Interest

Both Robust Optimization and Stochastic Optimization are special cases of DRO.

Stochastic Optimization:

$$\min_{x \in X} c^\top x + \mathbb{E}_{\xi \sim \mathbb{P}_0}[Q(x, \xi)]$$

$$\mathcal{P} = \{\mathbb{P}_0\}$$

Robust Optimization:

$$\min_{x \in X} c^\top x + \max_{\xi \in \mathcal{Z}} Q(x, \xi)$$

$$\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z})\}$$

$\Longrightarrow$ **How should we select $\mathcal{P}$?**

Distributionally Robust Optimization

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

Interest

Both Robust Optimization and Stochastic Optimization are special cases of DRO.

Stochastic Optimization:

$$\min_{x \in X} c^\top x + \mathbb{E}_{\xi \sim \mathbb{P}_0}[Q(x, \xi)]$$

$$\mathcal{P} = \{\mathbb{P}_0\}$$

Robust Optimization:

$$\min_{x \in X} c^\top x + \max_{\xi \in \mathcal{Z}} Q(x, \xi)$$

$$\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z})\}$$

$\Longrightarrow$ **How should we select $\mathcal{P}$?**

Distributionally Robust Optimization

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

Interest

Both Robust Optimization and Stochastic Optimization are special cases of DRO.

Stochastic Optimization:

$$\min_{x\in X} c^\top x + \mathbb{E}_{\xi\sim\mathbb{P}_0}[Q(x,\xi)]$$

$$\mathcal{P} = \{\mathbb{P}_0\}$$

Robust Optimization:

$$\min_{x\in X} c^\top x + \max_{\xi\in\mathcal{Z}} Q(x,\xi)$$

$$\mathcal{P} = \{\mathbb{Q}\in\mathcal{B}(\mathcal{Z})\}$$

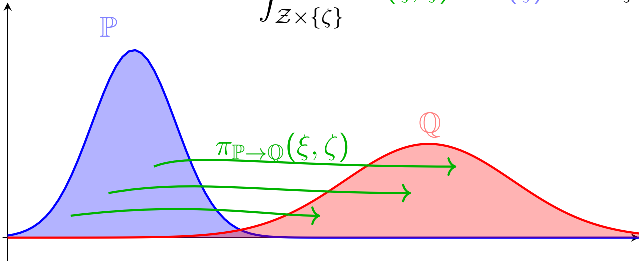
$\Longrightarrow$ **How should we select $\mathcal{P}$?**

Wassertein-distance definition

$$(W_p(\mathbb{Q},\mathbb{P}))^p = \inf_{\pi \in \mathcal{P}(\mathcal{Z},\mathcal{Z})} \int_{\mathcal{Z}\times\mathcal{Z}} \|\xi - \zeta\|^p d\pi(\xi,\zeta)$$

$$s.t. \quad \int_{\{\xi\}\times\mathcal{Z}} d\pi(\xi,\zeta) = \mathbb{Q}(\xi) \qquad \forall \xi \in \mathcal{Z}$$

$$\int_{\mathcal{Z}\times\{\zeta\}} d\pi(\xi,\zeta) = \mathbb{P}(\zeta) \qquad \forall \zeta \in \mathcal{Z}$$



$$\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q},\mathbb{P}) \leq \theta\}$$

Wasserstein distance-based ambiguity sets

$$\min_{x \in X} \quad c^\top x + \max_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\xi \sim \mathbb{Q}}[Q(x,\xi)] \qquad (\mathcal{D})$$

Idea: $\mathcal{P}$ has to include distributions "close" to the empirical one.

$$\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$$

- $\mathcal{Z}$: Support of the distributions.
- W_p: Wasserstein distance of order p defined with the cost function $c(\xi,\zeta) = \|\xi - \zeta\|$.
- $\mathbb{P}_0 = \frac{1}{N}\sum_{i \in [N]} \delta_{\zeta_i}$: empirical distribution

Wasserstein distance-based ambiguity sets

$$\min_{x \in X} \quad c^\top x + \max_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\xi \sim \mathbb{Q}}[Q(x, \xi)] \qquad (\mathcal{D})$$

Idea: $\mathcal{P}$ has to include distributions "close" to the empirical one.

$$\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$$

- $\mathcal{Z}$: Support of the distributions.
- W_p: Wasserstein distance of order p defined with the cost function $c(\xi, \zeta) = \|\xi - \zeta\|$.
- $\mathbb{P}_0 = \frac{1}{N} \sum_{i \in [N]} \delta_{\zeta_i}$: empirical distribution

Stochastic Optimization:

$$\min_{x \in X} c^\top x + \mathbb{E}_{\xi \sim \mathbb{P}_0}[Q(x, \xi)]$$

$$\theta = 0$$

Robust Optimization:

$$\min_{x \in X} c^\top x + \max_{\xi \in \mathcal{Z}} Q(x, \xi)$$

$$\theta = +\infty$$

Wasserstein distance-based ambiguity sets

$$\min_{x \in X} \quad c^\top x + \max_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\xi \sim \mathbb{Q}}[Q(x,\xi)] \qquad (\mathcal{D})$$

Idea: $\mathcal{P}$ has to include distributions "close" to the empirical one.

$$\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$$

- $\mathcal{Z}$: Support of the distributions.
- W_p: Wasserstein distance of order p defined with the cost function $c(\xi,\zeta) = \|\xi - \zeta\|$.
- $\mathbb{P}_0 = \frac{1}{N}\sum_{i\in[N]} \delta_{\zeta_i}$: empirical distribution

Reformulation of problem ($\mathcal{D}$) (Gao and Kleywegt 2016)

$$\min_{x \in X, \lambda \geq 0} \quad c^\top x + \lambda\theta^p + \frac{1}{N}\sum_{i\in[N]} \max_{\xi \in \mathcal{Z}} \left(Q(x,\xi) - \lambda\|\xi - \zeta_i\|^p\right)$$

