Duality and Discrete Optimization

Ted Ralphs[1]

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

Alibaba, DAMO Academy, 23 June 2022

Outline

Outline

What is Duality?

- Duality is a concept that is pervasive in mathematics but it can be hard to define ("I don't know what it is, but I know it when I see it!").
- Various notions of duality also arise in optimization and much of the theory underlying computational methods emerges from it.
- Many of the well-known "dualities" that arise in optimization and mathematics in general are closely connected.
- In fact, almost all such duality concepts can be seen as roughly "isomorphic."
- In a sense, any one can be derived from any other.

Duality Concepts

The following are duality concepts that play a role in optimization theory.

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
- Inverse problems: Functions and inverses, inverse optimization

Decision Problems and Complexity

- One way of connecting the theory of computation to other parts of mathematics is by formulating computational problems as problems about sets.
- We confine ourselves to problems in the *polynomial hierarchy* (PH), which is the categorization typically used for classifying optimization problems.
- This scheme applies only 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".
- In the theory of computation, the *formal proof* that the answer given by an algorithm is correct is called a *certificate*.
- 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.
- Formal proofs are constructed using the logic of a specific *formal system*.
- Mathematical optimization is a formal system for proving theorems about sets.

Theorems About Sets

- Let $\mathcal{S} = \{x \in \mathbb{Q}^n \mid P(x)\}$, where $P : \mathbb{Q}^n \rightarrow \{TRUE, FALSE\}$.
- The simplest question we can ask is whether $\mathcal{S}$ is non-empty.

$$\mathcal{S} \stackrel{?}{=} \emptyset. \tag{1}$$

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

$$\mathcal{S}(f, K) := \{x \in \mathcal{S} \mid f(x) < K\} \tag{2}$$

is non-empty is the *decision version* of the optimization problem

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

Constructing Proofs

- What do proofs of such 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.
  - Nonconvex sets are described using disjunctive logic: the *union* of convex sets is nonconvex (in general).
- This is why there is a short proof that a point is *not* in a convex set.
  - The Farkas Lemma and the separating hyperplane theorem in convex analysis provide methods for generating such proofs.
  - There is a short proof of emptiness for any set described as the intersection of simple convex sets, e.g., half-spaces.
- Proving a point is not in a nonconvex set is hard, which is why we can't expect short proofs of emptiness for disjunctive unions of convex sets.

Short Proofs of Emptiness

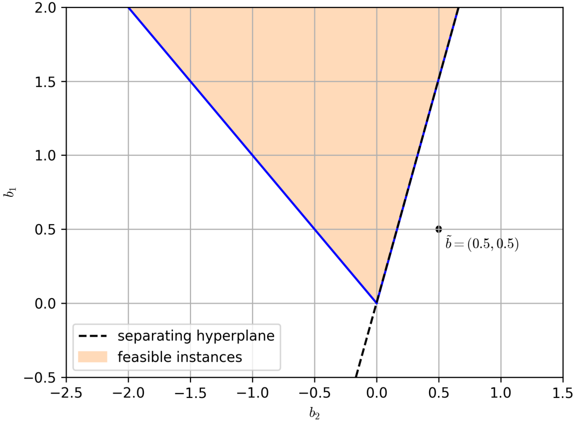
- In the case of convex sets, we can use a duality argument to obtain short proofs of emptiness.
- Consider the case of a polyhedron.

$$\mathcal{P} = \{x \in \mathbb{Q}^n_+ \mid Ax = \tilde{b}\} \tag{3}$$

- **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 $C = \{b \in \mathbb{Q}^m \mid \exists x \in \mathbb{Q}^n_+, Ax = b\} = \{b \in \mathbb{Q}^m : b^\top u \leq 0 \; \forall u \in C^*\}$, where $C^* = \{u \in \mathbb{Q}^m : A^\top u \leq 0\}$ (the *polar* of 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

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



Languages

- On one level, 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. $\nexists x \in \mathcal{S} : f(x) < K \Leftrightarrow \forall x \in \mathcal{S} : f(x) \geq K$

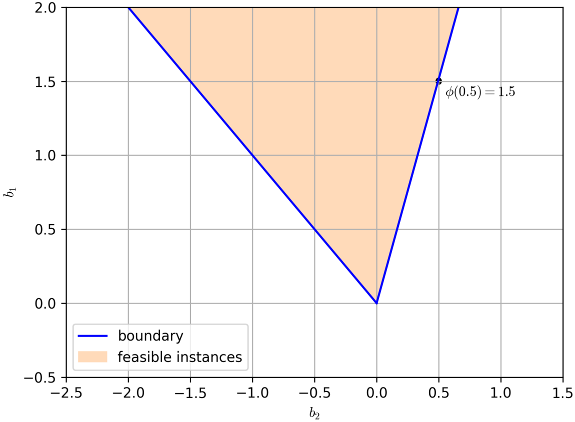
- The fact that one of these statements is universally quantified is the reason why short proofs of optimality cannot be expected in general.

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 $N = \{2, \dots, m\}$ 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 $b^* = \min\{b_1 \in \mathbb{Q} \mid b \in C\}$.
- To prove optimality, we need to show that $(b^*, \tilde{b}_N)$ is not only a member of C, but on its *boundary*.
- The proof is only slightly modified: $\exists u \in \mathbb{Q}^m, A^\top u \leq 0, (b^*, \tilde{b}_N)^\top u = 0, u_1 < 0$.
  - 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 C.
- The vector u is a solution to the usual LP dual problem.

Example

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

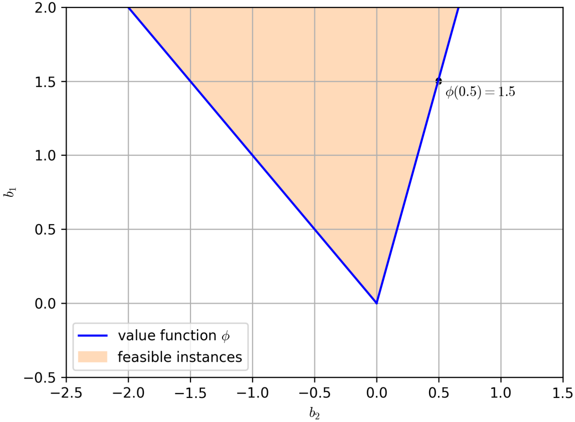


Interpreting

- It is not only the theorems in the class that are parametrically related, the proofs are themselves parametric.
- For example, the boundary of the cone C describes a parametric collection of proofs and has several nice interpretations.
  - The boundary can be interpreted as specifying the *value function* of the associated optimization problem.
  - The solution to the LP dual problem is a (sub)gradient of this function.
  - Alternatively, the boundary also encodes the way constraints can be traded off against each other (the *Pareto frontier*).
- The "dual price" of a given constraint has an economic interpretation when the constraints are interpreted as allocating resources.

Example

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



Outline

Mathematical Optimization

- The general form of a *mathematical optimization problem* is:

Form of a General Mathematical Optimization Problem

$$
\begin{array}{rrcll}
z_{MP} = \min & f(x) & & & \\
\text{s.t.} & g_i(x) & \leq & b_i, & 1 \leq i \leq m \\
& x & \in & X &
\end{array}
\tag{MP}
$$

where $X \subseteq \mathbb{R}^n$ may be a discrete set.

- The function f is the *objective function*, while g_i is the *constraint function* associated with constraint i.
- Our primary goal is to compute the optimal value z_{MP}.
- However, we may want to obtain some auxiliary information as well.
- More importantly, we may want to develop parametric forms of (MP) in which the input data are the output of some other function or process.

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 questions.

(Mixed Integer) Linear Optimization

- For this part of the talk, we focus on (single-level) mixed integer linear optimization problems (MILPs).

