Integer Programming
ISE 418

Lecture 8

Dr. Ted Ralphs

Reading for This Lecture

- Nemhauser and Wolsey Sections II.3.1-11.3.3, II.3.6
- "Duality for Mixed-Integer Linear Programs," Güzelsoy and Ralphs

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

Setup

- We focus on mixed integer linear optimization problems, although the concepts we discuss are much more general.
- *Note we are switching to the equality form of constraints (the* standard form *for LPs) and minimization for this lecture.*
- Thus, we consider the problem

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

where

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

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

Economic Interpretation

- The economic viewpoint interprets the variables as representing possible *activities* in which one can engage at specific numeric levels.
- We interpret the constraints as representing available *resources* so that the i^{th} row a^i of A represents the rate at which resource i will be consumed by each activity.
- Similarly, the j^{th} column A_j of A represents the rate at which activity j consumes each resource.
- The feasible set $\mathcal{S}$ represents combinations of activities that can engaged in simultaneously, given the vector of resources b.
- The space in which $\mathcal{S}$ and the vectors of activities live is the *primal space*.

The Dual Space

- We may also consider the problem from the point of view 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 activity?
  - What new activities would provide additional profit?
- The *dual space* is associated with *resources* and and is the space in which we can frame these questions.
- The dual space has a relatively straightforward economic interpretation when the activity levels exist on a continuum (the LP case).
- The dual space is is not as easy to interpret once we introduce the idea that the activity levels must be from a discrete set.

Quick Review of Concepts from LP

- Recall that there always exists an optimal solution that is *basic*.
- We construct basic solutions by
  - Choosing a *basis* B of m linearly independent columns of A.
  - Solving the system $Bx_B = b$ to obtain the values of the *basic variables*.
  - Setting remaining variables to value 0.
- If $x_B \geq 0$, then the associated basic solution is *feasible*.
- With respect to any basic feasible solution, it is easy to determine the impact of increasing a given activity.
- The *reduced cost*

$$\bar{c}_j = c_j - c_B^\top B^{-1} A_j.$$

  of (nonbasic) variable j tells us how the objective function value changes if we increase the level of activity j by one unit.
- It follows that a basic feasible solution is optimal if and only if the reduced costs are all non-positive.

Marginal Prices

- From the resource (dual) perspective, the quantity $u = c_B^\top B^{-1}$ is a vector that tells us the marginal economic value of each resource.
- In other words, $c_B^\top B^{-1}\Delta b$ is the marginal amount by which the objective value would change if we augmented the available resources by Δb.
- Thus, u can be interpreted as a vector of (linear) *prices* for the resources, with $u^\top b$ the economic worth of the bundle b.
- This give us an economic interpretation of strong duality.
- There exist prices u^* for which the value $(u^*)^\top b$ of the bundle of resources b is the same as the profit $c^\top x^*$ from the optimal activity vector $x^* \in \mathcal{S}$.
- In economics, u^* are the *market-clearing prices*.

The LP Value Function

- To construct a duality theory for MILPs, we need a more general notion of "dual prices."
- The first step in understanding this more general point of view is to consider the so-called *value function*, defined by

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

  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.

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 **exactly equals** 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.

$$\min \sum_{j=1}^{n} p_j x_j$$

$$\text{s.t.} \sum_{j=1}^{n} w_j x_j = W$$

$$0 \le x_i \le 1 \qquad \forall i$$

- What is the optimal solution?

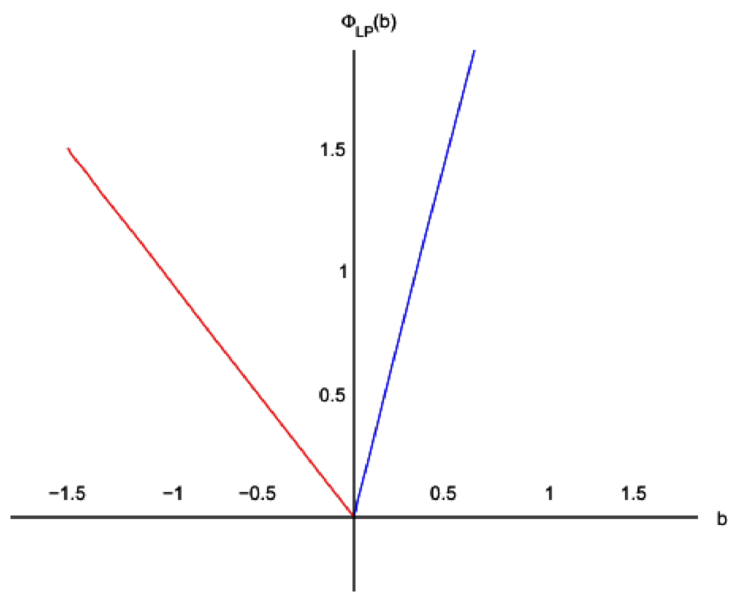
Example (cont'd)

Example 1

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

□

Figure 1: Value Function for Example 1



Economic Interpretation of the Value Function

- Consider a member $u \in \partial\phi_{LP}(b)$ of the subdifferential of ϕ_{LP} at b.
- Since ϕ_{LP} is convex, its (sub)gradients are *linear under-estimators* and can be used to derive bounds on the optimal value for any $\beta \in \mathbb{R}^m$.
- The quantity $u^\top \Delta b$ represents (an estimate of) the marginal change in the optimal value if we change the resource level by Δb.
- In other words, u can be interpreted as a vector of the *marginal values of the resources*.
- The (sub)gradient u of ϕ thus seems to play a role similar to a solution to the LP dual.
- This is not a coincidence!
- The subdifferential at 0 *is* the feasible set for the LP dual and the subdifferential at b is the set of optimal solutions of the associated dual!
- We can observe these properties in Example 1.
  - The dual solutions of this LP are exactly the subdifferential at 0.
  - The gradients are the optimal dual solutions for $\beta \neq 0$.

The Dual Optimization Problem

- For convex functions f, the subdifferential at x is exactly the set of linear underestimators that are tangent to f at x.
- We can thus determine a (sub)gradient of ϕ_{LP} at b using optimization: find the subgradient that yields the maximum bound at b.
- 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 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

$$g(\mu, \beta) = \begin{cases} \mu^\top \beta, & \text{if } c^\top - \mu^\top A \geq \mathbf{0}^\top, \\ -\infty, & \text{otherwise,} \end{cases} \tag{1}$$

  with parameters μ and β, then for fixed $u \in \mathbb{R}^m$ such that $c^\top \geq u^\top A$. $g(u, \beta)$ 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.

