Multistage Robust Mixed Integer Optimization with Adaptive Partitions

Iain Dunning
Dimitris Bertsimas

Operations Research Center
Massachusetts Institute of Technology
Cambridge, MA

ISMP, July 2015

Motivation

"...the original problem that started my research is still outstanding - namely the problem of planning or scheduling dynamically over time, particularly planning dynamically under uncertainty. If such a problem could be successfully solved it could eventually through better planning contribute to the well-being and stability of the world."

- George Dantzig, in "History of Mathematical Programming", 1991
- Applications: inventory control, supply chain flexibility, project management, unit commitment, facility location/expansion, air traffic control, portfolio construction, financial instruments...

Introduction

- Decisions: **continuous** (e.g. how much stock?), **discrete** (e.g. operate power plant?)
- Structure through constraints, objective: linear, quadratic...
- Difficulties arises from modeling **uncertainty**
  - How to represent it?
    - Good short-term estimates, but long-term?
  - Adaptability of decisions
    - Some **here-and-now**, **wait-and-see** for later decisions
  - Must be **tractable** = "good" solutions for effort invested

Fully-adaptive multistage robust optimization problem

Our model: **adaptive robust optimization**

$$z_{full} = \min_{\mathbf{x}\in\mathcal{X}} \max_{\boldsymbol{\xi}\in\Xi} \sum_{t=1}^{T} \mathbf{c}^t(\boldsymbol{\xi}) \cdot \mathbf{x}^t\left(\boldsymbol{\xi}^1,..,\boldsymbol{\xi}^{t-1}\right)$$

$$\text{subject to} \sum_{t=1}^{T} \mathbf{A}^t(\boldsymbol{\xi}) \cdot \mathbf{x}^t\left(\boldsymbol{\xi}^1,..,\boldsymbol{\xi}^{t-1}\right) \leq \mathbf{b}(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} = \left(\boldsymbol{\xi}^1,..,\boldsymbol{\xi}^T\right) \in \Xi$$

- T time stages, $t = 1$ is here-and-now
- Uncertain parameters $\boldsymbol{\xi}^t$ for each time stage t
  - Uncertainty set Ξ, captures structure across time
  - Today: $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$ affine in $\boldsymbol{\xi}$. Paper: more general
- Adaptive decisions $\mathbf{x}^t$ for each time $t \geq 2$
- Deterministic & integrality constraints in $\mathcal{X}$: AMIO

A Hierarchy of Adaptability

One extreme: **static policy**

- Future decisions cannot adapt - all here & now
- Conservative, very tractable

Other extreme: **fully adaptive policy**

- Intractable in complexity sense, in practice?
- Column-and-row generation approach e.g. (Zeng, Zhao 2013)
- Unit commitment problem (Bertsimas et al. 2013)

A Hierarchy of Adaptability

Linear decision rules, a.k.a. affine adaptability

- Applied to RO in (Ben-tal et. al. 2004)
- **Good**: problem class, simple, can be optimal (Bertsimas, Iancu, Parillo 2010) (Bertsimas & Goyal 2012) (Bertsimas & Bidkhori 2014)
- **Bad**: no discrete recourse, changes problem structure
- Extensions: deflected linear decision rules (Chen et al. 2008), polynomial adaptability (Bertsimas et al. 2010)

Piecewise linear decision rules

- (Bertsimas & Georghiou 2014, 2015): piecewise linear for continuous, piecewise constant for discrete
- (2015): cutting-plane based
- (2014): reformulation, good results for multistage.

A Hierarchy of Adaptability

Finite adaptability

$$\mathbf{x}^2(\boldsymbol{\xi}) = \begin{cases} \mathbf{x}_1^2, & \forall \boldsymbol{\xi} \in \Xi_1, \\ \vdots & \\ \mathbf{x}_K^2, & \forall \boldsymbol{\xi} \in \Xi_K \end{cases}$$

How to pick the partitions?

