Generalized Duality and Value Functions (Three Points of View)

Ted Ralphs[1]

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

Dagstuhl Workshop, 3 November 2022

ISE Industrial and Systems Engineering

COR@L
COMPUTATIONAL OPTIMIZATION RESEARCH AT LEHIGH

Agenda

1. A quick introduction to the challenges of multilevel/multistage optimization.
2. The remainder of the talk will be to support the “bold claim” that there *is* a rich duality theory that applies far beyond the continuous, convex case.
3. Three views of generalized duality.
    1. “Traditional” (value functions)
    2. Complexity-theoretic (automated theorem-proving, etc.)
    3. Projection-based (Benders)

General Bold Claim

Most (all?) (duality-based) techniques exploited for solving “tractable” classes of problems can be (and are being) generalized to classes that are “intractable.”

General Setting: Two-Stage Mixed Integer Optimization

- We have the following general formulation:

2SMILP

$$z_{\text{2SMILP}} = \min_{x \in \mathcal{P}_1} \left\{ c^\top x + \Xi(x) \right\}, \tag{2SMILP}$$

where

$$\mathcal{P}_1 = \left\{ x \in X \mid A^1 x = b^1 \right\}$$

is the *first-stage feasible region* with $X = \mathbb{Z}_+^{r_1} \times \mathbb{R}_+^{n_1 - r_1}$, $A^1 \in \mathbb{Q}^{m_1 \times n_1}$, and $b^1 \in \mathbb{R}^{m_1}$.

- Ξ is the *risk function* that represents the impact of future "uncertainty."
- The "uncertainty" can arise, e.g., due to stochasticity or due to the fact that Ξ represents the reaction of a competitor.
- Regardless of the source, what matters algorithmically is the structure and properties of Ξ, including how easily it can be evaluated.

Bilevel (Integer) Linear Optimization

In the case of general bilevel optimization, we have

Bilevel Risk Function (Optimistic)

$$\Xi(x) = \min \left\{ d^1 y \mid y \in \operatorname{argmin}\{d^2 \hat{y} \mid \hat{y} \in \mathcal{P}_2(b^2 - A^2 x) \cap Y\} \right\}$$

where $A^2 \in \mathbb{Q}^{m_2 \times n_1}$, and $b^2 \in \mathbb{R}^{m_2}$, $\mathcal{P}_2(\beta) = \left\{ y \in \mathbb{R}_+ \mid G^2 y \geq \beta \right\}$, and $Y = \mathbb{Z}^{p_2} \times \mathbb{R}^{n_2 - p_2}$.

Alternatively, the more familiar and equivalent form of the problem is

Mixed Integer Bilevel Linear Optimization Problem (MIBLP)

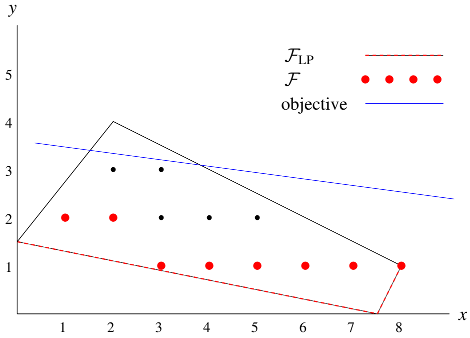
$$\min \left\{ cx + d^1 y \mid x \in \mathcal{P}_1 \cap X, y \in \operatorname{argmin}\{d^2 \hat{y} \mid \hat{y} \in \mathcal{P}_2(b^2 - A^2 x) \cap Y\} \right\} \tag{MIBLP}$$

Challenges

- Many “intuitive” statements are not true.
- We apparently cannot use the “dualization” trick to reformulate.
- Most interesting problems have decision versions that are complete for Σ_2^{P}, so there is not much hope for “tractability.”
- However, if it’s not intractable, where’s the fun?

Example

Example 1 Moore and Bard [1990]



$$
\begin{aligned}
\min_{x \in \mathbb{Z}_+} \quad & -x - 10y \\
\text{s.t.} \quad & y \in \operatorname{argmin} \{ y : \\
& \quad -25x + 20y \leq 30 \\
& \quad x + 2y \leq 10 \\
& \quad 2x - y \leq 15 \\
& \quad 2x + 10y \geq 15 \\
& \quad y \in \mathbb{Z}_+ \}
\end{aligned}
$$

"It's All Just Duality"

Quote from the Internet: *Duality* is a woefully overloaded mathematical term for a relation that groups elements of a set into "dual" pairs.
Bold claim: Many (most?) duality concepts can be seen as roughly "isomorphic".

Duality Concepts

- Sets: Projection/complement, intersection/union
- Conic duality: Cones and their duals, convexity/nonconvexity
- Farkas duality: Theorems of the alternative, empty/non-empty
- Complexity: Languages and their complements (NP vs. co-NP)
- Quantifier duality: Existential versus universal quantification
- De Morgan duality: Conjunction versus disjunction
- Weyl-Minkowski duality: V representation versus H representation
- Polarity: Optimization versus separation
- Dual problems: Primal and dual problems in optimization
- Inverses: Functions and inverses, inverse optimization inverses

Quick Review of "Classical" Duality

- Thus, we consider the problem

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

where

$$\mathcal{S} = \{x \in \mathbb{Z}_+^r \times \mathbb{R}_+^{n-r} \mid Ax = b\}, \tag{1}$$

with $c \in \mathbb{Q}^n, A \in \mathbb{Q}^{m \times n}$, and $b \in \mathbb{Q}^m$.

- The *value function* associated with the base instance (MILP) is

$$\phi(\beta) = \min_{x \in \mathcal{S}(\beta)} c^\top x \tag{VF}$$

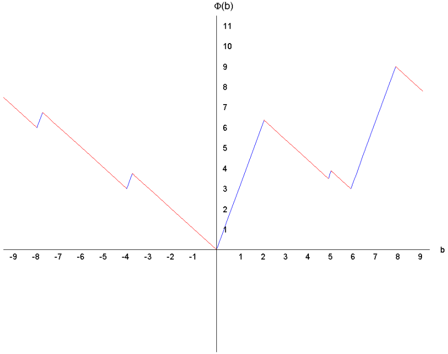
for $\beta \in \mathbb{R}^m$, where $\mathcal{S}(\beta) = \{x \in \mathbb{Z}_+^r \times \mathbb{R}_+^{n-r} \mid Ax = \beta\}$.

