Solving a convex quadratic maximization problem appearing in some distributionally robust problem

Mathis Azéma (ENPC, Paris, France)

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

EUROPT

June 30th, 2025

Context

$$\begin{aligned} \max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + c^\top x + \|x\|_2^2 \\ \text{s.t.} \quad & Ax + Ty = b, \\ & x \geq 0,\ y \geq 0 \end{aligned}$$

→ Appears as the **subproblem in a Benders' decomposition** for a Distributionally Robust Optimization (DRO) problem.

→ Solving it to optimality provides the **tightest possible upper bound**.

→ The problem is **NP-hard**.

→ Not necessary to solve it at optimality at each iteration!

Context

$$\max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad d^\top y + c^\top x + \|x\|_2^2$$
$$\text{s.t.} \quad Ax + Ty = b,$$
$$x \geq 0,\ y \geq 0$$

→ Appears as the **subproblem in a Benders' decomposition** for a Distributionally Robust Optimization (DRO) problem.

→ Solving it to optimality provides the **tightest possible upper bound**.

→ The problem is **NP-hard**.

→ Not necessary to solve it at optimality at each iteration!

Context

$$\max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad d^\top y + c^\top x + \|x\|_2^2$$
$$\text{s.t.} \quad Ax + Ty = b,$$
$$x \geq 0,\ y \geq 0$$

→ Appears as the **subproblem in a Benders' decomposition** for a Distributionally Robust Optimization (DRO) problem.

→ Solving it to optimality provides the **tightest possible upper bound**.

→ The problem is **NP-hard**.

→ Not necessary to solve it at optimality at each iteration!

Context

$$\max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad d^\top y + c^\top x + \|x\|_2^2$$
$$\text{s.t.} \quad Ax + Ty = b,$$
$$x \geq 0,\ y \geq 0$$

→ Appears as the **subproblem in a Benders' decomposition** for a Distributionally Robust Optimization (DRO) problem.

→ Solving it to optimality provides the **tightest possible upper bound**.

→ The problem is **NP-hard**.

→ Not necessary to solve it at optimality at each iteration!

Contents

Contents

Deterministic UC Problem

$\mathcal{M}$: set of units

T: time horizon

$\xi \in \mathbb{R}^T$: demand vector

x_i: commitment variables (binary).

y_i: production variables (continuous).

$$
\begin{aligned}
\min_{x_i, y_i} \quad & \sum_{i \in \mathcal{M}} c_i^\top x_i + \sum_{i \in \mathcal{M}} b_i^\top y_i \\
s.t. \quad & x_i \in X_i, && \forall i \in \mathcal{M} \\
& y_i \in Y_i(x_i) && \forall i \in \mathcal{M} \\
& \sum_{i \in \mathcal{M}} y_i = \xi
\end{aligned}
$$

2-stage UC Problem under uncertainty

Uncertainty on the demand.

2-stage assumption:

- 1^{st} stage: commitment variables (binary)
- 2^{nd} stage: production variables (continuous)

$$
\begin{aligned}
\min_{x_i, y_i} \quad & c^\top x + b^\top y \\
s.t. \quad & x \in X \\
& y_i \in Y_i(x_i) \qquad \forall i \in \mathcal{M} \\
& \sum_{i \in \mathcal{M}} y_i = \boxed{\xi}\,?
\end{aligned}
$$

2-stage UC Problem under uncertainty

Uncertainty on the demand.

2-stage assumption:

- 1^{st} stage: commitment variables (binary)
- 2^{nd} stage: production variables (continuous)

$$\begin{aligned}
\min_{x_i, y_i} \quad & c^\top x + b^\top y \\
s.t. \quad & x \in X \\
& y_i \in Y_i(x_i) \qquad \forall i \in \mathcal{M} \\
& \sum_{i \in \mathcal{M}} y_i = \xi
\end{aligned}$$

2-stage UC Problem under uncertainty

Uncertainty on the demand.

2-stage assumption:

- 1^{st} stage: commitment variables (binary)
- 2^{nd} stage: production variables (continuous)

$$\begin{aligned} \min_{x_i} \quad & c^\top x + \min_{\substack{y_i : y_i \in Y_i(x_i) \\ \sum_{i \in \mathcal{M}} y_i = \xi}} b^\top y \\ s.t. \quad & x \in X \end{aligned}$$

2-stage UC Problem under uncertainty

Uncertainty on the demand.

2-stage assumption:

- 1^{st} stage: commitment variables (binary)
- 2^{nd} stage: production variables (continuous)

$$\begin{aligned} \min_{x_i} \quad & c^\top x + Q(x,\xi) \\ s.t. \quad & x \in X \end{aligned}$$

$Q(x,\xi)$: recourse value representing the optimal cost of the second stage, considering which units are on or off (first stage) and the demand ξ

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

Distributionnally 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:

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

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

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

Distributionnally 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:

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

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

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

Distributionnally 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:

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

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

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

Wasserstein distance-based ambiguity sets

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

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

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

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

$p = 1$:	$p = 2$:
- Reformulation with linear programs - Actually, it is the same problem as with $\mathcal{P} = \{\mathbb{P}_0\}$.	- Reformulation with non convex quadratic programs. - Leads to better solutions !

Wasserstein distance-based ambiguity sets

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

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

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

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

$p = 1$:	$p = 2$:
- Reformulation with linear programs - Actually, it is the same problem as with $\mathcal{P} = \{\mathbb{P}_0\}$.	- Reformulation with non convex quadratic programs. - Leads to better solutions !

