Lecture 6

Duality of LPs and Applications*

Last lecture we introduced duality of linear programs. We saw how to form duals, and proved both the weak and strong duality theorems. In this lecture we will see a few more theoretical results and then begin discussion of applications of duality.

6.1 More Duality Results

6.1.1 A Quick Review

Last time we saw that if the primal $(\mathcal{P})$ is

$$\begin{aligned} \max \quad & c^\top x \\ s.t. \quad & Ax \leq b \end{aligned}$$

then the dual $(\mathcal{D})$ is

$$\begin{aligned} \min \quad & b^\top y \\ s.t. \quad & A^\top y = c \\ & y \geq 0. \end{aligned}$$

This is just one form of the primal and dual and we saw that the transformation from one to the other is completely mechanical. The duality theorem tells us that if $(\mathcal{P})$ and $(\mathcal{D})$ are a primal-dual pair then we have one of the three possibilities

1. Both $(\mathcal{P})$ and $(\mathcal{D})$ are infeasible.
2. One is infeasible and the other is unbounded.
3. Both are feasible and if x^* and y^* are optimal solutions to $(\mathcal{P})$ and $(\mathcal{D})$ respectively, then $c^\top x^* = b^\top y^*$.

*Lecturer: Anupam Gupta. Scribe: Deepak Bal.

6.1.2 A Comment about Complexity

Note that the duality theorem (and equivalently, the Farkas Lemma) puts several problems related to LP feasibility and solvability in $\mathsf{NP} \cap$ co-NP.

E.g., Consider the question of whether the equational form LP $Ax = b, x \geq 0$ is feasible. If the program is feasible, we may efficiently verify this by checking that a "certificate" point satisfies the equations. By taking this point to be a vertex and appealing to Hwk1 (Problem 4), we see that we may represent this certificate point in size polynomial in the size of the input. On the other hand, if the LP is infeasible, then Farkas Lemma (Form 1 from Lecture 5) says we can find a $y \in \mathbb{R}^m$ with $y^\top A \geq 0$ and $y^\top b < 0$. Again appealing to Homework 1, we may find a succinctly represented solution to this set of equations, thus providing a "certificate" for the infeasibility of the original LP.

We can similarly ask for whether the value of the LP $\max\{c^\top x \mid Ax \leq b\}$ is at least δ or not. Again, if we have n variables and m equations, we can convert this general-form LP into an equivalent equational form LP with $O(m+n)$ constraints and variables, and whose size is not much more. Now, if there is a solution with value at least δ, we can show a BFS x^* for this equivalent LP—this will have polynomial size, for the same reasons. And if there is no such solution of value δ or higher, there is a solution to the dual $\min\{b^\top y \mid A^\top y = c, y \geq 0\}$ of value strictly less than δ and we can give this dual solution. (Again this "proof" will be polynomial-sized.) Hence the decision problem "is the value of this maximization LP at least δ" is in $\mathsf{NP} \cap$ co-NP.

6.1.3 Duality from Lagrange Multipliers

Suppose we have the problem $(\mathcal{P})$

$$\begin{array}{rl} \max & c^\top x \\ s.t. & Ax \leq b \end{array}$$

where as usual, the constraints are $a_i x \leq b_i$ for $i \in [m]$. Let $K = \{x \mid Ax \leq b\}$. Now consider the situation wher we are allowed to violate the constraints, but we penalize a violation of the ith constraint at a rate of $\lambda_i \geq 0$ per unit of violation. Let $\lambda = (\lambda_1 \ldots, \lambda_m)^\top$ and define

$$g(x, \lambda) := c^\top x + \sum_{i \in [m]} \lambda_i (b_i - a_i x).$$

Then we see that for each feasible $x \in K$, $\lambda \geq 0$, we get $g(x, \lambda) \geq c^\top x$. So now letting x be unconstrained we have that

$$g(\lambda) := \max_{x \in \mathbb{R}^n} g(x, \lambda) \geq OPT(\mathcal{P}).$$

In other words, for each λ, $g(\lambda)$ provides an upper bound on the optimal value of the LP. Naturally, we may ask for the best upper bound acheived in this way, *i.e.*,

$$g* = \min_{\lambda \geq 0} g(\lambda).$$

Putting together our definitions, we get

$$\begin{aligned} g* &= \min_{\lambda \geq 0} \max_{x} \left\{ c^\top x + \lambda^\top (b - Ax) \right\} \\ &= \min_{\lambda \geq 0} \left(b^\top \lambda + \max_{x} \left\{ (c^\top - \lambda^\top A) x \right\} \right) \end{aligned}$$

If $c^\top - \lambda^\top A$ has any non-zero entries, then the maximum over all x is ∞ which gives us a useless upper bound. Hence we really should only consider λ which satisfy $A^\top \lambda = c$. So all in all, this is

$$\begin{aligned} \min \quad & b^\top \lambda \\ s.t. \quad & A^\top \lambda = c \\ & \lambda \geq 0. \end{aligned}$$

which is the dual! So we see that the technique of Lagrange multipliers in this context is really just a form of duality. We will return to Lagrange multipliers later when dealing with more general convex optimization problems.

