Separation, Inverse Optimization, and Decomposition: Some Observations

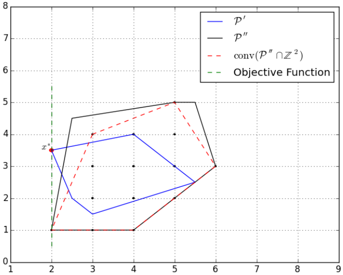
Ted Ralphs[1]
Joint work with:
Aykut Bulut[1]

[1]COR@L Lab, Department of Industrial and Systems Engineering, Lehigh University

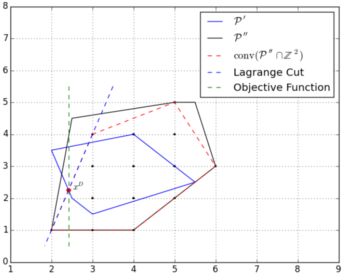
MOA 2016, Beijing, China, 27 June 2016

What Is This Talk About?

- Duality in integer programming.
- Connecting some concepts.
    - Separation problem
    - Inverse optimization
    - Decomposition methods
    - Primal cutting plane algorithms for MILP
- A review of some “well-known”(?) classic results.



(a)



(b)

Setting

- We focus on the case of the mixed integer linear optimization problem (MILP), but many of the concepts are more general.

$$z_{IP} = \min_{x \in \mathcal{S}} c^\top x, \qquad \text{(MILP)}$$

where, $c \in \mathbb{R}^n$, $\mathcal{S} = \{x \in \mathbb{Z}^r \times \mathbb{R}^{n-r} \mid Ax \leq b\}$ with $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$.

- For most of the talk, we consider the case $r = n$ and $\mathcal{P}$ bounded for simplicity.

Duality in Mathematical Optimization

- It is difficult to define precisely what is meant by "duality" in general mathematics, though the literature is replete with various "dualities."
  - Set Theory and Logic (De Morgan Laws)
  - Geometry (Pascal's Theorem & Brianchon's Theorem)
  - Combinatorics (Graph Coloring)
- In optimization, duality is *the* central concept from which much theory and computational practice emerges.

Forms of Duality in Optimization

- NP versus co-NP (computational complexity)
- Separation versus optimization (polarity)
- Inverse optimization versus forward optimization
- Weyl-Minkowski duality (representation theorem)
- Economic duality (pricing and sensitivity)
- Primal/dual functions/problems

Economic Interpretation of Duality

- The economic viewpoint interprets the variables as representing possible *activities* in which one can engage at specific numeric levels.
- The constraints represent available *resources* so that $g_i(\hat{x})$ represents how much of resource i will be consumed at activity levels $\hat{x} \in X$.
- With each $\hat{x} \in X$, we associate a *cost* $f(\hat{x})$ and we say that $\hat{x}$ is *feasible* if $g_i(\hat{x}) \leq b_i$ for all $1 \leq i \leq m$.
- The space in which the vectors of activities live is the *primal space*.
- On the other hand, we may also want to consider the problem from the view point of the *resources* in order to ask questions such as

  - How much are the resources "worth" in the context of the economic system described by the problem?
  - What is the marginal economic profit contributed by each existing activity?
  - What new activities would provide additional profit?

- The *dual space* is the space of *resources* in which we can frame these

What is Duality Used For?

- One way of viewing duality is as a tool for *transformation*.
  - Primal $\Rightarrow$ Dual
  - H-representation $\Rightarrow$ V-representation
  - Membership $\Rightarrow$ Separation
  - Upper bound $\Rightarrow$ Lower bound
  - Primal solutions $\Rightarrow$ Valid inequalities
- Optimization methodologies exploit these dualities in various ways.
  - Solution methods based on primal/dual bounding
  - Generation of valid inequalities
  - Inverse optimization
  - Sensitivity analysis, pricing, warm-starting

Duality in Integer Programming

- The following generalized *dual* can be associated with the base instance (MILP) (see [4])

$$\max \ \{F(b) \mid F(\beta) \leq \phi_D(\beta), \ \beta \in \mathbb{R}^m, F \in \Upsilon^m\} \qquad \text{(D)}$$

where $\Upsilon^m \subseteq \{f \mid f : \mathbb{R}^m \rightarrow \mathbb{R}\}$ and ϕ_D is the *(dual) value function* associated with the base instance (MILP), defined as

$$\phi_D(\beta) = \min_{x \in \mathcal{S}(\beta)} c^\top x \qquad \text{(DVF)}$$

for $\beta \in \mathbb{R}^m$, where $\mathcal{S}(\beta) = \{x \in \mathbb{Z}^r \times \mathbb{R}^{n-r} \mid \ Ax \leq \beta\}$.