DRO Unit Commitment problem

Idea: $\mathcal{P}$ has to include distributions "close" to the empirical one.
Wasserstein distance-based ambiguity sets

$$\mathcal{P} = \left\{ \mathbb{Q} \,|\, supp(\mathbb{Q}) \subset \mathbb{R}^T,\ W_2(\mathbb{Q}, \mathbb{P}_0) \leq \theta \right\}, \qquad \mathbb{P}_0 = \frac{1}{N} \sum_{i \in [N]} \delta_{\zeta_i}$$

$$\begin{aligned} \min_x \quad & c^\top x + \max_{\substack{\mathbb{Q}:\ W_2(\mathbb{Q},\mathbb{P}_0) \leq \theta \quad [\lambda] \\ supp(\mathbb{Q}) \subset \mathbb{R}^T}} \mathbb{E}_{\mathbb{Q}}[Q(x, \xi)] \\ s.t. \quad & x \in X \end{aligned}$$

DRO Unit Commitment problem

Idea: $\mathcal{P}$ has to include distributions "close" to the empirical one.
Wasserstein distance-based ambiguity sets

$$\mathcal{P} = \left\{ \mathbb{Q} \,|\, supp(\mathbb{Q}) \subset \mathbb{R}^T,\ W_2(\mathbb{Q}, \mathbb{P}_0) \leq \theta \right\}, \qquad \mathbb{P}_0 = \frac{1}{N} \sum_{i \in [N]} \delta_{\zeta_i}$$

$$\begin{aligned} \min_{x, \lambda \geq 0} \quad & c^\top x + \lambda \theta^2 + \frac{1}{N} \sum_{i=1}^{N} \sup_{\xi \in \mathbb{R}^T} \left(Q(x, \xi) - \lambda \|\xi - \zeta_i\|_2^2 \right) \\ s.t. \quad & x \in X \end{aligned}$$

DRO Unit Commitment problem

Idea: $\mathcal{P}$ has to include distributions "close" to the empirical one.
Wasserstein distance-based ambiguity sets

$$\mathcal{P} = \left\{ \mathbb{Q} \,|\, supp(\mathbb{Q}) \subset \mathbb{R}^T,\ W_2(\mathbb{Q}, \mathbb{P}_0) \leq \theta \right\}, \qquad \mathbb{P}_0 = \frac{1}{N} \sum_{i \in [N]} \delta_{\zeta_i}$$

$$\min_{x, \lambda \geq 0} \quad c^\top x + \lambda\theta^2 + \frac{1}{N} \sum_{i=1}^{N} \max_{\xi \in \mathbb{R}^T, k \in [K]} \left(\alpha_k^\top \xi + \beta_k^\top x + \gamma_k - \lambda \|\xi - \zeta_i\|^2 \right)$$

$$s.t. \quad x \in X$$

DRO Unit Commitment problem

Idea: $\mathcal{P}$ has to include distributions "close" to the empirical one.
Wasserstein distance-based ambiguity sets

$$\mathcal{P} = \left\{ \mathbb{Q} \,|\, supp(\mathbb{Q}) \subset \mathbb{R}^T,\ W_2(\mathbb{Q}, \mathbb{P}_0) \leq \theta \right\}, \qquad \mathbb{P}_0 = \frac{1}{N} \sum_{i \in [N]} \delta_{\zeta_i}$$

$$\min_{x, \lambda \geq 0} \quad c^\top x + \lambda \theta^2 + \frac{1}{N} \sum_{i=1}^{N} \max_{k \in [K]} \left(\beta_k^\top x + \gamma_k + \max_{\xi \in \mathbb{R}^T} (\alpha_k^\top \xi - \lambda \|\xi - \zeta_i\|^2) \right)$$

$$s.t. \quad x \in X$$

DRO Unit Commitment problem

Idea: $\mathcal{P}$ has to include distributions “close” to the empirical one.
Wasserstein distance-based ambiguity sets

$$\mathcal{P} = \left\{ \mathbb{Q} \,|\, supp(\mathbb{Q}) \subset \mathbb{R}^T,\; W_2(\mathbb{Q}, \mathbb{P}_0) \leq \theta \right\}, \qquad \mathbb{P}_0 = \frac{1}{N} \sum_{i \in [N]} \delta_{\zeta_i}$$

$$\begin{aligned} \min_{x, \lambda \geq 0} \quad & c^\top x + \lambda \theta^2 + \frac{1}{N} \sum_{i=1}^{N} \max_{k \in [K]} \left(\beta_k^\top x + \gamma_k + \alpha_k^\top \zeta_i + \frac{\|\alpha_k\|_2^2}{4\lambda} \right) \\ s.t. \quad & x \in X \end{aligned}$$

DRO Unit Commitment problem

Idea: $\mathcal{P}$ has to include distributions "close" to the empirical one.
Wasserstein distance-based ambiguity sets

$$\mathcal{P} = \left\{ \mathbb{Q} \,|\, supp(\mathbb{Q}) \subset \mathbb{R}^T,\ W_2(\mathbb{Q}, \mathbb{P}_0) \leq \theta \right\}, \qquad \mathbb{P}_0 = \frac{1}{N} \sum_{i \in [N]} \delta_{\zeta_i}$$

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

DRO Unit Commitment problem

Idea: $\mathcal{P}$ has to include distributions "close" to the empirical one.
Wasserstein distance-based ambiguity sets

$$\mathcal{P} = \left\{ \mathbb{Q} \,|\, supp(\mathbb{Q}) \subset \mathbb{R}^T,\ W_2(\mathbb{Q}, \mathbb{P}_0) \leq \theta \right\}, \qquad \mathbb{P}_0 = \frac{1}{N} \sum_{i \in [N]} \delta_{\zeta_i}$$

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

$\implies$ **Benders' algorithm**

Benders algorithm DRO

Master Problem:
Subset of constraints

(x^*, λ^*)