6.1.4 Complementary Slackness

Often times, the following theorem is very useful.

Theorem 6.1. *Suppose we have the primal dual pair $(\mathcal{P}), (\mathcal{D})$ from Section 6.1.1. If $(\mathcal{P}), (\mathcal{D})$ are both feasible with x^*, y^* feasible solutions, then following are equivalent*

1. *x^*, y^* are both optimal.*
2. *$c^\top x^* = b^\top y^*$.*
3. *$(y^*)^\top (Ax^* - b) = 0$*

In words, property 3 means that at optimality, either a dual variable is 0 or its corresponding inequality is tight (or both). Equivalently, for all constraints $i \in [m]$, if $y_i^* > 0$, then $a_i x = b_i$. Here we use the non-negativity of y^* and the fact that x^* is feasible.

Proof. 1 and 2 are equivalent by the duality theorem. We will prove 2 and 3 are equivalent. Suppose 2 holds. Then $c^\top x^* = (y^*)^\top b$ and on the other hand $c^\top x^* = (y^*)^\top A x^*$ since y^* is feasible. This holds if and only if $(y^*)^\top (Ax^* - b) = 0$ which is 3. □

6.2 Applications of Duality

In this section we will discuss two applications of duality. First the max-flow/min-cut theorem which was discussed in Lecture 3 without mention of duality. Then we will discuss König's Theorem on bipartite graphs.

6.2.1 Max-Flow = Min-Cut

In this problem, we are given a directed graph $G = (V, A)$ with two "special" vertices $s, t \in V$ called the source and sink. We are also given capacities c_e for all $e \in A$. The max-flow problem (or more formally, the max-(s,t)-flow problem) is to find an assignment of flows on the edges which obeys the capacities and maximizes the total amount of flow from s to t. For our purposes, we will formulate this differently than in Lecture 3.

Let P_{st} be the set of all paths from s to t. Note that P_{st} likely has size large comparaed to the number of nodes and arcs. Let f_p represent the flow assigned to path $p \in P_{st}$. Then the max-flow problem, which we will consider our primal problem $(\mathcal{P})$ is formulated as

$$\begin{aligned} \max \quad & \sum_{p \in P_{st}} f_p \\ s.t. \quad & \sum_{p \ni e} f_p \le c_e \quad \forall e \in A \\ & f_p \ge 0 \quad \forall p \in P_{st}. \end{aligned}$$

Note in this formulation, there may be exponentially many variables, but according to earlier results, in any BFS there will be at most $|A|$ many non-zero variables. The dual formulation $(\mathcal{D})$ is then

$$\begin{aligned} \min \quad & \sum_{e \in A} c_e x_e \\ s.t. \quad & \sum_{e \in p} x_e \ge 1 \quad \forall p \in P_{st} \\ & x_e \ge 0 \quad \forall e \in A. \end{aligned}$$

We may think of x_e as the length of the edge e. Thus $c_e x_e$ represents the "volume" of the edge e. So this dual problem is saying, find a "volume-minimizing" assignment of lengths to the edges so that every s-t path has length at least 1. The duality theorem tells us that the max flow (optimal value for $(\mathcal{P})$) is equal to this value. But our goal is to show that max-flow is equal to the min-(s,t)-cut! So we'd better show that this dual value actually equals the min-(s,t)-cut (which we call the min-cut in the rest of the discussion, for brevity).

Soon, we will see that this dual actually has 0-1 BFS's. With this information it is obvious that $(\mathcal{D})$ will represent a minimum cut. Let us ignore this for now though, and prove the result with what we have.

For an s-t cut $(S, \overline{S})$, let $E(S, \overline{S})$ represent the edges crossing the cut and let $c(S, \overline{S})$ represent the sum of capacities of edges crossing the cut. Then for any (s,t)-cut $(S, \overline{S})$, we can let $x_e = 1$ for all $e \in E(S, \overline{S})$ and $x_e = 0$ for all others. Then this is clearly feasible for $(\mathcal{D})$. Consequently we have that

$$OPT(\mathcal{D}) \le \text{min-}(s,t)\text{-cut}.$$

Now we must show the other, less trivial direction.