- We call F^* *strong* for this instance if F^* is a *feasible* dual function and $F^*(b) = \phi_D(b)$.

The Membership Problem

Membership Problem

Given $x^* \in \mathbb{R}^n$ and polyhedron $\mathcal{P}$, determine whether $x^* \in \mathcal{P}$.

For $\mathcal{P} = \text{conv}(\mathcal{S})$, the membership problem can be formulated as the following LP.

$$\min_{\lambda \in \mathbb{R}^{\mathcal{E}}_{+}} \left\{ 0^\top \lambda \;\middle|\; E\lambda = x^*, 1^\top \lambda = 1 \right\} \qquad \text{(MEM)}$$

where $\mathcal{E}$ is the set of extreme points of $\mathcal{P}$ and E is a matrix whose columns are in correspondence with the members of $\mathcal{E}$.

- When (MEM) is feasible, then we have a proof that $x^* \in \mathcal{P}$.
- When (MEM) is infeasible, we obtain a separating hyperplane.
- It is perhaps not too surprising that the dual of (MEM) is a variant of the *separation problem*.

The Separation Problem

Separation Problem

Given a polyhedron $\mathcal{P}$ and $x^* \in \mathbb{R}^n$, either certify $x^* \in \mathcal{P}$ or determine (π, π_0), a valid inequality for $\mathcal{P}$, such that $\pi x^* > \pi_0$.

For $\mathcal{P}$, the separation problem can be formulated as the dual of (MEM).

$$\max \left\{ \pi x^* - \pi_0 \;\middle|\; \pi^\top x \leq \pi_0 \; \forall x \in \mathcal{E}, (\pi, \pi_0) \in \mathbb{R}^{n+1} \right\} \qquad \text{(SEP)}$$

where $\mathcal{E}$ is the set of extreme points of $\mathcal{P}$.

- Note that we need some appropriate normalization.
- Assuming 0 is in the interior of $\mathcal{P}$, we can take $\pi_0 = 1$.
- In this case, we are optimizing over the *1-polar* of $\mathcal{P}$.
- This is equivalent to changing the objective of (MEM) to $\min 1^\top \lambda$.

The 1-Polar

Assuming 0 is in the interior of $\mathcal{P}$, the set of all inequalities valid for $\mathcal{P}$ is

$$\mathcal{P}^* = \left\{ \pi \in \mathbb{R}^n \;\middle|\; \pi^\top x \leq 1 \; \forall x \in \mathcal{P} \right\} \tag{1}$$

and is called its *1-polar*.

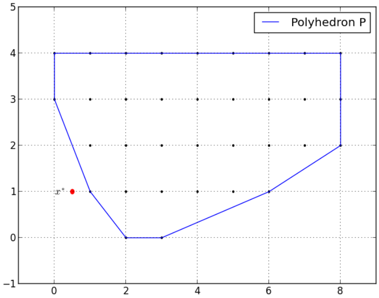
Properties of the 1-Polar

- $\mathcal{P}^*$ is a polyhedron;
- $\mathcal{P}^{**} = \mathcal{P}$;
- $x \in \mathcal{P}$ if and only if $\pi^\top x \leq 1 \; \forall \pi \in \mathcal{P}^*$;
- If $\mathcal{E}$ and $\mathcal{R}$ are the extreme points and extreme rays of $\mathcal{P}$, respectively, then
$$\mathcal{P}^* = \left\{ \pi \in \mathbb{R}^n \;\middle|\; \pi^\top x \leq 1 \; \forall x \in \mathcal{E}, \pi^\top r \leq 0 \; \forall r \in \mathcal{R} \right\}.$$

- A converse of the last result also holds.
- Separation can be interpreted as optimization over the polar.

