Duality for Discrete Optimization: Theory and Applications

Ted Ralphs[1]
Joint work with Suresh Bolusani[1], Scott DeNegre[3], Menal Güzelsoy[2], Anahita Hassanzadeh[4], Sahar Tahernejad[1]

[1]COR@L Lab, Department of Industrial and Systems Engineering, Lehigh University
[2]SAS Institute, Advanced Analytics, Operations Research R & D [3]The Hospital for Special Surgery [4]Climate Corp

Joint Mathematics Meeting, Baltimore, MD, 16 January 2019

LEHIGH UNIVERSITY
COR@L

Outline

Mathematical Optimization

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

Form of a General Mathematical Optimization Problem

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

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.

What is Duality?

- Duality is a central concept from which much theory and computational practice emerges in optimization.
- Many of the well-known "dualities" that arise in optimization are closely connected.
- This talk focuses on one particular kind of duality.

Forms of Duality in Optimization

- NP versus co-NP (computational complexity)
- Separation versus optimization (polarity)
- Inverse optimization versus forward optimization
- Weyl-Minkowski duality (representation theorem)
- Conic duality
- Gauge/Lagrangian/Fenchel duality
- Primal/dual functions/problems

- There are a number of other slide decks and papers about duality on my Web site, including an extended version of this talk.

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.

Outline

(Mixed Integer) Linear Optimization

- We focus on mixed integer linear optimization problems, although the concepts we discuss are much more general.

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

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 make the concepts outlined earlier 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*.
- 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 $\beta \in \mathbb{R}^m$.

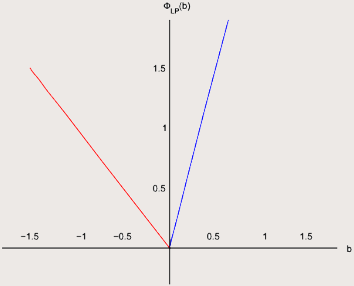
A Small Example

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 for Example 1



Outline

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 $\beta \in \Omega = \{\beta \in \mathbb{R}^m \mid \mathcal{S}(\beta) = \emptyset\}$.

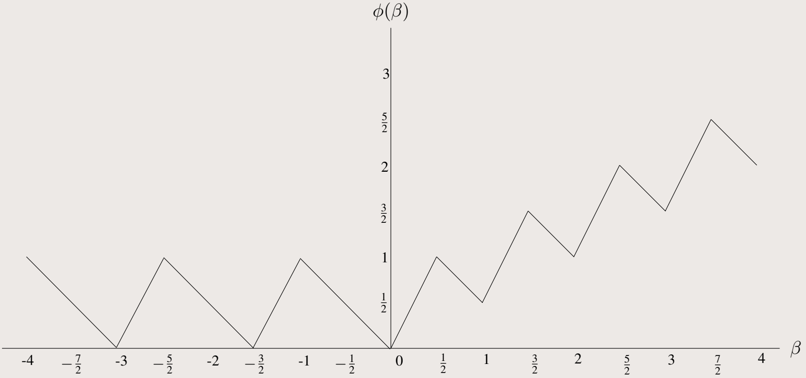
Another Example

Example 2

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

□

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]

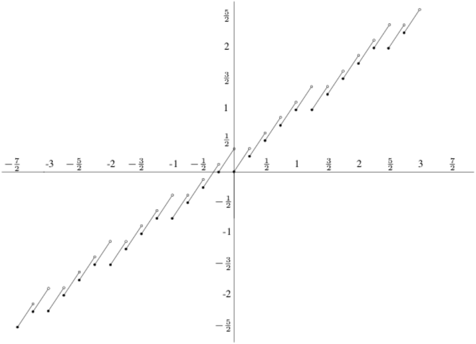
Properties of the MILP Value Function

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

Example 3

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

□

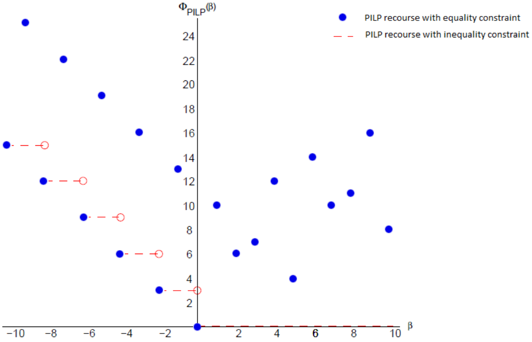


Example: MILP Value Function (Pure Integer)