Theorem 6.2. *Suppose x is a solution of $(\mathcal{D})$ of value $c^\top x$. Then there exists an (s,t)-cut $(S, \overline{S})$ such that $c(S, \overline{S}) \le c^\top x$.*

Proof. As above, we may interpret the x_e's as edge lengths. Let $d(v)$ be the shortest path distance from s to v for all $v \in V$ according to the lengths x_e. The x_e's are all non-negative so this is well defined. Note that $d(s) = 0$ and $d(t) \geq 1$ by the set of constraints in $(\mathcal{D})$.

Consider $\rho \in [0, 1)$. Let $S_\rho = \{v \in V \mid d(v) \leq \rho\}$. Then $(S_\rho, \overline{S}_\rho)$ is a feasible s-t cut in G. Now suppose ρ is chosen from $[0, 1)$ according to the uniform distribution. Then if we can show that

$$\mathbf{E}[c(S_\rho, \overline{S}_\rho)] \leq c^\top x$$

we will be done since this would imply that there exists a ρ with $c(S_\rho, \overline{S}_\rho) \leq c^\top x$. Note that

$$\mathbf{E}[c(S_\rho, \overline{S}_\rho)] = \sum_{e \in A} c_e \cdot \mathbf{Pr}[e \in E(S_\rho, \overline{S}_\rho)]$$

by linearilty of expectation. Let $e = (u, v)$ and let ρ^* be the smallest value so that $u \in S_{\rho^*}$. Then $\forall \rho \geq \rho^* + x_e$, $v \in S_\rho$. So $\mathbf{Pr}[u \in S_\rho,\ v \notin S_\rho] \leq x_e$, so

$$\mathbf{E}[c(S_\rho, \overline{S}_\rho)] = \sum_{e \in A} c_e \cdot \mathbf{Pr}[e \in E(S_\rho, \overline{S}_\rho)] \leq \sum_{e \in A} c_e \cdot x_e = c^\top x \qquad \square$$

So we have min-cut $\leq OPT(\mathcal{D})$, which proves that indeed max-flow is equal to min-cut by the duality theorem. In fact, we have proved that the polytope for $(\mathcal{D})$ is integral. Theorem 6.2 says that for any feasible solution x to the min-cut LP, and any cost vector c, there exists an integer s-t cut $(S_\alpha, \overline{S_\alpha})$ with cost at most $c^\top x$. Note that this s-t cut corresponds to an integer vector $y \in \mathbb{R}^{|A|}$ where $y_e = 1 \iff e \in E(S_\alpha, \overline{S}_\alpha)$ and $y_e = 0$ otherwise. This y is also feasible for the cut LP.

To see why the polyhedron K of $(\mathcal{D})$ is integer, consider any vertex x of K. By the definition of vertex, there is some cost function such that x is the unique minimizer for $\min\{c^\top x \mid x \in K\}$. But since $c^\top y \leq c^\top x$, and $y \in K$, it follows that $x = y$ and hence x is integral.

You may want to think about what information you can conclude about optimal flows/cuts using complementary slackness. E.g., we get that the paths carrying flow are all shortest paths according to the edge length x_e's: they all must have length 1. Similarly, if an edge has non-zero length according to the optimal dual solution, then it must be saturated in an optimal primal solution. (In fact, in *every* optimal primal solution.)

6.2.2 König's Theorem for Bipartite Graphs

Given an undirected bipartite graph $G = (U, V, E)$, a *matching* is a set of edges which do not intersect at any vertices. A *vertex cover* is a set of vertices S such that for all $e \in E$, $e \cap S \neq \emptyset$. Even though a vertex cover is covering edges, it is called a vertex cover because it is a set of vertices. To clarify, a vertex cover is a set of vertices: this is how one should keep from getting confused.

Theorem 6.3 (Kónig's Theorem)**.** *For bipartite graph $G = (U, V, E)$,*

$$\max\{|M| \ :\ M \text{ is a matching of } G\} = \min\{|S| \ :\ S \text{ is a vertex cover of } G\}$$

Proof. Let MM and MM_{LP} represent the cardinality of the maximum matching, the optimal value of the maximum matching LP relaxation. Similarly, let VC and VC_{LP} denote the cardinality of the minimum vertex cover, and the optimal value of the vertex cover LP relaxation respectively. So we have that MM_{LP} is given by

$$\begin{aligned} \max \quad & \sum_{(i,j)\in E} x_{ij} \\ s.t. \quad & \sum_{j:(i,j)\in E} x_{ij} \leq 1 \quad \forall i \in U \\ & \sum_{i:(i,j)\in E} x_{ij} \leq 1 \quad \forall j \in V \\ & x_{ij} \geq 0 \quad \forall (i,j) \in E. \end{aligned}$$