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

- (LPD) is the usual LP dual problem and we have shown that its optimal solutions are the (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.

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

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.

$$\mathcal{E} \equiv \left\{ c_B A_E^{-1} \mid E \text{ is the index set of a dual feasible bases of A} \right\} \qquad (2)$$

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

What is the Importance in This Context?

- 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
  - $c^\top x^* = u^\top b$,

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

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

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

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

Example

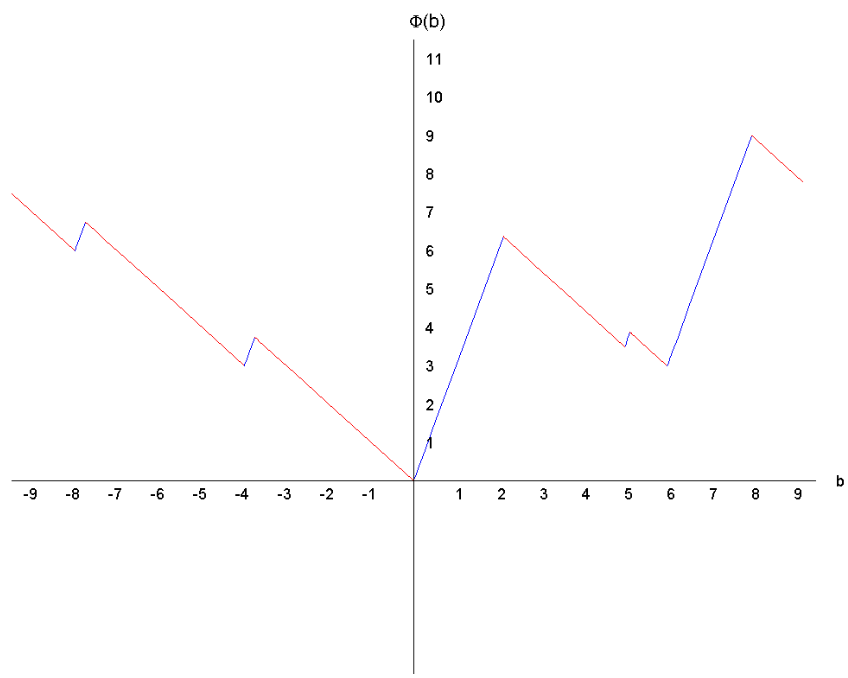
Example 2

$$\phi(\beta) = \min\ 3x_1 + \frac{7}{2}x_2 + 3x_3 + 6x_4 + 7x_5 + 5x_6$$

$$\text{s.t. } 6x_1 + 5x_2 - 4x_3 + 2x_4 - 7x_5 + x_6 = \beta$$

$$x_1, x_2, x_3 \in \mathbb{Z}_+,\ x_4, x_5, x_6 \in \mathbb{R}_+$$

□



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

Continuous and Integer Restriction of an MILP

Consider the general form of the value function

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

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

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

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

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

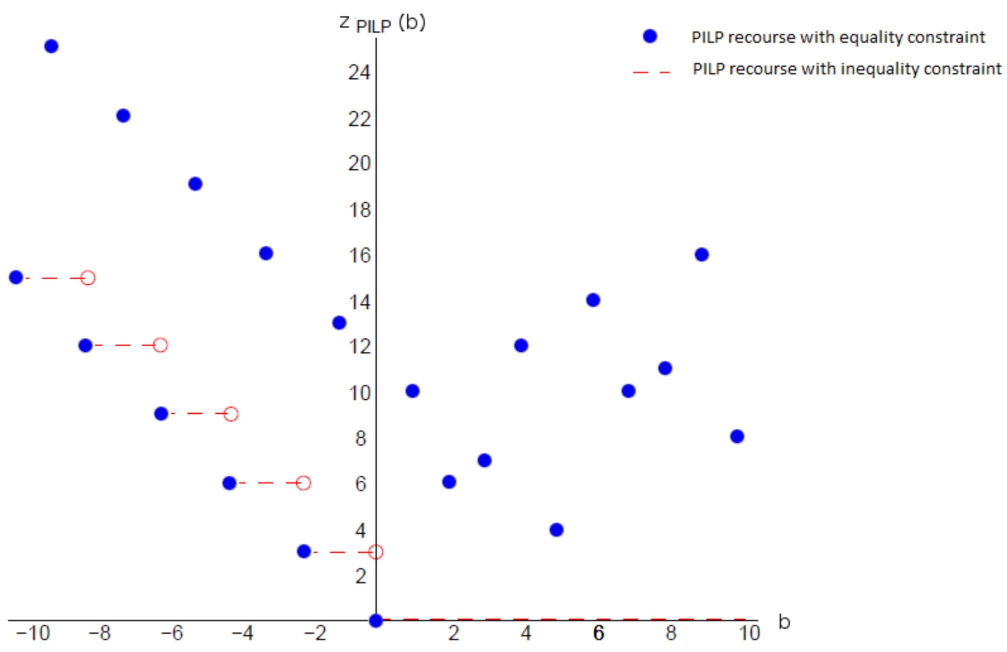
for $I = \{1, \ldots, r_2\}$.

Value Function of Integer Restriction (Example 2)

Example 3

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

□

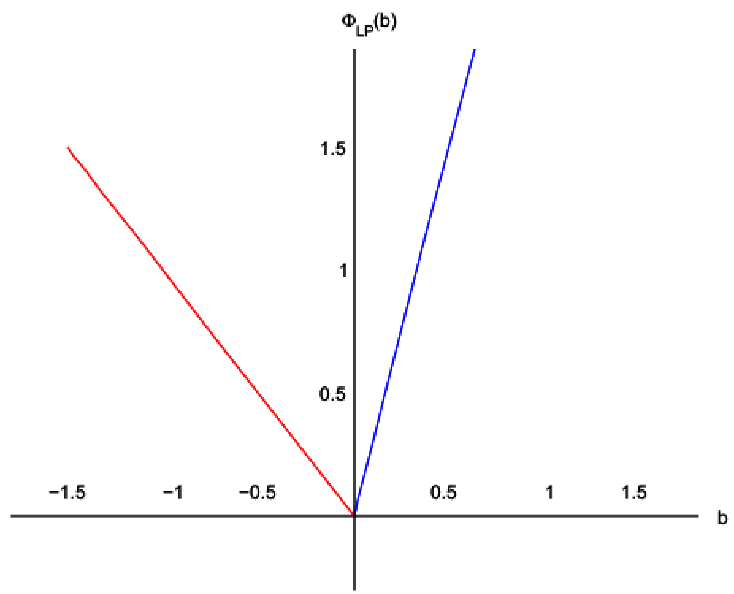


Value Function of Continuous Restriction (Example 2)