Example 4

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

□

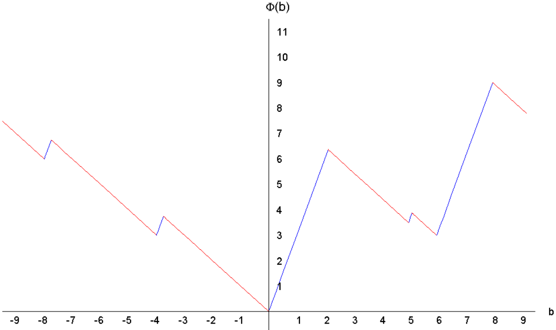


Another Example

Example 5

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

□



Continuous and Integer Restriction of an MILP

Consider the general form of the second-stage 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*:

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

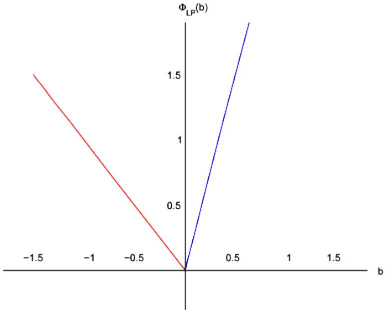
for $I = \{1, \ldots, r_2\}$.

Value Function of the Continuous Restriction

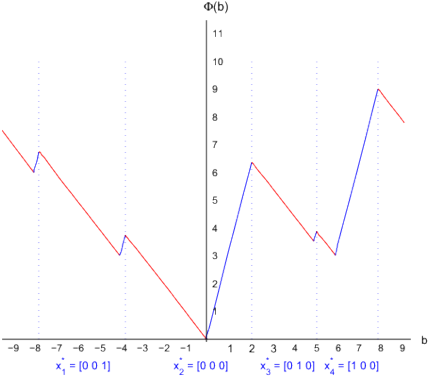
Example 6

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

□



Example 7



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 finite set* $\mathcal{S} \subseteq Y$ *such that*

$$\phi(\beta) = \min_{x_I \in \mathcal{S}} \{c_I^\top x_I + \phi_C(\beta - A_I x_I)\}. \quad (2)$$

Outline

Dual Bounding Functions

- A *dual function* $F : \mathbb{R}^m \to \mathbb{R}$ is one that satisfies $F(\beta) \le \phi(\beta)$ 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).

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

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

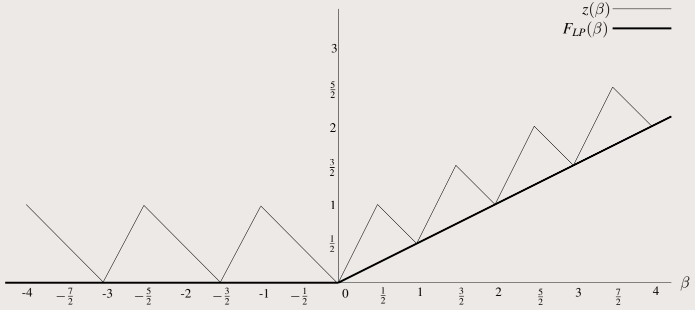
Example 8

$$F_{LP}(d) = \min \quad vd, \quad \text{s.t} \quad 0 \geq v \geq -\tfrac{1}{2}, \text{ and} \quad v \in \mathbb{R}, \tag{3}$$

