Lecture 4: Lagrangian duality for discrete optimization

Ann-Brith Strömberg
Michael Patriksson

22 March 2004

A reminder of nice properties in the convex case

- Example I (explicit dual) ($\boldsymbol{x}^* = (2,2), \mu^* = 4, f^* = 8$)

$$\begin{aligned} f^* = \text{minimum} \quad & f(\boldsymbol{x}) = x_1^2 + x_2^2, \\ \text{subject to} \quad & g(\boldsymbol{x}) = -x_1 - x_2 + 4 \leq 0, \\ & x_1, x_2 \geq 0 \end{aligned}$$

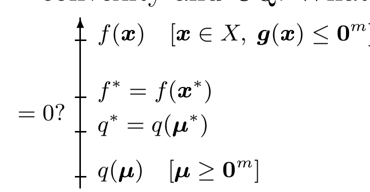
- Let $X := \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1,\ x_2 \geq 0\} = \mathbb{R}^2_+$.

$$\begin{aligned} q(\mu) &= \underset{\boldsymbol{x} \in X}{\text{minimum}} \left\{ x_1^2 + x_2^2 + \mu \cdot (-x_1 - x_2 + 4) \right\} \\ &= 4\mu + \underset{\boldsymbol{x} \in X}{\text{minimum}} \left\{ x_1^2 + x_2^2 - \mu x_1 - \mu x_2 \right\} \\ &= 4\mu + \underset{x_1 \geq 0}{\text{minimum}} \left\{ x_1^2 - \mu x_1 \right\} + \underset{x_2 \geq 0}{\text{minimum}} \left\{ x_2^2 - \mu x_2 \right\}. \end{aligned}$$

- For a fixed value of $\mu \geq 0$, the minimum of $L(\boldsymbol{x}, \mu)$ over $\boldsymbol{x} \in X$ is attained at $x_1(\mu) = \frac{\mu}{2}$, $x_2(\mu) = \frac{\mu}{2}$.
- This implies that $q(\mu) = L(\boldsymbol{x}(\mu), \mu) = \ldots = 4\mu - \frac{\mu^2}{2}$ for all $\mu \geq 0$. The dual function q is concave and differentiable.
- $f^* = f(\boldsymbol{x}^*) = 8 = q^*$.
- $\boldsymbol{x}(\mu^*) = \boldsymbol{x}^*$.

Weak duality! Strong duality?

- We know that the primal optimal solution is obtained from the Lagrangian dual optimal solution under convexity and CQ. What happens otherwise?



- How do we generate optimal solutions in the case of a positive duality gap?

A first example where the duality gap is non-zero

- Example II ($\boldsymbol{x}^* = (0,1,1)^{\mathrm{T}}$, $f^* = 17$)

$$\begin{aligned} f^* = \text{minimum} \quad & f(\boldsymbol{x}) = 3x_1 + 7x_2 + 10x_3, \\ \text{subject to} \quad & x_1 + 3x_2 + 5x_3 \geq 7, \\ & x_j \in \{0,1\}, \quad j = 1,2,3. \end{aligned}$$

- Let $X := \{\, \boldsymbol{x} \in \mathbb{R}^3 \mid x_j \in \{0,1\},\ j = 1,2,3 \,\} = B^3$.
- Let $g(\boldsymbol{x}) := 7 - x_1 - 3x_2 - 5x_3$.

$$\begin{aligned} q(\mu) &= 7\mu + \underset{\boldsymbol{x} \in X}{\text{minimum}}\ \{(3-\mu)x_1 + (7-3\mu)x_2 + (10-5\mu)x_3\} \\ &= 7\mu + \underset{x_1 \in \{0,1\}}{\text{minimum}}\ \{(3-\mu)x_1\} + \underset{x_2 \in \{0,1\}}{\text{minimum}}\ \{(7-3\mu)x_2\} \\ &\quad + \underset{x_3 \in \{0,1\}}{\text{minimum}}\ \{(10-5\mu)x_3\} \end{aligned}$$

- $X(\mu)$ is obtained by setting $x_j(\mu) = 1(0)$ when the objective coefficient is $< (\geq) 0$.