$(\alpha^*, \beta^*, \gamma^*)$

Subproblem for each scenario i:

$$\begin{aligned} \max_{\alpha,\beta,\gamma} \quad & \alpha^\top \zeta_i + \beta^\top x^* + \gamma + \frac{\lambda^* \|\alpha\|_2^2}{4} \\ s.t. \quad & (\alpha, \beta, \gamma) \in \Lambda \end{aligned}$$

Benders algorithm DRO

Master Problem:
Subset of constraints

(x^*, λ^*)

Subproblem for each scenario i:

$$\max_{\alpha,\beta,\gamma} \quad \alpha^\top \zeta_i + \beta^\top x^* + \gamma + \frac{\lambda^* \|\alpha\|_2^2}{4}$$
$$s.t. \quad (\alpha, \beta, \gamma) \in \Lambda$$

$(\alpha^*, \beta^*, \gamma^*)$

Convex maximization problem (NP-Hard)

Comparison risk-neutral/ DRO subproblems

Time spent on instance with 50 units and 1 random demand
$\implies$ dimension of the quadratic term: $T = 24$

Subproblem Risk-Neutral for each scenario i:

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

$\downarrow$

LP

$\downarrow$

$\approx 0.002s$

Subproblem DRO for each scenario i:

$$\max_{\alpha,\beta,\gamma} \quad \alpha^\top \zeta_i + \beta^\top x^* + \gamma + \frac{\lambda^* \|\alpha\|_2^2}{4}$$
$$s.t. \quad (\alpha, \beta, \gamma) \in \Lambda$$

$\downarrow$

Convex maximization problem (NP-Hard)

$\downarrow$

$\approx 0.5s$

- 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 on instance with 50 units and 1 random demand
$\implies$ dimension of the quadratic term: $T = 24$

Subproblem Risk-Neutral for each scenario i:

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

$\downarrow$ LP $\downarrow$ $\approx 0.002s$

Subproblem DRO for each scenario i:

$$\begin{aligned} \max_{\alpha,\beta,\gamma} \quad & \alpha^\top \zeta_i + \beta^\top x^* + \gamma + \frac{\lambda^* \|\alpha\|_2^2}{4} \\ s.t. \quad & (\alpha, \beta, \gamma) \in \Lambda \end{aligned}$$

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

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

Contents

Notations

Convex maximization problem: ("Hard" to solve)

$$
(\mathcal{P}) \quad \begin{aligned} \max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + c^\top x + \|x\|_2^2 \\ \text{s.t.} \quad & Ax + Ty = b, \\ & x \geq 0,\ y \geq 0 \end{aligned}
$$

$\boxed{\text{Optimal value: } v^*, \qquad \text{Optimal solution: } (x^*, y^*)}$ $\leftarrow$ Optimal cut

Linearization: ("Easy" to solve)

$$
(\mathcal{P}_\ell(\bar{x})) \quad \begin{aligned} \max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + (c + 2\bar{x})^\top x - \|\bar{x}\|_2^2 \\ \text{s.t.} \quad & Ax + Ty = b, \\ & x \geq 0,\ y \geq 0 \end{aligned}
$$

$\boxed{\text{Optimal value: } v^\ell(\bar{x}), \qquad \text{Optimal solution: } s^\ell(\bar{x})}$ $\leftarrow$ Valid cut

Notations

Convex maximization problem: ("Hard" to solve)

$$(\mathcal{P}) \quad \begin{aligned} \max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + c^\top x + \|x\|_2^2 \\ \text{s.t.} \quad & Ax + Ty = b, \\ & x \geq 0,\ y \geq 0 \end{aligned}$$

$\boxed{\text{Optimal value: } v^*, \qquad \text{Optimal solution: } (x^*, y^*)}$ $\leftarrow$ Optimal cut

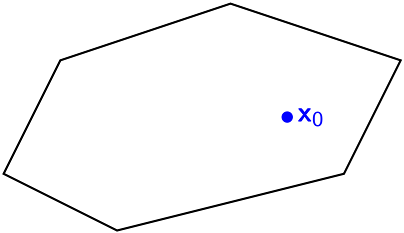
Linearization: ("Easy" to solve)

$$(\mathcal{P}_\ell(\bar{x})) \quad \begin{aligned} \max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + (c + 2\bar{x})^\top x - \|\bar{x}\|_2^2 \\ \text{s.t.} \quad & Ax + Ty = b, \\ & x \geq 0,\ y \geq 0 \end{aligned}$$

$\boxed{\text{Optimal value: } v^\ell(\bar{x}), \qquad \text{Optimal solution: } s^\ell(\bar{x})}$ $\leftarrow$ Valid cut

Frank-Wolfe Algorithm

$$f(x, y) = d^\top y + c^\top x + \|x\|_2^2$$



Choice of the step length α_k by solving:

$$\max_{\alpha \in [0,1]} f((x_k, y_k) + \alpha(p_k^x, p_k^y))$$

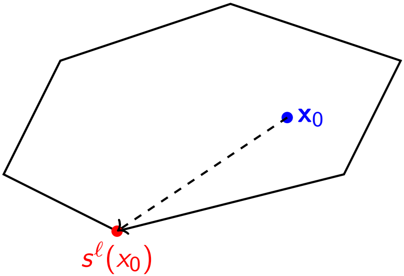
$\implies$ By convexity, $\alpha_k \in \{0, 1\} \implies (x_{k+1}, y_{k+1}) = s^\ell(x_k)$.

$\implies$ The sequence $(f(x_k, y_k))_{k \geq 1}$ is strictly increasing.

$\implies$ Frank-Wolfe Algorithm **terminates**.

