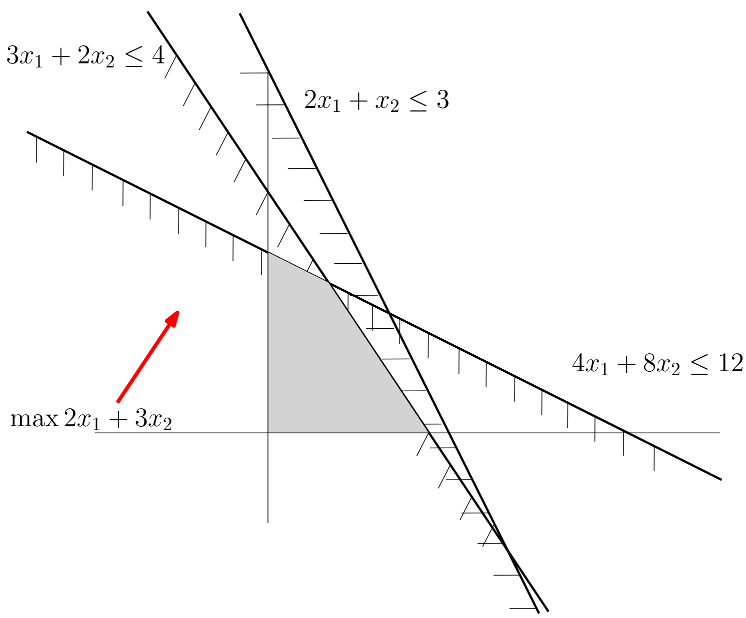
In this lecture we discuss the general notion of Linear Programming Duality, a powerful tool that you should definitely master.

1 Linear Programming Duality

Consider the following LP

$$\begin{aligned} P = \max(&2x_1 + 3x_2) \\ \text{s.t.} \quad &4x_1 + 8x_2 \le 12 \\ &2x_1 + x_2 \le 3 \\ &3x_1 + 2x_2 \le 4 \\ &x_1, x_2 \ge 0 \end{aligned} \tag{1}$$



In an attempt to solve P we can produce upper bounds on its optimal value.

- Since $2x_1 + 3x_2 \le 4x_1 + 8x_2 \le 12$, we know $\text{OPT}(P) \le 12$. (The first inequality uses that $2x_1 \le 4x_1$ because $x_1 \ge 0$, and similarly $3x_2 \le 8x_2$ because $x_2 \ge 0$.)
- Since $2x_1 + 3x_2 \le \frac{1}{2}(4x_1 + 8x_2) \le 6$, we know $\text{OPT}(P) \le 6$.
- Since $2x_1 + 3x_2 \le \frac{1}{3}((4x_1 + 8x_2) + (2x_1 + x_2)) \le 5$, we know $\text{OPT}(P) \le 5$.

In each of these cases we take a positive[1] linear combination of the constraints, looking for better and better bounds on the maximum possible value of $2x_1 + 3x_2$.

[1]Why positive? If you multiply by a negative value, the sign of the inequality changes.

How do we find the tighest upper bound that can be achieved as a linear combination of the constraints? This is just another algorithmic problem, and we can systematically solve it, by letting y_1, y_2, y_3 be the (unknown) coefficients of our linear combination. Then we must have

$$\begin{aligned} 4y_1 + 2y_2 + 3y_3 &\geq 2 \\ 8y_1 + y_2 + 2y_3 &\geq 3 \\ y_1, y_2, y_3 &\geq 0 \\ \text{and we seek } \min(12y_1 + 3y_2 + 4y_3) & \end{aligned} \tag{2}$$

This too is an LP! We refer to this LP (2) as the "dual" and the original LP 1 as the "primal". We designed the dual to serve as a method of constructing an upper bound on the optimal value of the primal, so if y is a feasible solution for the dual and x is a feasible solution for the primal, then $2x_1 + 3x_2 \leq 12y_1 + 3y_2 + 4y_3$. If we can find two feasible solutions that make these equal, then we know we have found the optimal values of these LP.

In this case the feasible solutions $x_1 = \frac{1}{2}, x_2 = \frac{5}{4}$ and $y_1 = \frac{5}{16}, y_2 = 0, y_3 = \frac{1}{4}$ give the same value 4.75, which therefore must be the optimal value.