Example 4

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

□



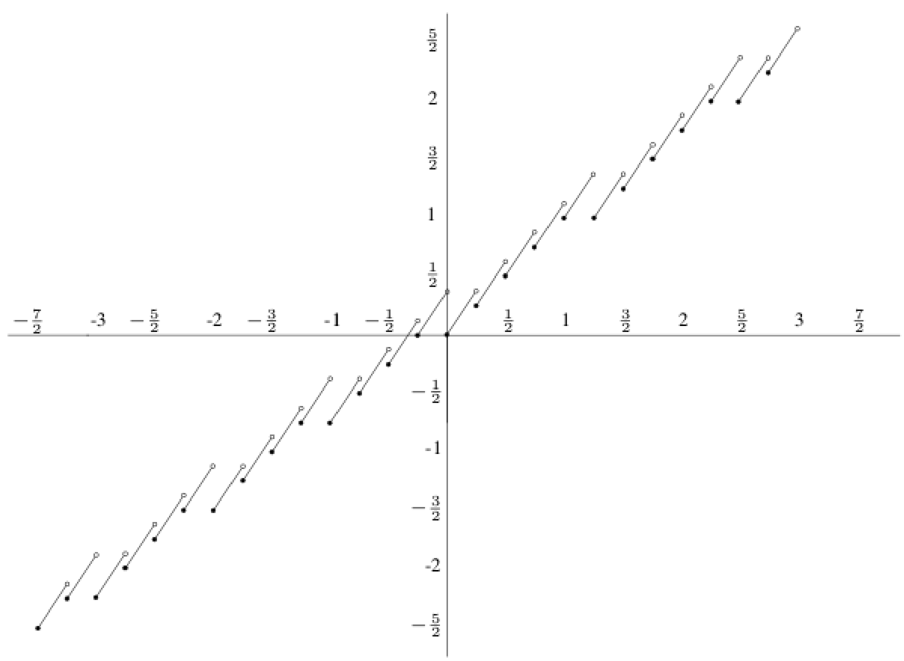
General Properties of the MILP Value Function

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

Example 5

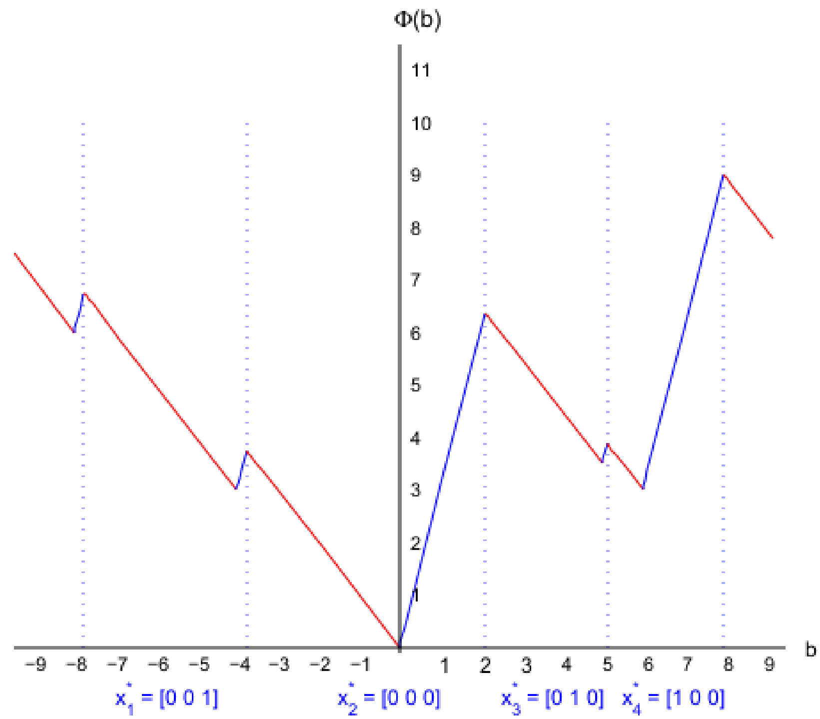
$$
\begin{aligned}
\phi(\beta) = \min\ & x_1 - \frac{3}{4}x_2 + \frac{3}{4}x_3 \\
\text{s.t. } & \frac{5}{4}x_1 - x_2 + \frac{1}{2}x_3 = \beta \\
& x_1, x_2 \in \mathbb{Z}_+,\ x_3 \in \mathbb{R}_+
\end{aligned}
\tag{Ex2.MILP}
$$

□



Points of Strict Local Convexity (Finite Representation)

Example 6



Theorem 1. *Under the assumption that* $\{\beta \in \mathbb{R}^{m_2} \mid \phi_I(\beta) < \infty\}$ *is finite, there exists a (minimal) finite set* $\mathcal{S}$ *such that*

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

Generalized Dual Problem

- A *dual function* $F : \mathbb{R}^m \rightarrow \mathbb{R}$ 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-EQ).

