Separation, Inverse Optimization, and Decomposition: Connecting the Dots

Ted Ralphs[1]
Joint work with Aykut Bulut[2]

[1]COR@L Lab, Department of Industrial and Systems Engineering, Lehigh University
[2]The MathWorks, Natick, MA

23rd Workshop on Combinatorial Optimization, Aussois, 9 January 2019

ISE Industrial and Systems Engineering

COR@L
COMPUTATIONAL OPTIMIZATION RESEARCH AT LEHIGH

Setting

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

$$z_{IP} = \max_{x \in \mathcal{S}} c^\top x, \tag{MILP}$$

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

- Duality is a central concept from which much theory and computational practice emerges in optimization.
- Many of the well-known “dualities” that arise in optimization are closely connected.
- This talk will illustrate a few of these connections.

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)
- Conic duality
- Gauge/Lagrangian/Fenchel duality
- Primal/dual functions/problems

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.

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

  where $\mathcal{E}$ is the set of extreme points of $\mathcal{P}$ and E is a matrix whose columns are 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 well-known 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).

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

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

- Note that we need some appropriate normalization.

Normalizing

- Assuming 0 is in the interior of $\mathcal{P}$, we can normalize by taking $\pi_0 = 1$.
- In this case, we are optimizing over the *1-polar* of $\mathcal{P}$.
- This is equivalent to changing the objective of (MEM) to $\min 1^\top \lambda$.
- In this case, (MEM) becomes the problem of evaluating the gauge function of $\mathcal{P}$ at x^*.
- If the result is greater than one, x^* is not in $\mathcal{P}$, otherwise it is.

The 1-Polar

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

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

and is called its *1-polar*.

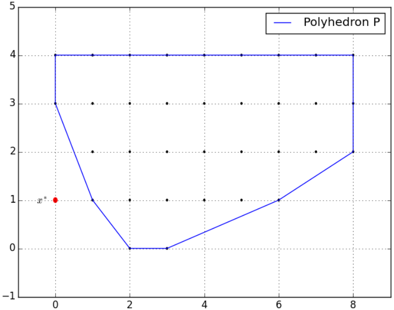
Properties of the 1-Polar

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

- 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.
- 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).
- This is the basis for a number of separation techniques in the literature.
- Example



Separation Example: Iteration 1

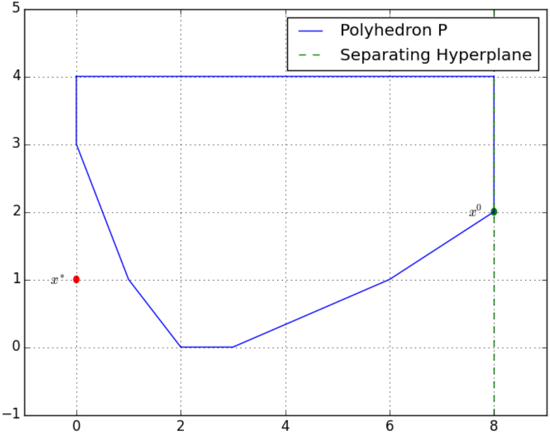


Figure: Separating x^* from $\mathcal{P}$ (Iteration 1)

Separation Example: Iteration 2

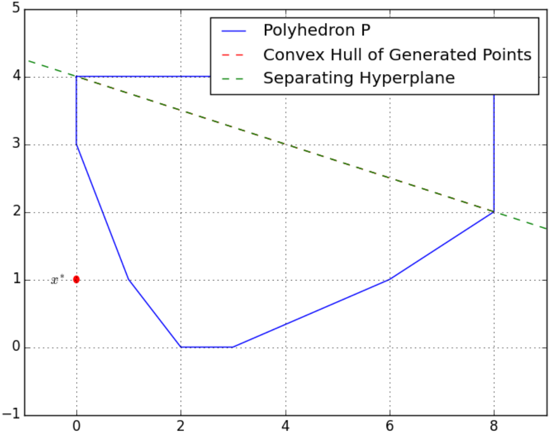


Figure: Separating x^* from $\mathcal{P}$ (Iteration 2)