Benders' decomposition

$$\min_{x \in X, \lambda \geq 0} c^\top x + \lambda \theta^p + \frac{1}{N} \sum_{i \in [N]} \max_{\xi \in \mathcal{Z}} \left(Q(x, \xi) - \lambda \|\xi - \zeta_i\|^p \right)$$

Benders' decomposition

$$\min_{x\in X,\lambda\geq 0,z} c^\top x + \lambda\theta^p + \frac{1}{N}\sum_{i\in[N]} z_i$$

$$\text{s.t.} \quad z_i \geq \max_{\xi\in\mathcal{Z}} \left(Q(x,\xi) - \lambda\|\xi - \zeta_i\|^p\right), \forall i$$

Benders' decomposition

$$\min_{x\in X,\lambda\geq 0,z} c^\top x + \lambda\theta^p + \frac{1}{N}\sum_{i\in[N]} z_i$$

$$\text{s.t.} \quad z_i \geq \max_{\xi\in\mathcal{Z}} \left(Q(x,\xi) - \lambda\|\xi-\zeta_i\|^p\right), \ \forall i$$

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

Benders' decomposition

$$\min_{x \in X, \lambda \geq 0, z} c^\top x + \lambda \theta^p + \frac{1}{N} \sum_{i \in [N]} z_i$$

$$\text{s.t.} \quad z_i \geq \max_{\xi \in \mathcal{Z}, k \in [K]} \left(\alpha_k^\top \xi + \beta_k^\top x + \gamma_k^\top d - \lambda \|\xi - \zeta_i\|^p \right), \forall i$$

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

Benders' decomposition

$$\min_{x \in X, \lambda \geq 0, z} c^\top x + \lambda \theta^p + \frac{1}{N} \sum_{i \in [N]} z_i$$

$$\text{s.t.} \quad z_i \geq \max_{\xi \in \mathcal{Z}} \left(\alpha_k^\top \xi + \beta_k^\top x + \gamma_k^\top d - \lambda \|\xi - \zeta_i\|^p \right), \forall i, \forall k$$

Benders' decomposition

$$\min_{x \in X, \lambda \geq 0, z} c^\top x + \lambda \theta^p + \frac{1}{N} \sum_{i \in [N]} z_i$$

$$\text{s.t.} \quad z_i \geq \beta_k^\top x + \gamma_k^\top d + \max_{\xi \in \mathcal{Z}} \left(\alpha_k^\top \xi - \lambda \|\xi - \zeta_i\|^p \right), \forall i, \forall k$$

Benders' decomposition

$$\min_{x \in X, \lambda \geq 0, z} c^\top x + \lambda \theta^p + \frac{1}{N} \sum_{i \in [N]} z_i$$

$$\text{s.t.} \quad z_i \geq \beta_k^\top x + \gamma_k^\top d + \underbrace{\max_{\xi \in \mathcal{Z}} \left(\alpha_k^\top \xi - \lambda \|\xi - \zeta_i\|^p \right)}_{f_i(\alpha_k, \lambda)}, \; \forall i, \; \forall k$$

Benders' decomposition

$$\min_{x\in X,\lambda\geq 0,z} c^\top x + \lambda\theta^p + \frac{1}{N}\sum_{i\in[N]} z_i$$

$$\text{s.t.}\quad z_i \geq \beta_k^\top x + \gamma_k^\top d + \underbrace{\max_{\xi\in\mathcal{Z}}\left(\alpha_k^\top \xi - \lambda\|\xi - \zeta_i\|^p\right)}_{f_i(\alpha_k,\lambda)},\ \forall i,\ \forall k$$

Proposition

The function f_i is convex.

Subproblem

$$\max_{\alpha,\beta,\gamma} \beta^\top x^* + \gamma^\top d + f_i(\alpha,\lambda^*)$$
$$\text{s.t.}\ (\alpha,\beta,\gamma)\in\Lambda$$

New constraint

$$z_i \geq \beta_k^\top x^* + \gamma_k^\top d + f_i(\alpha_k,\lambda^*) + \partial f_i(\alpha_k,\lambda^*)(\lambda-\lambda^*)$$

Contents

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 1, \|.\| = \|.\|_1, \mathcal{Z} = \mathbb{R}^T}$$

$$f_i(\alpha, \lambda) = \max_{\xi \in \mathbb{R}^T} \alpha^\top \xi - \lambda \|\xi - \zeta_i\|_1$$
$$= \alpha^\top \zeta_i + \mathbb{I}_{\|\alpha\|_\infty \leq \lambda}$$

Master Problem

$$\min_{x \in X, \lambda \geq 0, z} c^\top x + \lambda \theta^p + \frac{1}{N} \sum_{i \in [N]} z_i$$
$$\text{s.t.} \quad z_i \geq \beta_k^\top x + \gamma_k^\top d + f_i(\alpha_k, \lambda), \forall i, \forall k$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 1, \|.\| = \|.\|_1, \mathcal{Z} = \mathbb{R}^T}$$

$$\begin{aligned} f_i(\alpha, \lambda) &= \max_{\xi \in \mathbb{R}^T} \alpha^\top \xi - \lambda \|\xi - \zeta_i\|_1 \\ &= \alpha^\top \zeta_i + \mathbb{I}_{\|\alpha\|_\infty \leq \lambda} \end{aligned}$$

Master Problem

$$\min_{x \in X, \lambda \geq 0, z} c^\top x + \lambda \theta^p + \frac{1}{N} \sum_{i \in [N]} z_i$$

$$\text{s.t.} \quad z_i \geq \beta_k^\top x + \gamma_k^\top d + \alpha_k^\top \zeta_i + \mathbb{I}_{\|\alpha_k\|_\infty \leq \lambda}, \forall i, \forall k$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 1, \|.\| = \|.\|_1, \mathcal{Z} = \mathbb{R}^T}$$

$$\begin{aligned} f_i(\alpha, \lambda) &= \max_{\xi \in \mathbb{R}^T} \alpha^\top \xi - \lambda \|\xi - \zeta_i\|_1 \\ &= \alpha^\top \zeta_i + \mathbb{I}_{\|\alpha\|_\infty \leq \lambda} \end{aligned}$$