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

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

- **Note that we are using the equality form of constraints to simplify the presentation**.
- In this context, we can make the economic concepts just discussed more concrete.
- We can think of each row of A as representing a resource and each column as representing an activity or product.
- For each activity, resource consumption is a linear function of activity level.
- We first consider the case $r = 0$, which is the case of the (continuous) linear optimization problem (LP).

The LP Value Function

- Of central importance in duality theory for linear optimization is the *value function*, defined by

$$\phi_{LP}(\beta) = \min_{x \in \mathcal{S}(\beta)} c^\top x, \tag{LPVF}$$

  for a given $\beta \in \mathbb{R}^m$, where $\mathcal{S}(\beta) = \{x \in \mathbb{R}^n_+ \mid Ax = \beta\}$.
- We let $\phi_{LP}(\beta) = \infty$ if $\beta \in \Omega = \{\beta \in \mathbb{R}^m \mid \mathcal{S}(\beta) = \emptyset\}$.
- The value function returns the optimal value as a parametric function of the right-hand side vector, which represents available resources.

Economic Interpretation of the Value Function

- What information is encoded in the value function?
  - Consider the gradient $u = \phi'_{LP}(\beta)$ at β for which ϕ_{LP} is continuous.
  - The quantity $u^\top \Delta b$ represents the marginal change in the optimal value if we change the resource level by Δb.
  - In other words, it can be interpreted as a vector of the *marginal costs of the resources*.
  - This is also known as the *dual solution vector*, but we should really think of it as a linear function.
- In the LP case, the gradient is a *linear under-estimator* of the value function and can thus be used to derive bounds on the optimal value for any $\beta \in \mathbb{R}^m$.

Small Example: Fractional Knapsack Problem

- We are given a set $N = \{1, \dots n\}$ of items and a capacity W.
- There is a profit p_i and a size w_i associated with each item $i \in N$.
- We want a set of items that maximizes profit subject to the constraint that their total size does not exceed the capacity.
- In this variant of the problem, we are allowed to take a fraction of an item.
- For each item i, let variable x_i represent the fraction selected.

Fractional Knapsack Problem

$$
\begin{aligned}
\min \sum_{j=1}^{n} p_j x_j \\
\text{s.t.} \sum_{j=1}^{n} w_j x_j \leq W \\
0 \leq x_i \leq 1 \qquad \forall i
\end{aligned}
\tag{4}
$$

- What is the optimal solution?

Generalizing the Knapsack Problem

- Let us consider the value function of a (generalized) knapsack problem.
- To be as general as possible, we allow sizes, profits, and even the capacity to be negative.
- We also take the capacity constraint to be an equality.
- This is a proper generalization.

Example 1

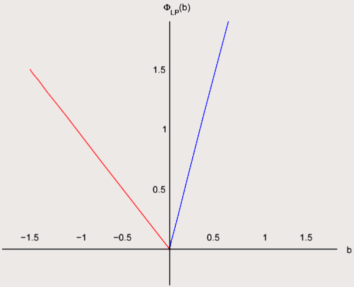
$$\begin{aligned} \phi_{LP}(\beta) = \min \quad & 6y_1 + 7y_2 + 5y_3 \\ \text{s.t.} \quad & 2y_1 - 7y_2 + y_3 = \beta \\ & y_1, y_2, y_3, \in \mathbb{R}_+ \end{aligned}$$

□

Value Function of the (Generalized) Knapsack Problem

- Now consider the value function ϕ_{LP} of Example 1.
- What do the gradients of this function represent?

Value Function for Example 1



The Dual Optimization Problem

- Can we calculate the gradient of ϕ_{LP} at b directly?
- Note that for any $\mu \in \mathbb{R}^m$, we have

$$\begin{aligned} \min_{x \geq 0} \left[c^\top x + \mu^\top (b - Ax)\right] &\leq c^\top x^* \; + \mu^\top (b - Ax^*) \\ &= c^\top x^* \\ &= \phi_{LP}(b) \end{aligned}$$

  and thus we have a lower bound on $\phi_{LP}(b)$.
- With some simplification, we can obtain a more explicit form for this bound.

$$\begin{aligned} min_{x \geq 0} \left[c^\top x + \mu^\top (b - Ax)\right] &= \mu^\top b + min_{x \geq 0}(c^\top - \mu^\top A)x \\ &= \begin{cases} \mu^\top b, & \text{if } c^\top - \mu^\top A \geq \mathbf{0}^\top, \\ -\infty, & \text{otherwise,} \end{cases} \end{aligned}$$

The Dual Problem (cont'd)

- If we now interpret this quantity as a function

$$g(u,\beta) = \begin{cases} u^\top \beta, & \text{if } c^\top - u^\top A \geq \mathbf{0}^\top, \\ -\infty, & \text{otherwise,} \end{cases}$$

  with parameters u and β, then for fixed first parameter, $g(\cdot,\beta)$ is a linear under-estimator of ϕ_{LP}.

- An LP dual problem is obtained by computing the $u \in \mathbb{R}^m$ that gives the under-estimator yielding the strongest bound for a fixed b.

LP Dual Problem

$$\begin{aligned} \max_{\mu \in \mathbb{R}^m} g(\mu,\cdot) = \max \quad & b^\top \mu \\ \text{s.t.} \quad & \mu^\top A \leq c^\top \end{aligned} \tag{LPD}$$

- An optimal solution to (LPD) is a (sub)gradient of ϕ_{LP} at b.

Combinatorial Representation of the LP Value Function

- Note that the feasible region of (LPD) does not depend on b.
- From the fact that there is always an extremal optimum to (LPD), we conclude that the LP value function can be described combinatorially.

Combinatorial Representation of the LP Value Function

$$\phi_{LP}(\beta) = \max_{u \in \mathcal{E}} u^\top \beta \qquad \text{(LPVF)}$$

for $\beta \in \mathbb{R}^m$, where $\mathcal{E}$ is the set of extreme points of the *dual polyhedron* $\mathcal{D} = \{u \in \mathbb{R}^m \mid u^\top A \leq c^\top\}$ (assuming boundedness).

- Alternatively, $\mathcal{E}$ is also in correspondence with the dual feasible bases of A.

$$\mathcal{E} \equiv \left\{ c_B A_E^{-1} \mid E \text{ is the index set of a dual feasible bases of A} \right\}$$

- Thus, we see that $\mathrm{epi}(\phi_{LP})$ is a polyhedral cone and whose facets correspond to dual feasible bases of A.

What is the Importance?

- The dual problem is important is because it gives us a set of *optimality conditions*.
- For a given $b \in \mathbb{R}^m$, whenever we have
  - $x^* \in \mathcal{S}(b)$,
  - $u \in \mathcal{D}$, and
  - $c^\top x^* = u^\top b = \phi_{LP}(b)$,

  then x^* is optimal!
- This means we can write down a set of constraints involving the value function that ensure optimality.
- This set of constraints can then be embedded inside another optimization problem.

The MILP Value Function

- We now generalize the notions seen so far to the MILP case.
- This is quite natural by building on the concept of LP duality we've just developed.
- We start by defining the *value function* associated with the base instance (MILP), which is

MILP Value Function

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

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: The (Mixed) Binary Knapsack Problem

- We now consider a further generalization of the previously introduced knapsack problem.
- In this problem, we must take some of the items either fully or not at all.
- In the example, we allow all of the previously introduced generalizations.