- *A priori*, e.g. (Vayanos et al. 2011)
- Fix K & optimize, e.g. (Bertsimas & Caramanis 2010), (Hanasusanto et al. 2014)
- Optimizing directly results in difficult MIO

Motivation for our approach: two-stage, static policy

- Continuous $\mathcal{X}$, polyhedral Ξ, objective certain, feasible, bounded

$$
\begin{aligned}
z_{static} = \min_{\mathbf{x}\in\mathcal{X},z} \quad & z \\
\text{subject to} \quad & \mathbf{c}^1 \cdot \mathbf{x}^1 + \mathbf{c}^2 \cdot \mathbf{x}^2 \leq z \\
& \mathbf{a}_i^1(\boldsymbol{\xi}) \cdot \mathbf{x}^1 + \mathbf{a}_i^2(\boldsymbol{\xi}) \cdot \mathbf{x}^2 \leq b_i(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \Xi,\ i \in \mathcal{I}.
\end{aligned}
$$

- Solve with cutting plane method
- Let $\mathcal{A}_i$ = set of *active uncertain parameters* $\hat{\boldsymbol{\xi}}$ = zero slack cuts
- Now consider creating two partitions Ξ_1, Ξ_2

Two-stage, two partitions

$$\begin{aligned} z_{part} = \min_{\mathbf{x}\in\mathcal{X},z} \quad & z \\ \text{subject to} \quad & \mathbf{c}^1 \cdot \mathbf{x}^1 + \mathbf{c}^2 \cdot \mathbf{x}_j^2 \le z \qquad \forall j \in \{1,2\} \\ & \mathbf{a}_i^1(\boldsymbol{\xi}) \cdot \mathbf{x}^1 + \mathbf{a}_i^2(\boldsymbol{\xi}) \cdot \mathbf{x}_j^2 \le b_i(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \Xi_j, j \in \{1,2\}, i \in \mathcal{I}, \end{aligned}$$

We hope $z_{part} < z_{static}$. Will it be? Let $\mathcal{A} = \bigcup_i \mathcal{A}_i$

Theorem

If $\mathcal{A}$ satisfies either $\mathcal{A} \subset \Xi_1$ or $\mathcal{A} \subset \Xi_2$ then $z_{part} = z_{static}$. Otherwise $z_{part} \le z_{static}$.

Two-stage, two partitions

Proof.

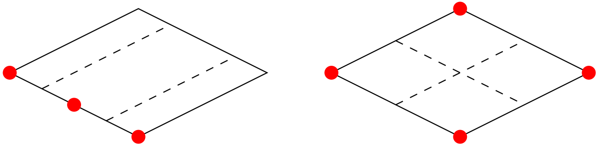
If $\mathcal{A} \subset \Xi_1$ (or $\mathcal{A} \subset \Xi_2$), then exact same constraints for one partition are valid, so no improvement for partition, and no improvement overall. □

- Need partitioning scheme that satisfies this
- One option: one $\hat{\boldsymbol{\xi}} \in \mathcal{A}$ per partition
- Use a *Voronoi diagram*: define partition $\Xi\left(\hat{\boldsymbol{\xi}}_i\right)$

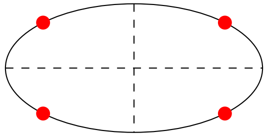
$$= \Xi \cap \left\{ \boldsymbol{\xi} \,\middle|\, \left\| \hat{\boldsymbol{\xi}}_i - \boldsymbol{\xi} \right\|_2 \leq \left\| \hat{\boldsymbol{\xi}}_j - \boldsymbol{\xi} \right\|_2 \quad \forall \hat{\boldsymbol{\xi}}_j \in \mathcal{A},\ \hat{\boldsymbol{\xi}}_i \neq \hat{\boldsymbol{\xi}}_j \right\}$$