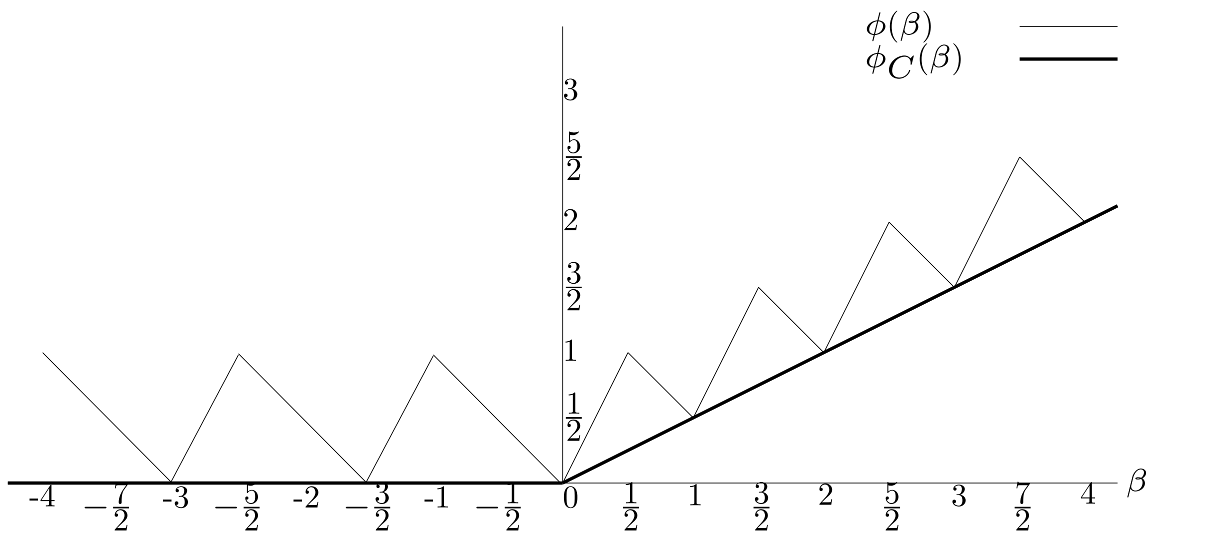
$$\max \{F(b) : F(\beta) \leq \phi(\beta),\ \beta \in \mathbb{R}^m, F \in \Upsilon^m\} \qquad \text{(D)}$$

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

- 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 $\Upsilon^m \equiv \{f \mid f : \mathbb{R}^m \rightarrow \mathbb{R}\}$. Why?

Example: LP Relaxation Dual Function

- The simplest dual function for any MILP is the value function of its LP relaxation.
- It is easy to show that such a function is the convex envelope of the MILP value function.
- It is the strongest convex dual function we can construct.



The Subadditive Dual

By considering that

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

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!

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

Theorem 2. [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.

- We can generalize other properties obtained using LP duality.
  - Complementary slackness conditions
  - Farkas Lemma

Optimality Conditions

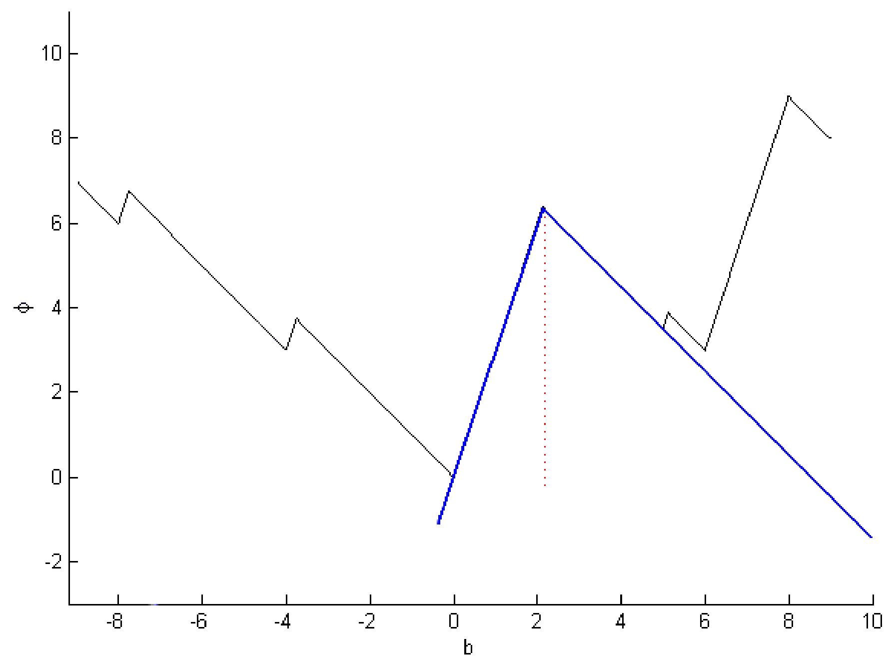
- One reason the dual problem is important is because it gives us a set of *optimality conditions*.

  Theorem 3. [Optimality conditions for (MILP-EQ)] *If* $x^* \in \mathcal{S}$, F^* *is feasible for* (D), *and* $c^\top x^* = F^*(b)$, *then* x^* *is an optimal solution to* (MILP-EQ) *and* F^* *is an optimal solution to* (D).

- These are the optimality conditions achieved in the branch-and-bound algorithm for MILP that prove the optimality of the primal solution.
- The branch-and-bound tree encodes a solution to the dual.

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

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

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

By LP duality, we then have that:

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

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

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

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

Example: Dual Function from Branch and Bound

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

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

- Suppose we evaluate $\phi(5.5)$ by solving the instance with fixed right-hand side by LP-based branch-and-bound.
- Solving the root LP relaxation, we obtain a solution in which $x_2 = 1.1$ and the optimal dual 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 $x_2 \leq 1$ and $x_2 \geq 2$, respectively.
- The problem is solved after this single branching, since $c_6/a_6 < c_1/a_1$ so that $x_1 = x_3 = 0$ in any optimal solution when $\beta > 0$.

Example: Dual Function from Branch and Bound

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

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

- Note that we have added the bound constraints explicitly and the upper bounds on all variables are initially taken to be a "big-M" value.

- Then, the following are the nodal dual functions.

