Duality

Uri Feige

November 17, 2011

1 Linear programming duality

1.1 The diet problem revisited

Recall the diet problem from Lecture 1. There are n foods, m nutrients, and a person (the *buyer*) is required to consume at least b_i units of nutrient i (for $1 \le i \le m$). Let a_{ij} denote the amount of nutrient i present in one unit of food j. Let c_i denote the cost of one unit of food item i. One needs to design a diet of minimal cost that supplies at least the required amount of nutrients. This gives the following linear program.

minimize $c^T x$
subject to
$Ax \ge b$
$x \ge 0$

Now assume that some other person (the *seller*) has a way of supplying the nutrients directly, not through food. (For example, the nutrients may be vitamins, and the seller may sell vitamin pills.) The seller wants to charge as much as he can for the nutrients, but still have the buyer come to him to buy nutrients. A plausible constraint in this case is that the price of nutrients is such that it is never cheaper to buy a food in order to get the nutrients in it rather than buy the nutrients directly. If y is the vector of nutrient prices, this gives the constraints $A^T y \le c$. In addition, we have the nonnegativity constrain $y \ge 0$. Under these constraints the seller wants to set the prices of the nutrients in a way that would maximize the sellers profit (assuming that the buyer does indeed buy all his nutrients from the seller). This gives the the following *dual* LP:

maximize $b^T y$
Subject to
$A^T y \le c$
$y \ge 0$

As one can replace any food by its nutrients and not pay more, one gets *weak duality*, namely, the dual provides a lower bound for the primal. Weak duality goes beyond the diet problem and holds even if A, b, c have some entries that are negative. That is, for every pair of feasible solutions to the primal and dual LPs we have:

$$b^T y \le (Ax)^T y = x^T A^T y \le x^T c = c^T x \qquad (1)$$

In particular, weak duality implies that if the optimal value of the primal is unbounded then the dual is infeasible, and if the optimal value of the dual is unbounded, then the primal is infeasible.

Assume that there is a pair of solutions x^* and y^* for which the values of the primal and dual LPs are equal, namely $c^T x^* = b^T y^*$. Then necessarily both x^* and y^* are optimal solutions to their respective LPs. In economics, the vector y^* is refereed to as *shadow prices*. These optimal solutions need to satisfy the inequalities of (1) with equality. This gives the following *complementary slackness* conditions:

$$(Ax^* - b)^T y^* = 0 \tag{2}$$

$$(c - A^T y^*)^T x^* = 0 \tag{3}$$

Condition (2) has the following economic interpretation. If a certain nutrient is in surplus in the optimal diet, then its shadow price is free (a *free good*). Condition (3) can be interpreted to say that if a food is overpriced (more expensive than the shadow price of its nutrients) then this food does not appear in the optimal diet.

1.2 Derivation of dual for standard form

Consider a (primal) linear program in standard form.

minimize $c^T x$
suject to
$Ax = b$
$x \geq 0$

Let us derive a lower bound on its value. The lower bound is obtained by removing the constraints $Ax = b$, and replacing them by a charging mechanism for violated constraints. There is a price y_i for one unit of violation of constraint $A_i x = b_i$. The following lower bound is valid for every vector y of prices (because feasible solutions of the primal LP give the same value for this new LP):

minimize (over x**)** $c^T x + y^T(b - Ax)$
suject to
$x \geq 0$

Rearranging the objective function, one gets:

minimize (over x**)** $y^T b + (c^T - y^T A)x$

Unless $c^T - y^T A \geq 0^T$, this lower bound can be made $-\infty$, and hence useless. So we shall restrict attention to y satisfying $y^T A \leq c^T$.

If we now remove the nonnegative $(c^T - y^T A)x$ from the objective function, we still have a lower bound. Moreover, this lower bound no longer depends on x, so we may remove the minimization over x, and the constraints $x \geq 0$. Hence for every y satisfying $y^T A \leq c^T$, $y^T b$ is a lower bound on the primal LP. What is the best (highest) lower bound that we can get by this approach? This is simply the solution to the dual LP:

maximize (over y**)** $b^t y$
subject to
$A^T y \leq c$

Observe that complementary slackness condition (3) can be extracted from the above development. Complementary slackness condition (2) is trivially true for LPs in standard form.