Exercise 1: The dual LP is a *minimization* LP, where the constraints are of the form $lhs_i \geq rhs_i$. You can try to give *lower* bounds on the optimal value of this LP by taking positive linear combinations of these constraints. E.g., argue that

$$12y_1 + 3y_2 + 4y_3 \geq 4y_1 + 2y_2 + 3y_3 \geq 2$$

(since $y_i \geq 0$ for all i) and

$$12y_1 + 3y_2 + 4y_3 \geq 8y_1 + y_2 + 2y_3 \geq 3$$

and

$$12y_1 + 3y_2 + 4y_3 \geq \frac{2}{3}(4y_1 + 2y_2 + 3y_3) + (8y_1 + y_2 + 2y_3) \geq \frac{4}{3} + 3 = 4\frac{1}{3}.$$

Formulate the problem of finding the best lower bound obtained by linear combinations of the given inequalities as an LP. Show that the resulting LP is the same as the primal LP 1.

Exercise 2: Consider the "primal" LP below on the left:

$$\begin{aligned} P = \max(7x_1 - x_2 + 5x_3) & \\ \text{s.t.} \quad x_1 + x_2 + 4x_3 &\leq 8 \\ 3x_1 - x_2 + 2x_3 &\leq 3 \\ 2x_1 + 5x_2 - x_3 &\leq -7 \\ x_1, x_2, x_3 &\geq 0 \end{aligned} \qquad \begin{aligned} D = \min(8y_1 + 3y_2 - 7y_3) & \\ \text{s.t.} \quad y_1 + 3y_2 + 2y_3 &\geq 7 \\ y_1 - y_2 + 5y_3 &\geq -1 \\ 4y_1 + 2y_2 - y_3 &\geq 5 \\ y_1, y_2, y_3 &\geq 0 \end{aligned}$$

Show that the problem of finding the best upper bound obtained using linear combinations of the constraints can be written as the LP above on the right (the "dual" LP). Also, now formulate the problem of finding a lower bound for the dual LP. Show this lower-bounding LP is just the primal (P).

Exercise 3: In the examples above, the maximization LPs had constraints of the form $lhs_i \leq rhs_i$, and the rhs were all scalars, so taking positive linear combinations gave us $blah \leq number$, i.e., an *upper* bound as we wanted. However, suppose the primal LP has some "nice" constraints $lhs_i \leq rhs_i$ and others are "not nice" $lhs_i \geq rhs_i$, e.g., like the left one below. Show that the dual has non-positive variables for the non-nice constraints. For example,

$$\begin{aligned} P = \max(7x_1 - x_2 + 5x_3) & \\ \text{s.t.} \quad x_1 + x_2 + 4x_3 &\leq 8 \\ 3x_1 - x_2 + 2x_3 &\geq 3 \\ x_1, x_2, x_3 &\geq 0 \end{aligned} \qquad \begin{aligned} D = \min(8y_1 + 3y_2) & \\ \text{s.t.} \quad y_1 + 3y_2 &\geq 7 \\ y_1 - y_2 &\geq -1 \\ 4y_1 + 2y_2 &\geq 5 \\ y_1 \geq 0, \; y_2 &\leq 0 \end{aligned}$$

Another way is to replace $lhs_i \geq rhs_i$ in P by the equivalent constraint $(-lhs_i) \leq (-rhs_i)$ and get to an LP P' with only nice constraints. Show that the dual D' for P' is equivalent to the dual D for P.

1.1 The Method

Consider the examples/exercises above. In all of them, we started off with a "primal" maximization LP:

$$\begin{aligned} \text{maximize } & \mathbf{c}^T\mathbf{x} \\ \text{subject to } & A\mathbf{x} \le \mathbf{b} \\ & \mathbf{x} \ge \mathbf{0}, \end{aligned} \tag{3}$$

The constraint $\mathbf{x} \ge \mathbf{0}$ is just short-hand for saying that the $\mathbf{x}$ variables are constrained to be non-negative.[2] And to get the best upper bound we generated a "dual" minimization LP:

$$\begin{aligned} \text{minimize } & \mathbf{r}^T\mathbf{y} \\ \text{subject to } & P\mathbf{y} \ge \mathbf{q} \\ & \mathbf{y} \ge \mathbf{0}, \end{aligned}$$

