# Lecture 10: Duality in Linear Programs

*Lecturer: Ryan Tibshirani* *Scribes: Hongyang Zhang, Rahul Ladhania, Yi Wei*



This lecture's notes illustrate some uses of various LaTeX macros. Take a look at this and imitate.

## 1.1 Review of Proximal Gradient Descent

Consider the problem

$$\min_x g(x) + h(x) \tag{1.1}$$

with $g$, $h$ convex, $g$ differentiable, and $h$ "simple" in so much as

$$\text{prox}_t(x) = \underset{z}{\text{argmin}} \frac{1}{2t}\|x - z\|_2^2 + h(x) \tag{1.2}$$

is computable. Proximal gradient descent: let $x^{(0)} \in \mathbb{R}^n$, repeat:

$$x^{(k)} = \text{prox}_{t_k}(x^{(k-1)} - t_k \nabla g(x^{(k-1)})), \quad k = 1, 2, 3, ... \tag{1.3}$$

The step sizes $t_k$ are chosen to be fixed and small, or by backtracking.

If $\nabla g$ is Lipschitz with constant $L$, then this has convergence rate $O(1/\epsilon)$. We can accelerate this to optimal rate $O(1/\sqrt{\epsilon})$.

## 1.2 Lower Bounds in Linear Programs

Suppose we want to find lower bound on the optimal value in our convex problem, $B \leq \min_x f(x)$. For example, consider the following simple linear program

$$\min_{x,y} x + y, \quad \text{subject to} \;\; x + y \geq 2, \; x, y \geq 0. \tag{1.4}$$

It is easy to see that the lower bound is $B = 2$, because one of the constraints is exactly the same as the objective function.

Let us try other problems. Suppose the linear program is

$$\min_{x,y} x + 3y, \quad \text{subject to} \;\; x + y \geq 2, \; x, y \geq 0. \tag{1.5}$$

We can have a lower bound for the objective function by $x + 3y = (x + y) + 2y \geq 2$.

More generally, for linear program

$$\min_{x,y} px + qy, \quad \text{subject to} \;\; x + y \geq 2,\; x, y \geq 0, \tag{1.6}$$

the constraint can be equivalently represented as

$$ax + ay \geq 2a,\; ax \geq 0,\; cy \geq 0,\; a, b, c \geq 0. \tag{1.7}$$

Adding them together, we have that

$$(a + b)x + (a + c)y \geq 2a. \tag{1.8}$$

Letting $a + b = p$ and $a + c = q$, we obtain that the lower bound of linear program (1.6) is given by $B = 2a$, for any $a$, $b$, and $c$ such that

$$a + b = p,\; a + c = q,\; a, b, c \geq 0. \tag{1.9}$$

But what is the best we can do? We can maximize our lower bound over all possible $a, b, c$:

$$\max_{a,b,c} 2a, \quad \text{subject to} \;\; a + b = p,\; a + c = q,\; a, b, c \geq 0. \tag{1.10}$$

This is also called the dual linear program of primal problem (1.6). Note that the number of dual variables is the number of primal constraints.

Now let us try another one:

$$\min_{x,y} px + qy, \quad \text{subject to} \;\; x \geq 0,\; y \leq 1,\; 3x + y = 2. \tag{1.11}$$

The constraint of the linear program can be equivalently represented as

$$ax \geq 0,\; -by \geq -b,\; 3cx + cy = 2c,\; a \geq 0,\; b \geq 0. \tag{1.12}$$

Adding them together, we have

$$(a + 3c)x + (-b + c)y \geq -b + 2c. \tag{1.13}$$

Let $p = a + 3c$ and $q = -b + c$, we obtain the dual problem

$$\max_{a,b,c} 2c - b, \quad \text{subject to} \;\; a + 3c = p,\; -b + c = q,\; a, b \geq 0, \tag{1.14}$$

as desired.

## 1.3 Example: Max Flow and Min Cut Problems

This is a nice example of duality in linear programs. The max flow and the min cut problems are essentially equivalent because of duality.

First, let us define what the max flow and min cut problems are.