- Again, we let $\phi(\beta) = \infty$ if $\beta \in \Omega = \{\beta \in \mathbb{R}^m \mid \mathcal{S}(\beta) = \emptyset\}$.

Example

Example 2

$$\phi(\beta) = \min\ 3x_1 + \frac{7}{2}x_2 + 3x_3 + 6x_4 + 7x_5 + 5x_6$$
$$\text{s.t. } 6x_1 + 5x_2 - 4x_3 + 2x_4 - 7x_5 + x_6 = \beta$$
$$x_1, x_2, x_3 \in \mathbb{Z}_+,\ x_4, x_5, x_6 \in \mathbb{R}_+$$



The structure of this function is inherited from two related functions.

Continuous and Integer Restrictions

Consider the general form of the value function

$$\begin{aligned}\phi(\beta) = \min\ & c_I^\top x_I + c_C^\top x_C \\ \text{s.t. } & A_I x_I + A_C x_C = \beta, \\ & x \in \mathbb{Z}_+^{r_2} \times \mathbb{R}_+^{n_2 - r_2}\end{aligned} \tag{VF}$$

The structure is inherited from that of the *continuous restriction*:

$$\begin{aligned}\phi_C(\beta) = \min\ & c_C^\top x_C \\ \text{s.t. } & A_C x_C = \beta, \\ & x_C \in \mathbb{R}_+^{n_2 - r_2}\end{aligned} \tag{CR}$$

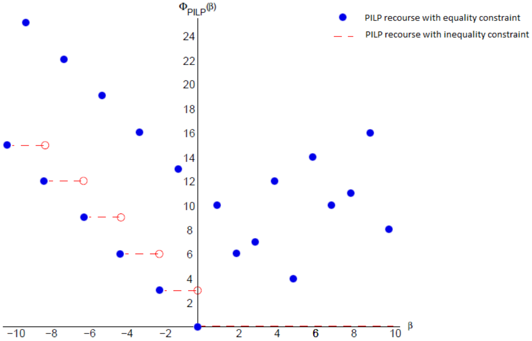
for $C = \{r_2 + 1, \ldots, n_2\}$ and the similarly defined *integer restriction*:

$$\begin{aligned}\phi_I(\beta) = \min\ & c_I^\top x_I \\ \text{s.t. } & A_I x_I = \beta \\ & x_I \in \mathbb{Z}_+^{r_2}\end{aligned} \tag{IR}$$

for $I = \{1, \ldots, r_2\}$.

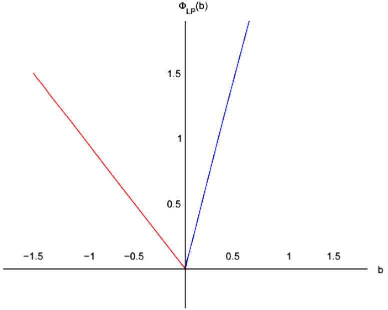
Example 3

$$\phi(\beta) = \min\ 3x_1 + \frac{7}{2}x_2 + 3x_3 + 6x_4 + 7x_5 + 5x_6$$
$$\text{s.t. } 6x_1 + 5x_2 - 4x_3 + 2x_4 - 7x_5 + x_6 = \beta$$
$$x_1, x_2, x_3, x_4, x_5, x_6 \in \mathbb{Z}_+$$



Example 4

$$\phi_C(\beta) = \min 6y_1 + 7y_2 + 5y_3$$
$$\text{s.t. } 2y_1 - 7y_2 + y_3 = \beta$$
$$y_1, y_2, y_3 \in \mathbb{R}_+$$



Discrete Representation of the Value Function

For $\beta \in \mathbb{R}^{m_2}$, we have that

$$\begin{aligned} \phi(\beta) = \min\ & c_I^\top x_I + \phi_C(\beta - A_I x_I) \\ \text{s.t. } & x_I \in \mathbb{Z}_+^{r_2} \end{aligned} \tag{2}$$

- From this we see that the value function is comprised of the minimum of a set of translations of ϕ_C.
- The set of shifts, along with ϕ_C describe the value function exactly.
- For $\hat{x}_I \in \mathbb{Z}_+^{r_2}$, let

$$\phi_C(\beta, \hat{x}_I) = c_I^\top \hat{x}_I + \phi_C(\beta - A_I \hat{x}_I)\ \forall \beta \in \mathbb{R}^{m_2}. \tag{3}$$

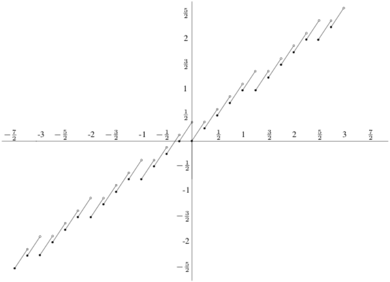
- Then we have that $\phi(\beta) = \min_{x_I \in \mathbb{Z}_+^{r_2}} \phi_C(\beta, \hat{x}_I)$.

General Properties of the MILP Value Function

The value function is subadditive, non-convex, lower semi-continuous, and piecewise polyhedral.

Example 5

$$\phi(\beta) = \min x_1 - \frac{3}{4}x_2 + \frac{3}{4}x_3$$
$$\text{s.t. } \frac{5}{4}x_1 - x_2 + \frac{1}{2}x_3 = \beta \qquad \text{(Ex2.MILP)}$$
$$x_1, x_2 \in \mathbb{Z}_+,\ x_3 \in \mathbb{R}_+$$



Generalized Dual

- A *dual function* $F : \mathbb{R}^m \to \mathbb{R}$ is one that satisfies $F(\beta) \leq \phi(\beta)$ for all $\beta \in \mathbb{R}^m$.
- How to select such a function?
- We may choose one that is easy to construct/evaluate or for which $F(b) \approx \phi(b)$.
- This results in the following generalized *dual* associated with the base instance (MILP).

$$\max \{F(b) : F(\beta) \leq \phi(\beta),\ \beta \in \mathbb{R}^m, F \in \Upsilon^m\} \tag{D}$$

where $\Upsilon^m \subseteq \{f \mid f : \mathbb{R}^m \to \mathbb{R}\}$

- We call F^* *strong* for this instance if F^* is a *feasible* dual function and $F^*(b) = \phi(b)$.
- Under mild conditions, this is a strong dual when $\Upsilon^m \equiv \{f \mid f : \mathbb{R}^m \to \mathbb{R}\}$.
- The value function itself is typically a strong dual function.