□

which can be written explicitly as

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



Outline

The Subadditive Dual

By considering that

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

the generalized dual problem can be rewritten as

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

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 $\Upsilon^m \equiv \Gamma^m \equiv \{F \text{ is subadditive} \mid F : \mathbb{R}^m \to \mathbb{R}, F(0) = 0\}$, 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

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

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

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

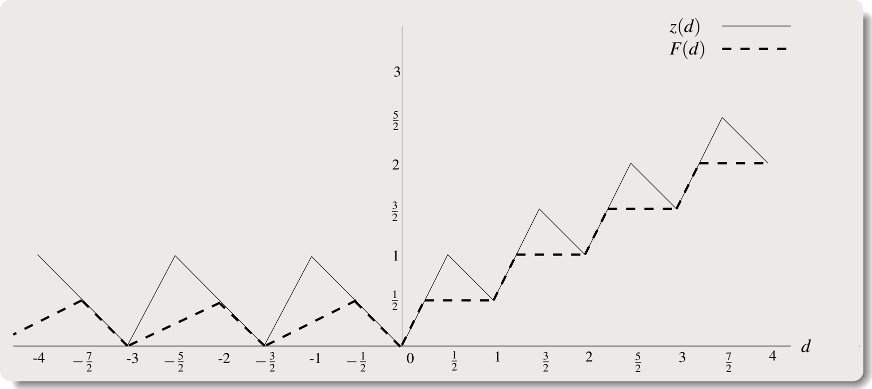
$$\begin{array}{rrl} \max & F(b) & \\ & F(1) & \leq \frac{1}{2} \\ & F(-\frac{3}{2}) & \leq 0 \\ & \bar{F}(1) & \leq 2 \\ & \bar{F}(-1) & \leq 1 \\ & F \in \Gamma^1. & \end{array} \quad .$$

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 9



- 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

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

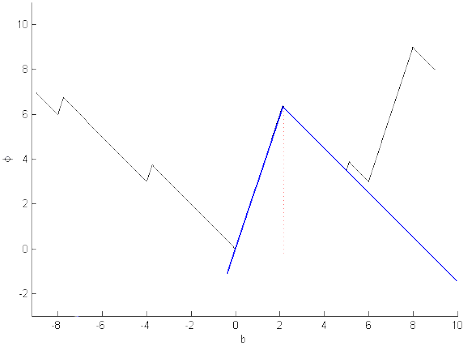
Optimality conditions for (MILP)

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.

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:

$$\begin{aligned} \phi^t(\beta) = \min\ & c^\top x \\ \text{s.t. } & Ax = \beta, \\ & l^t \le x \le u^t, x \ge 0 \end{aligned} \tag{4}$$

The dual at node t:

$$\begin{aligned} \phi^t(\beta) = \max\ & \{\pi^t \beta + \underline{\pi}^t l^t + \bar{\pi}^t u^t\} \\ \text{s.t. } & \pi^t A + \underline{\pi}^t + \bar{\pi}^t \le c^\top \\ & \underline{\pi} \ge 0, \bar{\pi} \le 0 \end{aligned} \tag{5}$$

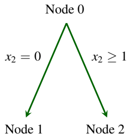
We obtain the following strong dual function:

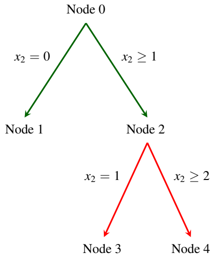
$$\min_{t \in T} \{\hat{\pi}^t \beta + \hat{\underline{\pi}}^t l^t + \hat{\bar{\pi}}^t u^t\}, \tag{6}$$

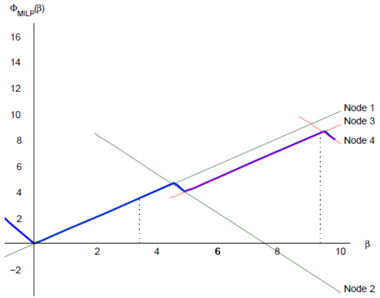
where $(\hat{\pi}^t, \hat{\underline{\pi}}^t, \hat{\bar{\pi}}^t)$ is an optimal solution to the dual (5).