For the max flow problem, given a directed graph, $G = (V, E)$, $c_{ij}$ is the capacity of the edge - the maximum amount of, say, a commodity that one can push through that edge. We have the flow over the graph, which is a vector $f$ that is assigned to all the edges in the graph. The flow across any edge $f_{ij}$ has to be non-negative (one cannot go against the direction of the edge) and one cannot push through more than the capacity $c_{ij}$ of that edge. Then, there is the important balancing property that the flow going into the node has to be equal to the flow coming out of the node. That is true for all the nodes except for the source ($v_s$) and the sink ($v_t$) nodes. Mathematically, it is represented thus:

$$f_{ij} \geq 0, (i,j) \in E$$
$$f_{ij} \leq c_{ij}, (i,j) \in E$$
$$\sum_{(i,k)\in E} f_{ik} = \sum_{(k,j)\in E} f_{ij}, k \in V \setminus \{s,t\}$$

In the problem, therefore, we want to maximize the total flow leaving the source, subject to satisying the aforementioned properties. Think of it as distributing some commodity across the graph: pushing it from the source to the sink.

$$\begin{aligned} \max_{f\in\mathbb{R}^{|E|}} \quad & \sum_{(s,j)\in E} f_{sj} \\ \text{subject to} \quad & f_{ij} \geq 0, f_{ij} \leq c_{ij} \text{for all } (i,j) \in E \\ & \sum_{(i,k)\in E} f_{ik} = \sum_{(k,j)\in E} f_{ij}, \text{for all } k \in V \setminus \{s,t\} \end{aligned} \tag{1.15}$$

Now, we can derive the dual of this, as this is just a linear program. We will not go through the dual derivation steps in class (it is for the homework).

Note that

$$\sum_{(i,j)\in E} (-a_{ij}f_{ij} + b_{ij}(f_{ij} - c_{ij})) + \sum_{k\in V\setminus\{s,t\}} x_k( \sum_{(i,k)\in E} f_{ik} - \sum_{(k,j)\in E} f_{kj}) \leq 0$$

for any $a_{ij}$,$b_{ij} \geq 0$, (i,j)$\in E$, and x$_k$, $k \in V \setminus \{s,t\}$

Rearrange as

$$\sum_{(i,j)\in E} M_{ij}(a,b,x) \leq \sum_{(i,j)\in E} b_{ij}c_{ij}$$

where $M_{ij}(a,b,x)$ collects terms multiplying $f_{ij}$

Want to make LHS in previous inequality equal to primal objective,

$$\text{i.e., } \left\{ \begin{array}{ll} M_{sj} = b_{sj} - a_{sj} + x_j & \text{want this} = 1 \\ M_{it} = b_{it} - a_{it} + x_i & \text{want this} = 0 \\ M_{ij} = b_{ij} - a_{ij} + x_j - x_i & \text{want this} = 0 \end{array} \right.$$

We have shown that primal optimal value $\leq \sum_{(i,j)\in E} b_{ij}c_{ij}$

subject to $a, b, x$ satisfying constraints. And thus, we end with this dual problem:

$$\begin{aligned} \min_{b\in\mathbb{R}^{|E|}, x\in\mathbb{R}^{|V|}} \quad & \sum_{(i,j)\in E} b_{ij}c_{ij} \\ \text{subject to} \quad & b_{ij} + x_j - x_i \geq 0 \quad \text{for all } (i,j) \in E \\ & b \geq 0, x_s = 1, x_t = 0 \end{aligned} \tag{1.16}$$

This is similar to what we did before. In the examples we did earlier, we started with a minimization problem and got a maximization dual problem. Here, we begin with a maximization problem, and end with a dual minimization problem. This is equivalent to deriving upper bound for the maximization problem. So here, we end up with our dual as another linear program - interestingly, this problem is the linear program relaxation of what is called the min cut problem.

What is the min cut problem? It is another well studied problem over graphs. And we have the same setup: we have directed edges, and capacity over the edges. In this problem, we want to divide our nodes into two groups, such that the sum of the capacites of the edges crossing the cut is the smallest over all possible configurations of the nodes in the two groups.

That is an integer problem. We are assigning nodes to one group or the other. We have binary variables in the optimization problem, which is written down here:

$$\begin{aligned} \min_{b\in\mathbb{R}^{|E|},x\in\mathbb{R}^{|V|}} \quad & \sum_{(i,j)\in E} b_{ij}c_{ij} \\ \text{subject to} \quad & b_{ij} \geq x_i - x_j \\ & b_{ij}, x_i, x_j \in \{0,1\} \\ & \text{for all } i,j \end{aligned} \tag{1.17}$$

It turns out that the dual of the max flow problem is just the LP relaxation of the integer min flow problem. So it takes the binary variables and allows them to be anything between 0 and 1 rather than be equal to 0 or 1.