$$= \Xi \cap \left\{ \boldsymbol{\xi} \,\middle|\, \left(\hat{\boldsymbol{\xi}}_j - \hat{\boldsymbol{\xi}}_i \right) \cdot \boldsymbol{\xi} \leq \frac{1}{2} \left(\hat{\boldsymbol{\xi}}_j - \hat{\boldsymbol{\xi}}_i \right) \cdot \left(\hat{\boldsymbol{\xi}}_j + \hat{\boldsymbol{\xi}}_i \right) \quad \forall \hat{\boldsymbol{\xi}}_j \in \mathcal{A}, \hat{\boldsymbol{\xi}}_i \neq \hat{\boldsymbol{\xi}}_j \right\}$$

- Ξ polyhedral $\rightarrow \Xi\left(\hat{\boldsymbol{\xi}}_i\right)$ polyhedral



$$\Xi_P = \left\{ \boldsymbol{\xi} \,\middle|\, \left\| \left[\frac{1}{2}\xi_1, \xi_2 \right] \right\|_1 \leq 1 \right\}$$



$$\Xi_E = \left\{ \boldsymbol{\xi} \,\middle|\, \left\| \left[\frac{1}{2}\xi_1, \xi_2 \right] \right\|_2 \leq 1 \right\}$$

Active uncertain parameters

- Generalize: discrete, general convex Ξ, reformulation
- Problem: might be *no* zero-slack constraints
- Given $\left(\bar{\mathbf{x}}^1, \bar{\mathbf{x}}^2\right)$, let

$$\mathcal{A}_i = \arg\min_{\boldsymbol{\xi} \in \Xi} \left\{ b_i\left(\boldsymbol{\xi}\right) - \mathbf{a}_i^1\left(\boldsymbol{\xi}\right) \cdot \bar{\mathbf{x}}^1 - \mathbf{a}_i^2\left(\boldsymbol{\xi}\right) \cdot \bar{\mathbf{x}}^2 \right\}$$

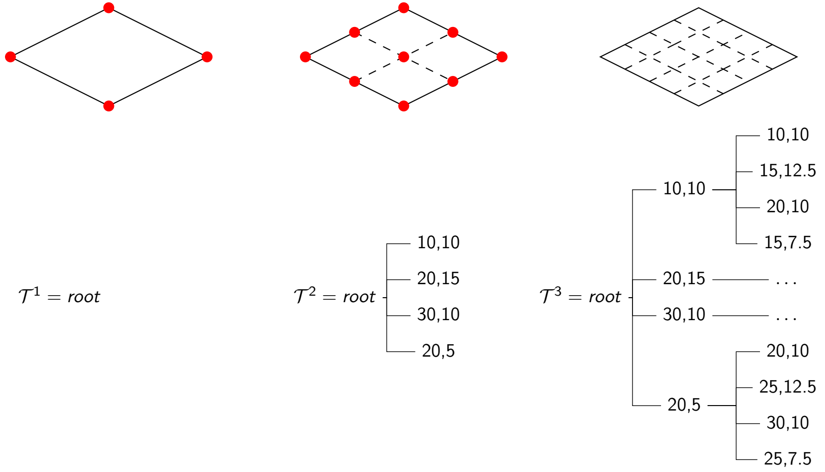
- Linear function, convex Ξ, $\mathcal{A}_i$ convex set?
- Generalized Voronoi $\rightarrow$ nonconvex partitions (Lee & Drysdale 1981)
- Select arbitrarily from $\mathcal{A}_i$, e.g. center, random, problem-specific