Frank-Wolfe Algorithm

$$f(x, y) = d^\top y + c^\top x + \|x\|_2^2$$



Choice of the step length α_k by solving:

$$\max_{\alpha \in [0,1]} f((x_k, y_k) + \alpha(p_k^x, p_k^y))$$

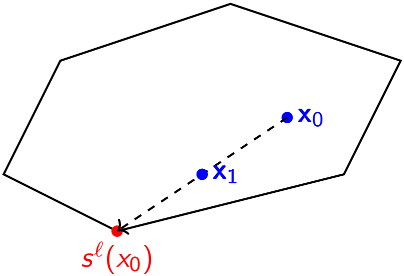
$\implies$ By convexity, $\alpha_k \in \{0, 1\} \implies (x_{k+1}, y_{k+1}) = s^\ell(x_k)$.

$\implies$ The sequence $(f(x_k, y_k))_{k \geq 1}$ is strictly increasing.

$\implies$ Frank-Wolfe Algorithm **terminates**.

Frank-Wolfe Algorithm

$$f(x,y) = d^\top y + c^\top x + \|x\|_2^2$$



Choice of the step length α_k by solving:

$$\max_{\alpha \in [0,1]} f((x_k, y_k) + \alpha(p_k^x, p_k^y))$$

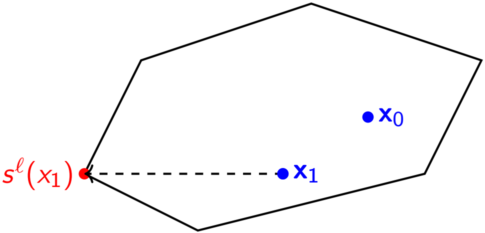
$\implies$ By convexity, $\alpha_k \in \{0,1\} \implies (x_{k+1}, y_{k+1}) = s^\ell(x_k)$.

$\implies$ The sequence $(f(x_k, y_k))_{k\geq 1}$ is strictly increasing.

$\implies$ Frank-Wolfe Algorithm **terminates**.

Frank-Wolfe Algorithm

$$f(x, y) = d^\top y + c^\top x + \|x\|_2^2$$



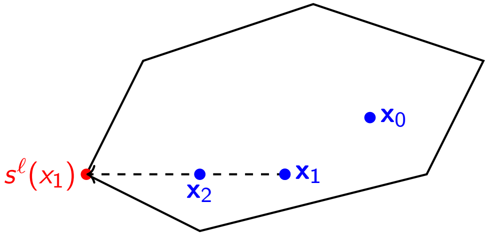
Choice of the step length α_k by solving:

$$\max_{\alpha \in [0,1]} f((x_k, y_k) + \alpha(p_k^x, p_k^y))$$

$\implies$ By convexity, $\alpha_k \in \{0, 1\} \implies (x_{k+1}, y_{k+1}) = s^\ell(x_k)$.
$\implies$ The sequence $(f(x_k, y_k))_{k \geq 1}$ is strictly increasing.
$\implies$ Frank-Wolfe Algorithm **terminates**.

Frank-Wolfe Algorithm

$$f(x,y) = d^\top y + c^\top x + \|x\|_2^2$$



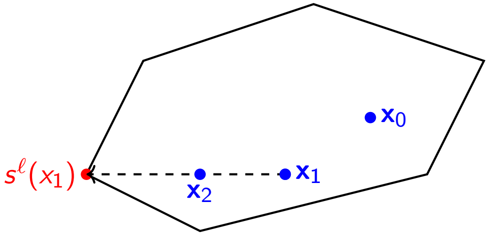
Choice of the step length α_k by solving:

$$\max_{\alpha\in[0,1]} f((x_k,y_k)+\alpha(p_k^x,p_k^y))$$

$\implies$ By convexity, $\alpha_k \in \{0,1\} \implies (x_{k+1},y_{k+1}) = s^\ell(x_k)$.
$\implies$ The sequence $(f(x_k,y_k))_{k\geq 1}$ is strictly increasing.
$\implies$ Frank-Wolfe Algorithm **terminates**.

Frank-Wolfe Algorithm

$$f(x,y) = d^\top y + c^\top x + \|x\|_2^2$$



Choice of the step length α_k by solving:

$$\max_{\alpha \in [0,1]} f((x_k, y_k) + \alpha(p_k^x, p_k^y))$$

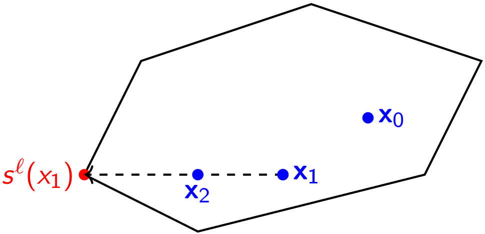
$\implies$ By convexity, $\alpha_k \in \{0,1\} \implies (x_{k+1}, y_{k+1}) = s^\ell(x_k)$.

$\implies$ The sequence $(f(x_k, y_k))_{k \geq 1}$ is strictly increasing.

$\implies$ Frank-Wolfe Algorithm **terminates**.

Frank-Wolfe Algorithm

$$f(x,y) = d^\top y + c^\top x + \|x\|_2^2$$



Choice of the step length α_k by solving:

$$\max_{\alpha \in [0,1]} f((x_k, y_k) + \alpha(p_k^x, p_k^y))$$

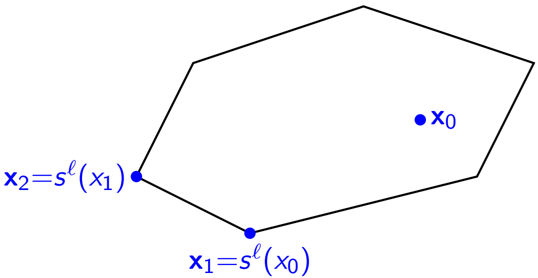
$\Longrightarrow$ By convexity, $\alpha_k \in \{0,1\} \Longrightarrow (x_{k+1}, y_{k+1}) = s^\ell(x_k)$.