1.3 Strong duality

If the primal is feasible and its optimal solution is bounded, then the lower bound given by the dual is tight. Namely, the value of the dual is equal it this case to value of the primal. We shall show this for LPs in standard form.

The proof we give is based on the stopping condition for the simplex algorithm. Recall that at that point $X_B = B^{-1}b$, and the reduced costs are nonnegative, namely $c_j - c_B^T B^{-1} A_j \geq 0$, or in matrix notation:

$c^T - c_B^T B^{-1} A \geq 0^T$.

Let us "guess" the following solution to the dual: $y^T = c_B^T B^{-1}$. Then

$y^T A = c_B^T B^{-1} A \leq c^T$

satisfying the constraints of the dual. Hence y is a feasible solution to the dual. Its value is:

$b^T y = y^T b = c_B^T B^{-1} b = c_B^T x_B = c^T x$,

implying that the lower bound is tight.

1.4 LPs in general form

The following table explains how to obtain the dual of a primal LP that is in general form. Here A_j denotes a row of matrix A and A^j denotes a column.

$\min c^T x$		$\max b^T y$
$A_i x \geq b_i$	$i \in I^+$	$y_i \geq 0$
$A_i x = b_i$	$i \in I^=$	y_i free
$x_j \geq 0$	$j \in J^+$	$y^T A^j \leq c_j$
x_j free	$j \in J^=$	$y^T A^j = c_j$

Note that the dual of the dual is the primal. (Verifying this is left as homework.)

Weak and strong duality apply also in this case. More specifically, if the optimum to the primal is bounded, then so is the optimum to the dual, and vice versa. If the optimum to one of the LPs is unbounded, then the other is not feasible. It may also happen that neither one of them is feasible, as can be shown by the following example:

minimize $-x_1 + 2x_2$
subject to
$x_1 - x_2 = -1$
$-x_1 + x_2 = 2$

Coincidently, its dual happens to be the same:

maximize $-y_1 + 2y_2$
subject to
$y_1 - y_2 = -1$
$-y_1 + y_2 = 2$

1.5 A geometric view of the dual

Consider the primal LP

minimize $c^T x$
subject to
$Ax \geq b$

Its dual is:

maximize $b^t y$
subject to
$A^T y = c$
$y \geq 0$

Hence in a feasible solution to the dual, c is expressed as a positive combitation of the rows of A. The rows of A are the normals to the hyperplanes defining the primal LP, pointing into the polyhedron. In the optimal solution to the dual, we get more (value for the objective function) by including those normals with highest values of b. Intuitively, those are the constraints that are hardest to satisfy in the primal.

1.6 Some easy consequences of duality

Using duality it is a simple matter to show that the following problem is in $NP \bigcap coNP$:

Given a minimization LP, does it have a feasible solution of value at most k?

A witness for this statement is a feasible solution to the LP with value at most k. A refutation is a solution for the dual of value higher than k. Both solutions have complexity polynomial in the input size.

We can also show that finding a feasible solution to an LP is essentially as hard as finding the optimal solution (when the optimum is bounded). Given an LP, consider a new LP composed of all constraints of the dual and the primal, and the additional constraints that both objective functions must have equal value. Its size is polynomially related to the size of the original LP. A feasible solution to this new LP is an optimal solution to the original LP.

1.7 Farkas' lemma

Farkas' lemma states that given a system of equation $Ax = b$ for which we seek a nonnegative solution (meaning $x \geq 0$), then either the system has a solution, or (and this is an *exclusive or*) there is the following type of explanation for the insolvability of the system: there is some vector y such that $y^T b > 0$, but $y^T A \leq 0$.

Geometrically, this vector y is a normal to a hyperplane that separates b from the columns of A.

To prove Farkas' lemma we use the following LP primal-dual pair (though it would be fair to note that historically, Farkas' lemma predated LP duality).

minimize $0^T x$
subject to
$Ax = b$
$x \geq 0$

and

maximize $b^t y$
subject to

$y^t A \leq 0$

If the primal is feasible, its optimal value is 0, which upper bounds the dual, refuting the second alternative.

