Benders 102 - Acceleration techniques

Kevin Dalmeijer, Mathieu Tanneau

October 7, 2021

Georgia Institute of Technology

Foreword

Ground rules

- First make it work, then make it fast

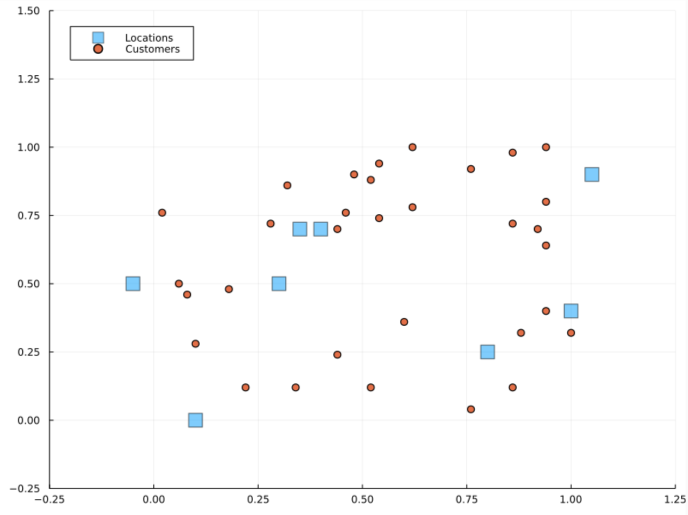
No acceleration technique unless you have a working Benders!

- Systematically benchmark your code on multiple, hard, instances
- Don't improve anything that's not the bottleneck

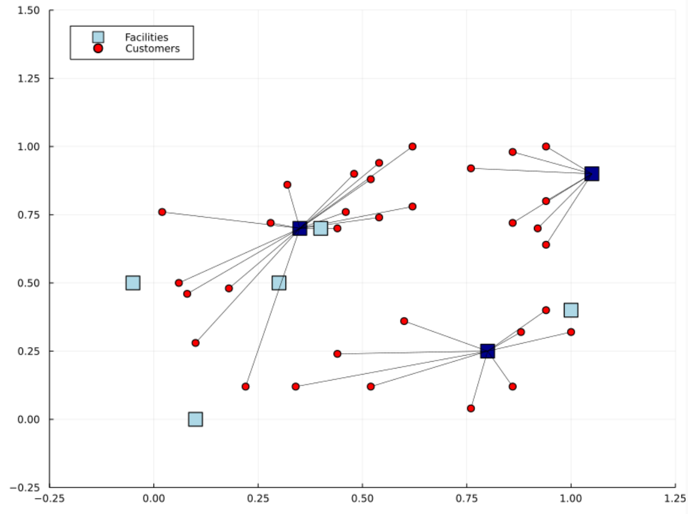
Outline

- Background on solving MIPs
- Pareto-optimal cuts
- Separating fractional points (Benders at the root)
- Stabilization techniques
- Cut aggregation

+ demos on a basic Benders implementation
We'll get a 100x speedup



Facility location problem (instance data)



Facility location problem (solution)

Background on MIPs

There are 2 sides to solving MIPs

	The **primal** side	The **dual** side
What?	*Find* an optimal solution	*Prove* it's optimal
How?	Heuristics, Branching (rare)	Cuts + B&B
When?	Root + nodes (periodically)	Root (a lot) + B&B
Rules?	None besides feasibility	Valid relaxations
% time?	Usually low	Usually high (close last ‰)

What's hard: finding good solutions (primal)? Or closing the gap (dual)?

No free lunch $\implies$ it depends on the instance!