Subproblem solutions and the dual function

$\mu \in$	$x_1(\mu)$	$x_2(\mu)$	$x_3(\mu)$
$[-\infty, 2]$	0	0	0
$[2, \frac{7}{3}]$	0	0	1
$[\frac{7}{3}, 3]$	0	1	1
$[3, \infty]$	1	1	1

$$q(\mu) = \begin{cases} 7\mu, & \mu \in [-\infty, 2] \\ 2\mu + 10, & \mu \in [2, \frac{7}{3}] \\ -\mu + 17, & \mu \in [\frac{7}{3}, 3] \\ -2\mu + 20, & \mu \in [3, \infty] \end{cases}$$

- q concave; non-differentiable at break points $\mu \in \{2, \frac{7}{3}, 3\}$.
- To the left (right) of the optimal solution the derivative of q is non-negative (non-positive). To the left (right) of the optimal solution the subproblem solutions $\boldsymbol{x}(\mu)$ are infeasible (feasible). (Check that the derivative equals the value of the constraint function!)
- The one-variable function q has a "derivative" which is anti-monotone (decreasing); this is a property of every concave function of one variable.
- $\mu^* = \frac{7}{3}$, $q^* = q(\mu^*) = \frac{44}{3} = 14\frac{2}{3}$. Positive duality gap!
- $X(\mu^*) = \{(0,0,1)^{\mathrm{T}}, (0,1,1)^{\mathrm{T}}\} \ni \boldsymbol{x}^*$.

A second example where the duality gap is non-zero

- Example III ($\boldsymbol{x}^* = (2,1)^{\mathrm{T}}$, $f^* = -3$)

$$\begin{aligned} f^* = \text{minimum} \quad & f(\boldsymbol{x}) = -2x_1 + x_2, \\ \text{subject to} \quad & x_1 + x_2 - 3 = 0, \\ & \boldsymbol{x} \in X = \big\{(0,0)^{\mathrm{T}}, (0,4)^{\mathrm{T}}, (4,4)^{\mathrm{T}}, (4,0)^{\mathrm{T}}, \\ & \qquad (1,2)^{\mathrm{T}}, (2,1)^{\mathrm{T}}\big\}. \end{aligned}$$

- $L(\boldsymbol{x}, \mu) = -3\mu + (-2+\mu)x_1 + (1+\mu)x_2$.
- Observe! $\mu \in \mathbb{R}$!

$$X(\mu) = \begin{cases} \{(4,4)^{\mathrm{T}}\}, & \mu < -1 \\ \{(4,4)^{\mathrm{T}}, (4,0)^{\mathrm{T}}\}, & \mu = -1 \\ \{(4,0)^{\mathrm{T}}\}, & \mu \in (-1,2) \\ \{(4,0)^{\mathrm{T}}, (0,0)^{\mathrm{T}}\}, & \mu = 2 \\ \{(0,0)^{\mathrm{T}}\}, & \mu > 2 \end{cases}$$

$$q(\mu) = \begin{cases} -4+5\mu, & \mu \leq -1 \\ -8+\mu, & \mu \in [-1,2] \\ -3\mu, & \mu \geq 2 \end{cases}$$

- $\mu^* = 2$; $q^* = q(\mu^*) = -6$; $q^* < f^*$, $\boldsymbol{x}^* \notin X(\mu^*)$.
- The set $X(\mu^*)$ does not even contain a feasible solution!

Strong duality—repetition

The following three statements are equivalent:

(a) $(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ is a saddle point to L

(b) i. $f(\boldsymbol{x}^*) + (\boldsymbol{\mu}^*)^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}^*) = \text{minimum}_{\boldsymbol{x}\in X}\,\{f(\boldsymbol{x}) + (\boldsymbol{\mu}^*)^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})\}$ $[\Longleftrightarrow \boldsymbol{x}^* \in X(\boldsymbol{\mu}^*)]$

ii. $(\boldsymbol{\mu}^*)^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}^*) = 0$

iii. $\boldsymbol{g}(\boldsymbol{x}^*) \leq \boldsymbol{0}^m$

(c) $f^* = f(\boldsymbol{x}^*) = q(\boldsymbol{\mu}^*) = q^*$.

$\Longrightarrow$ Method for finding an optimal solution:

1) Solve the Lagrangian dual problem $\Longrightarrow \boldsymbol{\mu}^*$;
2) Find a vector $\boldsymbol{x}^* \in X$ which satisfies (b).

- When does this work? What to do if it doesn't?
- Let's study the convex case first.
- Clearly, it only works if the problem has a zero duality gap. Even in the case of a zero duality gap, it is not always trivial to find an optimal primal solution in this way, because the set $X(\boldsymbol{\mu}^*)$ is normally not explicitly given or available—given a value of $\boldsymbol{\mu}$ we normally get one element of the set $X(\boldsymbol{\mu})$.

- A good example was given in Lecture 3—Example II (the 2-variable LP problem). Imagine using the simplex method for solving each LP subproblem. Then, we only get extreme points of X, and $\boldsymbol{x}^*$ was, in this case, an extreme point of $X \cap \{\, \boldsymbol{x} \in \mathbb{R}^2 \mid g(\boldsymbol{x}) \leq 0 \,\}$ (since it is an LP!) but *not* an extreme point of X!
- Several ways out from this *non-coordinability*:
- (1) Remember all the points $\boldsymbol{x}(\boldsymbol{\mu}_k) \in X(\boldsymbol{\mu}_k)$ visited, and at the end solve the LP problem which finds the best point in their convex hull which is also feasible in the original problem. This is the Dantzig–Wolfe (DW) decomposition method.