Separation Using an Optimization Oracle

- We can solve (SEP) using a cutting plane algorithm that separates intermediate solutions from the 1-polar.
- The separation problem for the 1-polar of $\mathcal{P}$ is precisely a linear optimization problem over $\mathcal{P}$.
- We can visualize this in the dual space as column generation wrt (MEM).
- Example



Separation Example: Iteration 1

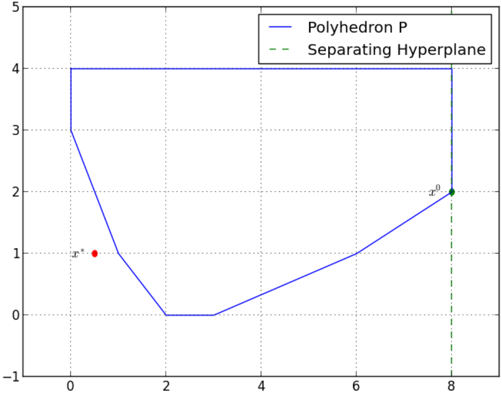


Figure: Separating x^* from $\mathcal{P}$ (Iteration 1)

Separation Example: Iteration 2

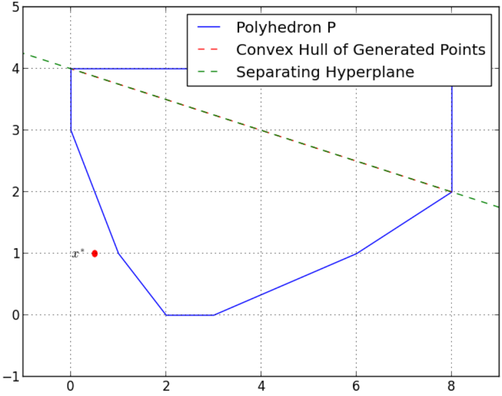


Figure: Separating x^* from $\mathcal{P}$ (Iteration 2)

Separation Example: Iteration 3

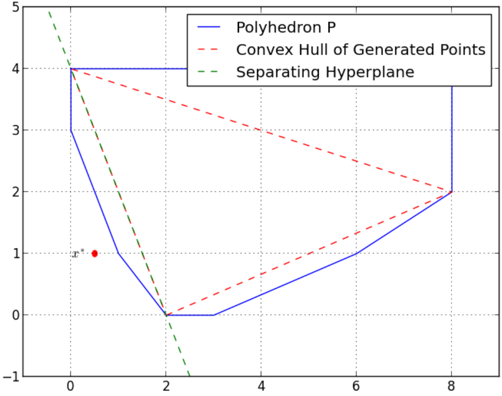


Figure: Separating x^* from $\mathcal{P}$ (Iteration 3)

Separation Example: Iteration 4

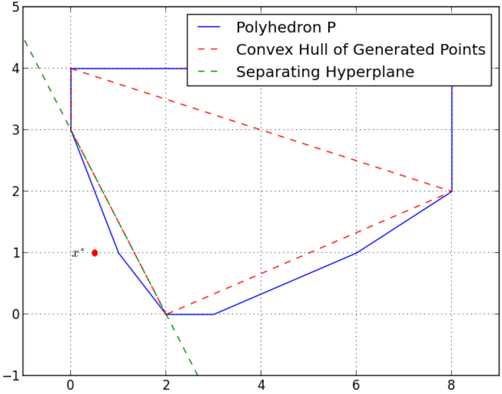


Figure: Separating x^* from $\mathcal{P}$ (Iteration 4)

Separation Example: Iteration 5

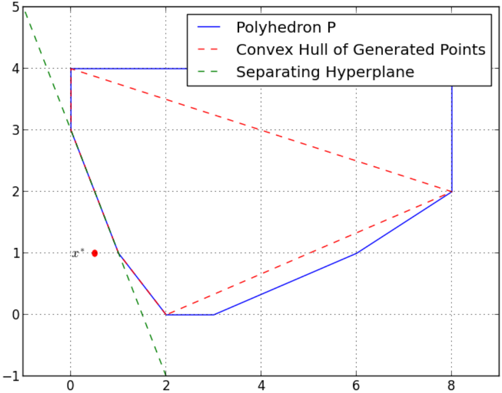


Figure: Separating x^* from $\mathcal{P}$ (Iteration 5)

Inverse Problems

What is an inverse problem?

Given a function, an inverse problem is that of determining *input* that would produce a given *output*.

- The input may be partially specified.
- We may want an answer as close as possible to a given *target*.

- This is precisely the mathematical notion of the inverse of a function.
- A *value function* is a function whose value is the optimal solution of an optimization problem defined by the given input.
- The inverse problem with respect to an optimization problem is to evaluate the inverse of a given *value function*.

Why is Inverse Optimization Useful?