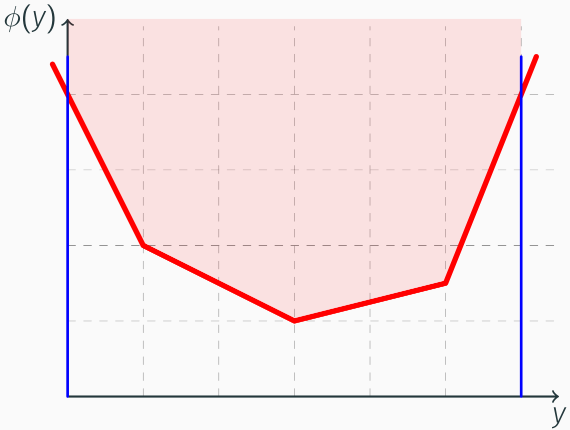
Textbook Benders: recap

$$\underbrace{\begin{array}{rl} \min_{\mathbf{y},\mathbf{x}} & c^T\mathbf{y} + d^T\mathbf{x} \\ s.t. & \mathbf{y} \in \mathcal{Y} \\ & A\mathbf{y} + B\mathbf{x} \geq b \end{array}}_{\text{original problem}} \longrightarrow \underbrace{\begin{array}{rl} \min_{\mathbf{y},\theta} & c^T\mathbf{y} + \theta \\ s.t. & \mathbf{y} \in \mathcal{Y} \\ & \text{[Benders cuts]} \end{array}}_{\text{master problem}} + \underbrace{\begin{array}{rl} \min_{\mathbf{x}} & d^T\mathbf{x} \\ s.t. & B\mathbf{x} \geq b - A\mathbf{y} \end{array}}_{\text{sub-problem}}$$

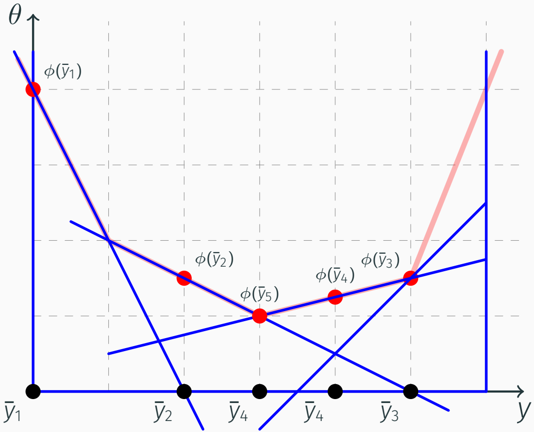
- Solve the master problem $\longrightarrow$ get an integer solution
- Solve the sub-problem $\longrightarrow$ get a cut
- Repeat

Pareto-optimal cuts

$$\min_{y} \quad \phi(y)$$
$$s.t. \quad 0 \leq y \leq 6$$
$$y \in \mathbb{Z}$$



Iter.	$\bar{z}$	$\underline{z}$
0	$+\infty$	$-\infty$
1	4.00	0.00
2	1.50	0.00
3	1.50	0.00
4	1.25	0.50
5	1.00	1.00



Back to iteration 3...

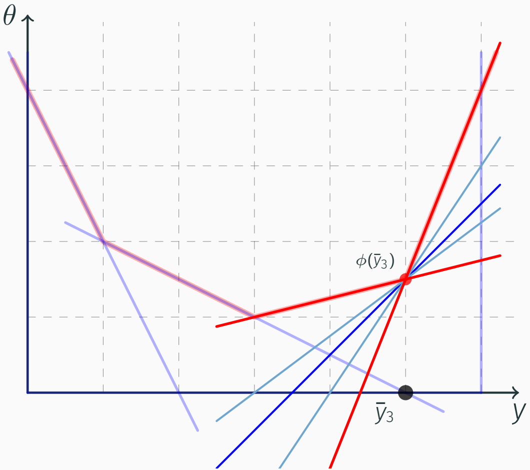
- Solution: $\bar{y}_3 = 5$ and $\phi(\bar{y}_3) = 1.5$
- Cut: $\theta \geq 1.5 + 1.00 \times (y - \bar{y}_3)$
- Other valid cuts

$$\theta \geq 1.5 + 0.25 \times (y - \bar{y}_3)$$

$$\theta \geq 1.5 + 0.75 \times (y - \bar{y}_3)$$

$$\theta \geq 1.5 + 1.75 \times (y - \bar{y}_3)$$

$$\theta \geq 1.5 + 2.50 \times (y - \bar{y}_3)$$



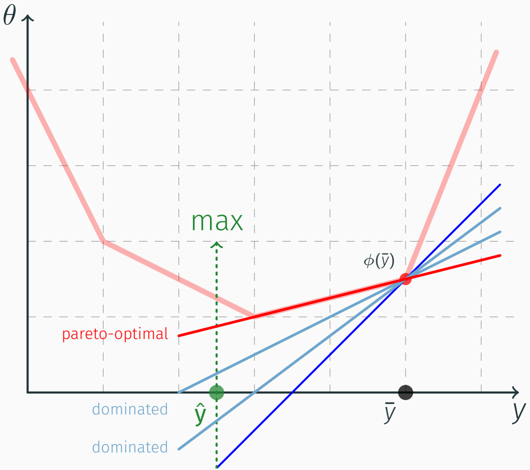
Pareto-optimal cuts [MW81]

