Lecture 10: Linear programming duality and sensitivity

Michael Patriksson

17 February 2005

The canonical primal–dual pair

$\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, and $\boldsymbol{c} \in \mathbb{R}^n$

$$\begin{aligned} \text{maximize} \quad & z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}, \\ & \boldsymbol{x} \ge \boldsymbol{0}^n \end{aligned} \tag{1}$$

and

$$\begin{aligned} \text{minimize} \quad & w = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} \\ \text{subject to} \quad & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \ge \boldsymbol{c}, \\ & \boldsymbol{y} \ge \boldsymbol{0}^m \end{aligned} \tag{2}$$

The dual of the LP in standard form

$$
\begin{array}{llr}
\text{minimize} & z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} & \text{(P)} \\
\text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, & \\
 & \boldsymbol{x} \geq \boldsymbol{0}^n &
\end{array}
$$

and

$$
\begin{array}{llr}
\text{maximize} & w = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} & \text{(D)} \\
\text{subject to} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \leq \boldsymbol{c}, & \\
 & \boldsymbol{y} \ \text{free} &
\end{array}
$$

Rules for formulating dual LPs

- We say that an inequality is *canonical* if it is of $\leq$ [respectively, $\geq$] form in a maximization [respectively, minimization] problem
- We say that a variable is *canonical* if it is ≥ 0
- The rule is that the dual variable [constraint] for a primal constraint [variable] is canonical if the other one is canonical. If the direction of a primal constraint [sign of a primal variable] is the opposite from the canonical, then the dual variable [dual constraint] is also opposite from the canonical

- Further, the dual variable [constraint] for a primal equality constraint [free variable] is free [an equality constraint]
- Summary:

primal/dual constraint		**dual/primal variable**
canonical inequality	$\Longleftrightarrow$	≥ 0
non-canonical inequality	$\Longleftrightarrow$	≤ 0
equality	$\Longleftrightarrow$	unrestricted (free)

Weak Duality Theorem

- *If $\boldsymbol{x}$ is a feasible solution to* (P) *and $\boldsymbol{y}$ a feasible solution to* (D), *then $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \geq \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}$.*
- Similar relation for the primal–dual pair (2)–(1): the max problem never has a higher objective value.
- *Proof.*
- Corollary: If $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}$ for a feasible primal–dual pair $(\boldsymbol{x}, \boldsymbol{y})$ then they must be optimal.

Strong Duality Theorem

- In the book, strong duality is established for the pair (P) and (D). Here, we establish the result for the pair (1), (2).
- *If one of the problems* (1) *and* (2) *has a finite optimal solution, then so does its dual, and their optimal objective values are equal.*
- *Proof idea.* Suppose w.l.o.g. that it is the primal max problem that has a finite optimal solution. State an optimal BFS representing an optimal extreme point $\boldsymbol{x}$, and thereafter establish (a) that from the optimality conditions that the reduced costs are non-positive, we

can construct a dual vector of the form $\boldsymbol{y}^{\mathrm{T}} = \boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{B}^{-1}$ which is feasible in (2). (b) We show that the objective values $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$ and $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}$ are equal. Hence, the dual vector must be optimal in its problem by the Corollary.

Complementary Slackness Theorem

- *Let $\boldsymbol{x}$ be a feasible solution to* (1) *and $\boldsymbol{y}$ a feasible solution to* (2). *Then $\boldsymbol{x}$ is optimal to* (1) *and $\boldsymbol{y}$ optimal to* (2) *if and only if*

$$x_j(c_j - \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}_{\cdot j}) = 0, \qquad j = 1, \ldots, n, \tag{3a}$$

$$y_i(\boldsymbol{A}_{i\cdot}\boldsymbol{x} - b_i) = 0, \qquad i = 1 \ldots, m, \tag{3b}$$

  where $\boldsymbol{A}_{\cdot j}$ is the j^{th} column of $\boldsymbol{A}$, $\boldsymbol{A}_{i\cdot}$ the i^{th} row of $\boldsymbol{A}$.