Master Problem

$$\begin{aligned} \min_{x \in X, \lambda \geq 0, z} \quad & c^\top x + \lambda \theta^p + \frac{1}{N} \sum_{i \in [N]} z_i \\ \text{s.t.} \quad & z_i \geq \beta_k^\top x + \gamma_k^\top d + \alpha_k^\top \zeta_i, \ \forall i, \ \forall k \\ & \|\alpha_k\|_\infty \leq \lambda, \ \forall k \end{aligned}$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 1, \|.\| = \|.\|_1, \mathcal{Z} = \mathbb{R}^T}$$

$$f_i(\alpha, \lambda) = \max_{\xi \in \mathbb{R}^T} \alpha^\top \xi - \lambda \|\xi - \zeta_i\|_1$$
$$= \alpha^\top \zeta_i + \mathbb{I}_{\|\alpha\|_\infty \leq \lambda}$$

Master Problem

$$\min_{x \in X, z} c^\top x + \theta^p \max_{k \in K}(\|\alpha_k\|_\infty) + \frac{1}{N} \sum_{i \in [N]} z_i$$

$$\text{s.t.} \quad z_i \geq \beta_k^\top x + \gamma_k^\top d + \alpha_k^\top \zeta_i, \forall i, \forall k$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 1, \|.\| = \|.\|_1, \mathcal{Z} = \mathbb{R}^T}$$

$$\begin{aligned} f_i(\alpha, \lambda) &= \max_{\xi \in \mathbb{R}^T} \alpha^\top \xi - \lambda \|\xi - \zeta_i\|_1 \\ &= \alpha^\top \zeta_i + \mathbb{I}_{\|\alpha\|_\infty \leq \lambda} \end{aligned}$$

Master Problem

$$\min_{x \in X, z} c^\top x + \frac{1}{N} \sum_{i \in [N]} z_i$$

$$\text{s.t.} \quad z_i \geq \beta_k^\top x + \gamma_k^\top d + \alpha_k^\top \zeta_i, \forall i, \forall k$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \le \theta\}$

$$\boldsymbol{p = 1, \|.\| = \|.\|_1, \mathcal{Z} = \mathbb{R}^T}$$

$$\begin{aligned} f_i(\alpha, \lambda) &= \max_{\xi \in \mathbb{R}^T} \alpha^\top \xi - \lambda \|\xi - \zeta_i\|_1 \\ &= \alpha^\top \zeta_i + \mathbb{I}_{\|\alpha\|_\infty \le \lambda} \end{aligned}$$

Master Problem

$$\min_{x \in X, z} c^\top x + \frac{1}{N} \sum_{i \in [N]} z_i \quad \rightarrow \text{No radius of DRO ball!}$$

$$\text{s.t.} \quad z_i \ge \beta_k^\top x + \gamma_k^\top d + \alpha_k^\top \zeta_i, \; \forall i, \; \forall k$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 1, \|.\| = \|.\|_1, \mathcal{Z} = [\underline{\xi}, \bar{\xi}]}$$

$$f_i(\alpha, \lambda) = \max_{\xi} \alpha^\top \xi - \lambda \|\xi - \zeta_i\|_1$$
$$\text{s.t. } \xi \in [\underline{\xi}, \bar{\xi}]$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 1, \|.\| = \|.\|_1, \mathcal{Z} = [\underline{\xi}, \bar{\xi}]}$$

$$f_i(\alpha, \lambda) = \max_{\xi} \alpha^\top \xi - \lambda 1^\top u$$
$$\text{s.t. } \xi \in [\underline{\xi}, \bar{\xi}],\ u \geq \xi - \zeta_i,\ u \geq \zeta_i - \xi$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 1, \|.\| = \|.\|_1, \mathcal{Z} = [\underline{\xi}, \bar{\xi}]}$$

$$f_i(\alpha, \lambda) = \max_{\xi} \alpha^\top \xi - \lambda 1^\top u$$
$$\text{s.t. } \xi \in [\underline{\xi}, \bar{\xi}],\ u \geq \xi - \zeta_i,\ u \geq \zeta_i - \xi$$

Subproblem

$$\max_{\alpha,\beta,\gamma} \beta^\top x^* + \gamma^\top d + f_i(\alpha, \lambda^*)$$
$$\text{s.t. } (\alpha, \beta, \gamma) \in \Lambda$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 1, \|.\| = \|.\|_1, \mathcal{Z} = [\underline{\xi}, \bar{\xi}]}$$

$$f_i(\alpha, \lambda) = \max_{\xi} \alpha^\top \xi - \lambda 1^\top u$$
$$\text{s.t. } \xi \in [\underline{\xi}, \bar{\xi}],\ u \geq \xi - \zeta_i,\ u \geq \zeta_i - \xi$$

Subproblem

$$\max_{\alpha,\beta,\gamma,\xi,u} \beta^\top x^* + \gamma^\top d + \alpha^\top \xi - \lambda 1^\top u$$
$$\text{s.t. } (\alpha, \beta, \gamma) \in \Lambda$$
$$\xi \in [\underline{\xi}, \bar{\xi}]$$
$$u \geq \xi - \zeta_i$$
$$u \geq \zeta_i - \xi$$

→ Linearization to obtain a MILP!