Inputs

- Current master solution $\bar{y}$
- "Core point" $\hat{y} \in \text{relint}(\mathcal{Y})$

Recipe

- Solve sub-problem, get $\phi(\bar{y})$
- Fix SP objective to $\phi(\bar{y})$
- Maximize cut violation at $\hat{y}$
- Return the cut



Benders at the root

Textbook Benders: recap

$$
\underbrace{\begin{array}{rl} \min_{\mathbf{y},\mathbf{x}} & c^T\mathbf{y} + d^T\mathbf{x} \\ s.t. & \mathbf{y} \in \mathcal{Y} \\ & A\mathbf{y} + B\mathbf{x} \geq b \end{array}}_{\text{original problem}}
\longrightarrow
\underbrace{\begin{array}{rl} \min_{\mathbf{y},\theta} & c^T\mathbf{y} + \theta \\ s.t. & \mathbf{y} \in \mathcal{Y} \\ & \text{[Benders cuts]} \end{array}}_{\text{master problem}}
+
\underbrace{\begin{array}{rl} \min_{\mathbf{x}} & d^T\mathbf{x} \\ s.t. & B\mathbf{x} \geq b - A\mathbf{y} \\ & \end{array}}_{\text{sub-problem}}
$$

- Solve the master problem $\longrightarrow$ get an integer solution
- Solve the sub-problem $\longrightarrow$ get a cut
- Repeat

Textbook Benders: Limitations

$$\underbrace{\begin{array}{rl} \min_{\mathbf{y},\mathbf{x}} & c^T\mathbf{y} + d^T\mathbf{x} \\ s.t. & \mathbf{y} \in \mathcal{Y} \\ & A\mathbf{y} + B\mathbf{x} \geq b \end{array}}_{\text{original problem}} \longrightarrow \underbrace{\begin{array}{rl} \min_{\mathbf{y},\theta} & c^T\mathbf{y} + \theta \\ s.t. & \mathbf{y} \in \mathcal{Y} \\ & \text{[Benders cuts]} \end{array}}_{\text{master problem}} + \underbrace{\begin{array}{rl} \min_{\mathbf{x}} & d^T\mathbf{x} \\ s.t. & B\mathbf{x} \geq b - A\mathbf{y} \end{array}}_{\text{sub-problem}}$$

- Solving the *integer* master takes time
- Why solve an *approximate* model to optimality?
- The sub-problem remains an LP no matter the value of $\mathbf{y}$!

We can compute cuts at fractional points!

Benders at the root

Phase I (root)

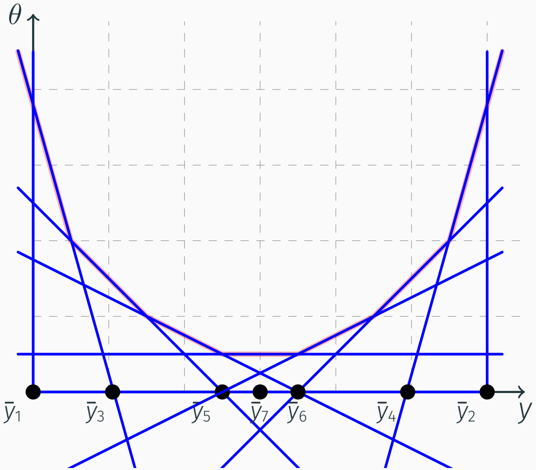
- Relax integrality constraints $\longrightarrow$ master is LP
- Solve root relaxation with Benders

Phase II (branch)

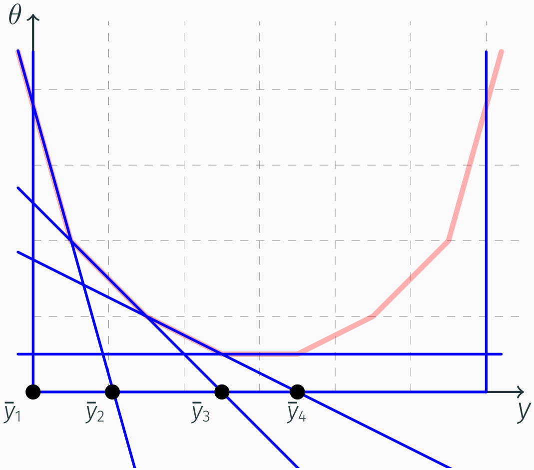
- Re-instate integrality constraints
- Only separate cuts at integer points

Stabilization techniques

Iter.	$\bar{Z}$	$\underline{Z}$
0	$+\infty$	$-\infty$
1	3.80	0.00
2	3.80	0.00
3	1.50	0.00
4	1.50	0.00
5	0.50	0.00
6	0.50	0.00
7	0.50	0.25
8	0.50	0.50



Iter.	$\bar{z}$	$\underline{z}$
0	$+\infty$	$-\infty$
1	3.80	0.00
2	1.50	0.00
3	0.50	0.00
4	0.50	0.00
5	0.50	0.50



Why?

