Decision Rules in Adjustable Robust Optimization

Ayşe N. Arslan

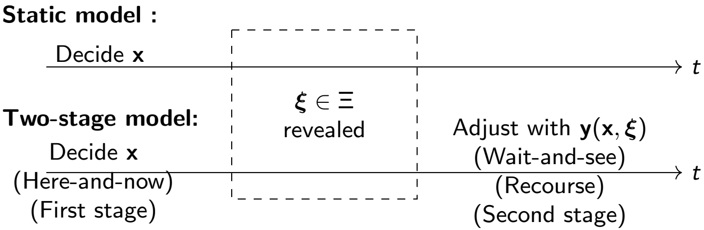
Centre Inria de l'Universite de Bordeaux

Optimization at the Second Level
Tutorial
April 2024-CIRM

Outline

1. Introduction
2. Two-stage RO
   - Affine decision rules
   - Piecewise affine decision rules
   - Piecewise constant decision rules
3. Multi-stage RO
4. Conclusions

Static vs. Two-stage RO



Static vs. Two-stage RO

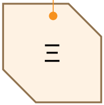
Locate healthcare centers and allocate population

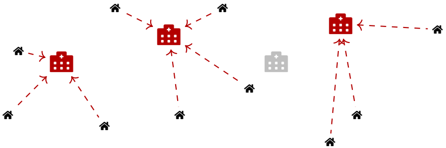
$\mathbf{x}$

$f(\xi, \mathbf{x}) =$ number of untreated patients

ξ

in anticipation of possible disasters





Static model

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\xi\in\Xi} f(\xi, \mathbf{x})$$

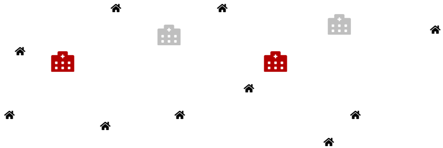
Static vs. Two-stage RO

First-stage decision

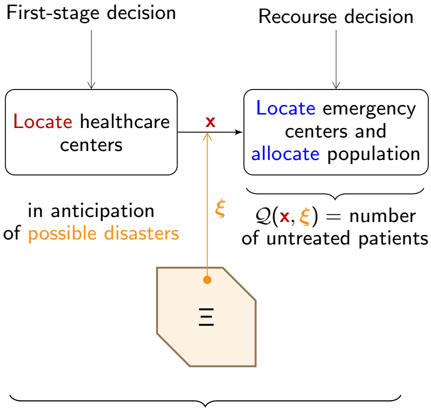
Locate healthcare centers $\xrightarrow{\mathbf{x}}$

in anticipation of possible disasters

Ξ



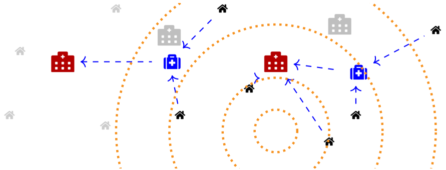
Static vs. Two-stage RO



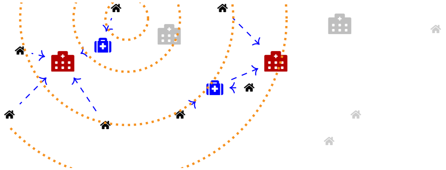
Recourse problem

$$\mathcal{Q}(\mathbf{x}, \xi) = \min \{ f(\xi, \mathbf{x}, \mathbf{y}) \mid \mathbf{y} \in \mathcal{Y}(\mathbf{x}, \xi) \}$$

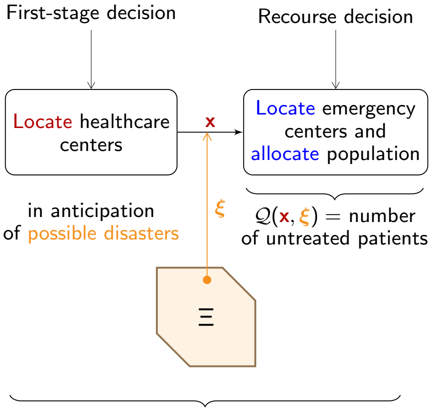
Scenario 1



Scenario 2

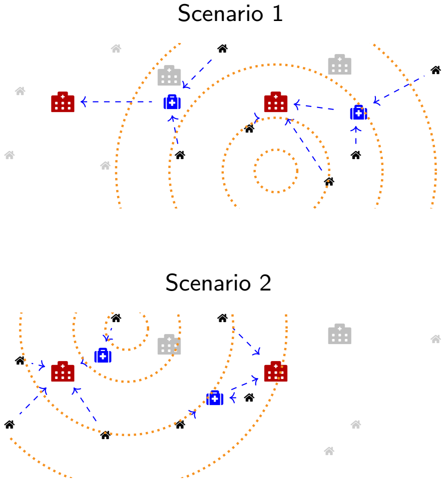


Static vs. Two-stage RO



Two-stage model

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\xi \in \Xi} \mathcal{Q}(\mathbf{x}, \xi)$$



Outline

Formally

- $\mathcal{X} \subseteq \mathbb{R}^{n_x^c} \times \mathbb{Z}^{n_x^i}$: first-stage feasible region ($n_x = n_x^c + n_x^i$)
- $\boldsymbol{\xi}$: uncertain vector
- $\Xi \subseteq \mathbb{R}^{n_\xi}$: uncertainty set, non-empty and compact
- $\mathcal{Y} \subseteq \mathbb{R}^{n_y^c} \times \mathbb{Z}^{n_y^i}$: recourse feasible region ($n_y = n_y^c + n_y^i$)

$$z_{\mathrm{2ARO}} = \min_{\mathbf{x} \in \mathcal{X}} \quad \mathbf{c}^\top \mathbf{x} + \sup_{\boldsymbol{\xi} \in \Xi} \mathcal{Q}(\mathbf{x}, \boldsymbol{\xi}) \tag{2ARO}$$

with

$$\begin{aligned} \mathcal{Q}(\mathbf{x}, \boldsymbol{\xi}) = \min \quad & \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y} \\ \text{s.t.} \quad & \mathbf{W}(\boldsymbol{\xi})\mathbf{y} \leq \mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x} \\ & \mathbf{y} \in \mathcal{Y} \end{aligned}$$

Attention!

- In adjustable models only here-and-now ($\mathbf{x}$) decisions are retained.
- Recourse decisions ($\mathbf{y}$) are used to guide the here-and-now decisions.

Formally

Assumptions

- Ξ is polyhedral.
- $\mathcal{X}$ and $\mathcal{Y}$ contain only linear constraints.
- Data is affine in $\boldsymbol{\xi}$:

$$\mathbf{h}(\boldsymbol{\xi}) = \begin{bmatrix} 1+\xi_1 \\ 1-\xi_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ \xi_1 \\ \xi_2 \end{bmatrix} = \mathbf{H}\tilde{\boldsymbol{\xi}}$$

$$\mathbf{T}(\boldsymbol{\xi}) = \begin{bmatrix} 2+\xi_1-\xi_2 & 1-\xi_1+\xi_2 \\ 3-\xi_2 & 2-\xi_1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ 0 & -1 \end{bmatrix} \xi_1 + \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix} \xi_2$$

$$= \mathbf{T_0} + \sum_{\mathbf{i=1}}^{\mathbf{2}} \mathbf{T_i}\xi_\mathbf{i} \rightarrow \sum_{\mathbf{i=0}}^{\mathbf{2}} \mathbf{T_i}\xi_\mathbf{i} \text{ with } \xi_0 = 1$$

Min-max-min representation

$$\min_{\mathbf{x}\in\mathcal{X}} \quad \mathbf{c}^\top \mathbf{x} + \sup_{\boldsymbol{\xi}\in\Xi} \quad \min_{\mathbf{y}\in\mathcal{Y}} \quad \mathbf{f}(\xi)^\top \mathbf{y}$$
$$\text{s.t.} \quad \mathbf{W}(\xi)\mathbf{y} \leq \mathbf{h}(\xi) - \mathbf{T}(\xi)\mathbf{x}$$

- If $\mathcal{Y}$ is polyhedral and $\mathbf{f}$ and $\mathbf{W}$:
  - Benders-like (supporting hyperplane) approach [Thiele et al., 2009].
  - Constraint-and-column generation [Zeng and Zhao, 2013].
  - Separation problems are bilinear programming problems.

Monolithic (functional) representation

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot)} \quad & \mathbf{c}^\top \mathbf{x} + \sup_{\boldsymbol{\xi}\in\Xi} \quad \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) \\
\text{s.t.} \quad & \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \le \mathbf{h}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi}\in\Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi}\in\Xi
\end{aligned}
$$

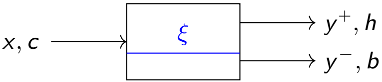
- $\mathbf{y}(\cdot) : \Xi \to \mathcal{Y}$ are functionals to be optimized (also called the policies).

Attention!

