Separation, Inverse Optimization, and Decomposition: Some Observations

Ted Ralphs
Joint work with:
Aykut Bulut

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

PKU Workshop on Optimization, Beijing, China, 11 August 2018
What Is This Talk About?

- Duality in integer programming (and more generally)
- Connecting some concepts.
  - Separation problem
  - Inverse optimization
  - Decomposition methods
  - Primal cutting plane algorithms for MILP

Ralphs, Bulut (COR@L Lab)
Separation, Inverse Optim., and Decomposition
We focus on the case of the *mixed integer linear optimization problem* (MILP), but many of the concepts are more general.

\[
z_{IP} = \max_{x \in \mathcal{S}} c^\top x, \quad \text{(MILP)}
\]

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.
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
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 Güzelsoy and Ralphs [2007])

\[
\min \{ F(b) \mid F(\beta) \geq \phi_D(\beta), \ \beta \in \mathbb{R}^m, F \in \Upsilon^m \} \quad \text{(D)}
\]

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

\[
\phi_D(\beta) = \max_{x \in S(\beta)} c^\top x \quad \text{(DVF)}
\]

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

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

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

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

\[
\begin{array}{l}
\min_{\lambda \in \mathbb{R}^E_+} \\ \left\{ 0^\top \lambda \mid E\lambda = x^*, \ 1^\top \lambda = 1 \right\}
\end{array}
\]  

(MEM)

where \( E \) is the set of extreme points of \( P \) and \( E \) is a matrix whose columns are the members of \( E \).

- When (MEM) is feasible, then we have a proof that \( x^* \in P \).
- 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 $(\pi, \pi_0)$, 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).

$$
\max \left\{ \pi x^* - \pi_0 \mid \pi^T x \leq \pi_0 \ \forall x \in \mathcal{E}, (\pi, \pi_0) \in \mathbb{R}^{n+1} \right\} \quad \text{(SEP)}
$$

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 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 $\min 1^\top \lambda$. We can interpret this essentially as how much we need to expand $\mathcal{P}$ in order to include $x^*$. If the result is more 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

\[
\mathcal{P}^* = \left\{ \pi \in \mathbb{R}^n \mid \pi^\top x \leq 1 \ \forall x \in \mathcal{P} \right\}
\]

(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
  \[
  \mathcal{P}^* = \left\{ \pi \in \mathbb{R}^n \mid \pi^\top x \leq 1 \ \forall x \in \mathcal{E}, \pi^\top r \leq 0 \ \forall r \in \mathcal{R} \right\}.
  \]

- 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
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)
Figure: Separating $x^*$ from $\mathcal{P}$ (Iteration 4)
Figure: Separating $x^*$ from $\mathcal{P}$ (Iteration 5)
What is an inverse problem?

Given a function, an inverse problem is that of determining \textit{input} that would produce a given \textit{output}.

- The input may be partially specified.
- We may want an answer as close as possible to a given \textit{target}.

- This is precisely the mathematical notion of the inverse of a function.
- A \textit{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 \textit{value function}.
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 $\phi_P$, defined as

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

(\text{PVF})

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

$$
\min \left\{ f(d) \mid d^\top x^0 = \phi_P(d) \right\},
$$

(\text{INV})

The classical objective function is taken to be $f(d) = \|c - d\|$, where $c \in \mathbb{R}^n$ is a given target.
The feasible set of the inverse problem is the set of objective vectors that make $x^0$ optimal.

This is precisely the dual of $\text{cone}(\mathcal{S} - \{x^0\})$, which is, roughly, a translation of the polyhedron described by the inequalities binding at $x^0$.

Figure: $\text{conv}(\mathcal{S})$ 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

\[ \mathcal{D} = \left\{ d \in \mathbb{R}^n \mid d^\top x \leq d^\top x^0 \ \forall x \in \mathcal{S} \right\} \quad \text{(IFS)} \]

- 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 corresponds to 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 relaxation* with respect to \( x^0 \).
Formulating the Inverse Problem

General Formulation

\[
\begin{align*}
\min & \quad f(d) \\
\text{s.t.} & \quad d^\top x \leq d^\top x^0 \quad \forall x \in \mathcal{E} \\
\end{align*}
\]

(INVMP)

- With \( f(d) = \|c - d\| \), this can be linearized for \( \ell_1 \) and \( \ell_\infty \) norms.
- The separation problem for the feasible region is again optimization over \( \text{conv}(\mathcal{S}) \).
It should be clear that inverse optimization and separation are very closely related.

