Cutting Plane Selection with Analytic Centers and Multiregression

Mathieu Besançon @matbesancon - besancon@zib.de,
Mark Turner, Timo Berthold, Thorsten Koch

February 23, 2023

Zuse Institute Berlin - AIS²T
TU Berlin, Fair Isaac Deutschland GmbH

Outline

Table of Contents

MIP notation

Mixed-Integer Linear Program (MILP):

$$\underset{\mathbf{x}}{\operatorname{argmin}}\{\mathbf{c}^\intercal \mathbf{x} \;\mid\; \mathbf{A}\mathbf{x} \leq \mathbf{b}, \;\; \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}, \;\; \mathbf{x} \in \mathbb{Z}^{|\mathcal{J}|} \times \mathbb{R}^{n-|\mathcal{J}|}\}$$

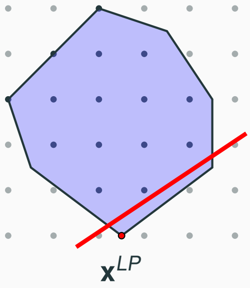
- $\mathbf{c} \in \mathbb{R}^n$ - Objective coefficient vector
- $\mathbf{A} \in \mathbb{R}^{m \times n}$ - Constraint matrix
- $\mathbf{b} \in \mathbb{R}^m$ - RHS constraint vector
- $\mathbf{l}, \mathbf{u} \in \{\mathbb{R}, -\infty, \infty\}^n$ - Lower and upper variable bound vectors
- $\mathcal{J} \subseteq \{1, \ldots, n\}$ - Set of indices of integer variables

Cut definition

A *cut* is a linear constraint that does not increase the optimal value when added.

Cut $(\alpha, \beta) \in \mathbb{R}^n \times \mathbb{R}$ noted:

$$\alpha^\top \mathbf{x} \leq \beta,$$



Cut selection

Purpose of cuts: tighten the convex relaxation (intentionally vague).

Cuts generated in *separation rounds*: separate and re-solve.
Most separators cheaper than relaxation solves:
$\rightarrow$ MIP solvers generate more cuts than used.

Key trade-off:
Adding all cuts $\Rightarrow$ Expensive LPs, numerically unstable
$\Leftrightarrow$
Adding no cut $\Rightarrow$ More nodes needed to solve

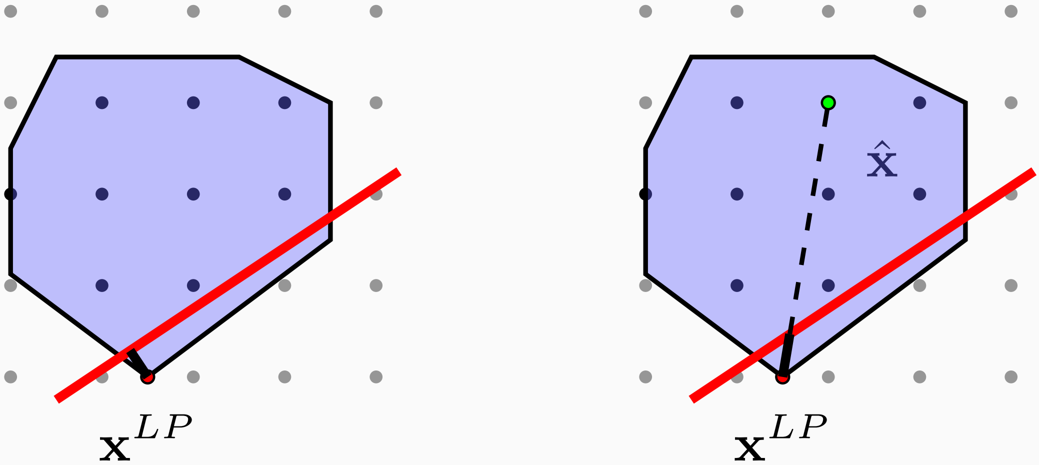
Cut scoring

Different aspects for selecting a subset of cuts - here those of SCIP:

- *Strength*: how much are we removing from the relaxation?
- *Parallelism* / orthogonality (to previous cuts),
- *Sparsity*,
- *Objective parallelism*.

This talk $\Rightarrow$ What is a good measure or proxy for strength?

