Lecture 5: Lagrangian duality for discrete optimization

Ann-Brith Strömberg

10 September 2008

A reminder of nice properties in the convex case

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

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

- Let $X := \{\mathbf{x} \in \mathbb{R}^2 \mid x_1,\ x_2 \geq 0\} = \mathbb{R}^2_+$.
- $L(\mathbf{x}, \mu) = x_1^2 + x_2^2 + \mu \cdot (-x_1 - x_2 + 4)$

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

- For a fixed value of $\mu \geq 0$, the minimum of $L(\mathbf{x}, \mu)$ over $\mathbf{x} \in X$ is attained at $x_1(\mu) = \frac{\mu}{2}$, $x_2(\mu) = \frac{\mu}{2}$

$\Rightarrow$ $q(\mu) = L(\mathbf{x}(\mu), \mu) = \ldots = 4\mu - \frac{\mu^2}{2}$ for all $\mu \geq 0$.
The dual function q is concave and differentiable.

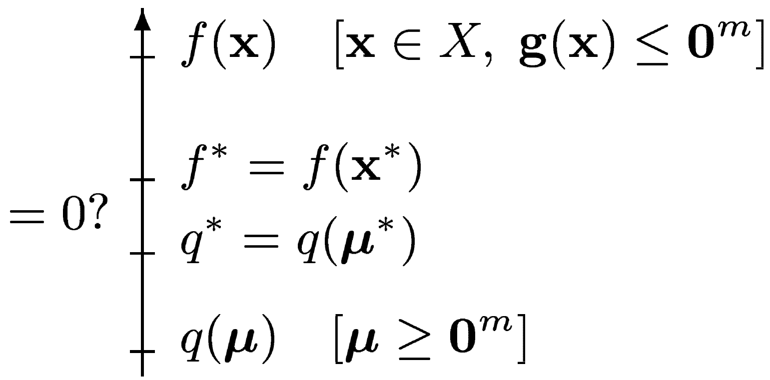
- $f^* = f(\mathbf{x}^*) = 8 = q^*$

- $\mu^* = 4$, $\mathbf{x}(\mu^*) = \mathbf{x}^* = (2, 2)$

Weak duality! Strong duality?

- The primal optimal solution is obtained from the Lagrangian dual optimal solution under convexity and CQ. (For non-strictly convex f we have to deal with the non-coordinability, though)

- What happens otherwise?



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

A first example where the duality gap is non-zero

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

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

- Let $X := \{\, \mathbf{x} \in \mathbb{R}^3 \mid x_j \in \{0, 1\},\ j = 1, 2, 3 \,\} = B^3$
- Let $g(\mathbf{x}) := 7 - x_1 - 3x_2 - 5x_3$
- $L(\mathbf{x}, \mu) = 3x_1 + 7x_2 + 10x_3 + \mu \cdot (7 - x_1 - 3x_2 - 5x_3)$

The dual function is computed according to

$$q(\mu) = 7\mu + \underset{\boldsymbol{x}\in X}{\text{minimum}}\ \{(3-\mu)x_1 + (7-3\mu)x_2 + (10-5\mu)x_3\}$$

$$= 7\mu + \underset{x_1\in\{0,1\}}{\text{minimum}}\ \{(3-\mu)x_1\} + \underset{x_2\in\{0,1\}}{\text{minimum}}\ \{(7-3\mu)x_2\}$$

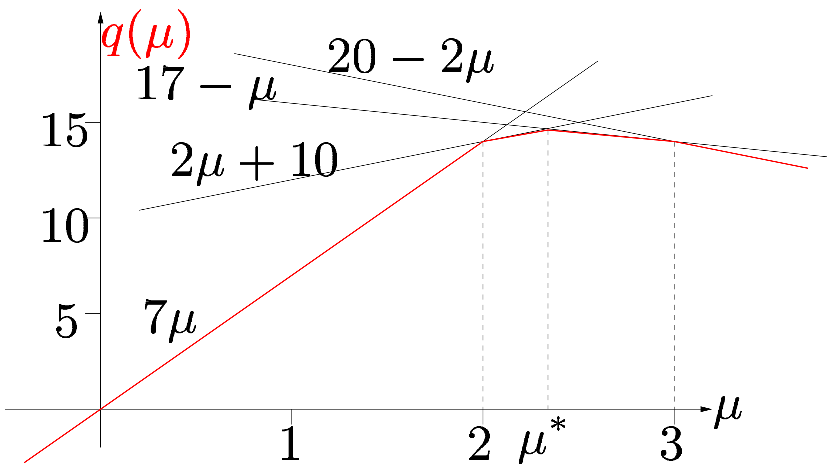
$$+ \underset{x_3\in\{0,1\}}{\text{minimum}}\{(10-5\mu)x_3\}$$