The important thing is: this matrix P, and vectors $\mathbf{q}, \mathbf{r}$ are not just any vectors. Look carefully: $P = A^T$. $\mathbf{q} = \mathbf{c}$ and $\mathbf{r} = \mathbf{b}$. The dual is in fact:

$$\begin{aligned} \text{minimize } & \mathbf{y}^T\mathbf{b} \\ \text{subject to } & \mathbf{y}^T A \ge \mathbf{c}^T \\ & \mathbf{y} \ge \mathbf{0}, \end{aligned} \tag{4}$$

And if you take the dual of (4) to try to get the best lower bound on this LP, you'll get (3). *The dual of the dual is the primal.* The dual and the primal are best upper/lower bounds you can obtain as linear combinations of the inputs.

The natural question is: maybe we can obtain better bounds if we combine the inequalities in more complicated ways, not just using linear combinations. Or do we obtain optimal bounds using just linear combinations? In fact, we get optimal bounds using just linear combinations, as the next theorems show.

1.2 The Theorems

It is easy to show that the dual (4) provides an upper bound on the value of the primal (3):

Theorem 1 (Weak Duality) *If $\mathbf{x}$ is a feasible solution to the primal LP (3) and $\mathbf{y}$ is a feasible solution to the dual LP (4) then*

$$\mathbf{c}^T\mathbf{x} \le \mathbf{y}^T\mathbf{b}.$$

Proof: This is just a sequence of trivial inequalities that follow from the LPs above:

$$\mathbf{c}^T\mathbf{x} \le (y^T A)\mathbf{x} = y^T(A\mathbf{x}) \le y^T b.$$

■

The amazing (and deep) result here is to show that the dual actually gives a perfect upper bound on the primal (assuming some mild conditions).

[2] We use the convention that vectors like $\mathbf{c}$ and $\mathbf{x}$ are column vectors. So $\mathbf{c}^T$ is a row vector, and thus $\mathbf{c}^T\mathbf{x}$ is the same as the inner product $\mathbf{c} \cdot \mathbf{x} = \sum_i c_i x_i$. We often use $\mathbf{c}^T\mathbf{x}$ and $\mathbf{c} \cdot \mathbf{x}$ interchangeably. Also, $\mathbf{a} \le \mathbf{b}$ means component-wise inequality, i.e., $a_i \le b_i$ for all i.

Theorem 2 (Strong Duality Theorem) *Suppose the primal LP (3) is feasible (i.e., it has at least one solution) and bounded (i.e., the optimal value is not ∞). Then the dual LP (4) is also feasible and bounded. Moreover, if $\mathbf{x}^*$ is the optimal primal solution, and $\mathbf{y}^*$ is the optimal dual solution, then*

$$\mathbf{c}^T\mathbf{x}^* = (\mathbf{y}^*)^T\mathbf{b}.$$

In other words, the maximum of the primal equals the minimum of the dual.

Why is this useful? If I wanted to prove to you that $\mathbf{x}^*$ was an optimal solution to the primal, I could give you the solution $\mathbf{y}^*$, and you could check that $\mathbf{x}^*$ was feasible for the primal, $\mathbf{y}^*$ feasible for the dual, and they have equal objective function values.

This min-max relationship is like in the case of s-t flows: the maximum of the flow equals the minimum of the cut. Or like in the case of zero-sum games: the payoff for the maxmin-optimum strategy of the row player equals the (negative) of the payoff of the maxmin-optimal strategy of the column player. Indeed, both these things are just special cases of strong duality!

We will not prove Theorem 2 in this course, though the proof is not difficult. But let's give a geometric intuition of why this is true in the next section.

1.3 The Geometric Intuition for Strong Duality

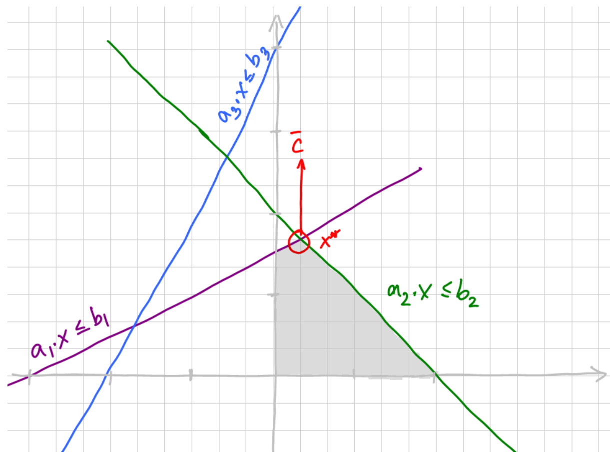
To give a geometric view of the strong duality theorem, consider an LP of the following form:

$$\begin{aligned} \text{maximize } & \mathbf{c}^T\mathbf{x} \\ \text{subject to } \quad & A\mathbf{x} \le \mathbf{b} \\ & \mathbf{x} \ge 0 \end{aligned} \tag{5}$$