Cut scoring: efficacy & directed cutoff distance



Other measures

Expected improvement:

$$\texttt{expimprov}(\alpha,\beta;\mathbf{c},\mathbf{x}^{LP}) := ||\mathbf{c}|| \cdot \frac{\alpha^\top \mathbf{c}}{||\alpha||||\mathbf{c}||} \cdot \texttt{eff}(\alpha,\beta;\mathbf{x}^{LP})$$

Actual improvement:
Re-solve the relaxation with the cut added a.k.a. *look-ahead*, analogue to strong branching.

Existing literature for cut selection

Older comprehensive computational experiments on cut selection:

- [Achterberg, 2007, Andreello et al., 2007, Wesselmann and Stuhl, 2012],
- Focus on cheap ranking measures, mostly **efficacy**,
- Filtering **parallel** cuts is most important.

Question recently revisited with machine learning angle:

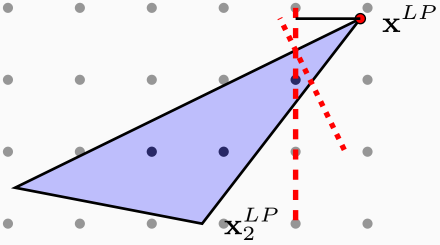
- [Baltean-Lugojan et al., 2019, Paulus et al., 2022]: learning objective change (strong cutting/look-ahead),
- [Tang et al., 2020]: selection of Gomory cuts with neural networks,
- [Turner et al., 2022]: learn to weigh different scores.

Summary: Improvement is possible, but non-trivial and difficult to generalize.

What is wrong with efficacy?

Efficacy $\approx$ how far does a cut remove current LP solution?

Infeasible projection on cut:
what is removed from relaxation?



Dual degeneracy:
Arbitrary LP solution on optimal face

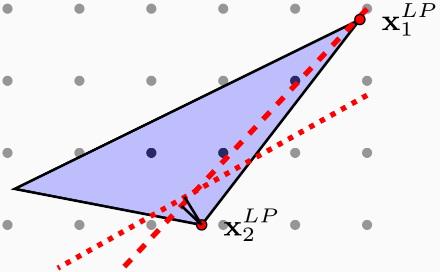
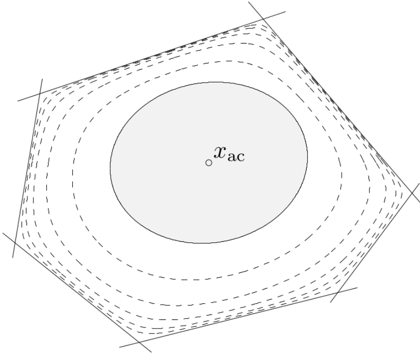


Table of Contents

Analytic Centers

Solution to log-barrier minimization:

$$\arg\min_{\mathbf{x}} - \sum_i \log(\mathbf{b}_i - \mathbf{A}_i\mathbf{x}) - \sum_j (\log(\mathbf{x}_j - \mathbf{l}_j) + \log(\mathbf{u}_j - \mathbf{x}_j))$$



- "Central-most" point of the polytope.
- Easily extendable to conic optimization.
- Robust to bad conditioning (unlike Chebyshev)
- Not independent of the formulation for a given set (e.g. redundant inequalities) $\rightarrow$ importance of presolving

Figure: Convex Optimization, Boyd & Vandenberghe

Analytic centers in MIPs

- Presolving: deducing tight bounds, fixing variables [Berthold et al., 2018]
- Branching: which variables are close to integer values [Berthold et al., 2018]
- Heuristics: exploiting the AC as central direction [Baena and Castro, 2011, Naoum-Sawaya, 2014]
- Cut generation: [Fischetti and Salvagnin, 2009]