$\Longrightarrow$ The sequence $(f(x_k, y_k))_{k \geq 1}$ is strictly increasing.

$\Longrightarrow$ Frank-Wolfe Algorithm **terminates**.

Frank-Wolfe Algorithm

$$f(x, y) = d^\top y + c^\top x + \|x\|_2^2$$



Choice of the step length α_k by solving:

$$\max_{\alpha \in [0,1]} f((x_k, y_k) + \alpha(p_k^x, p_k^y))$$

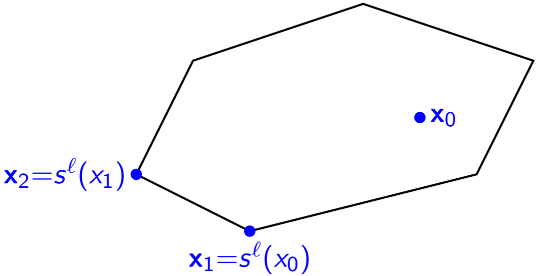
$\implies$ By convexity, $\alpha_k \in \{0, 1\} \implies (x_{k+1}, y_{k+1}) = s^\ell(x_k)$.

$\implies$ The sequence $(f(x_k, y_k))_{k \geq 1}$ is strictly increasing.

$\implies$ Frank-Wolfe Algorithm **terminates**.

Frank-Wolfe Algorithm

$$f(x,y) = d^\top y + c^\top x + \|x\|_2^2$$



Choice of the step length α_k by solving:

$$\max_{\alpha \in [0,1]} f((x_k, y_k) + \alpha(p_k^x, p_k^y))$$

$\implies$ By convexity, $\alpha_k \in \{0,1\} \implies (x_{k+1}, y_{k+1}) = s^\ell(x_k)$.

$\implies$ The sequence $(f(x_k, y_k))_{k \geq 1}$ is strictly increasing.

$\implies$ Frank-Wolfe Algorithm **terminates**.

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

Linearized problem with $\bar{x} = 0$

$$
\begin{aligned}
\max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + (c + 2\bar{x})^\top x - \|\bar{x}\|_2^2 \\
\text{s.t.} \quad & Ax + Ty = b, \\
& x \geq 0, \; y \geq 0
\end{aligned}
$$

Stochastic Optimization

$$
\begin{aligned}
\max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + c^\top x \\
\text{s.t.} \quad & Ax + Ty = b, \\
& x \geq 0, \; y \geq 0
\end{aligned}
$$

- The DRO problem is in general “close” to the empirical one.
- Starting with $x_0 = 0$ is in general a good choice.

2^{nd} starting point strategy: Sampling

Definition (Attraction field of a face F)

The attraction field of a face F at $\bar{x}$ is the set of points $P(F)$ such that the linearized problem at $\bar{x}$ has F as the set of optimal solutions.

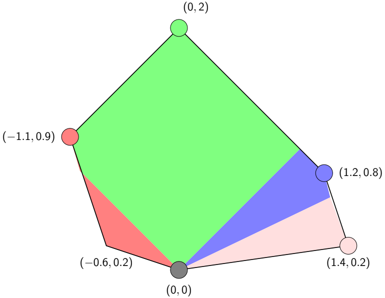


Figure: Maximizing $\|x\|_2^2$

2^{nd} starting point strategy: Sampling

Definition (Attraction field of a face F)

The attraction field of a face F at $\bar{x}$ is the set of points $P(F)$ such that the linearized problem at $\bar{x}$ has F as the set of optimal solutions.

Proposition

The set $P(F)$ is a **polyhedron**. $P(x^*)$ has a **non-zero measure**. A point is a local optimum if and only if it belongs to its attraction field.

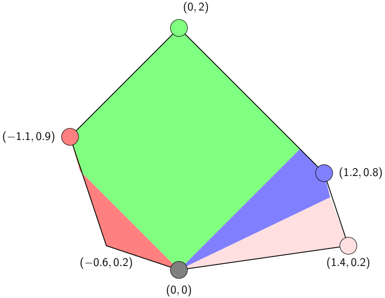


Figure: Maximizing $\|x\|_2^2$

2^{nd} starting point strategy: Sampling

Definition (Attraction field of a face F)

The attraction field of a face F at $\bar{x}$ is the set of points $P(F)$ such that the linearized problem at $\bar{x}$ has F as the set of optimal solutions.

Proposition

The set $P(F)$ is a **polyhedron**. $P(x^*)$ has a **non-zero measure**. A point is a local optimum if and only if it belongs to its attraction field.

- FW algorithm finds a local optimum.
- Randomly draw the starting point.

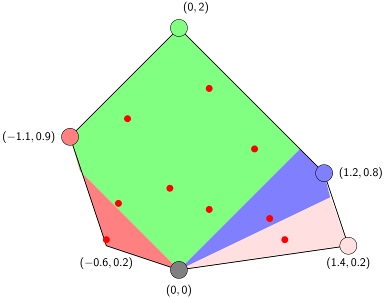


Figure: Maximizing $\|x\|_2^2$

3^{rd} starting point strategy: Discretize the space

Proposition

Let $\bar{x}$ and $v(x,y)$ defined by :

$$v(x,y) = d^\top y + c^\top x + \|x\|_2^2$$

the following inequality holds:

$$v^\ell(\bar{x}) \leq v(s^\ell(\bar{x})) \leq v^* \leq v^\ell(\bar{x}) + \|\bar{x} - x^*\|^2.$$

Corollary

Let $\varepsilon > 0$ and $X \subset \mathbb{R}^n$ such that:

$$\max_{x \in P} dist(x, X) \leq \varepsilon \qquad dist(x, X) = \min_{\bar{x} \in X} \|x - \bar{x}\|_2$$

