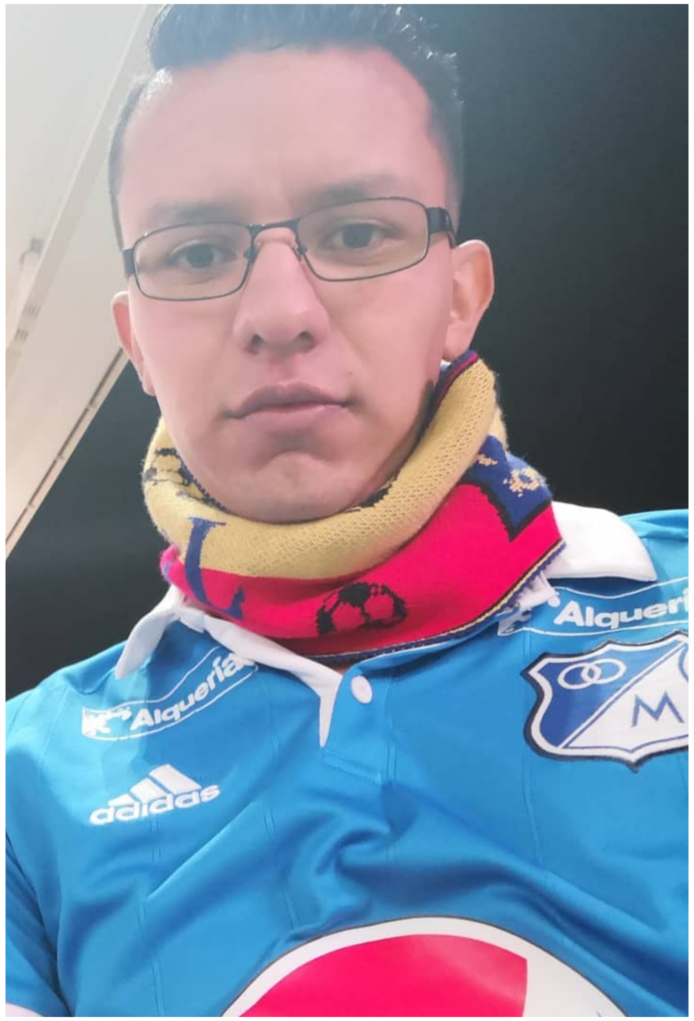
# *Benders Adaptive-Cuts Method for Two-Stage Stochastic Programs*

**Eduardo Moreno**
Universidad Adolfo Ibáñez, Santiago, Chile
(Currently at ESSEC Business School / Google Research, Paris, France)

# Joint work with...

Daniel Espinoza
Google Research



Cristian Ramirez-Pico
PhD Student - UAI

Ivana Ljubic
ESSEC Paris

# Optimization under uncertainty

# Two-stage stochastic problems

- First-stage decisions $x$ (before the uncertainty is revealed)
- Second-stage decisions $y$ (*a posteriori* recourse actions)

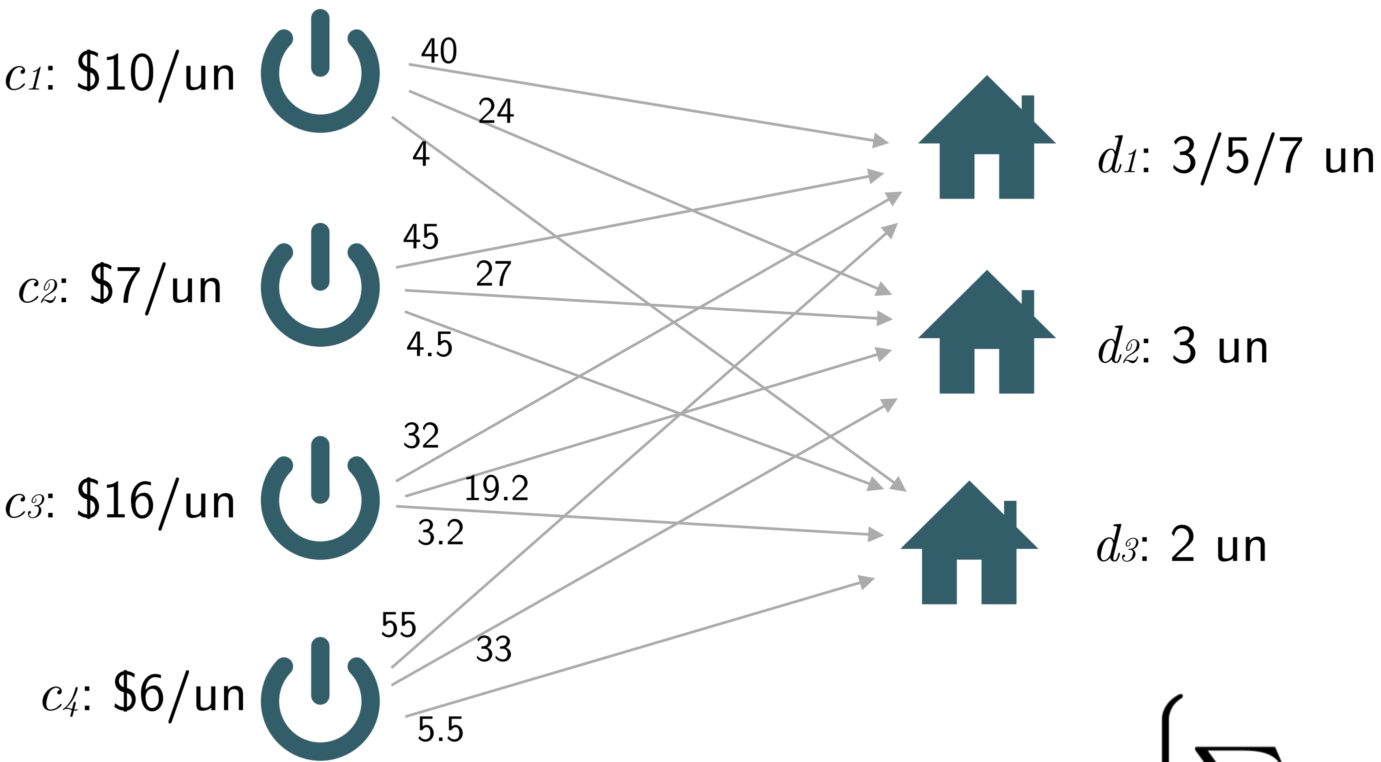
$$\min cx$$
$$Ax = b$$
$$x \geq 0$$

$$\rightarrow$$

$$\min c^\intercal x + \mathbb{E}\left[\mathcal{Q}(x,\xi)\right]$$
$$Ax = b$$
$$x \geq 0$$

$$\mathcal{Q}(x,\xi) := \min q^\xi y$$
$$T^\xi x + W^\xi y = h^\xi$$
$$x \geq 0$$

# Example: LandS (1988)




Problem: To decide the installed capacity given an uncertain future demand

Budget $b$ : $120
Min. Capac $m$ : 12

$d_1$: 3 (30%), 5 (40%), 7 (30%)

$$\min_{x \geq 0} \left\{ \sum_{i \in \mathbb{I}} c_i x_i + \mathbb{E}\left[\mathcal{Q}(x, \xi)\right] : \sum_{i \in \mathbb{I}} x_i \geq m, \sum_{i \in \mathbb{I}} c_i x_i \leq b \right\}$$

$$\mathcal{Q}(x, \xi) := \min_{y \geq 0} \sum_{i \in \mathbb{I}} \sum_{j \in \mathbb{J}} f_{ij} y_{ij}$$

$$\text{s.t} \sum_{j \in \mathbb{J}} y_{ij} \leq x_i, \quad \forall\, i \in \mathbb{I}$$

$$\sum_{i \in \mathbb{I}} y_{ij} \geq d_j^{\xi}, \quad \forall\, j \in \mathbb{J}$$

# Which one is better?

| | $d_1$: 3 un<br>$293 | $d_1$: 5 un<br>$378.7 | $d_1$: 7 un<br>$469.3 | |
|---|---|---|---|---|
| $c_1$: $10/un | 0 | 0.83 | 4.17 | 2.67 |
| $c_2$: $7/un | 3 | 3 | 3 | 4 |
| $c_3$: $16/un | 3 | 4.17 | 2.83 | 3.33 |
| $c_4$: $6/un | 6 | 4 | 2 | 2 |
| Scen1: | $293 | $294.4 | $297.8 | $295.4 |
| Scen2: | $395 | $378.7 | $381.3 | $380.3 |
| Scen3: | $501 | $480.7 | $469.3 | $470.3 |
| Expected cost | $396 | $384 | $383 | $381.9 |

$$\min \sum_{i\in\mathbb{I}} c_i x_i + \sum_{s=1,2,3} p^s \sum_{j\in\mathbb{J}} f_{ij} y_{ij}^s$$

$$\sum_{i\in\mathbb{I}} c_i x_i \le b$$

$$\sum_{i\in\mathbb{I}} x_i \ge m$$

$$\sum_{j\in\mathbb{J}} y_{ij}^s \le x_i, \quad \forall\, i \in \mathbb{I}, \forall\, s = 1,2,3$$

$$\sum_{i\in\mathbb{I}} y_{ij}^s \ge d_j^s, \quad \forall\, j \in \mathbb{J}, \forall\, s = 1,2,3$$

# Two-stage stochastic models

- First-stage decisions $x$ (before the uncertainty is revealed)
- Second-stage decisions $y$ (*a posteriori* recourse actions)

$$\min_{x \in \mathcal{X}} c^\intercal x + \sum_{s \in \mathcal{S}} p^s Q(x, \xi^s)$$

where

$$\begin{aligned} Q(x, \xi^s) := \min q^s y \\ T^s x + W^s y = h^s \\ y \geq 0 \end{aligned}$$

- Assumption: **Fixed recourse** (only $T^s$ and $h^s$ are random, i.e. fixed $W$ and $q$ for all scenarios)

# Decomposition methods

- Problem has a very particular structure of the problem

$$
\begin{pmatrix}
A & & & & \\
T^1 & W & & & \\
T^2 & & W & & \\
\vdots & & & \ddots & \\
T^n & & & & W
\end{pmatrix}
\begin{pmatrix} x \\ y^1 \\ y^2 \\ \vdots \\ y^n \end{pmatrix}
=
\begin{pmatrix} b \\ h^1 \\ h^2 \\ \vdots \\ h^n \end{pmatrix}
$$

- Decomposition algorithms exploit this structure by solving subproblems independently and *changing* the first stage problem

# Benders Decomposition (aka L-shape)

- Benders idea: To approximate the second stage by cuts

$$\min_{x \in \mathcal{X}} \quad c^{\intercal} x + \sum_{s \in \mathcal{S}} p^s \theta^s$$

$$Q(x, \xi^s) \leq \theta^s \qquad \forall s \in \mathcal{S}$$

# Benders Decomposition (aka L-shape)

- Benders idea: To approximate the second stage by cuts

$$\min_{x \in \mathcal{X}} \quad c^\intercal x + \sum_{s \in \mathcal{S}} p^s \theta^s$$

$$Q(x, \xi^s) \leq \theta^s \qquad \forall s \in \mathcal{S}$$

- Using the dual subproblems, we can construct cuts

$$\mathcal{Q}(x, \xi^s) := \min_{y \geq 0} q^\intercal y$$

$$T^s x + W y = h^s$$

# Benders Decomposition (aka L-shape)

- Benders idea: To approximate the second stage by cuts

$$\min_{x\in\mathcal{X}} \quad c^\intercal x + \sum_{s\in\mathcal{S}} p^s\theta^s$$

$$Q(x,\xi^s) \le \theta^s \qquad \forall s \in \mathcal{S}$$

- Using the dual subproblems, we can construct cuts

$$\mathcal{Q}(x,\xi^s) := \min_{y\ge 0} q^\intercal y \qquad \Longleftrightarrow \qquad Q(x,\xi^s) = \max_{\lambda^s\in\mathbb{R}^p} \ (h^s - T^s x)^\intercal \lambda^s$$

$$T^s x + W y = h^s \qquad\qquad W^\intercal \lambda^s \le q$$

# Benders Decomposition (aka L-shape)

- Benders idea: To approximate the second stage by cuts

$$\min_{x\in\mathcal{X}} \quad c^\intercal x + \sum_{s\in\mathcal{S}} p^s\theta^s$$

$$Q(x,\xi^s) \leq \theta^s \qquad \forall s\in\mathcal{S}$$

- Using the dual subproblems, we can construct cuts

$$\mathcal{Q}(x,\xi^s) := \min_{y\geq 0} q^\intercal y \qquad\qquad Q(x,\xi^s) = \max_{\lambda^s\in\mathbb{R}^p} \; (h^s - T^s x)^\intercal \lambda^s$$

$$T^s x + W y = h^s \qquad\qquad W^\intercal \lambda^s \leq q$$

obtaining

$$\min_{x\in\mathcal{X}} \quad c^\intercal x + \sum_{s\in\mathcal{S}} p^s\theta^s$$

$$\left\{ \max_{\lambda^s\in\mathbb{R}^p, W^\intercal\lambda^s\leq q} (h^s - T^s x)^\intercal \lambda^s \right\} \leq \theta^s \qquad \forall s\in\mathcal{S}$$

# Benders Decomposition (aka L-shape)

- Benders idea: To approximate the second stage by cuts

$$\min_{x \in \mathcal{X}} \quad c^\intercal x + \sum_{s \in \mathcal{S}} p^s \theta^s$$

$$Q(x, \xi^s) \leq \theta^s \qquad \forall s \in \mathcal{S}$$

- Using the dual subproblems, we can construct cuts

$$\mathcal{Q}(x, \xi^s) := \min_{y \geq 0} q^\intercal y \quad T^s x + W y = h^s \quad \Longleftrightarrow \quad Q(x, \xi^s) = \max_{\lambda^s \in \mathbb{R}^p} (h^s - T^s x)^\intercal \lambda^s \quad W^\intercal \lambda^s \leq q$$

obtaining

$$\min_{x \in \mathcal{X}} \quad c^\intercal x + \sum_{s \in \mathcal{S}} p^s \theta^s$$

$$(h^s - T^s x)^\intercal \hat{\lambda}^s \leq \theta^s \quad \forall\ \hat{\lambda}^s \in \mathrm{XP}(\Lambda),\ \forall\ s \in \mathcal{S}$$

where $\mathrm{XP}(\mathit{\Lambda})$ are the extreme points of $\Lambda = \{\lambda : W^\intercal \lambda \leq q\}$

# Benders Decomposition (aka L-shape)

- Problem has a very particular structure of the problem

$$\begin{pmatrix} A & & & & \\ T^1 & W & & & \\ T^2 & & W & & \\ \vdots & & & \ddots & \\ T^n & & & & W \end{pmatrix} \begin{pmatrix} x \\ y^1 \\ y^2 \\ \vdots \\ y^n \end{pmatrix} = \begin{pmatrix} b \\ h^1 \\ h^2 \\ \vdots \\ h^n \end{pmatrix}$$

- Decompose by fixing $x$ and adding cuts

# Benders Decomposition (aka L-shape)

- Problem has a very particular structure of the problem

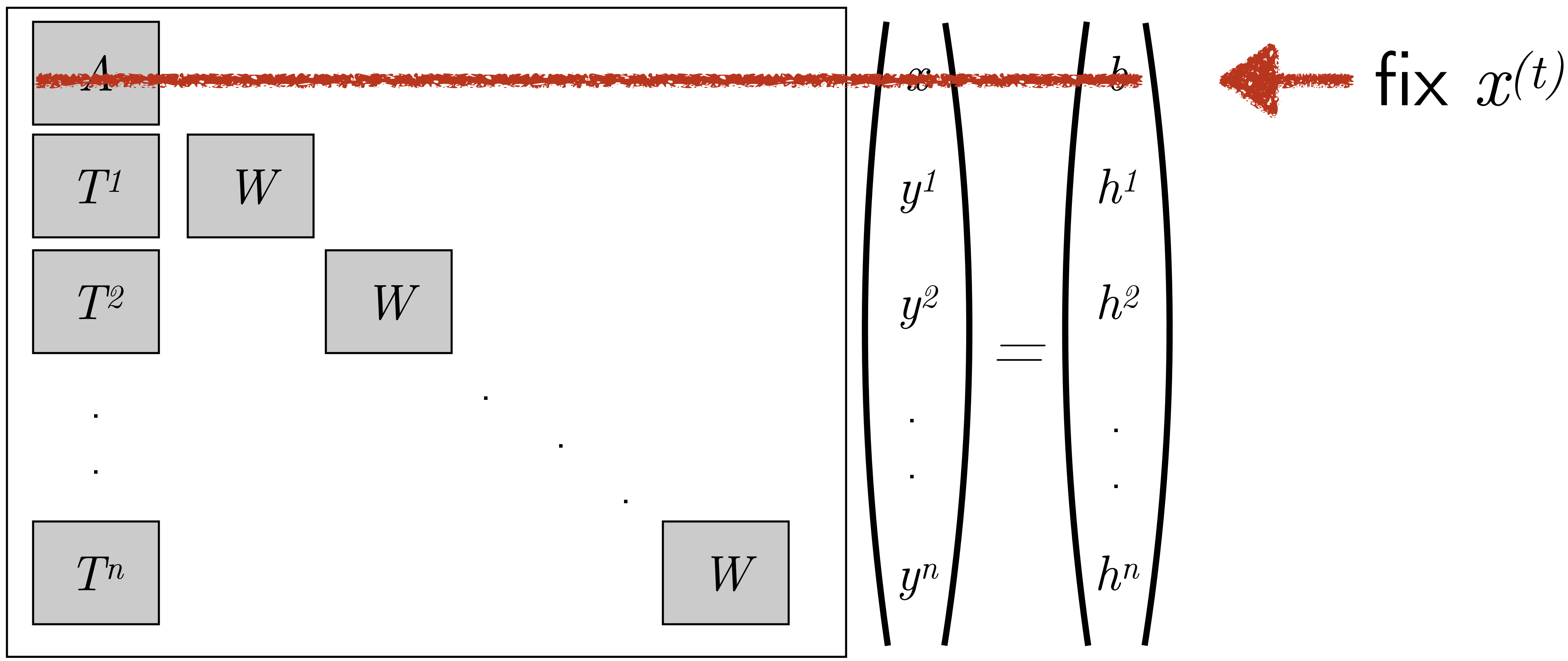
$$\begin{pmatrix} A & & & & \\ T^1 & W & & & \\ T^2 & & W & & \\ \vdots & & & \ddots & \\ T^n & & & & W \end{pmatrix} \begin{pmatrix} x \\ y^1 \\ y^2 \\ \vdots \\ y^n \end{pmatrix} = \begin{pmatrix} b \\ h^1 \\ h^2 \\ \vdots \\ h^n \end{pmatrix}$$

← fix $x^{(t)}$

- Decompose by fixing $x$ and adding cuts

# Benders Decomposition (aka L-shape)

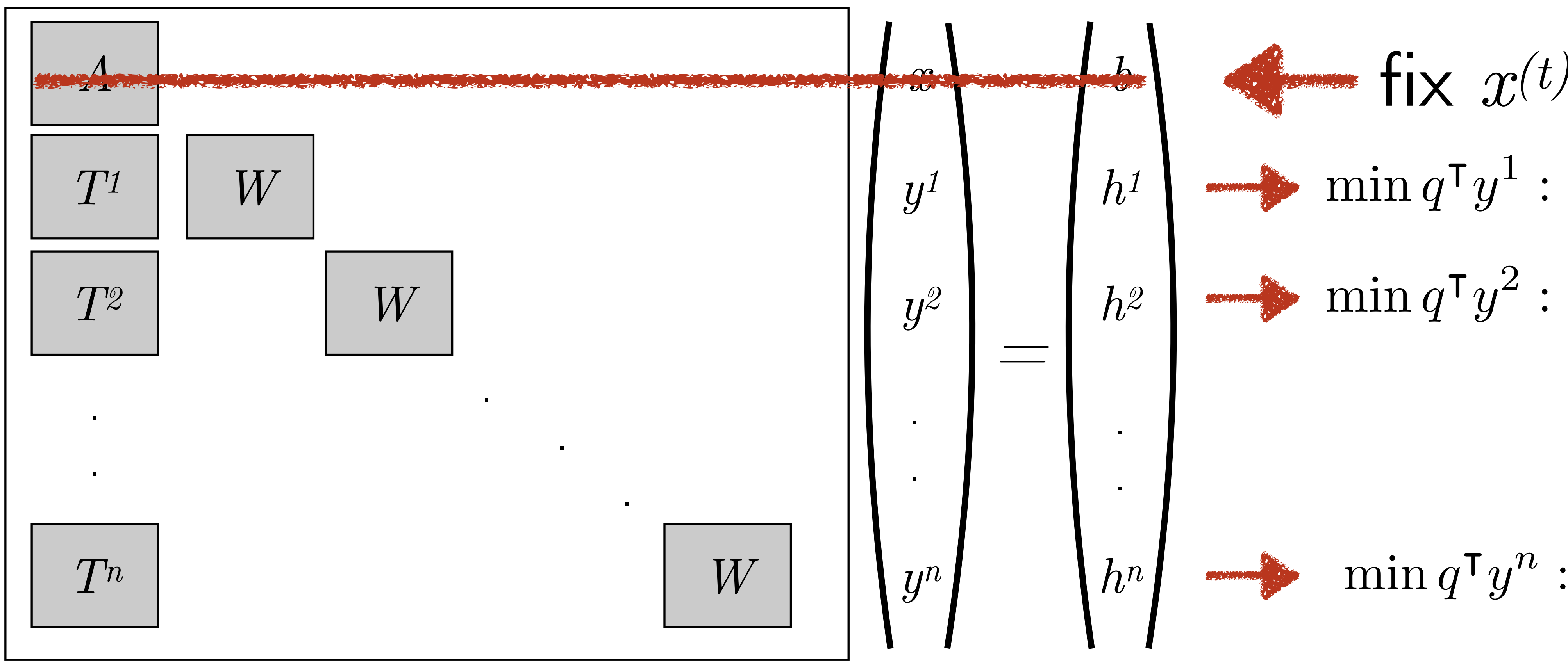
- Problem has a very particular structure of the problem




- Decompose by fixing $x$ and adding cuts

# Benders Decomposition (aka L-shape)

- Problem has a very particular structure of the problem




$$\min q^\intercal y^1 : W y^1 = h^1 - T^1 x^{(t)}$$

$$\min q^\intercal y^2 : W y^2 = h^1 - T^2 x^{(t)}$$

$$\min q^\intercal y^n : W y^n = h^1 - T^n x^{(t)}$$

- Decompose by fixing $x$ and adding cuts

# Benders Decomposition (aka L-shape)

- Benders idea: To approximate the second stage by cuts

$$\min_{x \in \mathcal{X}} \quad c^\intercal x + \sum_{s \in \mathcal{S}} p^s \theta^s$$

Optimality cuts: $(h^s - T^s x)^\intercal \hat{\lambda}^s \leq \theta^s \quad \forall\ \hat{\lambda}^s \in \mathrm{XP}(\Lambda),\ \forall\ s \in \mathcal{S}$

Feasibility cuts: $(h^s - T^s x)^\intercal \tilde{\lambda}^s \leq 0 \quad \forall\ \tilde{\lambda}^s \in \mathrm{XR}(\Lambda),\ \forall\ s \in \mathcal{S}$

# Benders Decomposition (aka L-shape)

- Benders idea: To approximate the second stage by cuts

$$\min_{x \in \mathcal{X}} \quad c^\intercal x + \sum_{s \in \mathcal{S}} p^s \theta^s$$

Optimality cuts: $(h^s - T^s x)^\intercal \hat{\lambda}^s \leq \theta^s \quad \forall\ \hat{\lambda}^s \in \mathrm{XP}(\Lambda),\ \forall\ s \in \mathcal{S}$

Feasibility cuts: $(h^s - T^s x)^\intercal \tilde{\lambda}^s \leq 0 \quad \forall\ \tilde{\lambda}^s \in \mathrm{XR}(\Lambda),\ \forall\ s \in \mathcal{S}$

- Two approaches for optimality cuts:
  - Multi-cut Benders: one cut for each scenario
  - Single-cut Benders: one cut aggregating all scenarios

# Benders Decomposition (aka L-shape)

- Benders idea: To approximate the second stage by cuts

$$\min_{x\in\mathcal{X}} \quad c^\intercal x + \sum_{s\in\mathcal{S}} p^s\theta^s$$

Optimality cuts: $(h^s - T^s x)^\intercal \hat{\lambda}^s \leq \theta^s \quad \forall\ \hat{\lambda}^s \in \mathrm{XP}(\Lambda),\ \forall\ s \in \mathcal{S}$

Feasibility cuts: $(h^s - T^s x)^\intercal \tilde{\lambda}^s \leq 0 \quad \forall\ \tilde{\lambda}^s \in \mathrm{XR}(\Lambda),\ \forall\ s \in \mathcal{S}$

- Two approaches for optimality cuts:
  - Multi-cut Benders: one cut for each scenario
  - Single-cut Benders: one cut aggregating all scenarios $\underbrace{\sum_{s\in\mathbb{S}} p^s\theta^s}_{\Theta}$

$$\sum_{s\in S} p^s (h^s - T^s x)^\intercal \hat{\lambda}^s \leq \Theta \quad (\hat{\lambda}^1, \ldots, \hat{\lambda}^{\mathcal{S}}) \in \mathrm{XP}(\Lambda)^{\mathcal{S}}$$

# Benders Decomposition (aka L-shape)

Advantages:

- Small Master problem (omit recourse variables)
- Proved convergence to the optimal solution
- Particularly useful if subproblems can be solved analytically

**But**, if there are too many scenarios $\mathcal{S}$

- We need to solve $|\mathcal{S}|$ subproblems at each iteration
- Master problem can became large again (one additional cut per scenario at each step)
- We can use Single-cut Benders, but convergence can be very slow.

# Our idea

- Can we **aggregate** the scenarios in order to have a smaller problem, and iteratively **disaggregate** them to obtain the exact* optimal solution?