- (2) Construct a primal sequence as a simple convex combination of the points $\boldsymbol{x}(\boldsymbol{\mu}_k) \in X(\boldsymbol{\mu}_k)$ visited. Compared to DW, we do not solve any extra optimization problems, and virtually no extra memory is needed. On the other hand, DW converges finitely for LP problems, which this technique does not. Read the paper by Larsson, Patriksson, and Strömberg (1999).
- (3) Introduce non-linear price functions for the constraints, instead of the linear one given by Lagrangian relaxation. $\Longrightarrow$ Augmented Lagrangian methods.

Linear integer optimization: The strength of the Lagrangian relaxation

- Comparison with a continuous (LP) relaxation:

$$\begin{array}{rlcrl} v_{LP} = \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} & \leq & v^* = \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} & & \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\ & \boldsymbol{D}\boldsymbol{x} \leq \boldsymbol{d} & & & \boldsymbol{D}\boldsymbol{x} \leq \boldsymbol{d} \\ & \boldsymbol{x} \in \mathbb{R}^n_+ & & & \boldsymbol{x} \in \mathbb{Z}^n_+ \end{array}$$

- Let $X = \{\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^K\}$ be the set of points in $X = \{\, \boldsymbol{x} \in \mathbb{Z}^n_+ \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \,\}$.

$$\begin{aligned} v_L &= \max_{\boldsymbol{\mu} \geq \boldsymbol{0}} \left(\min_{\boldsymbol{x} \in X} \left[\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x} - \boldsymbol{d}) \right] \right) \\ &= \max_{\boldsymbol{\mu} \geq \boldsymbol{0}} \left(\min_{k=1,\ldots,K} \left[\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^k + \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^k - \boldsymbol{d}) \right] \right) \\ &= \max_{\boldsymbol{\mu} \geq \boldsymbol{0}, \theta \in \mathbb{R}} \left\{ \theta \,\middle|\, \theta - (\boldsymbol{D}\boldsymbol{x}^k - \boldsymbol{d})^{\mathrm{T}}\boldsymbol{\mu} \leq \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^k, \quad k = 1, \ldots, K \right\} \end{aligned}$$

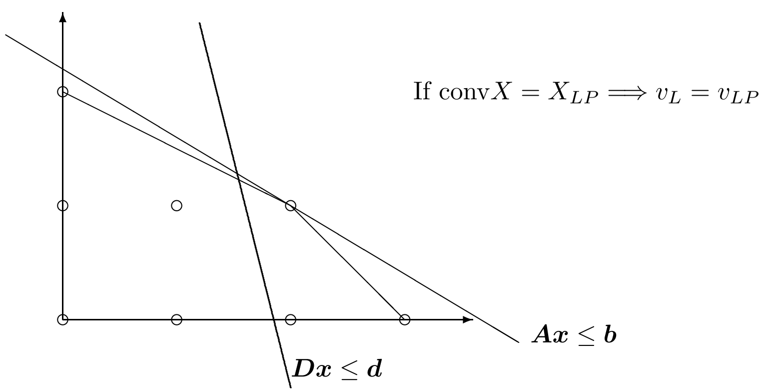
- Introduce dual variables y_k. Continuing,

$$
\begin{aligned}
v_L = \min\ & \sum_{k=1}^{K} (\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^k) y_k \\
\text{s.t.}\ & \sum_{k=1}^{K} y_k = 1 \\
& \sum_{k=1}^{K} (\boldsymbol{D}\boldsymbol{x}^k - \boldsymbol{d}) y_k \le \boldsymbol{0} \iff \boldsymbol{D} \underbrace{\sum_{k=1}^{K} \boldsymbol{x}^k y_k}_{\in \operatorname{conv} X} \le \boldsymbol{d} \underbrace{\sum_{k=1}^{K} y_k}_{=1} \\
& y_k \ge 0, \qquad k = 1, \dots, K \\
= v_C := \min\ & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\
\text{s.t.}\ & \boldsymbol{D}\boldsymbol{x} \le \boldsymbol{d}, \\
& \boldsymbol{x} \in \operatorname{conv} X
\end{aligned}
$$

- Hence, Lagrangian relaxation is a convexification!
- Generating primal solutions through, for example, Dantzig–Wolfe decomposition, or the ergodic sequence method (Larsson, Patriksson, and Strömberg, 1999), yields a solution to a primal LP problem which is the same as the original IP problem where, however, X is replaced its convex hull conv X.
- $v^* \ge v_C = v_L \ge v_{LP}$.