Example 2

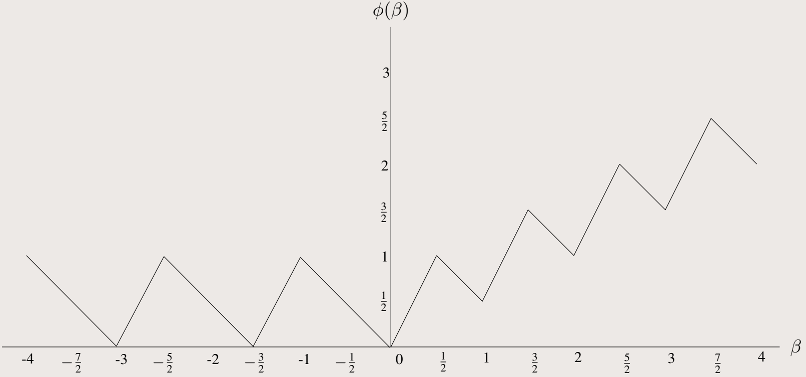
$$
\begin{aligned}
\phi(\beta) = \min \quad & \tfrac{1}{2}x_1 + 2x_3 + x_4 \\
\text{s.t} \quad & x_1 - \tfrac{3}{2}x_2 + x_3 - x_4 = \beta \\
& x_1, x_2 \in \mathbb{Z}_+, x_3, x_4 \in \mathbb{R}_+.
\end{aligned}
\tag{5}
$$

□

Value Function for (Generalized) Mixed Binary Knapsack

- Below is the value function of the optimization problem in Example 2.
- How do we interpret the structure of this function?

Value Function for Example 2



Related Work on Value Function

Duality

- Johnson [1973, 1974, 1979]
- Jeroslow [1979]
- Wolsey [1981]
- Güzelsoy and Ralphs [2007], Güzelsoy [2009]

Structure and Construction

- Blair and Jeroslow [1977b, 1979, 1982, 1984, 1985], Blair [1995]
- Kong et al. [2006]
- Güzelsoy and Ralphs [2008], Hassanzadeh and Ralphs [2014]

Sensitivity and Warm Starting

- Ralphs and Güzelsoy [2005, 2006], Güzelsoy [2009]
- Gamrath et al. [2015]

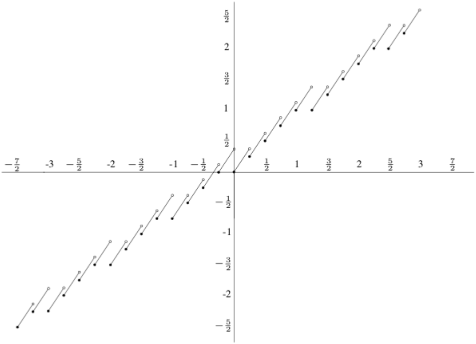
Properties of the MILP Value Function

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

Example 3

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

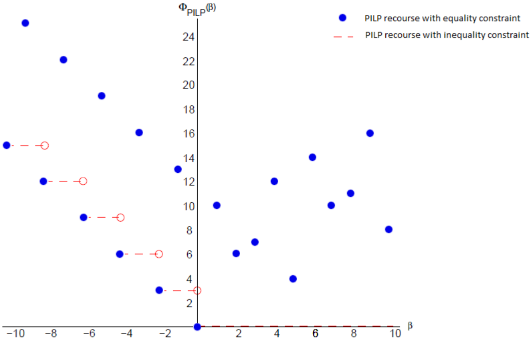
□



Example 4

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

□

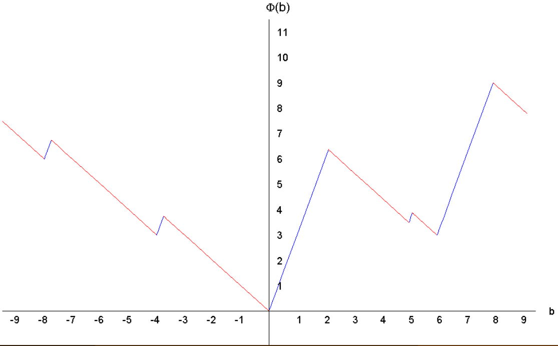


Another Example

Example 5

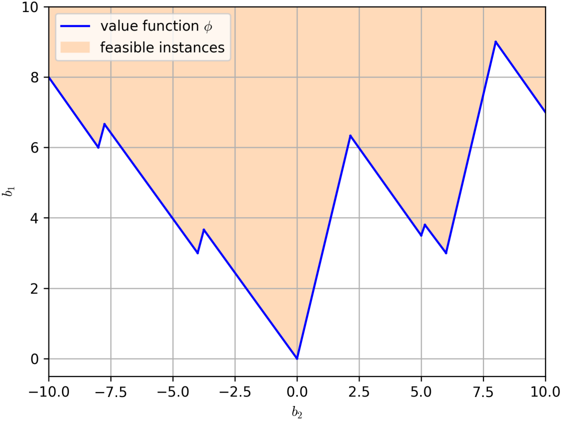
$$\phi(\beta) = \min 3x_1 + \frac{7}{2}x_2 + 3x_3 + 6x_4 + 7x_5 + 5x_6$$
$$\text{s.t. } 6x_1 + 5x_2 - 4x_3 + 2x_4 - 7x_5 + x_6 = \beta$$
$$x_1, x_2, x_3 \in \mathbb{Z}_+,\ x_4, x_5, x_6 \in \mathbb{R}_+$$

□



Another Example (cont'd)

As before, the value function represents the boundary between feasible and infeasible instances in a parametric family.



Continuous and Integer Restriction of an MILP

The structure of the value function is inherited from two related functions.

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

The two functions are the *continuous restriction*:

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

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

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

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

Discrete Representation of the Value Function

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

$$\begin{aligned} \phi(\beta) = \min\ & c_I^\top x_I + \phi_C(\beta - A_I x_I) \\ \text{s.t. } & x_I \in \mathbb{Z}^r_+ \end{aligned} \tag{6}$$

- 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 $\hat{x}_I \in \mathbb{Z}^r_+$, let

$$\phi_C(\beta, \hat{x}_I) = c_I^\top \hat{x}_I + \phi_C(\beta - A_I \hat{x}_I) \ \forall \beta \in \mathbb{R}^m. \tag{7}$$

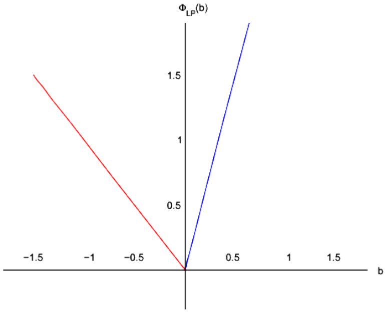
- Then we have that $\phi(\beta) = \min_{x_I \in \mathbb{Z}^r_+} \phi_C(\beta, \hat{x}_I)$.

Value Function of the Continuous Restriction

Example 6

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

$\square$



Related Results

- From the basic structure outlined, we can derive many other useful results.

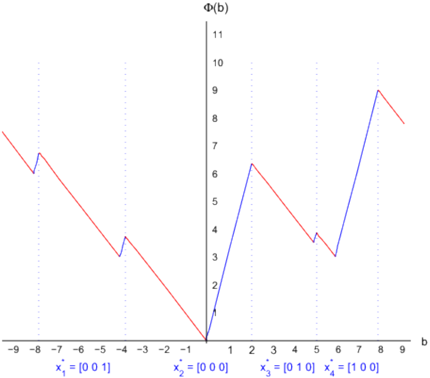
Proposition 1. *[Hassanzadeh and Ralphs, 2014] The gradient of* ϕ *on a neighborhood of a differentiable point is a unique optimal dual feasible solution to* (CR).

Proposition 2. *[Hassanzadeh and Ralphs, 2014] If* ϕ *is differentiable over a connected set* $\mathcal{N} \subseteq \mathbb{R}^m$*, then there exists* $x_I^* \in \mathbb{Z}^r$ *and* $E \in \mathcal{E}$ *such that* $\phi(b) = c_I^\top x_I^* + \nu_E^\top (b - A_I x_I^*)$ *for all* $b \in \mathcal{N}$.