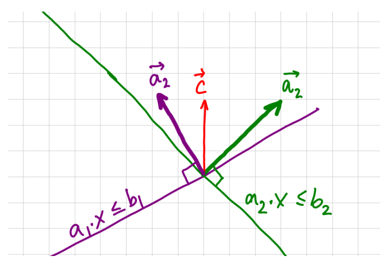
For concreteness, let's take the following 2-dimensional LP:

$$\begin{aligned} \text{maximize } & x_2 \\ \text{subject to } & -x_1 + 2x_2 \le 3 \\ & x_1 + x_2 \le 2 \\ & -2x_1 + x_2 \le 4 \\ & x_1, x_2 \ge 0 \end{aligned}$$

If $\mathbf{c} := (0, 1)$, then the objective function wants to maximize $\mathbf{c} \cdot \mathbf{x}$, i.e., to go as far up in the vertical direction as possible. As we have already argued before, the optimal point $\mathbf{x}^*$ must be obtained at the intersection of two constraints for this 2-dimensional problem (n tight constraints for n dimensions). In this case, these happen to be the first two constraints.



If $\mathbf{a}_1 = (-1, 2), b_1 = 3$ and $\mathbf{a}_2 = (1, 1), b_2 = 2$, then $\mathbf{x}^*$ is the (unique) point $\mathbf{x}$ satisfying both $\mathbf{a}_1 \cdot \mathbf{x} = b_1$ and $\mathbf{a}_2 \cdot \mathbf{x} = b_2$. Indeed, we're being held down by these two constraints. Geometrically, this means that $\mathbf{c} = (0, 1)$ lies "between" these the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ that are normal (perpendicular) to these constraints.



Consequently, $\mathbf{c}$ can be written as a positive linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$. (It "lies in the cone formed by $\mathbf{a}_1$ and $\mathbf{a}_2$.") I.e., for some positive values y_1 and y_2,

$$\mathbf{c} = y_1 \mathbf{a}_1 + y_2 \mathbf{a}_2.$$

Great. Now, take dot products on both sides with $\mathbf{x}^*$. We get

$$\begin{aligned} \mathbf{c} \cdot \mathbf{x}^* &= (y_1\, \mathbf{a}_1 + y_2\, \mathbf{a}_2) \cdot \mathbf{x}^* \\ &= y_1(\mathbf{a}_1 \cdot \mathbf{x}^*) + y_2(\mathbf{a}_2 \cdot \mathbf{x}^*) \\ &= y_1 b_1 + y_2 b_2 \end{aligned}$$

Defining $\mathbf{y} = (y_1, y_2, 0, \ldots, 0)$, we get

$$\text{optimal value of primal} = \mathbf{c} \cdot \mathbf{x}^* = \mathbf{b} \cdot \mathbf{y} \geq \text{value of dual solution } \mathbf{y}.$$

The last inequality follows because

- the $\mathbf{y}$ we found satisfies $\mathbf{c} = y_1\mathbf{a}_1 + y_2\mathbf{a}_2 = \sum_i y_i \mathbf{a}_i = A^T\mathbf{y}$, and hence $\mathbf{y}$ satisfies the dual constraints $\mathbf{y}^T A \geq \mathbf{c}^T$ by construction.

In other words, $\mathbf{y}$ is a feasible solution to the dual, has value $\mathbf{b} \cdot \mathbf{y} \leq \mathbf{c} \cdot \mathbf{x}^*$. So the *optimal* dual value cannot be less. Combined with weak duality (which says that $\mathbf{c} \cdot \mathbf{x}^* \leq \mathbf{b} \cdot \mathbf{y}$), we get strong duality

$$\mathbf{c} \cdot \mathbf{x}^* = \mathbf{b} \cdot \mathbf{y}.$$

Above, we used that the optimal point was constrained by two of the inequalities (and that these were not the non-negativity constraints). The general proof is similar: for n dimensions, we just use that the optimal point is constrained by n tight inequalities, and hence $\mathbf{c}$ can be written as a positive combination of n of the constraints (possibly some of the non-negativity constraints too).