- In the following:

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \ge \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi}\in\Xi \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \le \mathbf{h}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi}\in\Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi}\in\Xi
\end{aligned}
$$

Example: Newsvendor problem



$$\xi \in \Xi = \left\{ |\xi - \bar{\xi}| \leq \rho\bar{\xi} \right\} \text{ for } \bar{\xi} \geq 0,\ 0 < \rho < 1$$

- Order quantity x
- Excess quantity $y^+(\xi)$
- Shortage quantity $y^-(\xi)$
- Order cost c, return cost $0 < h$ and shortage cost $b > 0$

Example: Newsvendor problem

- The problem reads:

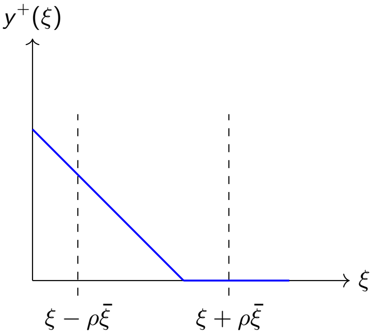
$$\min_{x\geq 0} \quad cx + \max_{\xi\in\Xi} \quad \min_{y^+\geq 0, y^-\geq 0} \quad hy^+ + by^-$$
$$y^+ - y^- = x - \xi$$

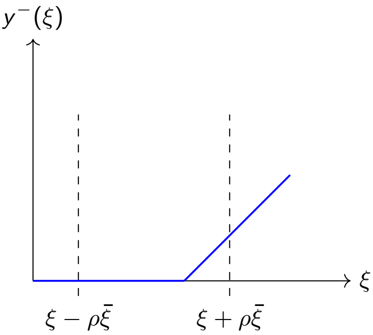
- Optimal policies can be written as:

$$y^+(\xi) = \max\{x-\xi, 0\}$$
$$y^-(\xi) = \max\{0, \xi - x\}.$$

Example: Newsvendor problem

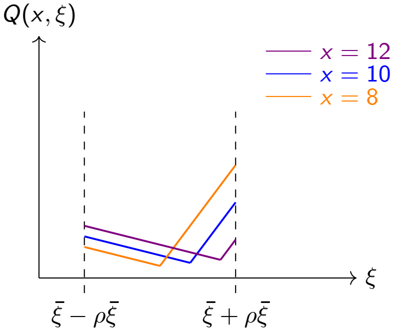
- Optimal policies a function of ξ when $x = 10$:





Example: Newsvendor problem

- For given c, h, b:



- When x is low worst case is $\bar{\xi} + \rho\bar{\xi}$, otherwise worst case is $\bar{\xi} - \rho\bar{\xi}$.
- Optimality is achieved for x such that:

$$Q(x, \bar{\xi} + \rho\bar{\xi}) = Q(x, \bar{\xi} - \rho\bar{\xi})$$

What do we know about optimal policies?[1]

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \boxed{\mathbf{f}(\boldsymbol{\xi})^\top} \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \boxed{\mathbf{W}(\boldsymbol{\xi})} \mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \boxed{\mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y}} && \forall \boldsymbol{\xi} \in \Xi
\end{aligned}
$$

- $\mathcal{Y}$ is polyhedral and $\mathbf{f}$ and $\mathbf{W}$ are deterministic $\rightarrow$ the optimal policy is a (continuous) piecewise affine function of $\boldsymbol{\xi}$ [Bemporad et al., 2003].
- We, however, do not know how many pieces are needed to describe an optimal policy.

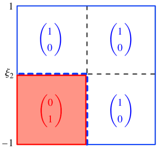
[1] Example from Hanasusanto et al., 2016

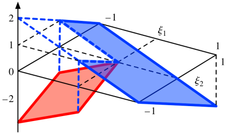
What do we know about optimal policies?[1]

$$\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi} \in \Xi
\end{aligned}$$

- In the general case the optimal policy is not necessarily a continuous function.
- The set of realizations for which a solution is feasible and optimal can be non-convex and non-closed.

$$\begin{aligned}
\sup_{\boldsymbol{\xi}\in\mathbb{R}^2 \mid -1\leq\boldsymbol{\xi}\leq 1} \quad \min_{\mathbf{y}\in\{0,1\}^2} \quad & (\xi_1+\xi_2)(y_2-y_1) \\
\text{s.t.} \quad & y_1 + y_2 = 1 \\
& y_1 \geq \xi_1 \\
& y_1 \geq \xi_2
\end{aligned}$$





[1] Example from Hanasusanto et al., 2016

Idea behind decision rules

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi} \in \Xi
\end{aligned}
$$

Idea

- Restrict the form of $\mathbf{y}(\boldsymbol{\xi})$ to a simple family of functions.
- Optimize within the chosen family.
- Most common restrictions: affine, piecewise affine/constant.

Attention!

- Being restrictions, decision rule approximations may be infeasible.
- If feasible, they provide feasible solutions (primal bounds) to (2ARO).
- They typically improve upon the static robust solution.

Static robust optimization viewed as a decision rule

$$\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi} \in \Xi
\end{aligned}$$

Idea

- Restrict $\mathbf{y}(\boldsymbol{\xi})$ to be a constant function of $\boldsymbol{\xi}$

$$y_i(\boldsymbol{\xi}) = y_i \qquad \boldsymbol{\xi} \in \Xi$$

where $y_i \in \mathcal{Y}$ can be continuous or integer.

Static robust optimization viewed as a decision rule

$$\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}\in\mathcal{Y},\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta && \text{(Stat)}\\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y} & \forall \boldsymbol{\xi} \in \Xi &\\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y} \leq \mathbf{h}(\boldsymbol{\xi}) & \forall \boldsymbol{\xi} \in \Xi &
\end{aligned}$$

- (Stat) is a static robust optimization problem, it is semi-infinite.
- Can be reformulated as a monolithic LP/MILP through LP duality.
- Write:

$$\theta \geq \max_{\boldsymbol{\xi}\in\Xi} \ \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}$$

$$\max_{\boldsymbol{\xi}\in\Xi} \ \mathbf{T}_i(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}_i(\boldsymbol{\xi})\mathbf{y} - \mathbf{h}_i(\boldsymbol{\xi}) \leq 0 \qquad \forall i \in [m]$$

  and reformulate.
- Can be solved through scenario generation as well.

Attention!

- Most decision rule approximations aim to reduce two- and multi-stage models to static (single-stage) models due to the "numerical tractability" of these problems.

Affine decision rules[2]

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi} \in \Xi
\end{aligned}
$$

Assumptions

- $\mathcal{Y}$ is polyhedral
- $\mathbf{f}$ and $\mathbf{W}$ are deterministic

Idea

- Restrict $\mathbf{y}(\boldsymbol{\xi})$ to be an affine function of $\boldsymbol{\xi}$

$$y_i(\boldsymbol{\xi}) = \alpha_{i0} + \boldsymbol{\alpha}_i^\top \boldsymbol{\xi} \quad \forall i \in [n_y] \;\rightarrow\; \mathbf{y} = \mathbf{A}\boldsymbol{\xi}.$$

- Optimize over $\mathbf{A} \in \mathbb{R}^{n_y \times (n_\xi + 1)}$ to obtain the best such function.

[2] Ben-Tal et al., 2004

Affine decision rules[2]

$$
\begin{aligned}
z_{\mathrm{AFF}} = \min_{\mathbf{x}\in\mathcal{X},\mathbf{A}\in\mathbb{R}^{n_y\times(n_\xi+1)},\theta\in\mathbb{R}} \quad & \mathbf{c}^\top\mathbf{x} + \theta && \text{(AFF)} \\
\text{s.t.} \quad & \theta \geq \mathbf{f}^\top\mathbf{A}\boldsymbol{\xi} && \forall\boldsymbol{\xi}\in\Xi \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}\mathbf{A}\boldsymbol{\xi} \leq \mathbf{h}(\boldsymbol{\xi}) && \forall\boldsymbol{\xi}\in\Xi \\
& \mathbf{A}\boldsymbol{\xi} \in \mathcal{Y} && \forall\boldsymbol{\xi}\in\Xi
\end{aligned}
$$

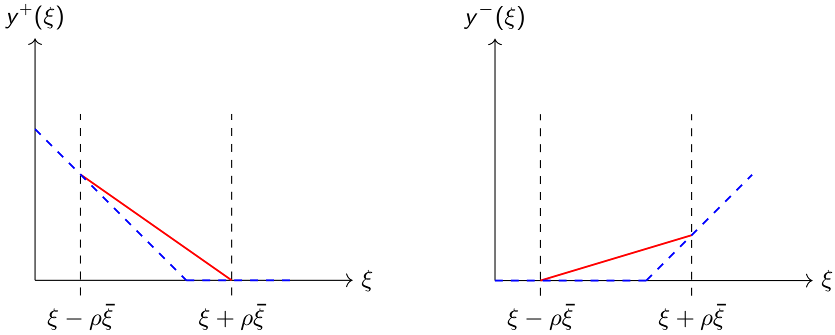
Remark

- (AFF) is a static robust optimization problem, it is semi-infinite.
- Can be solved using either reformulation or scenario generation.
- Polynomial number of variables and constraints added through the decision rule and the resulting reformulation.

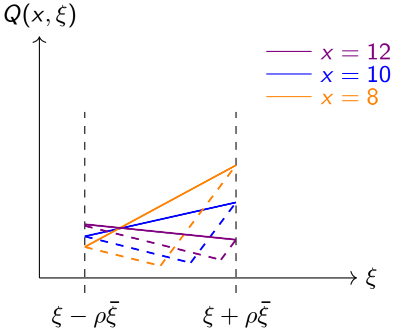
[2] Ben-Tal et al., 2004

Example: Newsvendor (Cont'd)

- Let $y^+(\xi) = \alpha_0^1 + \alpha_1^1 \xi$ and $y^-(\xi) = \alpha_0^2 + \alpha_1^2 \xi$.
- Optimal *affine* recourse quantities as a function of ξ when $x = 10$:



Example: Newsvendor (Cont'd)



- When x is low worst case is $\bar{\xi} + \rho\bar{\xi}$, otherwise worst case is $\bar{\xi} - \rho\bar{\xi}$.
- Optimality is achieved at equality.
- Optimal x value and the worst-case cost is the same as in the exact solution[3].

[3] proved more generally in Bertsimas et al., 2010.

On the quality of affine decision rules

- The quality of a decision rule is measured based on the relative gap:

$$100 \times \frac{|z_{\text{AFF}} - z_{\text{Dual}}|}{|z_{\text{Dual}}|}$$

  where z_{Dual} is a dual bound such that $z_{\text{Dual}} \leq z_{2\text{ARO}}$.
- Dual bounds can be categorized as a priori/a posteriori.
- Some a priori bounds (under $\mathbf{c} \geq 0, \mathbf{f} \geq 0$):
  1. Bertsimas and Goyal (2012) bound the absolute gap between z_{AFF} and $z_{2\text{ARO}}$ as a function of n_ξ.
  2. Bertsimas and Bidkhori (2015) express bounds as functions of the geometric properties of the uncertainty set.
- Some a posteriori bounds:
  1. Kuhn et al. (2011) obtain dual bounds by applying affine decision rules to Lagrangian multipliers of a dual problem.
  2. Hadjiyiannis (2011) propose an a posteriori bounding problem based on discretization.