Separation Example: Iteration 3

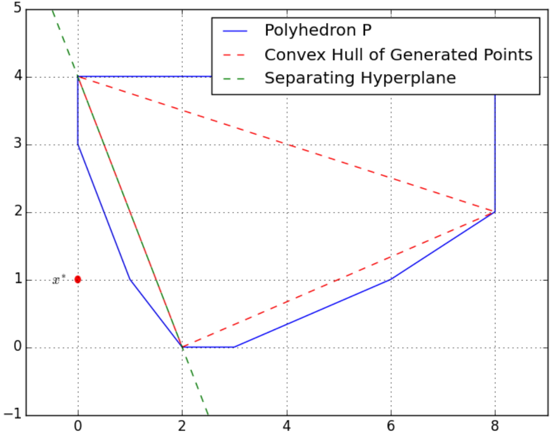


Figure: Separating x^* from $\mathcal{P}$ (Iteration 3)

Separation Example: Iteration 4

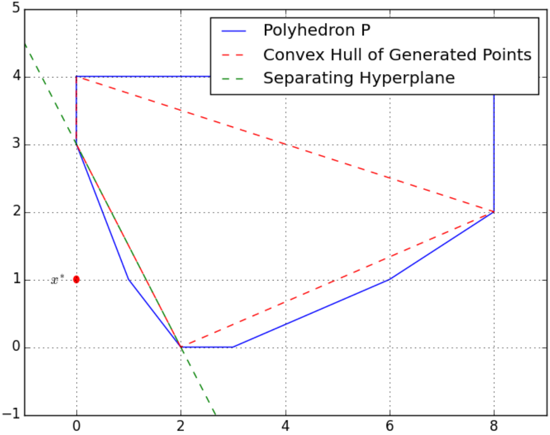


Figure: Separating x^* from $\mathcal{P}$ (Iteration 4)

Separation Example: Iteration 5

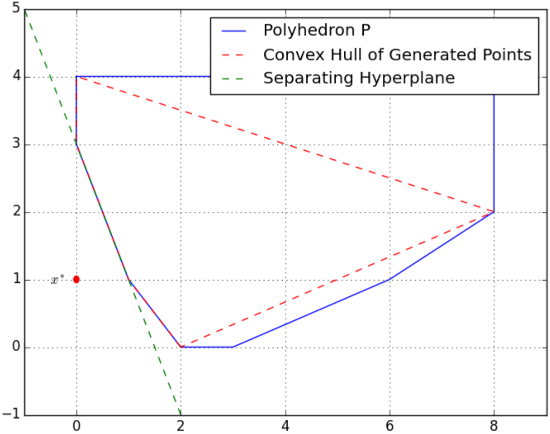


Figure: Separating x^* from $\mathcal{P}$ (Iteration 5)

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.

A Value Function

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

$$\phi_P(d) = \operatorname{argmax}_{x \in \mathcal{S}} d^\top x = \operatorname{argmax}_{x \in \operatorname{conv}(\mathcal{S})} d^\top x \ \forall d \in \mathbb{R}^n. \tag{PVF}$$

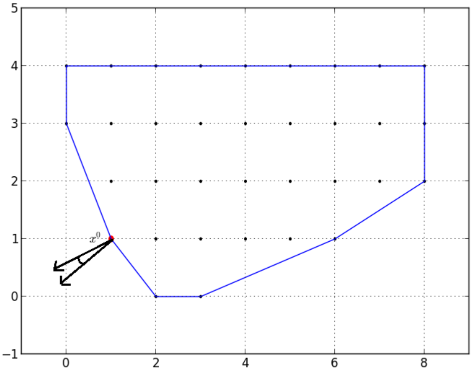
- With respect to a given $x^0 \in \mathcal{S}$, the inverse problem is defined as

$$\min_{d \in \phi^{-1}(x^0)} f(d) \tag{INV}$$

- Feasible solutions are objective function vectors in the preimage of x^0.
- The classical objective function is taken to be $f(d) = \|c - d\|$, where $c \in \mathbb{R}^n$ is a given target.
- Note that this form of the objective makes normalization a bit trickier (more later).

A Small Example

- The feasible set of the inverse problem is the set of objective vectors that make x^0 optimal.
- This is the polar of $\mathrm{cone}(\mathcal{S} - \{x^0\})$, which is a translation of the radial cone of $\mathcal{P}$ at vertex x^0, which we'll denote as $\mathcal{P}^{x^0}$.