- This last result can be extended to subset of the domain over which ϕ is convex.
- Over such a region, ϕ coincides with the value function of a translation of the continuous restriction.
- Putting all of together, we get a practical finite representation...

Points of Strict Local Convexity (Finite Representation)

Example 7



Theorem 1. *[Hassanzadeh and Ralphs, 2014]* $\square$
Under the assumption that $\{\beta \in \mathbb{R}^m \mid \phi_I(\beta) < \infty\}$ *is finite, there exists a finite set* $\mathcal{S} \subseteq Y$ *such that*

$$\phi(\beta) = \min_{x_I \in \mathcal{S}} \{c_I^\top x_I + \phi_C(\beta - A_I x_I)\}. \tag{8}$$

Interpretation

- It is only possible to get a unique linear price function for resource vectors where the value function is differentiable.
- This only happens when the continuous restriction has a unique dual solution at the current resource vector.
- Otherwise, there is no linear price function that will be valid in an epsilon neighborhood of the current resource vector.
- When this function has a gradient, its value is determined only by the continuous part of the problem!
- Thus, these prices reflect behavior over only a very localized region for which the discrete part of the solution remains constant.
- In the case of the generalized knapsack problem, the differentiable points have the following two properties:
  - the continuous part of the solution is non-zero (and unique); and
  - The discrete part of the solution is unique.

Outline

Dual Bounding Functions

- 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 choose 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).

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

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 $F^*(b) = \phi(b)$.
- This dual instance always has a solution F^* that is strong if the value function is bounded and $\Upsilon^m \equiv \{f \mid f : \mathbb{R}^m \to \mathbb{R}\}$. Why?

Example: LP Relaxation Dual Function

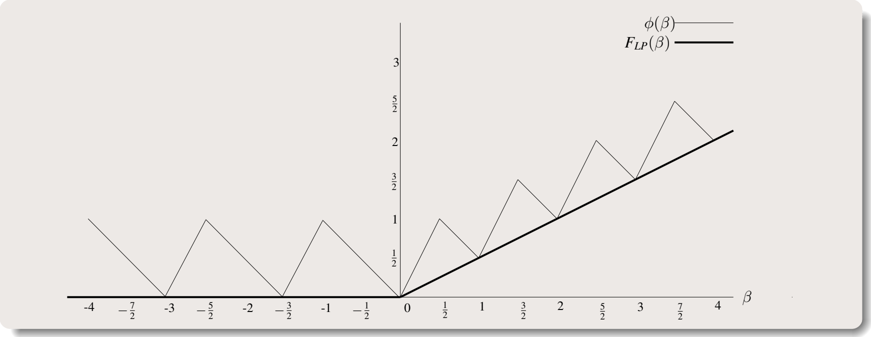
Example 8

$$F_{LP}(d) = \min \quad vd,$$
$$\text{s.t} \quad 0 \geq v \geq -\tfrac{1}{2}, \text{ and} \tag{9}$$
$$v \in \mathbb{R},$$

□

which can be written explicitly as

$$F_{LP}(\beta) = \begin{cases} 0, & \beta \leq 0 \\ -\frac{1}{2}\beta, & \beta > 0 \end{cases}.$$



The Subadditive Dual

By considering that

$$F(\beta) \leq \phi(\beta) \; \forall \beta \in \mathbb{R}^m \quad \Longleftrightarrow \quad F(\beta) \leq c^\top x \, , \; x \in \mathcal{S}(\beta) \; \forall \beta \in \mathbb{R}^m$$
$$\Longleftrightarrow \quad F(Ax) \leq c^\top x \, , \; x \in \mathbb{Z}^n_+ ,$$

the generalized dual problem can be rewritten as

$$\max \left\{ F(\beta) : F(Ax) \leq cx, \; x \in \mathbb{Z}^r_+ \times \mathbb{R}^{n-r}_+, \; F \in \Upsilon^m \right\}.$$

Can we further restrict Υ^m and still guarantee a strong dual solution?

- The class of linear functions? NO!
- The class of convex functions? NO!
- The class of subadditive functions? YES!

See [Johnson, 1973, 1974, 1979, Jeroslow, 1979] for details.

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. Why?
- 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 the dual problem above 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

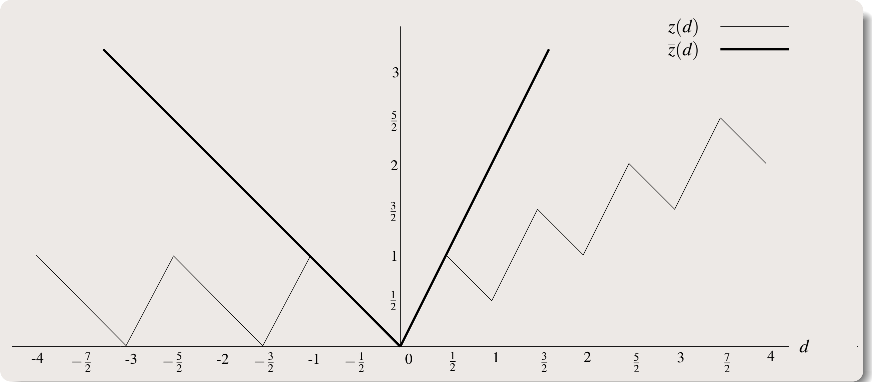
$$
\bar{F}(\beta) = \limsup_{\delta \rightarrow 0^+} \frac{F(\delta\beta)}{\delta} \quad \forall \beta \in \mathbb{R}^m.
$$

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

Example: Upper β-directional Derivative

- The upper β-directional derivative is βu, where u is the gradient at $\epsilon\beta$ for sufficiently small ϵ.
- Recall that u is also the (unique) solution to the dual of the continuous restriction.
- Therefore, the problem reduces to the LP dual around the origin (and globally, in the continuous case).

Example 9



Weak Duality

Weak Duality Theorem

Let x be a feasible solution to the primal problem and let F be a feasible solution to the subadditive dual. Then, $F(b) \leq c^\top x$.

Proof.

Corollary

For the primal problem and its subadditive dual:

1. If the primal problem (resp., the dual) is unbounded then the dual problem (resp., the primal) is infeasible.
2. If the primal problem (resp., the dual) is infeasible, then the dual problem (resp., the primal) is infeasible or unbounded.

Strong Duality

Strong Duality Theorem

If the primal problem (resp., the dual) has a finite optimum, then so does the subadditive dual problem (resp., the primal) and they are equal.

Outline of the Proof. Show that the value function ϕ or an extension of ϕ is a feasible dual function.

- Note that ϕ satisfies the dual constraints.
- $\Omega \equiv \mathbb{R}^m$: $\phi \in \Gamma^m$.
- $\Omega \subset \mathbb{R}^m$: $\exists\ \phi_e \in \Gamma^m$ with $\phi_e(\beta) = \phi(\beta)\ \forall \beta \in \Omega$ and $z_e(\beta) < \infty\ \forall \beta \in \mathbb{R}^m$.