2 Example #1: Zero-Sum Games

Consider a 2-player zero-sum game defined by an n-by-m payoff matrix R for the row player. That is, if the row player plays row i and the column player plays column j then the row player gets payoff R_{ij} and the column player gets $-R_{ij}$. To make this easier on ourselves (it will allow us to simplify things a bit), let's assume that all entries in R are positive (this is really without loss of generality since as pre-processing one can always translate values by a constant and this will just change the game's value to the row player by that constant). We saw we could write this as an LP:

- Variables: $v, p_1, p_2, \ldots, p_n$.
- Maximize v,
- Subject to:

  $p_i \geq 0$ for all rows i,

  $\sum_i p_i = 1$,

  $\sum_i p_i R_{ij} \geq v$, for all columns j.

To put this into the form of (3), we can replace $\sum_i p_i = 1$ with $\sum_i p_i \leq 1$ since we said that all entries in R are positive, so the maximum will occur with $\sum_i p_i = 1$, and we can also safely add in the constraint $v \geq 0$. We can also rewrite the third set of constraints as $v - \sum_i p_i R_{ij} \leq 0$. This then gives us an LP in the form of (3) with

$$\mathbf{x} = \begin{array}{|c|}\hline v \\ p_1 \\ p_2 \\ \ldots \\ p_n \\ \hline\end{array}, \mathbf{c} = \begin{array}{|c|}\hline 1 \\ 0 \\ 0 \\ \ldots \\ 0 \\ \hline\end{array}, \mathbf{b} = \begin{array}{|c|}\hline 0 \\ 0 \\ \ldots \\ 0 \\ 1 \\ \hline\end{array}, \text{and } A = \begin{array}{|c|ccc|}\hline 1 & & & \\ 1 & & -R^T & \\ \ldots & & & \\ 1 & & & \\ \hline 0 & 1 & \ldots & 1 \\ \hline\end{array}.$$

I.e., maximizing $\mathbf{c}^T\mathbf{x}$ subject to $A\mathbf{x} \leq \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$.

We can now write the dual, following (4). Let $\mathbf{y}^T = (y_1, y_2, \ldots, y_{m+1})$. We now are asking to minimize $\mathbf{y}^T\mathbf{b}$ subject to $\mathbf{y}^T A \geq c^T$ and $\mathbf{y} \geq \mathbf{0}$. In other words, we want to:

- Minimize y_{m+1},
- Subject to:

  $y_1 + \ldots + y_m \geq 1$,

  $-y_1 R_{i1} - y_2 R_{i2} - \ldots - y_m R_{im} + y_{m+1} \geq 0$ for all rows i,

  or equivalently,

  $y_1 R_{i1} + y_2 R_{i2} + \ldots + y_m R_{im} \leq y_{m+1}$ for all rows i.

So, we can interpret y_{m+1} as the value to the row player, and $y_1, \ldots, y_m$ as the randomized strategy of the column player, and we want to find a randomized strategy for the column player that minimizes y_{m+1} subject to the constraint that the row player gets *at most* y_{m+1} no matter what row he plays. Now notice that we've only required $y_1 + \ldots + y_m \geq 1$, but since we're minimizing and the R_{ij}'s are positive, the minimum will happen at equality.

Notice that the fact that the maximum value of v in the primal is equal to the minimum value of y_{m+1} in the dual follows from strong duality. Therefore, the minimax theorem is a corollary to the strong duality theorem.

3 Example #2: Shortest Paths

Duality allows us to write problems in multiple ways, which often gives us power and flexibility. For instance, let us see two ways of writing the shortest s-t path problem, and why they are equal.

Here is an LP for computing an s-t shortest path with respect to the edge lengths $\ell(u, v) \geq 0$:

$$\begin{aligned} \max \quad & d_t \\ \text{subject to} \quad & d_s = 0 \\ & d_v - d_u \leq \ell(u, v) \qquad \forall (u, v) \in E \end{aligned} \tag{6}$$

The constaints are the natural ones: the shortest distance from s to s is zero. And if the s-u distance is d_u, the s-v distance is at most $d_u + \ell(u, v)$ — i.e., $d_v \leq d_u + \ell(u, v)$. It's like putting strings of length $\ell(u, v)$ between u, v and then trying to send t as far from s as possible—the farthest you can send t from s is when the shortest s-t path becomes tight.