Nested partitioning

- How to partition *again*?
- Create tree $\mathcal{T}$ of $\hat{\boldsymbol{\xi}}$, $\mathcal{A}$ for a partition added as children of parent $\hat{\boldsymbol{\xi}}_i$

$$\Xi\left(\hat{\boldsymbol{\xi}}_i\right) = \left\{\boldsymbol{\xi} \,\middle|\, \left\|\hat{\boldsymbol{\xi}}_i - \boldsymbol{\xi}\right\|_2 \leq \left\|\hat{\boldsymbol{\xi}}_j - \boldsymbol{\xi}\right\|_2 \quad \forall \hat{\boldsymbol{\xi}}_j \in \mathit{Siblings}\left(\hat{\boldsymbol{\xi}}_i\right)\right\}$$
$$\cap \left\{\boldsymbol{\xi} \,\middle|\, \left\|\mathit{Parent}\left(\hat{\boldsymbol{\xi}}_i\right) - \boldsymbol{\xi}\right\|_2 \leq \left\|\hat{\boldsymbol{\xi}}_j - \boldsymbol{\xi}\right\|_2 \quad \forall \hat{\boldsymbol{\xi}}_j \in \mathit{Siblings}\left(\mathit{Parent}\left(\hat{\boldsymbol{\xi}}_i\right)\right)\right\}$$
$$\cdots \cap \Xi$$

Nested partitioning illustrated



Two-stage partition-and-bound algorithm

1. **Initialize**. Let $\mathcal{T}^1 \leftarrow$ initial tree, iteration $k \leftarrow 1$
2. **Solve**. Solve the partitioned problem, $\forall \hat{\boldsymbol{\xi}}_j \in Leaves\left(\mathcal{T}^k\right)$:

$$z_{alg}\left(\mathcal{T}^k\right) = \min_{\mathbf{x}\in\mathcal{X},z} \quad z$$

$$\text{subject to} \quad \mathbf{c}^1(\boldsymbol{\xi})\cdot\mathbf{x}^1 + \mathbf{c}^2(\boldsymbol{\xi})\cdot\mathbf{x}^2_j \le z \qquad \forall \boldsymbol{\xi} \in \Xi\left(\hat{\boldsymbol{\xi}}_j\right), \forall \hat{\boldsymbol{\xi}}_j$$

$$\mathbf{a}^1_i(\boldsymbol{\xi})\cdot\mathbf{x}^1 + \mathbf{a}^2_i(\boldsymbol{\xi})\cdot\mathbf{x}^2_j \le b_i(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \Xi\left(\hat{\boldsymbol{\xi}}_j\right), \forall \hat{\boldsymbol{\xi}}_j, i \in \mathcal{I},$$

3. **Grow**. $\mathcal{T}^{k+1} \leftarrow \mathcal{T}^k$. $\forall \hat{\boldsymbol{\xi}}_j \in Leaves\left(\mathcal{T}^{k+1}\right)$, add children for each $\hat{\boldsymbol{\xi}}$ in $\mathcal{A}$ for solution & constraints for partition $\Xi\left(\hat{\boldsymbol{\xi}}_j\right)$
4. **Bound**. Calculate $z_{lower}\left(\mathcal{T}^{k+1}\right)$ for fully adaptive, terminate if bound gap $\frac{\left(z_{alg}-z_{lower}\right)}{\left|z_{lower}\right|} \le \epsilon_{gap}$. Otherwise $k \leftarrow k+1$, go to Step 2.

Two-stage lower bound

- Sample-based bound of (Hadjiyiannis et al. 2011)
- Proposition: solve

$$\begin{aligned} z_{lower}\left(\mathcal{T}\right) = \min_{\mathbf{x}\in\mathcal{X},z} \quad & z \\ \text{subject to} \quad & \mathbf{c}^1\left(\hat{\boldsymbol{\xi}}_j\right)\cdot\mathbf{x}^1 + \mathbf{c}^2\left(\hat{\boldsymbol{\xi}}_j\right)\cdot\mathbf{x}^2_j \le z & \forall\hat{\boldsymbol{\xi}}_j \in \mathcal{T} \\ & \mathbf{a}^1_i\left(\hat{\boldsymbol{\xi}}_j\right)\cdot\mathbf{x}^1 + \mathbf{a}^2_i\left(\hat{\boldsymbol{\xi}}_j\right)\cdot\mathbf{x}^2_j \le b_i\left(\hat{\boldsymbol{\xi}}_j\right) & \forall\hat{\boldsymbol{\xi}}_j \in \mathcal{T}, i \in I \end{aligned}$$

  Then $z_{lower}\left(\mathcal{T}\right) \le z_{full}$.