Gamboa et al. (2021)

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 2, \|.\| = \|.\|_2, \mathcal{Z} = \mathbb{R}^T}$$

$$f_i(\alpha, \lambda) = \max_{\xi \in \mathbb{R}^T} \alpha^\top \xi - \lambda \|\xi - \zeta_i\|_2^2$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 2, \|.\| = \|.\|_2, \mathcal{Z} = \mathbb{R}^T}$$

$$f_i(\alpha, \lambda) = \alpha^\top \zeta_i + \frac{\|\alpha\|_2^2}{4\lambda}$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 2, \|.\| = \|.\|_2, \mathcal{Z} = \mathbb{R}^T}$$

$$f_i(\alpha, \lambda) = \alpha^\top \zeta_i + \frac{\|\alpha\|_2^2}{4\lambda}$$

Master Problem

$$\min_{x \in X, z, \lambda \geq 0} \quad c^\top x + \lambda\theta^2 + \frac{1}{N}\sum_{i \in [N]} z_i$$

$$\text{s.t.} \quad z_i \geq \beta_k^\top x + \gamma_k^\top d + f_i(\alpha, \lambda^*), \quad \forall i, \forall k$$

Subproblem

$$\max_{\alpha, \beta, \gamma} \beta^\top x^* + \gamma^\top d + f_i(\alpha, \lambda^*)$$

$$\text{s.t.} \, (\alpha, \beta, \gamma) \in \Lambda$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 2, \|.\| = \|.\|_2, \mathcal{Z} = \mathbb{R}^T}$$

$$f_i(\alpha, \lambda) = \alpha^\top \zeta_i + \frac{\|\alpha\|_2^2}{4\lambda}$$

Master Problem

$$\min_{x \in X, z, \lambda \geq 0} \quad c^\top x + \lambda\theta^2 + \frac{1}{N}\sum_{i \in [N]} z_i$$

$$\text{s.t.} \quad z_i \geq \beta_k^\top x + \gamma_k^\top d + \alpha_k^\top \zeta_i + \frac{\|\alpha_k\|_2^2}{4\lambda}, \quad \forall i, \forall k$$

Subproblem

$$\max_{\alpha, \beta, \gamma} \beta^\top x^* + \gamma^\top d + \alpha^\top \zeta_i + \frac{\|\alpha\|_2^2}{4\lambda^*}$$

$$\text{s.t.} \, (\alpha, \beta, \gamma) \in \Lambda$$

Examples: $\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\}$

$$\boldsymbol{p = 2, \|.\| = \|.\|_2, \mathcal{Z} = \mathbb{R}^T}$$

$$f_i(\alpha, \lambda) = \alpha^\top \zeta_i + \frac{\|\alpha\|_2^2}{4\lambda}$$

Master Problem

$$\begin{aligned} \min_{x \in X, z, \lambda \geq 0} \quad & c^\top x + \frac{\theta^2}{\lambda} + \frac{1}{N} \sum_{i \in [N]} z_i \\ \text{s.t.} \quad & z_i \geq \beta_k^\top x + \gamma_k^\top d + \alpha_k^\top \zeta_i + \frac{\lambda \|\alpha_k\|_2^2}{4}, \quad \forall i, \forall k \end{aligned}$$

Subproblem

$$\max_{\alpha, \beta, \gamma} \beta^\top x^* + \gamma^\top d + \alpha^\top \zeta_i + \frac{\lambda^* \|\alpha\|_2^2}{4}$$

$$\text{s.t.} \ (\alpha, \beta, \gamma) \in \Lambda$$

Comparison Wasserstein distances

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

Definition (Worst-case Distribution)

A worst-case distribution $\mathbb{P}^*$ is a distribution in $\mathcal{P}$ satisfying:

$$\max_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\xi \sim \mathbb{Q}}[Q(x, \xi)] = \mathbb{E}_{\xi \sim \mathbb{P}^*}[Q(x, \xi)]$$

- It is useful to analyze this to identify which constraints to add to $\mathcal{P}$ to improve the model.

Comparison Wasserstein distances

$$\min_{x \in X} \quad c^\top x + \max_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\xi \sim \mathbb{Q}}[Q(x,\xi)] \qquad Q(x,\xi) = \max_{k \in [K]} \alpha_k^\top \xi + \beta_k^\top x + \gamma_k^\top d$$

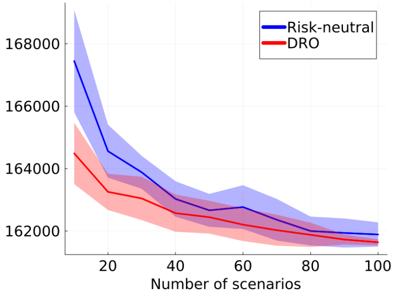
$$\mathcal{P} = \{\mathbb{Q} \in \mathcal{B}(\mathcal{Z}) \mid W_p(\mathbb{Q}, \mathbb{P}_0) \leq \theta\} \quad [\lambda] \qquad \mathbb{P}_0 = \frac{1}{N} \sum_{i \in [N]} \delta_{\zeta_i}$$

p	Norm	$\mathcal{Z}$	$\mathbb{P}^*$	$\text{supp}(\mathbb{P}^*)$
1	$\|.\|_1$	$\mathbb{R}^T$	$\nexists$	-
1	$\|.\|_1$	$[\underline{\xi}, \bar{\xi}]$	$\exists$	$\xi_t \in \{\zeta_{i,t}, \underline{\xi}_t, \bar{\xi}_t\}$
2	$\|.\|_2$	$\mathbb{R}^T$	$\exists$	$\xi \in \bigcup_{k \in [K]} \{\zeta_i + \frac{\lambda^* \alpha_k}{2}\}$

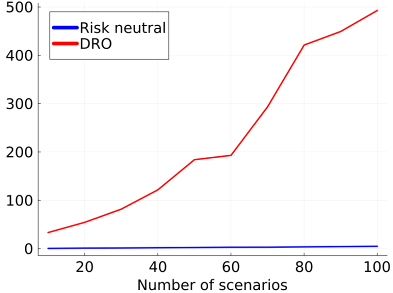
- It is the only one with a worst-case distribution whose support is unpredictable (depends on λ^*).