The Subadditive Dual

- Let a function F be defined over a domain V. Then F is subadditive if $F(v_1) + F(v_2) \geq F(v_1 + v_2) \forall v_1, v_2, v_1 + v_2 \in V$.
- Note that the value function z is subadditive over Ω.
- If $\Upsilon^m \equiv \Gamma^m \equiv \{F \text{ is subadditive} \mid F : \mathbb{R}^m \rightarrow \mathbb{R}, F(0) = 0\}$, we can rewrite (D) as the *subadditive dual*

$$
\begin{aligned}
\max \quad & F(b) \\
& F(a^j) \leq c_j \quad j = 1, ..., r, \\
& \bar{F}(a^j) \leq c_j \quad j = r+1, ..., n, \text{ and} \\
& F \in \Gamma^m,
\end{aligned}
$$

where the function $\bar{F}$ is defined by

$$
\bar{F}(\beta) = \limsup_{\delta \rightarrow 0^+} \frac{F(\delta\beta)}{\delta} \;\; \forall \beta \in \mathbb{R}^m.
$$

- Here, $\bar{F}$ is the *upper β-directional derivative* of F at zero.

Discrete Farkas Lemma

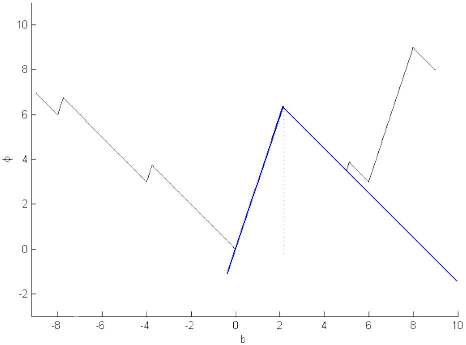
For the primal problem, exactly one of the following holds:

1. $\mathcal{S} \neq \emptyset$
2. There is an $F \in \Gamma^m$ with $F(a^j) \leq 0, j = 1, ..., n$, and $F(b) > 0$.

Proof. Let $c = 0$ and apply strong duality to the subadditive dual.

Generating Dual Functions

- The easiest way to get a dual function is to take the value function of a relaxation.
- The value function of the LP relaxation is the convex envelope of the value function of an MILP.
- From this we can visually see the integrality "gap."
- To get a strong dual function, we need something piecewise linear.



Importance

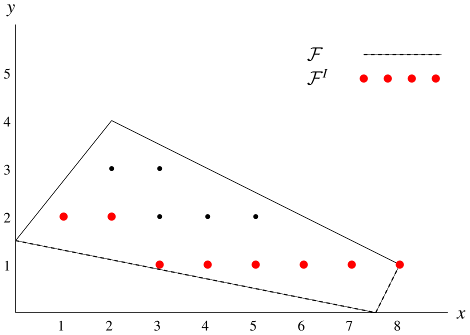
- Provides optimality conditions for MILPs that generalize those from the LP case.

  Theorem 1 *[Optimality conditions for* (MILP)*] If $x^* \in \mathcal{S}$, F^* is feasible for* (D)*, and $c^\top x^* = F^*(b)$, then x^* is an optimal solution to* (MILP) *and F^* is an optimal solution to* (D).

- These are the optimality conditions achieved in the branch-and-bound algorithm for MILP that prove the optimality of the primal solution.
- The branch-and-bound tree encodes a solution to the dual.
- Provides a way of "dualizing" the second-stage and later problems.

Value Function Reformulation (⇒ Generalized Benders)

More generally, we can reformulate (MIBLP) as



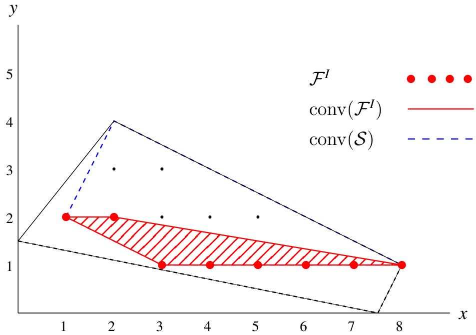
$$
\begin{aligned}
\min \quad & c^1 x + d^1 y \\
\text{subject to} \quad & A^1 x \le b^1 \\
& G^2 y \ge b^2 - A^2 x \\
& d^2 y \le \phi(b^2 - A^2 x) \\
& x \in X, y \in Y,
\end{aligned}
$$

where ϕ is the value function of the second-stage problem.

- This is, in principle, a standard mathematical program.
- Note that the second-stage variables need to appear in the formulation in order to enforce feasibility.
- These can be projected out to obtain a “pure Benders” reformulation.

Polyhedral Reformulation (⇒ Cutting Plane Method)

Convexification considers the following conceptual reformulation.



$$\begin{aligned} \min \quad & c^1x + d^1y \\ \text{s.t.} \quad & (x, y) \in \text{conv}(\mathcal{F}^I) \end{aligned}$$

where $\mathcal{F}^I = \{(x, y) \mid x \in \mathcal{P}_1 \cap X, y \in \text{argmin}\{d^2y \mid y \in \mathcal{P}_2(x) \cap Y\}\}$

- To get bounds, we optimize over a relaxed feasible region.
- We iteratively approximate the true feasible region with linear inequalities, just as in standard algorithms for solving MILPs.

Disjunctive Reformulation (⇒ Branch and Cut)

Theorem 2 *A set $\mathcal{F} \subseteq \mathbb{R}^n$ is MILP representable if and only if there exist rational polytopes $\mathcal{P}_1, \ldots, \mathcal{P}_k$ and vectors $r^1, \ldots, r^t \in \mathbb{Z}^n$ such that*

$$\mathcal{F} = \bigcup_{i=1}^{k} (\mathcal{P}_i + \text{intcone}\{r^1, \ldots, r^t\}) \tag{4}$$

Theorem 3 *Basu et al. [2018] $\mathcal{F}$ is MIBLP-representable if and only if*

$$\text{cl}(\mathcal{F}) = \bigcup_{i=1}^{k} \mathcal{S}_i, \tag{5}$$

where $\mathcal{S}_i$ are MILP-representable.