First, note that the inequality

\[ \pi^\top x \leq \pi_0 \]  

(PI)

is valid for \( \mathcal{P} \) if and only if \( \pi_0 \geq \phi_P(\pi) \).

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 \( \phi^\pi \) is a **dual function** for (MILP) when the objective function is taken as \( \pi \).

The feasible set of (INV) can be seen as the set of all valid primal inequalities that are tight at \( x^0 \).
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 Padberg and Grötschel [1985] called the *primal separation problem* (see also Lodi and Letchford [2003]).

Their original idea was to separate $x^*$ with an inequality binding at the current incumbent.

Taking $x^0$ to be 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 $d^\top x^0 = 1$ (Note: For this normalization, 0 must be in the interior of $\text{conv}(\mathcal{S})$)

(INVMP) is then precisely the separation problem for the corner relaxation with respect to $x^0$ (alternatively, the conic hull of $\mathcal{S} - \{x^0\}$).
Roughly speaking, the dual of (INVMP) is the membership problem for cone($\mathcal{S} - \{x^0\}$).

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

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

With the normalization, this becomes

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

We can interpret the value of $\alpha$ as the amount by which we need to shift $x^*$ along the direction $x^0$ in order for it to be inside 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.
• We can use an algorithm almost identical to the one from earlier.
• We now generate inequalities valid for the corner relaxation associated with $x^0$. 

![Graph showing 2D optimization problem with a blue curve and a point marked with an 'x'.]
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)
Theorem 1  Bulut and Ralphs [2015] Inverse MILP optimization problem under $\ell_\infty/\ell_1$ 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**

\[ K(\gamma) = \{ d \in \mathbb{R}^n \mid \| c - d \| \leq \gamma \} \]

\[ X(\gamma) = \{ x \in S \mid \exists d \in K(\gamma) \text{ s.t. } d^\top (x^0 - x) > 0 \}, \]

\[ K^*(\gamma) = \{ x \in \mathbb{R}^n \mid d^\top (x^0 - x) \geq 0 \forall d \in K(\gamma) \}. \]

**Inverse MILP Decision Problem (INVD)**

**Inputs:** \( \gamma, c, x^0 \in S \) and MILP feasible set \( S \).

**Problem:** Decide whether \( K(\gamma) \cap D \) 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.
As usual, we divide the constraints into two sets.

\[
\begin{align*}
\text{max } & \quad c^\top x \\
\text{s.t. } & \quad A'x \leq b' \text{ (the “nice” constraints)} \\
& \quad A''x \leq b'' \text{ (the “complicating” constraints)} \\
& \quad x \in \mathbb{Z}^n
\end{align*}
\]

\[
\mathcal{P}' = \{ x \in \mathbb{R}^n \mid A'x \leq b' \}, \\
\mathcal{P}'' = \{ x \in \mathbb{R}^n \mid A''x \leq b'' \}, \\
\mathcal{P} = \mathcal{P}' \cap \mathcal{P}'', \\
\mathcal{S} = \mathcal{P} \cap \mathbb{Z}^n \text{, and} \\
\mathcal{S}_R = \mathcal{P}' \cap \mathbb{Z}^n.
\]
Reformulation

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

\[
\begin{align*}
\text{max} & \quad c^\top x \\
\text{s.t.} & \quad \sum_{s \in \mathcal{E}} \lambda_s s = x \\
& \quad A''x \leq b'' \\
& \quad \sum_{s \in \mathcal{E}} \lambda_s = 1 \\
& \quad \lambda \in \mathbb{R}_+^\mathcal{E} \\
& \quad x \in \mathbb{Z}^n
\end{align*}
\]

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

- 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 aforementioned relaxation, we obtain a formulation for the so-called decomposition bound.

