Learning to Cut in Integer Programming

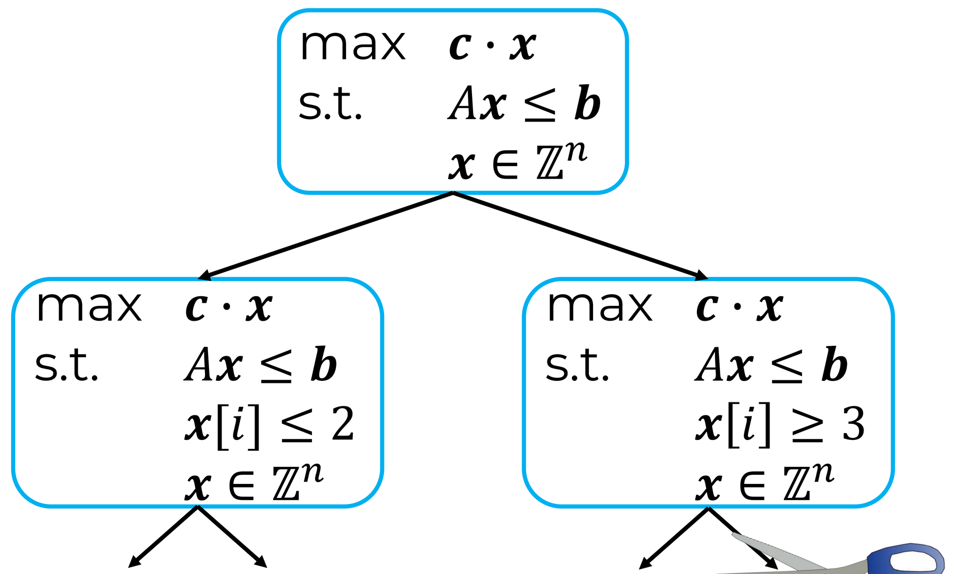
Siddharth Prasad (CMU)

Joint with Nina Balcan (CMU), Tuomas Sandholm (CMU), and Ellen Vitercik (Stanford)

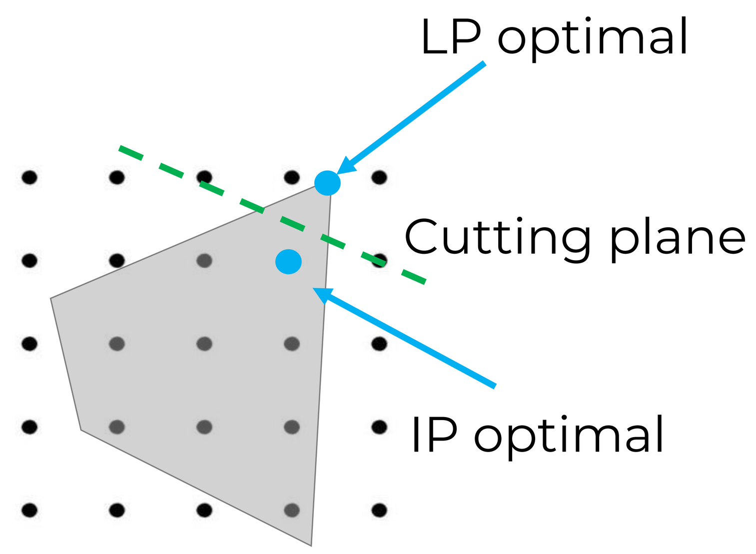
Branch-and-cut for integer programming

LP guidance to do informed search through feasible set

Choose variable i to branch on:
add constraints $\boldsymbol{x}[i] \leq \lfloor \boldsymbol{x}^*_{\mathrm{LP}}[i] \rfloor$, $\boldsymbol{x}[i] \geq \lceil \boldsymbol{x}^*_{\mathrm{LP}}[i] \rceil$



Cutting planes:
Responsible for breakthrough speedups of IP solvers



Prune subtrees if LP relaxation is:
- Integral, or
- Worse than best integral solution found so far

Our contribution:
First formal theory for using ML to select cutting planes

Learning to cut

Best cutting planes for **routing** problems likely not suited for **scheduling**

Application domain modeled by distribution over IPs

Key question: Sample complexity
If cut yields small B&C tree size on average over a training set...