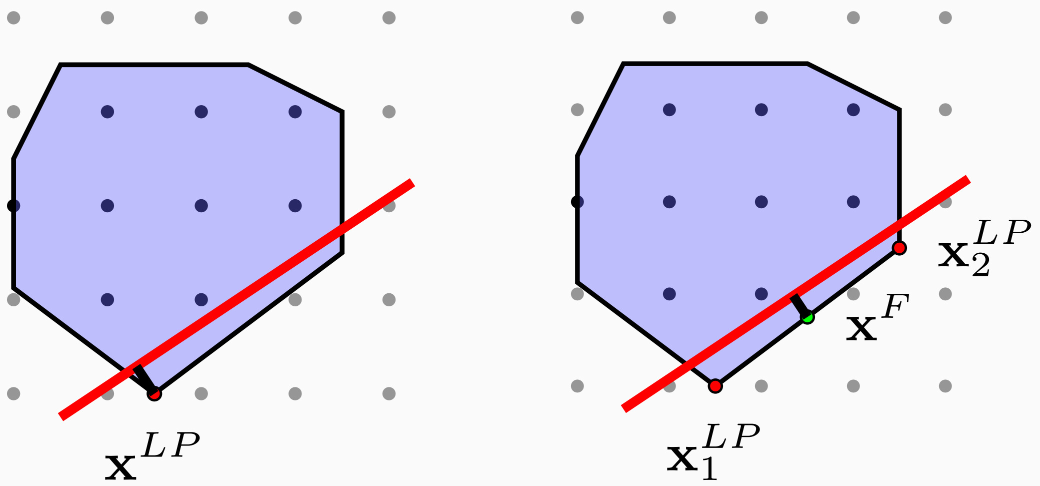
Analytic centers: the computational case

Analytic centers: (almost) a “co-product” of LP solves with Interior Points.

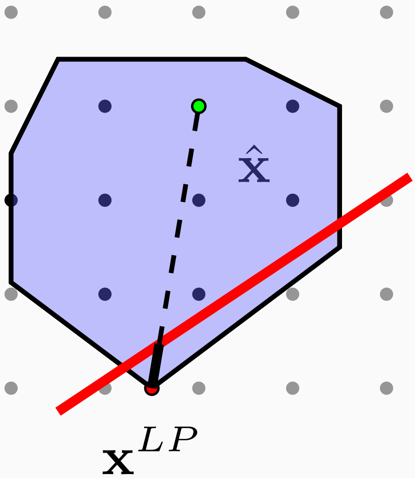
- Original LP with feasibility objective: AC of polyope
- Direct LP solve without cross-over: AC of optimal face.

Latter crucial to exploit dual degeneracy of current relaxation.

New efficacy measures



New directed cutoff measure



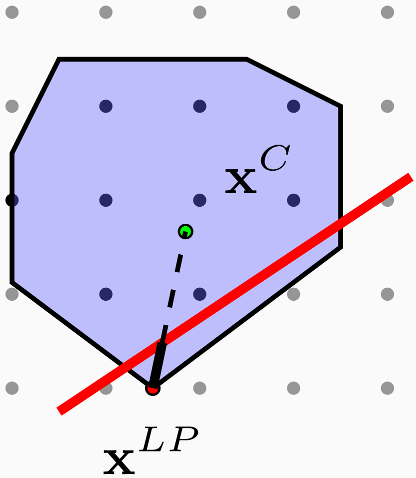


Table of Contents

Cut dominance consistency

Cut $(A) = (\alpha_A, \beta_A)$ *dominates* cut $(B) = (\alpha_B, \beta_B)$
$\Rightarrow$ points of the polytope cut by (A) are cut by (B), $\exists$ point cut by (A) not cut by (B).

Dominance consistency:
Given a relaxation point to cut off $\mathbf{x}$, a distance measure $d(\mathbf{x}, \alpha_X, \beta_X)$ is *dominance-consistent* w.r.t. cuts (A) and (B) if
$d(\mathbf{x}, \alpha_A, \beta_A) > d(\mathbf{x}, \alpha_B, \beta_B) \ \Rightarrow \ (A)$ is not dominated by (B).

Results

Projection-based methods are dominance-consistent if cut with smallest measure has **feasible projection** and positive measure.

Directed cutoff measures are dominance-consistent **if all cuts separate** the relaxation point.

Proof: white board depending on time
TL;DR: we draw circles.

Table of Contents

Setup and questions

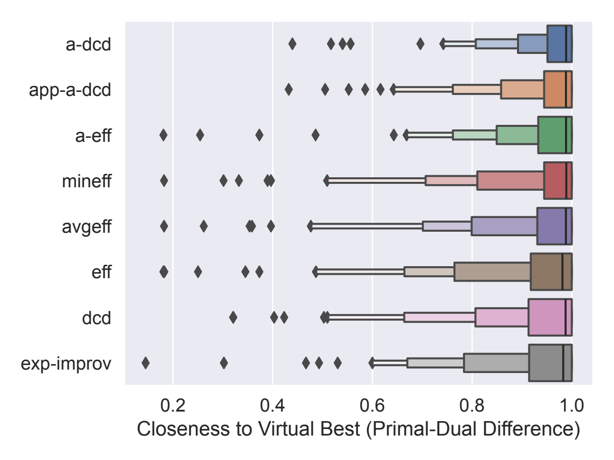
Experiments are performed on 162 MIPLIB instances with SCIP 8.
Analytic centers computed in the transformed (presolved) space.
Separation at the root node only.