- $X(\mu)$ is obtained by setting

$$x_j(\mu) = \begin{cases} 1 \\ 0 \end{cases} \text{when the objective coefficient is} \begin{cases} \leq 0 \\ \geq 0 \end{cases}$$

Subproblem solutions and the dual function

$\mu \in$	$x_1(\mu)$	$x_2(\mu)$	$x_3(\mu)$	$\mathbf{g}(\mathbf{x}(\mu))$	$q(\mu)$
$[-\infty, 2]$	0	0	0	7	7μ
$[2, \frac{7}{3}]$	0	0	1	2	$2\mu + 10$
$[\frac{7}{3}, 3]$	0	1	1	-1	$-\mu + 17$
$[3, \infty]$	1	1	1	-2	$-2\mu + 20$



- q concave, non-differentiable at break points $\mu \in \{2, \frac{7}{3}, 3\}$.

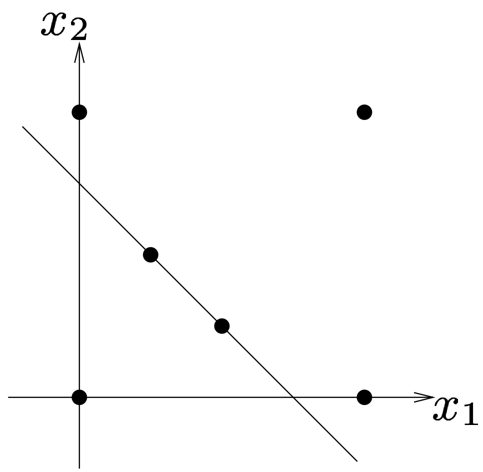
	$\mu < \mu^*$	$\mu > \mu^*$
slope of $q(\mu)$	> 0	< 0
$\mathbf{x}(\mu)$	infeasible	feasible

- Check that the slope equals the value of the constraint function!
- The one-variable function q has a "derivative" which is non-increasing; this is a property of every concave function of one variable.
- $\mu^* = \frac{7}{3}$, $q^* = q(\mu^*) = \frac{44}{3} = 14\frac{2}{3}$.
- Recall: $\mathbf{x}^* = (0, 1, 1)$, $f^* = 17$
- A positive duality gap!
- $X(\mu^*) = \{(0, 0, 1), (0, 1, 1)\} \ni \mathbf{x}^*$

Another example with non-zero duality gap

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

$$
\begin{aligned}
f^* = \min \quad & f(\mathbf{x}) = -2x_1 + x_2 \\
\text{s.t.} \quad & x_1 + x_2 - 3 = 0, \\
& \mathbf{x} \in X = \{(0,0),(0,4),(4,4),(4,0),(1,2),(2,1)\}
\end{aligned}
$$



- $L(\mathbf{x}, \mu) = -3\mu + (-2+\mu)x_1 + (1+\mu)x_2$.

- Observe! $\mu \in \mathbb{R}$ (the relaxed constraint is an equality: $g(\mathbf{x}) = 0$)
- $X(\mu) =$
  $\operatorname{argmin}\{(-2+\mu)x_1 + (1+\mu)x_2 \,|\, \mathbf{x} \in \{(0,0),(0,4),(4,4),(4,0),(1,2),(2,1)\}\}$

$$X(\mu) = \begin{cases} \{(4,4)\}, & \mu < -1 \\ \{(4,4),(4,0)\}, & \mu = -1 \\ \{(4,0)\}, & \mu \in (-1,2) \\ \{(4,0),(0,0)\}, & \mu = 2 \\ \{(0,0)\}, & \mu > 2 \end{cases} \qquad q(\mu) = \begin{cases} -4+5\mu, & \mu \le -1 \\ -8+\mu, & \mu \in [-1,2] \\ -3\mu, & \mu \ge 2 \end{cases}$$

- $\mu^* = 2$; $q^* = q(\mu^*) = -6 < f^* = -3$, $\mathbf{x}^* = (2,1) \notin X(\mu^*)$.
- The set $X(\mu^*)$ does not even contain a feasible solution!

Strong duality—repetition

The following three statements are equivalent:

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

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

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

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

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

$\Longrightarrow$ Method for finding an optimal solution:

1) Solve the Lagrangian dual problem $\Longrightarrow \boldsymbol{\mu}^*$;

2) Find a vector $\mathbf{x}^* \in X$ which satisfies (b).