- Background: **Adaptive Parition Method**
(Espinoza & M. 2014, Song & Luedtke 2015, van Ackooij, de Oliveira & Song 2018, Pay & Song 2020, M. & Ramirez-Pico 2022, Forcier & Leclère 2022)

# Aggregating all scenarios

Fixing first-stage, we solve each problem independently

$$\min_{x \in \mathcal{X}} \quad c^\intercal x + \sum_{s \in \mathcal{S}} p^s \theta^s$$

$$(h^s - T^s x)^\intercal \hat{\lambda}^s \leq \theta^s$$

$$Q(x, \xi^s) = \max_{\lambda^s \in \mathbb{R}^p} \; (h^s - T^s x)^\intercal \lambda^s$$

$$W^\intercal \lambda^s \leq q$$

# Aggregating all scenarios

Fixing first-stage, we solve each problem independently

$$\min_{x \in \mathcal{X}} \quad c^\intercal x + \sum_{s \in \mathcal{S}} p^s \theta^s$$

$$(h^s - T^s x)^\intercal \hat{\lambda}^s \leq \theta^s$$

$$Q(x, \xi^s) = \max_{\lambda^s \in \mathbb{R}^p} \; (h^s - T^s x)^\intercal \lambda^s$$

$$W^\intercal \lambda^s \leq q$$

Idea: to force dual solutions to be the same for all scenarios

$$\max_{\lambda \in \mathbb{R}^p} \; \sum_{s \in \mathcal{S}} p_s (h^s - T^s x)^\intercal \lambda$$

$$W^\intercal \lambda \leq q$$

# Aggregating all scenarios

Fixing first-stage, we solve each problem independently

$$\min_{x \in \mathcal{X}} \quad c^\intercal x + \sum_{s \in \mathcal{S}} p^s \theta^s$$

$$(h^s - T^s x)^\intercal \hat{\lambda}^s \leq \theta^s$$

$$Q(x, \xi^s) = \max_{\lambda^s \in \mathbb{R}^p} \; (h^s - T^s x)^\intercal \lambda^s$$

$$W^\intercal \lambda^s \leq q$$

Idea: to force dual solutions to be the same for all scenarios

$$\max_{\lambda \in \mathbb{R}^p} \; \sum_{s \in \mathcal{S}} p_s (h^s - T^s x)^\intercal \lambda$$

$$W^\intercal \lambda \leq q$$

$\Rightarrow$

$$\max_{\lambda \in \mathbb{R}^p} \; (\bar{h} - \bar{T} x)^\intercal \lambda$$

$$W^\intercal \lambda \leq q$$

UNIVERSIDAD ADOLFO IBÁÑEZ

# Aggregating all scenarios

Fixing first-stage, we solve each problem independently

$$\min_{x \in \mathcal{X}} \quad c^\intercal x + \sum_{s \in \mathcal{S}} p^s \theta^s$$

$$(h^s - T^s x)^\intercal \hat{\lambda}^s \leq \theta^s$$

$$Q(x, \xi^s) = \max_{\lambda^s \in \mathbb{R}^p} \ (h^s - T^s x)^\intercal \lambda^s$$

$$W^\intercal \lambda^s \leq q$$

Idea: to force dual solutions to be the same for all scenarios

$$\max_{\lambda \in \mathbb{R}^p} \sum_{s \in \mathcal{S}} p_s (h^s - T^s x)^\intercal \lambda$$

$$W^\intercal \lambda \leq q$$

→

$$\max_{\lambda \in \mathbb{R}^p} \ (\bar{h} - \bar{T} x)^\intercal \lambda$$

$$W^\intercal \lambda \leq q$$

no need to solve all scenarios!

# Aggregating all scenarios

Fixing first-stage, we solve each problem independently

$$\min_{x\in\mathcal{X}} \quad c^\intercal x + \sum_{s\in\mathcal{S}} p^s\theta^s$$

$$(h^s - T^s x)^\intercal \hat{\lambda}^s \leq \theta^s$$

$$Q(x,\xi^s) = \max_{\lambda^s\in\mathbb{R}^p} \ (h^s - T^s x)^\intercal \lambda^s$$

$$W^\intercal \lambda^s \leq q$$

Idea: to force dual solutions to be the same for all scenarios

$$\max_{\lambda\in\mathbb{R}^p} \sum_{s\in\mathcal{S}} p_s(h^s - T^s x)^\intercal \lambda$$

$$W^\intercal \lambda \leq q$$

$$\max_{\lambda\in\mathbb{R}^p} \ (\bar{h} - \bar{T}x)^\intercal \lambda$$

$$W^\intercal \lambda \leq q$$

no need to solve all scenarios!

and we add one Bender cut using these duals

$$\sum_{s\in\mathcal{S}} p_s(h^s - T^s x)^\intercal \hat{\lambda} \leq \sum_{s\in\mathcal{S}} p_s\theta^s$$

# Partially aggregating scenarios

Let $\mathscr{P}$ a partition of the scenarios. For each subset of scenarios $P \in \mathscr{P}$ we can force equality among them

$$\max_{\lambda \in \mathbb{R}^p} \frac{1}{\sum_{s \in P} p_s} \sum_{s \in P} p_s (h^s - T^s x)^\intercal \lambda^P \quad W^\intercal \lambda^P \leq q \quad \Rightarrow \quad \max_{\lambda \in \mathbb{R}^p} (h^P - T^P x)^\intercal \lambda^P \quad W^\intercal \lambda^P \leq q$$

where $h^P$ and $T^P$ are the "weighted average" coefficients for scenarios in $P$

# Partially aggregating scenarios

Let $\mathscr{P}$ a partition of the scenarios. For each subset of scenarios $P \in \mathscr{P}$ we can force equality among them

$$\max_{\lambda \in \mathbb{R}^p} \frac{1}{\sum_{s \in P} p_s} \sum_{s \in P} p_s (h^s - T^s x)^\intercal \lambda^P \qquad \Rightarrow \qquad \max_{\lambda \in \mathbb{R}^p} (h^P - T^P x)^\intercal \lambda^P$$

$$W^\intercal \lambda^P \leq q \qquad\qquad\qquad\qquad W^\intercal \lambda^P \leq q$$

where $h^P$ and $T^P$ are the "weighted average" coefficients for scenarios in $P$

And, we add a Bender cut for each subset

$$p^P (h^P - T^P x)^\intercal \hat{\lambda}^P \leq \sum_{s \in P} p_s \theta^s \qquad \forall P \in \mathcal{P}$$

# Partially aggregating scenarios

Let $\mathscr{P}$ a partition of the scenarios. For each subset of scenarios $P \in \mathscr{P}$ we can force equality among them

$$\max_{\lambda \in \mathbb{R}^p} \frac{1}{\sum_{s \in P} p_s} \sum_{s \in P} p_s (h^s - T^s x)^\intercal \lambda^P \qquad W^\intercal \lambda^P \leq q$$

$$\Rightarrow$$

$$\max_{\lambda \in \mathbb{R}^p} (h^P - T^P x)^\intercal \lambda^P \qquad W^\intercal \lambda^P \leq q$$

where $h^P$ and $T^P$ are the "weighted average" coefficients for scenarios in $P$

And, we add a Bender cut for each subset

$$p^P (h^P - T^P x)^\intercal \hat{\lambda}^P \leq \sum_{s \in P} p_s \theta^s \qquad \forall P \in \mathcal{P}$$

Solving $|\mathscr{P}|$ subproblems, and adding $|\mathscr{P}|$ cuts to the master problem

# Theoretical background : Valid cuts

Proposition:

$$p^P \cdot (h^P - T^P x)^\intercal \hat{\lambda}^P \leq \sum_{s \in P} p^s \theta^s \qquad \hat{\lambda}^P \in \mathrm{XP}(\Lambda)$$

is a valid constraint for the Benders problem for any $P$

Proof: let $\hat{\lambda}^s$ be optimal dual solutions for each subproblem

$$\begin{aligned}\sum_{s \in P} p^s \cdot (h^s - T^s x)^\intercal \hat{\lambda}^s &\geq \sum_{s \in P} p^s \cdot (h^s - T^s x)^\intercal \hat{\lambda}^P \\ &= \left( \left( \sum_{s \in P} p^s h^s \right) - \left( \sum_{s \in P} p^s T^s \right) x \right)^\intercal \hat{\lambda}^P \\ &= p^P \cdot (h^P - T^P x)^\intercal \hat{\lambda}^P\end{aligned}$$

then

$$p^P \cdot (h^P - T^P x)^\intercal \hat{\lambda}^P \leq \sum_{s \in P} p^s (h^s - T^s x)^\intercal \hat{\lambda}^s \leq \sum_{s \in P} p^s \theta^s$$

# How to aggregate scenarios?

What are we doing?

$$\min_{x\in\mathcal{X}} c^\intercal x + \mathbb{E}\left[\mathcal{Q}(x,\xi))\right] = \min_{x\in\mathcal{X}} c^\intercal x + \sum_{P\in\mathcal{P}} \mathbb{E}\left[\mathcal{Q}(x,\xi)|P\right]\cdot\mathbb{P}(P)$$

$$\stackrel{?}{=} \min_{x\in\mathcal{X}} c^\intercal x + \sum_{P\in\mathcal{P}} \mathcal{Q}(x,\mathbb{E}\left[\xi|P\right])\cdot\mathbb{P}(P)$$

($\geq$ by Jensen's Inequality)

($\leq$ ??)

# How to choose the right partition $\mathscr{P}$?

Proposition: let $x^*$ optimal solution of using only aggregated Benders problem, for a partition satisfying

$$\left(\sum_{s\in P} p^s\right) \cdot \sum_{s\in P} p^s \left(h^{s\intercal}\hat{\lambda}^s\right) = \left(\sum_{s\in P} p^s h^s\right)^\intercal \left(\sum_{s\in P} p^s \hat{\lambda}^s\right) \qquad \text{for all } P \in \mathcal{P}$$

$$\left(\sum_{s\in P} p^s\right) \cdot \sum_{s\in P} p^s \left(T^s x^\intercal \hat{\lambda}^s\right) = \left(\sum_{s\in P} p^s T^s x\right)^\intercal \left(\sum_{s\in P} p^s \hat{\lambda}^s\right) \qquad \text{for all } P \in \mathcal{P}$$

Then $x^*$ is an optimal solution of the original problem

# How to choose the right partition $\mathscr{P}$?

Proposition: let $x^*$ optimal solution of using only aggregated Benders problem, for a partition satisfying

$$\left(\sum_{s\in P} p^s\right) \cdot \sum_{s\in P} p^s \left(h^{s\intercal}\hat{\lambda}^s\right) = \left(\sum_{s\in P} p^s h^s\right)^\intercal \left(\sum_{s\in P} p^s \hat{\lambda}^s\right) \qquad \text{for all } P \in \mathcal{P}$$

$$\left(\sum_{s\in P} p^s\right) \cdot \sum_{s\in P} p^s \left(T^s x^\intercal \hat{\lambda}^s\right) = \left(\sum_{s\in P} p^s T^s x\right)^\intercal \left(\sum_{s\in P} p^s \hat{\lambda}^s\right) \qquad \text{for all } P \in \mathcal{P}$$

*(Multiplying and then taking average)*

Then $x^*$ is an optimal solution of the original problem

# How to choose the right partition $\mathscr{P}$?

Proposition: let $x^*$ optimal solution of using only aggregated Benders problem, for a partition satisfying

$$\left(\sum_{s\in P} p^s\right) \cdot \sum_{s\in P} p^s \left(h^{s\intercal}\hat{\lambda}^s\right) = \left(\sum_{s\in P} p^s h^s\right)^\intercal \left(\sum_{s\in P} p^s \hat{\lambda}^s\right) \qquad \text{for all } P \in \mathcal{P}$$

$$\left(\sum_{s\in P} p^s\right) \cdot \sum_{s\in P} p^s \left(T^s x^\intercal \hat{\lambda}^s\right) = \left(\sum_{s\in P} p^s T^s x\right)^\intercal \left(\sum_{s\in P} p^s \hat{\lambda}^s\right) \qquad \text{for all } P \in \mathcal{P}$$

*(Multiplying and then taking average)* *(Taking average and then multiplying)*

Then $x^*$ is an optimal solution of the original problem

# How to choose the right partition $\mathscr{P}$?

Proposition: let $x^*$ optimal solution of using only aggregated Benders problem, for a partition satisfying

$$\left(\sum_{s\in P} p^s\right) \cdot \sum_{s\in P} p^s \left(h^{s\intercal}\hat{\lambda}^s\right) = \left(\sum_{s\in P} p^s h^s\right)^\intercal \left(\sum_{s\in P} p^s \hat{\lambda}^s\right) \qquad \text{for all } P \in \mathcal{P}$$

$$\left(\sum_{s\in P} p^s\right) \cdot \sum_{s\in P} p^s \left(T^s x^\intercal \hat{\lambda}^s\right) = \left(\sum_{s\in P} p^s T^s x\right)^\intercal \left(\sum_{s\in P} p^s \hat{\lambda}^s\right) \qquad \text{for all } P \in \mathcal{P}$$

*(Multiplying and then taking average)* *(Taking average and then multiplying)*

Then $x^*$ is an optimal solution of the original problem

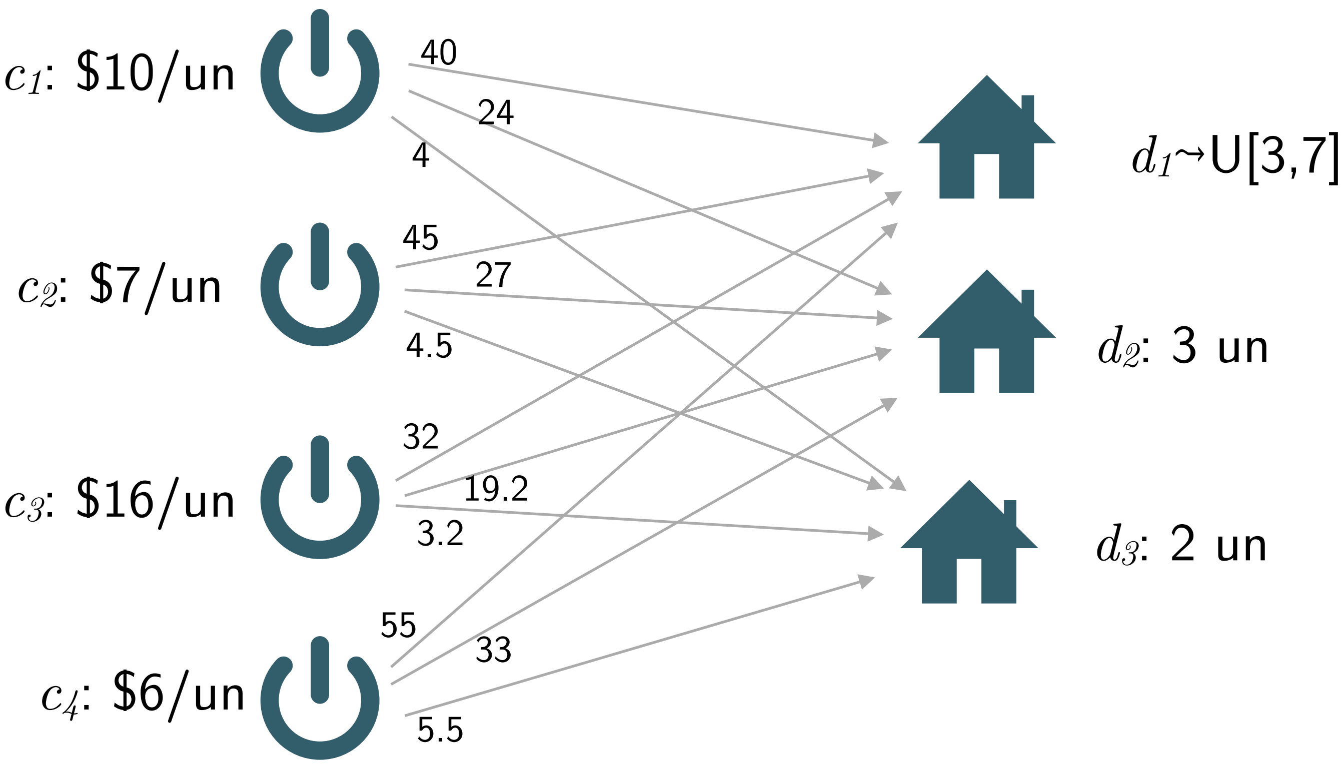
These conditions are satisfied for example when:

- All dual variables λ have the same value in $P$
- All $h^s$ or $T^s$ have the same value in $P$
- A combination of both cases

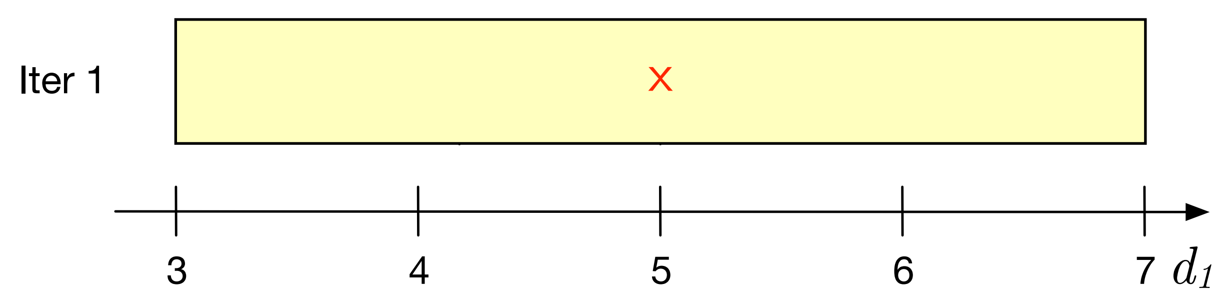
# New method: Benders Adaptive Partitions

1. Set $\mathscr{P} = \{\mathscr{S}\}$
2. Solve MP $:= \min_{x \in \mathscr{X}, \theta \in \mathbb{R}^{|\mathscr{S}|}} c^\intercal x + \sum_{s \in \mathscr{S}} p^s \theta^s$ and get solution $x^{(t)}$ and a lower bound
3. For each $P \in \mathscr{P}$ compute $p^P$, $h^P$ , $T^P$ and solve $Q(x^{(t)}, \xi^P)$
   1. Infeasible: add feasibility cut $p^P \cdot (h^P - T^P x)^\intercal \hat{\lambda}^P \leq 0$
   2. Feasible: add optimality cut $p^P \cdot (h^P - T^P x)^\intercal \hat{\lambda}^P \leq \sum_{s \in P} p^s \theta^s$
4. If a cut was added, go to Step 2. If not,
   1. Solve $Q(x^{(t)}, \xi^s)$ for each scenario and get an upper bound
   2. **Refine** the partition trying to satisfy the conditions of the proposition and go to Step 2.

# Example: LandS (1988)




Budget $b$ : \$120
Min. Capac $m$ : 12



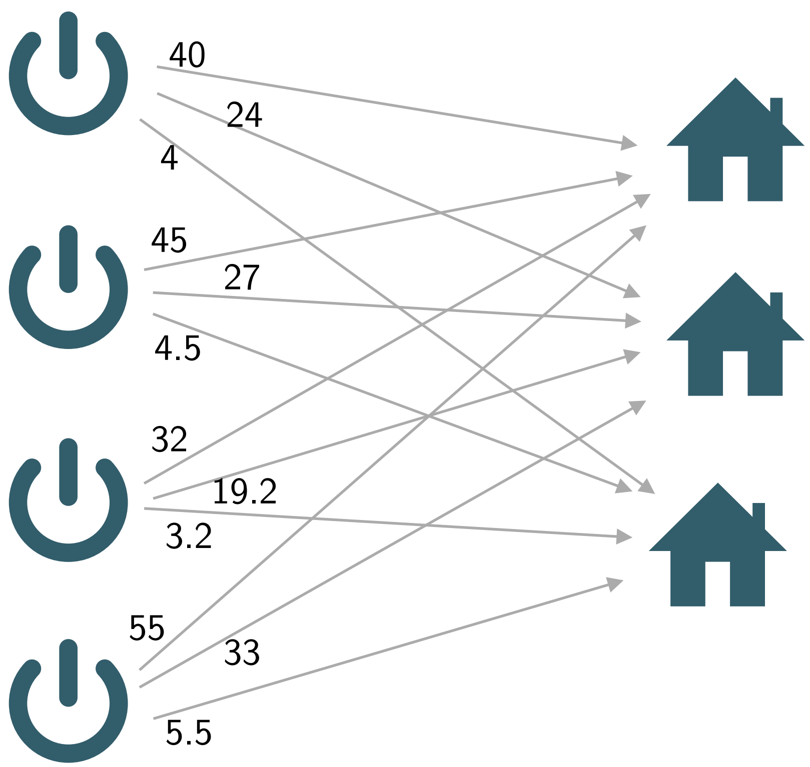
# Example: LandS (1988)

$c_1$: \$10/un

$c_2$: \$7/un

$c_3$: \$16/un

$c_4$: \$6/un




$d_1 \sim$ U[3,7]

$d_2$: 3 un

$d_3$: 2 un

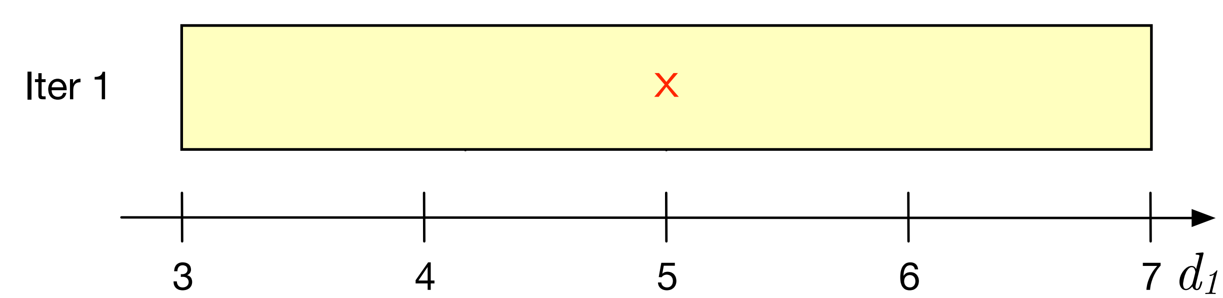
$$\mathcal{Q}(x,\xi) := \min_{y \geq 0} \sum_{i \in \mathbb{I}} \sum_{j \in \mathbb{J}} f_{ij} y_{ij}$$

$$\text{s.t} \sum_{j \in \mathbb{J}} y_{ij} \leq x_i, \quad \forall\, i \in \mathbb{I}$$

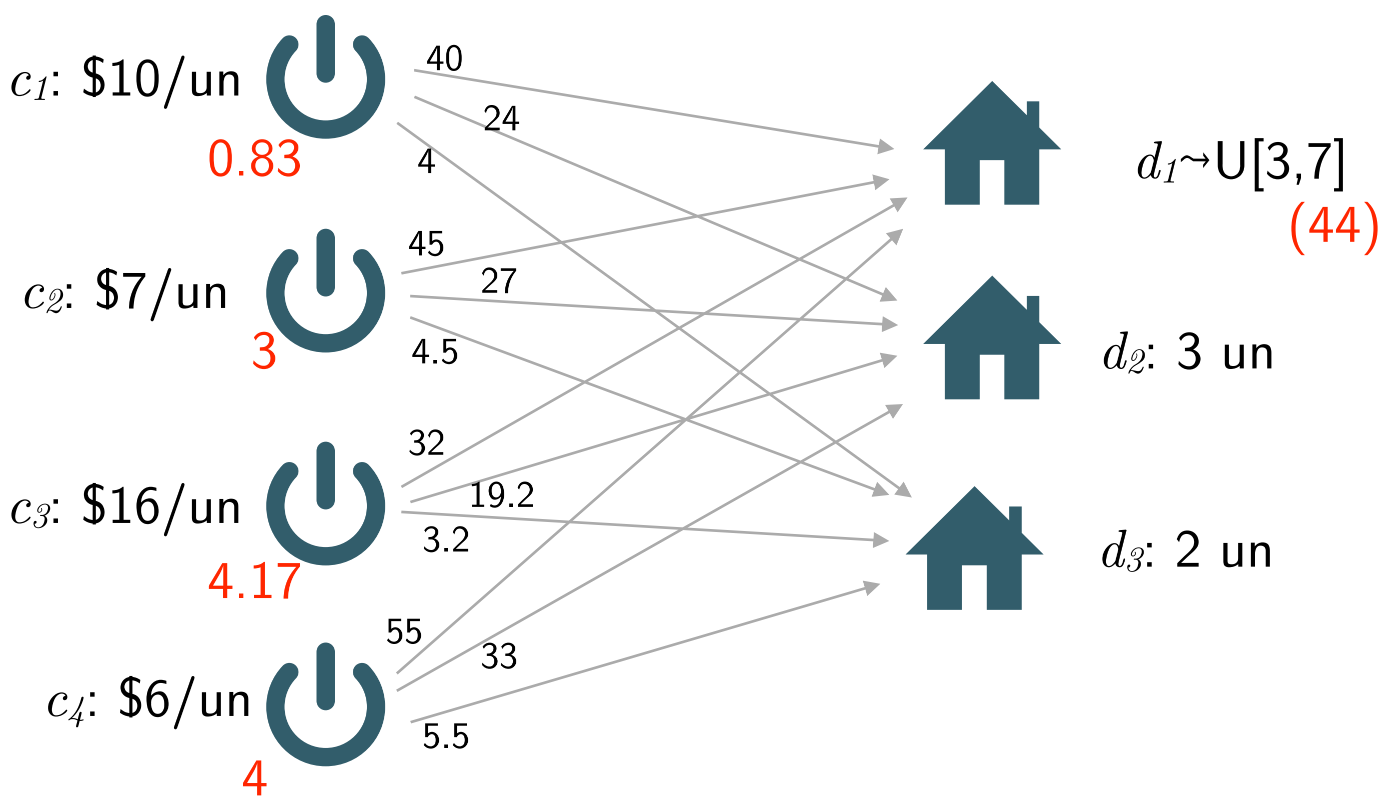
$$\sum_{i \in \mathbb{I}} y_{ij} \geq d_j^{\xi}, \quad \forall\, j \in \mathbb{J}$$

Budget $b$ : \$120

Min. Capac $m$ : 12



# Example: LandS (1988)



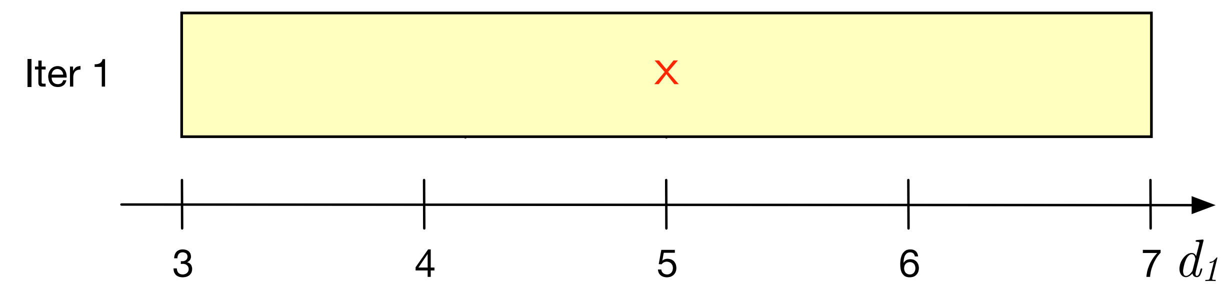

$$\mathcal{Q}(x,\xi) := \min_{y\geq 0} \sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{J}} f_{ij} y_{ij}$$

$$\text{s.t} \sum_{j\in\mathbb{J}} y_{ij} \leq x_i, \quad \forall\, i \in \mathbb{I}$$

