Lecture notes 5: Duality in applications

Vincent Conitzer

We have already seen how to take the dual of a linear program in general form. However, when we are solving a problem using linear programming, it can be very enlightening to take the dual of the linear program *for that particular problem.* Typically, in the context of the problem under study, it is possible to give a natural interpretation to the dual variables, and this also often leads to natural interpretations of weak/strong duality and complementary slackness. We will illustrate this by considering the duals of some applications that we studied previously.

1 Combinatorial auctions

Let us again consider the combinatorial auction winner determination problem. Because we want to consider linear programs rather than (mixed) integer programs, we assume that bids are partially acceptable. We have the following linear program formulation from before:

maximize $\sum_b v_b x_b$
subject to
$(\forall i \in I)\ \sum_b a_{ib} x_b \leq 1$
$(\forall b)\ x_b \in [0, 1]$

We recall that a_{ib} is an indicator parameter that is set to 1 if item i occurs in bid b, and to 0 otherwise. When taking duals of linear programs, it helps to simplify the formulation as much as possible. In this case, assuming that each bid bids on at least one item, the item constraints already ensure that $x_b \leq 1$ for each bid. So, we can simplify the program to:

maximize $\sum_b v_b x_b$
subject to
$(\forall i \in I)\ \sum_b a_{ib} x_b \leq 1$
$(\forall b)\ x_b \geq 0$

To get started, let us first consider the example instance from before. There are 4 items, A, B, C, D, and we receive the following bids: $(\{A, B\}, 4), (\{B, C\}, 5), (\{A, C\}, 4), (\{A, B, D\}, 7), (\{D\}, 1)$. The abstract linear program above instantiates to:

maximize $4x_1 + 5x_2 + 4x_3 + 7x_4 + x_5$
subject to
$x_1 + x_3 + x_4 \leq 1$
$x_1 + x_2 + x_4 \leq 1$
$x_2 + x_3 \leq 1$
$x_4 + x_5 \leq 1$
$x_1, x_2, x_3, x_4, x_5 \geq 0$

We recall that the optimal solution to this linear program is to set $x_1 = 0, x_2 = 0.5, x_3 = 0.5, x_4 =$

$0.5, x_5 = 0.5$, for a total value of 8.5. The dual of this linear program is:

minimize $y_A + y_B + y_C + y_D$
subject to
$y_A + y_B \geq 4$
$y_B + y_C \geq 5$
$y_A + y_C \geq 4$
$y_A + y_B + y_D \geq 7$
$y_D \geq 1$
$y_A, y_B, y_C, y_D \geq 0$

We can use complementary slackness to help us find the optimal solution to this dual. Because the variables x_2, x_3, x_4, x_5 are set to positive values in the optimal solution to the primal, the last four constraints of the dual (not considering the nonnegativity constraints) must be binding. The unique solution to the system of equalities
$y_B + y_C = 5$
$y_A + y_C = 4$
$y_A + y_B + y_D = 7$
$y_D = 1$

is to set $y_A = 2.5, y_B = 3.5, y_C = 1.5, y_D = 1$. Indeed, this solution gives us $y_A + y_B + y_C + y_D = 8.5$, as expected.

But what do these dual variables *mean*? To figure this out, let us first take the dual of the abstract linear program, which results in:

minimize $\sum_{i \in I} y_i$
subject to
$(\forall b)\ \sum_{i \in I} a_{ib} y_i \geq v_b$
$(\forall i \in I)\ y_i \geq 0$

Let us think of the y_i as *prices* of the items i. The main constraint in the dual then says that for each bid, the value of that bid is at most the sum of the prices of the items in that bid. That is, the bids are (weakly) underbidding the prices of the items. The objective is simply the sum of all prices. Then, weak duality means that if the prices are set in such a way that the constraints hold, then the auctioneer will never be able to get a revenue greater than the sum of the prices. This makes perfect sense intuitively: every bidder is bidding at most these prices on her desired bundle of items, so the auctioneer cannot expect to get more than the sum of these prices! Strong duality means that there is in fact a feasible way to set the prices such that the auctioneer can obtain the sum of these prices. Complementary slackness states that if a bid is (partially) accepted in the optimal solution ($x_b > 0$), then it must bid exactly the sum of the optimal prices of the items in that bid ($\sum_{i \in I} a_{ib} y_i = v_b$). It also states that if a price is set to a positive value in the optimal solution ($y_i > 0$), then this item must be sold completely in the optimal solution ($\sum_b a_{ib} x_b = 1$). Conversely, complementary slackness states that if both of these hold (for feasible solutions), then the solutions are in fact optimal.