- As tree grows, upper bound and lower bound both improving

Incorporating affine adaptability

- If $A^2(\boldsymbol{\xi}) = \bar{A}^2$, can substitute in affine policy

$$\mathbf{x}^2(\boldsymbol{\xi}) = \mathbf{F}\boldsymbol{\xi} + \mathbf{g}$$

- But observe: $\mathbf{F}, \mathbf{g}$ can also be wait-and-see
- Associate different affine policy with each partition, e.g.

$$\mathbf{x}^2(\boldsymbol{\xi}) = \begin{cases} \mathbf{F}_1\boldsymbol{\xi} + \mathbf{g}_1, & \boldsymbol{\xi} \in \Xi\left(\hat{\boldsymbol{\xi}}_1\right), \\ \mathbf{F}_2\boldsymbol{\xi} + \mathbf{g}_2, & \boldsymbol{\xi} \in \Xi\left(\hat{\boldsymbol{\xi}}_2\right), \end{cases}$$

- Piecewise affine continuous, piecewise constant discrete

Convergence Properties

Proposition

The upper bound $z_{alg}\left(\mathcal{T}^k\right)$ will never increase as k increases.

Proof.

Follows from the "nested" nature of partitions. □

Proposition

The upper bound may not improve for any finite k

Proof.

Consider problem that "requires" partitions $[0, \frac{1}{3}]$ and $[\frac{1}{3}, 1]$, but our method will only produce partitions at 2^{-p} intervals □

Multistage problems

- Must respect *nonanticipativity*
- Applying current scheme blindly results in nonadaptive solutions
- Make partitioning scheme time-stage-aware: $\Xi\left(\hat{\xi}_i\right) =$

$$\left\{\xi \,\middle|\, \left\|\hat{\xi}_i^{t_{i,j}} - \xi^{t_{i,j}}\right\|_2 \le \left\|\hat{\xi}_j^{t_{i,j}} - \xi^{t_{i,j}}\right\|_2 \quad \forall \hat{\xi}_j \in Siblings\left(\hat{\xi}_i\right)\right\}$$
$$\cap \left\{\xi \,\middle|\, \left\|Parent\left(\hat{\xi}_i\right)^{t'_{i,j}} - \xi^{t'_{i,j}}\right\|_2 \le \left\|\hat{\xi}_j^{t'_{i,j}} - \xi^{t'_{i,j}}\right\|_2 \right.$$
$$\left. \forall \hat{\xi}_j \in Siblings\left(Parent\left(\hat{\xi}_i\right)\right)\right\} \cdots \cap \Xi.$$

where $t_{i,j}$ for $\hat{\xi}_i$ and $\hat{\xi}_j$ is $\arg\min_t \left\{\hat{\xi}_i^t \ne \hat{\xi}_j^t\right\}$,