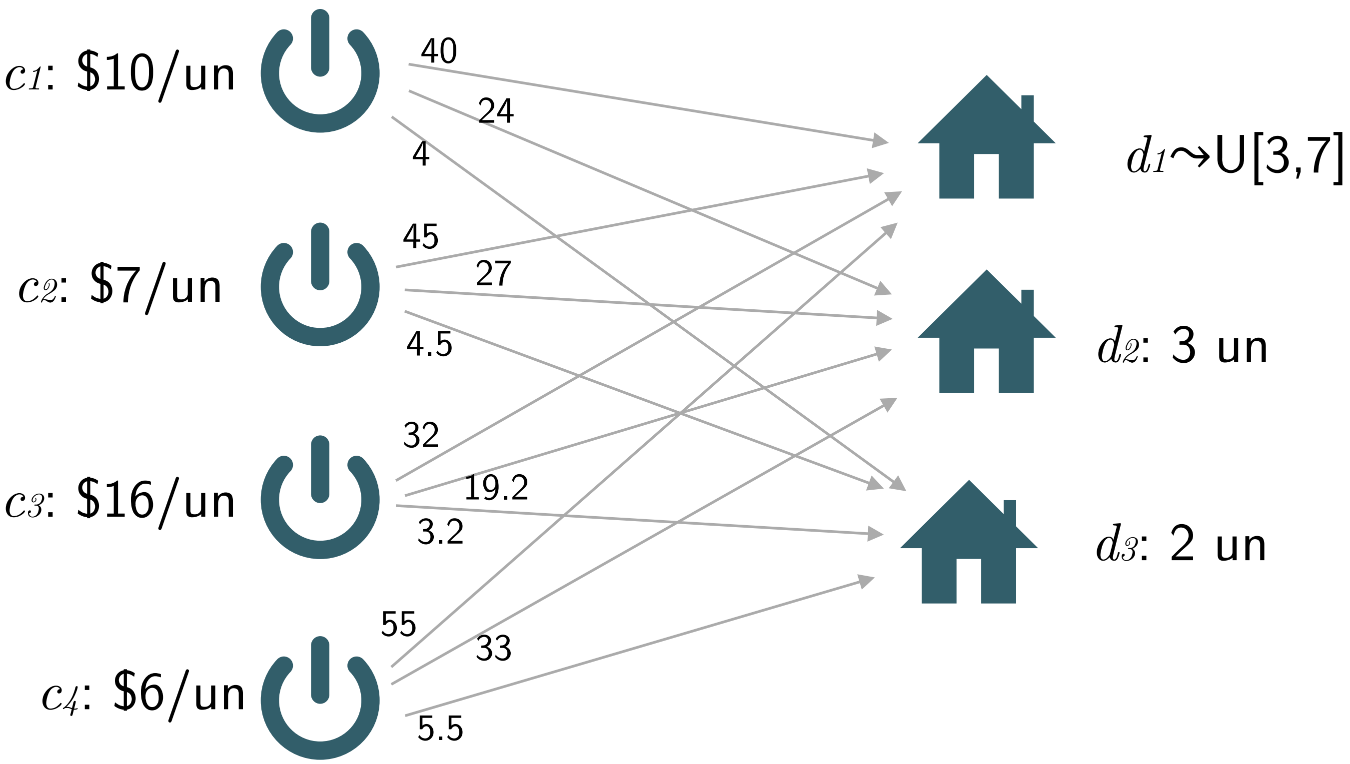
$$\sum_{i\in\mathbb{I}} y_{ij} \geq d_j^{\xi}, \quad \forall\, j \in \mathbb{J}$$

Budget $b$ : $120

Min. Capac $m$ : 12



# Example: LandS (1988)

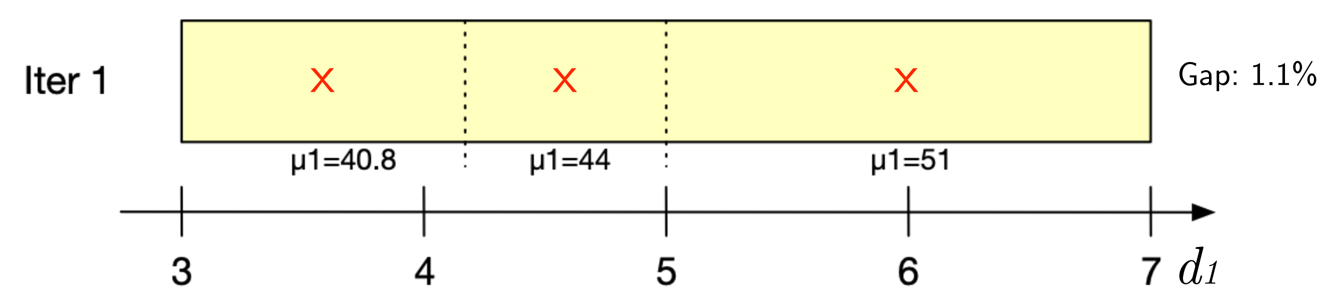



$$\mathcal{Q}(x,\xi) := \min_{y \geq 0} \sum_{i \in \mathbb{I}} \sum_{j \in \mathbb{J}} f_{ij} y_{ij}$$

$$\text{s.t} \sum_{j \in \mathbb{J}} y_{ij} \leq x_i, \quad \forall\, i \in \mathbb{I}$$

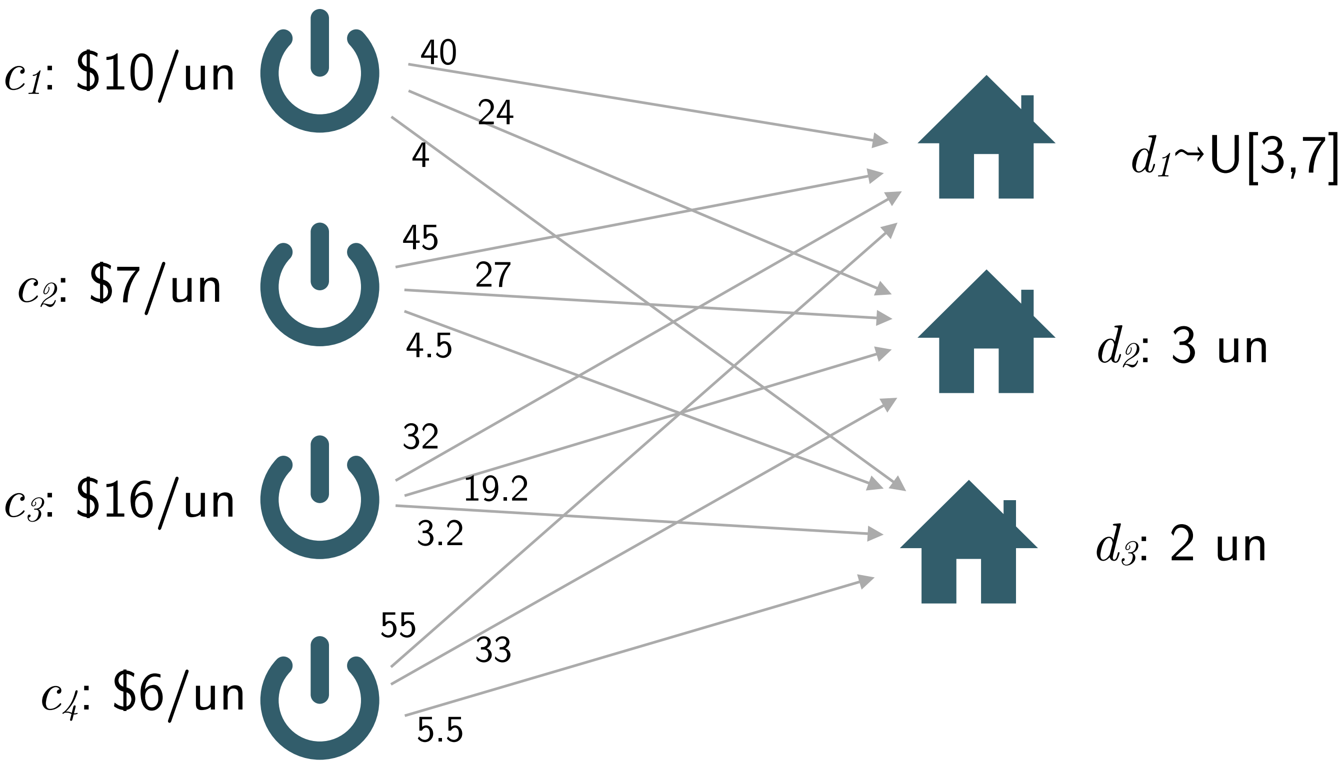
$$\sum_{i \in \mathbb{I}} y_{ij} \geq d_j^{\xi}, \quad \forall\, j \in \mathbb{J}$$

Budget $b$ : \$120
Min. Capac $m$ : 12



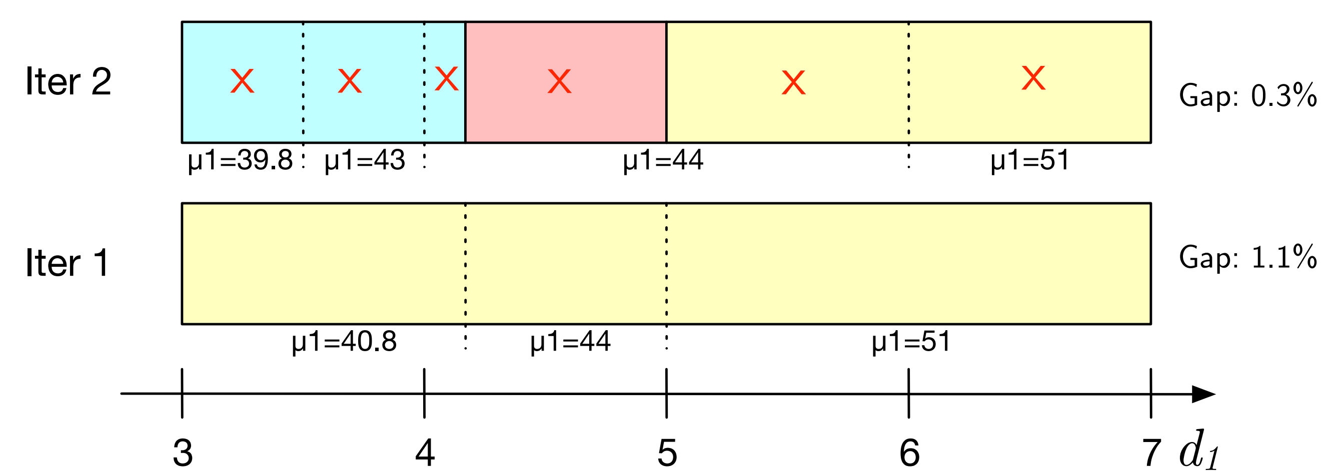

Segments with same duals can be estimated using sensitivity analysis

# Example: LandS (1988)



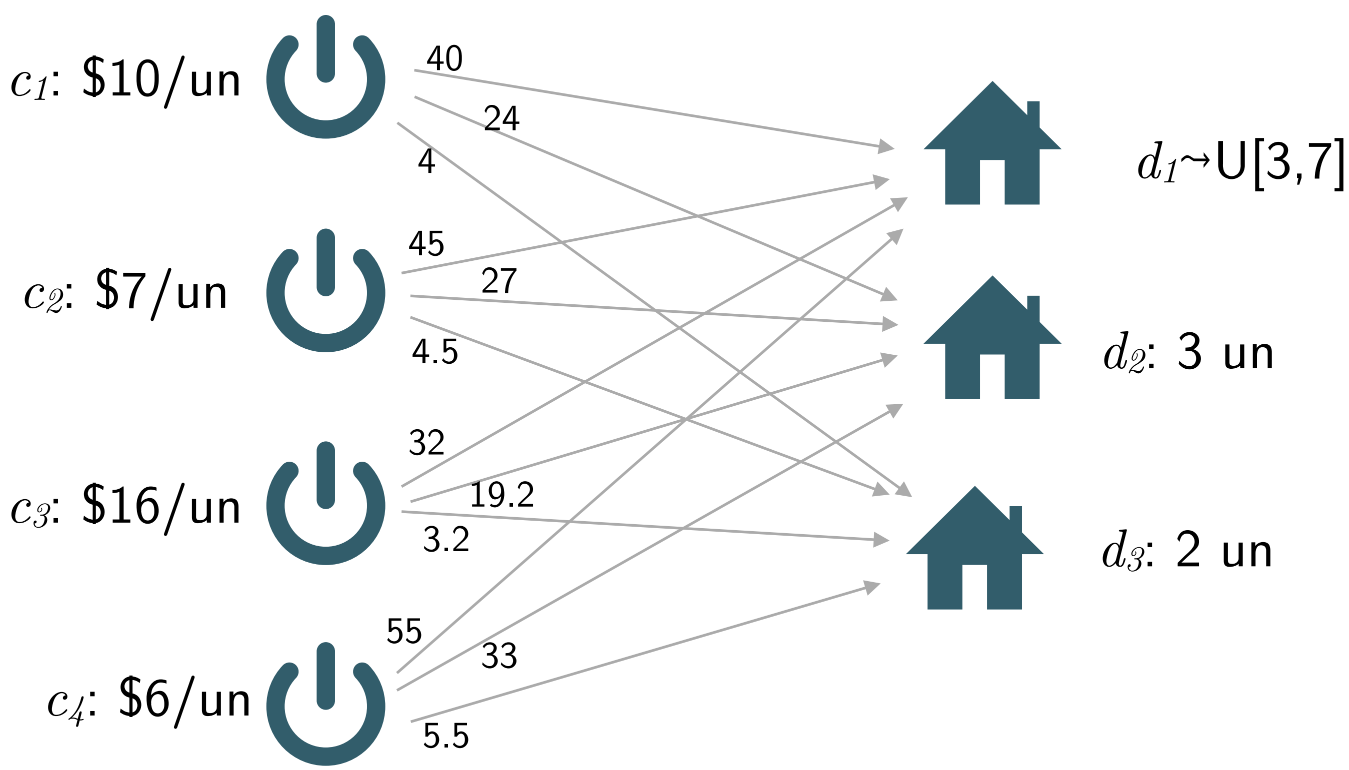

Budget $b$ : \$120

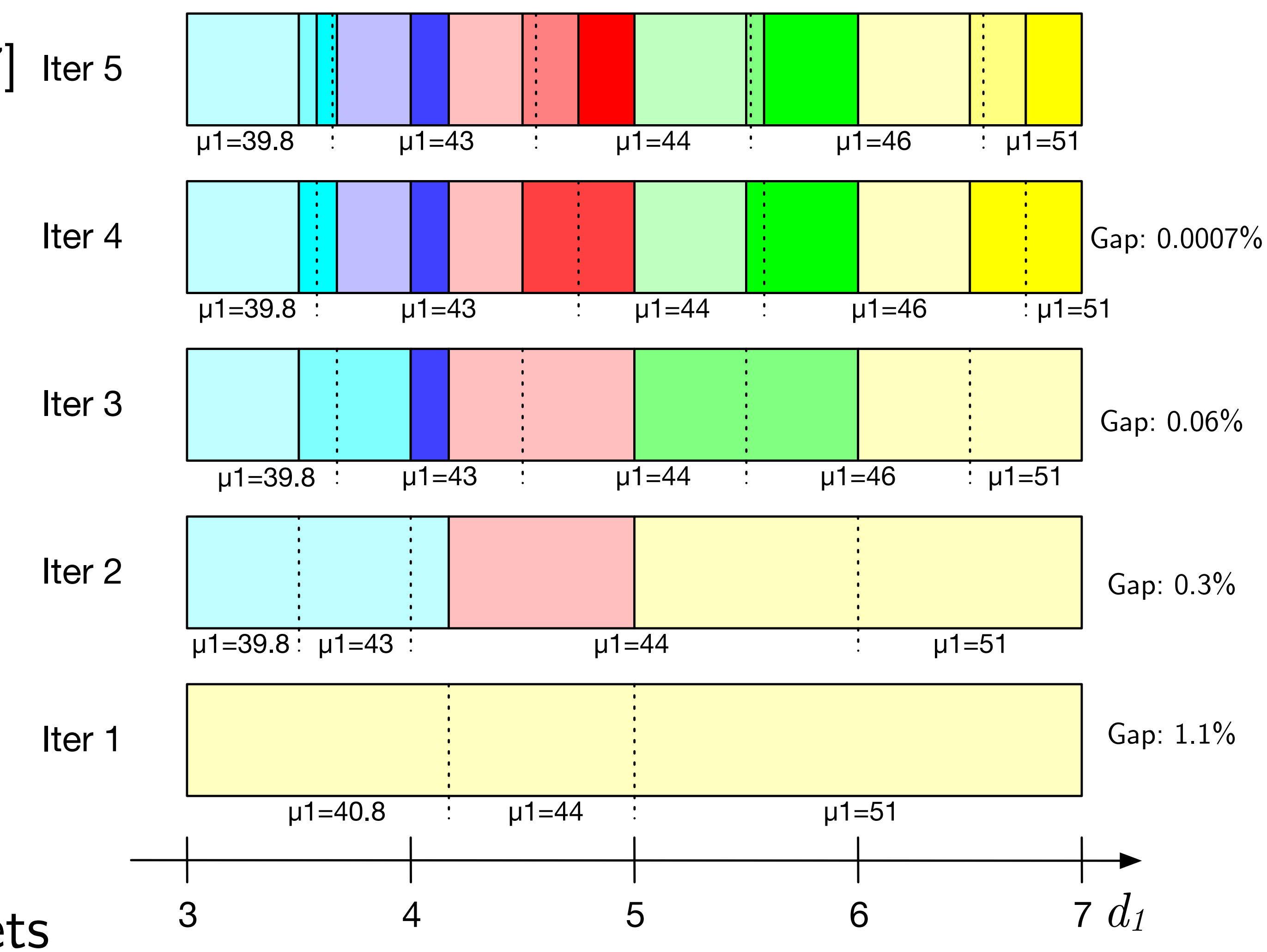
Min. Capac $m$ : 12




Segments with same duals can be estimated using sensitivity analysis

# Example: LandS (1988)



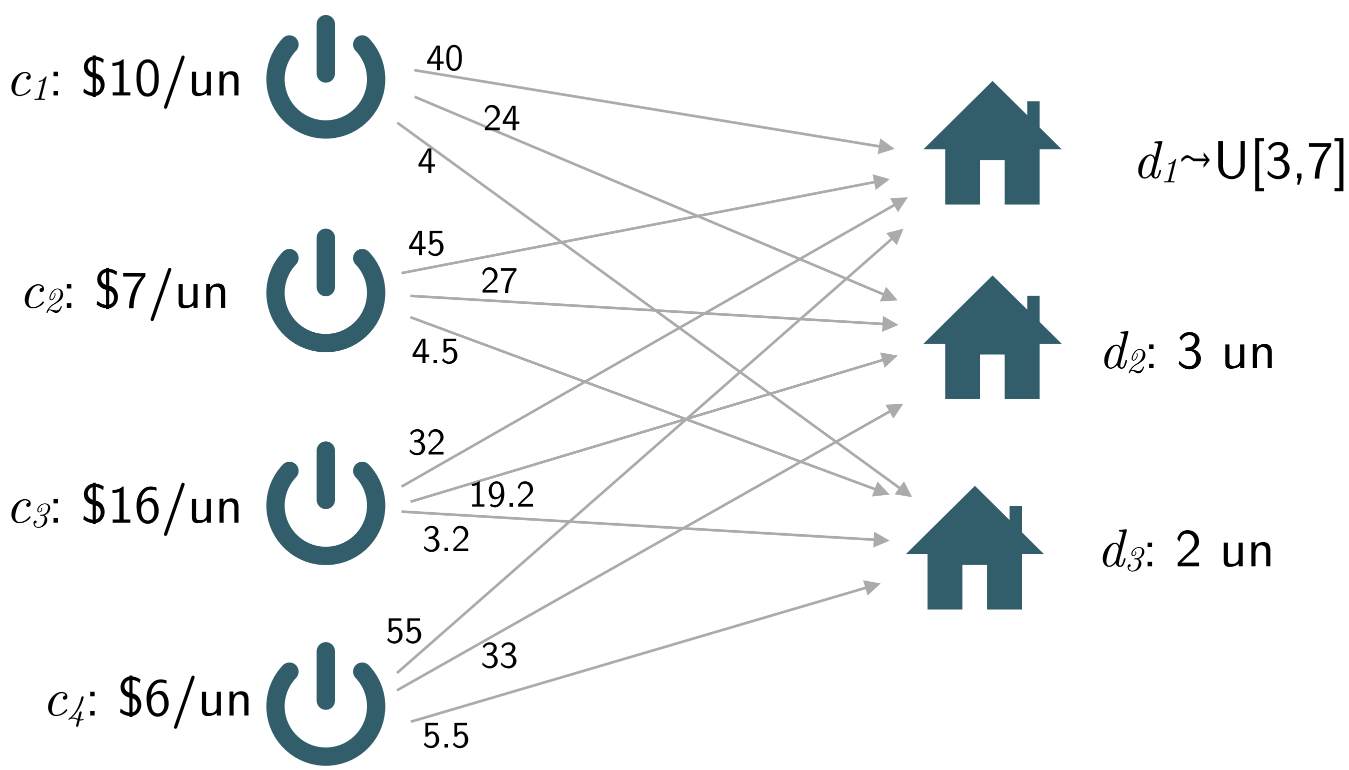

Budget $b$ : \$120

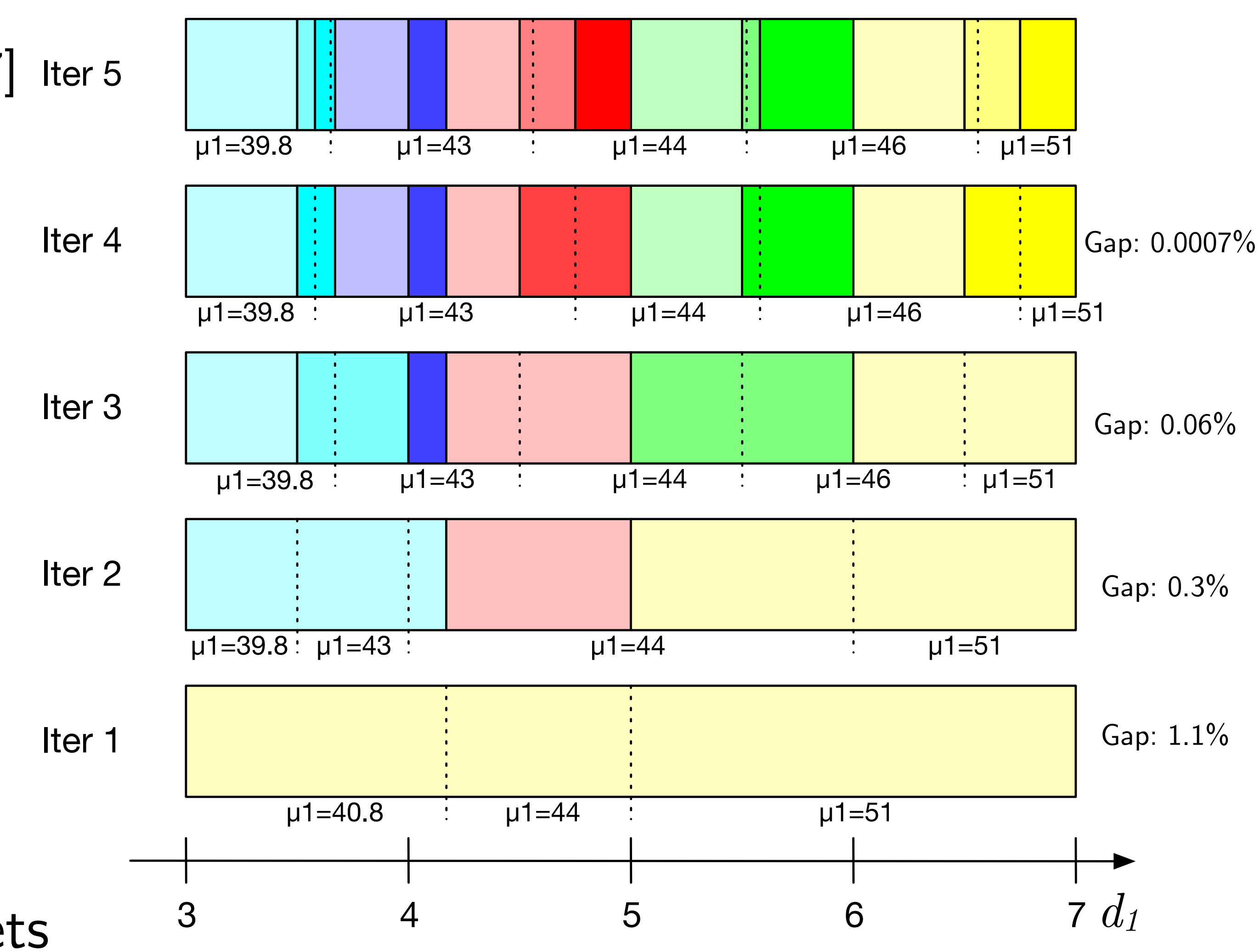
Min. Capac $m$ : 12

In 5 iterations, we get a partition with 14 subsets of scenario, and solution with gap < $10^{-5}$