On the quality of affine decision rules

- Let $\hat{\Xi} \subseteq \Xi$ be a finite subset of realizations.
- Then the following relaxation provides a dual bound:

$$
\begin{aligned}
\min_{\substack{\mathbf{x}\in\mathcal{X},\theta\in\mathbb{R}\\ \mathbf{y}^1,\ldots,\mathbf{y}^{|\hat{\Xi}|}\in\mathcal{Y}}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi}^k)^\top \mathbf{y}^k && \forall k \in [|\hat{\Xi}|] \\
& \mathbf{T}(\boldsymbol{\xi}^k)\mathbf{x} + \mathbf{W}(\boldsymbol{\xi}^k)\mathbf{y}^k \leq \mathbf{h}(\boldsymbol{\xi}^k) && \forall k \in [|\hat{\Xi}|]
\end{aligned}
$$

- But how do we choose $|\hat{\Xi}|$ in a meaningful way?

Hadjiyiannis et al. (2011)

- Solve (AFF) to optimality, let $(\mathbf{x}^*, \mathbf{A}^*, \theta^*)$ be an optimal solution.
- Extract the "binding" scenarios by solving:

$$
\max_{\boldsymbol{\xi}\in\Xi} \ \mathbf{f}^\top \mathbf{A}^* \boldsymbol{\xi} - \theta^*
$$

$$
\max_{\boldsymbol{\xi}\in\Xi} \ \mathbf{T}_i(\boldsymbol{\xi})\mathbf{x}^* + \mathbf{W}_i \mathbf{A}^* \boldsymbol{\xi} - \mathbf{h}_i(\boldsymbol{\xi}) \qquad \forall i \in [m]
$$

- Constitute $\hat{\Xi}$ of binding scenarios.

When affine rules do not work well [4]

- Consider

$$
\begin{aligned}
\min_{x} \quad x + \max_{||\boldsymbol{\xi}||_1 \leq 1} \quad \min_{\boldsymbol{y}} \quad & 0 \\
\text{s.t.} \quad & y_i \geq |\xi_i| & \forall i \in [2] \\
& \sum_{i \in [2]} y_i \leq x
\end{aligned}
$$

- Since $y_i \geq |\xi_i|$, we have that $\sum_{i \in [2]} y_i \geq \sum_{i \in [2]} |\xi_i| = ||\boldsymbol{\xi}||_1$.
- This implies that $x \geq ||\boldsymbol{\xi}||_1 \longrightarrow x \geq 1$.
- The optimal solution is given by $x = 1$, $y_i(\xi) = |\xi_i|$ for $i \in [2]$.

[4] from Chen and Zhang, 2009

When affine rules do not work well [4]

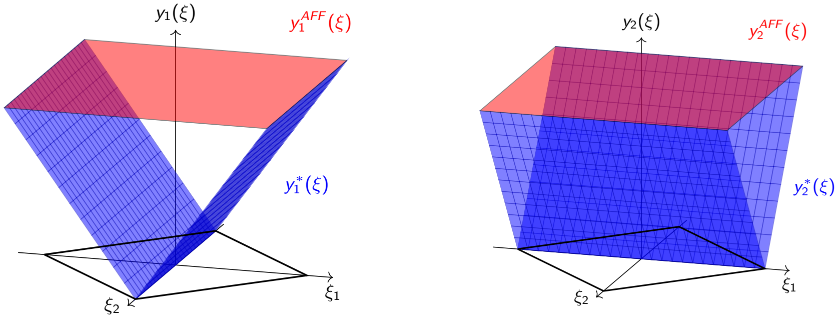
- Consider now the ADR $y_i = \alpha_0^i + \sum_{j\in[2]} \alpha_j^i \xi_j$ for $i \in [2]$.
- We have

$$\begin{aligned} \min_{x,\boldsymbol{\alpha}_0\in\mathbb{R}^2,\mathbf{A}\in\mathbb{R}^{2\times 2}} \quad x + \max_{||\boldsymbol{\xi}||_1\le 1} \quad \min \quad & 0 \\ \text{s.t.} \quad & \alpha_0^i + \sum_{j\in[2]} \alpha_j^i \xi_j \ge |\xi_i| \qquad \forall i \in [2] \\ & \sum_{i\in[2]} \left(\alpha_0^i + \sum_{j\in[2]} \alpha_j^i \xi_j \right) \le x \end{aligned}$$

- The optimal solution is given by $x = 2$, $\alpha_0^i = 1$ for $i \in [2]$, $\alpha_j^i = 0$ for $i, j \in [2]$.
- Corresponding *affine* recourse is $y_i = 1$ for $i \in [2]$.

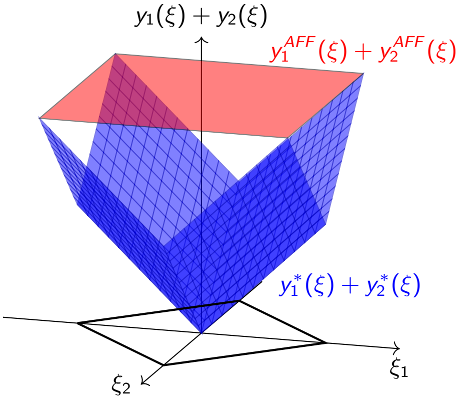
[4] from Chen and Zhang, 2009

When affine rules do not work well [4]



[4] from Chen and Zhang, 2009

When affine rules do not work well [4]



[4] from Chen and Zhang, 2009

Lifted affine decision rules

Splitting-based affine rules [Chen and Zang, 2009]

- Let $\xi_i = \xi_i^+ - \xi_i^-$ for $i \in [n_\xi]$.
- Restrict $\boldsymbol{y}(\boldsymbol{\xi})$ to be an affine function of $\boldsymbol{\xi}^+$ and $\boldsymbol{\xi}^-$

$$y_i(\boldsymbol{\xi}) = \alpha_{i0} + \boldsymbol{\alpha}_i^{+\top}\boldsymbol{\xi}^+ + \boldsymbol{\alpha}_i^{-\top}\boldsymbol{\xi}^- \quad \forall i \in [n_y] \longrightarrow \boldsymbol{y}(\boldsymbol{\xi}) = \mathbf{A}^+\boldsymbol{\xi}^+ + \mathbf{A}^-\boldsymbol{\xi}^-.$$

- Optimize $\mathbf{A}^+ \in \mathbb{R}^{n_y \times (n_\xi+1)}, \mathbf{A}^- \in \mathbb{R}^{n_y \times (n_\xi+1)}$ to find the best such function.

More general idea (Lifted affine rules)

- Let $\Xi^L \subseteq \mathbb{R}^{n_\xi + n'_\xi}$ such that $proj_{\boldsymbol{\xi}}\, \Xi^L = \Xi$.
- Apply the affine decision rule in the lifted uncertainty space.

Remark

- Can be used to represent non-linear functions as linear functions in a lifted space. [Georghiou et al., 2015]
- The lifting operators generally do not conserve the convexity of Ξ.
- Requires intrinsic knowledge of the problem at hand.

Piecewise affine decision rules

$$\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi} \in \Xi
\end{aligned}$$

Idea

- Partition the uncertainty set into K subsets, *i.e.*,

$$\Xi = \bigcup_{k=1}^{K} \Xi_k.$$

- Define one affine policy over each subset

$$\mathbf{y}(\boldsymbol{\xi}) = \mathbf{A}^k \boldsymbol{\xi} \quad \forall k \in [K], \boldsymbol{\xi} \in \Xi_k.$$

Piecewise affine decision rules

$$\mathbf{y}(\boldsymbol{\xi}) = \begin{cases} \mathbf{A}^1\boldsymbol{\xi} & \boldsymbol{\xi} \in \Xi_1 \\ \mathbf{A}^2\boldsymbol{\xi} & \boldsymbol{\xi} \in \Xi_2 \\ \vdots & \\ \mathbf{A}^K\boldsymbol{\xi} & \boldsymbol{\xi} \in \Xi_K \end{cases}$$

Remark

- How many subsets do we create?
- Is the number of subsets fixed? Does it evolve throughout the solution process?
- Are subsets chosen a priori to optimization or not? (implicit/explicit design)?
- How do we handle realizations at the intersection of subsets?

Piecewise affine decision rules [5]

$$\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi} \in \Xi
\end{aligned}$$

Assumptions

- $\mathcal{Y}$ is polyhedral
- **f** and **W** are deterministic

Result

For every continuous piecewise linear function $f(\cdot) : \mathbb{R}^{n_\xi} \to \mathbb{R}$ defined over a polyhedron $\Xi \subseteq \mathbb{R}^{n_\xi}$ there exists $P \in \mathbb{N}_+$, and $\overline{\alpha}_p \in \mathbb{R}^{n_\xi}$, $\underline{\alpha}_p \in \mathbb{R}^{n_\xi}$ for $p \in [P]$ such that:

$$f(\boldsymbol{\xi}) = \max\{\overline{\boldsymbol{\alpha}}_1^\top \boldsymbol{\xi}, \ldots, \overline{\boldsymbol{\alpha}}_P^\top \boldsymbol{\xi}\} - \max\{\underline{\boldsymbol{\alpha}}_1^\top \boldsymbol{\xi}, \ldots, \underline{\boldsymbol{\alpha}}_P^\top \boldsymbol{\xi}\}.$$

[5] Bertsimas and Georghiou, 2015

Piecewise affine decision rules [5]

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi} \in \Xi
\end{aligned}
$$

Assumptions

- $\mathcal{Y}$ is polyhedral
- $\mathbf{f}$ and $\mathbf{W}$ are deterministic

Idea

- Restrict $\mathbf{y}(\boldsymbol{\xi})$ to be a piecewise affine function of $\boldsymbol{\xi}$ with P fixed:

$$
y_i(\boldsymbol{\xi}) = \max\{\overline{\boldsymbol{\alpha}}_1^{i\top}\boldsymbol{\xi}, \ldots, \overline{\boldsymbol{\alpha}}_P^{i\top}\boldsymbol{\xi}\} - \max\{\underline{\boldsymbol{\alpha}}_1^{i\top}\boldsymbol{\xi}, \ldots, \underline{\boldsymbol{\alpha}}_P^{i\top}\boldsymbol{\xi}\} \qquad \forall i \in [n_y]
$$

- Optimize $\overline{\mathbf{A}}_p, \underline{\mathbf{A}}_p \in \mathbb{R}^{n_y \times (n_\xi + 1)}$ for $p = 1, \ldots, P$ to obtain the best such function.
- P controls the level of detail $\to P = 1$ gives ADR.