Inverse optimization is useful when we can observe the result of solving an optimization problem and we want to know what the input was.

Example: Consumer preferences

- Let's assume consumers are rational and are making decisions by solving an underlying optimization problem.
- By observing their choices, we try ascertain their utility function.

Example: Analyzing seismic waves

- We know that the path of seismic waves travels along paths that are optimal with respect to some physical model of the earth.
- By observing how these waves travel during an earthquake, we can infer things about the composition of the earth.

Formal Setting

We consider the inverse of the *(primal) value function* ϕ_P, defined as

$$\phi_P(d) = \min_{x \in \mathcal{S}} d^\top x = \min_{x \in \text{conv}(\mathcal{S})} d^\top x \; \forall d \in \mathbb{R}^n. \quad \text{(PVF)}$$

With respect to a given $x^0 \in \mathcal{S}$, the inverse problem is defined as

$$\min \left\{ f(d) \;\middle|\; d^\top x^0 = \phi_P(d) \right\}, \quad \text{(INV)}$$

The classical objective function is taken to be $f(d) = \|c - d\|$, where $c \in \mathbb{R}^n$ is a given target.

A Small Example

- The feasible set of the inverse problem is the set of objective vectors that make x^0 optimal.
- This is precisely the dual of cone$(\mathcal{S} - \{x^0\})$, which is, roughly, a translation of the polyhedron described by the inequalities binding at x^0.

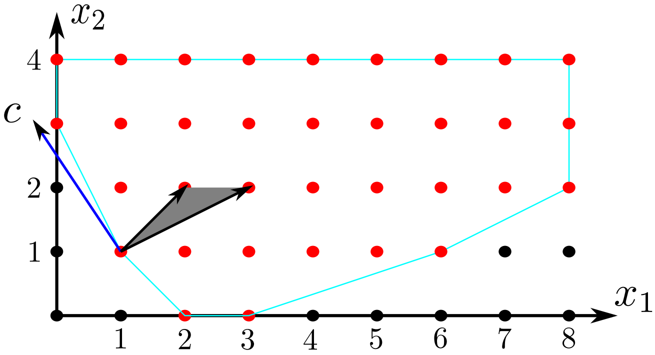


Figure: conv$(\mathcal{S})$ and cone $\mathcal{D}$ of feasible objectives

Inverse Optimization as a Mathematical Program

- To formulate as a mathematical program, we need to represent the implicit constraints of (INV) explicitly.
- The cone of feasible objective vectors can be described as

$$\mathcal{D} = \left\{ d \in \mathbb{R}^n \;\middle|\; d^\top x^0 \leq d^\top x \; \forall x \in \mathcal{S} \right\} \tag{IFS}$$

- Since $\mathcal{P}$ is bounded, we need only the inequalities corresponding to extreme points of $\text{conv}(\mathcal{S})$.
- This set of constraints is exponential in size, but we can generate them dynamically, as we will see.
- Note that this is the set of inequalities valid for $\mathcal{S}$ that are binding at x^0.
- Alternatively, it is the set of all inequalities valid for the so-called *corner polyhedron* with respect to x^0.

Formulating the Inverse Problem

General Formulation

$$\begin{aligned} \min \quad & f(d) \\ s.t. \quad & d^\top x^0 \leq d^\top x \qquad \forall x \in \mathcal{E} \end{aligned} \tag{INVMP}$$

- With $f(d) = \|c - d\|$, this can be linearized for ℓ_1 and ℓ_∞ norms.
- The separation problem for the feasible region is again optimization over $\text{conv}(\mathcal{S})$.

Separation and Inverse Optimization

- It should be clear that inverse optimization and separation are very closely related.
- First, note that the inequality

$$\pi^\top x \geq \pi_0 \tag{PI}$$

  is valid for $\mathcal{P}$ if and only if $\pi_0 \leq \phi_P(\pi)$.
- We refer to inequalities of the form (PI) for which $\pi_0 = \phi_P(\pi)$ as *primal inequalities*.
- This is as opposed to *dual inequalities* for which $\pi_0 = \phi^\pi(b)$, where ϕ^π is a *dual function* for (MILP) when the objective function is taken as π.
- The feasible set of (INV) can be seen as the set of all valid primal inequalities that are tight at x^0.