Segments with same duals can be estimated using sensitivity analysis

# Example: LandS (1988)




Budget $b$ : \$120

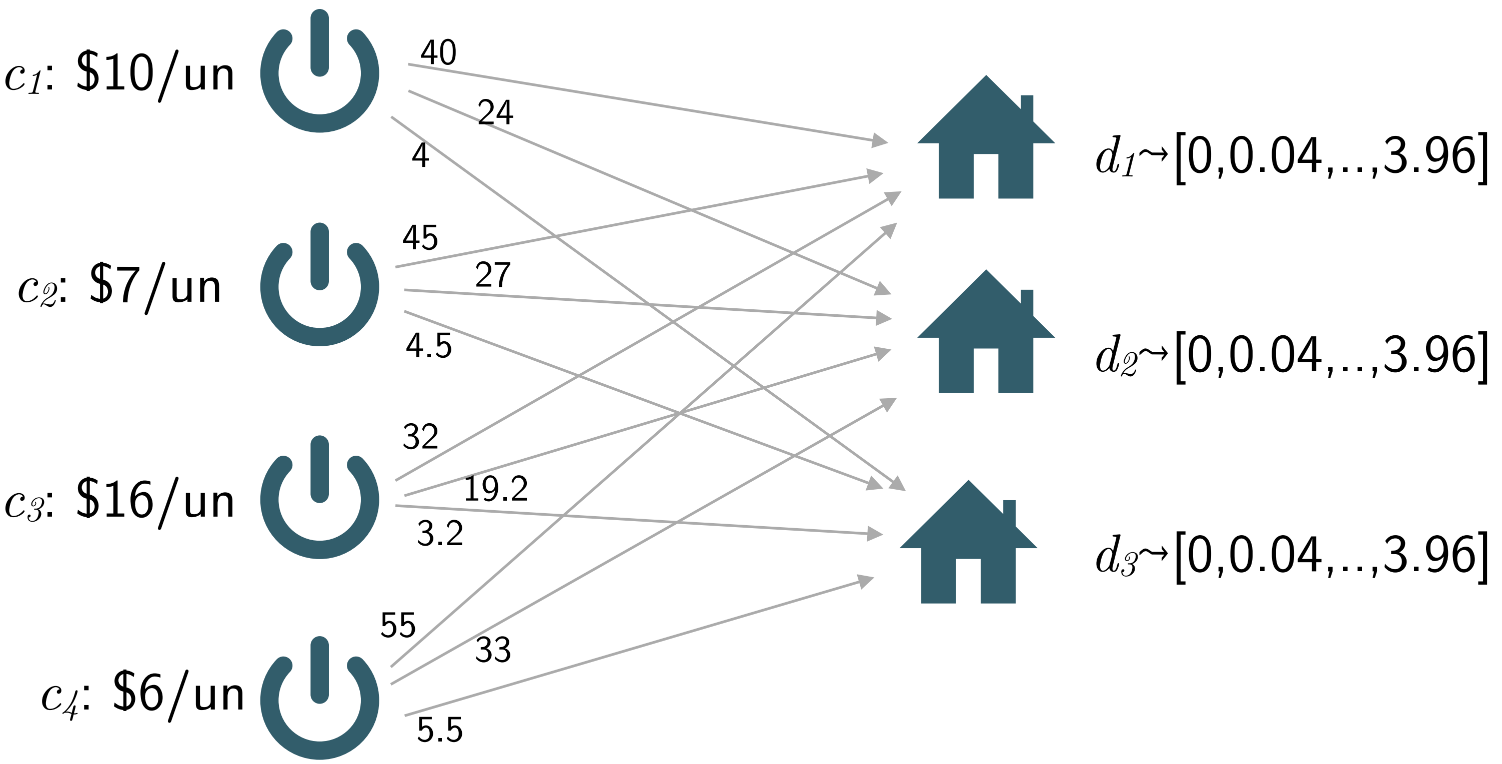
Min. Capac $m$ : 12

In 5 iterations, we get a partition with 14 subsets of scenario, and solution with gap < $10^{-5}$

Note: we just solve an stochastic problem with **continuous distribution (!)** by a series of discrete problems

Segments with same duals can be estimated using sensitivity analysis

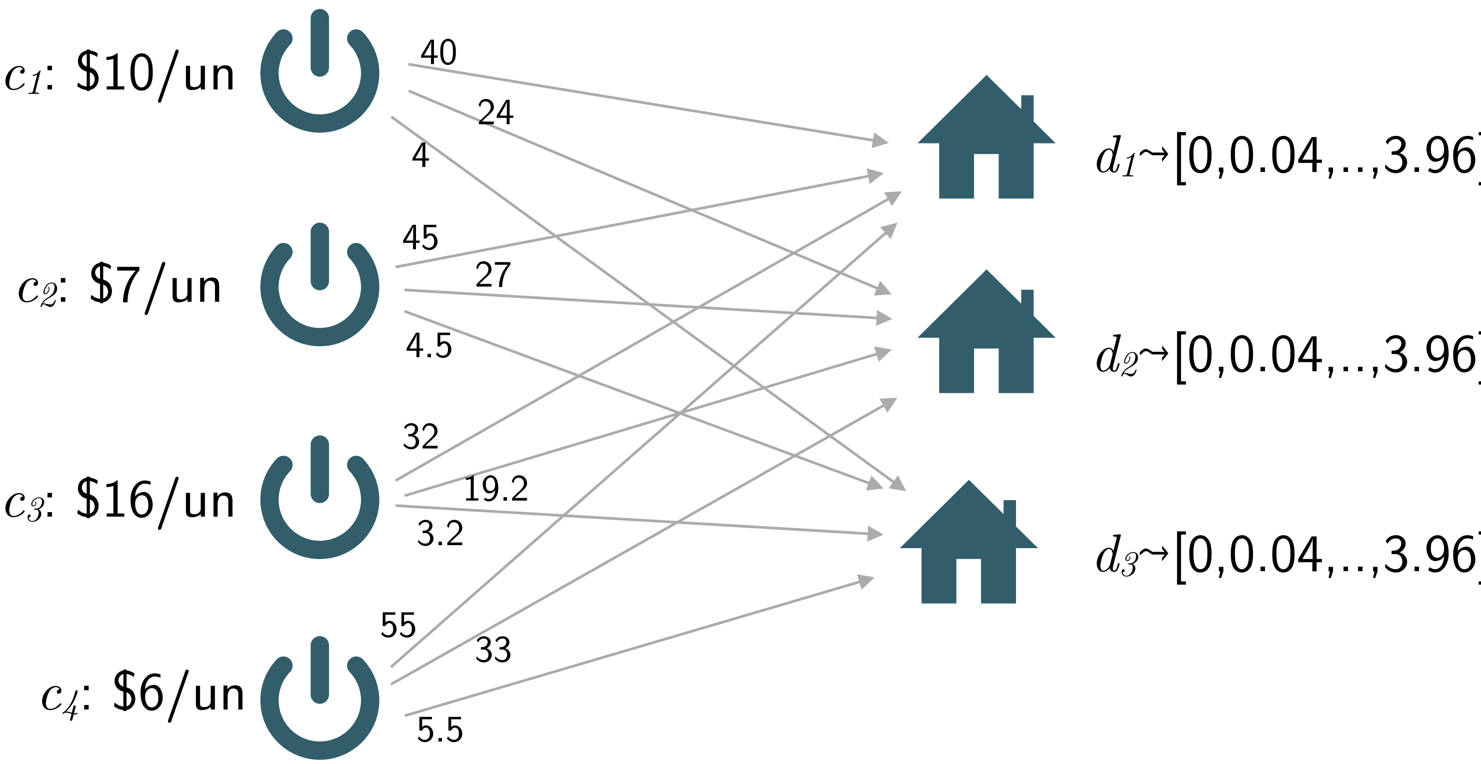
# Different version of LandS (Linderoth et al, 2006)




Budget $b$ : \$120

Min. Capac $m$ : 12

# Different version of LandS (Linderoth et al, 2006)

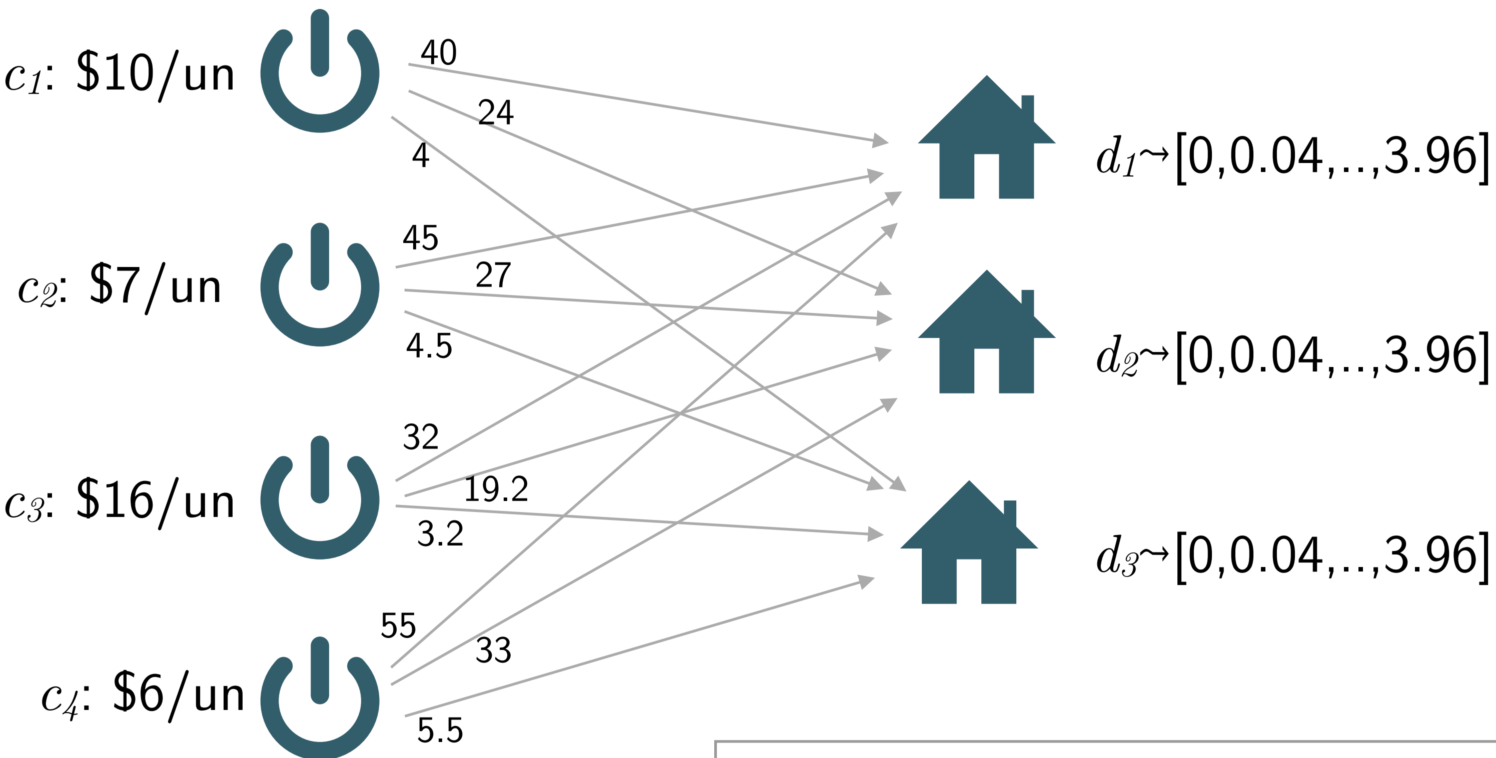



A total of 1'000'000 scenarios

Budget $b$ : $120

Min. Capac $m$ : 12

# Different version of LandS (Linderoth et al, 2006)




A total of 1'000'000 scenarios

Budget $b$ : $120

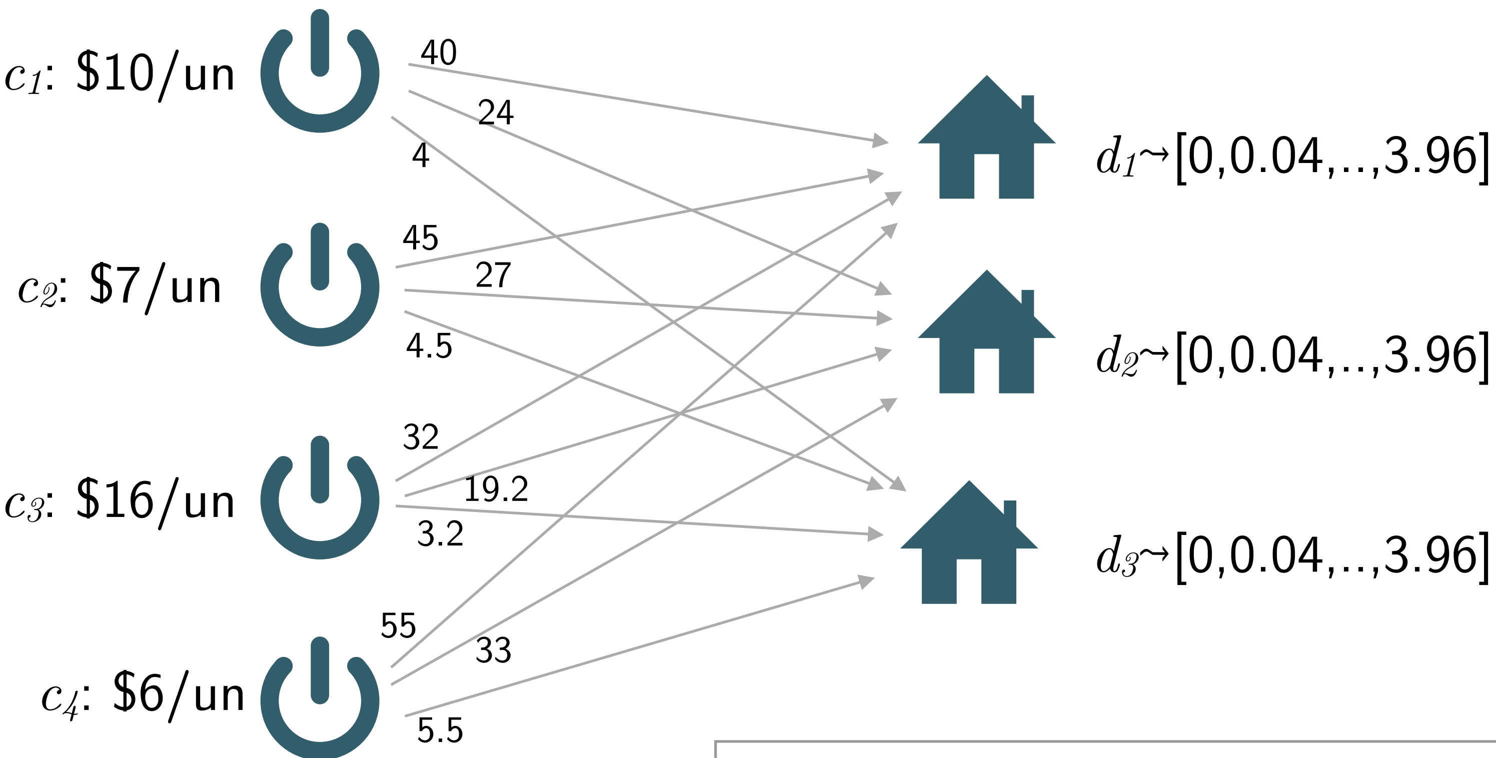
Min. Capac $m$ : 12

**True Optimal solution:**

x*=(0.84, 3.40, 1.88, 5.88)

Objective Value: 225.6294001

# Different version of LandS (Linderoth et al, 2006)




A total of 1'000'000 scenarios

Budget $b$ : $120

Min. Capac $m$ : 12

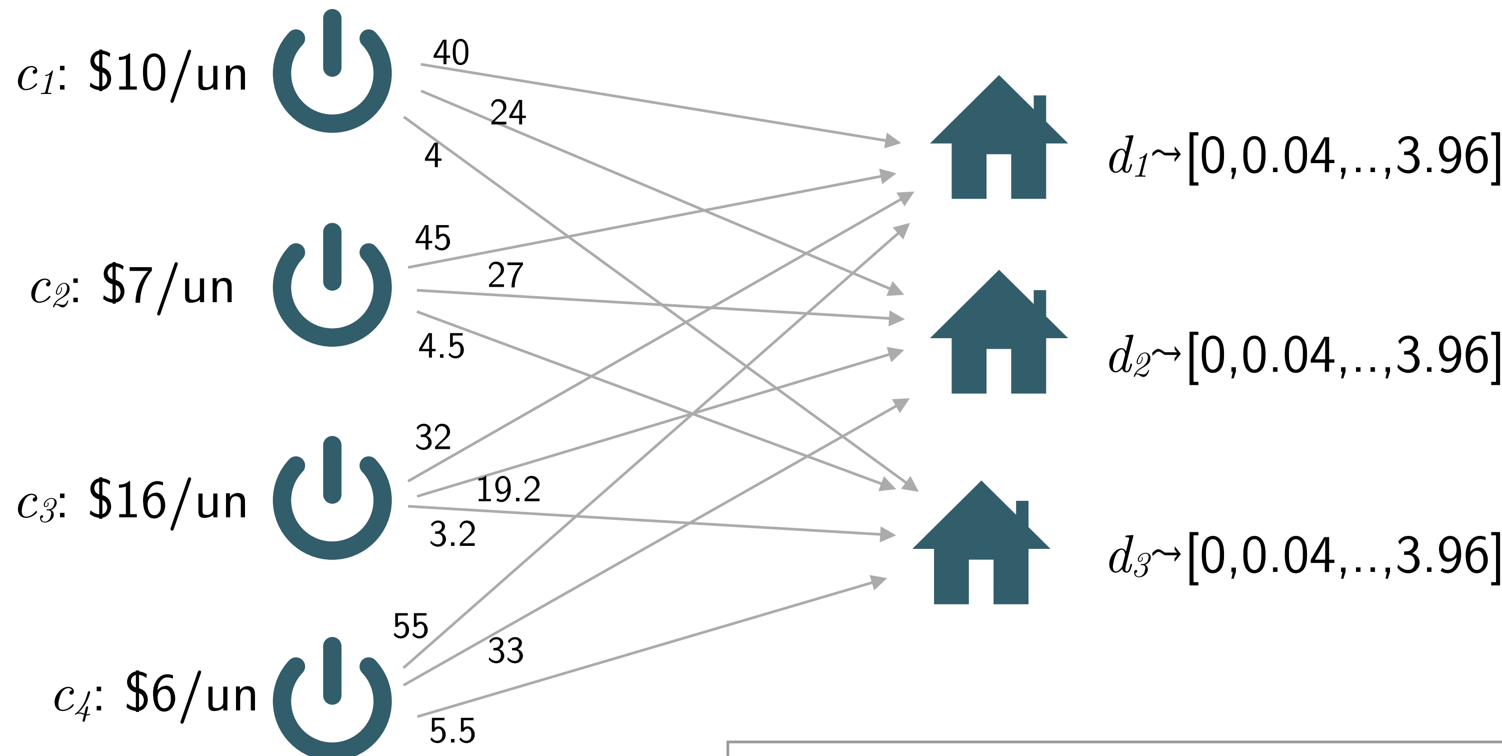
**True Optimal solution:**

x*=(0.84, 3.40, 1.88, 5.88)

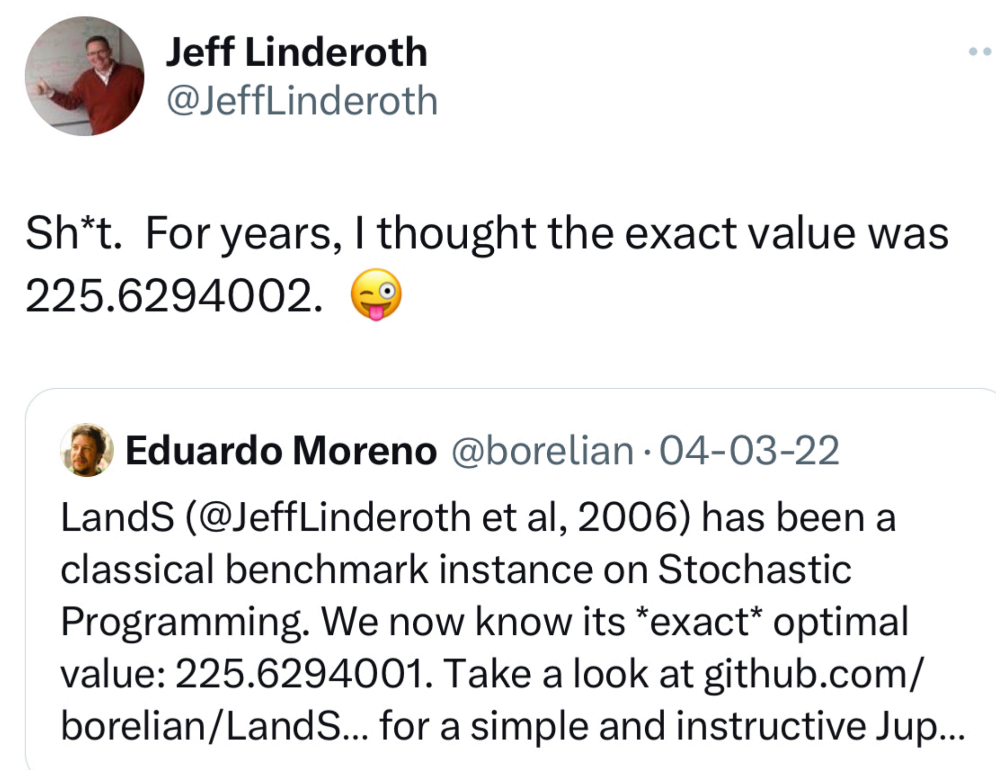
Objective Value: 225.6294001

- Required 8 iterations (refinements) of the algorithm
- Finished with a partition of size 1374
- Solution validated using an exact solver (`QSopt_ex`, Espinoza et al, 2007)
- Jupyter notebook available at https://github.com/borelian/LandS_exact_solution

# Different version of LandS (Linderoth et al, 2006)




A total of 1'000'000 scenarios




Budget $b$ : $120
Min. Capac $m$ : 12

**True Optimal solution:**
x*=(0.84, 3.40, 1.88, 5.88)
Objective Value: 225.6294001

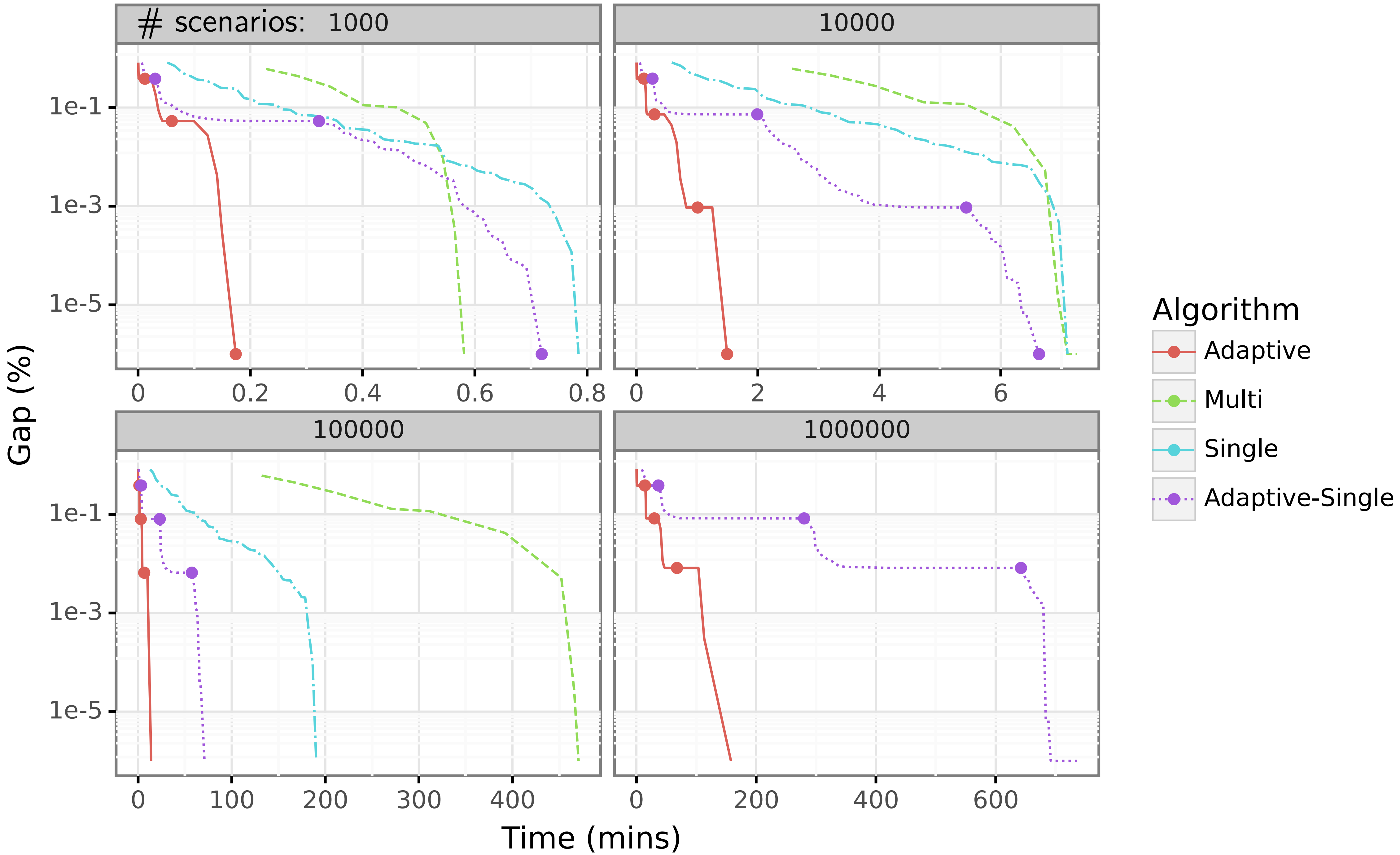
- Required 8 iterations (refinements) of the algorithm
- Finished with a partition of size 1374
- Solution validated using an exact solver (`QSopt_ex`, Espinoza et al, 2007)
- Jupyter notebook available at https://github.com/borelian/LandS_exact_solution

# Problem #1: Capacity Planning Problem

- Dataset: *Electricity Planning* instance (small) from https://people.orie.cornell.edu/huseyin/research/sp_datasets/sp_datasets.html
- 20 source nodes, 50 demand nodes, 10 resources
- 1,000, 10,000, 100,000 and 1,000,000 scenarios sampled from a discrete distribution for each demand node
- Implemented in Python using Gurobi as Solver
- Source codes available at https://github.com/borelian/AdaptiveBenders