[5] Bertsimas and Georghiou, 2015

Piecewise affine decision rules [5]

$$
\begin{aligned}
\min_{\substack{\mathbf{x}\in\mathcal{X},\theta\in\mathbb{R}\\ \color{red}{\overline{\mathbf{A}}_1,\dots,\overline{\mathbf{A}}_P}\\ \color{blue}{\underline{\mathbf{A}}_1,\dots,\underline{\mathbf{A}}_P}}} \quad & \mathbf{c}^\top\mathbf{x}+\theta && \text{(P-Aff-}\Xi\text{)}\\
\text{s.t.} \quad & \theta \geq \mathbf{f}^\top\mathbf{y}(\boldsymbol{\xi}) && \forall\boldsymbol{\xi}\in\Xi\\
& \mathbf{W}\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi})-\mathbf{T}(\boldsymbol{\xi})\mathbf{x} && \forall\boldsymbol{\xi}\in\Xi\\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathbb{R}_+^{n_y} && \forall\boldsymbol{\xi}\in\Xi\\
& y_i(\boldsymbol{\xi}) = \max\{\color{red}{\overline{\boldsymbol{\alpha}}_1^{i\top}}\boldsymbol{\xi},\dots,\color{red}{\overline{\boldsymbol{\alpha}}_P^{i\top}}\boldsymbol{\xi}\} - \max\{\color{blue}{\underline{\boldsymbol{\alpha}}_1^{i\top}}\boldsymbol{\xi},\dots,\color{blue}{\underline{\boldsymbol{\alpha}}_P^{i\top}}\boldsymbol{\xi}\} && \forall i\in[n_y],\forall\boldsymbol{\xi}\in\Xi
\end{aligned}
$$

Remark

- (P-Aff) is a static robust optimization problem but it is not convex.
- It can be solved through a scenario generation algorithm.

[5] Bertsimas and Georghiou, 2015

Piecewise affine decision rules [5]

- Let $\hat{\Xi} \subset \Xi$ be a finite subset of realizations, solve:

$$
\begin{aligned}
\min_{\substack{\mathbf{x}\in\mathcal{X},\theta\in\mathbb{R}\\ \overline{\mathbf{A}}_1,\ldots,\overline{\mathbf{A}}_P\\ \underline{\mathbf{A}}_1,\ldots,\underline{\mathbf{A}}_P}} \quad & \mathbf{c}^\top \mathbf{x} + \theta && (\text{P-Aff-}\hat{\Xi})\\
\text{s.t.} \quad & \theta \geq \mathbf{f}^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \hat{\Xi}\\
& \mathbf{W}\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x} && \forall \boldsymbol{\xi} \in \hat{\Xi}\\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi} \in \hat{\Xi}\\
& y_i(\boldsymbol{\xi}) = \max\{\overline{\boldsymbol{\alpha}}_1^{i\top}\boldsymbol{\xi},\ldots,\overline{\boldsymbol{\alpha}}_P^{i\top}\boldsymbol{\xi}\} - \max\{\underline{\boldsymbol{\alpha}}_1^{i\top}\boldsymbol{\xi},\ldots,\underline{\boldsymbol{\alpha}}_P^{i\top}\boldsymbol{\xi}\} && \forall i \in [n_y], \forall \boldsymbol{\xi} \in \Xi
\end{aligned}
$$

  by linearizing non-linear constraints.

- Let $\mathbf{x}^*$, $\overline{\mathbf{A}}^*$, $\underline{\mathbf{A}}^*$, θ^* be optimal for (P-Aff-$\hat{\Xi}$). Is it feasible for (P-Aff-Ξ)?

[5] Bertsimas and Georghiou, 2015

Piecewise affine decision rules [5]

- Solve one separation problem for each uncertain constraint.
- For instance, for the objective function:

$$\begin{aligned} \max_{\boldsymbol{\xi} \in \Xi} \quad & \mathbf{f}^\top \mathbf{y}(\boldsymbol{\xi}) - \theta^* \\ \text{s.t.} \quad & \mathbf{y}_i(\boldsymbol{\xi}) = \max\{\overline{\boldsymbol{\alpha}}_1^{i*\top}\boldsymbol{\xi}, \ldots, \overline{\boldsymbol{\alpha}}_P^{i*\top}\boldsymbol{\xi}\} - \max\{\underline{\boldsymbol{\alpha}}_1^{i*\top}\boldsymbol{\xi}, \ldots, \underline{\boldsymbol{\alpha}}_P^{i*\top}\boldsymbol{\xi}\} \quad \forall i \in [n_y] \end{aligned}$$

  by linearizing the non-linear constraints.
- If any violated scenarios are found $\hat{\Xi} \leftarrow \hat{\Xi} \cup \{\boldsymbol{\xi}^*\}$, otherwise STOP.
- Generated scenarios can be used for the dual bound.

Attention!

- Designs the best continuous piecewise affine function given P.
- Convergence can be slow.

[5] Bertsimas and Georghiou, 2015

Piecewise constant decision rules

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top\mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x} && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi} \in \Xi
\end{aligned}
$$

Idea

- Partition the uncertainty set into K subsets, *i.e.*,

$$\Xi = \bigcup_{k=1}^{K} \Xi_k.$$

- Define one recourse policy over each partition

$$\mathbf{y}(\boldsymbol{\xi}) = \mathbf{y}^k \quad \forall \boldsymbol{\xi} \in \Xi_k.$$

Attention!

- $\mathcal{Y}$ can contain continuous and integer variables.
- This allows handling **f** and **W** as affine functions of $\boldsymbol{\xi}$.

Piecewise constant decision rules

$$\mathbf{y}(\boldsymbol{\xi}) = \begin{cases} \mathbf{y}^1 & \boldsymbol{\xi} \in \Xi_1 \\ \mathbf{y}^2 & \boldsymbol{\xi} \in \Xi_2 \\ \vdots & \\ \mathbf{y}^K & \boldsymbol{\xi} \in \Xi_K \end{cases}$$

Remark

- How many subsets do we create?
- Is the number of subsets fixed? Does it evolve throughout the solution process?
- Are subsets chosen a priori to optimization or not? (implicit/explicit design)?
- How do we handle realizations at the intersection of subsets?

Piecewise constant decision rules

$$\mathbf{y}(\boldsymbol{\xi}) = \begin{cases} \mathbf{y}^1 & \boldsymbol{\xi} \in \Xi_1 \\ \mathbf{y}^2 & \boldsymbol{\xi} \in \Xi_2 \\ \vdots & \\ \mathbf{y}^K & \boldsymbol{\xi} \in \Xi_K \end{cases}$$

Remark

- Uncertainty set partitioning schemes:
  - Subsets are chosen before optimizing over $\mathbf{y}$.
  - Number of subsets increases throughout the algorithmic process.
  - Resulting problem is single-stage (static).
  - Algorithms are designed to create polyhedral subsets.
  - Chooses the worst solution for realizations at the intersection of subsets.
- Finite or $K-$adaptability:
  - Subsets are chosen while optimizing over $\mathbf{y}$.
  - Number of subsets is fixed to a given K.
  - Resulting problem is two-stage.
  - Subsets can be non-closed and non-convex.
  - Chooses the best solution for realizations at the intersection of subsets.

Uncertainty Set Partitioning[6]

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x} && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi} \in \Xi
\end{aligned}
$$

Idea

- Iteratively partition the uncertainty set into polyhedral subsets.
- Define one recourse policy over each subset.

[6] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

Uncertainty Set Partitioning[6]

- When there is no partitioning $\mathbf{y}(\boldsymbol{\xi}) = \mathbf{y} \in \mathcal{Y}$ for $\boldsymbol{\xi} \in \Xi$ $\rightarrow$ static robust problem:

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}\in\mathcal{Y},\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y} & \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y} \leq \mathbf{h}(\boldsymbol{\xi}) & \forall \boldsymbol{\xi} \in \Xi
\end{aligned}
$$

- Let $(\mathbf{x}^*, \mathbf{y}^*, \theta^*)$ be the static optimal solution.
- Extract the "binding" scenarios:

$$
\begin{aligned}
\mathcal{A}_0 &\in \operatorname{argmax}_{\boldsymbol{\xi}\in\Xi} \ \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}^* - \theta^* \\
\mathcal{A}_i &\in \operatorname{argmax}_{\boldsymbol{\xi}\in\Xi} \ \mathbf{T}_i(\boldsymbol{\xi})\mathbf{x}^* + \mathbf{W}_i\mathbf{y}^* - \mathbf{h}_i(\boldsymbol{\xi}) & \forall i \in [m]
\end{aligned}
$$

- Let $\mathcal{A} = \bigcup_{i=0}^{m} \{\mathcal{A}_i\}$.

Attention!

- For a partition to improve the current solution not all elements of $\mathcal{A}$ should be in the same subset.