Primal Separation

- Suppose we take $f(d) = d^\top x^0 - d^\top x^*$ for given $x^* \in \mathbb{R}^n$.
- Then this problem is something like the classical separation problem.
- This variant is what [6] called the *primal separation problem* (see also [5]).
- Their original idea was to separate x^* with an inequality binding at the current incumbent.
- Taking x^0 to tbe the current incumbent, this is exactly what we're doing.
- With this objective, we need a normalization to ensure boundedness, as before.
- A straightforward option is to take $d^\top x^0 = 1$.
- (INVMP) is then equivalent to the separation problem for the conic hull of $\mathcal{S} - \{x^0\}$.

Dual of the Inverse Problem

- Roughly speaking, the dual of (INVMP) is the membership problem for $\text{cone}(\mathcal{S} - \{x^0\})$.

$$\min_{\lambda \in \mathbb{R}^{\mathcal{E}}_{+}} \left\{ 0^\top \lambda \;\middle|\; \bar{E}\lambda = x^* - x^0 \right\} \qquad \text{(CMEM)}$$

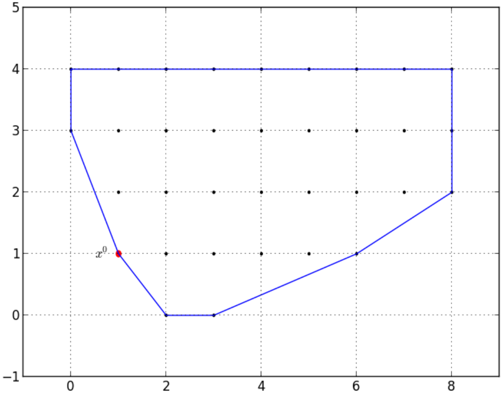
- With the normalization, this becomes

$$\min_{\lambda \in \mathbb{R}^{\mathcal{E}}_{+}} \left\{ \alpha \;\middle|\; \bar{E}\lambda + \alpha 1 = x^* - x^0 \right\}, \qquad \text{(CMEMN)}$$

where $\bar{E}$ is the set of extreme rays of $\text{cone}(\mathcal{S} - \{x^0\})$

Inverse Optimization with Forward Optimization Oracle

- We can use an algorithm almost identical to the one from earlier.
- We now generate inequalities valid for the corner polyhedron associated with x^0.



Inverse Example: Iteration 1

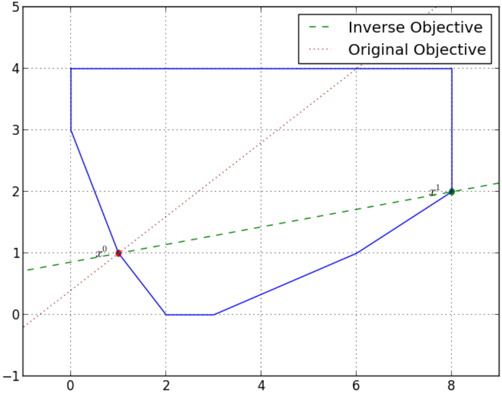


Figure: Solving the inverse problem for $\mathcal{P}$ (Iteration 1)

Inverse Example: Iteration 2

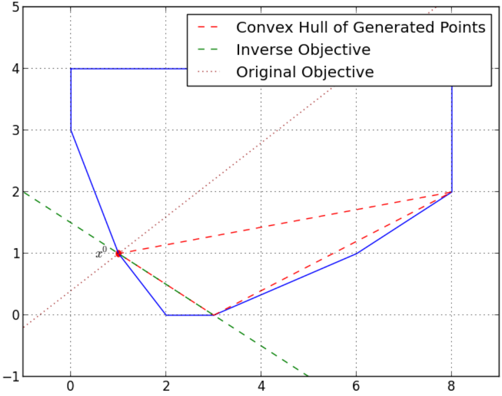


Figure: Solving the inverse problem for $\mathcal{P}$ (Iteration 3)

Inverse Example: Iteration 3

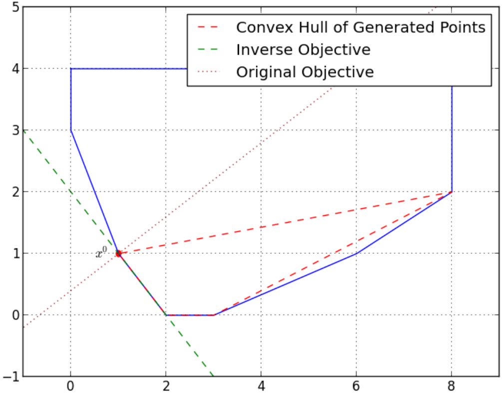


Figure: Solving the inverse problem for $\mathcal{P}$ (Iteration 3)

Solvability of Inverse MILP

Theorem 1 *[?] Inverse MILP optimization problem under ℓ_∞/ℓ_1 norm is solvable in time polynomial in the size of the problem input, given an oracle for the MILP decision problem.*

- This is a direct result of the well-known result of [3].
- GLS does not, however, tell us the formal complexity.