The strength of a Lagrangian dual problem

Since $X \subseteq \operatorname{conv} X \subseteq X_{LP} = \{\, \boldsymbol{x} \in \mathbb{R}^n_+ \mid \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b} \,\}$ we have that $v^* \ge v_L \ge v_{LP}$.



Integrality property

- If $\min_{\boldsymbol{x} \in X_{LP}} \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x} = \min_{\boldsymbol{x} \in \operatorname{conv} X} \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x}$, for all $\boldsymbol{p} \in \mathbb{R}^n$, that is, if the Lagrangian subproblem has the *integrality property*, then $v_L = v_{LP}$.
- Otherwise, v_L is a *better* bound on v^* than is v_{LP} is. $[v_L \ge v_{LP}.]$
- Integrality property $\underbrace{\iff}_{\text{often}}$ easy problem.
- Easy subproblem "$\Longrightarrow$" Bad bounds.
- Difficult subproblem "$\Longrightarrow$" Better bounds.

$\Longrightarrow$ The subproblem should *not* be such that it is too easy to solve!

The strength of the Lagrangian relaxation—An example

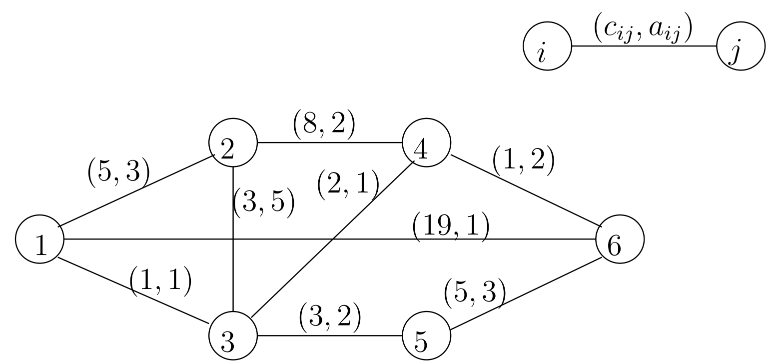
- Consider the *generalized assignment problem* (GAP) to

$$\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \\
\text{subject to} \quad & \sum_{i=1}^{m} x_{ij} = 1, \qquad j = 1,\dots,n, \quad (1) \\
& \sum_{j=1}^{n} a_{ij}x_{ij} \le b_i, \qquad i = 1,\dots,m, \quad (2) \\
& x_{ij} \in \{0,1\}, \quad \forall i,j.
\end{aligned}$$

- (1): Every job j must be performed on exactly one machine.
- (2): The total work done on machine i must not exceed the capacity of the machine.
- Lagrangian relax (1) $\Longrightarrow$ binary knapsack problem! (Difficult) $\Longrightarrow v_L^1$.
- Lagrangian relax (2) $\Longrightarrow$ Semi-assignment problem! (Easy!) $\Longrightarrow v_L^2 \le v_L^1$.
- We prefer the Lagrangian relaxation of (1), because we get much better bounds from the Lagrangian dual problem, and knapsack problems are relatively easy (as far as NP-complete problems go ...)

Questions on the network design problem

1. Formulate the minimum spanning tree problem (MST) as a network flow problem. [*Hint*: consider node 1 as a sink and all other nodes as sources with strength 1.]
2. Consider the graph below.



(a) Provide *all* the spanning trees of this graph explicitly. Calculate the sum of c_{ij} and a_{ij} for each tree. Which ones are feasible with respect to the *budget constraint*

$$\sum_{(i,j)\in\mathcal{T}} a_{ij} \le 10$$

(where $\mathcal{T}$ denotes a collection of links forming a spanning tree)? Which ones are optimal (minimal) with respect to the link costs c_{ij}?

(b) Utilize the solution in (a) to formulate this problem for a general graph.

(c) Formulate the MST problem as a binary, integer programming problem.

(d) Is there a polynomial algorithm for the problem in (b)? [*Hint*: utilize that the binary knapsack problem is

hard.]

3. Provide a polynomial *heuristic* for the problem which gives a feasible solution.
4. Provide a *local search* heuristic which improves a feasible solution.
5. Provide a *Lagrangian relaxation* algorithm.
   (a) Suggest a suitable relaxation.
   (b) How are the subproblems solved?
   (c) Suggest a primal feasibility heuristic.
   (d) Provide a complete Lagrangian relaxation scheme.
6. Suggest a *Branch & Bound* algorithm.

   (a) Suggest a suitable Lagrangian relaxation.
   (b) Suggest a proper branching rule.
   (c) Provide a complete B & B algorithm.
7. Apply some of these algorithms on the above example.