2 Combinatorial reverse auctions

Now let us consider combinatorial reverse auctions, for which we have the following linear program (as long as bids are partially acceptable).

minimize $\sum_b v_b x_b$
subject to
$(\forall i \in I)\ \sum_b a_{ib} x_b \geq 1$
$(\forall b)\ x_b \geq 0$

Again, we have removed the constraint that $x_b \leq 1$: in fact, it cannot help the auctioneer to set $x_b > 1$, because setting it to 1 will already cover all of the items in b. The dual becomes:

maximize $\sum_{i \in I} y_i$
subject to
$(\forall b)\ \sum_{i \in I} a_{ib} y_i \leq v_b$
$(\forall i \in I)\ y_i \geq 0$

Again, we interpret the y_i as prices. Now, the constraint says that the value of each bid is *at least* the sum of the prices of the items in that bid—that is, the bidders are (weakly) overbidding the prices. Weak duality means that the auctioneer will have to spend at least the sum of all the prices, which makes intuitive sense because every bid is overbidding the prices. Strong duality means that there is a feasible way of setting the prices so that the auctioneer only has to pay the sum of these prices. Complementary slackness states that if a bid is (partially) accepted in the optimal solution ($x_b > 0$), then it must bid exactly the sum of the optimal prices of the items in that bid ($\sum_{i \in I} a_{ib} y_i = v_b$). It also states that if a price is set to a positive value in the optimal solution ($y_i > 0$), then this item must not be overbought in the optimal solution, that is, exactly one of this item is bought ($\sum_b a_{ib} x_b = 1$).

3 Game theory

Let us consider zero-sum games again. We recall that a minimax strategy for the column player is a probability distribution over the columns that minimizes the maximum expected utility for the row player, where the maximum is taken over all rows. We have the following linear program for finding a minimax strategy for the column player.

minimize u
subject to
$(\forall i)\ u - \sum_j p_j u_1(i,j) \geq 0$
$\sum_j p_j = 1$
$(\forall j)\ p_j \geq 0$

We can assume that $u_1(i,j) > 0$ for all i, j. This is without loss of generality because if we add a constant to all utilities in the game, this does not affect the game strategically. Given this, it never helps to set $\sum_j p_j > 1$, because this will only increase u. Hence, we can equivalently reformulate the linear program as:

minimize u
subject to
$(\forall i)\ u - \sum_j p_j u_1(i,j) \geq 0$
$\sum_j p_j \geq 1$
$(\forall j)\ p_j \geq 0$

Let us now take the dual of this linear program. We use dual variables q_i for the row constraints in the primal, and v for the probability constraint in the primal.

maximize v
subject to
$(\forall j)\ v - \sum_i q_i u_1(i,j) \leq 0$
$\sum_i q_i \leq 1$
$(\forall i)\ q_i \geq 0$

As you can see, this dual program looks somewhat similar to the primal. In fact, it corresponds to the problem of finding a maximin strategy for the row player—that is, a probability distribution over the rows that maximizes the minimum expected utility for the row player, where the minimum is taken over all columns. The q_i are the probabilities on the rows, and v is the expected utility that the row player can guarantee herself. For each column j, there is a constraint that v can be at most the expected utility that the row player would get if the column player plays j; and the last constraint says that the total probability can be at most 1 (and in fact, an optimal solution will never use a total probability less than 1, because all the row player utilities are positive).