Formal Complexity of Inverse MILP

Sets

$$\mathcal{K}(\gamma) = \{d \in \mathbb{R}^n \mid \|c - d\| \leq \gamma\}$$
$$\mathcal{X}(\gamma) = \{x \in \mathcal{S} \mid \exists d \in \mathcal{K}(\gamma) \ s.t. \ d^\top(x - x^0) > 0\},$$
$$\mathcal{K}^*(\gamma) = \{x \in \mathbb{R}^n \mid d^\top(x - x^0) \geq 0 \ \forall d \in \mathcal{K}(\gamma)\}.$$

Inverse MILP Decision Problem (INVD)

Inputs: γ, c, $x^0 \in \mathcal{S}$ and MILP feasible set $\mathcal{S}$.
Problem: Decide whether $\mathcal{K}(\gamma) \cap \mathcal{D}$ is non-empty.

Theorem 2 *[?] INVD is* coNP*–complete.*

Theorem 3 *[?]Both* (MILP) *and* (INV) *optimal value problems are D^p–complete.*

Connections to Constraint Decomposition

As usual, we divide the constraints into two sets.

$$
\begin{aligned}
\min\ & c^\top x \\
\text{s.t. } A'x &\leq b' \text{ (the "nice" constraints)} \\
A''x &\leq b'' \text{ (the "complicating" constraints)} \\
x &\in \mathbb{Z}^n
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{P}' &= \{x \in \mathbb{R}^n \mid A'x \leq b'\}, \\
\mathcal{P}'' &= \{x \in \mathbb{R}^n \mid A''x \leq b''\}, \\
\mathcal{P} &= \mathcal{P}' \cap \mathcal{P}'', \\
\mathcal{S} &= \mathcal{P} \cap \mathbb{Z}^n, \text{ and} \\
\mathcal{S}_R &= \mathcal{P}' \cap \mathbb{Z}^n.
\end{aligned}
$$

Reformulation

- We replace the H-representation of the polyhedron $\mathcal{P}'$ with a V-representation of $\text{conv}(\mathcal{S}_R)$.

$$\begin{aligned} \min \quad & c^\top x && (2)\\ \text{s.t.} \quad & \sum_{s\in\mathcal{E}} \lambda_s s = x && (3)\\ & A''x \leq b'' && (4)\\ & \sum_{s\in\mathcal{E}} \lambda_s = 1 && (5)\\ & \lambda \in \mathbb{R}_+^{\mathcal{E}} && (6)\\ & x \in \mathbb{Z}^n && (7) \end{aligned}$$

where $\mathcal{E}$ is the set of extreme points of $\text{conv}(\mathcal{S}_R)$.

- If we relax the integrality consraints (7), then we can also drop (3) and we obtain a relaxation which is tractable.
- This relaxation may yield a bound better than that of the LP relaxation.

The Decomposition Bound

Using the aformentioned relaxation, we obtain a formulation for the so-called *decomposition bound*.

$$
\begin{aligned}
z_{\mathrm{IP}} &= \min_{x \in \mathbb{Z}^n} \left\{ c^\top x \;\middle|\; A'x \leq b', A''x \leq b'' \right\} \\
z_{\mathrm{LP}} &= \min_{x \in \mathbb{R}^n} \left\{ c^\top x \;\middle|\; A'x \leq b', A''x \leq b'' \right\} \\
z_{\mathrm{D}} &= \min_{x \in \mathrm{conv}(\mathcal{S}_R)} \left\{ c^\top x \;\middle|\; A''x \leq b'' \right\}
\end{aligned}
$$

$$
z_{\mathrm{IP}} \geq z_{\mathrm{D}} \geq z_{\mathrm{LP}}
$$

It is well-known that this bound can be computed using various decomposition-based algorithms:

- Lagrangian relaxation
- Dantzig-Wolfe decomposition
- Cutting plane method

Shameless plug: Try out DIP/DipPy!

