28 Duality in Linear Programming

As we have noted , a linear program is an optimization problem with linear constraints and a linear objective function.

If there are n variables and m constraints (in addition to constraints that each of the variables must be non-negative), we can write such an optimization problem as

$$\text{SUM}_k\ A_{jk}x_k\ + s_j\ - b_j = 0 \text{ for each } j$$

and each x_k and s_j is non-negative for each j and k, and we are to maximize $\text{SUM}_k\ v_k x_k$.

As we have also noted, the form of the equations here is one in which the s_j variables are *solved for* in the sense that each appears with coefficient 1 in exactly one equation. This form is convenient for evaluating these s variables when all the x variables are set to 0, so **we associate the origin in the x variables with this form for the equations**.

We can similarly associate every vertex or intersection of n linearly independent constraints with a form of the equations that is *solved for* all the variables which are not 0 at that vertex.

We saw that a vertex is a vertex of the allowed or feasible polytope if all the b parameters are non-negative, and that we can change the form of the equations to move the origin along an edge of that polytope by performing a **pivot operation**. This involves choosing one of the x variables, solving an appropriate equation for it, and using that equation to eliminate the given x variable from all equations and from the objective function.

If we write $-b_j = A_{j0}$ and $c_k = A_{0k}$, we find that the step of eliminating x_K from all the equations but the J equation takes the form

A_{jk} becomes $A_{jk} - A_{jK}A_{Jk}/A_{JK}$ for all j and k from 0 to m or n, except for J and K.

This operation has an obvious symmetry between row and column, which suggests that there is some kind of transposed problem that we can define, which will transform in the same way under this kind of pivot operation.

The symmetry between row and column manifests itself in a second way as well: the condition for feasibility is that the A_{j0} are negative, while the condition at a feasible vertex that the objective function is a minimum is that the A_{0k} are negative. Thus the condition for the maximum feasible point is symmetric as well.

There is indeed such a transposed problem hich, if defined as follows, has important and very useful properties. It is called the **dual** to the original LP, which is often referred to as the **primal** problem. (Of course the dual of the dual is the primal

again, so that the distinction between those is really about which problem we defined first.)

In the dual problem we define a second variable for each of the variables of the primal including the slack variables, here, **a y variable for each s variable** and **a t variable for each x variable** of the primal.

We define the dual to obey the conditions:

1. The dual equations transform under a pivot exactly as do the primal equations, so that the dual is uniquely determined by the primal equations, and does not depend on the form used to describe them
2. The solution conditions for the dual are the same as those for the primal, so that a solution for one determines a solution for the other.
3. Any feasible solution to the dual provides an upper bound to the primal solution, while any solution to the primal provides a lower bound to the dual solution.

All this can be accomplished by writing the dual equations as

$$\text{SUM}_j \ y_j A_{jk} \ - c_k \ - t_k = 0,$$

and requiring that the dual objective function, $\text{SUM}_j \ y_j \ b_j$ be **minimized at the dual solution, whose feasibility requires that the c_k are negative**. (We have to make the signs of the t_k variables what they are here so that the solution will occur for c_k negative, as it does in the primal problem. Minimizing here means that a solution corresponds to the b_j non-negative, again what we need here.)

Here the t_k are called **dual slack variables**, and the y_j are called **dual variables**. Again each are non-negative in the standard form for an LP.

With this form for the dual, not only do the equations transform in the same way under a pivot operation, **but both the primal and the dual objective functions change by the same amount, by $A_{J0} \ A_{0K}/A_{JK}$ in each pivo**t. A JK pivot interchanges the roles of x_K and s_J in the primal and of y_J and t_K in the dual. The resulting constant term in the objective function is therefore the same for the primal and dual problems at every vertex.

Notice that with these equations we can evaluate the double sum $\text{SUM}_{j,k} y_j A_{jk} x_k$ in either order, and draw the conclusion, after rearranging the four resulting terms:

$$\text{SUM}_j \ y_j b_j \ - \text{SUM}_k \ c_k \ x_k = \text{SUM}_j \ s_j \ y_j \ + \text{SUM}_k \ x_k \ t_k$$

The left hand side here is **the difference between the dual objective function and the primal objective function**. The right side is **always positive if the variables represent feasible solutions to the primal and dual problems.**

We can deduce from this equation not only that **any feasible dual point has dual objective function greater than the primal maximum, (and any primal feasible solution has objective function less than it**) but at the solution point, where the b_j are positive and the q_k are negative for each problem, and the objective functions are the same for both**, at least one of each pair of variables (x_k and t_k, or y_j and s_j) must be 0, in order that the right hand side here be 0.**

At this solution point the origin will not be the original origin since the x and s variables will be mixed up; but the *solved for* primal variables will be given by the b_j at that point, and the rest will be 0; the *solved for* dual variables will be given by the $-c_k$ and the rest of those will be 0.