In this context, weak duality means that if the row player can guarantee herself an expected utility of at least v, and the column player can guarantee that the row player has an expected utility of at most u, then $v \leq u$, which makes perfect sense. Strong duality means that the maximum expected utility that the row player can guarantee herself is in fact equal to the minimum expected utility that the column player can force the row player to have. That is, $\min_{(p_1,\ldots,p_n)} \max_i \sum_{j=1}^{n} p_j u_1(i,j) = \max_{(q_1,\ldots,q_m)} \min_j \sum_{i=1}^{m} q_i u_1(i,j)$—a result known as the *Minimax Theorem*. So, the Minimax Theorem follows as a corollary of strong duality. Complementary slackness means that if in a minimax strategy, the column player puts positive probability on a column ($p_j > 0$), then, if the row player plays a maximin strategy and the column player plays j, the row player will receive a utility of exactly v, the utility that she is guaranteeing herself ($v = \sum_i q_i u_1(i,j)$). In other words, column j is a *best response* to the row player's maximin strategy. Complementary slackness also implies that the same is true if we swap the roles of the row and column players ($q_i > 0 \Rightarrow u = \sum_j p_j u_1(i,j)$). This means that if the row player plays a maximin strategy and the column player plays a minimax strategy, then the strategies are in *Nash equilibrium*, that is, each player is (always) playing a best response against the other player. Conversely, it also means that if the players are playing a Nash equilibrium (in a two-player zero-sum game), they must be playing a maximin strategy and a minimax strategy, respectively. These are all fundamental results in the theory of zero-sum games that follow naturally from the theory of duality.

4 Markov decision processes

Let us consider the problem of finding an optimal policy for a Markov decision process (MDP) again. We previously used the following linear program for this:

minimize $\sum_s v_s^*$
subject to
$(\forall s,a)\ v_s^* - \gamma \sum_{s'} P(s,a,s') v_{s'}^* \geq R(s,a)$

Taking the dual, we get

maximize $\sum_{s,a} y_{s,a} R(s,a)$
subject to
$(\forall s)\ \sum_a y_{s,a} - \gamma \sum_{s'} \sum_a P(s',a,s) y_{s',a} = 1$
$(\forall s,a)\ y_{s,a} \geq 0$

Because the v_s^* variables are free, the corresponding constraints in the dual are equality constraints. How can we interpret this dual? In fact, the dual will be easier to interpret if we divide the primal objective by the number of states $|S|$. Hence, the primal objective becomes **minimize** $\sum_s v^*_s/|S|$—the average value of the states, or the expected value if the initial state is chosen uniformly at random. The dual can then be written as follows:

maximize $\sum_{s,a} y_{s,a} R(s,a)$
subject to
$(\forall s)\ \sum_a y_{s,a} = 1/|S| + \gamma \sum_{s'} \sum_a P(s', a, s) y_{s',a}$
$(\forall s, a)\ y_{s,a} \geq 0$

One of our interpretations of discounting is that in each round, the probability that the Markov decision process continues another round is γ. With this interpretation of discounting, we can give the following interpretation to the dual variables: $y_{s,a}$ is the expected total number of times that we are in state s and take action a. Under this interpretation, the expected number of times that we arrive in a given state s is $1/|S| + \gamma \sum_{s'} \sum_a P(s', a, s) y_{s',a}$, because we have a $1/|S|$ probability of the first state being s, and, every time that we are in a state s' and take action a, we have a probability of $\gamma P(s', a, s)$ that the MDP will continue another round and that we end up in s in this next round. The constraint states that this number must be equal to $\sum_a y_{s,a}$, which is the expected number of times that we leave from state s. Under this constraint, we seek to maximize $\sum_{s,a} y_{s,a} R(s,a)$, our total expected reward. (We do not have to discount in the objective because the discounting is already taken care of by the constraint, which enforces that there is only a γ probability of continuing another round.)

One could consider this dual linear program more natural than the primal; indeed, unlike the primal problem, feasible solutions to this dual problem correspond to objective values that can be obtained. So, perhaps it is more natural to think of this program as the primal, and the original one as the dual. This is a matter of taste.