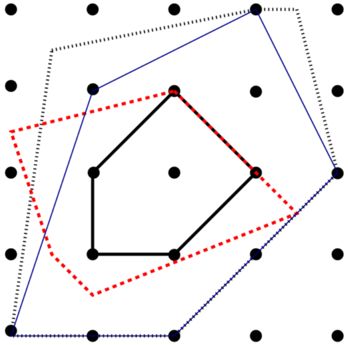
A framework for switching between various decomp-based algorithms.

Example

$$
\begin{aligned}
\min \quad & x_1 \\
-x_1 - x_2 &\geq -8, & (8)\\
-0.4x_1 + x_2 &\geq 0.3, & (9)\\
x_1 + x_2 &\geq 4.5, & (10)\\
3x_1 + x_2 &\geq 9.5, & (11)\\
0.25x_1 - x_2 &\geq -3, & (12)\\
7x_1 - x_2 &\geq 13, & (13)\\
x_2 &\geq 1, & (14)\\
-x_1 + x_2 &\geq -3, & (15)\\
-4x_1 - x_2 &\geq -27, & (16)\\
-x_2 &\geq -5, & (17)\\
0.2x_1 - x_2 &\geq -4, & (18)\\
x &\in \mathbb{Z}^n. & (19)
\end{aligned}
$$

Example (cont)

$$
\begin{aligned}
\mathcal{Q}' &= \{x \in \mathbb{R}^2 \mid x \text{ satisfies } (8) - (12)\}, \\
\mathcal{Q}'' &= \{x \in \mathbb{R}^2 \mid x \text{ satisfies } (13) - (18)\}, \\
\mathcal{Q} &= \mathcal{Q}' \cap \mathcal{Q}'', \\
\mathcal{S} &= \mathcal{Q} \cap \mathbb{Z}^n, \text{ and} \\
\mathcal{S}_R &= \mathcal{Q}' \cap \mathbb{Z}^n.
\end{aligned}
$$



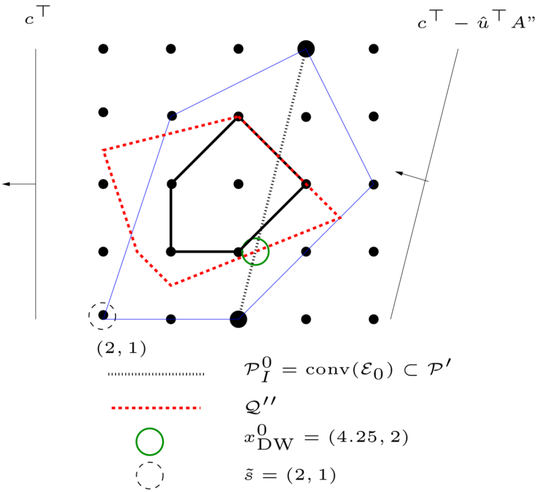
$\mathrm{conv}(S) = \mathrm{conv}\{x \in \mathbb{Z}^n \mid A'x \geq b', A''x \geq b''\}$

$\mathrm{conv}(\mathcal{S}_R) = \mathrm{conv}\{x \in \mathbb{Z}^n \mid A'x \geq b'\}$

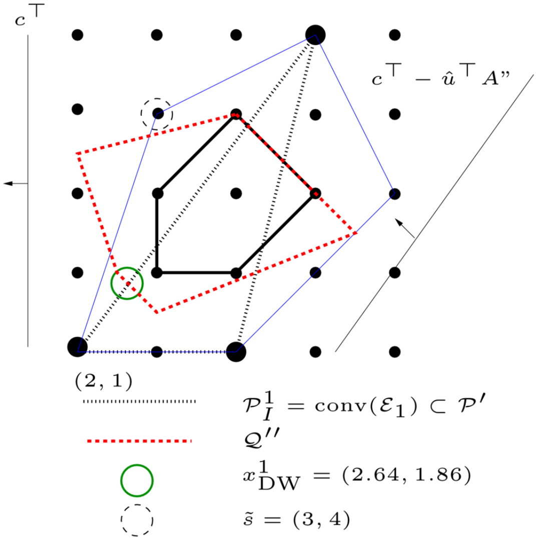
$\mathcal{Q}' = \{x \in \mathbb{R}^n \mid A'x \geq b'\}$