Multistage problems - nonanticipativity

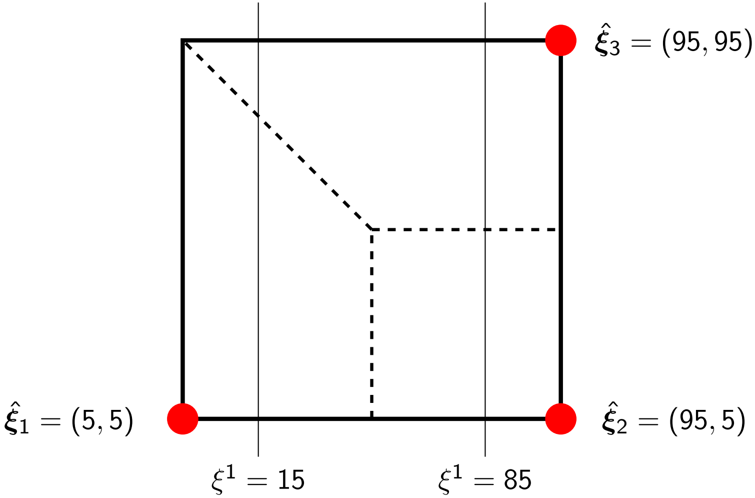
Proposition

If there exists $\psi = (\psi^1, \ldots, \psi^T) \in \Xi\left(\hat{\xi}_i\right)$ *and* $\phi = (\phi^1, \ldots, \phi^T) \in \Xi\left(\hat{\xi}_j\right)$ *such that* $\psi^s = \phi^s \quad \forall s \in \{1, \ldots, t-1\}$, *and at least one of* $\psi \in int\left(\Xi\left(\hat{\xi}_i\right)\right)$ *and* $\phi \in int\left(\Xi\left(\hat{\xi}_j\right)\right)$ *holds, then we must enforce nonanticipativity constraints for the corresponding decisions at time stage* t

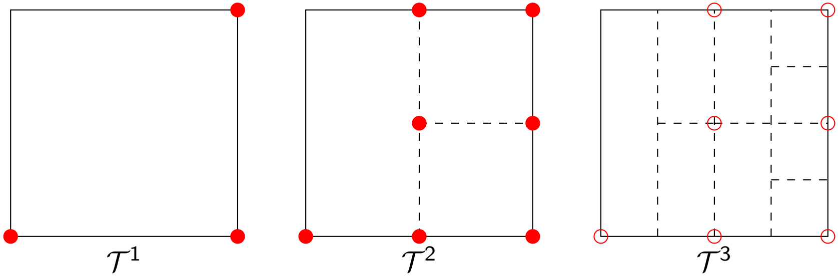
Proof.

All holds: $\exists$ partial realization of ξ which could lie in either. If the $=$ holds, but on boundary, then both decisions ok. If $\neq$ then partitions are distinguishable. □

Multistage problems - example



Multistage problems - example



Computational experiment - capital budgeting

From (Hanasusanto et al. 2014)

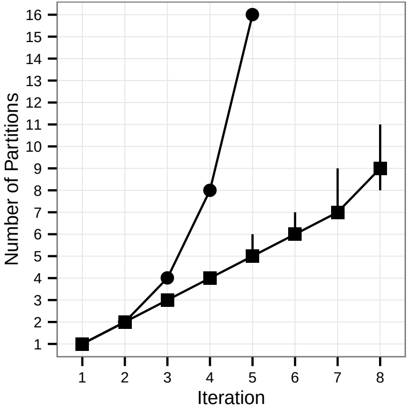
$$\begin{aligned}
\max_{z,\mathbf{x}} \quad & z \\
\text{subject to} \quad & \mathbf{r}(\boldsymbol{\xi}) \cdot \left(\mathbf{x}^1 + \theta \mathbf{x}^2(\boldsymbol{\xi})\right) \geq z && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{c}(\boldsymbol{\xi}) \cdot \left(\mathbf{x}^1 + \mathbf{x}^2(\boldsymbol{\xi})\right) \leq B && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{x}^1 + \mathbf{x}^2(\boldsymbol{\xi}) \leq \mathbf{e} && \forall \boldsymbol{\xi} \in \Xi \\
& \mathbf{x}^1, \mathbf{x}^2(\boldsymbol{\xi}) \in \{0,1\}^N && \forall \boldsymbol{\xi} \in \Xi,
\end{aligned}$$