Contents

Out of sample Costs



(a) Average costs out-of-sample



(b) Computational time (s)

Comparison risk-neutral/ DRO subproblems

Time spent in the subproblems

Subproblem Risk-Neutral for each scenario i:

$$\max_{\alpha,\beta,\gamma} \quad \alpha^\top \zeta_i + \beta^\top x^* + \gamma$$
$$s.t. \quad (\alpha,\beta,\gamma) \in \Lambda$$

↓ LP

↓ $\approx 0.002s$

Subproblem DRO for each scenario i:

$$\max_{\alpha,\beta,\gamma} \quad \alpha^\top \zeta_i + \beta^\top x^* + \gamma + \boldsymbol{f_i(\alpha,\lambda^*)}$$
$$s.t. \quad (\alpha,\beta,\gamma) \in \Lambda$$

↓ Convex maximization problem (NP-Hard)

↓ $\approx 0.5s$

- If we have 100 iterations and 100 scenarios, we have to solve 10,000 subproblems.
- **Idea:** Use non-optimal dual solutions!

Comparison risk-neutral/ DRO subproblems

Time spent in the subproblems

Subproblem Risk-Neutral for each scenario i**:**

$$\max_{\alpha,\beta,\gamma} \quad \alpha^\top \zeta_i + \beta^\top x^* + \gamma$$
$$s.t. \quad (\alpha,\beta,\gamma) \in \Lambda$$

↓ LP ↓ $\approx 0.002s$

Subproblem DRO for each scenario i**:**

$$\max_{\alpha,\beta,\gamma} \quad \alpha^\top \zeta_i + \beta^\top x^* + \gamma + \boldsymbol{f_i(\alpha,\lambda^*)}$$
$$s.t. \quad (\alpha,\beta,\gamma) \in \Lambda$$

↓ Convex maximization problem (NP-Hard) ↓ $\approx 0.5s$

- If we have 100 iterations and 100 scenarios, we have to solve 10,000 subproblems.
- **Idea:** Use non-optimal dual solutions!

Contents

Contents

Notations

Non convex subproblem: ("Hard" to solve)

$$\max_{\alpha,\beta,\gamma} \beta^\top x^K + \gamma^\top d + f_i(\alpha, \lambda^K)$$
$$\text{s.t. } (\alpha, \beta, \gamma) \in \Lambda$$

Optimal solution: (α^*, β^*) $\rightarrow$ Optimal cut

Idea

- Every solution (α, β) of the subproblem gives a valid cut for the master problem.
- It is not necessary to solve the subproblem to optimality at each iteration.

Notations

Non convex subproblem: ("Hard" to solve)

$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

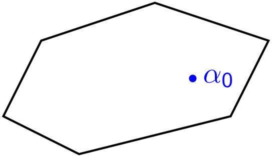
Optimal solution: (α^*, β^*) $\rightarrow$ Optimal cut

Idea

- Every solution (α, β) of the subproblem gives a valid cut for the master problem.
- It is not necessary to solve the subproblem to optimality at each iteration.

Difference-of-Convex Algorithm (DCA)

$$g(\alpha, \beta) = c^\top \beta + f(\alpha)$$

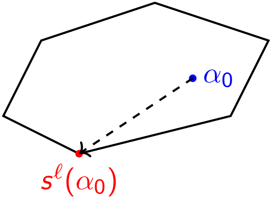


1. Choose a starting point α_0.
2. Compute $\partial f(\alpha_k)$.
3. Compute an optimal solution $s^\ell(\alpha_k) \in \arg\max_{(\alpha,\beta)\in\Lambda} c^\top \beta + \partial f(\alpha_k)^\top \alpha$.
4. $\alpha_{k+1} = s^\ell(\alpha_k)$

$\Longrightarrow$ The sequence $(g(\alpha_k, \beta_k))_{k\geq 1}$ is strictly increasing.
$\Longrightarrow$ Algorithm **terminates** and converges toward a local maximum.

Difference-of-Convex Algorithm (DCA)

$$g(\alpha, \beta) = c^\top \beta + f(\alpha)$$

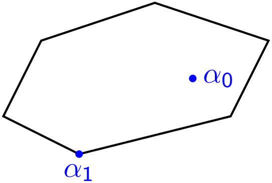


1. Choose a starting point α_0.
2. Compute $\partial f(\alpha_k)$.
3. Compute an optimal solution $s^\ell(\alpha_k) \in \arg\max_{(\alpha,\beta)\in\Lambda} c^\top \beta + \partial f(\alpha_k)^\top \alpha$.
4. $\alpha_{k+1} = s^\ell(\alpha_k)$

$\implies$ The sequence $(g(\alpha_k, \beta_k))_{k\geq 1}$ is strictly increasing.
$\implies$ Algorithm **terminates** and converges toward a local maximum.

Difference-of-Convex Algorithm (DCA)

$$g(\alpha, \beta) = c^\top \beta + f(\alpha)$$

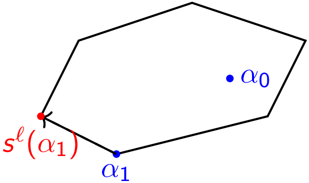


1. Choose a starting point α_0.
2. Compute $\partial f(\alpha_k)$.
3. Compute an optimal solution $s^\ell(\alpha_k) \in \arg\max_{(\alpha,\beta)\in\Lambda} c^\top \beta + \partial f(\alpha_k)^\top \alpha$.
4. $\alpha_{k+1} = s^\ell(\alpha_k)$

$\Longrightarrow$ The sequence $(g(\alpha_k, \beta_k))_{k\geq 1}$ is strictly increasing.
$\Longrightarrow$ Algorithm **terminates** and converges toward a local maximum.

Difference-of-Convex Algorithm (DCA)

$$g(\alpha, \beta) = c^\top \beta + f(\alpha)$$

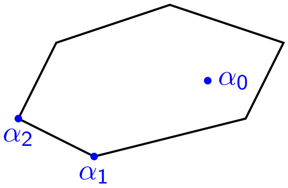


1. Choose a starting point α_0.
2. Compute $\partial f(\alpha_k)$.
3. Compute an optimal solution $s^\ell(\alpha_k) \in \arg\max_{(\alpha,\beta)\in\Lambda} c^\top \beta + \partial f(\alpha_k)^\top \alpha$.
4. $\alpha_{k+1} = s^\ell(\alpha_k)$

$\implies$ The sequence $(g(\alpha_k, \beta_k))_{k\geq 1}$ is strictly increasing.
$\implies$ Algorithm **terminates** and converges toward a local maximum.