- *Proof.*

Necessary and sufficient optimality conditions: Strong duality, the global optimality conditions, and the KKT conditions are equivalent for LP

- We have seen above that the following statement characterizes the optimality of a primal–dual pair $(\boldsymbol{x}, \boldsymbol{y})$:
- $\boldsymbol{x}$ is feasible in (1), $\boldsymbol{y}$ is feasible in (2), and complementarity holds.
- In other words, we have the following result (think of the KKT conditions!):

- *Take a vector* $\boldsymbol{x} \in \mathbb{R}^n$*. For* $\boldsymbol{x}$ *to be an optimal solution to the linear program* (1), *it is both necessary and sufficient that*

  (a) $\boldsymbol{x}$ *is a feasible solution to* (1)*;*

  (b) *corresponding to* $\boldsymbol{x}$ *there is a dual feasible solution* $\boldsymbol{y} \in \mathbb{R}^m$ *to* (2)*; and*

  (c) $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *together satisfy complementarity* (3)*.*

- This is precisely the same as the KKT conditions!
- Those who wishes to establish this—note that there are no multipliers for the "$\boldsymbol{x} \geq \mathbf{0}^n$" constraints, and in the KKT conditions there are. Introduce such a multiplier vector and see that it can later be eliminated.

- Further: suppose that $\boldsymbol{x}$ and $\boldsymbol{y}$ are feasible in (1) and (2). Then, the following are equivalent:

  (a) $\boldsymbol{x}$ and $\boldsymbol{y}$ have the same objective value;

  (b) $\boldsymbol{x}$ and $\boldsymbol{y}$ solve (1) and (2);

  (c) $\boldsymbol{x}$ and $\boldsymbol{y}$ satisfy complementarity.

The Simplex method and the global optimality conditions

- The Simplex method is remarkable in that it satisfies two of the three conditions at every BFS, and the remaining one is satisfied at optimality:
- $\boldsymbol{x}$ is feasible after Phase-I has been completed.
- $\boldsymbol{x}$ and $\boldsymbol{y}$ always satisfy complementarity. Why? If x_j is in the basis, then it has a zero reduced cost, implying that the dual constraint j has no slack. If the reduced cost of x_j is non-zero (slack in dual constraint j), then its value is zero.

- The feasibility of $\boldsymbol{y}^{\mathrm{T}} = \boldsymbol{c}_{\boldsymbol{B}}^{\mathrm{T}} \boldsymbol{B}^{-1}$ is not fulfilled until we reach an optimal BFS. How is the incoming criterion related to this? We introduce as an incoming variable that variable which has the best reduced cost. Since the reduced cost measures the dual feasibility of $\boldsymbol{y}$, this means that we select the most violated dual constraint; at the new BFS, that constraint is then satisfied (since the reduced cost then is zero). The Simplex method hence works to try to satisfy dual feasibility!

Farkas' Lemma revisited

- *Let* $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ *and* $\boldsymbol{b} \in \mathbb{R}^m$*. Then, exactly one of the systems*

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} &= \boldsymbol{b}, \\ \boldsymbol{x} &\geq \boldsymbol{0}^n, \end{aligned} \tag{I}$$

and

$$\begin{aligned} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} &\leq \boldsymbol{0}^n, \\ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} &> 0, \end{aligned} \tag{II}$$

has a feasible solution, and the other system is inconsistent.

- *Proof.*

Sensitivity analysis, I: Shadow prices are derivatives of a convex function!

- Suppose an optimal BFS is non-degenerate. Then, $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^* = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}^* = \boldsymbol{c}_B^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{b}$ varies linearly as a function of $\boldsymbol{b}$ around its given value.
- Non-degeneracy also implies that $\boldsymbol{y}^*$ is unique. Why?
- Perturbation function $\boldsymbol{b} \mapsto v(\boldsymbol{b})$ given by

$$\begin{aligned} v(\boldsymbol{b}) = \min\, & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} & = \max\, & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} & = \max_{k\in K} \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}_k. \\ \text{s.t. } & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, & \text{s.t. } & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \leq \boldsymbol{c} & \\ & \boldsymbol{x} \geq \boldsymbol{0}^n & & & \end{aligned}$$

K: set of DFS. v a piece-wise linear, convex function.