Define $v^\ell(X) = \max_{x \in X} v^\ell(x)$. Then, the following inequalities hold:

$$v^\ell(X) \leq v^* \leq \varepsilon^2 + v^\ell(X)$$

3^{rd} starting point strategy: Discretize the space

Proposition

Let $\bar{x}$ and $v(x, y)$ defined by :

$$v(x, y) = d^\top y + c^\top x + \|x\|_2^2$$

the following inequality holds:

$$v^\ell(\bar{x}) \leq v(s^\ell(\bar{x})) \leq v^* \leq v^\ell(\bar{x}) + \|\bar{x} - x^*\|^2.$$

Corollary

Let $\varepsilon > 0$ and $X \subset \mathbb{R}^n$ such that:

$$\max_{x \in P} dist(x, X) \leq \varepsilon \qquad dist(x, X) = \min_{\bar{x} \in X} \|x - \bar{x}\|_2$$

Define $v^\ell(X) = \max_{x \in X} v^\ell(x)$. Then, the following inequalities hold:

$$v^\ell(X) \leq v^* \leq \varepsilon^2 + v^\ell(X)$$

A MILP to solve the discretized problem

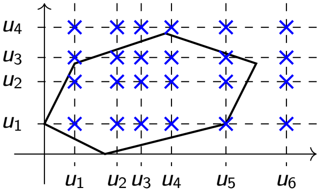
Corollary

Let $\varepsilon > 0$ and $X \subset \mathbb{R}^n$ such that:

$$\max_{x \in P} dist(x, X) \leq \varepsilon \qquad dist(x, X) = \min_{\bar{x} \in X} \|x - \bar{x}\|_2$$

Define $v^\ell(X) = \max_{x \in X} v^\ell(x)$. Then, the following inequalities hold:

$$v^\ell(X) \leq v^* \leq \varepsilon^2 + v^\ell(X)$$



$$\begin{aligned} v^\ell(X) = \max_{\bar{x},x,y} \quad & d^\top y + (c + 2\bar{x})^\top x - \|\bar{x}\|_2^2 \\ \text{s.t.} \quad & Ax + Ty = b, \\ & x \geq 0,\ y \geq 0, \bar{x} \in X \end{aligned}$$

$\Longrightarrow$ Linearization through binary variables
$\Longrightarrow$ MILP

A MILP to solve the discretized problem

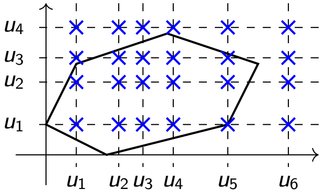
Corollary

Let $\varepsilon > 0$ and $X \subset \mathbb{R}^n$ such that:

$$\max_{x \in P} dist(x, X) \leq \varepsilon \qquad dist(x, X) = \min_{\bar{x} \in X} \|x - \bar{x}\|_2$$

Define $v^\ell(X) = \max_{x \in X} v^\ell(x)$. Then, the following inequalities hold:

$$v^\ell(X) \leq v^* \leq \varepsilon^2 + v^\ell(X)$$

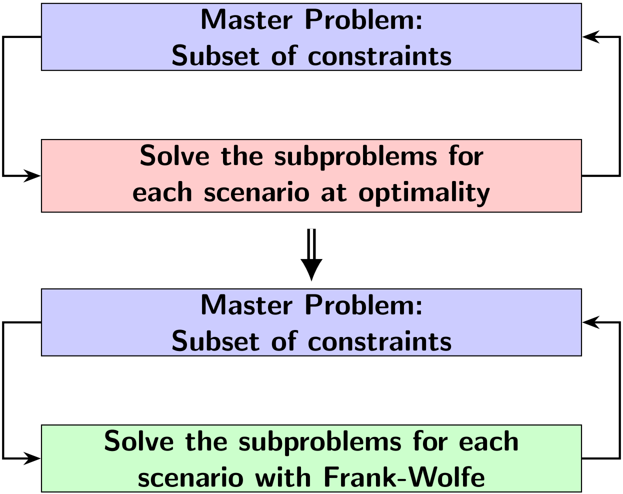


$$\begin{aligned} v^\ell(X) = \max_{\bar{x}, x, y} \quad & d^\top y + (c + 2\bar{x})^\top x - \|\bar{x}\|_2^2 \\ \text{s.t.} \quad & Ax + Ty = b, \\ & x \geq 0,\ y \geq 0, \bar{x} \in X \end{aligned}$$

$\Longrightarrow$ Linearization through binary variables

$\Longrightarrow$ MILP

Adapted Benders' Algorithm



Contents

KKT reformulation

$$\begin{aligned} \max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + c^\top x + \|x\|_2^2 \\ \text{s.t.} \quad & Ax + Ty = b, \\ & x \geq 0,\ y \geq 0 \end{aligned}$$

KKT conditions $\implies$ MILP.

→ **Advantage:** Provide optimal solution.

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

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

KKT reformulation

$$\begin{aligned} \max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + c^\top x + \|x\|_2^2 \\ \text{s.t.} \quad & Ax + Ty = b, \\ & x \geq 0,\ y \geq 0 \end{aligned}$$

KKT conditions $\implies$ MILP.

→ **Advantage:** Provide optimal solution.

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

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

KKT reformulation

$$
\begin{aligned}
\max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + c^\top x + \|x\|_2^2 \\
\text{s.t.} \quad & Ax + Ty = b, \\
& x \geq 0,\ y \geq 0
\end{aligned}
$$

KKT conditions $\implies$ MILP.

$\rightarrow$ **Advantage:** Provide optimal solution.

$\rightarrow$ **Drawback:** Large number of binary variables.