A Different View of Duality

- Now we switch to an entirely different point of view that ends in the same place.
- The goal is to explain some of the connections discussed earlier.
- We also aim to connect a bit more clearly with complexity-theoretic notions.

Decision Problems and Complexity

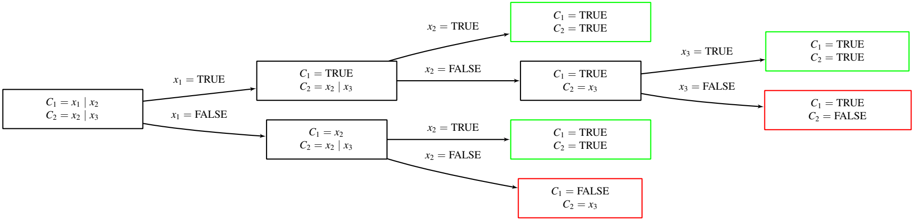
- We address problems in the *polynomial hierarchy* (PH).
- Technically, this classification applies to problems for which the result of a computation is "YES" or "NO."
- It is useful, however, to interpret such a problem as that of trying to prove a theorem, which must be either "TRUE" or "FALSE".
- By viewing the proof as part of the output, it is easier to see that this class of problems is in fact very rich.
- The notion of a proof is fundamental to how problems are classified in the PH—higher complexity means longer proofs are expected.

A Canonical Example: Satisfiability Game

- A canonical extensive-form game that illustrates many of the basic principles is the *k-player satisfiability game*.
  - k players determine the value of a set of Boolean variables with each in control of a specific subset.
  - In round i, player i determines the values of her variables.
  - Each player tries to choose values that force a certain end result, given that subsequent players may be trying to achieve the opposite result.
- Examples
  - k = 1: SAT
  - k = 2: The first player tries to choose values such that any choice by the second player will result in satisfaction.
  - k = 3: The first player tries to choose values such that the second player cannot choose values that will leave the third player without the ability to find satisfying values.
- Note that the odd players and the even players are essentially "working together" and the same game can be described with only two players.

Constructing a Proof

- This diagram illustrates the search for a proof as a tree.
- The nodes in green represent settings of the truth values that satisfy all the given clauses; red represents non-satisfying truth values.
  - With one player, the proof is any path to one of the green nodes.
  - With two players, the proof is a subtree in which there are no red nodes.
- The latter requires knowledge of *all* leaf nodes (important!).



More Formally

- More formally, we are given a Boolean formula with variables partitioned into k sets $X_1, \ldots, X_k$.
- For k odd, the SAT game can be formulated as

$$\exists X_1 \forall X_2 \exists X_3 \ldots ?X_k$$

- for even k, we have

$$\forall X_1 \exists X_2 \forall X_3 \ldots ?X_k$$

- A more general form of this problem, known as the *quantified Boolean formula problem* (QBF) allows an arbitrary sequence of quantifiers.

Theorems About Sets

- In optimization (and even more generally), the "theorems" we wish to prove or disprove can be formulated as statements about sets.
- Let $\mathcal{S} = \{x \in \mathbb{Q}^n \mid P(x)\}$, where $P : \mathbb{Q}^n \to \{TRUE, FALSE\}$.
- The simplest question we can ask is whether $\mathcal{S}$ is non-empty

$$\mathcal{S} \stackrel{?}{=} \emptyset.$$

- Given function f and constant K, the related question of

$$\mathcal{S}(f, K) := \{x \in \mathcal{S} \mid f(x) < K\} \stackrel{?}{=} \emptyset$$

is the *decision version* of the optimization problem

$$\min_{x \in \mathcal{S}} f(x) \tag{OPT}$$

Constructing Proofs

- What do proofs of theorems about sets look like?
  - Certifying $\mathcal{S} \neq \emptyset$ is easy: produce a point in the set.
  - Certifying $\mathcal{S} = \emptyset$ is more difficult in general.
- The difficulty of proving a set is empty is most easily seen by re-stating the theorems we are trying to prove/disprove, as follows.

$$\mathcal{S} \neq \emptyset \Leftrightarrow \exists x \in \mathcal{S}$$
$$\mathcal{S} = \emptyset \Leftrightarrow \forall x \in \mathbb{Q}^n \; x \notin \mathcal{S} \Leftrightarrow \forall x \in \mathbb{Q}^n \; x \in \bar{\mathcal{S}}$$

- The statement that a set is non-empty is *existentially quantified*, whereas the statement that a set is empty is *universally quantified*.
- Universally quantified statements are intuitively more difficult to prove than existentially quantified ones.

De Morgan Duality

- There is a duality between existential and universal quantifiers that can be seen as one of a number of generalized forms of De Morgan's Laws.

DeMorgan's Laws

- The complement of the union is the intersection of the complements.
- The complement of the intersection is the union of the complements.

- These laws can be used to equivalently formulate logical statements in different dual forms to aid in constructing proofs.

$$P(x)\ \forall x \in \mathcal{S} \Leftrightarrow \neg[\exists x \in \mathcal{S}\ \neg P(x)] \Leftrightarrow \neg \bigvee_{x \in \mathcal{S}} \neg P(x) \Leftrightarrow \bigwedge_{x \in \mathcal{S}} P(x)$$

$$\exists x \in \mathcal{S} : P(x) \Leftrightarrow \neg[\forall x \in \mathcal{S}\ P(x)] \Leftrightarrow \neg \bigwedge_{x \in \mathcal{S}} \neg P(x) \Leftrightarrow \bigvee_{x \in \mathcal{S}} P(x)$$

- Note also the duality between conjunction and disjunction.

Convexity and Nonconvexity

- Related dualities exist between between conjunction and disjunction, which are reflected in the way convex and nonconvex sets are described.
- Convex sets are described by conjunctive logic: the *intersection* of convex sets is convex.

  - There is always a short proof that a point is *not* in a convex set (separating hyperplane).
  - When a convex set is empty, the Farkas Lemma provides a short proof.

- Nonconvex sets are described using disjunctive logic: the *union* of convex sets is nonconvex (in general).

  - In general, there is no short proof that a point is not in a nonconvex set.
  - Similarly, we can't expect short proofs of emptiness for disjunctive unions of convex sets.