Example: Subadditive Dual

For the instance in Example 2, the subadditive dual

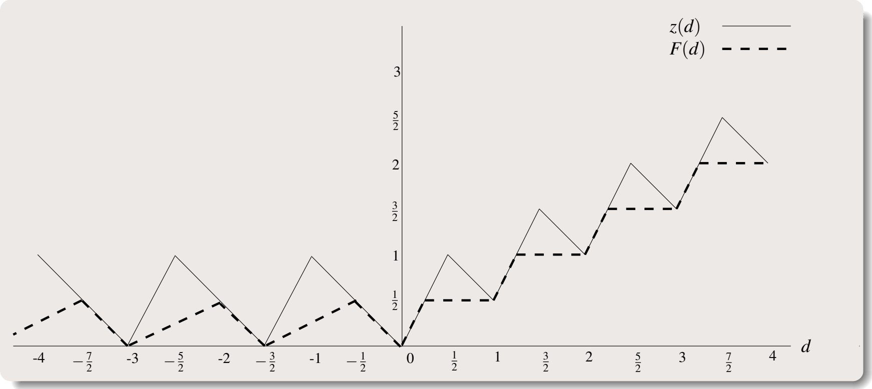
$$
\begin{array}{rrl}
\max & F(b) & \\
 & F(1) & \leq \frac{1}{2} \\
 & F(-\frac{3}{2}) & \leq 0 \\
 & \bar{F}(1) & \leq 2 \\
 & \bar{F}(-1) & \leq 1 \\
 & F \in \Gamma^1. &
\end{array}
\quad .
$$

and we have the following feasible dual functions:

1. $F_1(\beta) = \frac{\beta}{2}$ is an optimal dual function for $\beta \in \{0, 1, 2, ...\}$.
2. $F_2(\beta) = 0$ is an optimal function for $\beta \in \{..., -3, -\frac{3}{2}, 0\}$.
3. $F_3(\beta) = \max\{\frac{1}{2}\lceil \beta - \frac{\lceil \lceil \beta \rceil - \beta \rceil}{4} \rceil, 2d - \frac{3}{2}\lceil \beta - \frac{\lceil \lceil \beta \rceil - \beta \rceil}{4} \rceil\}$ is an optimal function for $b \in \{[0, \frac{1}{4}] \cup [1, \frac{5}{4}] \cup [2, \frac{9}{4}] \cup ...\}$.
4. $F_4(\beta) = \max\{\frac{3}{2}\lceil \frac{2\beta}{3} - \frac{2\lceil \lceil \frac{2\beta}{3} \rceil - \frac{2\beta}{3} \rceil}{3} \rceil - \beta, -\frac{3}{4}\lceil \frac{2\beta}{3} - \frac{2\lceil \lceil \frac{2\beta}{3} \rceil - \frac{2\beta}{3} \rceil}{3} \rceil + \frac{\beta}{2}\}$ is an optimal function for $b \in \{... \cup [-\frac{7}{2}, -3] \cup [-2, -\frac{3}{2}] \cup [-\frac{1}{2}, 0]\}$

Example: Feasible Dual Functions

Example 10



□

- Notice how different dual solutions are optimal for some right-hand sides and not for others.
- Only the value function is optimal for all right-hand sides.

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 $F(a^j) \geq 0, j = 1, ..., n$, and $F(b) < 0$.

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

Complementary Slackness [Wolsey, 1981]

For a given right-hand side b, let x^* and F^* be feasible solutions to the primal and the subadditive dual problems, respectively. Then x^* and F^* are optimal solutions if and only if

1. $x_j^*(c_j - F^*(a^j)) = 0, j = 1, ..., n$ and
2. $F^*(b) = \sum_{j=1}^n F^*(a^j)x_j^*$.

Proof. For an optimal pair we have

$$F^*(b) = F^*(Ax^*) = \sum_{j=1}^{n} F^*(a^j)x_j^* = cx^*. \tag{10}$$

Optimality Conditions

- One reason the dual problem is important is because it gives us a set of *optimality conditions*.

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-cut algorithm for MILP that prove the optimality of the primal solution.
- The branch-and-bound tree encodes a solution to the dual.

Outline

Approximating the Value Function

- In general, it is difficult to construct the value function explicitly.
- We therefore propose to approximate the value function by either primal (upper) or dual (lower) bounding functions.

Dual bounds

Derived by considering the value function of *relaxations* of the original problem or by constructing *dual functions* $\Rightarrow$ Relax constraints.

Primal bounds

Derived by considering the value function of *restrictions* of the original problem $\Rightarrow$ Fix variables.

Primal/Dual Bounding Functions

Dual (Bounding) Functions

Definition 1. A *dual (bounding) function* $F : \mathbb{R}^m \to \mathbb{R}$ is one that satisfies $F(\beta) \leq \phi(\beta)$ for all $\beta \in \mathbb{R}^m$.

Primal (Bounding) Functions

Definition 2. A *primal (bounding) function* $F : \mathbb{R}^m \to \mathbb{R}$ is one that satisfies $F(\beta) \geq \phi(\beta)$ for all $\beta \in \mathbb{R}^m$.

Strong Bounding Functions

Definition 3. A bounding function F is said to be *strong* at $b \in \mathbb{R}^m$ if $F(b) = \phi(b)$.

Strong Primal Bounding Functions

- Strong bounding functions can be used algorithmically both to construct the value function directly and to dynamically construct approximations.
- These approximations can be used in algorithms for multi-stage optimization.

Theorem 2. *Let x^* be an optimal solution to the primal problem with right-hand side b. Then $\phi_C(\beta, x_I^*)$ is a strong primal bounding function at b.*

- By repeatedly evaluating $\phi_I(\beta)$, we can obtain upper approximations (and eventually the full value function).

Benders-like Algorithm for Upper Approximation

Algorithm

Initialize: Let $\bar{\phi}(b) = \infty$ for all $b \in B$, $\Gamma^0 = \infty$, $x_I^0 = 0$, $S^0 = \{x_I^0\}$, and $k = 0$.
while $\Gamma^k > 0$ **do**:

- Let $\bar{\phi}(\beta) \leftarrow \min\{\bar{\phi}(\beta), \bar{\phi}(\beta; x_I^k)\}$ for all $\beta \in \mathbb{R}^m$.
- $k \leftarrow k + 1$.
- Solve

$$\begin{aligned} \Gamma^k = \max_{\beta \in \mathbb{R}^m} \ & \bar{\phi}(\beta) - c_I^\top x_I \\ \text{s.t. } & A_I x_I = b \\ & x_I \in \mathbb{Z}_+^r. \end{aligned} \tag{SP}$$

  to obtain x_I^k.
- Set $S^k \leftarrow S^{k-1} \cup \{x^k\}$

end while
return $\phi(b) = \bar{\phi}(b)$ for all $b \in B$.

Algorithm for Upper Approximation

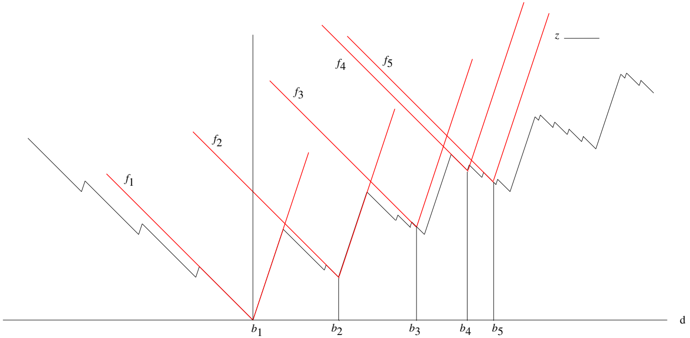


Figure 1: Upper bounding functions obtained at right-hand sides $b_i, i = 1, \ldots, 5$.

Formulating (SP)

Surprisingly, the "cut generation" problem (SP) can be formulated easily as an MINLP.

$$
\begin{aligned}
\Gamma^k = \ & \max \theta \\
& \text{s.t. } \theta + c_I^\top x_I \le c_I^\top x_I^i + (A_I x_I - A_I x_I^i)^\top \nu^i \quad i = 1, \dots, k-1 \\
& \qquad A_C^\top \nu^i \le c_C \quad i = 1, \dots, k-1 \\
& \qquad \nu^i \in \mathbb{R}^m \quad i = 1, \dots, k-1 \\
& \qquad x_I \in \mathbb{Z}_+^r.
\end{aligned} \tag{11}
$$

Sample Computational Results

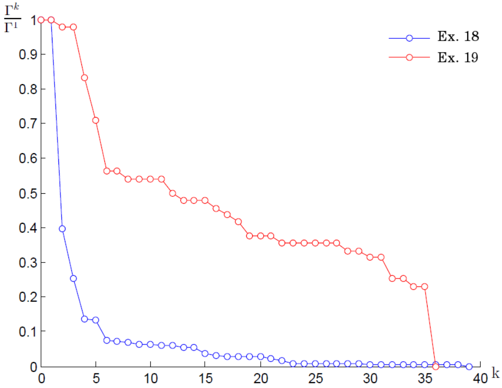


Figure 2: Normalized approximation gap vs. iteration number.

http://github.com/tkralphs/ValueFunction

Dual Bounding Functions Revisited

- 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 choose 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).