5 Maximum flow

Let us again consider the linear program for the maximum flow problem:

maximize $\sum_{w \in V:(s,w) \in E} f_{sw}$
subject to
$(\forall (v,w) \in E)\ f_{vw} \leq c_{vw}$
$(\forall v \in V - \{s,t\})\ \sum_{u \in V:(u,v) \in E} f_{uv} - \sum_{w \in V:(v,w) \in E} f_{vw} = 0$
$(\forall (v,w) \in E)\ f_{vw} \geq 0$

(Actually, we have changed the program slightly, maximizing the flow out of the source vertex s rather than maximizing the flow into the sink vertex t, but these two quantities must be the same.) The dual of this program is given as follows, where the y_{vw} correspond to the capacity constraints on the edges, and the g_v correspond to the flow constraints on the vertices (the latter are free variables because they correspond to equality constraints):

minimize $\sum_{(v,w) \in E} c_{vw} y_{vw}$
subject to
$(\forall (v,w) \in E, v \neq s, w \neq t)\ y_{vw} + g_w - g_v \geq 0$
$(\forall w : (s,w) \in E)\ y_{sw} + g_w \geq 1$
$(\forall v : (v,t) \in E)\ y_{vt} - g_v \geq 0$
$(\forall (v,w) \in E)\ y_{vw} \geq 0$

This dual is perhaps a little more difficult to interpret. It helps to know that it always has an optimal solution in which every variable is set to either 0 or 1 (we will show this later on). Given this, we can think of the g_v variables as specifying a subset of the vertices, where $g_v = 1$ indicates that v is inside the subset, and $g_v = 0$ indicates that v is outside of the subset. Because there are no variables g_s and g_t, let us, by convention, say that s is inside the subset and t is outside of the subset. Then, the constraints can be interpreted as follows: for an edge (v, w), if v is inside the subset and w is outside the subset, then we must set y_{vw} to 1 (which comes at a penalty of c_{vw}). Of course, we would like to minimize this penalty, but we will have to set some of the y_{vw} to 1: this is because s is inside the set and t is outside the set, so on any path from s to t, there will be some edge (v, w) such that v is inside the set and w is outside the set, so that y_{vw} must be set to 1. Hence, the edges (v, w) such that $y_{vw} = 1$ specify a *cut* in the graph: if these edges are removed, there is no longer any way to get from s to t. The vertices v with $g_v = 1$ are the vertices on s's side of the cut, and those with $g_v = 0$ are those on t's side of the cut. Conversely, it can be seen that any cut in fact corresponds to a feasible solution to the dual. Hence, the dual asks for the *minimum cut* in the graph, where the weight of a cut is the sum of the capacities of the edges in the cut.

Weak duality means that for any cut, the weight of that cut is an upper bound on the flow from s to t. This makes intuitive sense, because all flow from s to t must cross the cut somewhere, and hence the total capacity of the cut is an upper bound on our flow. Strong duality now means that there is in fact a cut whose weight is exactly equal to the maximum flow. This is known as the *Maximum Flow/Minimum Cut Theorem.* Complementary slackness implies that in optimal solutions: 1. If an edge has positive flow ($f_{vw} > 0$), then the edge does not go backwards across the cut ($y_{vw} + g_w - g_v = 0$, so $g_w \leq g_v$). 2. If an edge is in the cut ($y_{vw} > 0$), then the full capacity of that edge must be used in the flow ($f_{vw} = c_{vw}$).

6 Rank aggregation (the Kemeny rule)

Finally, we once again consider the problem of aggregating rankings according to the Kemeny rule. For this, we have the following integer program:

maximize $\sum_{a \neq b} n_{ab} x_{ab}$
subject to
$(\forall a, b : a \neq b)\ x_{ab} + x_{ba} = 1$
$(\forall a, b, c : a \neq b, b \neq c, c \neq a)\ x_{ab} + x_{bc} + x_{ca} \leq 2$
$(\forall a, b : a \neq b)\ x_{ab} \in \{0, 1\}$