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

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

  where $\mathcal{E}$ is the set of extreme points of $\operatorname{conv}(\mathcal{S})$.
- Note that this corresponds to the set of inequalities valid for $\mathcal{S}$ that are binding at x^0.
- These are referred to as *primal inequalities*.

Formulating the Inverse Problem

General Formulation

$$\begin{array}{rlll} \min & f(d) & & \\ s.t. & d^\top x \le d^\top x^0 & \forall x \in \mathcal{E} & \text{(INVMP)} \end{array}$$

- 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 $\mathrm{conv}(\mathcal{S})$.

Separation and Inverse Optimization

- Inverse optimization and separation are very closely related.
- The inverse MILP problem can also be formulated in terms of the 1-polar.
- To replicate the standard objective, we need to allow re-scaling of the normalized objective.

$$
\begin{aligned}
\min \quad & \|c - \alpha d\| \\
\text{s.t.} \quad & d \in \mathcal{P}^* \\
& d^\top x^0 = 1 \\
& \alpha \in \mathbb{R}_+.
\end{aligned}
\tag{INVMILP-1P}
$$

Primal Separation

- If we take $f(d) = d^\top x^0 - d^\top x^*$ for given $x^* \in \mathbb{R}^n$, then we get something like the classical separation problem.
- This variant is what Padberg and Grötschel [1985] called the *primal separation problem* (see also Lodi and Letchford [2003]).
- The original idea was to separate x^* with an inequality binding at the current incumbent.
- This is exactly what the inverse problem is doing with respect to x^0!
- As before, we need a normalization to ensure boundedness.
- Assuming again that 0 is in the interior of $\mathrm{conv}(\mathcal{S})$), we can take $d^\top x^0 = 1$.
- Then (INVMP) is precisely the separation problem for $\mathcal{P}^{x^0}$.

Dual of the Primal Separation Problem

- The dual of the primal separation problem is the membership problem for $\mathcal{P}^{x^0}$.

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

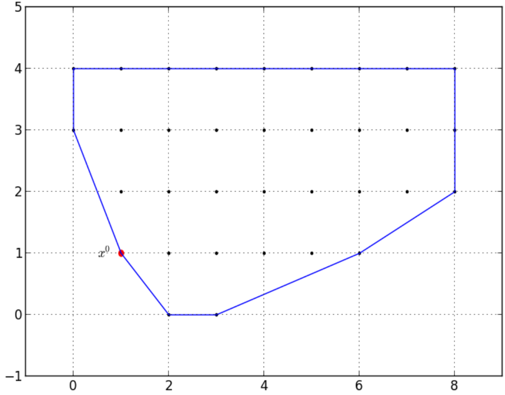
  where $\bar{E}$ is the set of extreme rays of $\mathcal{P}^{x^0}$.
- With the normalization, this becomes

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

- We can interpret the value of α as the amount by which we need to shift x^* along the direction x^0 in order for it to be inside $\text{cone}(\mathcal{S} - \{x^0\})$.
- If the optimal value is greater than one, then $x^* - x^0$ is not in the cone, otherwise it is.

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 relaxation associated with x^0.



Inverse Example: Iteration 1

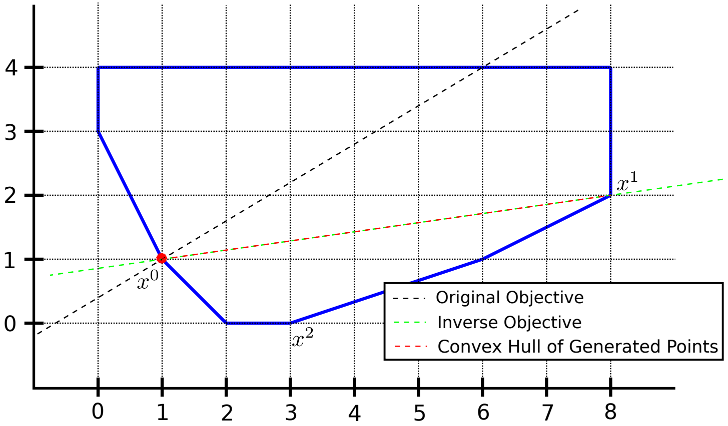


Figure: Solving the inverse problem for $\mathcal{P}$ (Iteration 1)