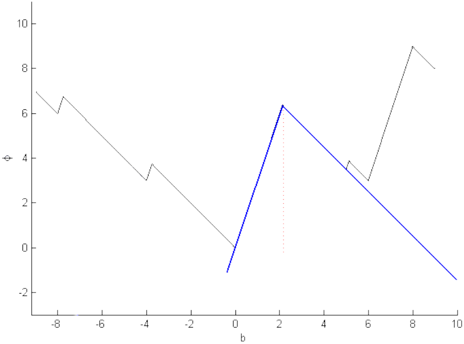
$$\max \{F(b) : F(\beta) \leq \phi(\beta),\ \beta \in \mathbb{R}^m, F \in \Upsilon^m\} \qquad \text{(D)}$$

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 $F^*(b) = \phi(b)$.
- This dual instance always has a solution F^* that is strong if the value function is bounded and $\Upsilon^m \equiv \{f \mid f : \mathbb{R}^m \to \mathbb{R}\}$. Why?

Dual Functions from Branch and Bound

- Recall that a *dual function* $F : \mathbb{R}^m \to \mathbb{R}$ is one that satisfies $F(\beta) \leq \phi(\beta)$ for all $\beta \in \mathbb{R}^m$.
- Observe that any branch-and-bound tree yields a lower approximation of the value function.



Dual Functions from Branch-and-Bound [Wolsey, 1981]

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

$$\begin{aligned} \phi^t(\beta) = \min\ & c^\top x \\ \text{s.t. } & Ax = \beta, \\ & l^t \le x \le u^t, x \ge 0 \end{aligned} \tag{12}$$

The dual at node t:

$$\begin{aligned} \phi^t(\beta) = \max\ & \{\pi^t \beta + \underline{\pi}^t l^t + \bar{\pi}^t u^t\} \\ \text{s.t. } & \pi^t A + \underline{\pi}^t + \bar{\pi}^t \le c^\top \\ & \underline{\pi} \ge 0, \bar{\pi} \le 0 \end{aligned} \tag{13}$$

We obtain the following strong dual function:

$$\min_{t \in T}\{\hat{\pi}^t \beta + \hat{\underline{\pi}}^t l^t + \hat{\bar{\pi}}^t u^t\}, \tag{14}$$

where $(\hat{\pi}^t, \hat{\underline{\pi}}^t, \hat{\bar{\pi}}^t)$ is an optimal solution to the dual (BB.LP.D).

Interpreting Branch and Bound as a Dual Method

- An alternative way of viewing branch and bound is simply as a method of iteratively refining a single overall disjunction (or dual function).
- The dual function arising from the branch-and-bound tree is

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

where $(\hat{\pi}^{t}, \underline{\hat{\pi}}^{t}, \bar{\hat{\pi}}^{t})$ is an optimal solution to the following dual at node t.

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

- When we branch, we remove one linear function from the above minimum and replace it with the minimum of two others.
- Depending on how we choose the disjunction, this will hopefully improve the bound yielded by the dual function.

Example: Branching as Dual Improvement

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

$$\begin{aligned} \phi(\beta) = \min\ & 6x_1 + 4x_2 + 3x_3 + 4x_4 + 5x_5 + 7x_6 \\ \text{s.t. } & 2x_1 + 5x_2 - 2x_3 - 2x_4 + 5x_5 + 5x_6 = \beta \\ & x_1, x_2, x_3 \in \mathbb{Z}_+, x_4, x_5, x_6 \in \mathbb{R}_+. \end{aligned} \tag{15}$$

- Suppose we evaluate $\phi(3.5)$ by solving the instance with right-hand side 3.5 by branch-and-bound.
- Solving the root LP relaxation, we obtain a solution in which $x_2 = 0.7$ and the optimal dual multipler 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 $x_2 \leq 0$ and $x_2 \geq 1$, respectively.
- Here, the problem is solved after this single branching.

Example: Dual Function from Branch and Bound

- Interpreting the branching in terms of dual functions, 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.

$$\begin{aligned}
\underline{\phi}^0_{\mathrm{LP}}(\beta) &= 0.8\beta \\
\underline{\phi}^1_{\mathrm{LP}}(\beta) &= \beta \\
\underline{\phi}^2_{\mathrm{LP}}(\beta) &= -1.5\beta + 11.5
\end{aligned}$$

- 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 $\underline{\phi}^{\mathcal{T}_1} = \min\{\underline{\phi}^1_{\mathrm{LP}}, \underline{\phi}^2_{\mathrm{LP}}\}$.

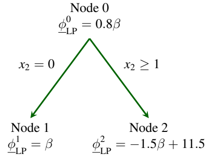
Example: Strengthening the Dual Function

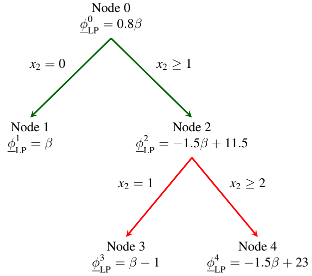
- The dual function can be strengthened by noting that the dual feasible region is the same for all nodes.
- We can therefore approximate the nodal value function by taking a max over all known dual solutions.
- Then we obtain

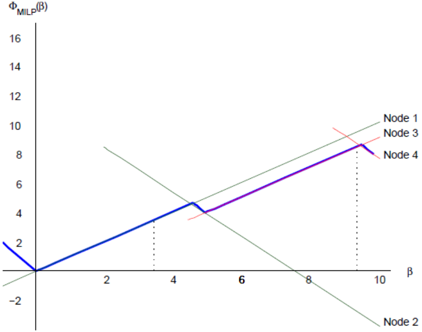
$$\min\{\max\{0.8\beta, \beta, -1.5\beta\}, \max\{0.8\beta, \beta, -1.5\beta + 11.5\}\} =$$
$$\min\{\max\{\beta, -1.5\beta\}, \max\{0.8\beta, -1.5\beta + 11.5\}\}$$

- Further, 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 $\phi(11.5)$ improves the approximation around that value of β.

Example: Iterative Refinement (cont'd)







Tree Representation of the Value Function

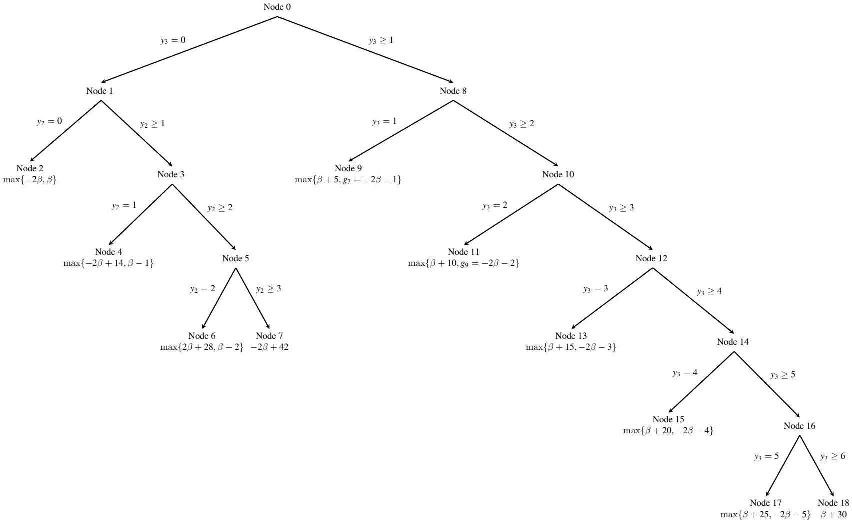
- Continuing the process, we eventually generate the entire value function.
- Consider the strengthened dual

$$\underline{\phi}^*(\beta) = \min_{t \in T} q_{I_t}^\top y_{I_t}^t + \phi_{N \setminus I_t}^t(\beta - W_{I_t} y_{I_t}^t), \tag{16}$$