$$\underline{\phi}^0_{\mathrm{LP}}(\beta) = 0.8\beta$$

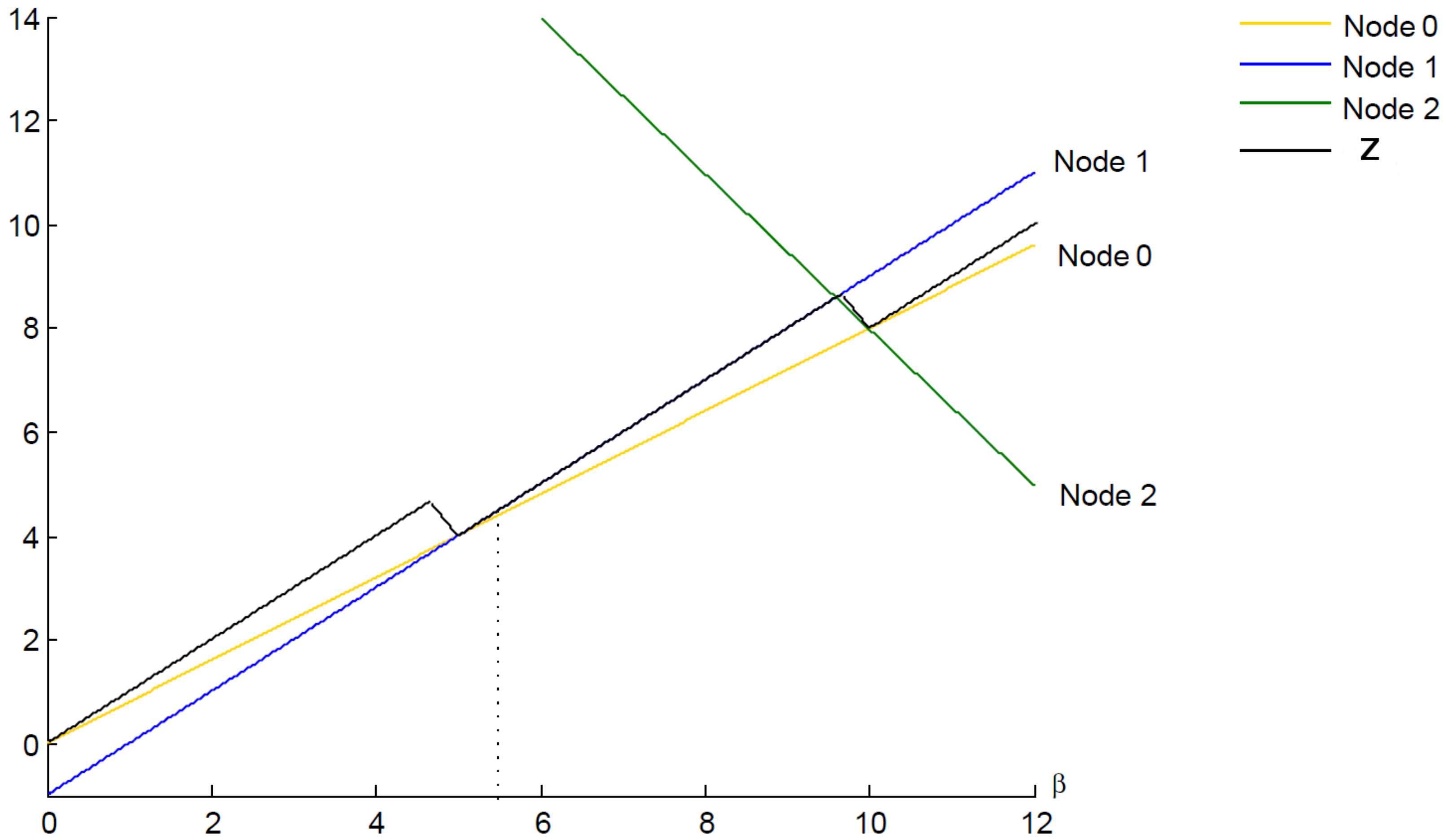
$$\underline{\phi}^1_{\mathrm{LP}}(\beta) = \beta - 1$$

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

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

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

Example: Visualizing the Dual Function



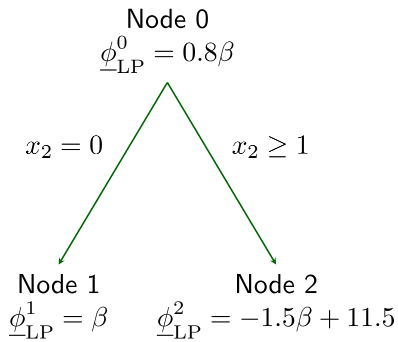
Strengthening the Dual Function

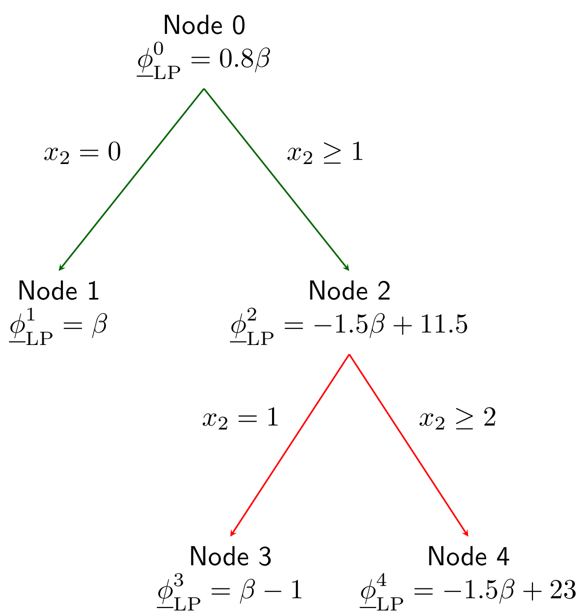
- The dual function can be strengthened by noting that the dual feasible region is the same for all nodes.
- We can therefore approximate the nodal value function by taking a max over all known dual solutions.
- Then we obtain

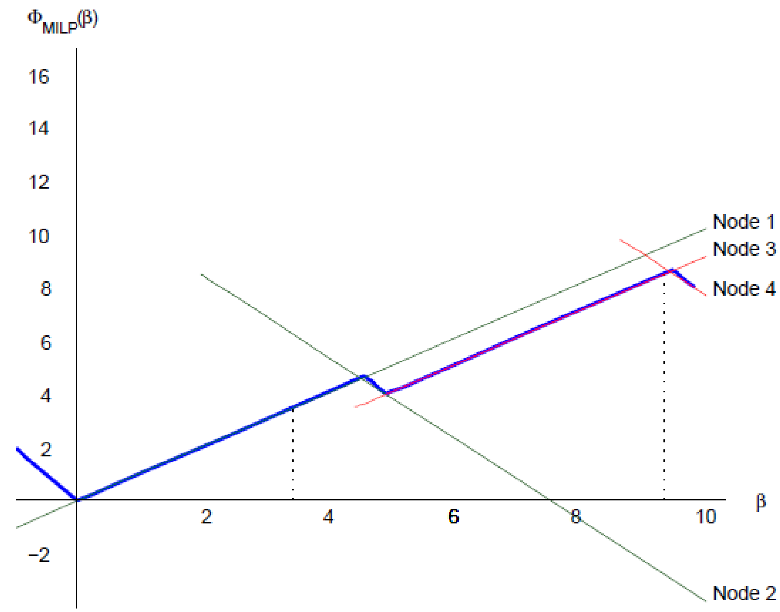
$$\min\{\max\{0.8\beta, \beta - 1, -1.5\beta\}, \max\{0.8\beta - M, \beta, -1.5\beta + 23\}\}$$

- Note the M, which is present because $\bar{\pi}_2^1 = -1$ and the implicit upper bound on x_2 is M in Node 1.
- By evaluating ϕ at a different right-hand side, but using the same tree as a starting point, we can begin to approximate the full value function.
- On the next slide, we show how evaluating at multiple right-hand sides can further improves the approximation.