Short Proofs of Emptiness

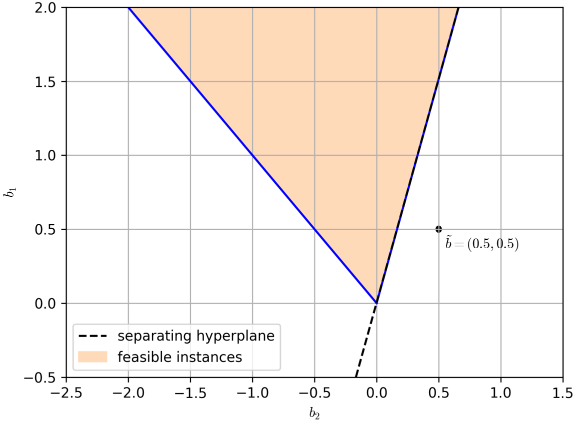
- In the case of convex sets, we can use a duality argument to obtain short proofs of emptiness.
- Consider the case of a polyhedron.

$$\mathcal{P} = \{x \in \mathbb{Q}^n_+ \mid Ax = \tilde{b}\}$$

- **Farkas Lemma**: $\mathcal{P} = \emptyset \Leftrightarrow \exists u \in \mathbb{Q}^m \; A^\top u \leq 0, \tilde{b}^\top u > 0$
- Equivalently, $S = \emptyset$ if and only if we can separate $\tilde{b}$ from the convex cone $\mathcal{C} = \{b \in \mathbb{Q}^m \mid \exists x \in \mathbb{Q}^n_+, Ax = b\} = \{b \in \mathbb{Q}^m : b^\top u \geq 0 \; \forall u \in \mathcal{C}^*\}$, where $\mathcal{C}^* = \{u \in \mathbb{Q}^m : A^\top u \geq 0\}$ (the *dual* of $\mathcal{C}$).
- One way to interpret this procedure is as follows.
  - We first lift the problem into a higher dimensional space by making b a vector of variables to obtain a related *non-empty* set.
  - Then project out the original variables and apply the separating hyperplane theorem.

Example

$$6y_1 + 7y_2 + 5y_3 = 1/2$$
$$2y_1 - 7y_2 + y_3 = 1/2$$
$$y_1, y_2, y_3 \in \mathbb{R}_+$$



Connection to Complexity

- On one hand, this is a "trick" for recasting a question of emptiness as one of non-emptiness (universal $\rightarrow$ existential), but there's a bigger picture.
- We are embedding a single theorem into a *parametric class* containing both TRUE and FALSE theorems.
- The questions we are asking is being re-cast as a question of where this theorem lies relative to the set of all TRUE theorems (in the class).
- To prove the theorem is FALSE, we separate it from the set of theorems that are TRUE—this is a "dual" proof based on a separation argument.
- In the terminology of complexity theory, the set of true theorems is called a *language*.

Proofs of Optimality

- The problem (OPT) is *not* a decision problem as stated.
- We can nevertheless build a proof that the optimal solution value is K using proofs for two related theorems.

1. $\exists x \in \mathcal{S} : f(x) = K$
2. $\nexists x \in \mathcal{S} : f(x) < K \Leftrightarrow \forall x \in \mathcal{S} : f(x) \geq K$

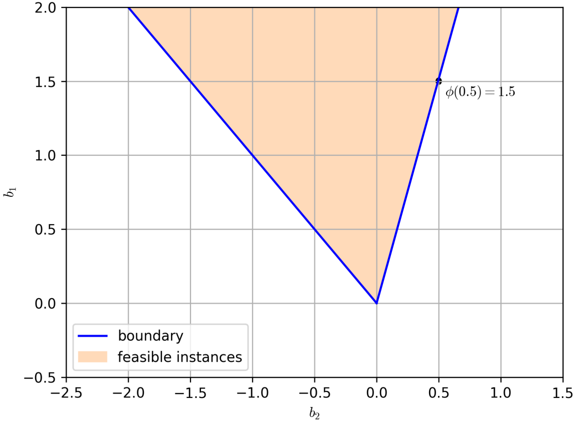
- The fact that one of these statements is universally quantified is the reason why short proofs of optimality cannot be expected in general.

Short Proofs of Optimality

- We consider the case of a linear optimization problem (LP).
- We can get an LP as follows.
  - Convert the first row of A from a constraint to the objective function.
  - Let $N = \{2, \ldots, m\}$ and $\tilde{b}_N \in \mathbb{Q}^{m-1}$ be all but the first element of $\tilde{b}$.
- The problem of finding the optimal value can then be recast as $b^* = \min\{b_1 \in \mathbb{Q} \mid b \in \mathcal{C}\}$.
- To prove optimality, we need to show that $(b^*, \tilde{b}_N)$ is not only a member of $\mathcal{C}$, but on its *boundary*.
- The proof is only slightly modified: $\exists u \in \mathbb{Q}^m, A^\top u \geq 0, (b^*, \tilde{b}_N)^\top u = 0, u_1 > 0$.
  - Assume u is scaled so that $u_1 = -1$.
  - Then we have $A_N^\top u_N \leq A_1^\top, (\tilde{b}_N)^\top u_N = b^*$.
  - This is equivalent to the usual LP optimality conditions, but also proves that $(b^*, \tilde{b}_N)$ is on the boundary of $\mathcal{C}$.
- The vector u is a solution to the usual LP dual problem.

Example

$$\min 6y_1 + 7y_2 + 5y_3$$
$$\text{s.t. } 2y_1 - 7y_2 + y_3 = 1/2$$
$$y_1, y_2, y_3 \in \mathbb{R}_+$$



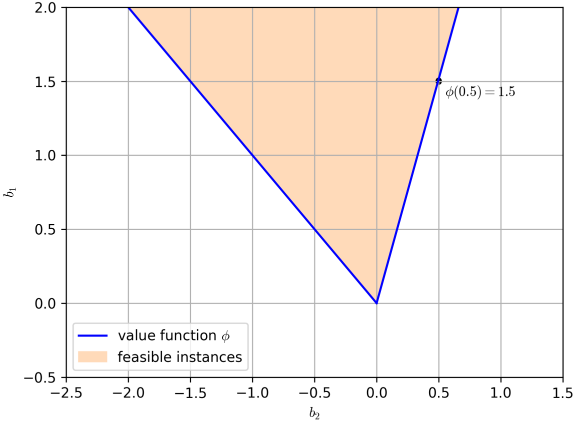
Interpreting

The boundary of the cone $\mathcal{C}$ describes a *parametric collection of proofs* and has several nice interpretations.