Training set			Fresh sample
max $\boldsymbol{c}_1 \cdot \boldsymbol{x}$ s.t. $A_1\boldsymbol{x} \leq \boldsymbol{b}_1$, $\boldsymbol{x} \in \mathbb{Z}^n$	⋯	max $\boldsymbol{c}_N \cdot \boldsymbol{x}$ s.t. $A_N\boldsymbol{x} \leq \boldsymbol{b}_N$, $\boldsymbol{x} \in \mathbb{Z}^n$	max $\boldsymbol{c} \cdot \boldsymbol{x}$ s.t. $A\boldsymbol{x} \leq \boldsymbol{b}$, $\boldsymbol{x} \in \mathbb{Z}^n$

...will it yield a small B&C tree on a fresh IP?

Sample complexity
- size of training set s.t. |empirical treesize – expected treesize| <= epsilon *for all cuts uniformly*
- grows linearly in ***pseudo-dimension***
- generalization of VC dimension

Cutting planes: Structure and sensitivity

Want to understand B&C tree as a *function* of cutting plane

Chvátal-Gomory Cuts:
$\boldsymbol{\alpha}_1\boldsymbol{x} \leq \beta_1$ of the form $\lfloor \boldsymbol{u}A \rfloor \boldsymbol{x} \leq \lfloor \boldsymbol{u}\boldsymbol{b} \rfloor, \boldsymbol{u} \in [\boldsymbol{0}, \boldsymbol{1}]^{\boldsymbol{m}}$

Theorem: For any IP, $O(\|A\|_{1,1} + \|\boldsymbol{b}\|_1 + n)$ hyperplanes partition $[0,1]^m$ into regions s.t. in each region the cutting plane given by $\boldsymbol{u}$ is the same

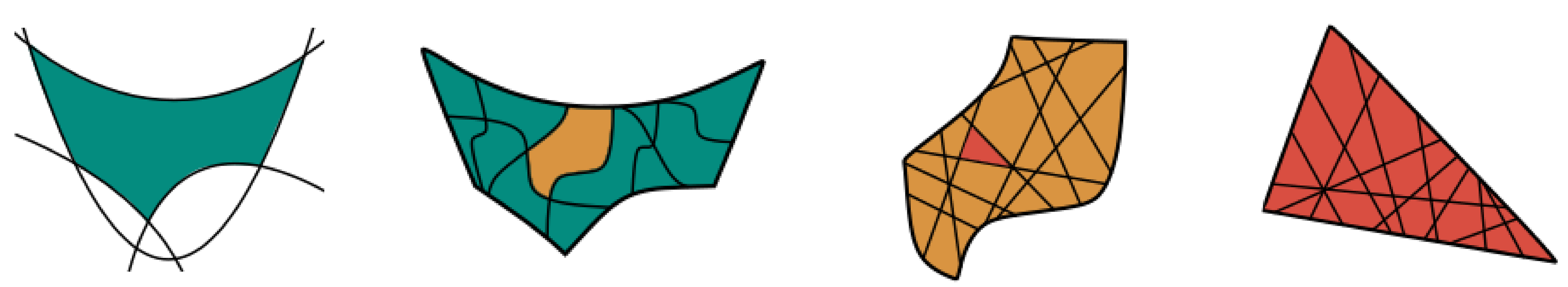
i.e. set of CG cuts "effectively finite"

More generally, when does B&C behave identically on:

max $\boldsymbol{c} \cdot \boldsymbol{x}$ s.t. $A\boldsymbol{x} \leq \boldsymbol{b}$, $\boldsymbol{\alpha}_1\boldsymbol{x} \leq \beta_1$, $\boldsymbol{x} \in \mathbb{Z}^n$ vs max $\boldsymbol{c} \cdot \boldsymbol{x}$ s.t. $A\boldsymbol{x} \leq \boldsymbol{b}$, $\boldsymbol{\alpha}_2\boldsymbol{x} \leq \beta_2$, $\boldsymbol{x} \in \mathbb{Z}^n$