This integer program is slightly different from the one we gave before: this one seeks to maximize the number of agreements rather than minimize the number of disagreements, but this is equivalent in the sense that the optimal rankings will be the same. If we take this program very literally, then there is a lot of redundancy in these constraints. For example, if A and B are distinct alternatives (elements to be ranked), then we have the constraint $x_{AB} + x_{BA} = 1$ as well as the constraint $x_{BA} + x_{AB} = 1$. Also, if A, B, and C are distinct alternatives, we have the following three cycle constraints: $x_{AB} + x_{BC} + x_{CA} \leq 2$, $x_{BC} + x_{CA} + x_{AB} \leq 2$, and $x_{CA} + x_{AB} + x_{BC} \leq 2$. If we have to introduce a dual variable for every redundant constraint, it will make our job more difficult, so we should try to write the program in a more concise way. Let $< a, b >$ denote a pair of distinct alternatives, with the understanding that $< a, b > = < b, a >$. Similarly, let $< a, b, c >$ denote a cycle of three distinct alternatives, with the understanding that $< a, b, c > = < b, c, a > = < c, a, b >$ (but we do *not* have $< a, b, c > = < a, c, b >$, since these two cycles have opposite orientations). Then, we can rewrite the program as:

maximize $\sum_{a \neq b} n_{ab} x_{ab}$
subject to
$(\forall < a, b >)\ x_{ab} + x_{ba} = 1$
$(\forall < a, b, c >)\ x_{ab} + x_{bc} + x_{ca} \leq 2$
$(\forall a, b : a \neq b)\ x_{ab} \in \{0, 1\}$

Before taking the dual, we first take the LP relaxation (so the result will only be an upper bound on the number of agreements that we can obtain). We do not need the constraint $x_{ab} \leq 1$ because this is already implied by $x_{ab} + x_{ba} = 1$ and $x_{ba} \geq 0$. Hence, we have the following primal linear program:

maximize $\sum_{a \neq b} n_{ab} x_{ab}$
subject to
$(\forall < a, b >)\ x_{ab} + x_{ba} = 1$
$(\forall < a, b, c >)\ x_{ab} + x_{bc} + x_{ca} \leq 2$
$(\forall a, b : a \neq b)\ x_{ab} \geq 0$

Taking the dual, we get:

minimize $\sum_{<a,b>} y_{<a,b>} + \sum_{<a,b,c>} 2y_{<a,b,c>}$
subject to
$(\forall a, b : a \neq b)\ y_{<a,b>} + \sum_{c \notin \{a,b\}} y_{<a,b,c>} \geq n_{ab}$
$(\forall < a, b, c >)\ y_{<a,b,c>} \geq 0$

Any feasible solution to this dual gives an upper bound to the LP relaxation of the rank aggregation problem, and hence it gives an upper bound on the original rank aggregation problem as well. There is no guarantee that we can get a feasible dual solution that has the same objective value as the optimal solution to the original problem (with integrality constraints), because the optimal objective value for the LP relaxation can be higher than the optimal solution value for the original problem.

But how can we interpret such an upper bound? For this, it is helpful to consider a graph whose vertices are the alternatives, and for any two alternatives a and b, there is an edge from a to b with weight n_{ab} (corresponding to n_{ab} rankings that rank a ahead of b, that is, n_{ab} opportunities for an agreement). Now, let us interpret setting $y_{<a,b>} = k$ as taking weight k away from each of the edges (a, b) and (b, a). We note that the final ranking can only agree with k of this $2k$ weight. Similarly, let us interpret setting $y_{<a,b,c>} = k$ as taking weight k away from each of the edges $(a, b), (b, c)$, and (c, a). Because these edges form a cycle, the final ranking at best agrees with two of these edges; that is, of the $3k$ weight, the final ranking can agree with at most $2k$. Now, in order to get a feasible solution to the dual, we must take weights away in this manner until we have taken all the weight away from the graph: $y_{<a,b>} + \sum_{c \notin \{a,b\}} y_{<a,b,c>}$ is the weight we take away from (a, b), and this is required to be at least n_{ab}. Each time that we take some weight away, we consider how much of this weight a final ranking could have agreed with, and add this to the objective. That is, if we increase $y_{<a,b>}$ by k, we add k to the objective, and if we increase $y_{<a,b,c>}$ by k, we add $2k$ to the objective. In fact, this results in the objective in the dual. In the end, we have accounted for all the weight, so our objective must be an upper bound on the weight that any final ranking agrees with; this corresponds to the weak duality property.