$\mathcal{Q}'' = \{x \in \mathbb{R}^n \mid A''x \geq b''\}$

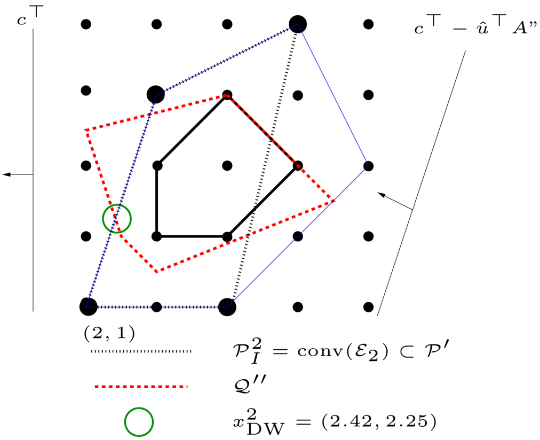
Geometry of Dantzig-Wolfe Decomposition



Geometry of Dantzig-Wolfe Decomposition



Geometry of Dantzig-Wolfe Decomposition



Lagrange Cuts

- [1] observed that for $u \in \mathbb{R}^m_-$, a *Lagrange cut* of the form

$$(c - uA'')^\top x \geq LR(u) - ub'' \tag{LC}$$

  is valid for $\mathcal{P}$.
- If we take u^* to be the optimal solution to the Lagrangian dual, then this inequality reduces to

$$(c - u^*A'')^\top x \geq z_D - ub'' \tag{OLC}$$

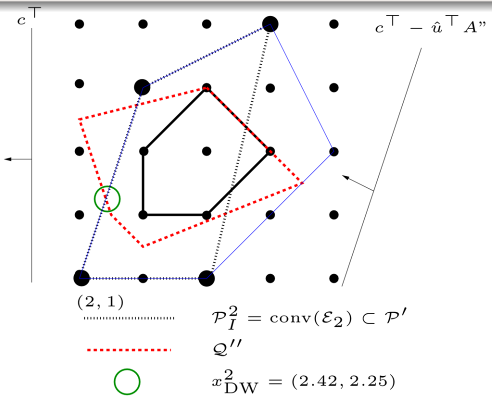
- If we now take

$$x^D \in \operatorname{argmin}\left\{c^\top x \mid A''x \leq b'', (c - u^*A'')^\top x \geq z_D - ub''\right\},$$

  then we have $c^\top x^D = z_D$.

Results

- The inequality (OLC) is a primal inequality for $\mathrm{conv}(\mathcal{S}_R)$ wrt x^D.
- $c - uA''$ is a solution to the inverse problem wrt $\mathrm{conv}(\mathcal{S}_R)$ and x^D.
- These properties also hold for $e \in \mathcal{E}$ such that $\lambda_e^* > 0$ in the RMP.



Conclusions and Future Work

- We gave a brief overview of connections between a number of different problems and methodologies.
- Exploring these connections may be useful to improving intuition and understanding.
- The connection to primal cutting plane algorithms is still largely unexplored, but this should lead to new algorithms for the inverse problem.
- We did not touch much on complexity, but it should be possible to generalize complexity results to the separation/optimization context.
- We believe GLS can be extended to show that inverse optimization forward optimization, and and separation are all complexity-wise equivalent.
- Much of that is discussed here can be further generalized to general computation via Turing machines (useful?).

Thank You!

References I

A. Boyd.
The lagrangian and other primal cutting planes for linear integer programming.
Technical Report TR90-03, Rice University, Department of Computational and Applied Mathematics, 1990.

A Bulut and T.K. Ralphs.
On the complexity of inverse mixed integer linear optimization.
Technical report, COR@L Laboratory Report 15T-001-R1, Lehigh University, 2015.

M. Grötschel, L. Lovász, and A. Schrijver.
Geometric Algorithms and Combinatorial Optimization, volume 2 of *Algorithms and Combinatorics*.
Springer, second corrected edition edition, 1993.

References II

M Güzelsoy and T K Ralphs.
Duality for mixed-integer linear programs.
International Journal of Operations Research, 4:118–137, 2007.

A. Lodi and A.N. Letchford.
Primal separation algorithms.
4OR, 1:209–224, 2003.

M.W. Padberg and M. Grötschel.
Polyhedral computations.
In E. L. Lawler, Jan Karel Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys, editors, *The Traveling Salesman Problem. A Guided Tour of Combinatorial Optimization*, chapter 9, pages 307–360. Wiley, 1985.