If the primal is not feasible, then the dual must be unbounded (because it is feasible, taking $y = 0$), proving the second alternative.

1.8 The minimax theorem

A simple example of a two-person zero-sum game is "stone, paper, scissors" (which I suppose is the translation of "even niyar umisparaim"). In such games, there are two players. Each player has a set of possible moves, and both players move simultaneously (without prior knowledge of the other player's move). A payoff matrix (known to both players) specifies the payoff to the column player as a function of the two moves. The payoff to the row player is $-A$. (The gain of one player is equal to the loss of the other).

What is the best move in such a game? In general, there is no best move. The payoff that results from a move changes based on the move of the other player. If one player knows the move of the other player, then its easy for him to play the game optimaly. The question is how to play the game when one does not know the move of the other player. Clearly, keeping your own move secret from the other player is important. The "mathematical" way of generating secrets is via randomness, and this calls for a randomized strategy, or in game theory terminology, a *mixed* strategy. What should this strategy optimize? The conservative view is to seek a strategy that carries the least amount of risk. Namely, whatever the move of the other player is, the expected payoff is maximized. This is a maxmin approach: maximize (over your choice of moves) the minimum (over all possible moves of the other player) expected payoff.

Finding the optimal mixed strategy can be formulated as a linear program. Let A be a payoff matrix for the column player whose columns specify moves for the maximizing player and rows specify moves for the minimizing player. The payoff matrix to the row player is then $-A$. Let x be a vector of probabilities for the possible moves of the maximizing player. Then we have the following LP:

maximize v (the maxmin payoff)

subject to

$1^T x = 1$ (the probabilities sum up to 1),

$Ax \geq v$ (for every move of the row player, the expected payoff is at least v),

$x \geq 0$ (probabilities are nonnegative).

For the minimizing player, let y be the vector of probabilities for his possible moves. Then we have:

maximize u (the maxmin payoff)

subject to

$1^T y = 1$ (the probabilities sum up to 1),

$-A^T y \geq u$ (for every move of the column player, the expected payoff to the row player is at least u),

$y \geq 0$ (probabilities are nonnegative).

The famous minimax theorem of Von Neumann says that the values of the optimal conservative mixed strategies for the two players are equal (to the negative of each other). The mixed strategies are in equilibrium. Each player can tell the other player his own mixed strategy (which is not the same as revealing your move in advance), and still, no (conservative) player will have an incentive to change his own mixed strategy.

The proof of the minimax theorem is a simple consequence of linear programming duality. (Historically, the minimax theorem was proved before LP duality.) To see this, observe that v and u are variables. Change the main constraint for the column player to $-Ax + v \leq 0$. For the row player, change the objective function to **minimize** $-u$, and the main constraint to $-A^T y - u \geq 0$.

The complementary slackness conditions say that a move has probability 0 unless it is optimal against the strategy of the other player.

Another point worth noticing is that the support of the mixed strategy of each player need not be larger than the number of moves available to the other player (as the optimum is attained at a vertex).

1.9 Yao's Lemma

A randomized algorithm can be viewed as a distribution over deterministic algorithms. The success probability (or expected running time) of a randomized algorithm is its success probability over a worst case input. Yao's lemma (an immediate application of the minimax theorem) says that there is a worst case distribution over inputs such that the average success probability of the best deterministic algorithm over this distribution is equal to the success probability of the best randomized algorithm over the worst case input. This principle is often used in order to prove lower bounds for randomized algorithms. One exhibits a distribution on which no deterministic algorithm does well.