Example: Iterative Refinement







Recall again that these pictures are for minimization!

Tree Representation of the Value Function

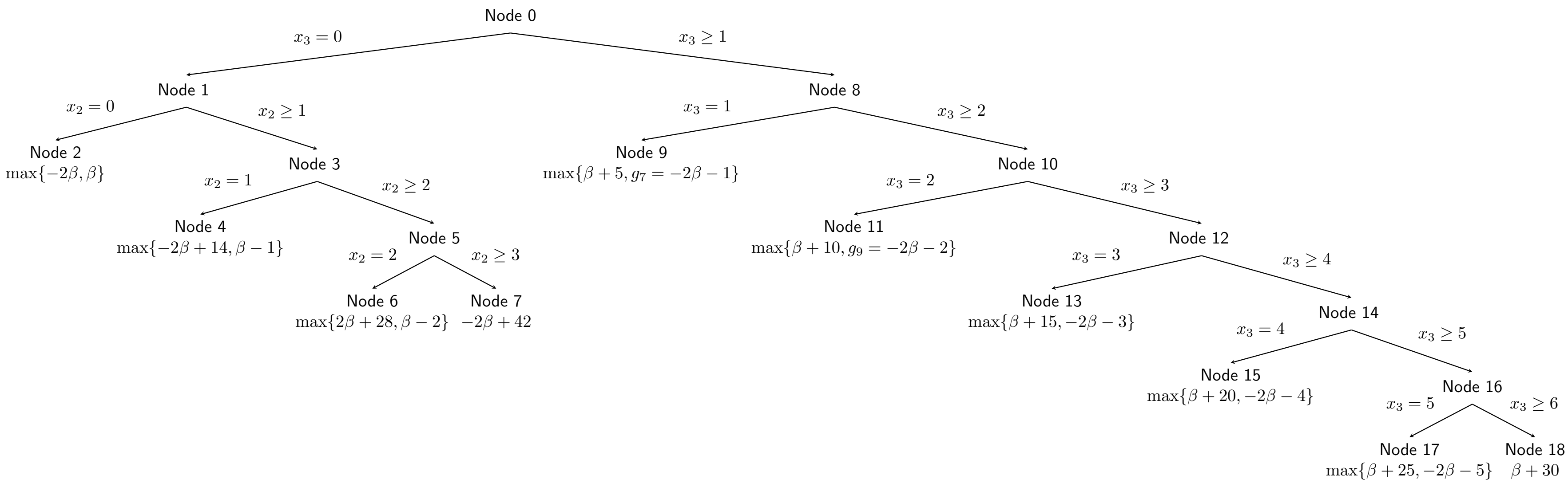
- Continuing the process, we eventually generate the entire value function.
- Consider the strengthened dual

$$\underline{\phi}^*(\beta) = \min_{t \in T} c_{I_t}^\top x_{I_t}^t + \phi_{N \setminus I_t}^t(\beta - A_{I_t} x_{I_t}^t),$$

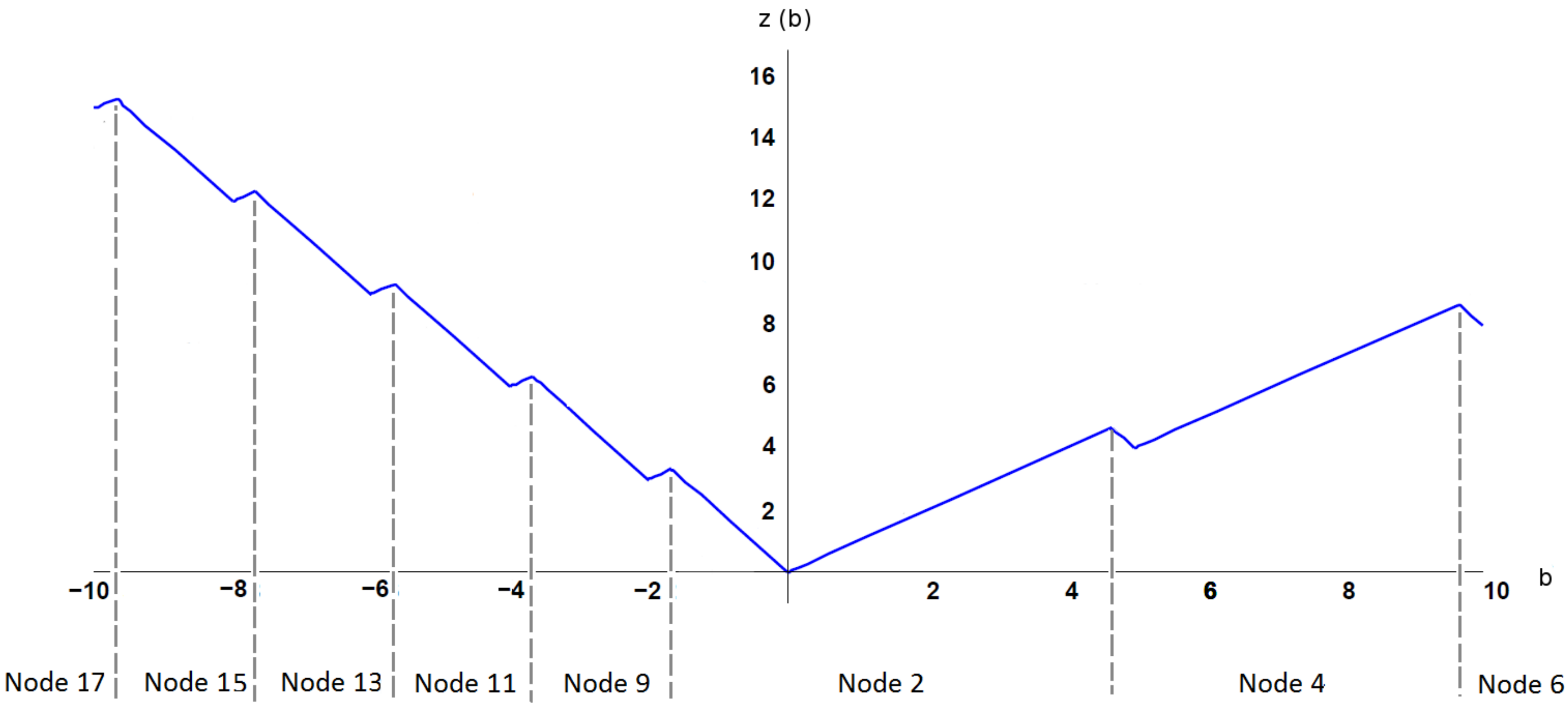
- I_t is the set of indices of fixed variables, $x_{I_t}^t$ are the values of the corresponding variables in node t.
- $\phi_{N \setminus I_t}^t$ is the value function of the linear optimization problem at node t, including only the unfixed variables.