Theorem: For any IP, $O\left(14^n(m+2n)^{3n^2}\tau^{5n^2}\right)$ degree ≤ 5 polynomial hypersurfaces partition $\mathbb{R}^{n+1}$ into regions s.t. in each region the execution of B&C is the same
τ = diameter of feasible polytope (roughly)

Proof. Derive piecewise closed form for LP optimum in terms of $\boldsymbol{\alpha}, \beta$. B&C actions defined by polynomial boundaries



LP optimum closed form | Invariant branching | Invariant LP integrality | Invariant B&C execution

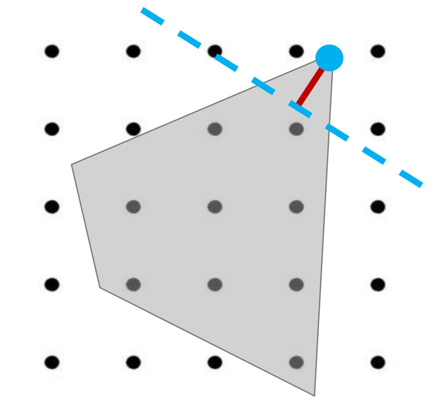
Gomory Mixed Integer Cuts:
Parameterized by $\boldsymbol{u} \in [-\boldsymbol{U}, \boldsymbol{U}]^{\boldsymbol{m}}$, more nuanced rounding than CG cuts, an "infinite" class

Theorem: For any IP, $O(nU^2\|A\|_1\|\boldsymbol{b}\|_1)$ hyperplanes, $2^{O(n^2)}(m+n)^{O(n^3)}\tau^{O(n^3)}$ degree ≤ 10 polynomial hypersurfaces partition $[-U,U]^m$ into regions s.t. in each region the execution of B&C is the same

Learning to cut

Pseudo-dimensions of class of tree-size functions
- Chvátal-Gomory: $O(m\log(\|A\|_{1,1} + \|\boldsymbol{b}\|_1 + n))$
- Gomory mixed integer: $O(mn^3\log(mn\tau\|A\|_1\|\boldsymbol{b}\|_1))$

Cut selection policies
Solvers often use *scoring rules* to choose from a pool of cuts

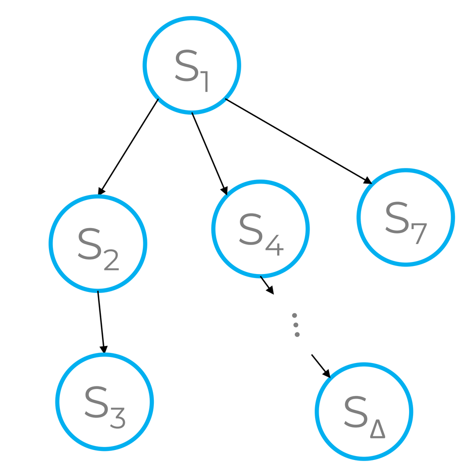


E.g., $\mathrm{score}(\boldsymbol{\alpha}^T\boldsymbol{x} \leq \beta)$ = distance between cut and $\boldsymbol{x}^*_{\mathrm{LP}}$

Given d scoring rules, learn mixture $\mu_1\mathrm{score}_1 + \cdots + \mu_d\mathrm{score}_d$

Theorem: Class of tree-size functions parameterized by $\boldsymbol{\mu}$ has pseudo-dim (for CG cuts) $O(dm\log(\|A\|_{1,1} + \|\boldsymbol{b}\|_1 + n))$

General tree search Model captures branching, cutting planes, node selection *simultaneously*



Theorem: b actions available at each node, take action that maximizes mixture of 2 scoring rules. Pseudo-dim = $O(\Delta^2 + \Delta\log b)$.

Distribution-dependent cut selection can help!
Tuning mixture of scoring rules to select **knapsack cover cuts** for various multiple-knapsack-problem distributions