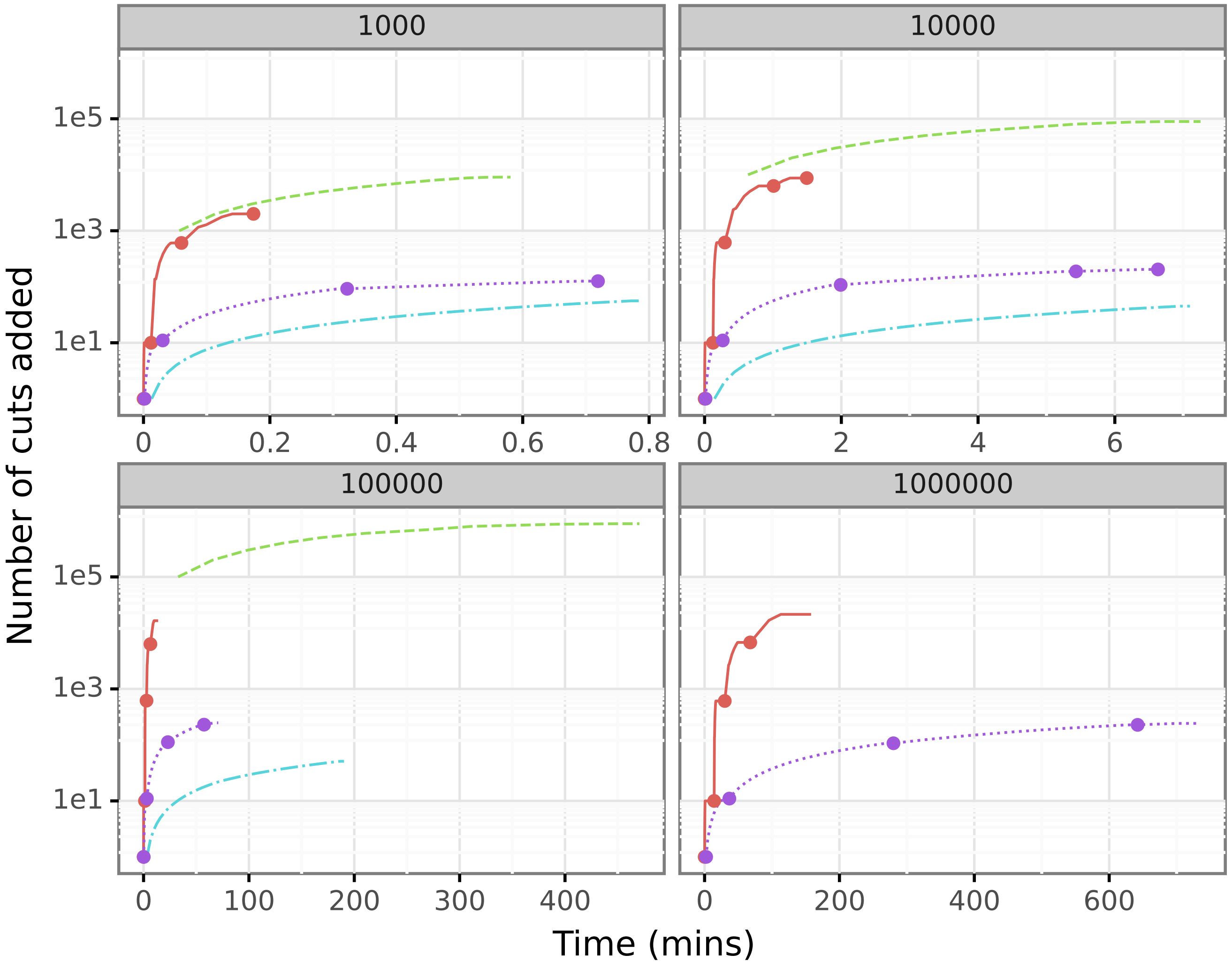
# Problem #1: Capacity Planning Problem

## Gap of incumbent solution vs optimal solution over time

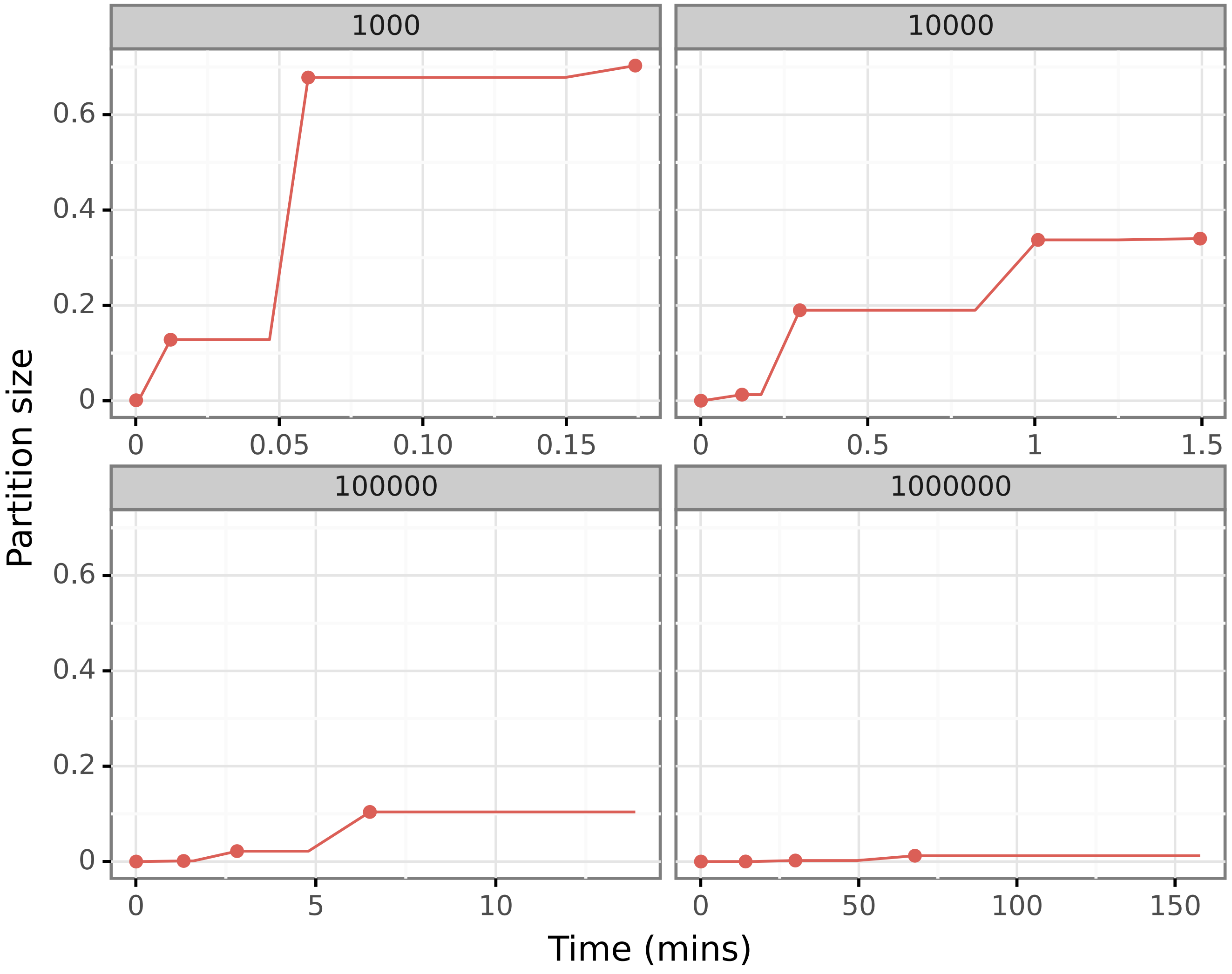


# Problem #1: Capacity Planning Problem

## # cuts over time




## Partition size over time



# Geometrical interpretation: Newsvendor

- We decide to buy a quantity $x_i$ of each items $i$. (1st stage)
  - If the demand is greater than $x_i$, we pay an opportunity cost of $p_i$ per unit.
  - If the demand is less than $x_i$, we pay a holding cost of $h_i$ per unit.

$$\min \sum_i c_i x_i + \mathbb{E}[p_i(\xi_i - x_i)^+ + h_i\,(x_i - \xi_i)^+]$$

# Geometrical interpretation: Newsvendor

- We decide to buy a quantity $x_i$ of each items $i$. (1st stage)
  - If the demand is greater than $x_i$, we pay an opportunity cost of $p_i$ per unit.
  - If the demand is less than $x_i$, we pay a holding cost of $h_i$ per unit.

$$\min \sum_i c_i x_i + \mathbb{E}[p_i(\xi_i - x_i)^+ + h_i\,(x_i - \xi_i)^+]$$

$$Q(x,\xi) = \min \sum_i p_i y_i^+ + h_i y_i^-$$

$$x_i + y_i^+ - y_i^- = \xi_i \quad \forall i$$

# Geometrical interpretation: Newsvendor

- We decide to buy a quantity $x_i$ of each items $i$. (1st stage)
  - If the demand is greater than $x_i$, we pay an opportunity cost of $p_i$ per unit.
  - If the demand is less than $x_i$, we pay a holding cost of $h_i$ per unit.

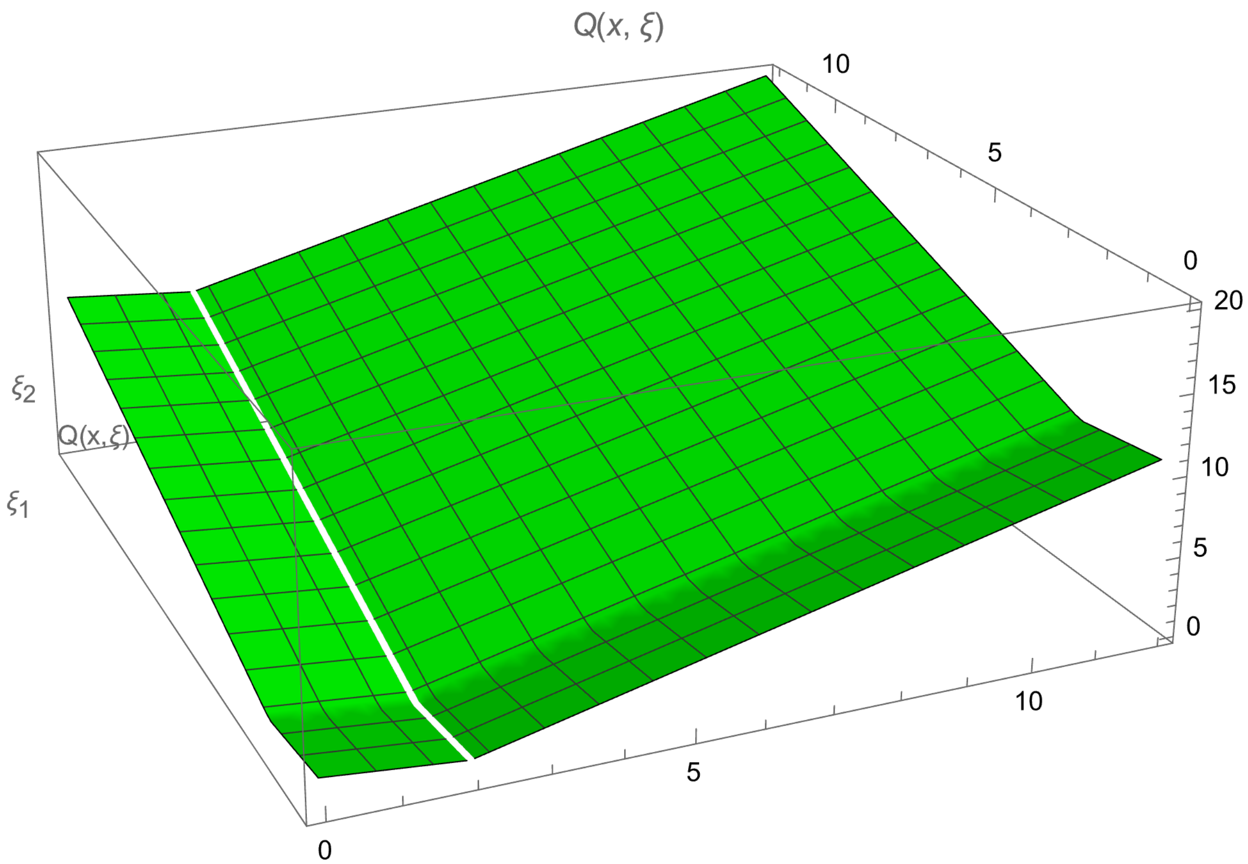
$$\min \sum_i c_i x_i + \mathbb{E}[p_i(\xi_i - x_i)^+ + h_i\ (x_i - \xi_i)^+]$$

$$Q(x,\xi) = \min \sum_i p_i y_i^+ + h_i y_i^-$$
$$x_i + y_i^+ - y_i^- = \xi_i \quad \forall i$$

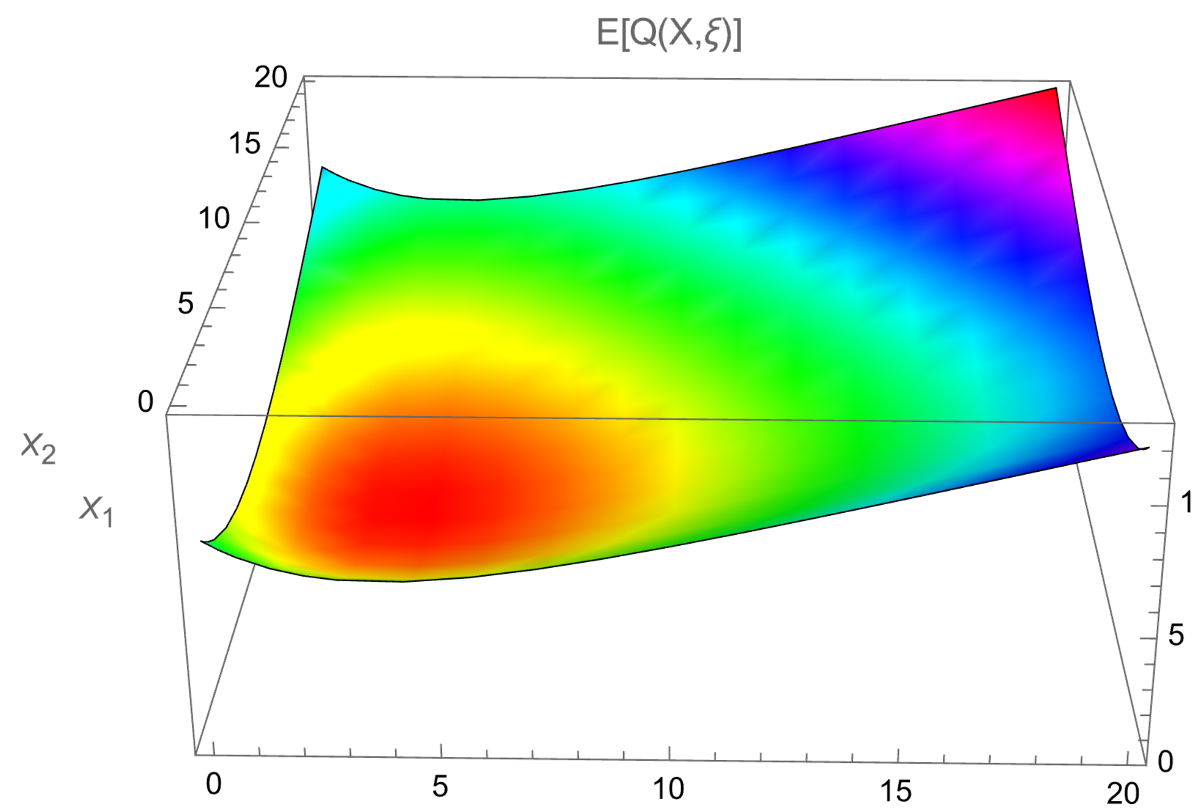
Dual solution: $\lambda_i = \begin{cases} -h_i & \xi_i < x_i \\ p_i & \xi_i \geq x_i \end{cases}$

# Example: Newsvendor

- 2 items, $p=1$, $h=0.4$.
- Demand: $\xi_1 \rightsquigarrow Exp\left(\frac{1}{3}\right), \xi_2 \rightsquigarrow Exp\left(\frac{1}{5}\right)$




$Q(x, \xi)$ for $x = (2,2)$

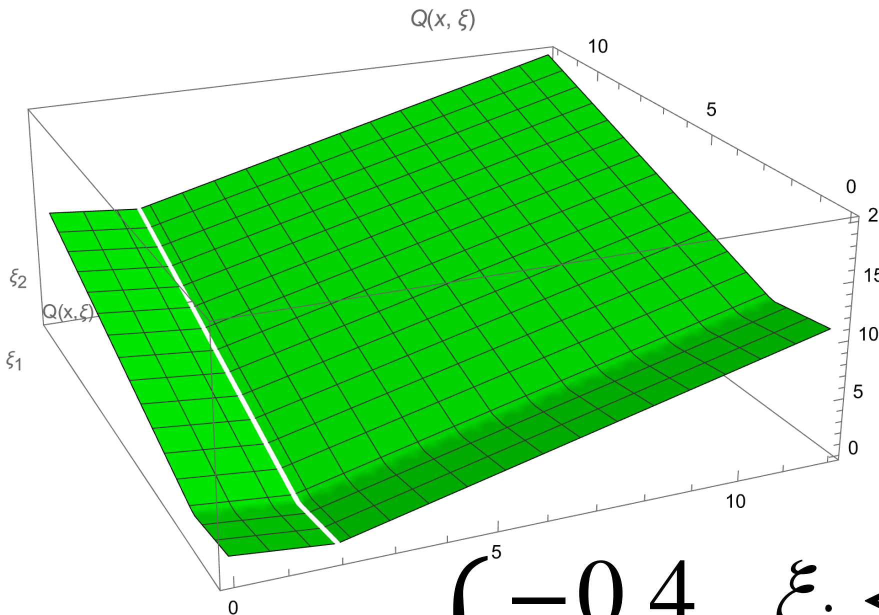



$\mathbb{E}\left[Q(x, \xi)\right]$

# Example: Newsvendor

"Single" Bender cut

$$(\xi_1 - x_1)\lambda_1 + (\xi_2 - x_2)\lambda_2 \leq \theta^\xi \quad \forall \xi$$




$$\lambda_i = \begin{cases} -0.4 & \xi_i < 2 \\ 1.0 & \xi_i \geq 2 \end{cases}$$

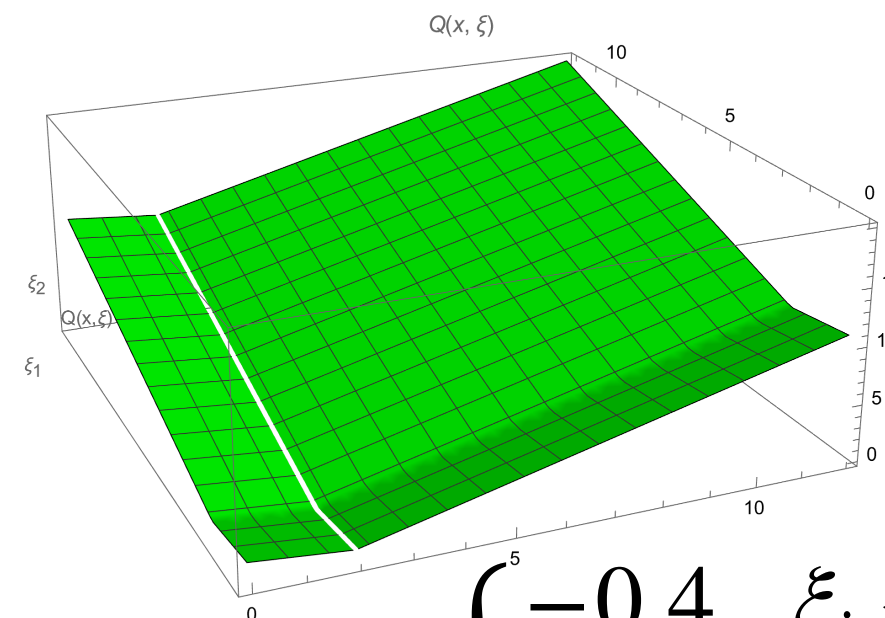
# Example: Newsvendor

"Single" Bender cut

$$(\xi_1 - x_1)\lambda_1 + (\xi_2 - x_2)\lambda_2 \leq \theta^{\xi} \quad \forall \xi$$

integrating over the distribution:

$$0.103 \cdot (23.3 - 3.1x1) + 0.134 \cdot (34.1 - 4.0x2) \leq \Theta$$

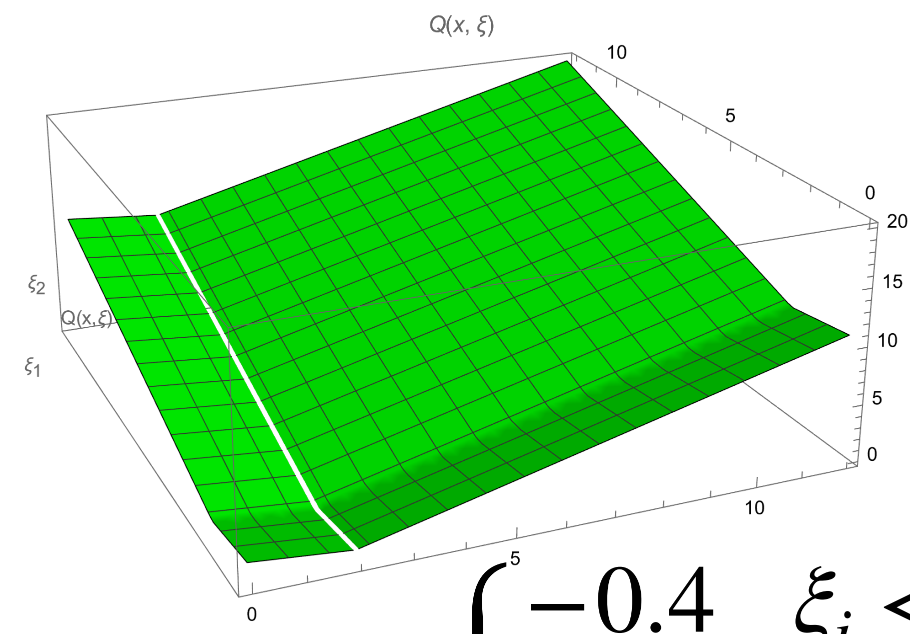



$$\lambda_i = \begin{cases} -0.4 & \xi_i < 2 \\ 1.0 & \xi_i \geq 2 \end{cases}$$

# Example: Newsvendor

"Single" Bender cut

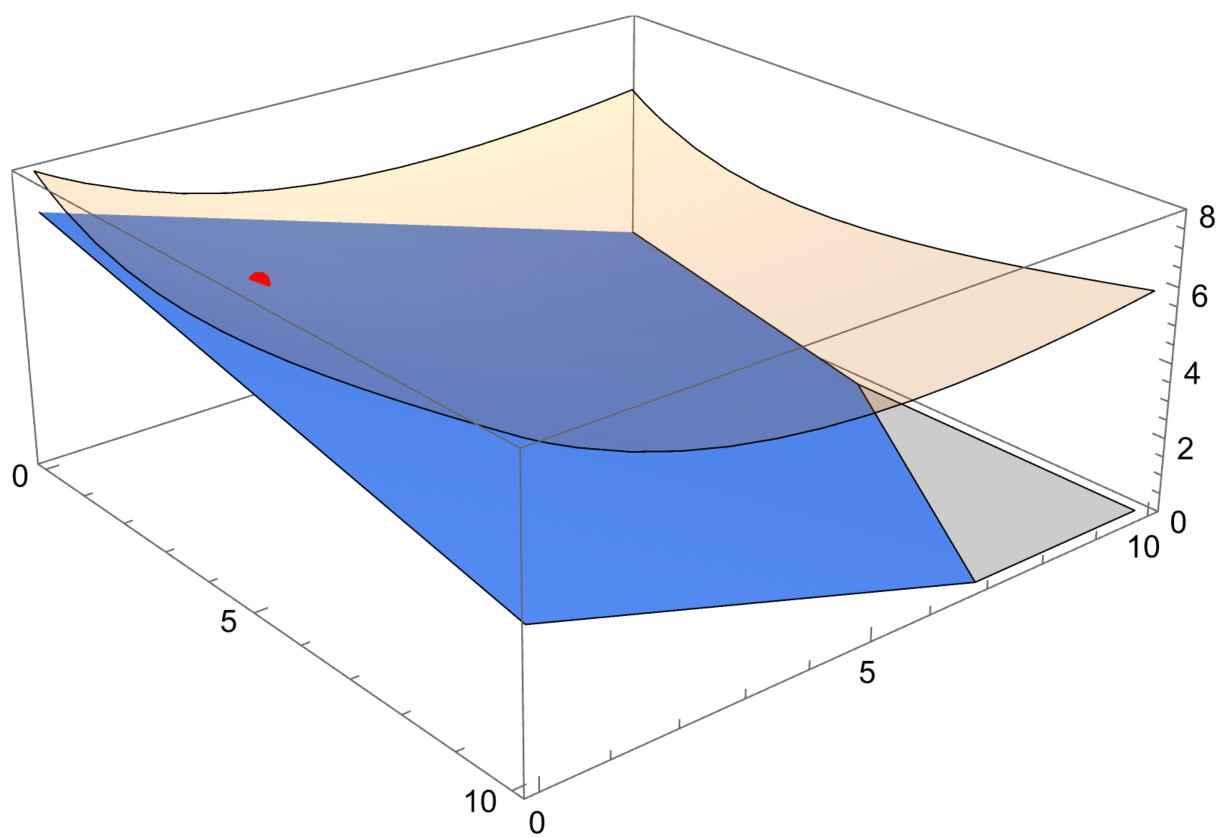
$$(\xi_1 - x_1)\lambda_1 + (\xi_2 - x_2)\lambda_2 \leq \theta^{\xi} \quad \forall \xi$$