- Most importantly, $\mathcal{C}$ is the epigraph of the value function and the boundary is its graph.
- The solution to the LP dual problem is a (sub)gradient of this function.
- The non-increasing part of this function is also the *Pareto frontier* when the constraints are interpreted as multiple objectives (roughly speaking).
- This last result has some important implications.

Example

$$\begin{aligned} \min\ & 6y_1 + 7y_2 + 5y_3 \\ \text{s.t. } & 2y_1 - 7y_2 + y_3 = 1/2 \\ & y_1, y_2, y_3 \in \mathbb{R}_+ \end{aligned}$$

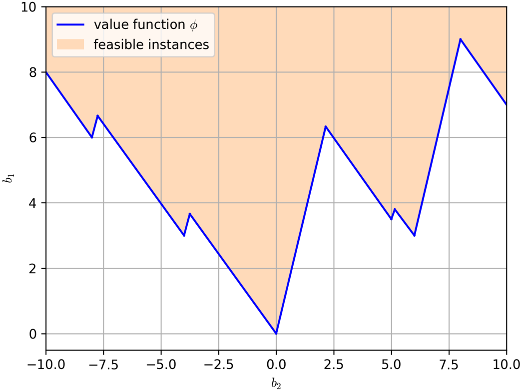


MILP Setting

- The very same logic extends easily to the MILP case.

Example

$$\phi(\beta) = \min 3x_1 + \frac{7}{2}x_2 + 3x_3 + 6x_4 + 7x_5 + 5x_6$$
$$\text{s.t. } 6x_1 + 5x_2 - 4x_3 + 2x_4 - 7x_5 + x_6 = \beta$$
$$x_1, x_2, x_3 \in \mathbb{Z}_+,\ x_4, x_5, x_6 \in \mathbb{R}_+$$



Dual Functions from Branch-and-Bound

Let T be set of the terminating nodes of the tree. Then in a terminating node $t \in T$ we solve:

$$\begin{aligned} \phi^t(\beta) = \min\ & c^\top x \\ \text{s.t. } & Ax = \beta, \\ & l^t \le x \le u^t, x \ge 0 \end{aligned} \tag{BB.VF}$$

By LP duality, we then have that:

$$\begin{aligned} \phi^t(\beta) = \max\ & \pi^t \beta + \underline{\pi}^t l^t + \bar{\pi}^t u^t \\ \text{s.t. } & \pi^t A + \underline{\pi}^t + \bar{\pi}^t \le c^\top \\ & \underline{\pi} \ge 0, \bar{\pi} \le 0 \end{aligned} \tag{BB.LP.D}$$

Finally, we obtain the following dual function, which is strong at b.

$$\underline{\phi}^T_{\text{LP}}(\beta) = \min_{t \in T} \underline{\phi}^t_{\text{LP}}(\beta) = \min_{t \in T} \{\hat{\pi}^t \beta + \hat{\underline{\pi}}^t l^t + \hat{\bar{\pi}}^t u^t\} \tag{BB.D}$$

where $(\hat{\pi}^t, \hat{\underline{\pi}}^t, \hat{\bar{\pi}}^t)$ is an optimal solution to the dual (BB.LP.D) at node t. Since $\underline{\phi}^T_{\text{LP}}(\beta) = \phi(b)$, this proves optimality of the final incumbent.

Example: Dual Function from Branch and Bound

- Recall the following value function associated with an MILP from earlier.

$$\begin{aligned} \phi(\beta) = \min\ & 6x_1 + 4x_2 + 3x_3 + 4x_4 + 5x_5 + 7x_6 \\ \text{s.t. } & 2x_1 + 5x_2 - 2x_3 - 2x_4 + 5x_5 + 5x_6 = \beta \\ & x_1, x_2, x_3 \in \mathbb{Z}_+, x_4, x_5, x_6 \in \mathbb{R}_+. \end{aligned} \tag{6}$$

- Suppose we evaluate $\phi(5.5)$ by solving the instance with fixed right-hand side by LP-based branch-and-bound.
- Solving the root LP relaxation, we obtain a solution in which $x_2 = 1.1$ and the optimal dual multiplier for the single constraint is $c_2/a_2 = 4/5 = 0.8$.
- We therefore branch on variable x_2 and obtain two subproblems, whose LP relaxations have the variable bounds $x_2 \leq 1$ and $x_2 \geq 2$, respectively.
- The problem is solved after this single branching, since $c_6/a_6 < c_1/a_1$ so that $x_1 = x_3 = 0$ in any optimal solution when $\beta > 0$.

Example: Dual Function from Branch and Bound

- To see how the branch-and-bound tree yields a dual function in this particular case, we have the following dual solutions.

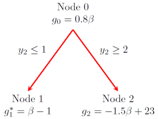
t	π^t	$\underline{\pi}^t$						$\bar{\pi}^t$					
0	0.8	4.4	0.0	4.6	5.6	1.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0
1	1.0	4.0	0.0	5.0	6.0	0.0	2.0	0.0	−1.0	0.0	0.0	0.0	0.0
2	−1.5	9.0	11.5	0.0	1.0	12.5	14.5	0.0	0.0	0.0	0.0	0.0	0.0

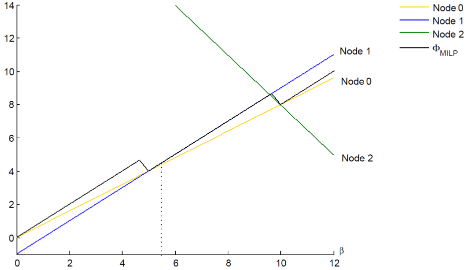
- Note that we have added the bound constraints explicitly and the upper bounds on all variables are taken to be a "big-M" value.
- Then, the following are the nodal dual functions.

$$\underline{\phi}^0_{\mathrm{LP}}(\beta) = 0.8\beta$$
$$\underline{\phi}^1_{\mathrm{LP}}(\beta) = \beta - 1$$
$$\underline{\phi}^2_{\mathrm{LP}}(\beta) = -1.5\beta + 23$$

- The initial (global) dual function in the root node is $\underline{\phi}^{\mathcal{T}_0} = \underline{\phi}^0_{\mathrm{LP}}$.
- After branching, the (global) dual function is $\underline{\phi}^{\mathcal{T}_1} = \min\{\underline{\phi}^1_{\mathrm{LP}}, \underline{\phi}^2_{\mathrm{LP}}\}$.