- When does this work? What if it doesn't?
- First the convex case (with zero duality gap).
- Even with a zero duality gap, it is not always trivial to find an optimal primal solution in this way, since the set $X(\boldsymbol{\mu}^*)$ is normally *not explicitly available*—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 $\mathbf{x}^*$ was, in this case, an extreme point of $X \cap \{ \mathbf{x} \in \mathbb{R}^2 \mid g(\mathbf{x}) \leq 0 \}$ (since it is an LP!) but *not* an extreme point of X!
- Several ways out from this *non-coordinability*:

(1) Remember all the points $\mathbf{x}(\boldsymbol{\mu}^k) \in X(\boldsymbol{\mu}^k)$ visited. At the end, solve an LP to find the best point in their convex hull—also feasible in the original problem. (*The Dantzig–Wolfe (DW) decomposition method.*)

(2) Construct a primal sequence as a *convex combination* of the points $\mathbf{x}(\boldsymbol{\mu}^k) \in X(\boldsymbol{\mu}^k)$ visited. We must not solve any extra optimization problems, and virtually no extra memory is needed (compare to DW).
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 (*Augmented Lagrangian methods*)

Linear integer optimization: The strength of the Lagrangian relaxation

- Compare with a continuous (LP) relaxation:

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

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

- Lagrangian relax the constraints $\mathbf{Dx} \le \mathbf{d}$:

$$v_L = \max_{\boldsymbol{\mu} \ge \mathbf{0}} \left(\min_{\boldsymbol{x} \in X} \left[\mathbf{c}^{\mathrm{T}}\mathbf{x} + \boldsymbol{\mu}^{\mathrm{T}}(\mathbf{Dx} - \mathbf{d}) \right] \right)$$

$$= \max_{\boldsymbol{\mu} \ge \mathbf{0}} \left(\min_{k=1,\ldots,K} \left[\mathbf{c}^{\mathrm{T}}\mathbf{x}^k + \boldsymbol{\mu}^{\mathrm{T}}(\mathbf{Dx}^k - \mathbf{d}) \right] \right)$$

[Picture of a piece-wise linear function of $\boldsymbol{\mu}$!]

$$= \max_{\boldsymbol{\mu} \ge \mathbf{0}, \theta \in \mathbb{R}} \left\{ \theta \,\middle|\, \theta - (\mathbf{Dx}^k - \mathbf{d})^{\mathrm{T}}\boldsymbol{\mu} \le \mathbf{c}^{\mathrm{T}}\mathbf{x}^k, \quad k = 1, \ldots, K \right\}$$

[Picture including θ!]

- Introduce (LP) dual variables y_k. Continuing,

$$v_L = \max_{\boldsymbol{\mu}\geq \mathbf{0},\theta\in\mathbb{R}} \left\{\theta \,\middle|\, \theta - (\mathbf{D}\mathbf{x}^k - \mathbf{d})^{\mathrm{T}}\boldsymbol{\mu} \leq \mathbf{c}^{\mathrm{T}}\mathbf{x}^k, \quad k = 1,\dots,K\right\}$$

$$v_L = \min \sum_{k=1}^{K} (\mathbf{c}^{\mathrm{T}}\mathbf{x}^k) y_k = \mathbf{c}^{\mathrm{T}} \underbrace{\sum_{k=1}^{K} \mathbf{x}^k y_k}_{\in \operatorname{conv} X}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} y_k = 1$$

$$\sum_{k=1}^{K} (\mathbf{D}\mathbf{x}^k - \mathbf{d}) y_k \leq \mathbf{0} \Longleftrightarrow \mathbf{D} \underbrace{\sum_{k=1}^{K} \boldsymbol{x}^k y_k}_{\in \operatorname{conv} X} \leq \mathbf{d} \underbrace{\sum_{k=1}^{K} y_k}_{=1}$$