\[
\begin{align*}
    z_{IP} &= \max_{x \in \mathbb{Z}^n} \left\{ c^\top x \mid A'x \leq b', A''x \leq b'' \right\} \\
    z_{LP} &= \max_{x \in \mathbb{R}^n} \left\{ c^\top x \mid A'x \leq b', A''x \leq b'' \right\} \\
    z_D &= \max_{x \in \text{conv}(S_R)} \left\{ c^\top x \mid A''x \leq b'' \right\}
\end{align*}
\]

\[z_{IP} \leq z_D \leq z_{LP}\]

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

\[ \begin{align*}
\text{max} & \quad -x_1 \\
-x_1 - x_2 & \geq -8, \quad (8) \\
-0.4x_1 + x_2 & \geq 0.3, \quad (9) \\
x_1 + x_2 & \geq 4.5, \quad (10) \\
3x_1 + x_2 & \geq 9.5, \quad (11) \\
0.25x_1 - x_2 & \geq -3, \quad (12) \\
7x_1 - x_2 & \geq 13, \quad (13) \\
x_2 & \geq 1, \quad (14) \\
-x_1 + x_2 & \geq -3, \quad (15) \\
-4x_1 - x_2 & \geq -27, \quad (16) \\
-x_2 & \geq -5, \quad (17) \\
0.2x_1 - x_2 & \geq -4, \quad (18) \\
x & \in \mathbb{Z}''.
\end{align*} \]
\[ Q' = \{ x \in \mathbb{R}^2 \mid x \text{ satisfies (8) -- (12)} \}, \]
\[ Q'' = \{ x \in \mathbb{R}^2 \mid x \text{ satisfies (13) -- (18)} \}, \]
\[ Q = Q' \cap Q'', \]
\[ S = Q \cap \mathbb{Z}^n, \] and
\[ S_R = Q' \cap \mathbb{Z}^n. \]
\[
\text{conv}(S) = \text{conv}\{x \in \mathbb{Z}^n \mid A' x \geq b', A'' x \geq b''\}
\]
\[
\text{conv}(S_R) = \text{conv}\{x \in \mathbb{Z}^n \mid A' x \geq b'\}
\]
\[
Q' = \{x \in \mathbb{R}^n \mid A' x \geq b'\}
\]
\[
Q'' = \{x \in \mathbb{R}^n \mid A'' x \geq b''\}
\]
Geometry of Dantzig-Wolfe Decomposition

\[ c^T \]

\[ c^T - \hat{u}^T A'' \]

\[
\mathcal{P}_I^0 = \text{conv}(\mathcal{E}_0) \subset \mathcal{P}'
\]

\[ Q'' \]

\[ x_{DW}^0 = (4.25, 2) \]

\[ \bar{s} = (2, 1) \]
Geometry of Dantzig-Wolfe Decomposition

\[ c^T = \hat{u}^T A'' \]

\[ (2, 1) \]

\[ \mathcal{P}^1_I = \text{conv}(\mathcal{E}_1) \subset \mathcal{P}' \]

\[ Q'' \]

\[ x_{DW}^1 = (2.64, 1.86) \]

\[ \tilde{s} = (3, 4) \]
Geometry of Dantzig-Wolfe Decomposition

\[ \mathcal{P}_{I}^{2} = \text{conv}(\mathcal{E}_{2}) \subseteq \mathcal{P}' \]

\[ Q'' \]

\[ x_{\text{DW}}^{2} = (2.42, 2.25) \]
Lagrange Cuts

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

\[ (c - uA'')^\top x \leq LR(u) - ub'' \]  

is valid for $\mathcal{P}$.

- If we take $u^*$ to be the optimal solution to the Lagrangian dual, then this inequality reduces to

\[ (c - u^*A'')^\top x \leq z_D - ub'' \]  

(OLC)

- If we now take

\[ x^D \in \arg\max \left\{ c^\top x \mid A''x \leq b'', (c - u^*A'')^\top x \leq z_D - ub'' \right\}, \]

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

Results

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

$P_I^2 = \text{conv}(\mathcal{E}_2) \subset P'$

$Q''$

$x_{DW}^2 = (2.42, 2.25)$
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 may lead to new algorithms for inverse problems.
- Much of that is discussed here can be further generalized to general computation via Turing machines (useful?).
Thank You!

[Photo of a person standing on the Great Wall of China]