Example: Visualizing the Dual Function





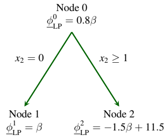
Strengthening the Dual Function

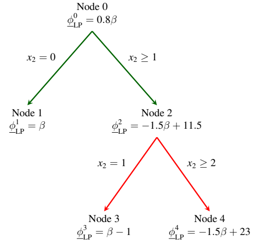
- The dual function can be strengthened by noting that the dual feasible region is the same for all nodes.
- We can therefore approximate the nodal value function by taking a max over all known dual solutions.
- Then we obtain

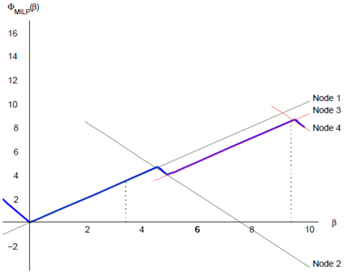
$$\min\{\max\{0.8\beta, \beta - 1, -1.5\beta\}, \max\{0.8\beta, \beta, -1.5\beta + 23\}\}$$

- By evaluating ϕ at a different right-hand side, but using the same tree as a starting point, we can begin to approximate the full value function.
- On the next slide, we show how evaluating at multiple right-hand sides can further improves the approximation.

Example: Iterative Refinement







Interpreting Branch and Bound as a "Separation Oracle"

- Notice that what we are doing is dynamically generating an outer description of the epigraph of the value function.
- The generated dual functions can be interpreted as "cutting functions."
- In fact, we have used this observation to develop a "cutting plane algorithm" for generating the value function.
- These can also be interpreted as Benders' cuts!
- This leads to our third way of viewing duality.

Projection (Benders)

- *General optimization* problem: $X \subseteq \mathbb{R}^{n_1}$, $Y \subseteq \mathbb{R}^{n_2}$

$$\min_{x \in X,\ y \in Y} \{f(x,y) \mid F(x,y) \geq 0\} \tag{GP}$$

- Key idea: Consider the *projection* of this problem onto the set X.

$$\min_{x \in X} \phi(x), \tag{GP-Proj}$$

where

$$\phi(x) = \min_{y \in Y} \{f(x,y) \mid F(x,y) \geq 0\}$$

- Benders' Algorithm: Approximate ϕ from below and improve the approximation iteratively.

The Framework

Approximate ϕ with lower-bounding function $\underline{\phi}$, and improve the approximation iteratively.

Master problem:

$$\min_{x \in X} \underline{\phi}(x),$$

Subproblem:

$$\phi(x) = \min_{y \in Y} \{f(x, y) \mid F(x, y) \geq 0\}$$

Generalized Benders' Decomposition Algorithm

1. Solve the *master problem* (lower bound)
   a) Construct $\underline{\phi}(x)$ such that it is strong at all previous iterates $\{x^i\}_{i=1}^{k-1}$.
   b) Solve the master problem to obtain an optimal solution x^k. Set $\text{LB}^k = \underline{\phi}(x^k)$.
2. Solve the *subproblem* (upper bound)
   a) Solve the subproblem with x^k to obtain an optimal solution y^k and *improve* $\underline{\phi}$. Set $\text{UB}^k = \phi(x^k)$.
   b) *Termination check*: $\text{LB}^k = \text{UB}^k$? If yes, STOP. If no, set $k \leftarrow k + 1$ and go to Step 1.

Standard Implementation

- Usually, one $\underline{\phi}^i$ per iteration i is generated.
- In iteration k, the master problem can be reformulated as

$$\min_{x \in X} z$$
$$\text{s.t. } z \geq \underline{\phi}^i(x) \quad 1 \leq i \leq k.$$

- The constraints $z \geq \underline{\phi}^i(x)$ for $1 \leq i \leq k$ are typically called *optimality constraints* in Benders' standard approach.
- If $f(x, y) = g(x) + h(y)$ and $F(x, y) = G(x) + H(y)$, then ϕ is usually expressed as a function of right-hand side, which is again the *value function*!

Master problem:

$$\min_{x \in X} g(x) + z$$
$$\text{s.t. } z \geq \underline{\phi}^i(-G(x)) \quad 1 \leq i \leq k$$

Subproblem (evaluate $\phi(-G(x^k))$):

$$\phi(\beta) = \min_{y \in Y} \{h(y) \mid H(y) \geq \beta\}$$

Benders for MIBLP

- Applying Benders to (MIBLP), the relevant value function is

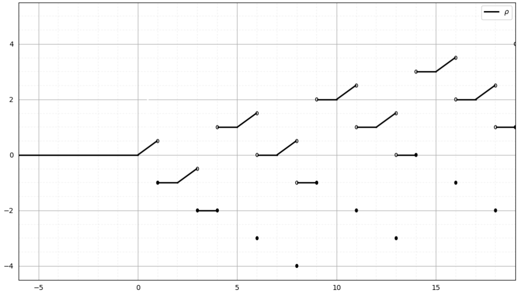
$$\rho(\beta) = \min \left\{ {d^1}^\top y \;\middle|\; y \in \arg\min \left\{ {d^2}^\top \check{y} \;\middle|\; \check{y} \in \mathcal{P}_2(\beta) \cap Y \right\} \right\}. \tag{7}$$

- Evaluating this even for one value of β looks rather difficult.
- However, it is “only” a lexicographic MILP.
- Nevertheless, these functions can get rather ugly!

Ugly!

Example 6

$$\rho(\beta) = \min \; -y_1 + y_2 - 5y_3 + y_4$$
$$\text{s.t. } (y_1, y_2, y_3, y_4) \in \arg\min \{2\check{y}_1 + 4\check{y}_2 + 3\check{y}_3 + 4\check{y}_4$$
$$\text{s.t. } 2\check{y}_1 + 5\check{y}_2 + 2\check{y}_3 + 2\check{y}_4 \geq \beta \qquad (8)$$
$$\check{y}_1, \check{y}_2, \check{y}_3 \in \mathbb{Z}_+, \; \check{y}_4 \in \mathbb{R}_+\}$$



To Infinity and Beyond!