Then the dual is

$$\min \quad \sum_{i\in U} y_i + \sum_{j\in V} z_j$$

$$\begin{aligned} s.t. \quad y_i + z_j &\geq 1 \quad \forall (i,j) \in E \\ y_i,\, z_j &\geq 0 \quad \forall (i,j) \in E. \end{aligned}$$

Adding in an integrality constraint to this gives us VC, since any vertex cover is feasible for this LP. Hence we define this dual to be VC_{LP}. So using the notations above to represent both the problem formulations and the optimum values, we now know, using duality theory that

$$MM \leq MM_{LP} = VC_{LP} \leq VC.$$

If we can show that the two inequalities are actually equalities, we would be done. In fact we will show that the BFS's of MM_{LP} and VC_{LP} are both integral.

Claim 6.4. *Any BFS of MM_{LP} is integral. Hence, $MM = MM_{LP}$ for bipartite graphs.*

Proof. We essentially did this in Lecture 3, except there we had equality constraints. So in that case, we always could find a cycle of fractional values. Here, this might not be the case. Suppose we have a fractional extreme point. If we find a cycle, proceed as in the Lecture 3 proof. We may only find a tree of fractional values. Similarly to the proof in Lecture 3, we alternately raise and lower the values by ϵ along a path from a leaf to a leaf on this tree. Choose ϵ small enough so that none of the constraints become violated after the adjustment. We can average these "complementary" solutions to contradict the extremity of the original point. □

Claim 6.5. *Any BFS of VC_{LP} is integral. Hence, $VC = VC_{LP}$ for bipartite graphs.*

Proof. Let y^* be an optimal solution to VC_{LP} chosen so that y^* has a maximum number of integer components. It does not make sense that $y*$ would have any component > 1, so assume all are ≤ 1. Let F be the set of fractional vertices. If $F = \emptyset$, we are done. WLOG, suppose $F \cap U$ is larger than or the same size as $F \cap V$. Let $\epsilon = \min\{y_i^* \mid i \in F \cap U\}$. Then subtract ϵ from all the components in $F \cap U$ and add ϵ to all the components in $F \cap V$.

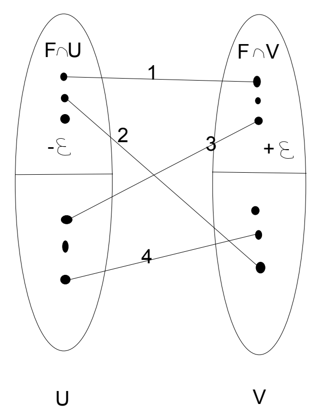

Figure 6.1: U and V

As seen in Figure 6.1, we need to check that constraints corresponding to edges of type 1 to 4 are still satisfied. Constraints of type 1 are not affected as epsilon is both added and subtracted. Constraints of type 4 are no affected at all and constraints of type 3 are trivially still satisfied. For constraints fo type 2, since the vertex in U was fractional and the vertex in V was not, the vertex in V must have had value 1! So subtracting ϵ from the U vertex will not violate the constraint. So we have a solution with objective function less than or equal to the original and with one less fractional component. This is a contradiction. Hence VC_{LP} has integer vertices and so $MM = MM_{LP} = VC_{LP} = VC$ and the theorem is proved. □

Putting it all together, we get that on bipartite graphs, the minimum cardinality vertex cover equals the maximum cardinality maximum matching. Note that this equality is false for general graphs (e.g., the 3-cycle shows a counterexample). □

An important aside: The proofs of Claims 6.4 and 6.5 show that the vertices of those LPs are integral: this fact is independent of what the objective function was. Indeed, such results immediately extend to weighted versions of the problems. E.g., we get that the *weighted* bipartite matching problem, where the edges have weights w_e, and the goal is to find the matching with the highest weight $\sum_{e \in M} w_e$, can be solved on bipartite graphs, just by finding a basic optimal solution to MM_{LP} with objective function $w^\top x$. Similarly, for the minimum weight vertex cover on bipartite graphs, we can seek to minimize $\sum_{i \in U} w_i y_i + \sum_{j \in V} w_j z_j$ subject to the constraints in VC_{LP}, and an optimal BFS gives us this min-weight vertex cover.

Another connection. Hall's theorem says that in a bipartite graph $G = (U, V, E)$, there is a matching M that matches all the vertices on the left (i.e. has cardinality $|U|$) if and only if every set $S \subseteq U$ on the left has at least $|S|$ neighbors on the right. König's theorem (which shows that the size of the maximum matching in G is precisely the size of the minimum vertex cover of G) is equivalent to Hall's theorem. We leave the proof for the reader.