[6] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

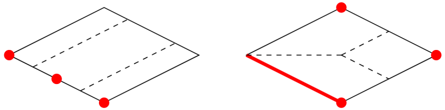
Uncertainty Set Partitioning[6]

- Postek and Den Hertog propose separating with a hyperplane $\boldsymbol{\beta}^\top \boldsymbol{\xi} = \beta$ such that at least one element of $\mathcal{A}$ is on either side, *i.e.*,

$$\exists \boldsymbol{\xi}_i, \boldsymbol{\xi}_j \in \mathcal{A} \text{ s.t. } \boldsymbol{\beta}^\top \boldsymbol{\xi}_i \leq \beta \text{ and } \boldsymbol{\beta}^\top \boldsymbol{\xi}_j \geq \beta.$$

- Bertsimas and Dunning propose Voronoi diagrams:

$$\Xi(\hat{\boldsymbol{\xi}}_i) = \{\boldsymbol{\xi} \in \Xi \mid ||\hat{\boldsymbol{\xi}}_i - \boldsymbol{\xi}|| \leq ||\hat{\boldsymbol{\xi}}_j - \boldsymbol{\xi}|| \quad \forall \hat{\boldsymbol{\xi}}_j \in \mathcal{A}\}$$



[6]Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

Uncertainty Set Partitioning[6]

- Let at iteration r:

$$\Xi = \bigcup_{k \in N_r} \Xi_{rk}.$$

- We solve:

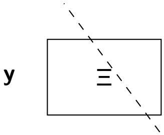
$$\begin{aligned} \min_{\substack{\mathbf{x}\in\mathcal{X},\theta\in\mathbb{R} \\ \mathbf{y}^{(r,1)},\dots,\mathbf{y}^{(r,N_r)}\in\mathcal{Y}}} \quad & \mathbf{c}^\top \mathbf{x} + \theta && \text{(Part-r)} \\ \text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}^{(r,k)} && \forall k \in [N_r], \boldsymbol{\xi} \in \Xi_{rk} \\ & \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y}^{(r,k)} \leq \mathbf{h}(\boldsymbol{\xi}) && \forall k \in [N_r], \boldsymbol{\xi} \in \Xi_{rk} \end{aligned}$$

Remark

- (Part-r) is a static robust optimization problem, it is semi-infinite.
- Under the polyhedral subsets assumption it can be reformulated as a monolithic LP/MILP.
- Binding scenarios $\mathcal{A}_k$ will be identified for each partition Ξ_{rk}.
- Identified binding scenarios can be used to obtain a dual bound.

[6] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

Uncertainty Set Partitioning[6]

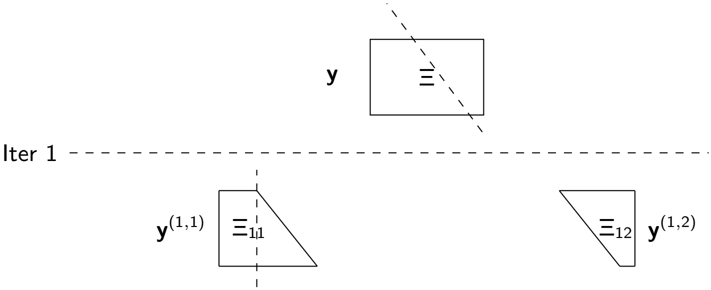


Remark

- Illustration of the algorithm with the partitioning approach of Postek and Den Hertog (2016).

[6] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

Uncertainty Set Partitioning[6]

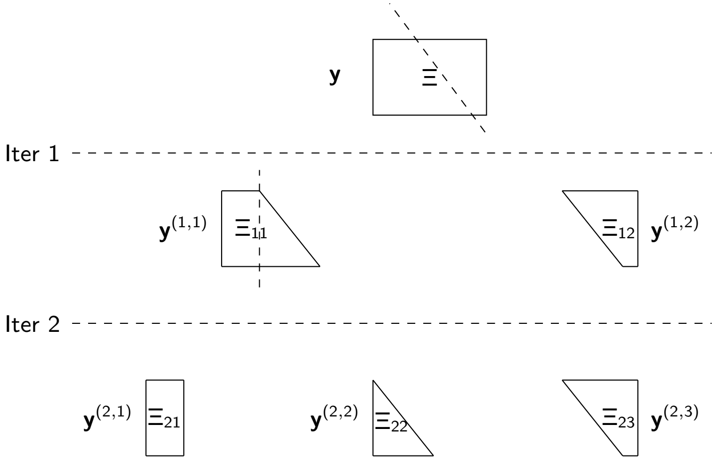


Remark

- Illustration of the algorithm with the partitioning approach of Postek and Den Hertog (2016).

[6] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

Uncertainty Set Partitioning[6]



Remark

- Illustration of the algorithm with the partitioning approach of Postek and Den Hertog (2016).

[6] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

An example where things do not go well[7]

- Consider for $\epsilon \in (0,1]$:

$$
\begin{aligned}
z(\epsilon) = \min_{y_1 \in \{0,1\}, y_2 \in \{0,1\}, z \in \mathbb{R}} \quad & x \\
\text{s.t.} \quad & x \geq y_1(\xi) + y_2(\xi) && \forall \xi \in [0,1] \\
& y_1(\xi) \geq \frac{\epsilon - \xi}{\epsilon} && \forall \xi \in [0,1] \\
& y_2(\xi) \geq \frac{-\epsilon + \xi}{\epsilon} && \forall \xi \in [0,1]
\end{aligned}
$$

- Optimal static solution $x = 2$ with $y_1, y_2 = 1$.
- Optimal solution $x = 1$ with policy:

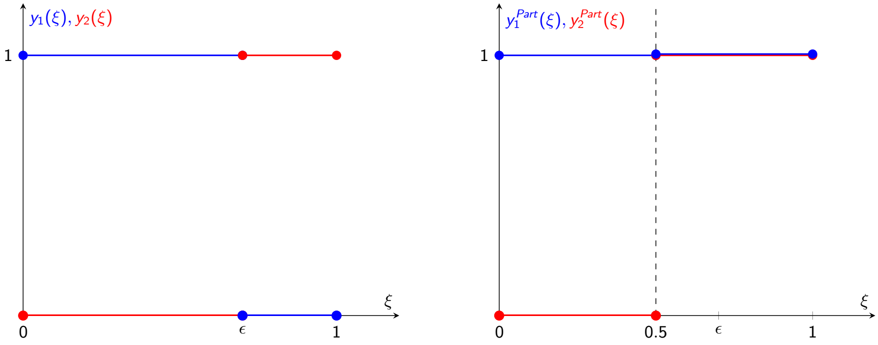
$$
y_1(\xi) = \begin{cases} 1 & 0 \leq \xi \leq \epsilon \\ 0 & \epsilon \leq \xi \leq 1 \end{cases} \qquad y_2(\xi) = \begin{cases} 0 & 0 \leq \xi \leq \epsilon \\ 1 & \epsilon \leq \xi \leq 1 \end{cases}
$$

Attention!

- Partitioning methods will never find an optimal solution or decrease the bound gap unless a subset $[a, b]$ such that $a = \epsilon$ or $b = \epsilon$ is created.

[7] from Bertsimas and Dunning, 2016

An example where things do not go well[7]



[7] from Bertsimas and Dunning, 2016

Combining affine rules with uncertainty set partitioning[8]

$$\begin{aligned}
\min_{\substack{\mathbf{x}\in\mathcal{X},\theta\in\mathbb{R}\\ \mathbf{y}^C(\cdot):\Xi\to\mathbb{R}^{n_C}\\ \mathbf{y}^I(\cdot):\Xi\to\mathbb{Z}^{n_I}}} \quad & \mathbf{c}^\top\mathbf{x}+\theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}_C(\boldsymbol{\xi})^\top \mathbf{y}^C(\boldsymbol{\xi}) + \mathbf{f}_I(\boldsymbol{\xi})^\top \mathbf{y}^I(\boldsymbol{\xi}) && \forall \boldsymbol{\xi}\in\Xi \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}_C(\boldsymbol{\xi})\mathbf{y}^C(\boldsymbol{\xi}) + \mathbf{W}_I(\boldsymbol{\xi})\mathbf{y}^I(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi}\in\Xi \\
& (\mathbf{y}^C(\boldsymbol{\xi}),\mathbf{y}^I(\boldsymbol{\xi})) \in \mathcal{Y} \subseteq \mathbb{R}^{n_c}\times\mathbb{Z}^{n_i} && \forall \boldsymbol{\xi}\in\Xi
\end{aligned}$$

Assumptions

- $\mathbf{f}_C$ and $\mathbf{W}_C$ are deterministic

Idea

- Let $\mathbf{y}^C(\cdot)$ be piecewise affine and $\mathbf{y}^I(\cdot)$ be piecewise constant.
- Iteratively partition the uncertainty set into polyhedral subsets.
- Define one recourse policy over each subset.

[8] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

Combining affine rules with uncertainty set partitioning[8]

- Let at iteration r:

$$\Xi = \bigcup_{k \in N_r} \Xi_{rk}.$$

- We solve:

$$\begin{aligned}
\min_{\substack{\mathbf{x}\in\mathcal{X},\theta\in\mathbb{R} \\ \mathbf{y}^{(r,1)},\ldots,\mathbf{y}^{(r,N_r)}\in\mathbb{Z}^{n_i} \\ \mathbf{A}^{(r,1)},\ldots,\mathbf{A}^{(r,N_r)}\in\mathbb{R}^{q}}} \quad & \mathbf{c}^\top \mathbf{x} + \theta && \text{(Part-r)} \\
\text{s.t.} \quad & \theta \geq \mathbf{f}_I(\boldsymbol{\xi})^\top \mathbf{y}^{(r,k)} + \mathbf{f}_C^\top \mathbf{A}^{(r,k)} \boldsymbol{\xi} && \forall k \in [N_r], \boldsymbol{\xi} \in \Xi_{rk} \\
& \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}_C \mathbf{A}^{(r,k)} \boldsymbol{\xi} + \mathbf{W}_I(\boldsymbol{\xi})\mathbf{y}^{(r,k)} \leq \mathbf{h}(\boldsymbol{\xi}) && \forall k \in [N_r], \boldsymbol{\xi} \in \Xi_{rk} \\
& (\mathbf{A}^{(r,k)}\boldsymbol{\xi}, \mathbf{y}^{(r,k)}) \in \mathcal{Y} && \forall k \in [N_r], \boldsymbol{\xi} \in \Xi_{rk}
\end{aligned}$$

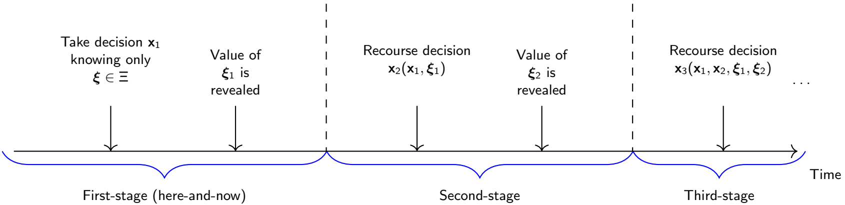
Remark

- (Part-r) is a static robust optimization problem, it is semi-infinite.
- Under the polyhedral subsets assumption it can be reformulated as a monolithic MILP.