1.10 Total unimodularity

Definition 1 *A matrix is totally unimodular if each square submatrix of it has a determinant that is either 0,1, or -1.*

Lemma 1 *If A is totally unimodular and A and b are integral, then all vertices of the polytope are integer. It particular, the optimal solution (with respect to any linear objective function) is integer.*

Proof: Consider the constraints that are tight in the optimal solution, and let x_B be the basic variables that correspond to them. Then $x_B = B^{-1}b$. Recall that Cramer's rule shows that $x_j = \frac{detB^j}{detB}$ where B^j denotes here the matrix B with column j replaced by b. By Cramer's rule and total unimodularity of A, x_B is integer. ($x_N = 0$.) $\square$

1.11 The max-flow min-cut theorem

The vertex-arc incidence matrix of a directed graph has vertices as its rows, arcs as its columns, and for arc $e = (i, j)$, its column has $+1$ in row i, -1 in row j, and 0 elsewhere.

Proposition 2 *The vertex-arc incidence matrix of a digraph is totally unimodular.*

Proof: Consider a square submatrix. If it has a 0 column or row, then its determinant is 0. So we can assume that each row and column contain at least one ± 1 entry. If there is a row (or column) with exactly one ± 1 entry, remove this row and the column on which the ± 1 lies. The determinant changes only by a multiplicative factor of ± 1. Eventually, either the determinant is seen to be ± 1, or every column has exactly two ± 1 entries. Summing up the rows 0ne gets the 0 vector, implying that the rows are linearly dependent and the submatrix has determinant 0. □

In the max flow problem, there is a directed graph with nonnegative capacities b_{ij} on its edges. One has to send a maximum flow from vertex 1 to vertex n, obeying capacity constraints and preservation of flow. Namely, one seeks the solution to the following linear program:

maximize x_{n1}
subject to:
$\sum_j x_{ij} - x_{ji} = 0$ for all $1 \leq i \leq n$,
$x_{ij} \leq b_{ij}$ for all i, j. Here x_{n1} is not constrained.
$x_{ij} \geq 0$ for all i, j.

Proposition 3 *If all capacities are integer, then the maximum flow is integer.*

Proof: We can add nonnegative slack variables to obtain constraints $x_{ij} + z_{ij} = b_{ij}$. The resulting constrain matrix is totally unimodular. □

In the min-cost flow problem, one has to route f units of flow from v_1 to v_n. Every edge has a cost c_{ij} per sending one unit of flow through it. The above LP can be modified by adding the constraint $x_{n1} = f$, and changing the objective function to **minimize** $\sum c_{ij} x_{ij}$. Again we find that when f is integer, then the minimum cost flow is integer (though its cost need not be integer, if the c_{ij} are noninteger).

The dual of the max-flow LP can be seen to be as follows. We associate a variable y_i with every vertex and y_{ij} with every arc except for $(n, 1)$. One gets:

minimize $\sum b_{ij} y_{ij}$
subject to
$y_n - y_1 \geq 1$
$y_{ij} + y_i - y_j \geq 0$
$y_{ij} \geq 0$.

Again, total unimodularity applies, and the optimal solution is integer. It is interesting to observe that in fact in every BFS all variables have values in $\{0, 1, -1\}$. This follows from observing that when applying Cramer's rule, not only the denominator but also the numerator is totally unimodular. Together with the nonnegativity constraints this implies that all y_{ij} are in $\{0, 1\}$. Let us observe that the optimal value of the LP is not affected by adding the constraint $y_i \geq 0$ for every vertex v_i (because the y_i are not part of the objective function, and all constraints remain satisfied if a fixed constant is added to all y_i). Hence without loss of generality we may assume that we have an optimal basic feasible solution that satisfies these additional nonnegativity constraints, which by total unimodularity (adding nonnegativity constraints does not affect total unimodularity) has also all its y_i variables in $\{0, 1\}$.

Now consider a directed cut in the graph, where on the one side V_0 we have the vertices with $y_i = 0$ (which includes vertex v_1), and on the other side V_1 we have vertices with $y_i = 1$ (which includes vertex v_n). The constraints of the LP imply that only edges from V_0 to V_1 need have $y_{ij} = 1$, whereas all other edges have $y_{ij} = 0$. Hence the value of the LP is equal to the value of a directed cut between v_1 and v_n. As the LP is a minimization problem, it gives the minimum directed cut. Hence strong LP duality implies the well known theorem that the maximum flow is equal to the minimum cut.