Duality and Warm Starting in MILP

Ted Ralphs[1]

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

ZIB, Berlin, Germany, 28 September 2016

Outline

Discrete (Linear) Optimization

- For the remainder of the talk, we focus on the case of the mixed integer linear optimization problem (MILP).

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

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

- In this context, we can consider the concepts outlined previously more concretely.
- We first consider the case $r = 0$, which is the case of the (continuous) linear optimization problem (LP).

The LP Value Function

- The price vectors we derived on the previous slide can be seen as the gradients of the *value function*

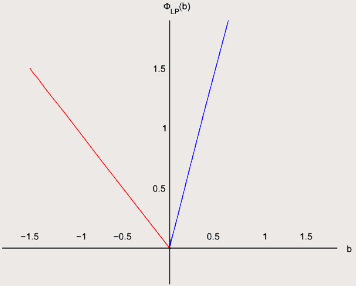
$$\phi_{LP}(\beta) = \min_{x \in \mathcal{S}(\beta)} c^\top x, \qquad \text{(LPVF)}$$

  of an LP, where for a given $\beta \in \mathbb{R}^m$, $\mathcal{S}(\beta) = \{x \in \mathbb{R}^n_+ \mid Ax = \beta\}$.
- We let $\phi_{LP}(\beta) = \infty$ if $\beta \in \Omega = \{\beta \in \mathbb{R}^m \mid \mathcal{S}(\beta) = \emptyset\}$.
- These gradients can be seen as *linear under-estimators* of the value function.
- The dual problems we'll consider are essentially aimed at producing such under-estimators.
- We'll generalize to *non-linear functions*.

LP Value Function Example

Example 1

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



The LP Dual

- To understand the structure of the value function in more detail, first note that it is easy to see ϕ_{LP} is convex.
- Now consider an optimal basis matrix B for the instance (**??**) (still assuming that $r = 0$).
  - The gradient of ϕ_{LP} at b is $\hat{u} = c_B B^{-1}$.
  - Since $\phi_{LP}(b) = \hat{u}^\top b$ and ϕ_{LP} is convex, we know that $\phi_{LP}(\beta) \geq \hat{u}^\top \beta$ for all $\beta \in \mathbb{R}^m$.
- The traditional LP dual problem can be viewed as that of finding a linear function that bounds the value function from below and has maximum value at b.

The LP Dual (cont'd)

- Recall that for any $u \in \mathbb{R}^m$, the following gives a lower bound on $\phi_{LP}(b)$.

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

- With some simplification, we can obtain an explicit form for this function.

$$
\begin{aligned}
g(u) \quad &= \quad min_{x \geq 0} \left[c^\top x + u^\top (b - Ax) \right] \\
&= \quad u^\top b + min_{x \geq 0} (c^\top - u^\top A) x
\end{aligned}
$$

- Note that

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

The LP Dual (cont'd)

- So we have

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

  which is again a linear under-estimator of the value function.

- An LP dual problem is obtained by computing the strongest linear under-estimator with respect to b.

LP Dual Problem

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

The MILP Value Function

- We now generalize the notions seen so far to the MILP case.
- The *value function* associated with the base instance (MILP) is

MILP Value Function

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

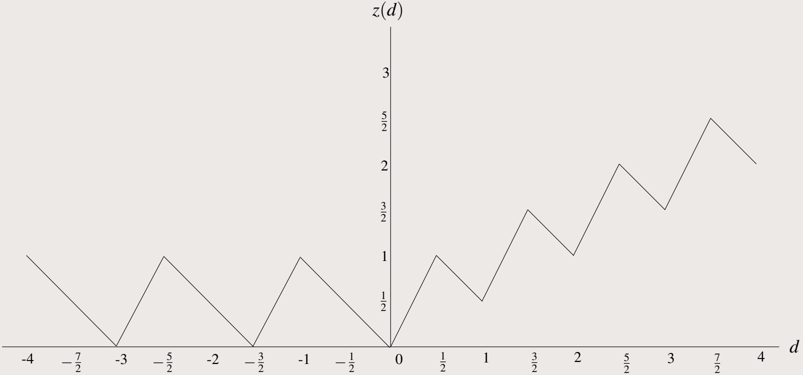
for $\beta \in \mathbb{R}^m$, where $\mathcal{S}(\beta) = \{x \in \mathbb{Z}^r_+ \times \mathbb{R}^{n-r}_+ \mid Ax = \beta\}$.

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

Example: MILP Value Function

Example 2

$$\phi(\beta) = \min \quad \frac{1}{2}x_1 + 2x_3 + x_4$$
$$s.t \quad x_1 - \frac{3}{2}x_2 + x_3 - x_4 = \beta \quad and \tag{1}$$
$$x_1, x_2 \in \mathbb{Z}_+, x_3, x_4 \in \mathbb{R}_+.$$



Dual 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?