Questions:

- Does the cut selection method influence root relaxation strength?
- Does it influence solver performance?
- Are the two tied?

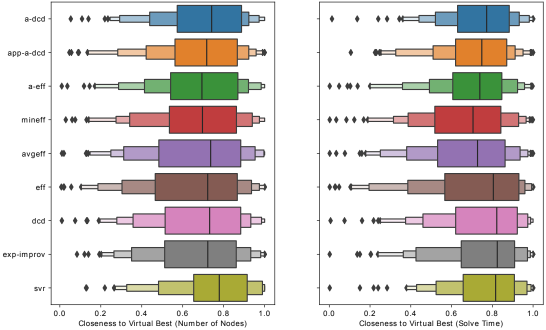
Root node experiments setup

Most measures are usually similar.
Effect on worst cases only.
Influence of performance variability:
→ “trickle down” effect of early cuts.



Tree experiments (1)

Analytic directed cutoff distance:→
best median # nodes.
Directed cutoff distance:
→ best median time, but more tail.
Choosing a scoring function matters.



Tree experiments (2)

Using approximate analytic center:
→ clear degradation.
26.5% of analytic centers infeasible with new cuts.
Efficacy: high variance of results.
Expected improvement & multi-solution efficacy: poor results overall.
90.4% of LP-infeasible projections.
87.5% dual degeneracy.

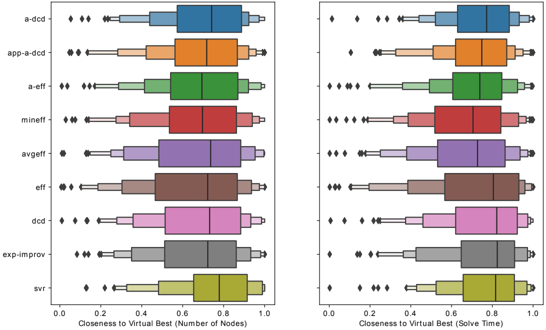


Table of Contents

Are instance features sufficient?

Given instance features, which score would produce the minimum tree size/runtime?

Features of the transformed problem:

- Dual degeneracy: % non-basic variables with zero-reduced cost;
- Primal degeneracy: % basic variables at bounds;
- Solution fractionality at root node;
- Thinness: % equality constraints;
- Density of the whole constraint matrix.

Learning to select cuts

Similar to all Learning to XYZ in MIP $\Rightarrow$ dual goal:

1. **Discovering** rules working well on representative instances;
2. **Building** practical adaptive algorithms within MIP solving.

1. can use sophisticated techniques, high-dimensional kernels, neural networks.
2. requires interpretable, light-weight models, simpler parameter space.

Multiregression model

Selecting best measure to score cuts:
Isn’t that a classification task?

Selecting one as the best is tricky for many instances:
$\rightarrow$ no cut selected, “obvious” cuts always selected, etc...

Ties allowed? What subset of measures included in “best”?

Multiregression model

Selecting best measure to score cuts:
Isn't that a classification task?

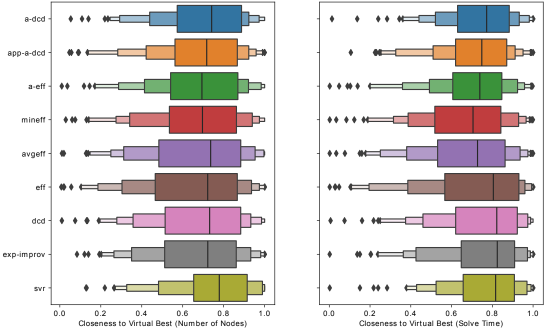
Selecting one as the best is tricky for many instances:
→ no cut selected, "obvious" cuts always selected, etc...

Ties allowed? What subset of measures included in "best"?