Here is another LP that also computes the s-t shortest path:

$$\begin{aligned} & \min \textstyle\sum_e \ell(e)\, y_e \\ \text{subject to} \quad & \textstyle\sum_{w:(s,w) \in E} y_{sw} = 1 \\ & \textstyle\sum_{v:(v,t) \in E} y_{vt} = 1 \\ & \textstyle\sum_{v:(u,v) \in E} y_{uv} = \sum_{v:(v,w) \in E} y_{vw} \qquad \forall w \in V \setminus \{s, t\} \\ & y_e \geq 0. \end{aligned} \tag{7}$$

In this one we're sending one unit of flow from s to t, where the cost of sending a unit of flow on an edge equals its length ℓ_e. Naturally the cheapest way to send this flow is along a shortest s-t path length. So both the LPs should compute the same value. Let's see how this follows from duality.

3.1 Duals of Each Other

Take the first LP. Since we're setting d_s to zero, we could hard-wire this fact into the LP. So we could rewrite (6) as

$$\begin{aligned} \max \quad & d_t \\ \text{subject to} \quad & d_v - d_u \leq \ell(u, v) \qquad \forall (u, v) \in E, s \notin \{u, v\} \\ & d_v \leq \ell(s, v) \qquad \forall (s, v) \in E \\ & -d_u \leq \ell(u, s) \qquad \forall (u, s) \in E \end{aligned} \tag{8}$$

Moreover, the distances are never negative for $\ell(u, v) \geq 0$, so we can add in the constraint $d_v \geq 0$ for all $v \in V$.

How to find an upper bound on the value of this LP? The LP is in the standard form, so we can do this mechanically. But let us do this from starting from the definition of the dual as the "best upper bound".

Let us define $E_s^{out} := \{(s,v) \in E\}$, $E_s^{in} := \{(u,s) \in E\}$, and $E^{rest} := E \setminus (E_s^{out} \cup \mathbb{E}_s^{in})$. For every arc $e = (u,v)$ we will have a variable $y_e \geq 0$. We want to get the best upper bound on d_t by linear combinations of the the constraints, so we should find a solution to

$$\sum_{e \in E^{rest}} y_{uv}\,(d_v - d_u) + \sum_{e \in E_s^{out}} y_{sv}\,d_v - \sum_{e \in E_s^{in}} y_{us}\,d_u \geq d_t \tag{9}$$

(this is like $\mathbf{y}^T A \geq c$) and the objective function is to

$$\text{minimize} \sum_{(u,v) \in E} y_{uv}\,\ell(u,v). \tag{10}$$

(This is like $\min \mathbf{y}^T \mathbf{b}$.) Great, the objective function (10) is exactly what we want, but what about the craziness in (9)? Just collect all copies of each of the variables d_v, and it now says

$$\sum_{v \neq s} d_v \left(\sum_{u:(u,v) \in E} y_{uv} - \sum_{w:(v,w) \in E} y_{vw} \right) \geq d_t.$$

First, this must be an equality at optimality (since otherwise we could reduce the y values). Moreover, these equalities must hold regardless of the d_v values, so this is really the same as

$$\begin{aligned} \sum_{u:(u,v) \in E} y_{uv} - \sum_{w:(v,w) \in E} y_{vw} &= 0 \qquad \forall v \notin \{s,t\}. \\ \sum_{u:(u,t) \in E} y_{ut} - \sum_{w:(t,w) \in E} y_{tw} &= 1. \end{aligned} \tag{11}$$

Summing all these inequalities for all nodes $v \in V \setminus \{s\}$ gives us the missing equality:

$$\sum_{w:(s,w) \in E} y_{sw} - \sum_{u:(u,s) \in E} y_{us} = 1.$$

Finally, observe that since there's flow conservation at all nodes, and the net unit flow leaving s and reaching t, this means we must have a possibly-empty circulation (i.e., flow going around in circles) plus one unit of s-t flow. Removing the circulation can only lower the objective function, so at optimality we're left with one unit of flow from s to t. This is precisely the LP (7), showing that the dual of LP (6) is LP (7), after a small amount of algebra.