$\rightarrow$ **Worst than Gurobi solving directly the quadratic problem**

PSD relaxation

$$
\begin{aligned}
\max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + c^\top x + \|x\|_2^2 \\
\text{s.t.} \quad & Ax + Ty = b, \\
& x \geq 0,\ y \geq 0
\end{aligned}
$$

PSD relaxation

$$
\begin{aligned}
\max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + c^\top x + \|x\|_2^2 \\
\text{s.t.} \quad & Ax + Ty = b, \\
& x \geq 0,\ y \geq 0
\end{aligned}
$$

PSD relaxation first order:

$$
\begin{aligned}
\max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m_+} \quad & d^\top y + c^\top x + Tr(X) \\
\text{s.t.} \quad & Ax + Ty = b, \\
& y \geq 0,\ X = \begin{bmatrix} 1 & x^\top \\ x & xx^\top \end{bmatrix}
\end{aligned}
$$

PSD relaxation

$$\begin{aligned} \max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + c^\top x + \|x\|_2^2 \\ \text{s.t.} \quad & Ax + Ty = b, \\ & x \geq 0, \; y \geq 0 \end{aligned}$$

PSD relaxation first order:

$$\begin{aligned} \max_{y\in\mathbb{R}^m_+} \quad & d^\top y + c^\top X_{1,.} + Tr(X) \\ \text{s.t.} \quad & AX_{1,.} + Ty = b \\ & X \succeq 0 \\ & y \geq 0 \end{aligned}$$

PSD relaxation

$$\begin{aligned}\max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad & d^\top y + c^\top x + \|x\|_2^2 \\ \text{s.t.} \quad & Ax + Ty = b, \\ & x \geq 0,\ y \geq 0\end{aligned}$$

PSD relaxation first order:

$$\begin{aligned}\max_{y\in\mathbb{R}^m_+} \quad & d^\top y + c^\top X_{1,.} + Tr(X) \\ \text{s.t.} \quad & AX_{1,.} + Ty = b \\ & X \succeq 0 \\ & y \geq 0\end{aligned}$$

- $\rightarrow$ **Advantage:** Tractable convex problem.
- $\rightarrow$ **Drawback:** Optimal value: $+\infty$
- $\rightarrow$ **Worst than Gurobi solving directly the quadratic problem**

PSD relaxation

$$\max_{x\in\mathbb{R}^n, y\in\mathbb{R}^m_+} \quad d^\top y + c^\top x + \|x\|_2^2$$
$$\text{s.t.} \quad Ax + Ty = b,$$
$$x \geq 0,\ y \geq 0$$

PSD relaxation second order:

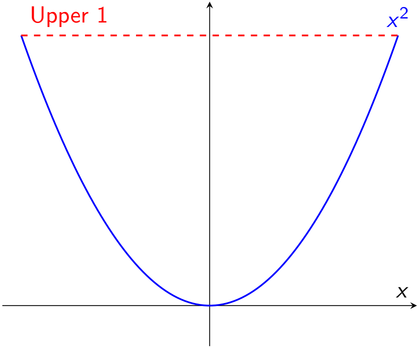
Idea: Consider a large PSD matrix that include product variables between x and y up to order 2.

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

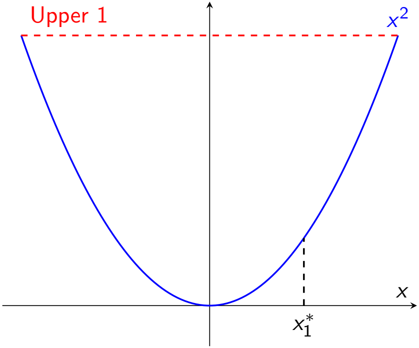
→ **Drawback:** Untractable with solvers like Mosek.