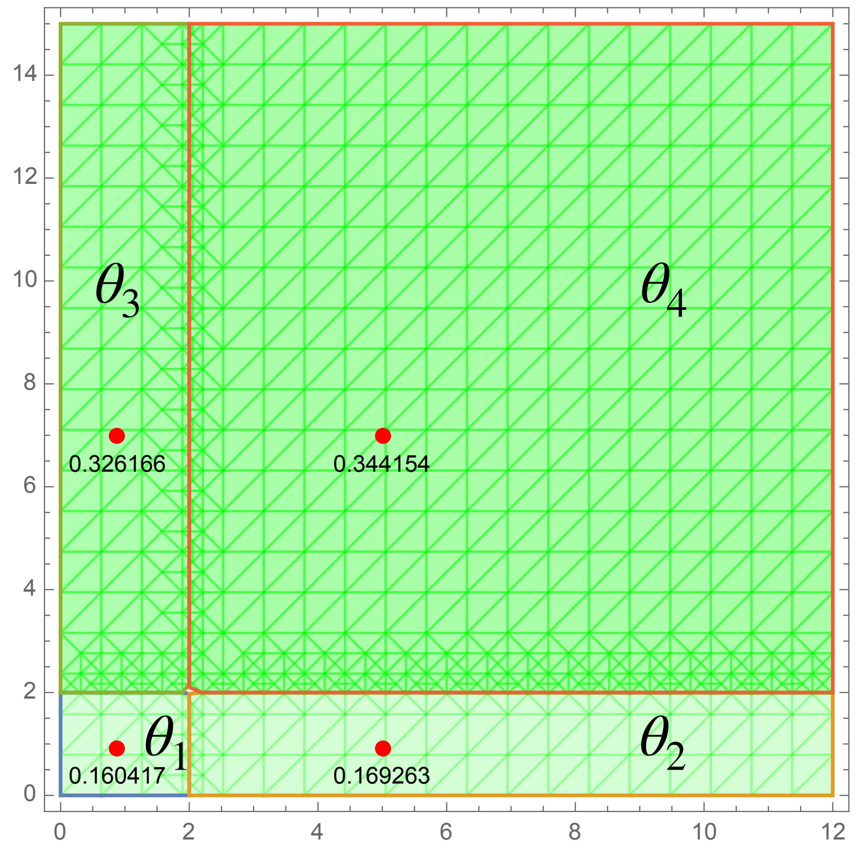

$$\lambda_i = \begin{cases} -0.4 & \xi_i < 2 \\ 1.0 & \xi_i \geq 2 \end{cases}$$

integrating over the distribution:

$$0.103 \cdot (23.3 - 3.1x1) + 0.134 \cdot (34.1 - 4.0x2) \leq \Theta$$

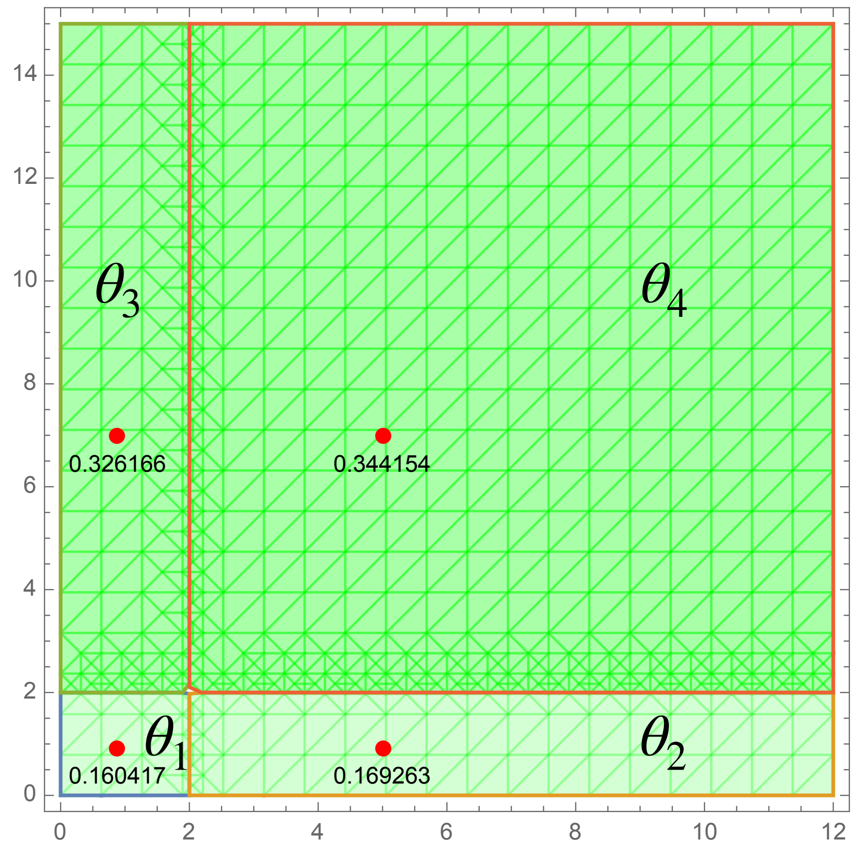


# Adaptive Bender cut




$$\lambda_i = \begin{cases} -0.4 & \xi_i < 2 \\ 1.0 & \xi_i \geq 2 \end{cases}$$

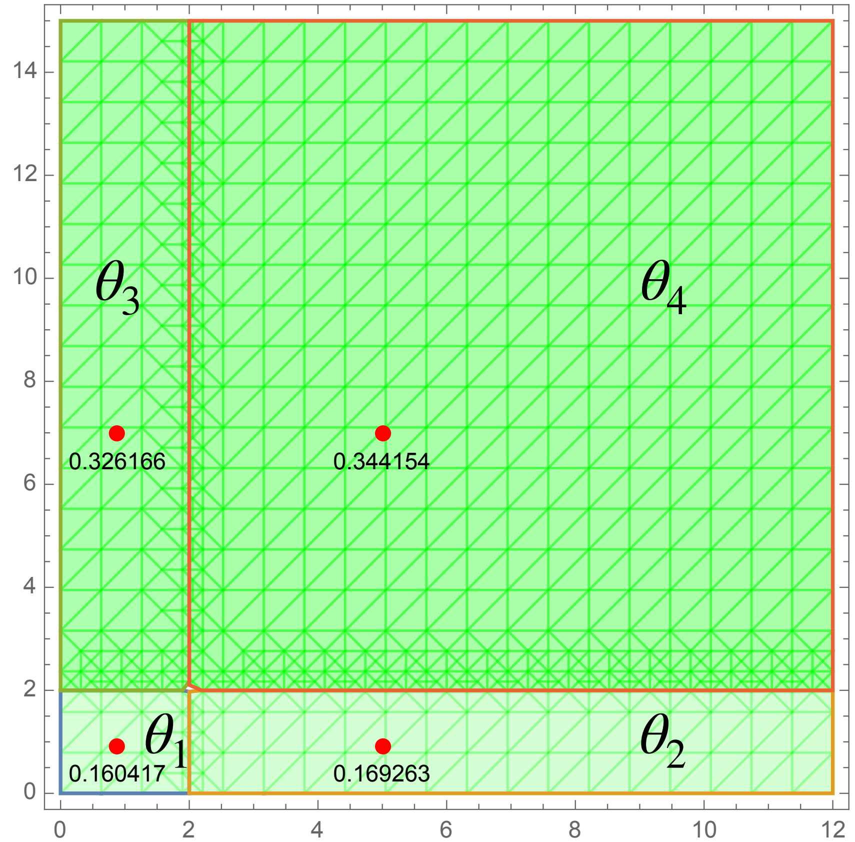
# Adaptive Bender cut




$$
\begin{aligned}
\min cx + 0.16\ \theta_1 + 0.17\ \theta_2 + 0.33\ \theta_3 + 0.34\ \theta_4 \\
-0.4(0.89 - x_1) - 0.4(0.93 - x_2) \le \theta_1 \\
5 - x_1 - 0.4(0.93 - x_2) \le \theta_2 \\
7 - 0.4(0.89 - x_1) - x_2 \le \theta_3 \\
12 - x_1 - x_2 \le \theta_4
\end{aligned}
$$

$$
\lambda_i = \begin{cases} -0.4 & \xi_i < 2 \\ 1.0 & \xi_i \ge 2 \end{cases}
$$

# Adaptive Bender cut




$$\min cx + 0.16\ \theta_1 + 0.17\ \theta_2 + 0.33\ \theta_3 + 0.34\ \theta_4$$

$$-0.4(0.89 - x_1) - 0.4(0.93 - x_2) \leq \theta_1$$

$$5 - x_1 - 0.4(0.93 - x_2) \leq \theta_2$$

$$7 - 0.4(0.89 - x_1) - x_2 \leq \theta_3$$

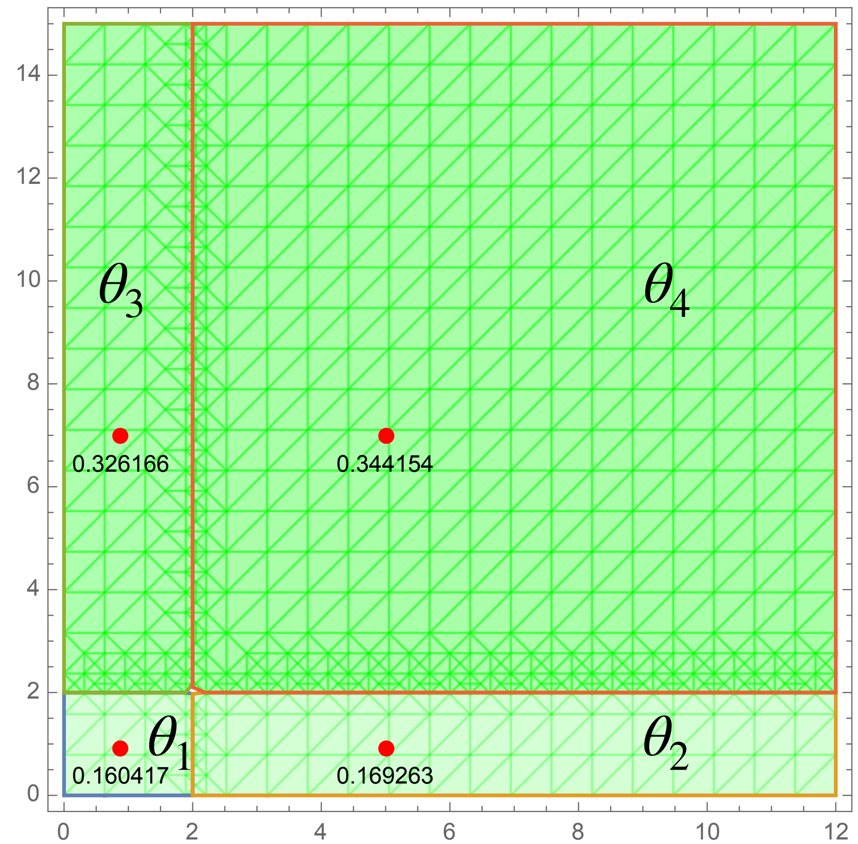
$$12 - x_1 - x_2 \leq \theta_4$$

equivalent to

$$\min cx + 0.103 \cdot (23.3 - 3.1\ x_1) + 0.134 \cdot (34.1 - 4.0\ x_2)$$

$$\lambda_i = \begin{cases} -0.4 & \xi_i < 2 \\ 1.0 & \xi_i \geq 2 \end{cases}$$

# Adaptive Bender cut

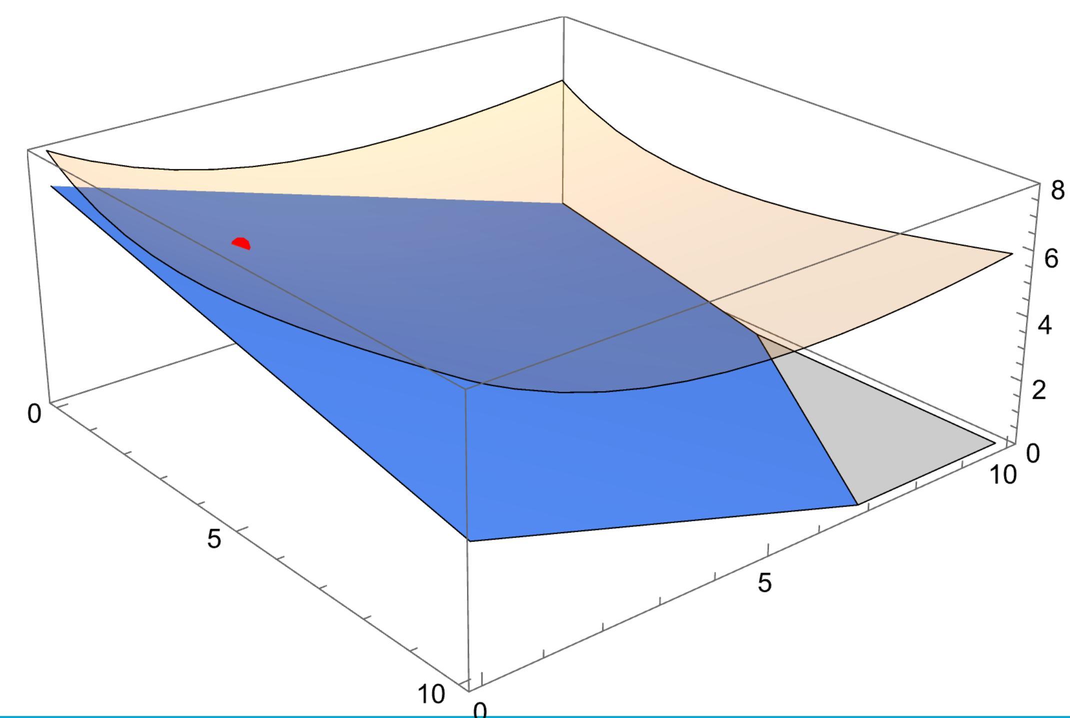



$$\min cx + 0.16\ \theta_1 + 0.17\ \theta_2 + 0.33\ \theta_3 + 0.34\ \theta_4$$
$$-0.4(0.89 - x_1) - 0.4(0.93 - x_2) \le \theta_1$$
$$5 - x_1 - 0.4(0.93 - x_2) \le \theta_2$$
$$7 - 0.4(0.89 - x_1) - x_2 \le \theta_3$$
$$12 - x_1 - x_2 \le \theta_4$$

equivalent to

$$\min cx + 0.103 \cdot (23.3 - 3.1\ x_1) + 0.134 \cdot (34.1 - 4.0\ x_2)$$

$$\lambda_i = \begin{cases} -0.4 & \xi_i < 2 \\ 1.0 & \xi_i \ge 2 \end{cases}$$



# New point: $x = (6,8)$

"Single" Bender cut

$$(\xi_1 - x_1)\lambda_1 + (\xi_2 - x_2)\lambda_2 \leq \theta^{\xi} \quad \forall \xi$$

integrating, we obtain

$$0.027(18.7 + 7.8\ x_1) + 0.04(41.5 + 2.9\ x_2) \leq \Theta$$

$$\lambda_1 = \begin{cases} -0.4 & \xi_i < 6 \\ 1.0 & \xi_i \geq 6 \end{cases}$$

$$\lambda_2 = \begin{cases} -0.4 & \xi_i < 8 \\ 1.0 & \xi_i \geq 8 \end{cases}$$

# New point: $x = (6,8)$

"Single" Bender cut

$$(\xi_1 - x_1)\lambda_1 + (\xi_2 - x_2)\lambda_2 \leq \theta^{\xi} \quad \forall \xi$$

integrating, we obtain

$$0.027(18.7 + 7.8\ x_1) + 0.04(41.5 + 2.9\ x_2) \leq \Theta$$

$$\lambda_1 = \begin{cases} -0.4 & \xi_i < 6 \\ 1.0 & \xi_i \geq 6 \end{cases}$$

$$\lambda_2 = \begin{cases} -0.4 & \xi_i < 8 \\ 1.0 & \xi_i \geq 8 \end{cases}$$

$$0.103\ (23.3 - 3.1\ x_1) + 0.134\ (34.1 - 4.0\ x_2) \leq \Theta$$

# New point: $x = (6,8)$

"Single" Bender cut

$$(\xi_1 - x_1)\lambda_1 + (\xi_2 - x_2)\lambda_2 \leq \theta^{\xi} \quad \forall \xi$$

integrating, we obtain

$$0.027(18.7 + 7.8\ x_1) + 0.04(41.5 + 2.9\ x_2) \leq \Theta$$

$$\lambda_1 = \begin{cases} -0.4 & \xi_i < 6 \\ 1.0 & \xi_i \geq 6 \end{cases}$$

$$\lambda_2 = \begin{cases} -0.4 & \xi_i < 8 \\ 1.0 & \xi_i \geq 8 \end{cases}$$

$$0.103\ (23.3 - 3.1\ x_1) + 0.134\ (34.1 - 4.0\ x_2) \leq \Theta$$

$$0.027(18.7 + 7.8\ x_1) + 0.04(41.5 + 2.9\ x_2) \leq \Theta$$

# New point: $x = (6,8)$

"Single" Bender cut

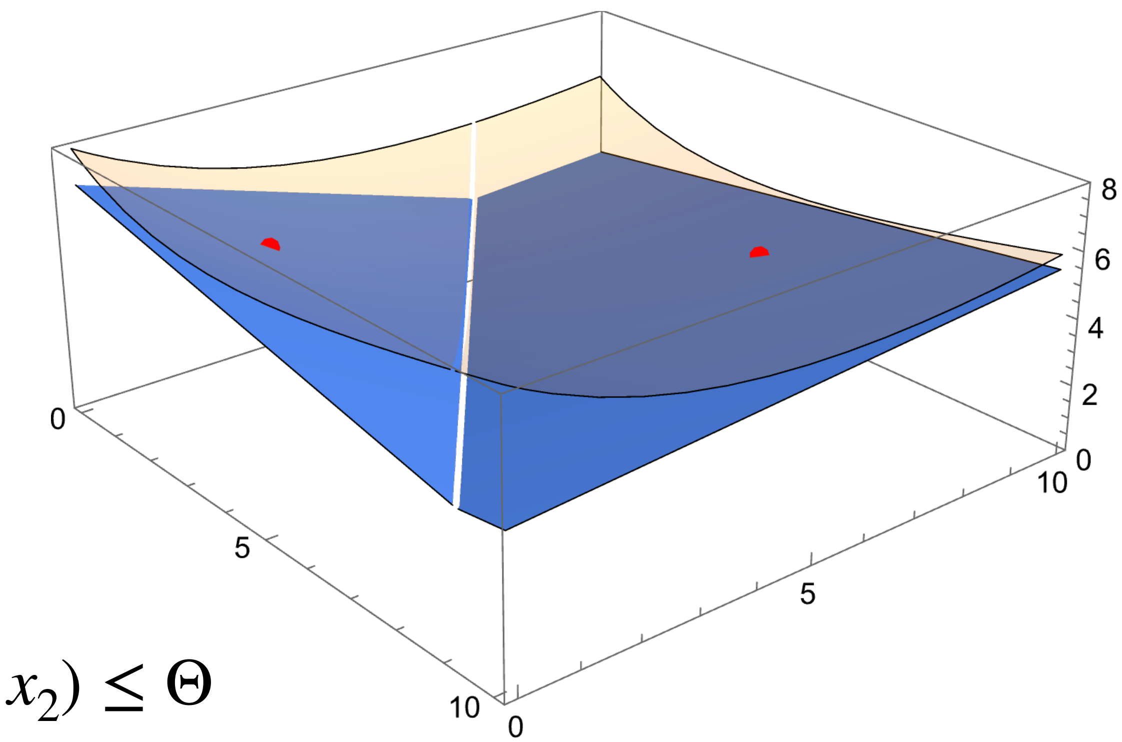
$$(\xi_1 - x_1)\lambda_1 + (\xi_2 - x_2)\lambda_2 \leq \theta^{\xi} \quad \forall \xi$$

$$\lambda_1 = \begin{cases} -0.4 & \xi_i < 6 \\ 1.0 & \xi_i \geq 6 \end{cases}$$

$$\lambda_2 = \begin{cases} -0.4 & \xi_i < 8 \\ 1.0 & \xi_i \geq 8 \end{cases}$$

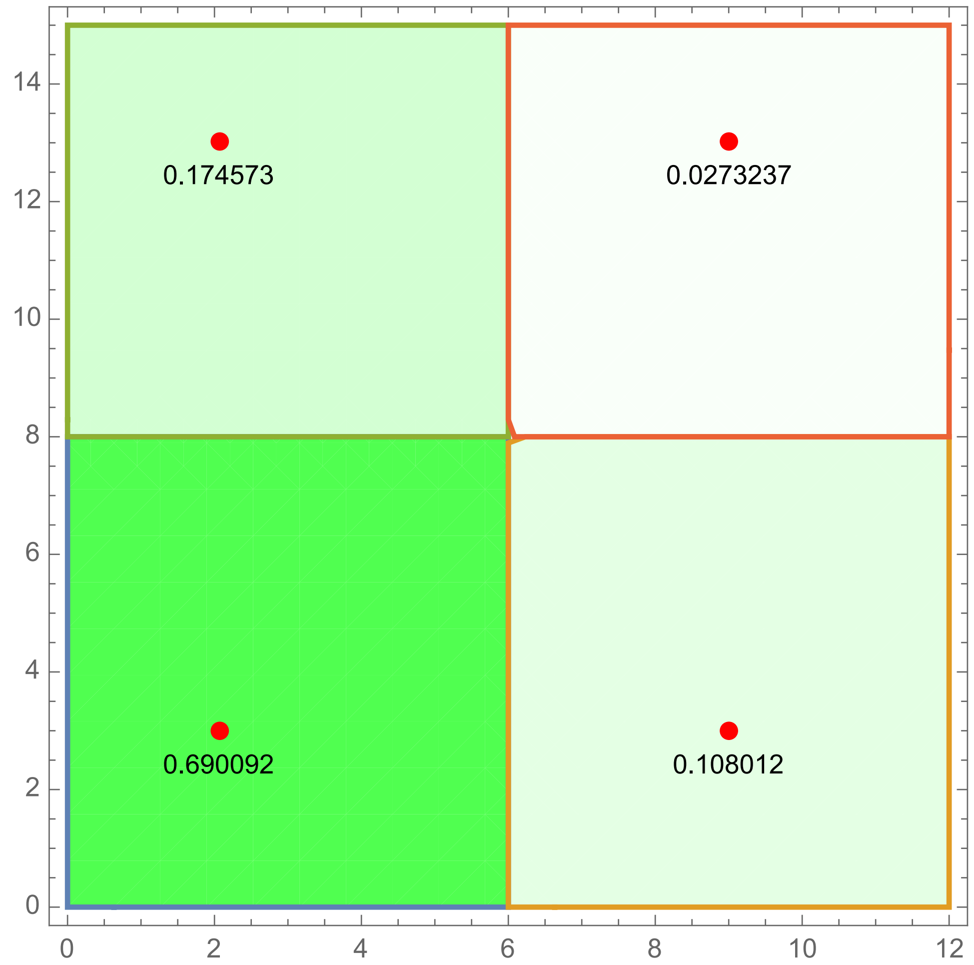
integrating, we obtain

$$0.027(18.7 + 7.8\ x_1) + 0.04(41.5 + 2.9\ x_2) \leq \Theta$$

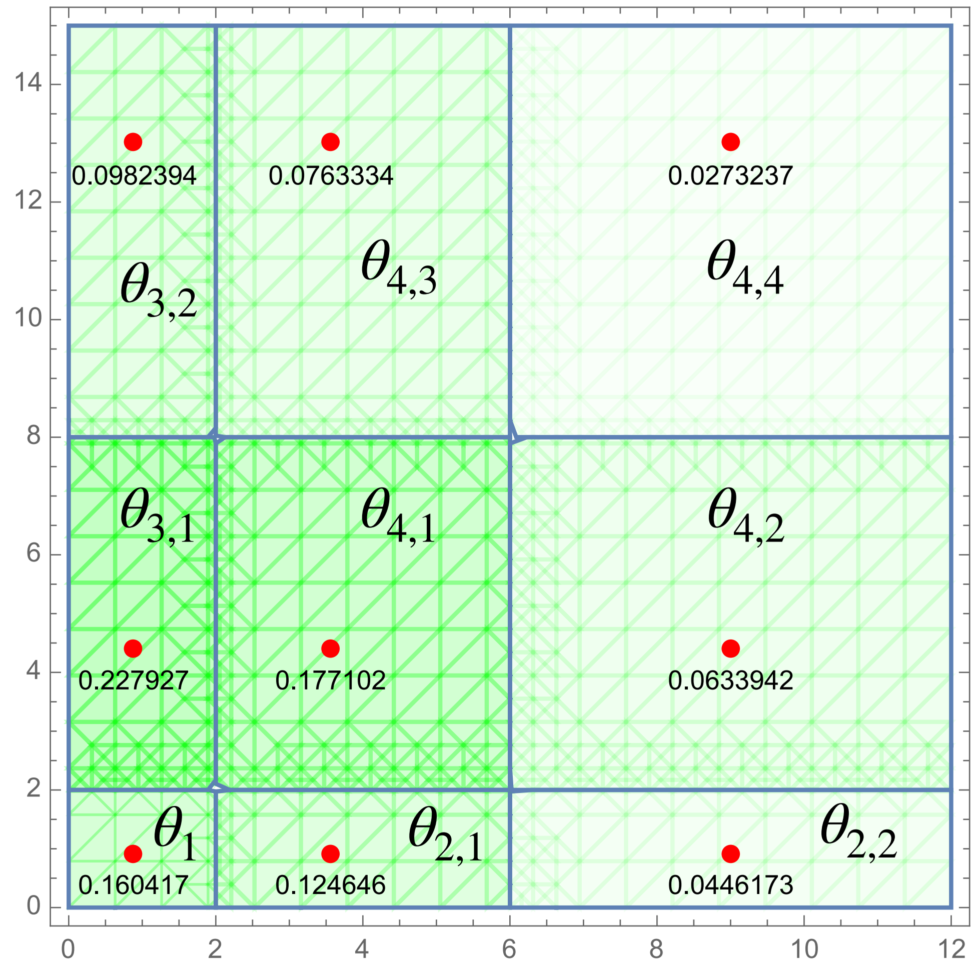



$$0.103\ (23.3 - 3.1\ x_1) + 0.134\ (34.1 - 4.0\ x_2) \leq \Theta$$
$$0.027(18.7 + 7.8\ x_1) + 0.04(41.5 + 2.9\ x_2) \leq \Theta$$