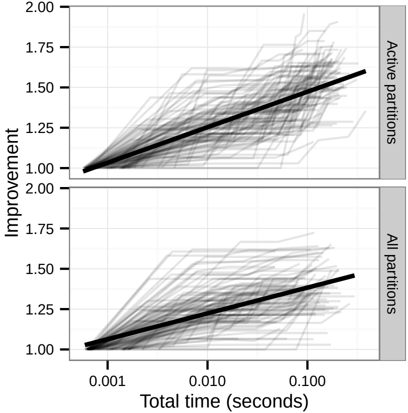
$$\Xi = \left\{ \boldsymbol{\xi} \,\middle|\, \boldsymbol{\xi} \in [-1,1]^4 \right\}$$

Measure bound gap versus time, and *improvement* versus time:

$$\frac{z_{alg}\left(\mathcal{T}^k\right) - z_{alg}\left(\mathcal{T}^1\right)}{z_{alg}\left(\mathcal{T}^1\right)}$$

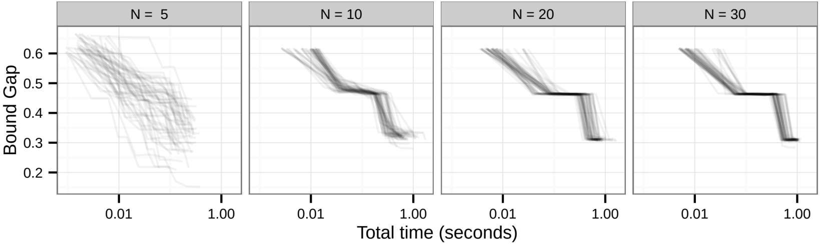
Computational experiment - capital budgeting

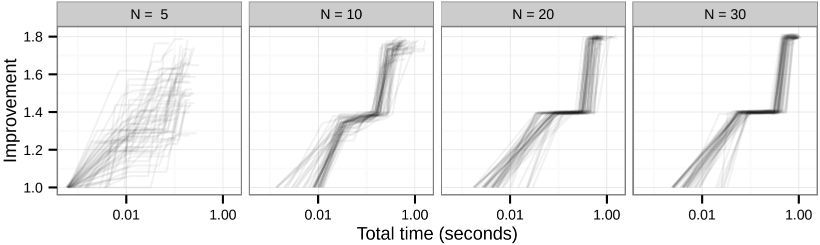




- Number of partitions is at most $(m+1)^{k-1}$ (!)
- Reduce by only subpartitioning the *active partitions*, i.e. partitions such that $\tilde{z}_j = z$

Computational experiment - capital budgeting





Computational experiment - lot sizing

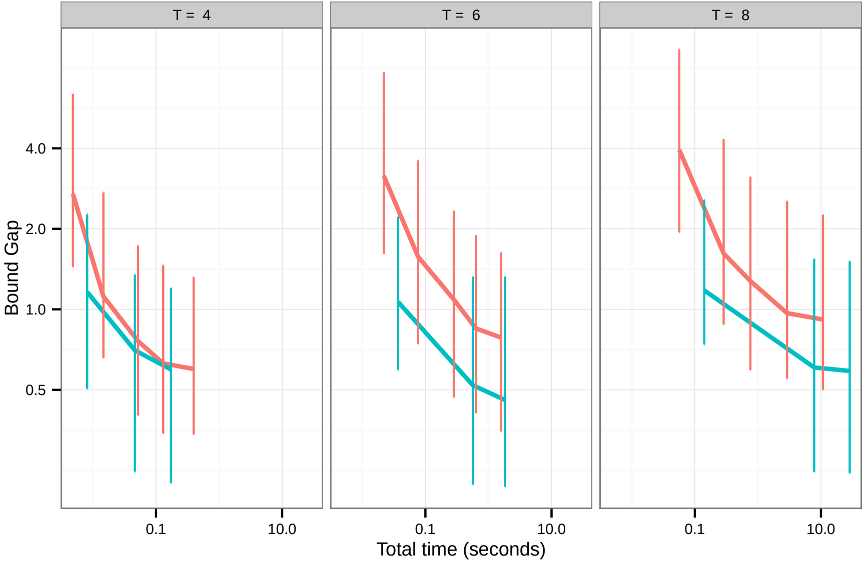
- From (Bertsimas & Georghiou 2015)
  - T time stages, must satisfy demand at all t
  - x^t continuous ordering decision before demand
  - y_n^t discrete ordering decision after demand
  - Holding costs, box uncertainty

    $$\Xi = \left\{ \boldsymbol{\xi} \,\middle|\, \xi^1 = 1,\ l^t \le \xi^t \le u^t \quad \forall t \in \{2, \ldots, T\} \right\}$$