Theorem 4. *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



Dual Functions from Disjunction

- The definition of the dual function arising from the branch-and-bound tree doesn't actually exploit the structure of the tree.
- To construct such a dual function, we only need to know that the collection of subproblems associated with the leaf nodes comprise a partition of the original feasible region.
- In fact, any such collection will do—it need not even have arisen from branch-and-bound.
- All we extract from the final tree is the *disjunction* it encodes
- Any disjunction can be used to derive a dual function of the form (BB.D).
- A disjunction that proves optimality of a given solution (by producing a sufficiently strong dual bound) is an *optimal disjunction*.
- We'll explore this topic further in the next lecture.

A More General Point of View: Theory of Computation

- In the theory of computation, it is common to consider only problems for which the result of a computation is "YES" or "NO."
- We can interpret such a problem as that of trying to prove a theorem, which must be either "TRUE" or "FALSE".
- By viewing the proof as part of the output, it is easier to see that this class of problems is in fact very rich.
- The notion of a proof is fundamental to how problems are classified in the theory of computational complexity.
- Higher complexity means longer proofs are expected.
- In the theory of computation, the *formal proof* that the answer given by an algorithm is correct is sometimes called a *certificate*.
- Formal proofs are constructed using the logic of a specific *formal system*.
- Mathematical optimization can be viewed as one such formal system.

Theorems About Sets

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

$$\mathcal{S} \stackrel{?}{=} \emptyset. \tag{3}$$

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

$$\mathcal{S}(f, K) := \{x \in \mathcal{S} \mid f(x) < K\} \stackrel{?}{=} \emptyset \tag{4}$$

is the *decision version* of the optimization problem

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

Constructing Proofs

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

$$\mathcal{S} \neq \emptyset \Leftrightarrow \exists x \in \mathcal{S} \tag{5}$$

$$\mathcal{S} = \emptyset \Leftrightarrow \forall x \in \mathbb{Q}^n \; x \notin \mathcal{S} \Leftrightarrow \forall x \in \mathbb{Q}^n \; x \in \bar{\mathcal{S}} \tag{6}$$

- 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.
  - 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.

$$P(x)\ \forall x \in \mathcal{S} \Leftrightarrow \neg[\exists x \in \mathcal{S}\ \neg P(x)] \Leftrightarrow \neg \bigvee_{x\in\mathcal{S}} \neg P(x) \Leftrightarrow \bigwedge_{x\in\mathcal{S}} P(x) \quad (7)$$

$$\exists x \in \mathcal{S} : P(x) \Leftrightarrow \neg[\forall x \in \mathcal{S}\ P(x)] \Leftrightarrow \neg \bigwedge_{x\in\mathcal{S}} \neg P(x) \Leftrightarrow \bigvee_{x\in\mathcal{S}} P(x) \quad (8)$$

- 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

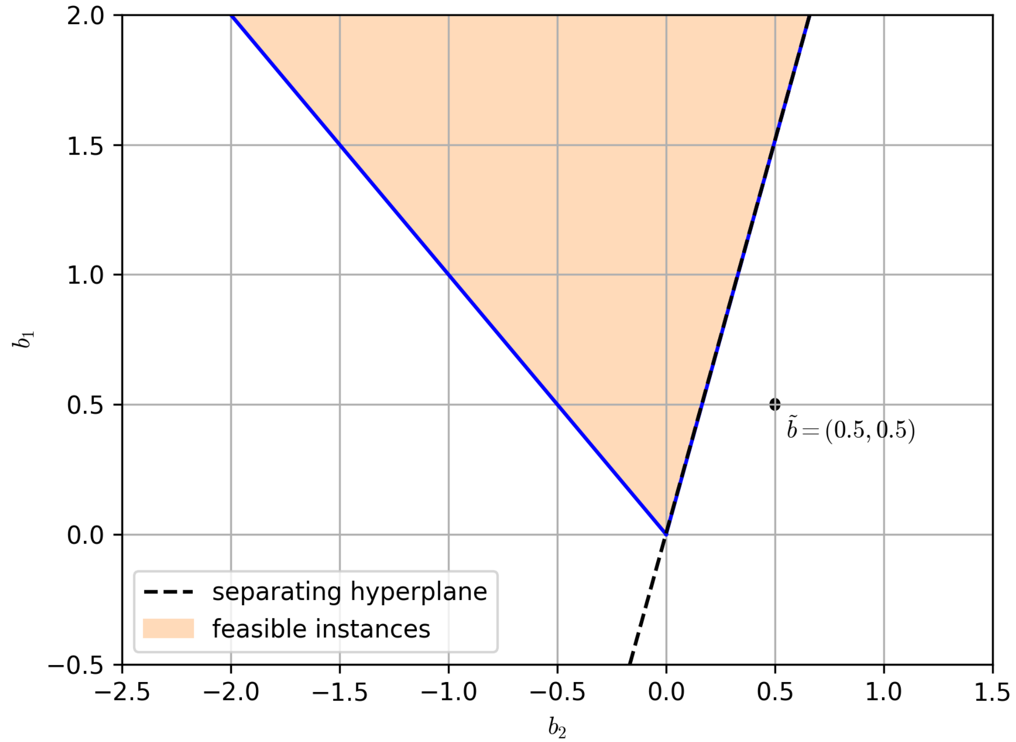
- In the case of convex sets, we can use a duality argument to obtain short proofs of emptiness.
- Consider the case of a polyhedron.