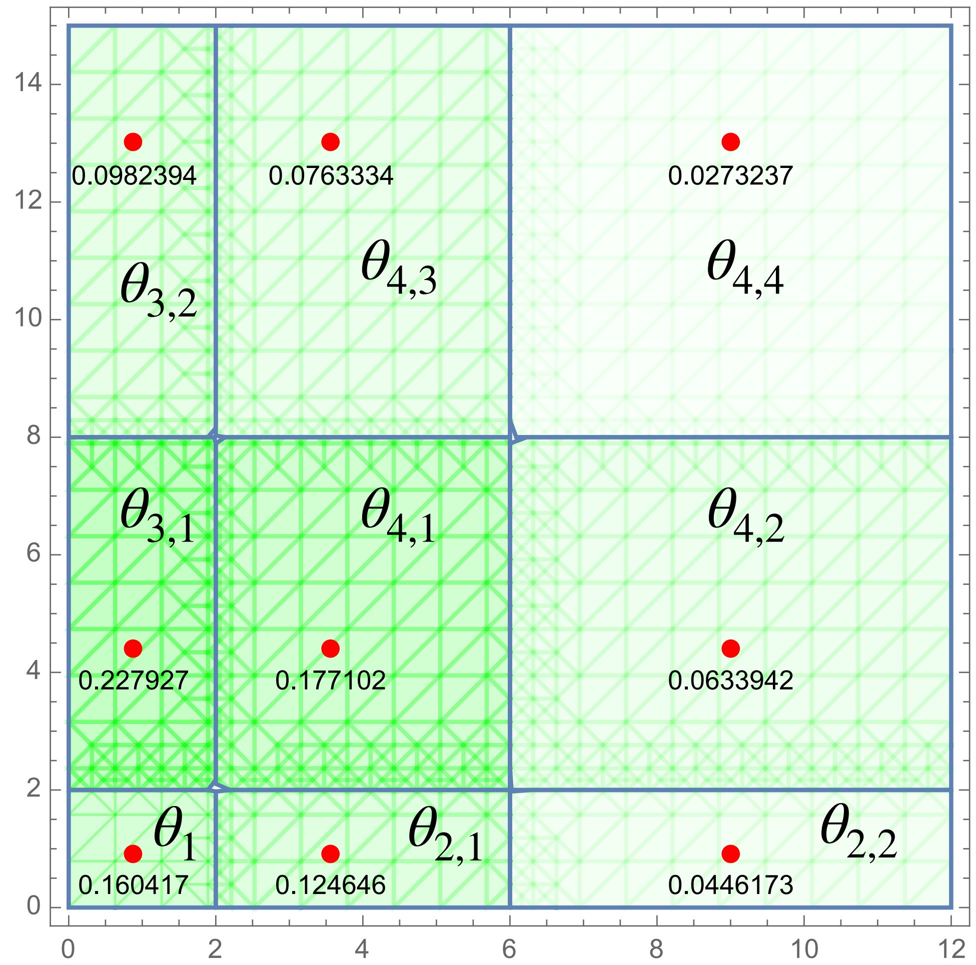
# Adaptive Bender cut



# Adaptive Bender cut

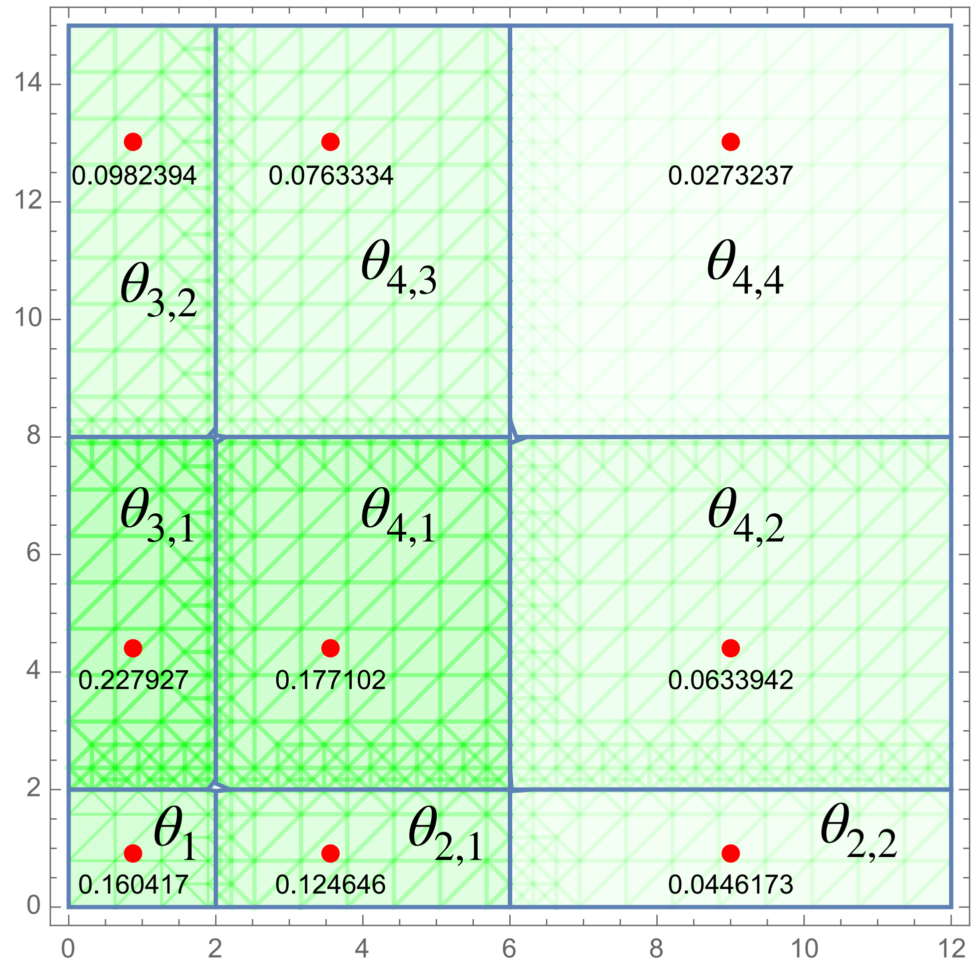


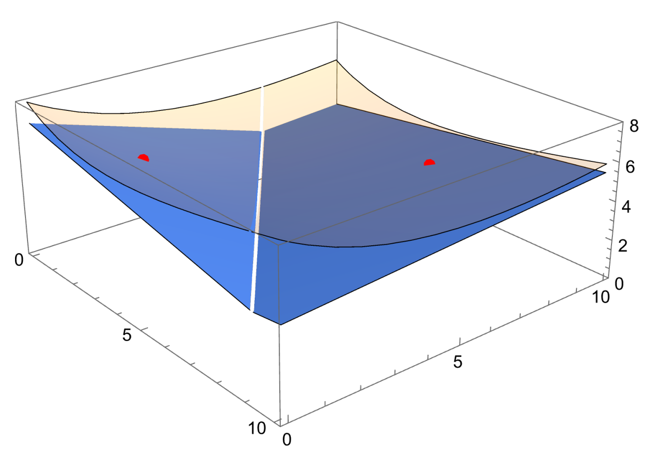
# Adaptive Bender cut




$$-0.73 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_1$$
$$-1.8 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{2,1}$$
$$8.63 - x_1 + 0.4\ x_2 \leq \theta_{2,2}$$
$$-2.12 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{3,1}$$
$$12.64 + 0.4\ x_1 - x_2 \leq \theta_{3,2}$$
$$-3.19 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{4,1}$$
$$7.23 - x_1 + 0.4\ x_2 \leq \theta_{4,2}$$
$$11.57 + 0.4\ x_1 - x_2 \leq \theta_{4,3}$$
$$22 - x_1 - x_2 \leq \theta_{4,4}$$

# Adaptive Bender cut



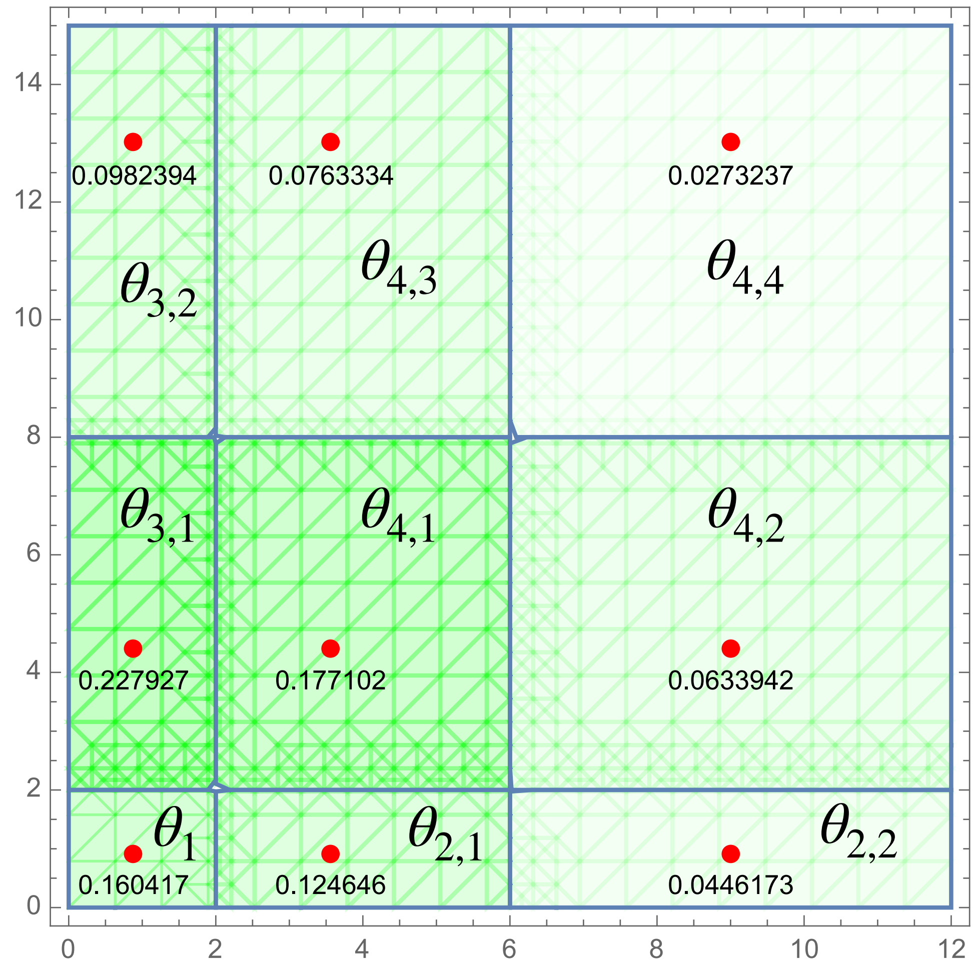



$$-0.73 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_1$$
$$-1.8 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{2,1}$$
$$8.63 - x_1 + 0.4\ x_2 \leq \theta_{2,2}$$
$$-2.12 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{3,1}$$
$$12.64 + 0.4\ x_1 - x_2 \leq \theta_{3,2}$$
$$-3.19 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{4,1}$$
$$7.23 - x_1 + 0.4\ x_2 \leq \theta_{4,2}$$
$$11.57 + 0.4\ x_1 - x_2 \leq \theta_{4,3}$$
$$22 - x_1 - x_2 \leq \theta_{4,4}$$

equivalent to

$$0.027(18.7 + 7.8\ x_1) + 0.04(41.5 + 2.9\ x_2) \leq \Theta$$

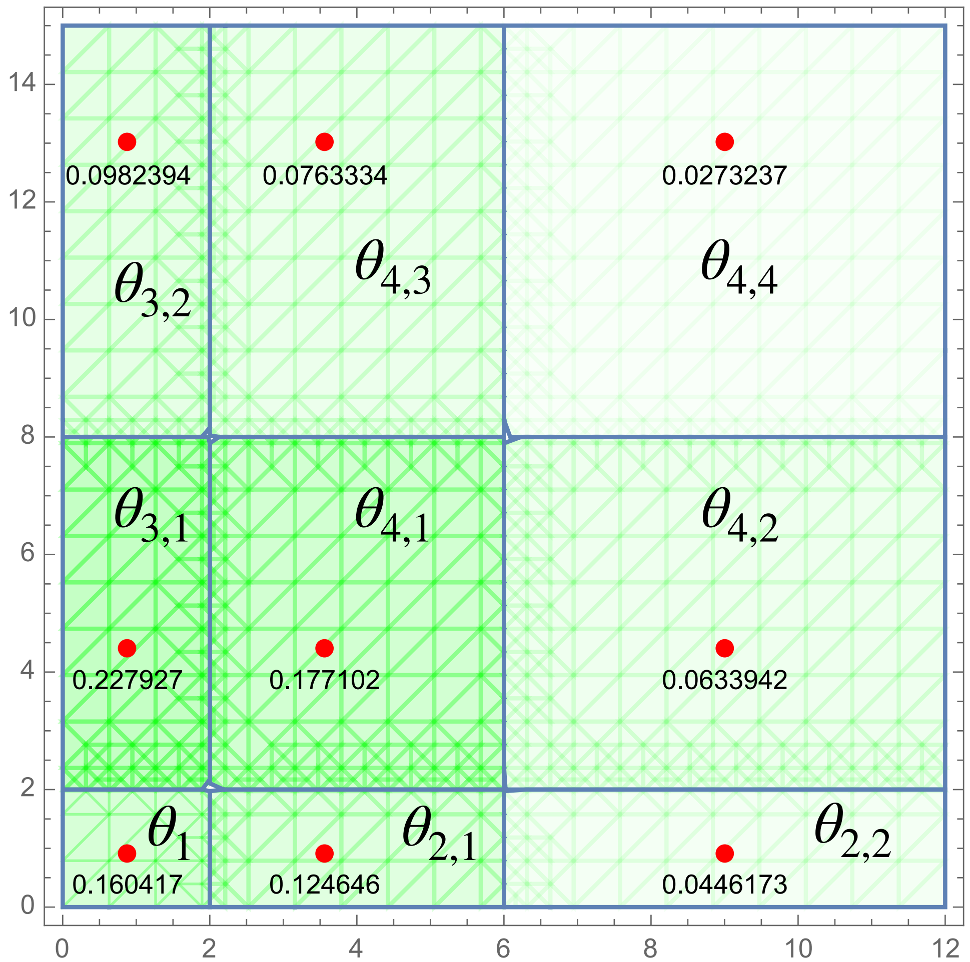
# Adaptive Bender cut




## But due to refinement

$$-0.73 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_1$$
$$-1.8 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{2,1}$$
$$8.63 - x_1 + 0.4\ x_2 \leq \theta_{2,2}$$
$$-2.12 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{3,1}$$
$$12.64 + 0.4\ x_1 - x_2 \leq \theta_{3,2}$$
$$-3.19 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{4,1}$$
$$7.23 - x_1 + 0.4\ x_2 \leq \theta_{4,2}$$
$$11.57 + 0.4\ x_1 - x_2 \leq \theta_{4,3}$$
$$22 - x_1 - x_2 \leq \theta_{4,4}$$

# Adaptive Bender cut



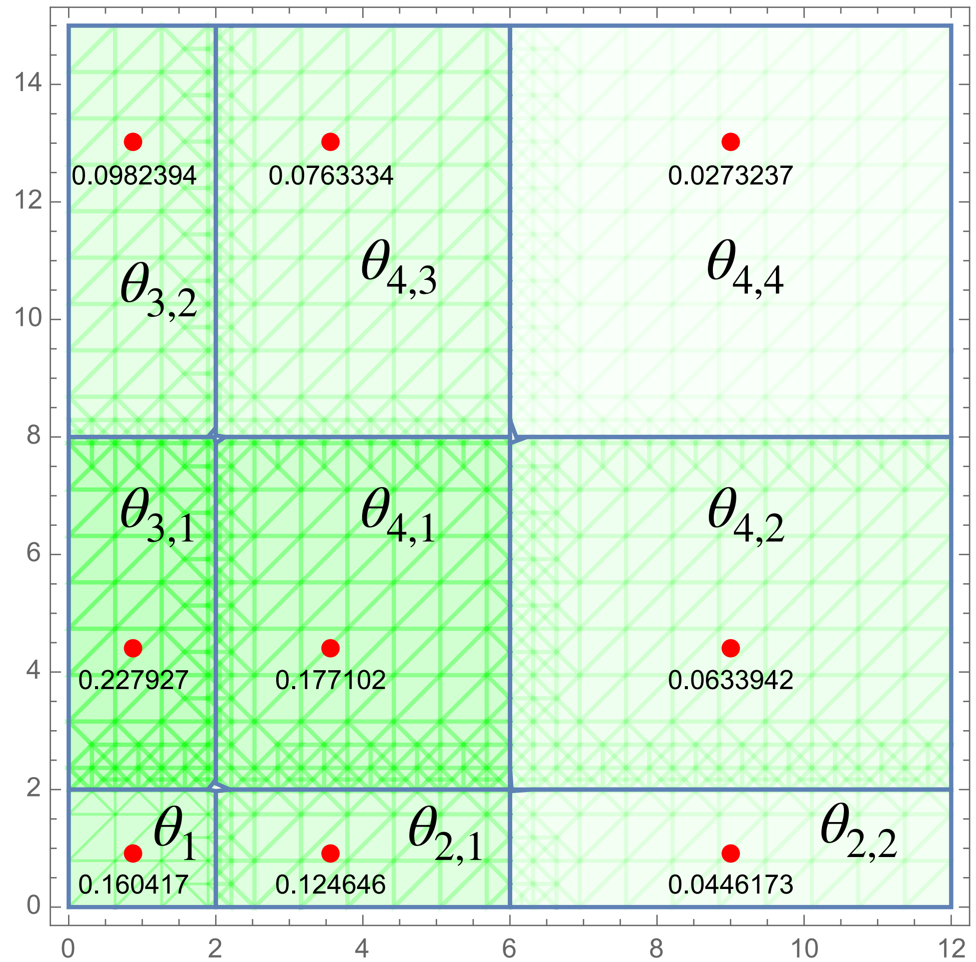

But due to refinement

$$-0.73 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_1$$
$$-1.8 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{2,1}$$
$$8.63 - x_1 + 0.4\ x_2 \leq \theta_{2,2}$$
$$-2.12 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{3,1}$$
$$12.64 + 0.4\ x_1 - x_2 \leq \theta_{3,2}$$
$$-3.19 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{4,1}$$
$$7.23 - x_1 + 0.4\ x_2 \leq \theta_{4,2}$$
$$11.57 + 0.4\ x_1 - x_2 \leq \theta_{4,3}$$
$$22 - x_1 - x_2 \leq \theta_{4,4}$$

$$-0.4(0.89 - x_1) - 0.4(0.93 - x_2) \leq \theta_1$$
$$5 - x_1 - 0.4(0.93 - x_2) \leq \theta_2$$
$$7 - 0.4(0.89 - x_1) - x_2 \leq \theta_3$$
$$12 - x_1 - x_2 \leq \theta_4$$

(Previous step cuts)

# Adaptive Bender cut

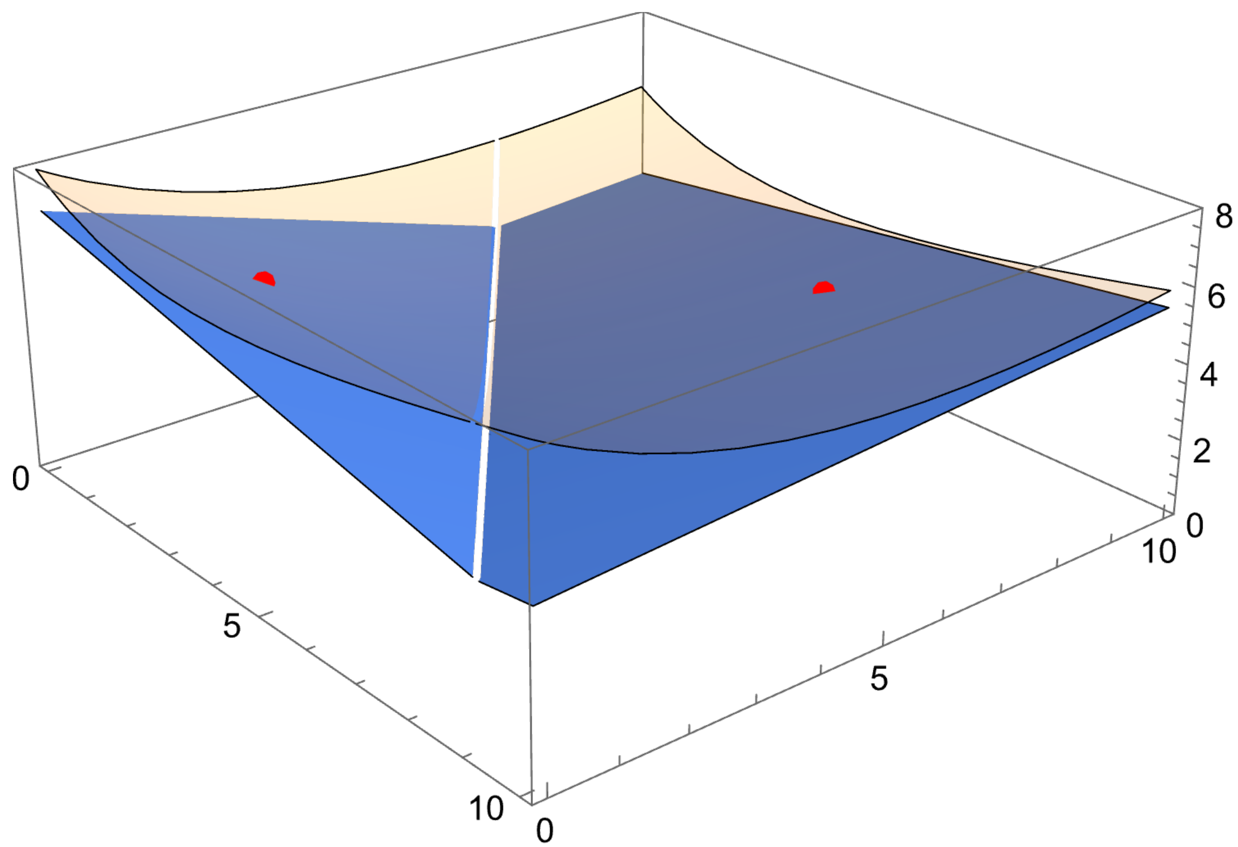



But due to refinement

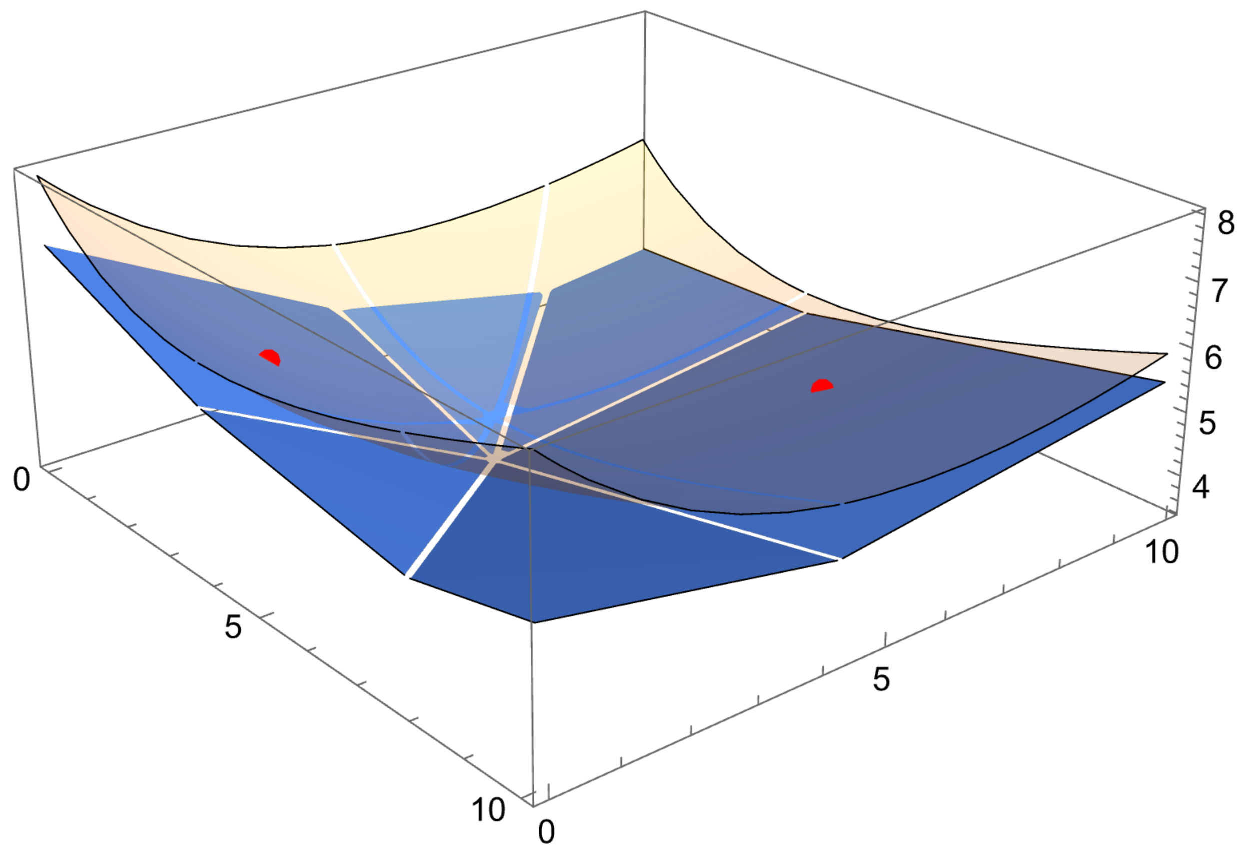
$$-0.73 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_1$$
$$-1.8 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{2,1}$$
$$8.63 - x_1 + 0.4\ x_2 \leq \theta_{2,2}$$
$$-2.12 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{3,1}$$
$$12.64 + 0.4\ x_1 - x_2 \leq \theta_{3,2}$$
$$-3.19 + 0.4\ x_1 + 0.4\ x_2 \leq \theta_{4,1}$$
$$7.23 - x_1 + 0.4\ x_2 \leq \theta_{4,2}$$
$$11.57 + 0.4\ x_1 - x_2 \leq \theta_{4,3}$$
$$22 - x_1 - x_2 \leq \theta_{4,4}$$

$$-0.4(0.89 - x_1) - 0.4(0.93 - x_2) \leq \theta_1$$
$$5 - x_1 - 0.4(0.93 - x_2) \leq 0.74\ \theta_{2,1} + 0.26\ \theta_{2,2}$$
$$7 - 0.4(0.89 - x_1) - x_2 \leq 0.70\ \theta_{3,1} + 0.30\ \theta_{3,2}$$
$$12 - x_1 - x_2 \leq 0.52\ \theta_{4,1} + 0.18\ \theta_{4,2} + 0.22\ \theta_{4,3} + 0.08\ \theta_{4,4}$$

# Adaptive Bender cut




Single cut Bender



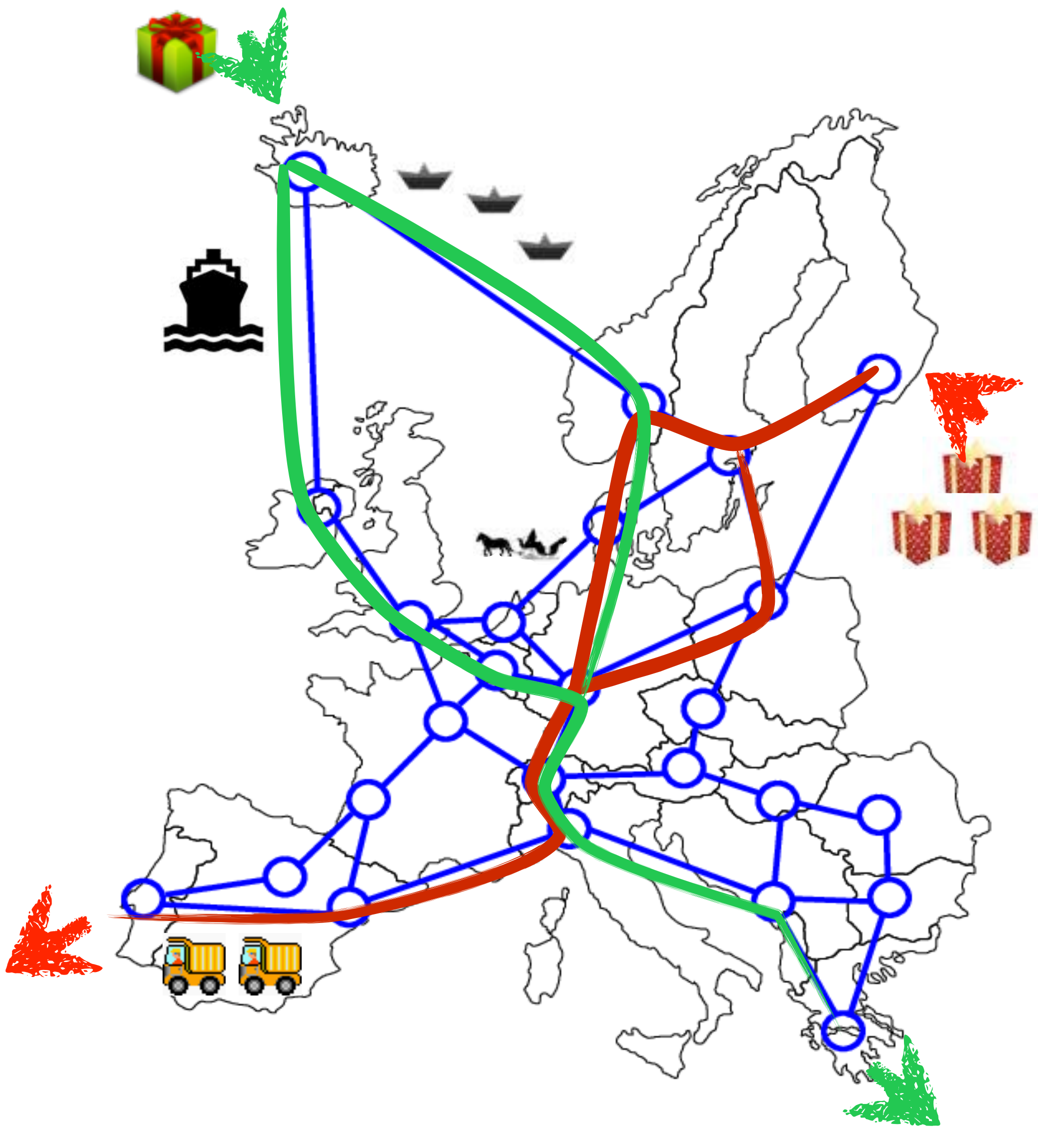

Adaptive cut Bender

# Conclusions

- Adaptive Benders combines the strengths of Benders multi and single cut:
  - a small problem on early iterations (like single-cut Benders) which is quicker to solve and requires fewer feasibility cuts.
  - a disaggregated problem on last iterations (like multi-cut Benders) to prove the optimality when required
- Can be used to solve problems with continuous distributions approximating them by discrete scenarios
- Still a Benders problem, so all common "tricks" applies

More details: Ramirez-Pico, C., Ljubić, I. and Moreno, E.,
*Benders Adaptive-Cuts Method for Two-Stage Stochastic Programs*.
Transportation Science 57(5) 115-1401, 2023.