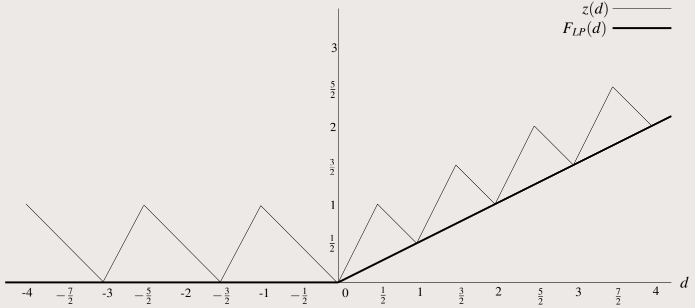
Example: LP Relaxation Dual Function

Example 3

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

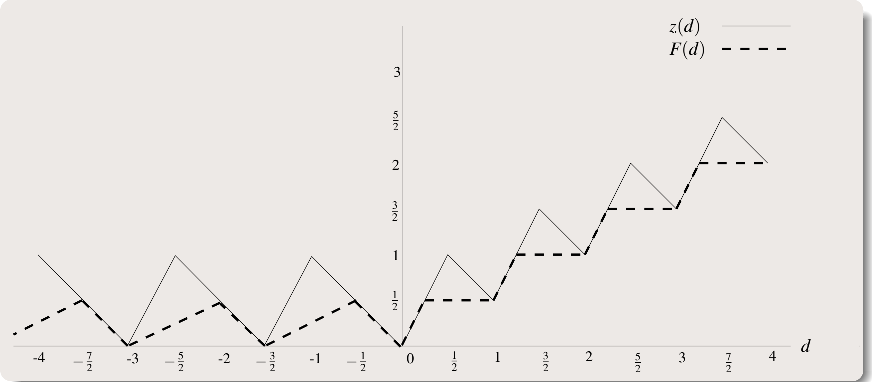
which can be written explicitly as

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



Example: Feasible Dual Functions

Example 4



- 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.

Optimality Conditions

- All widely used algorithms for MILP can be seen as constructing a strong dual function.
- This dual function provides the “proof” that result is correct.
- From the dual problem and these dual functions, we can straightforwardly derive optimality conditions.

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:

$$\begin{aligned} \phi^t(\beta) = \min\ & c^\top x \\ \text{s.t. } & Ax = \beta, \\ & l^t \le x \le u^t, x \ge 0 \end{aligned} \tag{3}$$

The dual at node t:

$$\begin{aligned} \phi^t(\beta) = \max\ & \{\pi^t \beta + \underline{\pi}^t l^t + \bar{\pi}^t u^t\} \\ \text{s.t. } & \pi^t A + \underline{\pi}^t + \bar{\pi}^t \le c^\top \\ & \underline{\pi} \ge 0, \bar{\pi} \le 0 \end{aligned} \tag{4}$$

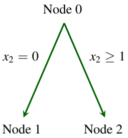
We obtain the following strong dual function:

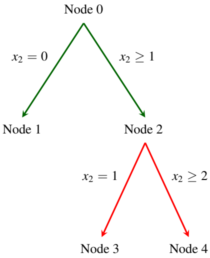
$$\min_{t \in T} \{\hat{\pi}^t \beta + \hat{\underline{\pi}}^t l^t + \hat{\bar{\pi}}^t u^t\}, \tag{5}$$

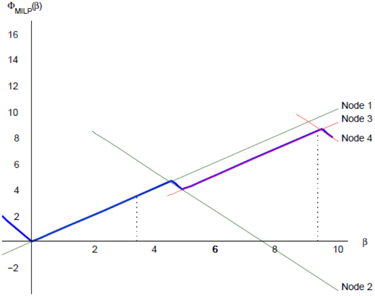
where $(\hat{\pi}^t, \hat{\underline{\pi}}^t, \hat{\bar{\pi}}^t)$ is an optimal solution to the dual (4).

Iterative Refinement

- The tree obtained from evaluating $\phi(\beta)$ yields a dual function strong at β.
- By solving for other right-hand sides, we obtain additional dual functions that can be aggregated.
- These additional solves can be done within the same tree, eventually yielding a single tree representing the entire function.