[8] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

Outline

Multi-stage RO



Formally

- T: number of decision stages
- $\mathcal{X}_t \subseteq \mathbb{R}^{n_t^c} \times \mathbb{Z}^{n_t^i}$: feasible region at stage $t \in [T]$
- $\Xi \subseteq \mathbb{R}^{n_\xi}$: uncertainty set
- $\boldsymbol{\xi}_t \in \Xi_t := \mathrm{proj}_{\mathbb{R}^{n_{\xi_t}}} \Xi \subseteq \mathbb{R}^{n_{\xi_t}}$: uncertain vector at stage $t \in [T]$
- $\boldsymbol{\xi}^t = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_t) \in \Xi^t := \mathrm{proj}_{\mathbb{R}^{n_{\xi^t}}} \Xi \subseteq \mathbb{R}^{n_{\xi^t}}$: history up to stage $t \in [T]$

$$\begin{aligned}
\min_{\mathbf{x}_1(\cdot),\ldots,\mathbf{x}_T(\cdot),\theta\in\mathbb{R}} \quad & \theta \\
\text{s.t.} \quad & \theta \geq \sum_{t\in[T]} \mathbf{f}_t(\boldsymbol{\xi}^t)^\top \mathbf{x}_t(\boldsymbol{\xi}^t) && \forall \boldsymbol{\xi}^T \in \Xi \\
& \mathbf{T}_t(\boldsymbol{\xi}^t)\mathbf{x}_{t-1}(\boldsymbol{\xi}^{t-1}) + \mathbf{W}_t(\boldsymbol{\xi}^t)\mathbf{x}_t(\boldsymbol{\xi}^t) \leq \mathbf{h}_t(\boldsymbol{\xi}^t) && \forall \boldsymbol{\xi}^T \in \Xi, t \in [T] \setminus \{1\} \\
& \mathbf{x}_t(\boldsymbol{\xi}^t) \in \mathcal{X}_t && \forall \boldsymbol{\xi}^T \in \Xi, t \in [T]
\end{aligned}$$

Attention!

- $\mathbf{x}_t(\cdot) : \Xi^t \to \mathcal{X}_t$ are functionals to be optimized.
- $\mathbf{x}_t$ are functions of $\boldsymbol{\xi}^t$ $\to$ non-anticipativity.

Affine and piecewise affine decision rules[9]

$$\begin{aligned}
\min_{\mathbf{x}_1(\cdot),\ldots,\mathbf{x}_T(\cdot),\theta\in\mathbb{R}} \quad & \theta \\
\text{s.t.} \quad & \theta \geq \sum_{t\in[T]} \mathbf{f}_t(\boldsymbol{\xi}^t)^\top \mathbf{x}_t(\boldsymbol{\xi}^t) && \forall \boldsymbol{\xi}^T \in \Xi \\
& \mathbf{T}_t(\boldsymbol{\xi}^t)\,\mathbf{x}_{t-1}(\boldsymbol{\xi}^{t-1}) + \mathbf{W}_t(\boldsymbol{\xi}^t)\,\mathbf{x}_t(\boldsymbol{\xi}^t) \leq \mathbf{h}_t(\boldsymbol{\xi}^t) && \forall \boldsymbol{\xi}^T \in \Xi, t \in [T] \setminus \{1\} \\
& \mathbf{x}_t(\boldsymbol{\xi}^t) \in \mathcal{X}_t && \forall \boldsymbol{\xi}^T \in \Xi, t \in [T]
\end{aligned}$$

Assumptions

- $\mathcal{X}_t$ is polyhedral for $t \in [T]$
- $\mathbf{f}_t, \mathbf{W}_t$ and $\mathbf{T}_t$ are deterministic for $t \in [T]$

[9] Ben-Tal et al., 2004 and Bertsimas and Georghiou, 2015

Affine and piecewise affine decision rules[9]

$$
\begin{aligned}
\min_{\mathbf{x}_1(\cdot),\ldots,\mathbf{x}_T(\cdot),\theta\in\mathbb{R}} \quad & \theta \\
\text{s.t.} \quad & \theta \geq \sum_{t\in[T]} \mathbf{f}_t(\boldsymbol{\xi}^t)^\top \mathbf{x}_t(\boldsymbol{\xi}^t) && \forall \boldsymbol{\xi}^T \in \Xi \\
& \mathbf{T}_t(\boldsymbol{\xi}^t)\,\mathbf{x}_{t-1}(\boldsymbol{\xi}^{t-1}) + \mathbf{W}_t(\boldsymbol{\xi}^t)\,\mathbf{x}_t(\boldsymbol{\xi}^t) \leq \mathbf{h}_t(\boldsymbol{\xi}^t) && \forall \boldsymbol{\xi}^T \in \Xi, t \in [T]\setminus\{1\} \\
& \mathbf{x}_t(\boldsymbol{\xi}^t) \in \mathcal{X}_t && \forall \boldsymbol{\xi}^T \in \Xi, t \in [T]
\end{aligned}
$$

Idea

- Restrict $\mathbf{x}_t(\boldsymbol{\xi}^t)$ to be a (continuous) (piecewise) affine function of $\boldsymbol{\xi}^t$

$$
\mathbf{x}_{it}(\boldsymbol{\xi}^t) = \max\{\overline{\boldsymbol{\alpha}}^{i\top}_{1t}\boldsymbol{\xi}^t, \ldots, \overline{\boldsymbol{\alpha}}^{i\top}_{Pt}\boldsymbol{\xi}^t\} - \max\{\underline{\boldsymbol{\alpha}}^{i\top}_{1t}\boldsymbol{\xi}^t, \ldots, \underline{\boldsymbol{\alpha}}^{i\top}_{Pt}\boldsymbol{\xi}^t\} \quad \forall i \in [n_{x_t}]
$$

  where $\overline{\boldsymbol{\alpha}}, \underline{\boldsymbol{\alpha}} \in \mathbb{R}^{n_{\xi^t}}$.

- We may also define an *information basis*, *e.g.*, $I_t = \{t, t-1\}$:

$$
\overline{\boldsymbol{\alpha}}, \underline{\boldsymbol{\alpha}} \in \mathbb{R}^{n_{\xi^t}} \longrightarrow \overline{\boldsymbol{\alpha}}, \underline{\boldsymbol{\alpha}} \in \mathbb{R}^{\xi_t + \xi_{t-1}}
$$

[9] Ben-Tal et al., 2004 and Bertsimas and Georghiou, 2015

Uncertainty set partitioning [10]

$$\begin{aligned}
\min_{\mathbf{x}_1(\cdot),\ldots,\mathbf{x}_T(\cdot),\theta\in\mathbb{R}} \quad & \theta \\
\text{s.t.} \quad & \theta \geq \sum_{t\in[T]} \mathbf{f}_t(\boldsymbol{\xi}^t)^\top \mathbf{x}_t(\boldsymbol{\xi}^t) && \forall \boldsymbol{\xi}^T \in \Xi \\
& \mathbf{T}_t(\boldsymbol{\xi}^t)\mathbf{x}_{t-1}(\boldsymbol{\xi}^{t-1}) + \mathbf{W}_t(\boldsymbol{\xi}^t)\mathbf{x}_t(\boldsymbol{\xi}^t) \leq \mathbf{h}_t(\boldsymbol{\xi}^t) && \forall \boldsymbol{\xi}^T \in \Xi, t \in [T] \setminus \{1\} \\
& \mathbf{x}_t(\boldsymbol{\xi}^t) \in \mathcal{X}_t && \forall \boldsymbol{\xi}^T \in \Xi, t \in [T]
\end{aligned}$$

Idea

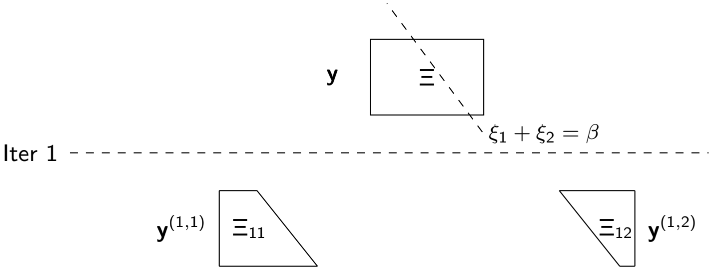
- Progressively partition the uncertainty set into polyhedral subsets making sure that the non-anticipativity constraints are respected.
- Define one recourse policy over each subset.

Attention!

- Can be mixed with affine decision rules for continuous variables under classical assumptions.

[10] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

Uncertainty set partitioning [10]



Attention!

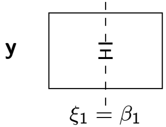
- If no additional constraints are imposed $\mathbf{y}^{(1,1)}$ and $\mathbf{y}^{(1,2)}$ can have different first- and second-stage components.
- This violates non-anticipativity.

Idea

- Keep track of the information used for partitioning.
- Impose non-anticipativity constraints explicitly.