# Problem #2: Stochastic multicommodity flow



$$\min_{x\in[0,1]^{|E|}} \sum_{ij\in E} f_{ij}x_{ij} + \sum_{s\in\mathcal{S}} p^s Q(x,\xi^s),$$

$$Q(\hat{x},\xi^s) := \min_{y\geq 0} \sum_{ij\in E}\sum_{k\in\mathbb{K}} c_{ijk}y^s_{ijk}$$

$$\sum_{j:ij\in E} y^s_{ijk} - \sum_{j:ji\in E} y^s_{jik} = \begin{cases} d^s_k & \text{if } i = O(k) \\ -d^s_k & \text{if } i = D(k) \\ 0 & \text{otherwise} \end{cases}$$

$$\sum_{k\in\mathbb{K}} y^s_{ijk} \leq u_{ij}\hat{x}_{ij} \qquad ij\in E$$

- First stage: arc capacities
- Second-stage: flow to satisfy demand

# Problem #2: Stochastic multicommodity flow

- Dual subproblem

$$Q(\hat{x},\xi^s) := \max_{\lambda^s\in\mathbb{R}^{|V||\mathbb{K}|},\mu^s\in\mathbb{R}_+^{|E|}} \sum_{k\in\mathbb{K}} d_k^s(\lambda^s_{O(k),k} - \lambda^s_{D(k),k}) - \sum_{ij\in E} (u_{ij}\hat{x}_{ij})\,\mu^s_{ij}$$

$$\lambda^s_{ik} - \lambda^s_{jk} - \mu^s_{ij} \leq c_{ijk} \qquad ij\in E,\ k\in\mathbb{K}$$

- Benders cuts

Multi-cut: $$\sum_{k\in\mathbb{K}} d_k^s\left(\hat{\lambda}^s_{O(k),k} - \hat{\lambda}^s_{D(k),k}\right) - \sum_{ij\in E} u_{ij}\hat{\mu}^s_{ij}x_{ij} \leq \theta^s \quad s\in\mathcal{S}$$

Single-cut: $$\sum_{s\in\mathcal{S}} p^s\left[\sum_{k\in\mathbb{K}} d_k^s\left(\hat{\lambda}^s_{O(k),k} - \hat{\lambda}^s_{D(k),k}\right) - \sum_{ij\in E} u_{ij}\hat{\mu}^s_{ij}x_{ij}\right] \leq \sum_{s\in\mathcal{S}} p^s\theta^s$$

Adaptive-cut: $$p^P\left[\sum_{k\in\mathbb{K}} d_k^P\left(\hat{\lambda}^P_{O(k),k} - \hat{\lambda}^P_{D(k),k}\right) - \sum_{ij\in E} u_{ij}\hat{\mu}^P_{ij}x_{ij}\right] \leq \sum_{s\in P} p^s\theta^s \qquad P\in\mathcal{P}$$

# Problem #2: Stochastic multicommodity flow

- Subproblem

$$Q(\hat{x}, \xi^s) := \min_{y \geq 0} \sum_{ij \in E} \sum_{k \in \mathbb{K}} c_{ijk} y^s_{ijk}$$

$$\sum_{j: ij \in E} y^s_{ijk} - \sum_{j: ji \in E} y^s_{jik} = \begin{cases} d^s_k & \text{if } i = O(k) \\ -d^s_k & \text{if } i = D(k) \\ 0 & \text{otherwise} \end{cases} \qquad i \in V, k \in \mathbb{K} \qquad (\lambda)$$

$$\sum_{k \in \mathbb{K}} y^s_{ijk} \leq u_{ij} \hat{x}_{ij} \qquad ij \in E \qquad (\mu)$$

- Required conditions for the partition

$$\sum_{k \in \mathcal{K}} \left[ \left( \sum_{s \in P} \frac{p^s}{p^P} d^s_k \right) \cdot \left( \sum_{s \in P} \frac{p^s}{p^P} \left( \lambda^s_{O(k),k} - \lambda^s_{D(k),k} \right) \right) \right] = \sum_{k \in \mathcal{K}} \left[ \sum_{s \in P} \frac{p^s}{p^P} \cdot d^s_k \cdot \left( \lambda^s_{O(k),k} - \lambda^s_{D(k),k} \right) \right]$$

Rule: Refine $P$ with scenarios having same vector of duals

$$\left( \lambda^s_{O(k),k} - \lambda^s_{D(k),k} \right)_{k \in \mathcal{K}}$$

# Problem #2: Stochastic multicommodity flow

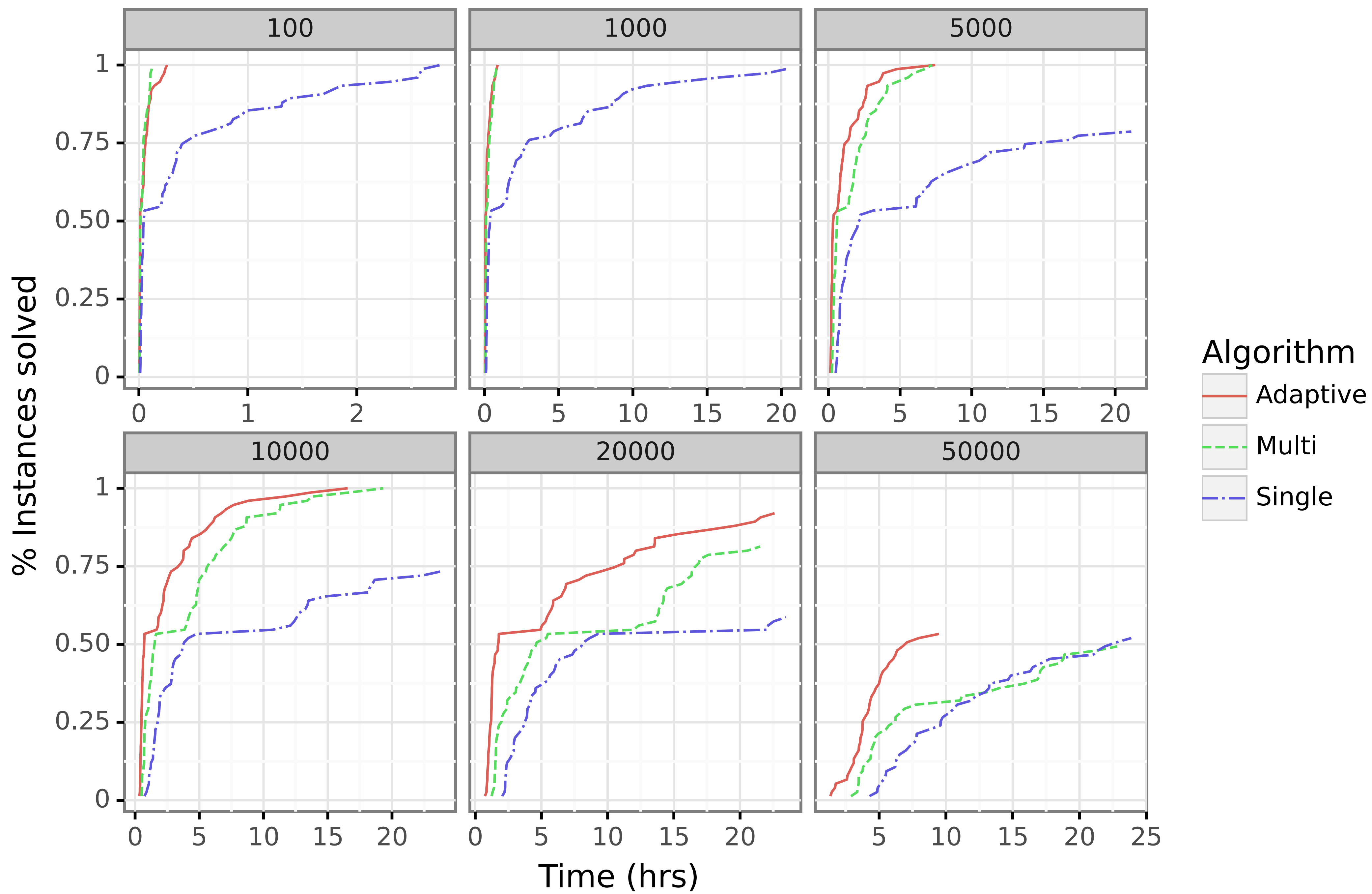
- Dataset: From Rahmaniani et al (2018)
  - Based on classic R instances (r04,r05,r06)
  - 5 configurations of costs and capacities (l1,l3,l5,l7,l9)
  - 5 different correlations (0% 20% 40% 60% 80%)
  - 100, 1,000, 5,000, 10,000, 20,000 and 50.000 scenarios
- Implemented in Python using Gurobi as Solver
- Source codes available at https://github.com/borelian/AdaptiveBenders

| $x$ | nodes | arcs | commodities |
|---|---|---|---|
| 01 | 10 | 35 | 10 |
| 01 | 10 | 35 | 25 |
| 03 | 10 | 35 | 50 |
| 04 | 10 | 60 | 10 |
| 05 | 10 | 60 | 25 |
| 06 | 10 | 60 | 50 |
| 07 | 10 | 82 | 10 |
| 08 | 10 | 83 | 25 |
| 09 | 10 | 83 | 50 |
| 10 | 20 | 120 | 40 |

| $y$ | fixed cost | capacity |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 5 | 1 |
| 3 | 10 | 1 |
| 4 | 1 | 2 |
| 5 | 5 | 2 |
| 6 | 10 | 2 |
| 7 | 1 | 8 |
| 8 | 5 | 8 |
| 9 | 10 | 8 |

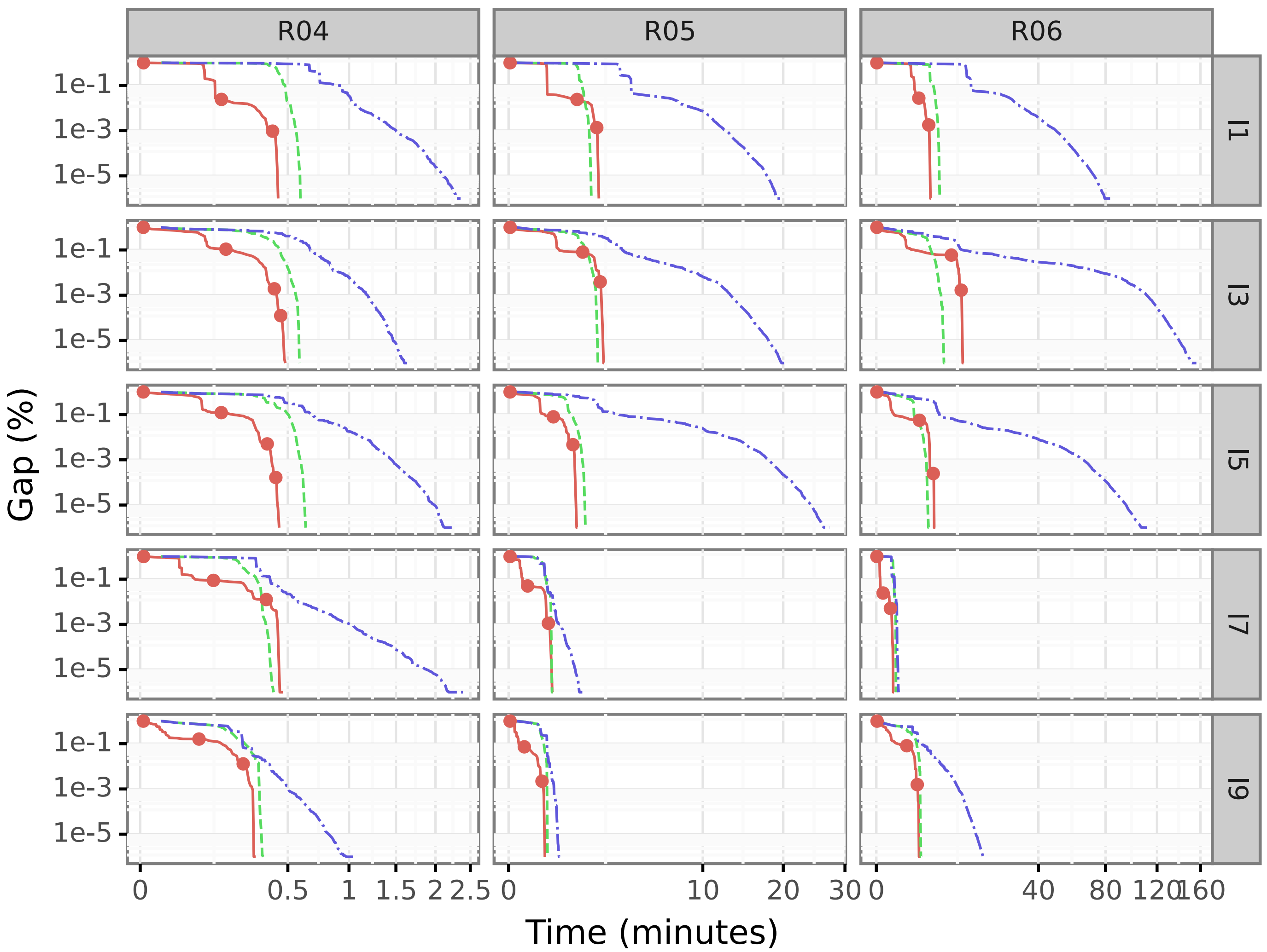
# Problem #2: Stochastic multicommodity flow

## Instances solved over time

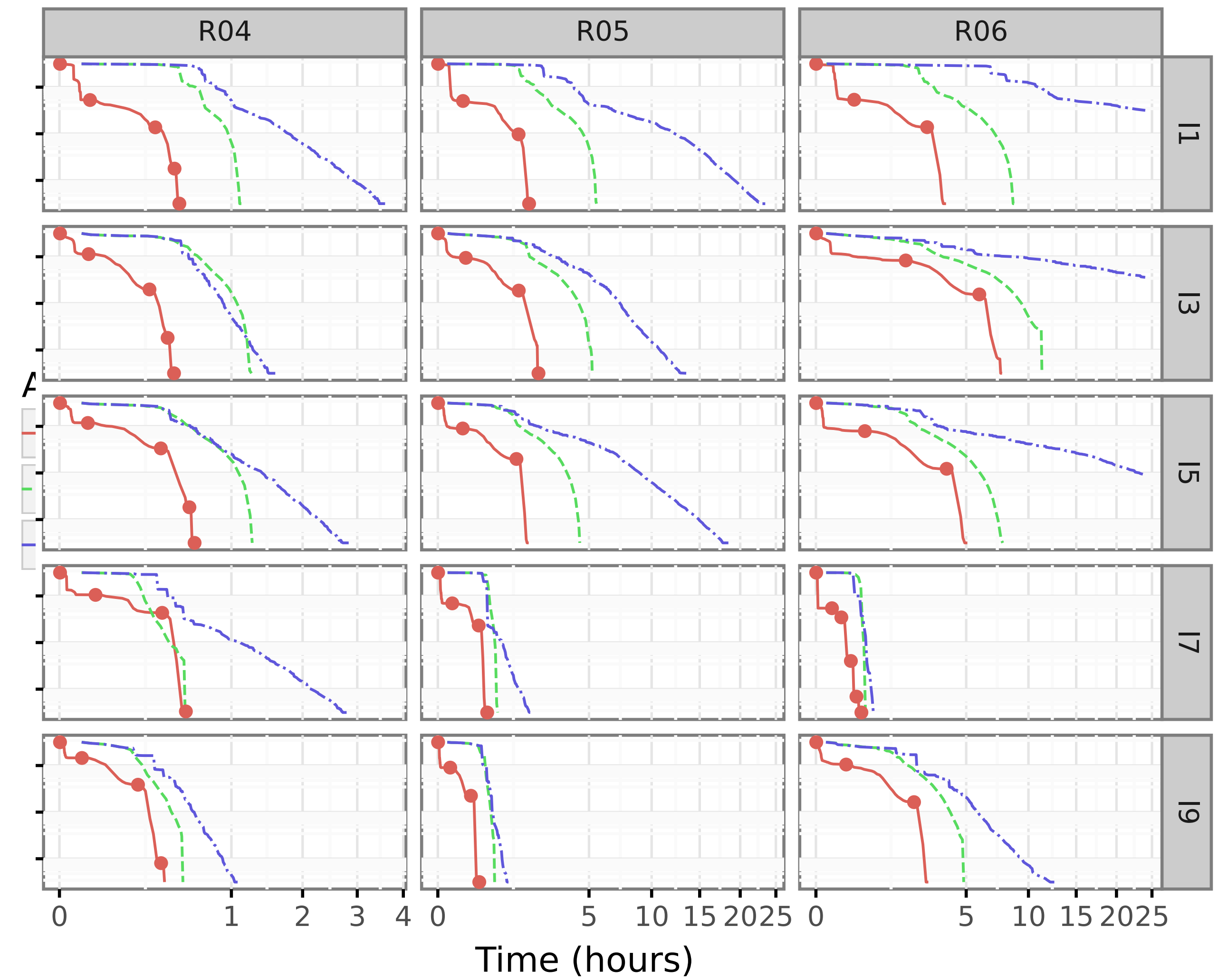


# Problem #2: Stochastic multicommodity flow

## Gap over time




100 scenarios




10.000 scenarios

# Problem #2: Stochastic multicommodity flow

| | | Adaptive | | | | | | Multi | | | | | | Single | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Inst | $\|\mathcal{S}\|$ | #Ref | #In | Time | Iter | FC | OC | #In | Time | | Iter | FC | OC | #In | Time | | Iter | FC | OC |
| R04 | 16 | 2.6 | 25 | 17.1 | 743.6 | 695 | 617 | 25 | 12.8 | (0.74) | 83.9 | 710 | 484 | 25 | 35.5 | (2.00) | 462.7 | 933 | 257 |
| | 100 | 3.2 | 25 | 28.4 | 715.3 | 872 | 1648 | 25 | 33.2 | (1.15) | 50.0 | 2283 | 1712 | 25 | 114.4 | (3.85) | 428.8 | 2608 | 235 |
| | 1000 | 3.3 | 25 | 153.3 | 752.2 | 2245 | 9997 | 25 | 264.7 | (1.70) | 38.8 | 16689 | 14410 | 25 | 911.1 | (5.66) | 385.8 | 16752 | 225 |
| | 5000 | 3.4 | 25 | 748.6 | 771.4 | 6384 | 41529 | 25 | 1540.9 | (2.00) | 35.6 | 74709 | 68595 | 25 | 4959.8 | (6.20) | 395.6 | 75877 | 222 |
| | 10000 | 3.5 | 25 | 1797.2 | 737.3 | 10309 | 79371 | 25 | 3849.2 | (2.04) | 34.2 | 140723 | 136964 | 25 | 8955.3 | (4.62) | 354.2 | 138664 | 205 |
| | 20000 | 3.6 | 25 | 4538.1 | 763.2 | 17703 | 149836 | 25 | 11490.8 | (2.39) | 35.2 | 295254 | 265130 | 25 | 19041.4 | (3.98) | 353.4 | 290202 | 206 |
| | 50000 | 3.8 | 25 | 18593.0 | 755.6 | 35164 | 333360 | 22 | 45655 | (2.21) | 30.1 | 634899 | 611159 | 24 | 49600.9 | (2.50) | 335.4 | 652731 | 203 |
| R05 | 16 | 2.3 | 25 | 66.6 | 1850.8 | 1608 | 1067 | 25 | 37.1 | (0.59) | 145.7 | 1172 | 800 | 25 | 201.2 | (2.59) | 1318.3 | 2188 | 489 |
| | 100 | 3.0 | 25 | 102.4 | 1722.9 | 1754 | 2492 | 25 | 89.5 | (0.95) | 85.8 | 3292 | 2534 | 25 | 881.2 | (5.01) | 1476.4 | 5079 | 462 |
| | 1000 | 3.0 | 25 | 407.7 | 1636.2 | 3113 | 13017 | 25 | 623.0 | (1.43) | 54.9 | 22406 | 19391 | 25 | 6212.7 | (7.85) | 1168.3 | 25441 | 392 |
| | 5000 | 3.0 | 25 | 2319.0 | 1680.7 | 7859 | 57436 | 25 | 4374.9 | (1.63) | 54.0 | 102603 | 94304 | 24 | 25334.5 | (6.44) | 908.4 | 102228 | 318 |
| | 10000 | 3.1 | 25 | 5583.0 | 1531.2 | 11908 | 106931 | 25 | 11375.8 | (1.79) | 50.8 | 196130 | 188977 | 20 | 33894.0 | (4.50) | 563.4 | 192746 | 252 |
| | 20000 | 3.1 | 25 | 16563.1 | 1537.8 | 20488 | 208087 | 25 | 34261.1 | (1.80) | 49.4 | 364854 | 375155 | 11 | 17073.4 | (2.52) | 126.8 | 256809 | 65 |
| | 50000 | 2.3 | 10 | 12351.0 | 376.2 | 27211 | 223703 | 10 | 14711 | (1.18) | 23.7 | 538463 | 284365 | 10 | 28116.5 | (2.19) | 73.4 | 525782 | 33 |
| R06 | 16 | 2.3 | 25 | 288.3 | 4327.5 | 3809 | 1752 | 25 | 124.1 | (0.48) | 234.1 | 2258 | 1261 | 25 | 1105.4 | (2.70) | 2730.4 | 4825 | 1083 |
| | 100 | 3.0 | 25 | 357.3 | 3882.0 | 3631 | 3842 | 25 | 240.9 | (0.78) | 115.2 | 4647 | 4210 | 25 | 3878.1 | (6.38) | 2441.2 | 7136 | 702 |
| | 1000 | 3.1 | 25 | 1385.4 | 3912.1 | 5344 | 22135 | 25 | 1610.1 | (1.25) | 73.2 | 29781 | 30440 | 24 | 23759.1 | (9.34) | 1613.3 | 30935 | 571 |
| | 5000 | 3.2 | 25 | 7922.9 | 3745.3 | 10027 | 95988 | 25 | 11396.0 | (1.57) | 71.1 | 127143 | 145203 | 10 | 13607.8 | (3.16) | 286.3 | 85420 | 89 |
| | 10000 | 3.3 | 25 | 19306.2 | 3956.2 | 15448 | 175679 | 25 | 27714.5 | (1.53) | 63.8 | 212647 | 279003 | 10 | 24954.2 | (2.76) | 266.1 | 151486 | 95 |
| | 20000 | 2.9 | 19 | 40475.0 | 2980.1 | 21944 | 292090 | 10 | 32340.8 | (1.54) | 41.6 | 322966 | 290521 | 8 | 36365.0 | (2.69) | 184.6 | 293469 | 69 |
| | 50000 | 1.5 | 5 | 6357.0 | 131.6 | 5566 | 37159 | 5 | 15964 | (2.52) | 23.4 | 635192 | 163153 | 5 | 22115.6 | (3.53) | 42.4 | 572869 | 14 |

Adaptive Benders

- More iterations (but "lighter")
- Fewer feasibility cuts
- Fewer optimality cuts than multi-cut Benders
- It requires to solve all subproblems, but only #Ref times (between 2 to 4)
- Note: Final partition size is almost all scenarios (98% median)