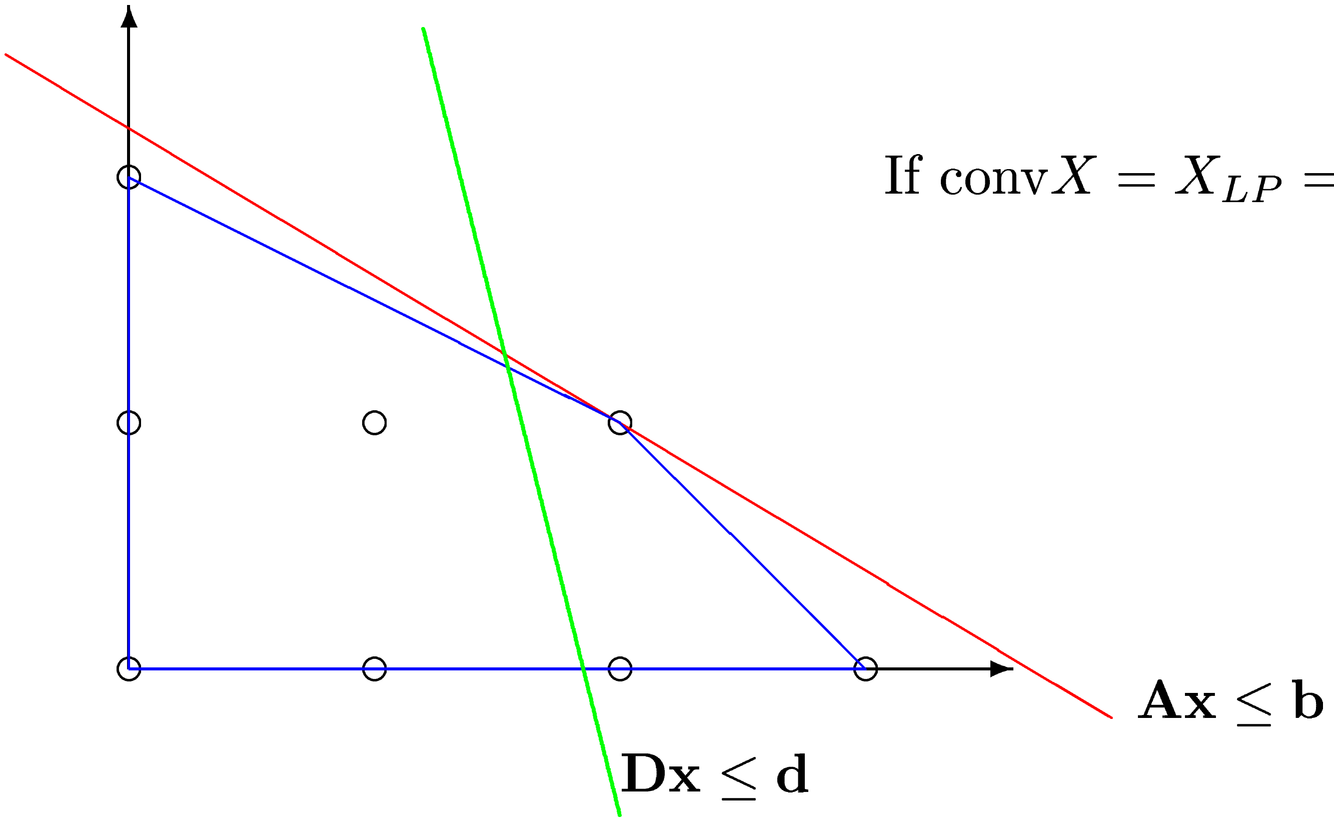
$$y_k \geq 0, \qquad k = 1,\dots,K$$

$$= v_C := \min \quad \left\{\mathbf{c}^{\mathrm{T}}\mathbf{x} \,\middle|\, \mathbf{D}\mathbf{x} \leq \mathbf{d}, \mathbf{x} \in \operatorname{conv} X\right\}$$

- 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 equivalent to the original IP problem where, however, X is replaced by $\operatorname{conv} X$.
- We conclude: $v^* \geq v_C = v_L \geq v_{LP}$.
  C: convexification of X
  L: Lagrangian dual
  LP: Linear programming relaxation

The strength of a Lagrangian dual problem

Since $X \subseteq \text{conv}X \subseteq X_{LP} = \{\, \mathbf{x} \in \mathbb{R}^n_+ \mid \mathbf{Ax} \leq \mathbf{b} \,\}$ we have that $v^* \geq v_L \geq v_{LP}$.



If $\text{conv}X = X_{LP} \Longrightarrow v_L = v_{LP}$

Integrality property

- If $\min_{\boldsymbol{x} \in X_{LP}} \mathbf{p}^{\mathrm{T}}\mathbf{x} = \min_{\boldsymbol{x} \in \mathrm{conv} X} \mathbf{p}^{\mathrm{T}}\mathbf{x}$, for all $\mathbf{p} \in \mathbb{R}^n$, 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} $[v_L \geq v_{LP}]$

- Integrality property $\underbrace{\Longleftrightarrow}_{\text{often}}$ easy problem

- Easy subproblem "$\Longrightarrow$" Bad bounds

- Difficult subproblem "$\Longrightarrow$" Better bounds

$\Longrightarrow$ The subproblem should *not* be *too easy* to solve!

The strength of the Lagrangian relaxation
An example

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

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

[Draw a bipartite graph!]

(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!) $\Longrightarrow v_L^2 \leq v_L^1$
- We prefer the Lagrangian relaxation of (1), since this gives much better bounds from the Lagrangian dual problem, and knapsack problems are relatively easy (as far as NP-complete problems go . . .)