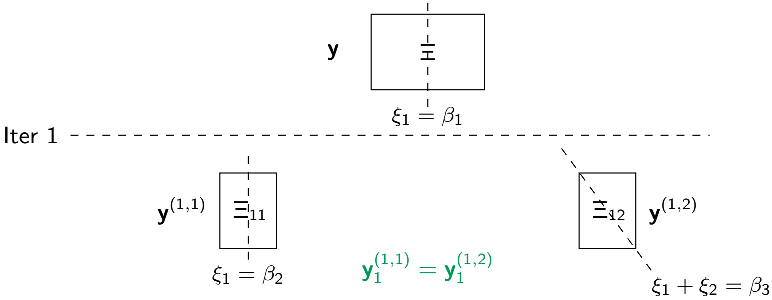
[10] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

Uncertainty set partitioning [10]



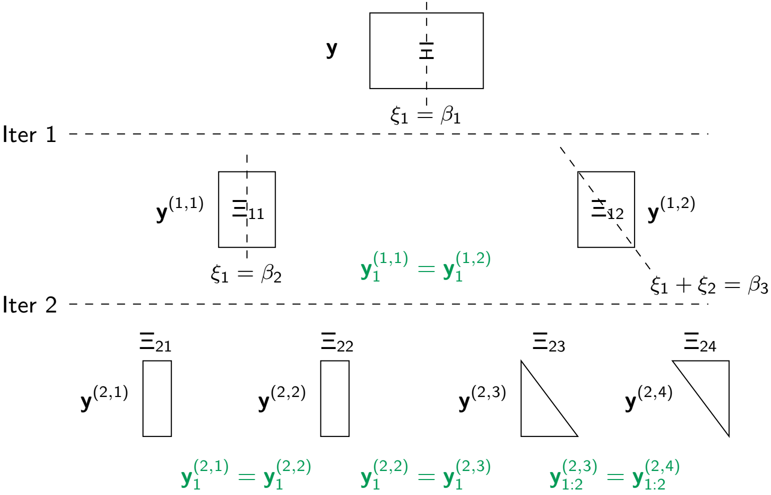
[10] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

Uncertainty set partitioning [10]



[10] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

Uncertainty set partitioning [10]



[10] Postek and Den Hertog, 2016 and Bertsimas and Dunning, 2016

Outline

1. Introduction
2. Two-stage RO
   - Affine decision rules
   - Piecewise affine decision rules
   - Piecewise constant decision rules
3. Multi-stage RO
4. Conclusions

Conclusions

- Decision rule approximations offer many possibilities for two- and multi-stage robust optimization problems.
- Here, we talked about:
  - In the continuous recourse case $\rightarrow$ affine and piecewise affine rules.
  - In the discrete recourse case $\rightarrow$ uncertainty set partitioning (piecewise constant rules).
- There are many other decision rule approximations, e.g.,:
  - $K-$adaptability, two-stage decisions rules...
- In some very special cases affine decision rules can be shown to be optimal. In other cases, their optimality gap can be bounded (empirically or theoretically).
- More work in the applied context and numerical improvements are needed in order to successfully use these results in practice.
- More work is needed on dual bounding problems in order to better evaluate the quality of obtained solutions.

Thank you for your attention!

Some references

- Ben-Tal, A., Goryashko, A., Guslitzer, E., and Nemirovski, A. (2004). Adjustable robust solutions of uncertain linear programs. Mathematical programming, 99(2), 351-376.
- Bertsimas, D., and Bidkhori, H. (2015). On the performance of affine policies for two-stage adaptive optimization: a geometric perspective. Mathematical Programming, 153(2), 577-594.
- Bertsimas, D., and de Ruiter, F. J. (2016). Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds. INFORMS Journal on Computing, 28(3), 500-511.
- Bertsimas, D., and Dunning, I. (2016). Multistage robust mixed-integer optimization with adaptive partitions. Operations Research, 64(4), 980-998.
- Bertsimas, D. and Georghiou, A. (2015). Design of near optimal decision rules in multistage adaptive mixed-integer optimization. Operations Research, 63(3), 610-627.
- Bertsimas, D., and Goyal, V. (2012). On the power and limitations of affine policies in two-stage adaptive optimization. Mathematical programming, 134(2), 491-531.
- Bertsimas, D., Iancu, D. A., and Parrilo, P. A. (2010). Optimality of affine policies in multistage robust optimization. Mathematics of Operations Research, 35(2), 363-394.

Some references

- Chen, X., and Zhang, Y. (2009). Uncertain linear programs: Extended affinely adjustable robust counterparts. Operations Research, 57(6), 1469-1482.
- Georghiou, A., Wiesemann, W., and Kuhn, D. (2015). Generalized decision rule approximations for stochastic programming via liftings. Mathematical Programming, 152, 301-338.
- Hadjiyiannis, M. J., Goulart, P. J., and Kuhn, D. (2011, December). A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. In 2011 50th IEEE Conference on Decision and Control and European Control Conference (pp. 7386-7391). IEEE.
- Hanasusanto, G. A., Kuhn, D., and Wiesemann, W. (2015). K-adaptability in two-stage robust binary programming. Operations Research, 63(4), 877-891.
- Postek, K., and den Hertog, D. (2016). Multistage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. INFORMS Journal on Computing, 28(3), 553-574.
- Subramanyam, A., Gounaris, C. E., and Wiesemann, W. (2020). K-adaptability in two-stage mixed-integer robust optimization. Mathematical Programming Computation, 12(2), 193-224.

A "dual" formulation and affine decision rules[11]

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot)} \quad & \mathbf{c}^\top \mathbf{x} + \sup_{\boldsymbol{\xi}\in\Xi} \boxed{\mathbf{f}(\boldsymbol{\xi})^\top} \mathbf{y}(\boldsymbol{\xi}) & \\
\text{s.t.} \quad & \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \boxed{\mathbf{W}(\boldsymbol{\xi})}\, \mathbf{y}(\boldsymbol{\xi}) \le \mathbf{h}(\boldsymbol{\xi}) & \forall \boldsymbol{\xi} \in \Xi \\
& \boxed{\mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y}} & \forall \boldsymbol{\xi} \in \Xi
\end{aligned}
$$

Assumptions

- $\mathcal{Y}$ is polyhedral
- $\mathbf{f}$ and $\mathbf{W}$ are deterministic

Idea

- Obtain a "dual" formulation through successive application of LP duality.
- Apply affine decision rules on this "dual" formulation.

[11] Bertsimas and de Ruiter, 2016

A "dual" formulation and affine decision rules[11]

Result

- The two-stage robust optimization problem,

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot):\Xi\to\mathbb{R}^{n_y}_+} \quad & \mathbf{c}^\top\mathbf{x} + \mathbf{f}^\top\mathbf{y}(\boldsymbol{\xi}) \\ \text{s.t.} \quad & \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y}(\boldsymbol{\xi}) \le \mathbf{H}\boldsymbol{\xi} \qquad \forall \boldsymbol{\xi} \in \Xi = \{\boldsymbol{\xi} \in \mathbb{R}^{n_\xi}_+ \mid \mathbf{D}\boldsymbol{\xi} \le \mathbf{d}\} \end{aligned}$$

is equivalent to the "dual" two-stage robust optimization problem,

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{X},\mathbf{u}(\cdot):\Pi\to\mathbb{R}^{m_\xi}_+} \quad & \mathbf{c}^\top\mathbf{x} + \max_{\boldsymbol{\pi}\in\Pi} \quad (-\mathbf{T}\mathbf{x})^\top\boldsymbol{\pi} + \mathbf{d}^\top\mathbf{u}(\boldsymbol{\pi}) \\ \text{s.t.} \quad & \mathbf{D}^\top\mathbf{u}(\boldsymbol{\pi}) \ge \mathbf{H}^\top\boldsymbol{\pi} \qquad \forall \boldsymbol{\pi} \in \Pi = \{\boldsymbol{\pi} \in \mathbb{R}^{m_y}_- \mid \mathbf{W}^\top\boldsymbol{\pi} \le \mathbf{f}\}. \end{aligned}$$

Idea

- Restrict $\mathbf{u}(\boldsymbol{\pi})$ to be an affine function of $\boldsymbol{\pi}$

$$u_i(\boldsymbol{\pi}) = \alpha_{i0} + \boldsymbol{\alpha}_i^\top\boldsymbol{\pi} \quad \forall i \in [n_y] \;\rightarrow\; \mathbf{u} = \mathbf{A}\boldsymbol{\pi}.$$

- Optimize over $\mathbf{A} \in \mathbb{R}^{m_\xi \times (m_y+1)}$ to obtain the best such function.

[11] Bertsimas and de Ruiter, 2016

A "dual" formulation and affine decision rules[11]

$$\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{A}\in\mathbb{R}^{m_\xi\times(m_y+1)}} \quad & \mathbf{c}^\top\mathbf{x} + \max_{\boldsymbol{\pi}\in\Pi} \quad (-\mathbf{T}\mathbf{x})^\top\boldsymbol{\pi} + \mathbf{d}^\top\mathbf{A}\boldsymbol{\pi} && \text{(D-AFF)} \\
\text{s.t.} \quad & \mathbf{D}^\top\mathbf{A}\boldsymbol{\pi} \geq \mathbf{H}\boldsymbol{\pi} && \forall\boldsymbol{\pi}\in\Pi \\
& \mathbf{A}\boldsymbol{\pi} \in \mathbb{R}_+^{m_\xi} && \forall\boldsymbol{\pi}\in\Pi
\end{aligned}$$

Attention!

- The optimal solutions of (AFF) and (D-AFF) give the same optimal $\mathbf{x}^*$ solution (and the same value).
- Numerical performance can be different depending on dimensions $n_y \times (n_\xi + 1)$ vs $m_\xi \times (m_y + 1)$.
- Dual bounds can be improved using the "dual" binding scenarios.

[11] Bertsimas and de Ruiter, 2016

K-adaptability[12]

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x} && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi} \in \Xi
\end{aligned}
$$

Idea

- Prepare K recourse policies $\mathbf{y}^1, \ldots, \mathbf{y}^K \in \mathcal{Y}$ in advance in the first-stage.
- Implement the best of $\mathbf{y}^1, \ldots, \mathbf{y}^K$ in the second-stage.

Attention!

- Implicitly designs the best piecewise constant functions for $\mathbf{y}^1, \ldots, \mathbf{y}^K \in \mathcal{Y}$.
- Can be mixed with affine decision rules for continuous variables under classical assumptions.