Inverse Example: Iteration 2

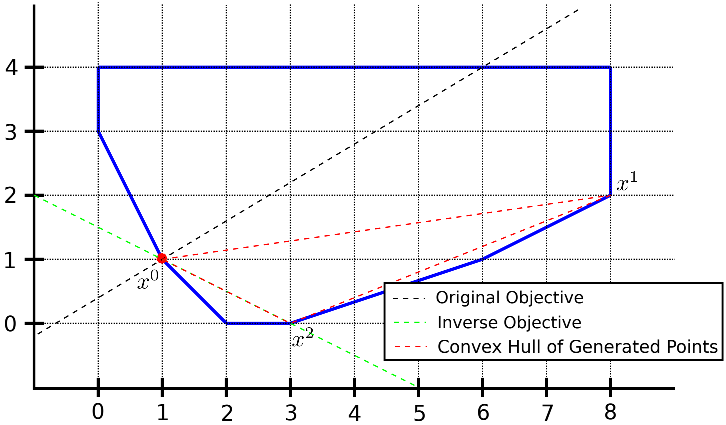


Figure: Solving the inverse problem for $\mathcal{P}$ (Iteration 3)

Inverse Example: Iteration 3

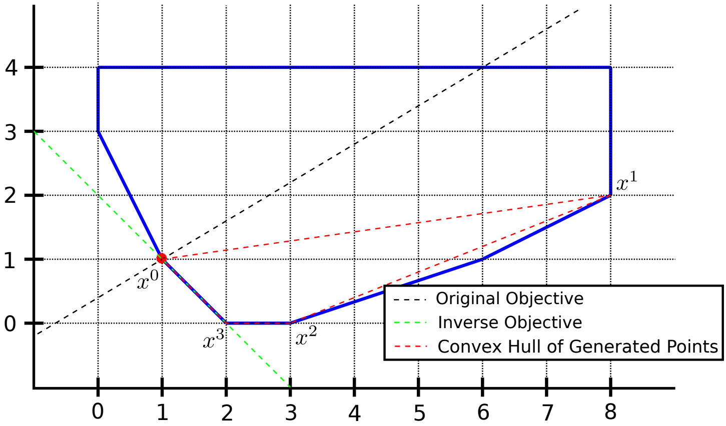


Figure: Solving the inverse problem for $\mathcal{P}$ (Iteration 3)

Tractability of Inverse MILP

Theorem 1 *Bulut and Ralphs [2015] 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 Grötschel et al. [1993].

Complexity of Inverse MILP

Sets

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

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 *Bulut and Ralphs [2015] INVD is* coNP*–complete.*

Theorem 3 *Bulut and Ralphs [2015]Both* (MILP) *and* (INV) *optimal value problems are* D^p*–complete.*

Connections to Constraint Decomposition

As usual, we divide the constraints into two sets.

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

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

Reformulation

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

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

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

- If we relax the integrality constraints (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*.

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

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

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.

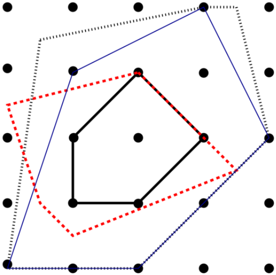
Example

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

Example (cont)

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

Constraint Decomposition in Integer Programming



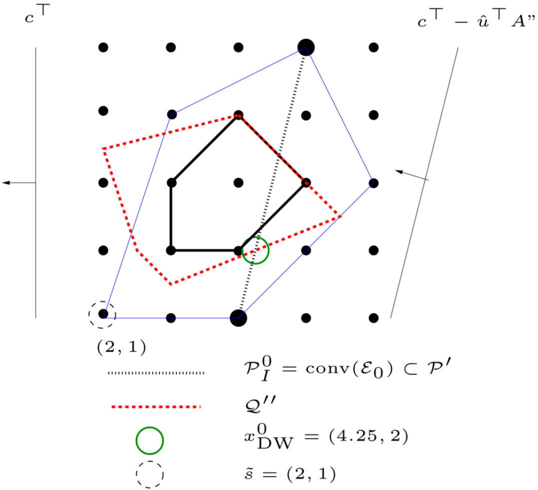
━━━ $\mathrm{conv}(S) = \mathrm{conv}\{x \in \mathbb{Z}^n \mid A'x \geq b', A''x \geq b''\}$