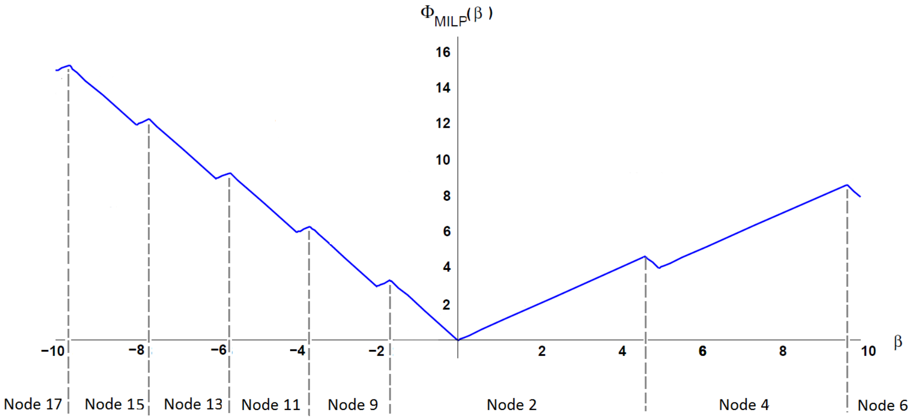
- I_t is the set of indices of fixed variables, $y_{I_t}^t$ are the values of the corresponding variables in node t.
- $\phi_{N \setminus I_t}^t$ is the value function of the linear optimization problem at node t, including only the unfixed variables.

  Theorem 3. *[Hassanzadeh and Ralphs, 2014] Under the assumption that* $\{\beta \in \mathbb{R}^m \mid \phi_I(\beta) < \infty\}$ *is finite, there exists a branch-and-bound tree with respect to which* $\underline{\phi}^* = \phi$.

Example of Value Function Tree



Correspondence of Nodes and Local Stability Regions



Dual Functions from the Cutting Plane Method

- Recall that there is a version of Gomory's Cutting Plane Method that yields a finite algorithm for ILPs.
- By tracking the operations undertaken to construct each inequality, we can obtain a different kind of (strong) dual function.
- Just as with branch-and-bound, the full value function can be obtained by taking the max over a collection of such dual functions.
- The operations needed are only the following simple ones.

$$\left.\begin{array}{l}\left.\begin{array}{l}\text{(i) rational multiplication}\\ \text{(ii) nonnegative combination}\\ \text{(iii) rounding}\end{array}\right\}\ \textit{Chvátal fcns.}\\ \text{(iv) taking the maximum}\end{array}\right\}\ \textit{Gomory fcns.}$$

- Note that the first three operations preserve subadditivity.

Chvátal and Gomory Functions

- Let $\mathcal{L}^m = \{f \mid f : \mathbb{R}^m \to \mathbb{R}, f \text{ is linear}\}$.
- Chvátal functions are the smallest set of functions $\mathscr{C}^m$ such that
  1. If $f \in \mathcal{L}^m$, then $f \in \mathscr{C}^m$.
  2. If $f_1, f_2 \in \mathscr{C}^m$ and $\alpha, \beta \in \mathbb{Q}_+$, then $\alpha f_1 + \beta f_2 \in \mathscr{C}^m$.
  3. If $f \in \mathscr{C}^m$, then $\lceil f \rceil \in \mathscr{C}^m$.
- Gomory functions are the smallest set of functions $\mathscr{G}^m \subseteq \mathscr{C}^m$ with the additional property that
  1. If $f_1, f_2 \in \mathscr{G}^m$, then $\max\{f_1, f_2\} \in \mathscr{G}^m$.

  It is easy to see that Chvátal functions are subadditive.

  Theorem 4. *For PILPs ($r = n$), if $\phi(0) = 0$, then there is a $g \in \mathscr{G}^m$ such that $g(d) = \phi(\beta)$ for all $d \in \mathbb{R}^m$ with $\mathcal{S}(d) \neq \emptyset$.*
- In fact, there is a one-to-one correspondence between ILP instances and Gomory functions!
- This result can be extended to MILPs by the addition of a correction term.
- The resulting form of the value is called the *Jeroslow Formula*.

Gomory's Procedure [Blair and Jeroslow, 1977a]

- For an ILP, there is a Chvátal function that is optimal to the subadditive dual.
- The procedure:
  In iteration k, we solve the following LP

$$
\begin{aligned}
\phi^{k-1}(\beta) = \min \quad & cx \\
\text{s.t.} \quad & Ax = \beta \\
& \textstyle\sum_{j=1}^{n} f^i(a_j)x_j \geq f^i(\beta) \qquad i = 1, ..., k-1 \\
& x \geq 0
\end{aligned}
$$

- The k^{th} cut, $k > 1$, is dependent on the RHS and written as:

$$f^k(\beta) = \left\lceil \sum_{i=1}^{m} \lambda_i^{k-1}\beta_i + \sum_{i=1}^{k-1} \lambda_{m+i}^{k-1} f^i(\beta) \right\rceil \quad \text{where} \quad \lambda^{k-1} = (\lambda_1^{k-1}, ..., \lambda_{m+k-1}^{k-1}) \geq 0$$

Gomory's Procedure (cont.)

- Assume that $b \in \Omega_{IP}$, $\phi(b) > -\infty$ and the algorithm terminates after $k+1$ iterations.
- If u^k is the optimal dual solution to the LP in the final iteration, then

$$F^k(\beta) = \sum_{i=1}^{m} u_i^k \beta_i + \sum_{i=1}^{k} u_{m+i}^k f^i(\beta),$$

is a Chvátal function with $F^k(b) = \phi(b)$ and furthermore, it is optimal to the subadditive dual problem.

Aside: Cuts from Subadditive Functions

- Not only can the cutting plane method be used to construct subadditive dual functions, subadditive functions can yield cuts!
- For *any* subadditive function ψ, the inequality

$$\sum_{i=1}^{n} \psi(a_i) \leq \psi(\beta),$$

  is valid.
- This is not very well-known and there is no clear statement of it with proof in the literature, it is not difficult to prove.

Branch and Cut

- We have seen it it easy to get a strong dual function from branch-and-bound.
- Note, however, that it's not subadditive in general.
- To obtain a subadditive function, we can include the variable bounds explicitly as constraints, but then the function may not be strong.
- For branch-and-cut, we have to take care of the cuts.
  - Case 1: We know the subadditive representation of each cut.
  - Case 2: We know the RHS dependency of each cut.
  - Case 3: Otherwise, we can use some proximity results or the variable bounds.

Case 1

If we know the subadditive representation of each cut:
At a node t, we solve the LP relaxation of the following problem

$$\begin{array}{rrcl} \phi^t(b) = \min & cx & & \\ \text{s.t} & Ax & \geq & b \\ & x & \geq & l^t \\ & -x & \geq & -\mathrm{g}^t \\ & H^t x & \geq & h^t \\ & x & \in & \mathbb{Z}_+^r \times \mathbb{R}_+^{n-r} \end{array}$$

where $\mathrm{g}^t,\ l^t \in \mathbb{R}^r$ are the branching bounds applied to the integer variables and $H^t x \geq h^t$ is the set of added cuts in the form

$$\sum_{j \in I} F_k^t(a_j^k) x_j + \sum_{j \in N \setminus I} \bar{F}_k^t(a_j^k) x_j \ \geq \ F_k^t(\sigma_k(b)) \qquad k = 1, ..., \nu(t),$$

$\nu(t)$: the number of cuts generated so far,
$a_j^k, j = 1, ..., n$: the columns of the problem that the k^{th} cut is constructed from,
$\sigma_k(b)$: is the mapping of b to the RHS of the corresponding problem.