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

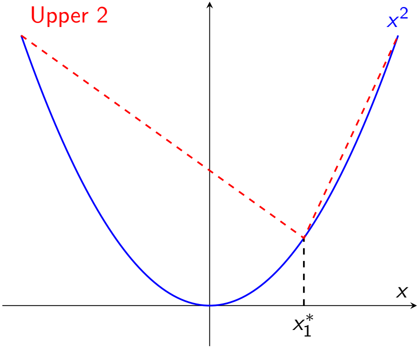
Piecewise linear upper approximation



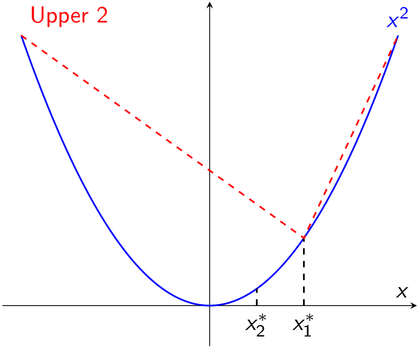
Piecewise linear upper approximation



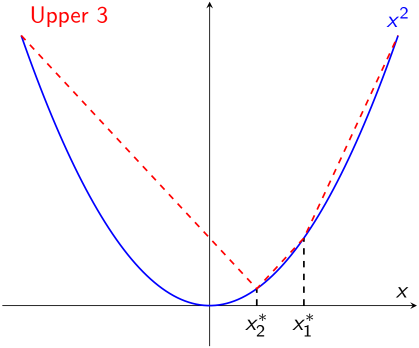
Piecewise linear upper approximation



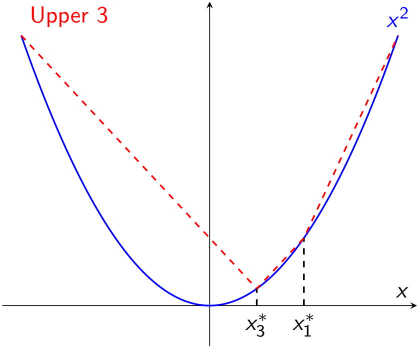
Piecewise linear upper approximation



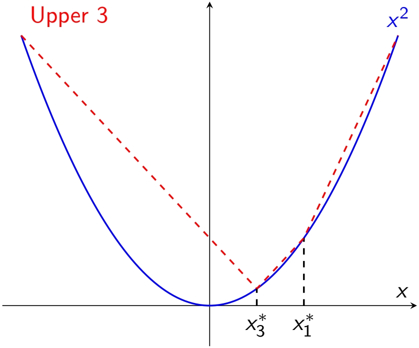
Piecewise linear upper approximation



Piecewise linear upper approximation

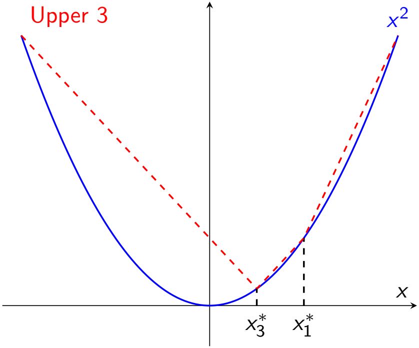


Piecewise linear upper approximation



MILPs.

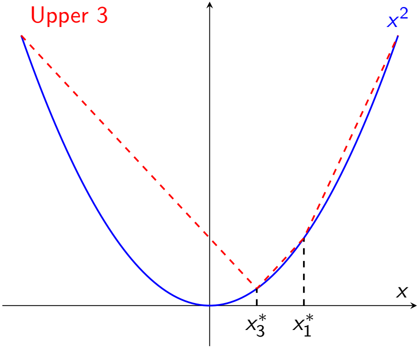
Piecewise linear upper approximation



MILPs.

Number of binary variables increases at each iteration

Piecewise linear upper approximation



MILPs.

Number of binary variables increases at each iteration

Convergence in a few iterations.

Contents

Instances

Source: SMS++/ EDF. Instances with generators (10, 50, 100) and only one random demand (dimension: $T = 24$)

Source IEEE: instance with network constraints:

54 generators and 10 wind farms $\Longrightarrow$ dimension of uncertainty: $10 \times T = 240$

Instances

Source: SMS++/ EDF. Instances with generators (10, 50, 100) and only one random demand (dimension: $T = 24$)

Source IEEE: instance with network constraints:

54 generators and 10 wind farms $\implies$ dimension of uncertainty: $10 \times T = 240$

Computational times

Units	10	50	100	54
ξ dimension	24	24	24	240
FW $(x_0 = 0)$	0.002s	0.002s	0.002s	0.04s
FW MILP $(x_0 \in \{0, M\}^T)$	0.01s	0.05s	0.1s	0.2s
MILP UB	0.1s	0.3s	0.5s	4s
Gurobi	0.2s	0.5s	0.75s	5s

Table: Time comparison (s)

Computational times

Units	10	50	100	54
ξ dimension	24	24	24	240
FW $(x_0 = 0)$	0.002s	0.002s	0.002s	0.04s
FW MILP $(x_0 \in \{0, M\}^T)$	0.01s	0.05s	0.1s	0.2s
MILP UB	0.1s	0.3s	0.5s	4s
Gurobi	0.2s	0.5s	0.75s	5s

Table: Time comparison (s)

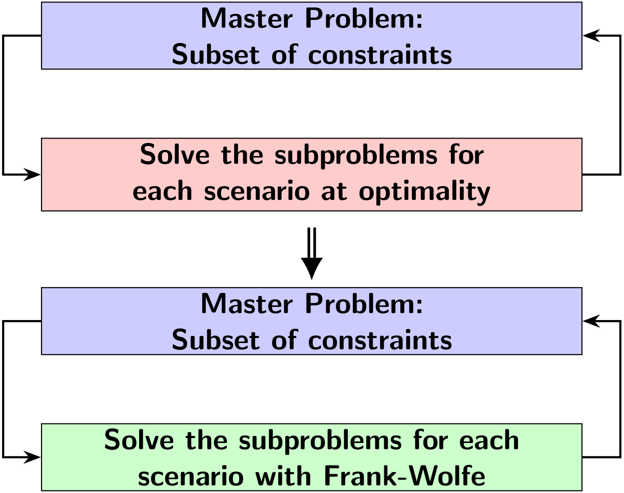
Number of binary variables: O(dimension of ξ $\times$ Number of pieces)

Computational times

Units	10	50	100	54
ξ dimension	24	24	24	240
FW $(x_0 = 0)$	0.002s	0.002s	0.002s	0.04s
FW MILP $(x_0 \in \{0, M\}^T)$	0.01s	0.05s	0.1s	0.2s
MILP UB	0.1s	0.3s	0.5s	4s
Gurobi	0.2s	0.5s	0.75s	5s

Table: Time comparison (s)

Convergence Analysis



Convergence Analysis

Units	10	50	100	54
ξ dimension	24	24	24	240
FW $(x_0 = 0)$	**3s**	**15s**	**40s**	**60s**
FW MILP $(x_0 \in \{0, M\}^T)$	30s	167s	400s	1400s
Gurobi	35s	300s	410s	-

Table: Computational time with 25 scenarios (0.1% gap)

Units	10	50	100	54
ξ dimension	24	24	24	240
FW $(x_0 = 0)$	0.3%	0.2%	0.3%	0.15%
FW MILP $(x_0 \in \{0, M\}^T)$	$< 0.1\%$	$< 0.1\%$	$< 0.1\%$	0.2%
Gurobi	$< 0.1\%$	$< 0.1\%$	$< 0.1\%$	-

Table: Real gap after "convergence" (25 scenarios)

Conclusion

Summary

- Analysis of an NP-hard problem that arises as the oracle problem in a Benders' decomposition for solving a DRO problem.
- Show how the Frank-Wolfe algorithm can be used to progress in the Benders' algorithm.
- Exploration of methods for calculating upper bounds.
- Evaluate the performance of the Adapted Benders' decomposition.