$$\mathcal{P} = \{x \in \mathbb{Q}^n_+ \mid Ax = \tilde{b}\} \tag{9}$$

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

Example

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



Meaning from Duality

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

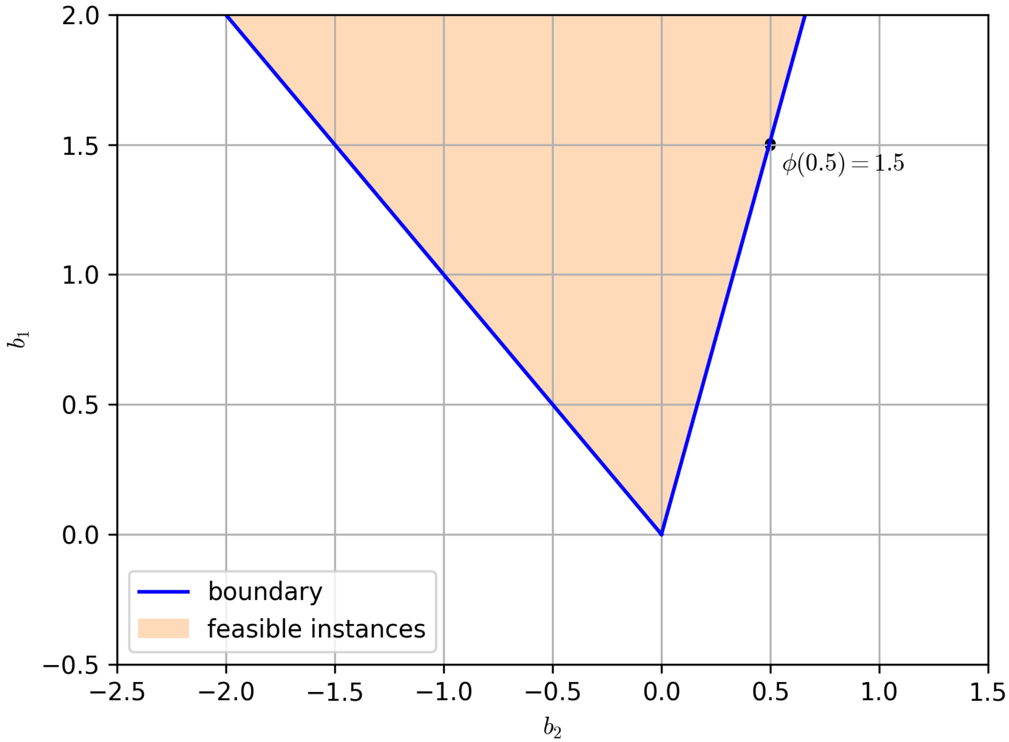
- The fact that one of these statements is universally quantified is the reason why short proofs of optimality cannot be expected in general.

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 $N = \{2, \ldots, m\}$ and $\tilde{b}_N \in \mathbb{Q}^{m-1}$ be all but the first element of $\tilde{b}$.
- The problem of finding the optimal value can then be recast as $b^* = \min\{b_1 \in \mathbb{Q} \mid b \in \mathcal{C}\}$.
- To prove optimality, we need to show that $(b^*, \tilde{b}_N)$ is not only a member of $\mathcal{C}$, but on its *boundary*.
- The proof is only slightly modified: $\exists u \in \mathbb{Q}^m, A^\top u \leq 0, (b^*, \tilde{b}_N)^\top u = 0, u_1 < 0$.
  - 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 $\mathcal{C}$.
- The vector u is a solution to the usual LP dual problem.

Example

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

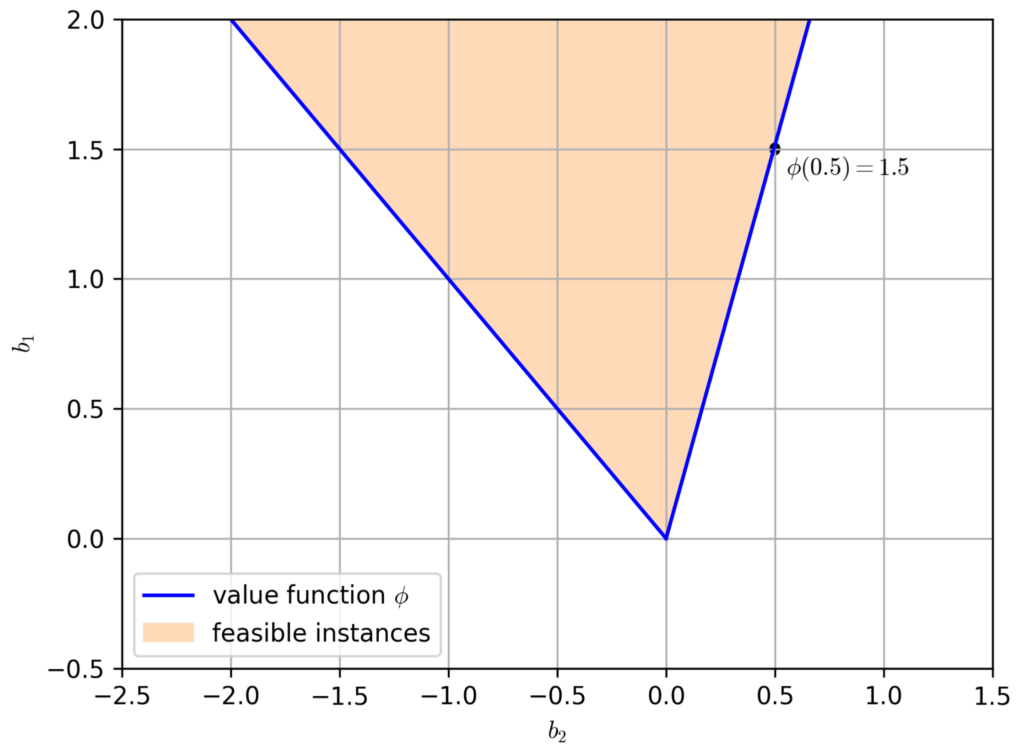


Interpreting

- It is not only the theorems in the class that are parametrically related, the proofs are themselves parametric.
- This parameterization can provide intuition to a human interpreter.
- For example, the boundary of the cone $\mathcal{C}$ 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*).
- Both of these interpretations provide a human-interpretable meaning to the result.
- The “dual price” of a given constraint has an economic interpretation when the constraints are interpreted as allocating resources.
- This provides a language for describing the result of a convex optimization to a human.

Example

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



The Value Function of an MILP

- The tree generated by an LP-based branch and bound encodes a kind of dual proof that generalizes the one seen earlier for LPs.
- Let us define the *value function* of an MILP as

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

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

- As before, this function can be viewed as encoding the boundary between between feasible and infeasible sets of instances.
- This boundary is the value function.

Example

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