Case 1

Let T be the set of leaf nodes, u^t, $\underline{u}^t$, $\overline{u}^t$ and w^t be the dual feasible solution used to prune $t \in T$. Then,

$$F(\beta) = \min_{t \in T}\{u^t\beta + \underline{u}^t l^t - \overline{u}^t g^t + \sum_{k=1}^{\nu(t)} {w^t}_k F_k^t(\sigma_k(\beta))\}$$

is an optimal dual function, that is, $\phi(b) = F(b)$.

- Again, we obtain a subadditive function if the variables are bounded.
- However, we may not know the subadditive representation of each cut.

Other Methods for Constructing Dual Functions

There are a wide range of other methods for constructing dual functions arising mainly from other solution algorithms.

- Explicit construction
  - The Value Function $\Rightarrow$ discussed today
  - Generating Functions
- Relaxations
  - Lagrangian Relaxation
  - Quadratic Lagrangian Relaxation
  - Corrected Linear Dual Functions
- Solution Algorithms
  - Cutting Plane Method $\Rightarrow$ discussed today
  - Branch-and-Bound Method $\Rightarrow$ discussed today
  - Branch-and-Cut Method $\Rightarrow$ discussed today

Representing/Embedding the Approximations

In practice, we generally want to embed these approximations in other optimization problems and doing this in a computationally efficient way is difficult.

1. The primal bounding functions we discussed can be represented by points of *strict local convexity*.
   - Embedding the approximation using this representation involves explicitly listing these points and choosing one (binary variables).
   - The corresponding continuous part of the function can be generated dynamically or can also be represented explicitly by dual extreme points.
2. The dual bounding functions must generally be represented explicitly in terms of their *polyhedral pieces*.
   - In this case, the points of strict local convexity are implicit and the selection is of the relevant piece or pieces.
   - This yields a much larger representation.

Outline

Warm Starting

- Many optimization algorithms can be viewed as iterative procedures for satisfying optimality conditions (based on duality).
- These conditions provide a measure of “distance from optimality.”
- Warm starting information is additional input data that allows an algorithm to quickly get “close to optimality.”
- In mixed integer linear optimization, the *duality gap* is the usual measure.
- As in linear programming, a feasible dual function may quickly reduce the gap.

What is a feasible dual function and where do we get one?

Valid Disjunctions

- Consider the implicit optimality conditions associated employed in branch and bound.
- Let $\mathcal{P}_1, \ldots, \mathcal{P}_s$ be a set of polyhedra whose union contains the feasible set which differ from $\mathcal{P}$ only in variable bounds.
- Let B^i be the optimal basis for the LP $\min_{x^i \in \mathcal{P}_i} c^\top x^i$.
- Then the following is a valid dual function

$$L(\beta) = \min\{c_{B^i}(B^i)^{-1}\beta + \gamma_i \mid 1 \leq i \leq s\}$$

  where γ_i is a constant factor associated with the nonbasic variables fixed at nonzero bounds.
- A similar function yields an upper bound.
- If this disjunction is the set of leaf nodes of a branch-and-bound tree, this can be used to "warm start" the computation.
- Alternatively, we can use this disjunction to strengthen the root relaxation in some way (disjunctive cuts, etc.).

Sensitivity Analysis

- Primal and dual bounding functions can be evaluated with modified problem data to obtain bounds on the optimal value in the obvious way.
- In the case of a branch-and-bound tree, the function

$$L(\beta) = \min\{c_{B^i}(B^i)^{-1}\beta + \gamma_i \mid 1 \leq i \leq s\}$$

  provides a valid lower bound as a function of the right-hand side.
- The corresponding upper bounding function is

$$U(c) = \min\{c_{B^i}(B^i)^{-1}b + \beta_i \mid 1 \leq i \leq s, \hat{x}^i \in \mathcal{S}\}$$

- These functions can be used for local sensitivity analysis, just as one would do in continuous linear optimization.
  - For changes in the right-hand side, the lower bound remains valid.
  - For changes in the objective function, the upper bound remains valid.
  - One can also make other modifications, such as adding variables or constraints.

Outline

Conclusions

- It is possible to generalize the duality concepts that are familiar to us from continuous linear optimization.
- Making any of it practical is difficult but we will see in the next lectures that this is possible in some cases.

References I

C.E. Blair. A closed-form representation of mixed-integer program value functions. *Mathematical Programming*, 71:127–136, 1995.

C.E. Blair and R.G. Jeroslow. The value function of a mixed integer program: I. *Discrete Mathematics*, 19(2):121–138, 1977a.

C.E. Blair and R.G. Jeroslow. The value function of a mixed integer program: I. *Discrete Mathematics*, 19:121–138, 1977b.

C.E. Blair and R.G. Jeroslow. The value function of a mixed integer program: II. *Discrete Mathematics*, 25:7–19, 1979.

C.E. Blair and R.G. Jeroslow. The value function of an integer program. *Mathematical Programming*, 23:237–273, 1982.

C.E. Blair and R.G. Jeroslow. Constructive characterization of the value function of a mixed-integer program: I. *Discrete Applied Mathematics*, 9:217–233, 1984.

C.E. Blair and R.G. Jeroslow. Constructive characterization of the value function of a mixed-integer program: II. *Discrete Applied Mathematics*, 10:227–240, 1985.

References II

G. Gamrath, B. Hiller, and J. Witzig. Reoptimization techniques for mip solvers. In *Proceedings of the* 14^{th} *International Symposium on Experimental Algorithms*, 2015.

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`.

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, COR@L Laboratory Report 14T-004, Lehigh University, 2014. URL `http://coral.ie.lehigh.edu/~ted/files/papers/MILPValueFunction14.pdf`.

Robert G Jeroslow. Minimal inequalities. *Mathematical Programming*, 17(1):1–15, 1979.

Ellis L Johnson. Cyclic groups, cutting planes and shortest paths. *Mathematical programming*, pages 185–211, 1973.

Ellis L Johnson. On the group problem for mixed integer programming. In *Approaches to Integer Programming*, pages 137–179. Springer, 1974.

Ellis L Johnson. On the group problem and a subadditive approach to integer programming. *Annals of Discrete Mathematics*, 5:97–112, 1979.

N. Kong, A.J. Schaefer, and B. Hunsaker. Two-stage integer programs with stochastic right-hand sides: a superadditive dual approach. *Mathematical Programming*, 108 (2):275–296, 2006.

References IV

T.K. Ralphs and M. Güzelsoy. The SYMPHONY Callable Library for Mixed Integer Programming. In *Proceedings of the Ninth INFORMS Computing Society Conference*, pages 61–76, 2005. doi: 10.1007/0-387-23529-9_5. URL `http://coral.ie.lehigh.edu/~ted/files/papers/SYMPHONY04.pdf`.

T.K. Ralphs and M. Güzelsoy. Duality and Warm Starting in Integer Programming. In *The Proceedings of the 2006 NSF Design, Service, and Manufacturing Grantees and Research Conference*, 2006. URL `http://coral.ie.lehigh.edu/~ted/files/papers/DMII06.pdf`.

L.A. Wolsey. Integer programming duality: Price functions and sensitivity analysis. *Mathematical Programming*, 20(1):173–195, 1981. ISSN 0025-5610.