Difference-of-Convex Algorithm (DCA)

$$g(\alpha, \beta) = c^\top \beta + f(\alpha)$$



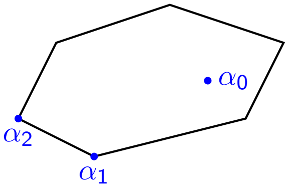
1. Choose a starting point α_0.
2. Compute $\partial f(\alpha_k)$.
3. Compute an optimal solution $s^\ell(\alpha_k) \in \arg\max_{(\alpha,\beta)\in\Lambda} c^\top \beta + \partial f(\alpha_k)^\top \alpha$.
4. $\alpha_{k+1} = s^\ell(\alpha_k)$

$\implies$ The sequence $(g(\alpha_k, \beta_k))_{k\geq 1}$ is strictly increasing.

$\implies$ Algorithm **terminates** and converges toward a local maximum.

Difference-of-Convex Algorithm (DCA)

$$g(\alpha, \beta) = c^\top \beta + f(\alpha)$$



1. Choose a starting point α_0.
2. Compute $\partial f(\alpha_k)$.
3. Compute an optimal solution $s^\ell(\alpha_k) \in \arg\max_{(\alpha,\beta)\in\Lambda} c^\top \beta + \partial f(\alpha_k)^\top \alpha$.
4. $\alpha_{k+1} = s^\ell(\alpha_k)$

$\implies$ The sequence $(g(\alpha_k, \beta_k))_{k\geq 1}$ is strictly increasing.
$\implies$ Algorithm **terminates** and converges toward a local maximum.

Contents

1^{st} starting point strategy: the non-DRO solution

$$f(\alpha) = \max_{\xi \in \mathcal{Z}} \alpha^\top \xi - \lambda \|\xi - \zeta_i\|^p$$
$$\partial f(0) \in \arg\max_{\xi \in \mathcal{Z}} -\lambda \|\xi - \zeta_i\|^p = \{\zeta_i\}$$

Linearized problem with $\alpha_0 = 0$

$$\begin{aligned} &\max c^\top \beta + \partial f(\alpha_0)^\top \alpha \\ &\text{s.t.} \quad (\alpha, \beta) \in \Lambda \end{aligned}$$

Stochastic Optimization

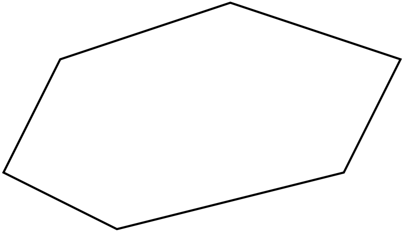
$$\begin{aligned} &\max c^\top \beta + \zeta_i^\top \alpha \\ &\text{s.t.} \quad (\alpha, \beta) \in \Lambda \end{aligned}$$

- The DRO problem is, in general, “close” to the empirical one.
- Starting with $\alpha_0 = 0$ is in general a good choice.

2^{nd} starting point strategy: Sampling

Proposition

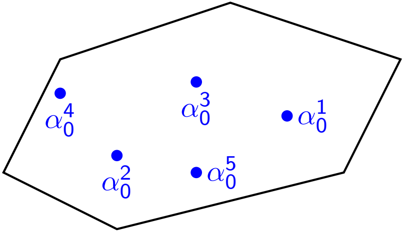
The set of starting points α_0 such that the DCA algorithm converges to a global maximum has a non-zero measure.



2^{nd} starting point strategy: Sampling

Proposition

The set of starting points α_0 such that the DCA algorithm converges to a global maximum has a non-zero measure.

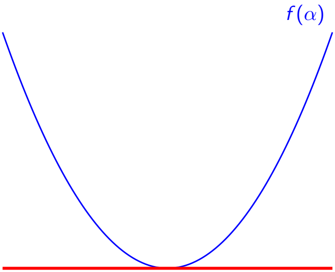


Randomly drawing the starting point ensures convergence almost surely.

3^{rd} starting point strategy: Lower approximation

Proposition

The DCA algorithm approximates the non-convex part by its tangent at α_0



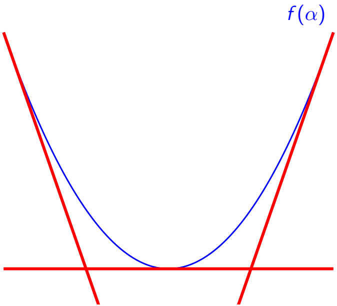
$$\max_{\alpha,\beta} c^\top \beta + f(\alpha) \quad \text{s.t. } (\alpha, \beta) \in \Lambda$$

$$\max_{\alpha,\beta} c^\top \beta + \partial f(\bar{\alpha})^\top \alpha \quad \text{s.t. } (\alpha, \beta) \in \Lambda$$

3^{rd} starting point strategy: Lower approximation

Proposition

The DCA algorithm approximates the non-convex part by its tangent at α_0



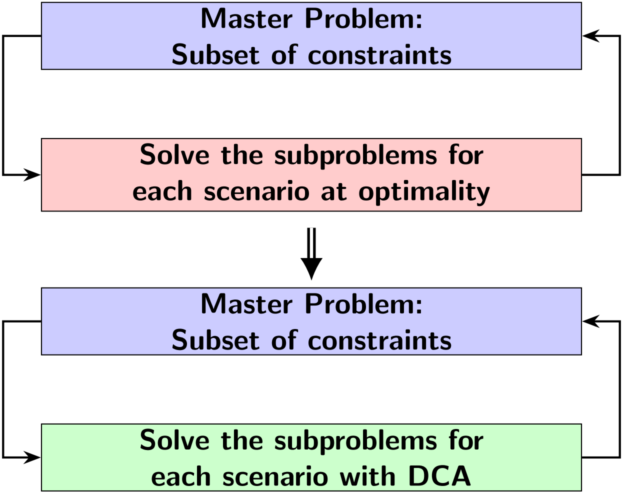
$$\max_{\alpha,\beta} c^\top \beta + f(\alpha) \quad \text{s.t.}\ (\alpha,\beta) \in \Lambda$$

$$\max_{\alpha,\beta} c^\top \beta + \partial f(\bar{\alpha})^\top \alpha \quad \text{s.t.}\ (\alpha,\beta) \in \Lambda$$

Improve the starting point by approximating the non-convex part by a piecewise linear approximation.