- Use affine for continuous decisions

Computational experiment - lot sizing

Method		T 4	6	8	10
Our method (2 iter.)	Gap (%)	13.0	10.3	11.6	14.9
	Time (s)	0.0	0.5	7.7	108.6
Our method (3 iter.)	Gap (%)	11.4	9.3	11.3	14.9
	Time (s)	0.2	2.0	52.4	309.3
Postek & den Hertog (2014)	Gap (%)	11.5	14.1	15.7	15.7
	Time (s)	0.4	1.6	10.8	77.8
Bertsimas & Georghiou (2015)	Gap (%)	17.2	34.5	37.6	-
	Time (s)	3381	9181	28743	-



T = 4
T = 6
T = 8
Bound Gap
4.0
2.0
1.0
0.5
0.1
10.0
Total time (seconds)

JuMPeR - https://github.com/IainNZ/JuMPeR.jl

```
rm = RobustModel()
@defVar(rm, obj <= 1000)
@defVar(rm, x[1:N], Bin)
@defVar(rm, y[1:num_leaf,1:N], Bin)
@defUnc(rm, -1 <= ξ[1:num_leaf,1:4] <= 1)
@setObjective(rm, Max, obj)

for j in 1:num_leaf
    cost   = [(1 + dot(Φ[i,:], ξ[j,:])/2) * c0[i]) for i=1:N]
    profit = [(1 + dot(Ψ[i,:], ξ[j,:])/2) * r0[i]) for i=1:N]

    @addConstraint(rm, obj[j] <=
        sum{ profit[i] * (x[i] + θ*y[j,i]), i=1:N} )

    @addConstraint(rm,
        sum{ cost[i] * (x[i] + y[j,i]), i=1:N} <= B)

    @addConstraint(rm, only_once[i=1:N],
        x[i] + y[j,i] <= 1)
```

Conclusions & future work

- Proposed method inspired by observations on structure
- Generalized to multistage, affine, & characterized performance
- Good solutions quickly
- Partition-and-bound method simple to implement
- c.f. with B-&-B for IP: cuts, branching rules, heuristics etc?
- Better partitioning, better use of Ξ structure
- Bertsimas, D. and Dunning, I. Multistage Robust Mixed Integer Optimization with Adaptive Partitions. Preprint available at `http://www.optimization-online.org/DB_HTML/2014/11/4658.html`

Extra: no convergence example

$$
\begin{aligned}
z(\epsilon) = \min_{x^2 \in \{0,1\}, y^2 \in \{0,1\}, z} \quad & z \\
\text{subject to} \quad & x^2(\xi) + y^2(\xi) \leq z && \forall \xi \in [0,1] \\
& x^2(\xi) \geq \frac{\epsilon - \xi}{\epsilon} && \forall \xi \in [0,1] \\
& y^2(\xi) \geq \frac{\epsilon + \xi - 1}{\epsilon} && \forall \xi \in [0,1],
\end{aligned}
$$

where $\epsilon \in [0,1]$. This problem has a fully adaptive solution of $z = 1$ and

$$
x^2(\xi) = \begin{cases} 1, & 0 \leq \xi \leq \epsilon, \\ 0, & \epsilon < \xi \leq 1, \end{cases} \qquad y^2(\xi) = \begin{cases} 0, & 0 \leq \xi \leq \epsilon, \\ 1, & \epsilon < \xi \leq 1. \end{cases}
$$