- Fact: v convex (and finite in a neighbourhood of $\boldsymbol{b}$) implies that v differentiable at $\boldsymbol{b}$ iff it has a unique subgradient there.
- Here: derivative w.r.t. $\boldsymbol{b}$ is $\boldsymbol{y}^*$, that is, the change in the optimal value from a change in the right-hand side $\boldsymbol{b}$ equals the dual optimal solution.

Sensitivity analysis, II: Perturbations in data

- How to find a new optimum through re-optimization when data has changed
- If an element of $\boldsymbol{c}$ changes, then the old BFS is feasible but may not be optimal. Check the new value of the reduced cost vector $\tilde{\boldsymbol{c}}$ and change the basis if some sign has changed.

- If an element of $\boldsymbol{b}$ changes, then the old BFS is optimal but may not be feasible. Check the new value of the vector $\boldsymbol{B}^{-1}\boldsymbol{b}$ and change the basis if some sign has changed. Since the BFS is infeasible but optimal, we use a dual version of the Simplex method: the *Dual Simplex method.*
- Find a negative basic variable $x_j \to$ outgoing basic variable x_s
- Choose among the non-basic variables for which the element $\boldsymbol{B}^{-1}\boldsymbol{N}_{sj} < 0$; means that the new basic variable will become positive.
- Choose the incoming variable so that $\tilde{\boldsymbol{c}}$ keeps its sign.

Decentralized planning

- Consider the following profit maximization problem:

$$\text{maximize } z = \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x} = \sum_{i=1}^{m} \boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{x}_i,$$

$$\text{s.t.} \quad \begin{pmatrix} \boxed{\boldsymbol{B}_1} & & & \\ & \boxed{\boldsymbol{B}_2} & & \\ & & \ddots & \\ & & & \boxed{\boldsymbol{B}_m} \\ & & \boxed{\quad\boldsymbol{C}\quad} & \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_m \end{pmatrix} \leq \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_m \\ \boldsymbol{c} \end{pmatrix},$$

$$\boldsymbol{x}_i \geq \boldsymbol{0}^{n_i}, \quad i = 1, \ldots, m,$$

for which we have the following interpretation:

- We have m independent subunits, responsible for finding their optimal production plan.
- While they are governed by their own objectives, we (the Managers) want to solve the overall problem of maximizing the company's profit.
- The constraints $\boldsymbol{B}_i\boldsymbol{x}_i \leq \boldsymbol{b}_i$, $\boldsymbol{x}_i \geq \mathbf{0}^{n_i}$ describe unit i's own production limits, when using their own resources.

- The units also use limited resources that are the same.
- The resource constraint is difficult as well as unwanted to enforce directly, because it would make it a *centralized planning* process.
- We want the units to maximize their own profits individually.
- But we must also make sure that they do not violate the resource constraints $\boldsymbol{C}\boldsymbol{x} \leq \boldsymbol{c}$.
- (This constraint is typically of the form $\sum_{i=1}^{m} \boldsymbol{C}_i\boldsymbol{x}_i \leq \boldsymbol{c}$.)
- How?

- ANSWER: Solve the LP dual problem!
- Generate from the dual solution the dual vector $\boldsymbol{y}$ for the joint resource constraint.
- Introduce an *internal price* for the use of this resource, equal to this dual vector.
- Let each unit optimize their own production plan, with an additional cost term.
- This will then be a *decentralized planning* process.

- Each unit i will then solve their own LP problem to

$$\begin{aligned} \underset{\boldsymbol{x}_i}{\text{maximize}}\ & [\boldsymbol{p}_i - \boldsymbol{C}_i^{\mathrm{T}}\boldsymbol{y}]^{\mathrm{T}}\boldsymbol{x}_i, \\ \text{subject to}\ & \boldsymbol{B}_i\boldsymbol{x}_i \leq \boldsymbol{b}_i, \\ & \boldsymbol{x} \geq \mathbf{0}^{n_i}, \end{aligned}$$

  resulting in an optimal production plan!
- *Decentralized planning*, is related to *Dantzig–Wolfe* decomposition, which is a general technique for solving large-scale LP by solving a sequence of smaller LP:s.
- More on such techniques in the Project course.