[12] Subramanyam et al., 2019

K-adaptability[12]

$$\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}^1,\dots,\mathbf{y}^K\in\mathcal{Y}} \quad \mathbf{c}^\top\mathbf{x} + \sup_{\boldsymbol{\xi}\in\Xi} \quad \min_{k\in[K]|\mathbf{W}(\boldsymbol{\xi})\mathbf{y}^k\leq\mathbf{h}(\boldsymbol{\xi})-\mathbf{T}(\boldsymbol{\xi})\mathbf{x}} \quad \mathbf{f}(\boldsymbol{\xi})^\top\mathbf{y}^k \qquad \text{(K-Adapt)}$$

Idea

- Write as a disjunctive semi-infinite programming problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{X},\mathbf{y}^1,\dots,\mathbf{y}^K\in\mathcal{Y}} \quad & \theta \\ \text{s.t.} \quad & \vee_{k\in[K]} \begin{bmatrix} \mathbf{c}^\top\mathbf{x} + \mathbf{f}(\boldsymbol{\xi})^\top\mathbf{y}^k \leq \theta \\ \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y}^\mathbf{k} \leq \mathbf{h}(\boldsymbol{\xi}) \end{bmatrix} \qquad \forall\boldsymbol{\xi}\in\Xi \end{aligned}$$

- Solve using scenario generation coupled with branch-and-bound.

[12] Subramanyam et al., 2019

K-adaptability[12]

- Let $\hat{\Xi} \subset \Xi$ a finite subset of Ξ.
- Partition $\hat{\Xi} = \cup_{k\in[K]} \hat{\Xi}_k$ such that:

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{X},\mathbf{y}^1,\dots,\mathbf{y}^K\in\mathcal{Y}} \quad & \theta && \text{(K-Adapt-}\hat{\Xi}\text{)} \\ \text{s.t.} \quad & \theta \geq \mathbf{c}^\top \mathbf{x} + \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}^k && \forall k \in [K], \forall \boldsymbol{\xi} \in \hat{\Xi}_k \\ & \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}(\boldsymbol{\xi})\mathbf{y}^\mathbf{k} \leq \mathbf{h}(\boldsymbol{\xi}) && \forall k \in [K], \forall \boldsymbol{\xi} \in \hat{\Xi}_k \end{aligned}$$

- Let $(\mathbf{x}^*, \mathbf{y}^{1*}, \dots, \mathbf{y}^{K*}, \theta^*)$ be an optimal solution.

Attention!

- Given $\hat{\Xi}$, finding the optimal subsets $\hat{\Xi}_k$, $k \in [K]$ at the same time as $\mathbf{y}^1, \dots, \mathbf{y}^K$ is NP-Hard.

[12] Subramanyam et al., 2019

K-adaptability[12]

- Separate $(\mathbf{x}^*, \mathbf{y}^{1*}, \ldots, \mathbf{y}^{K*}, \theta^*)$

$$
\begin{aligned}
\max \quad & \zeta \\
\text{s.t.} \quad & \sum_{\ell=0}^{L} z_{k\ell} = 1 & \forall k \in [K] \\
& z_{k0} = 1 \implies \zeta \leq \mathbf{c}^\top \mathbf{x}^* + \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}^{k*} - \theta^* & \forall k \in [K] \\
& z_{k\ell} = 1 \implies \zeta \leq \mathbf{T}_\ell(\boldsymbol{\xi})\mathbf{x}^* + \mathbf{W}_\ell(\boldsymbol{\xi})\mathbf{y}^{\mathbf{k}*} - \mathbf{h}_\ell(\boldsymbol{\xi}) & \forall \ell \in [L], k \in [K] \\
& \boldsymbol{\xi} \in \Xi \\
& \mathbf{z} \in \{0,1\}^{[K]\times(L+1)}
\end{aligned}
$$

- If $\zeta > 0$, create K branches

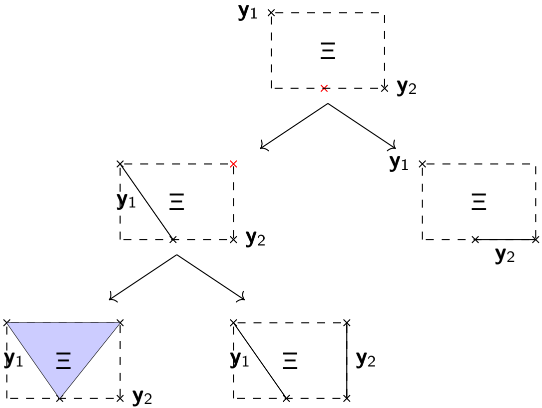
$$\{\Xi_1, \ldots, \Xi_k \cup \{\boldsymbol{\xi}^*\}, \ldots, \Xi_K\} \text{ for } k \in [K].$$

Attention!

- The separation problem can be solved as $(L+1)^K$ linear programs by enumerating over the vector $\mathbf{z}$.

[12]Subramanyam et al., 2019

K-adaptability[12]



Attention!

- The branch-and-bound tree implicitly enumerates over all possible partitions of generated scenarios.

[12]Subramanyam et al., 2019

Binary K-adaptability [13]

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}(\cdot),\theta\in\mathbb{R}} \quad & \mathbf{c}^\top \mathbf{x} + \theta \\
\text{s.t.} \quad & \theta \geq \mathbf{f}(\boldsymbol{\xi})^\top \mathbf{y}(\boldsymbol{\xi}) && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{W}(\boldsymbol{\xi})\mathbf{y}(\boldsymbol{\xi}) \leq \mathbf{h}(\boldsymbol{\xi}) - \mathbf{T}(\boldsymbol{\xi})\mathbf{x} && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{y}(\boldsymbol{\xi}) \in \mathcal{Y} && \forall \boldsymbol{\xi} \in \Xi
\end{aligned}
$$

Assumptions

- $\mathcal{Y} \subseteq \{0,1\}^{n_y}$
- **W**, **T** and **h** are deterministic

Idea

- Prepare K recourse policies $\mathbf{y}^1, \ldots, \mathbf{y}^K \in \mathcal{Y}$ in advance in the first-stage.
- Implement the best of $\mathbf{y}^1, \ldots, \mathbf{y}^K$ in the second-stage.

[13] Hanasusanto et al., 2015

Binary K-adaptability [13]

- We obtain:

$$\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}^k\in\mathcal{Y}} \quad \mathbf{c}^\top\mathbf{x} + \max_{\boldsymbol{\xi}\in\Xi} \quad \min_{k\in[K]} \quad \boldsymbol{\xi}^\top\mathbf{F}\mathbf{y}^k$$
$$\mathbf{W}\mathbf{y}^k \leq \mathbf{h} - \mathbf{T}\mathbf{x}$$

- The linking constraints can be moved to the first-stage:

$$\min_{\mathbf{x}\in\mathcal{X},\mathbf{y}^k\in\mathcal{Y}} \quad \mathbf{c}^\top\mathbf{x} \quad + \max_{\boldsymbol{\xi}\in\Xi} \quad \min_{k\in[K]} \quad \boldsymbol{\xi}^\top\mathbf{F}\mathbf{y}^k$$
$$\mathbf{W}\mathbf{y}^k \leq \mathbf{h} - \mathbf{T}\mathbf{x} \quad \forall k \in [K]$$

- Can be reformulated after writing:

$$\min_{\substack{\mathbf{x}\in\mathcal{X},\mathbf{y}^1,\dots,\mathbf{y}^K\in\mathcal{Y}\\ \mathbf{W}\mathbf{y}^k\leq\mathbf{h}-\mathbf{T}\mathbf{x},k\in[K]}} \mathbf{c}^\top\mathbf{x} + \quad \max_{\boldsymbol{\xi}\in\Xi} \quad \max_{\boldsymbol{\omega}} \left\{\boldsymbol{\omega} : \boldsymbol{\omega} \leq \boldsymbol{\xi}^\top\mathbf{F}\mathbf{y}^k, k \in [K]\right\}$$

[13] Hanasusanto et al., 2015

Binary K-adaptability [13]

$$\iff \quad \min_{\substack{\mathbf{x}\in\mathcal{X},\mathbf{y}^1,\ldots,\mathbf{y}^K\in\mathcal{Y}\\ \mathbf{W}\mathbf{y}^k\leq\mathbf{h}-\mathbf{T}\mathbf{x},k\in[K]}} \mathbf{c}^\top\mathbf{x} + \max_{\boldsymbol{\xi},\boldsymbol{\omega}}\left\{\boldsymbol{\omega} \;:\; \boldsymbol{\xi}\in\Xi,\; \boldsymbol{\omega}\leq\boldsymbol{\xi}^\top\mathbf{F}\mathbf{y}^k,\; k\in[K]\right\}$$

$$\underset{\Xi=\{\boldsymbol{\xi}\geq 0,\mathbf{D}\boldsymbol{\xi}\leq\mathbf{d}\}}{\overset{Dual}{\iff}} \begin{cases} \min & \mathbf{c}^\top\mathbf{x} + \sum_{l\in[m_2]} d_l\boldsymbol{\alpha}_l \\ \text{s.t.} & \mathbf{D_j}^\top\boldsymbol{\alpha} \geq \gamma_k\mathbf{F}\sum_{k\in[K]}\mathbf{y}_j^k \quad \forall j\in[n_2] \\ & \sum_{k\in[K]}\gamma_k = 1 \\ & \mathbf{W}\mathbf{y}^k \leq \mathbf{h}-\mathbf{T}\mathbf{x} \quad \forall k\in[K] \\ & \mathbf{y}^1,\ldots,\mathbf{y}^K\in\mathcal{Y} \\ & \mathbf{x}\in\mathcal{X},\boldsymbol{\gamma},\boldsymbol{\alpha}\geq 0 \end{cases}$$

Remark

- Bi-linear terms can be linearized using the McCormick envelope since $\mathbf{y}$ is assumed to be binary.

[13] Hanasusanto et al., 2015