- A conceptual extension of the generalized Benders' decomposition algorithm from MIBLPs to the general multistage case.
- An l-stage MILP can be formulated as a standard mathematical optimization problem with an $(l-1)$-stage value function constraint as follows.

$$\begin{aligned} \text{MMILP}^l(\beta) = \min\ & {d^{11}}^\top x^1 + {d^{12}}^\top x^2 + \ldots + {d^{1l}}^\top x^l \\ \text{s.t. } & A^{11}x^1 + A^{12}x^2 + \ldots + A^{1l}x^l \geq \beta \\ & x^1 \in X^1 \\ & (x^2, x^3, \ldots, x^l) \in \text{optimal set of } \text{MMILP}^{l-1}(b^2 - A^{21}x^1), \end{aligned} \tag{9}$$

- *$l - stage$* MILPs are hard for Σ_k^{P}.

Benders for Complexity Classes

- Further, the resulting recursive Benders algorithm reflects the recursive structure of the polynomial hierarchy itself.
- Surprisingly, there is a way to define the hierarchy in terms of projection and complement operations, as follows.

$$\Sigma_l^{\mathsf{P}} = \mathrm{proj}(\Pi_{l-1}^{\mathsf{P}})$$
$$\Pi_{l-1}^{\mathsf{P}} = \mathrm{comp}(\Sigma_{l-1}^{\mathsf{P}}).$$

Thanks To My Current and Former Ph.D Students!

- **Menal Güzelsoy**: R and Güzelsoy [2004], Güzelsoy and R [2007, 2008], Güzelsoy [2009]
- **Scott DeNegre**: DeNegre [2011], DeNegre and R [2009]
- **Anahita Hassanzadeh**: Hassanzadeh [2015], Hassanzadeh and R [2014a,b]
- **Sahar Tahernejad**: Tahernejad [2019], Tahernejad and R [2020], Tahernejad et al. [2020]
- **Suresh Bolusani**: Bolusani and R [2022], Bolusani et al. [2020]

References I

Amitabh Basu, Christopher Thomas Ryan, and Sriram Sankaranarayanan. Mixed-integer bilevel representability. *arXiv preprint arXiv:1808.03865*, 2018.

S. Bolusani and T.K. Ralphs. A Framework for Generalized Benders' Decomposition and Its Application to Multilevel Optimization. *Mathematical Programming Series B*, 2022. doi: 10.1007/s10107-021-01763-7. URL `https://arxiv.org/abs/2104.06496`.

S. Bolusani, S. Coniglio, T.K. Ralphs, and S. Tahernejad. A Unified Framework for Multistage Mixed Integer Linear Optimization. In S. Dempe and A. Zemkoho, editors, *Bilevel Optimization: Advances and Next Challenges*, chapter 18, pages 513–560. Springer, 2020. ISBN 978-3-030-52118-9. doi: 10.1007/978-3-030-52119-6. URL `https://arxiv.org/abs/2104.09003`.

S. DeNegre. *Interdiction and Discrete Bilevel Linear Programming*. PhD, Lehigh University, 2011. URL `http://coral.ie.lehigh.edu/~ted/files/papers/ScottDeNegreDissertation11.pdf`.

References II

S. DeNegre and T.K. Ralphs. A Branch-and-Cut Algorithm for Bilevel Integer Programming. In *Proceedings of the Eleventh INFORMS Computing Society Meeting*, pages 65–78, 2009. doi: 10.1007/978-0-387-88843-9_4. URL http://coral.ie.lehigh.edu/~ted/files/papers/BILEVEL08.pdf.

M. Güzelsoy. *Dual Methods in Mixed Integer Linear Programming*. PhD, Lehigh University, 2009. URL http://coral.ie.lehigh.edu/~ted/files/papers/MenalGuzelsoyDissertation09.pdf.

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.

M. Güzelsoy and T.K. Ralphs. The Value Function of a Mixed-integer Linear Program with a Single Constraint. Technical report, COR@L Laboratory, Lehigh University, 2008. URL http://coral.ie.lehigh.edu/~ted/files/papers/ValueFunction.pdf.

A. Hassanzadeh. *Two-Stage Stochastic Mixed Integer Optimization*. PhD, Lehigh, 2015.

References III

A. Hassanzadeh and T.K. Ralphs. On the Value Function of a Mixed Integer Linear Optimization Problem and an Algorithm for Its Construction. Technical Report 14T-004, COR@L Laboratory, Lehigh University, 2014a. URL http://coral.ie.lehigh.edu/~ted/files/papers/MILPValueFunction14.pdf.

A. Hassanzadeh and T.K. Ralphs. A Generalization of Benders' Algorithm for Two-Stage Stochastic Optimization Problems with Mixed Integer Recourse. Technical Report 14T-005, COR@L Laboratory, Lehigh University, 2014b. URL http://coral.ie.lehigh.edu/~ted/files/papers/SMILPGenBenders14.pdf.

J.T. Moore and J.F. Bard. The mixed integer linear bilevel programming problem. *Operations Research*, 38(5):911–921, 1990.

T.K. Ralphs and M. Güzelsoy. Duality, Warm Starting, and Sensitivity Analysis in Integer Programming. Technical Report 04T-022, COR@L Laboratory, Lehigh University, 2004. URL http://ise.lehigh.edu/sites/ise.lehigh.edu/files/04t{_}022.pdf.

References IV

S. Tahernejad. *Two-stage Mixed Integer Stochastic Bilevel Linear Optimization*. PhD, Lehigh University, 2019.

S. Tahernejad and T.K. Ralphs. Valid Inequalities for Mixed Integer Bilevel Optimization Problems. Technical Report 20T-013, COR@L Laboratory Report 20T-013, 2020. URL `http://coral.ie.lehigh.edu/~ted/files/papers/MibSCuts20.pdf`.

S. Tahernejad, T.K. Ralphs, and S.T. DeNegre. A Branch-and-Cut Algorithm for Mixed Integer Bilevel Linear Optimization Problems and Its Implementation. *Mathematical Programming Computation*, 12:529–568, 2020. doi: 10.1007/s12532-020-00183-6. URL `https://arxiv.org/abs/2104.09010`.