The problem becomes a MILP !

Adapted Benders' Algorithm



Contents

KKT reformulation ($f(\alpha) = \|\alpha\|_2^2$)

$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

- KKT conditions $\implies$ MILP.

→ **Advantage:** Provide optimal solution.

→ **Drawback:** Large number of binary variables.

→ **Worst than Gurobi solving directly the quadratic problem**

KKT reformulation ($f(\alpha) = \|\alpha\|_2^2$)

$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

- KKT conditions $\implies$ MILP.

→ **Advantage:** Provide optimal solution.

→ **Drawback:** Large number of binary variables.

→ **Worst than Gurobi solving directly the quadratic problem**

KKT reformulation ($f(\alpha) = \|\alpha\|_2^2$)

$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

- KKT conditions $\implies$ MILP.

→ **Advantage:** Provide optimal solution.

→ **Drawback:** Large number of binary variables.

→ **Worst than Gurobi solving directly the quadratic problem**

PSD relaxation ($f(\alpha) = \|\alpha\|_2^2$)

$$\max_{\alpha,\beta} c^\top \beta + \|\alpha\|_2^2$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

PSD relaxation ($f(\alpha) = \|\alpha\|_2^2$)

$$\max_{\alpha,\beta} c^\top \beta + \|\alpha\|_2^2$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

PSD relaxation first order:

$$\begin{aligned} \max \quad & c^\top \beta + Tr(X) \\ \text{s.t.} \quad & (\alpha, \beta) \in \Lambda \\ & X = \begin{bmatrix} 1 & \alpha^\top \\ \alpha & \alpha\alpha^\top \end{bmatrix} \end{aligned}$$

PSD relaxation ($f(\alpha) = \|\alpha\|_2^2$)

$$\max_{\alpha,\beta} c^\top \beta + \|\alpha\|_2^2$$
$$\text{s.t.}\ (\alpha, \beta) \in \Lambda$$

PSD relaxation first order:

$$\begin{aligned} \max \quad & c^\top \beta + Tr(X) \\ \text{s.t.} \quad & (\alpha, \beta) \in \Lambda \\ & X \succeq 0 \end{aligned}$$

PSD relaxation ($f(\alpha) = \|\alpha\|_2^2$)

$$\max_{\alpha,\beta} c^\top \beta + \|\alpha\|_2^2$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

PSD relaxation first order:

$$\begin{aligned} \max \quad & c^\top \beta + Tr(X) \\ \text{s.t.} \quad & (\alpha, \beta) \in \Lambda \\ & X \succeq 0 \end{aligned}$$

→ **Advantage:** Tractable convex problem.

→ **Drawback:** Optimal value: $+\infty$

→ **Worst than Gurobi solving directly the quadratic problem**

PSD relaxation ($f(\alpha) = \|\alpha\|_2^2$)

$$\max_{\alpha,\beta} c^\top \beta + \|\alpha\|_2^2$$
$$\text{s.t.}\ (\alpha, \beta) \in \Lambda$$

PSD relaxation second order:

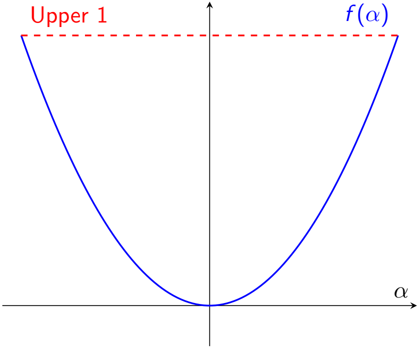
- **Idea:** Consider a large PSD matrix that include product variables between x and y up to order 2.

→ **Advantage:** Bounded convex problem and good relaxation.

→ **Drawback:** Untractable with solvers like Mosek.

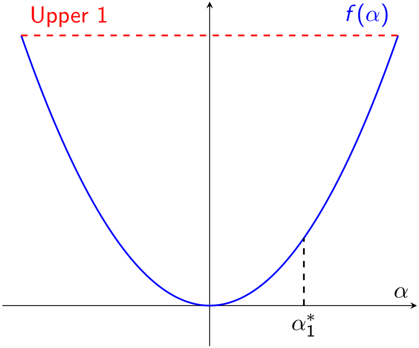
→ **Worse than Gurobi solving directly the quadratic problem**

Piecewise linear upper approximation



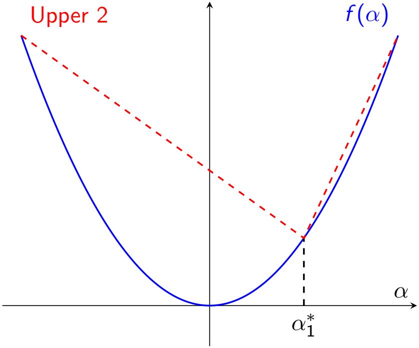
$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

Piecewise linear upper approximation



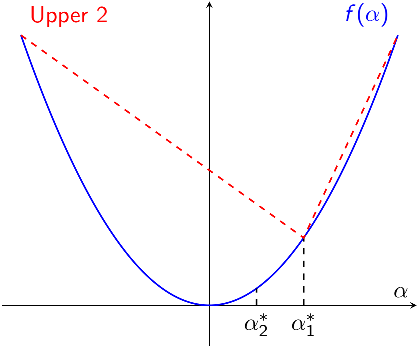
$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