- Avoid oscillating behaviors (exponentially bad in high dimension)
- Generate *better candidates* for separation
- Reach optimum faster

When?

Most useful at the root

How?

General idea: avoid large deviations from current solution

Many different techniques (will add references)

- Trust-region, Box-constraint
- Quadratic penalty (a.k.a *bundle methods*)
- Level method
- Interior-point stabilization [GGBM16]
- In-Out separation [FLS17, BST20]
- ...

Pick one you like and try it!

Dealing with many subproblems

Benders decomposition

$$
\begin{array}{llcll}
\min\limits_{\mathbf{y},\mathbf{x}} & c^T\mathbf{y} + \sum_i d_i^T \mathbf{x}_i & & \min\limits_{\mathbf{y},\theta} & c^T\mathbf{y} + \sum_i \theta_i \\
& \mathbf{y} \in \mathcal{Y} & & & \mathbf{y} \in \mathcal{Y} \\
& A_1\mathbf{y} + B_1\mathbf{x}_1 = b_1 & \longrightarrow & & \theta_1 \geq \phi_1(\mathbf{y}) \\
& \quad \vdots & & & \quad \vdots \\
& A_N\mathbf{y} + B_N\mathbf{x}_N = b_N & & & \theta_N \geq \phi_N(\mathbf{y})
\end{array}
$$

Each BD iteration adds (up to) N cuts $\Rightarrow$ master becomes large!

Multi-cut

$$\min_{\mathbf{y},\theta} \quad c^T\mathbf{y} + \sum_i \theta_i$$
$$\mathbf{y} \in \mathcal{Y}$$
$$\theta_1 \geq \phi_1(\mathbf{y})$$
$$\vdots$$
$$\theta_N \geq \phi_N(\mathbf{y})$$

Cut aggregation

$$\min_{\mathbf{y},\theta} \quad c^T\mathbf{y} + \sum_k \theta_k$$
$$\mathbf{y} \in \mathcal{Y}$$
$$\theta_k \geq \sum_{i \in \mathcal{N}_k} \phi_i(\mathbf{y})$$

Single-cut

$$\min_{\mathbf{y},\theta} \quad c^T\mathbf{y} + \theta$$
$$\mathbf{y} \in \mathcal{Y}$$
$$\theta_1 \geq \phi_1(\mathbf{y}) + \ldots + \phi_N(\mathbf{y})$$

where $\bigcup_k \mathcal{N}_k = \mathcal{N} = \{1, \ldots, N\}$ (sub-problems are *clustered*).

Pros and cons

	Multi-cut	Aggregated-cut	Single-cut
$\|\theta\|$	N	K	1
#cuts in MP	many	medium	few
MP size	largest	medium	smallest
#SP/cut	1	$\|\mathcal{N}_k\|$	N
cut strength	strongest	medium	weakest
Tail-off	lowest	medium	highest

How to cluster? Put similar sub-problems together.

Other strategies

When the number of sub-problems is large…

- Solve them in parallel
- *Partial cutting*: don’t solve all the sub-problems!
    - Keeps the size of MP under control
    - Typically adds much fewer cuts in total
    - Compatible with parallelization

References

- Pierre Bonami, Domenico Salvagnin, and Andrea Tramontani, *Implementing automatic benders decomposition in a modern mip solver.*, IPCO, 2020, pp. 78–90.
- Matteo Fischetti, Ivana Ljubić, and Markus Sinnl, *Redesigning benders decomposition for large-scale facility location*, Management Science **63** (2017), no. 7, 2146–2162.
- Jacek Gondzio, Pablo González-Brevis, and Pedro Munari, *Large-scale optimization with the primal-dual column generation method*, Mathematical Programming Computation **8** (2016), no. 1, 47–82.
- Thomas L Magnanti and Richard T Wong, *Accelerating benders decomposition: Algorithmic enhancement and model selection criteria*, Operations research **29** (1981), no. 3, 464–484.

That’s all folks!