By construction, we know value of max flow is less than or equal to that of the value of its dual. By construction, the dual here is an upper bound of the criterion in the primal, and as that is the relaxation of the min integer program, it is less than or equal to the capaxcity of the min cut. Note that by relaxing the integer min flow problem, we convexify the constraints, so the minimum will only get smaller. Thus, it is less than or equal to the capacity of the min cut.

Therefore, we have

value of max flow $\leq$ optimal value for LP relaxed min cut $\leq$ capacity of min cut

We have established a sequence of inequalities. A famous result called max flow min cut theorem is that these two things are the same - the capacity of the min cut and the value of the max flow problems are the same. And these problems are equivalent. Therefore, we have equalities for everything here. The min flow and max flow problems are equivalent.

Thus, this was just another example of strong duality. That primal and dual have exactly the same optimal values. The value of the max flow and capacity of the min cut are equal. In particular, from perspective of the current lecture, the dual problem did not have a gap betwen its criterion value and the primal problem.

## 1.4 Another Perspective on LP Duality

Primal:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & Gx \leq h \end{aligned}$$

Dual:

$$\max_{u,b} \quad -b^T u - h^T v$$

$$\text{subject to} \quad -A^T u - G^T v = c$$

$$v \geq 0$$

Let us introduce the second explanation of dual. For any $u$ and $v \geq 0$, and $x$ primal feasible, then for any $u$ and $v \geq 0$,

$$f^* \geq min_{x \in C}\ L(x,u,v) \geq min_x\ L(x,u,v) := g(u,v)$$

Note that

$$if\ c = -A^T u - G^T v,\ g(u,v) = -b^T u - h^T v$$

$$otherwise,\ g(u,v) = -\infty$$

Now we can maximize $g(u,v)$ over $u$ and $v \geq 0$ to get the tightest bound, and this gives exactly the dual LP as before. This last perspective is actually completely general and applies to arbitrary optimization problems (even nonconvex ones).
Example:
Setup: two players, J and R, and a payout matrix $P$
Game: if J chooses i and R chooses j, then J must pay R amount $P_{ij}$ (don't feel bad for J, this can be positive or negative).
They use mixed strategies, i.e., each will first specify a probability distribution, and then $P(J\ chooses\ i) = x_i,\ i = 1,...,m,\ P(R\ chooses\ j) = y_j,\ j = 1,...,n$. The expected payout then, from J to R, is $x^T P y$.
Now suppose that, because J is wiser, he will allow R to know his strategy x ahead of time. In this case, R will choose y to maximize $x^T P y$, which results in J paying off $max_{i=1,...,n}(P^T x)_i$.
J's best strategy is then to choose his distribution $x$ according to

$$\min_x \max_{i=1,\ldots,n} (P^T x)_i$$

$$\text{subject to} \ \ x \geq 0,\ 1^T x = 1$$

In an alternate universe, if R were somehow wiser than J, then he might allow J to know his strategy y beforehand.
By the same logic, R's best strategy is to choose his distribution $y$ according to

$$\max_y \min_{j=1,\ldots,m} (Py)_j$$

$$\text{subject to} \ \ y \geq 0,\ 1^T y = 1$$

Call R's expected payout in first scenario $f_1^*$, and expected payout in second scenario $f_2^*$. Because it is clearly advantageous to know the other player's strategy, $f_1^* \geq f_2^*$. But by Von Neumman's minimax theorem: we know that $f_1^* = f_2^*$.
Recast first problem as an LP:

$$\min_{s,t} \quad t$$

$$\text{subject to} \ \ x \geq 0,\ 1^T x = 1$$

$$P^T x \geq t$$

Now form what we call the Lagrangian:

$$L(x,t,u,v,y) = t - u^T x + v(1 - 1^T x) + y^T (P^T x - t1)$$

and what we call the Lagrange dual function:

$$g(u, v, y) = \min_{x,t} L(x, t, u, v, y) = v \ if \ 1 - 1^T y = 0,\ Py - u - v1 = 0$$

Hence dual problem, after eliminating slack variable u, is

$$\begin{aligned} \max_{y,v} \quad & v \\ \text{subject to} \quad & y \geq 0,\ 1^T y = 1 \\ & Py \geq v \end{aligned}$$

This is exactly the second problem, and therefore again we see that strong duality holds. In LPs, as we'll see, strong duality holds unless both the primal and dual are infeasible.