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

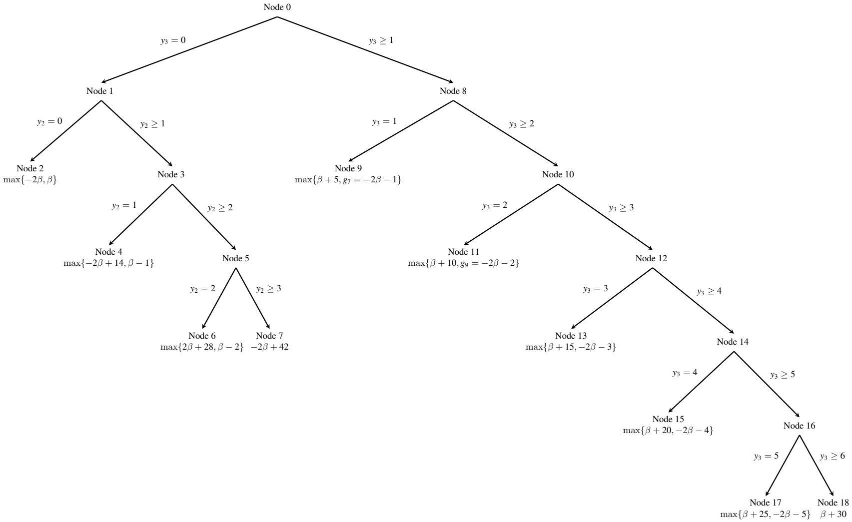
- Continuing the process, we eventually generate the entire value function.
- Consider the strengthened dual

$$\underline{\phi}^*(\beta) = \min_{t \in T} q_{I_t}^\top y_{I_t}^t + \phi_{N \setminus I_t}^t(\beta - W_{I_t} y_{I_t}^t), \tag{7}$$

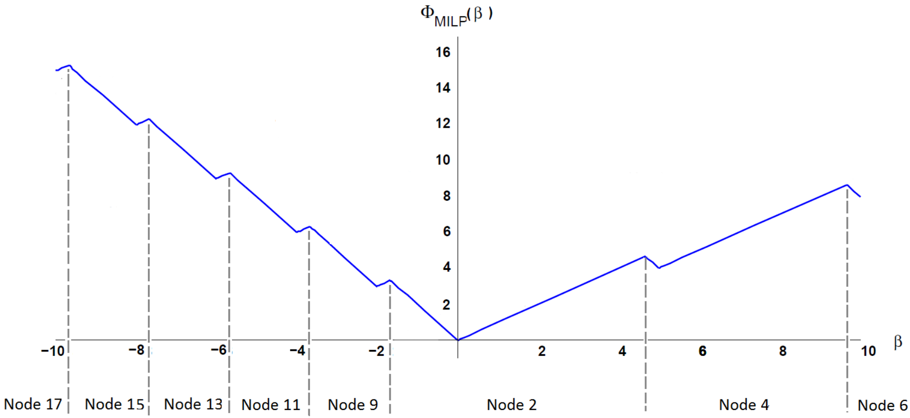
- I_t is the set of indices of fixed variables, $y_{I_t}^t$ are the values of the corresponding variables in node t.
- $\phi_{N \setminus I_t}^t$ is the value function of the linear optimization problem at node t, including only the unfixed variables.

Theorem 2. *[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



Conclusions

- Duality has a wide range of practical uses.
  - Sensitivity analysis
  - Warm starting
  - Parametric optimization
  - Multi-level/stochastic optimization
  - Benders decomposition
  - Parametric inequalities
  - ...
- It is possible to generalize the duality concepts that are familiar to us from continuous linear optimization.
- Making practical use of it is difficult but 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.