Piecewise linear upper approximation



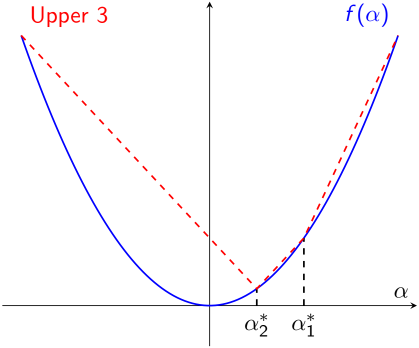
$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

Piecewise linear upper approximation



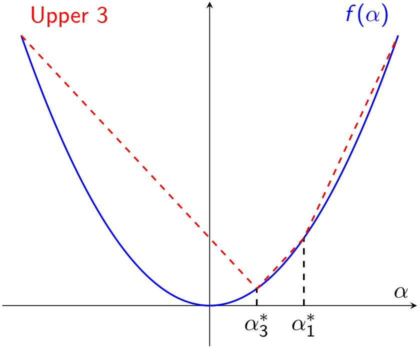
$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha,\beta) \in \Lambda$$

Piecewise linear upper approximation



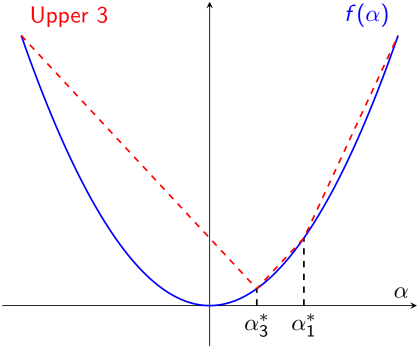
$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

Piecewise linear upper approximation



$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

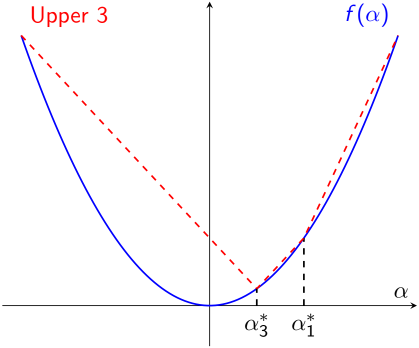
Piecewise linear upper approximation



$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

- MILPs.

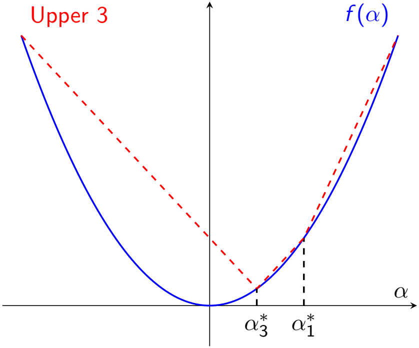
Piecewise linear upper approximation



$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha,\beta) \in \Lambda$$

- MILPs.
- Number of binary variables increases at each iteration

Piecewise linear upper approximation



$$\max_{\alpha,\beta} c^\top \beta + f(\alpha)$$
$$\text{s.t. } (\alpha, \beta) \in \Lambda$$

- MILPs.
- Number of binary variables increases at each iteration
- Convergence in a few iterations.

Contents

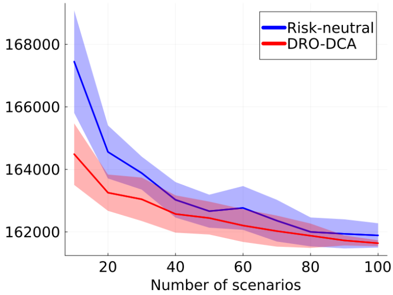
Methodology

→ Two test cases: one small, one large.
- **Small case**: 24 hours, 3 units, 6 buses
- **Large case**: 24 hours, 54 units, 118 buses

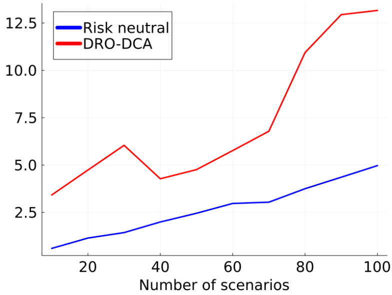
→ Before being able to solve the DRO problem, we first need to be able to solve the risk-neutral version using Benders' decomposition.

→ For the Unit Commitment, we use interval variables to model whether a unit is on or off over a time interval (paper in preparation).

Small test case

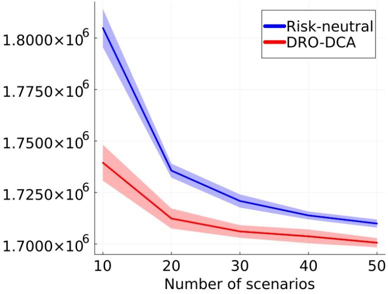


(a) Average costs out-of-sample

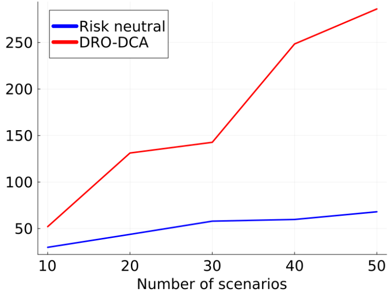


(b) Computational time (s)

Large test case



(a) Average costs out-of-sample



(b) Computational time (s)

Conclusion

What I presented:

→ A framework for distributionally robust unit commitment problems, highlighting why the 2-norm is more effective than the 1-norm.

→ A novel method to solve the subproblems using DCA.

→ Computational experiments showing that the DRO model outperforms the risk-neutral one.

Future work:

→ Conduct a large-scale comparison between DRO, RO, and SO models on a comprehensive benchmark.

→ Extend the framework to multi-stage decision problems.

Thank you for your attention!