Tree Representation of the Value Function

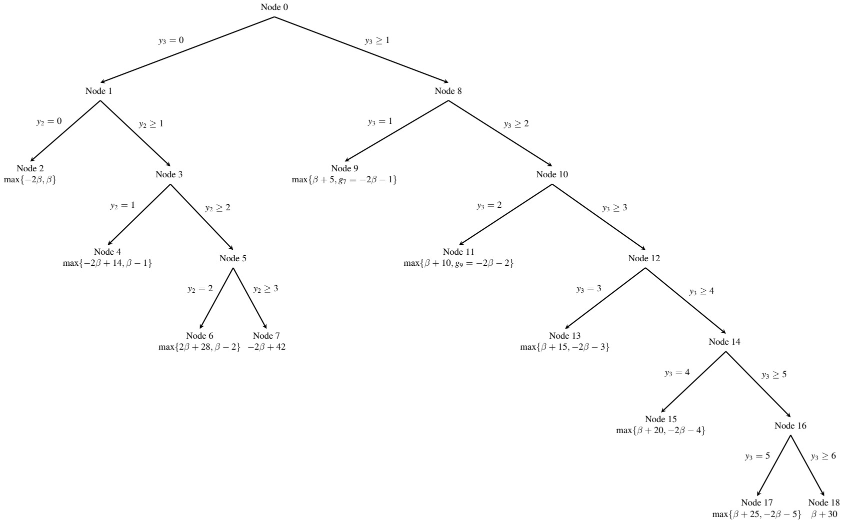
- Continuing the process, we eventually generate the entire value function.
- Consider the strengthened dual

$$\underline{\phi}^*(\beta) = \min_{t \in T} q_{I_t}^\top y_{I_t}^t + \phi_{N \setminus I_t}^t(\beta - W_{I_t} y_{I_t}^t), \tag{6}$$

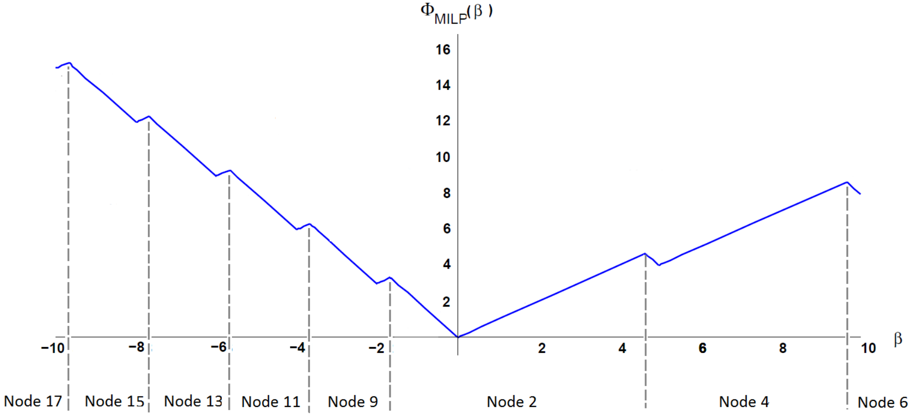
- I_t is the set of indices of fixed variables, $y_{I_t}^t$ are the values of the corresponding variables in node t.
- $\phi_{N \setminus I_t}^t$ is the value function of the linear optimization problem at node t, including only the unfixed variables.

Theorem 1 *[Hassanzadeh and Ralphs, 2014] Under the assumption that* $\{\beta \in \mathbb{R}^{m_2} \mid \phi_I(\beta) < \infty\}$ *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



Describing the Value Function by Parametric Inequalities

- For an ILP, it can be obtained by a finite number of limited operations on elements of the RHS:

(i) rational multiplication
(ii) nonnegative combination
(iii) rounding
(iv) taking the minimum

(i)–(iii): *Chvátal fcns.*
(i)–(iv): *Gomory fcns.*

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 $\max\{f_1, f_2\} \in \mathscr{G}^m$.

Theorem 2 *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 $\mathcal{S}(d) \neq \emptyset$.*

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, 1977]

- There is a Chvátal function that is optimal to the subadditive dual of an ILP with RHS $b \in \Omega_{IP}$ and $\phi(b) > -\infty$.
- The procedure:
  In iteration k, we solve the following LP

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

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

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

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.

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.
- A *starting dual function* may quickly reduce the gap.
- The most obvious choice for a starting function is to use the optimal function from a previous computation.

Exploiting Dual Functions

- There are many ways of exploiting these dual functionsin warm-starting.
- The dual functions are associated with a certain *relaxation*, which can be used in place of the initial LP relaxation.
- This relaxation can be further strengthened and refined by branch-and-cut, as usual.
- We can use these dual functions in the same way that we use the LP dual to perform pre-processing operations, such as bound tightening by *generalized reduced cost*.
- We can also derive *value function cuts* to be added to the LP relaxations to strenghthen initial bounds.

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.

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

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`.

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