Multi-output regression: # nodes method / # nodes best
Several attempted models: regression trees, support vectors, random forests.
Final choice: support vector regression with cubic kernel.

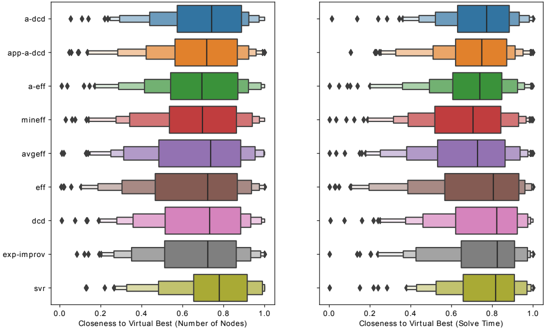
Results

12% fewer nodes on average compared to best method.
But 8% increase in time.



Results

12% fewer nodes on average compared to best method.
But 8% increase in time.
Learning with runtime much harder and not conclusive.
→ time-specific features?



Decision boundaries on projected space (with PCA)

Two PCA components: 71% of the explained variance.
Data points for validation instances.
Opacity
→ relative performance of the predicted measure.

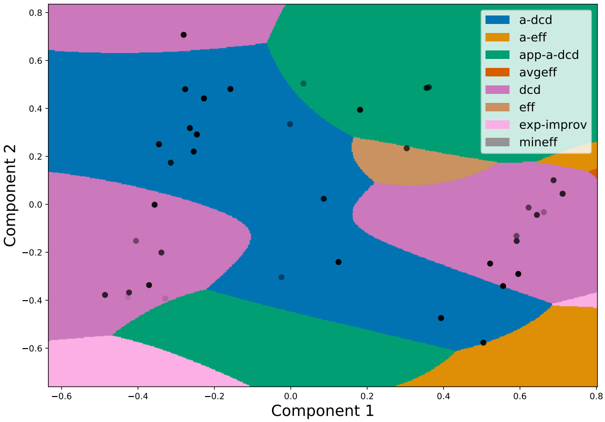


Table of Contents

Conclusion

Highlights:

- Scoring cuts for their selection matters
- Robust analytic measures avoid pitfalls of **dual degeneracy**, **infeasible projections**
- Ability to predict best score for tree size.

Future direction:

- Work on scoring individual cuts: how to pick subsets?
- Not all separation rounds may matter: finer adaptive method?
- Faster and exact analytic centers?

References i

Achterberg, T. (2007).
***Constraint integer programming*.**
PhD thesis, TU Berlin.

Andreello, G., Caprara, A., and Fischetti, M. (2007).
Embedding $\{0, 1/2\}$-cuts in a branch-and-cut framework: A computational study.
INFORMS Journal on Computing, 19(2):229–238.

Baena, D. and Castro, J. (2011).
Using the analytic center in the feasibility pump.
Operations Research Letters, 39(5):310–317.

Baltean-Lugojan, R., Bonami, P., Misener, R., and Tramontani, A. (2019).
Scoring positive semidefinite cutting planes for quadratic optimization via trained neural networks.
optimization-online preprint 2018/11/6943.

Berthold, T., Perregaard, M., and Mészáros, C. (2018).
Four good reasons to use an interior point solver within a MIP solver.
In *Operations Research Proceedings 2017*, pages 159–164. Springer.

Fischetti, M. and Salvagnin, D. (2009).
Yoyo search: a bisection cutting-plane method.

- Naoum-Sawaya, J. (2014).
  Recursive central rounding for mixed integer programs.
  Computers & operations research, 43:191–200.
- Paulus, M. B., Zarpellon, G., Krause, A., Charlin, L., and Maddison, C. (2022).
  Learning to cut by looking ahead: Cutting plane selection via imitation learning.
  In *International Conference on Machine Learning*, pages 17584–17600. PMLR.
- Tang, Y., Agrawal, S., and Faenza, Y. (2020).
  Reinforcement learning for integer programming: Learning to cut.
  In *International Conference on Machine Learning*, pages 9367–9376. PMLR.

Turner, M., Koch, T., Serrano, F., and Winkler, M. (2022).
Adaptive Cut Selection in Mixed-Integer Linear Programming.
arXiv preprint arXiv:2202.10962.

Wesselmann, F. and Stuhl, U. (2012).
Implementing cutting plane management and selection techniques.
Technical report, Technical report, University of Paderborn.