The dual problem is often written without the dual slack variables, just as the primal problem is often written without the s_j. The dual inequalities then read:

$$\text{SUM}_j \ y_j \ A_{jk} >= c_k.$$

Of course **at any feasible point for the primal problem other than the solution maximum itself**, the dual problem willl **not** be feasible at the corresponding dual point, and vice versa.

Exercise: The act of writing down explicit dual equations of a given LP is not complicated but it is quite unnatural to human beings, and you can expect to do it hesitantly and incorrectly the first three times you try to do it. Therefore please attempt to do it with random initial LP's at least four times.

2. Of What Use is the Dual?

The dual problem can be useful for solving LP's either because your origin is feasible for it rather than for the primal, or because there are fewer variables in it, or it just looks nicer.

Also the bound that the dual gives on solutions to the primal are sometimes useful in themselves.

There are even methods for attacking problems which use both the primal and the dual problems alternately.

Finally, the duality theorem, that primal and dual objective functions are the same at solutions, provides sometimes useful information as to the nature of the problems you are dealing with. Often the duality theorem contains a not obvious statement about the nature of the problems under consideration.

For example, flow problems in networks can be described as lps as we shall see. The dual of a flow problem is equivalent to the problem of finding a minimum **cut** in the

network. The statement that the amount of the maximum flow in a network is the size of a minimum cut, is an example of a non-trivial and useful duality statement.

3. Network Flow:

There are a number of practical problems related to flow of commodities or fluids in networks, which can be phrased as linear programs.

A network is another name for a graph. Suppose one has a graph and in it a source vertex and a sink vertex. A substance can flow in this network from its source to its sink, along edges of the graph. You can ask: **how should this flow be routed in the network to maximize it and how much flow can be achieved in a given network?**

There are typically two kinds of constraints in such a problem. There are limitations on how much flow can pass through each edge, and sometimes how much can pass through a vertex. Also there are conservation constraints at vertices other than the source and sink. These are all typically linear.

There are also variations in this problem in which there are several commodities with different sources and sinks. These can be modeled similarly as LPs.

For each directed edge e of the graph, let vfe be the vertex at its head and vte be the vertex at its tail.

(For convenience we split each edge into directed edges in each direction the way we tend to split roads).

`   Let the amount of flow in edge e be x_e. Then the constraints on this problem are

For each edge $0 <=$ x_e $<= C_e$ where C_e is the capacity of the edge.

For each vertex v other than the source and sink,

SUM (over edges with front end at v) x_v = SUM(over edges with tail at v)x_v.

and we want to maximize the flow into the sink vertex which is like the left hand side of this equation at the sink vertex.

What then is the dual to this LP?

We have a variable y_e for each equation corresponding to the capacity constraint on it, a variable z_v at each intermediate vertex corresponding to the conservation constraint there.

And what constraints do they appear in?

There is a constraint corresponding to each variable x_e; since x_e appears with coefficient 1 in its capacity constraint and with coefficient -1 (say) in its front end constraint, and 1 in its back end constraint, we get the dual constraint.

$$y_e \ >= \ z_{vfe} - z_{vte}$$

There is no z variable at the source or sink but the z variable at the sink is replaced by 1 coming from the coeffient of the edges leading into it in the objective functions.

We can interpret the z variables as a potential defined on the vertices, which is 0 at the source and 1 at the sink, and the edge variables must be at least the difference of potential across the edge,

The dual objective function here is $SUM_e \, C_e y_e$. If we find a minimum capacity cut between the source and the sink, and set y=0 on the source side of the cut, and y= 1 on the sink side, this sum becomes the minimum cut capacity.

On the other hand, if we define the potential on each edge as a linear function between its values on its ends, the points at which the potential takes on any value but 0 or 1 itself forms a cut, and the objective function will be the average of the capacity of these cuts.

Thus the minimum value of the objective function will be the minimum cut capacity, which tells us, by the duality theorem, that

The maximum flow in a network is the capacity of a minimum cut.

This result is a typical duality theorem, and these are often valuable facts.

Exercise: Suppose you have several commodities with different sources and sinks that can flow in the same network. Set up a linear programming formulation of this problem.
You can maximize the total value of the flows.