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

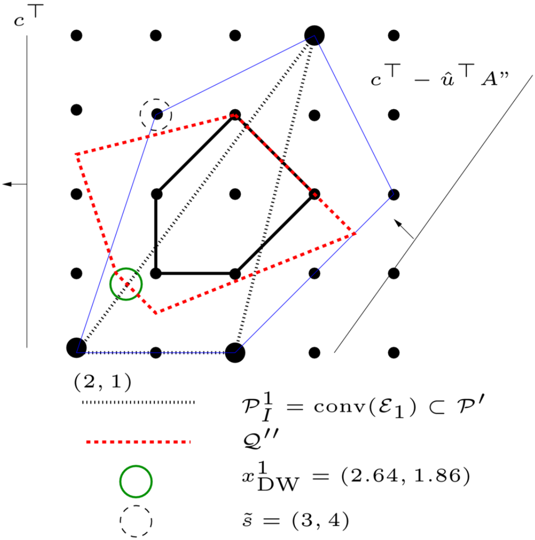
┈┈┈ $\mathcal{Q}' = \{x \in \mathbb{R}^n \mid A'x \geq b'\}$

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

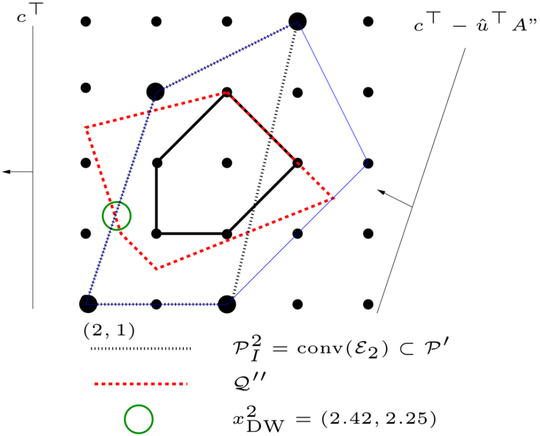
Geometry of Dantzig-Wolfe Decomposition



Geometry of Dantzig-Wolfe Decomposition



Geometry of Dantzig-Wolfe Decomposition



Lagrange Cuts

- Boyd [1990] observed that for $u \in \mathbb{R}^m_+$, a *Lagrange cut* of the form

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

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

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

- If we now take

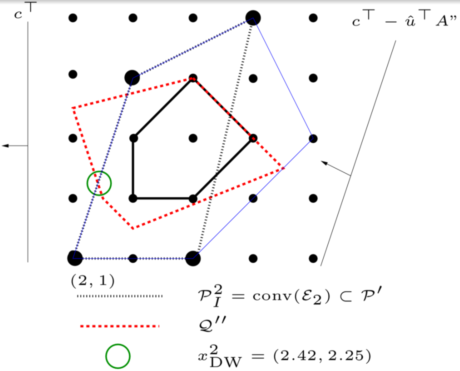
$$x^D \in \operatorname{argmax}\left\{c^\top x \mid A''x \leq b'', (c - u^*A'')^\top x \leq z_D - ub''\right\},$$

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

Connecting the Dots

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.



Future Work

Complexity

- There are further related problems and ways in which the current complexity framework can be extended/enhanced.
- The entire framework can be generalized and made much more abstract, but it's not clear how useful this would be.

Computation

- The hope is that some of these connections will have practical implications for solving both MILPs and inverse MILPs.
- With Matthias Walter, we are currently exploring practical algorithms for solving inverse problems.
- We have a working solver and a paper draft.

- Unfortunately, I have not found the time to write much of this down :(.
- Let me know if you're interested!

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-R3, Lehigh University, 2015. URL `http://coral.ie.lehigh.edu/~ted/files/papers/InverseMILP15.pdf`.

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

M. Güzelsoy and T.K. Ralphs. Duality for Mixed-Integer Linear Programs. *International Journal of Operations Research*, 4:118–137, 2007. URL `http://coral.ie.lehigh.edu/~ted/files/papers/MILPD06.pdf`.

References II

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.