10-725: Convex Optimization **Spring 2023**

Lecture 10: February 16th

Lecturer: Siva Balakrishnan

Today we'll depart the world of algorithms and return to talking about the structure of convex programs. Our focus will be on understanding the concept of *duality*. We'll see some uses of the concept of duality as we go along.

We will begin with a discussion of duality in linear programs. Often in LPs, the dual (also an LP) will be a nice reformulation of the original LP, so just writing down the dual will give you some insight into the original program. We'll also see that duality will give us an answer to a very basic question in optimization, given a candidate solution $\widehat{x}$ can we give a certificate of its optimality (we've done things like this before) and if it's not optimal can we give reasonable bounds on its sub-optimality, i.e. $f(\widehat{x}) - f(x^*)$.

10.1 Linear Programs

Linear programs (LPs) are a special sub-class of convex optimization problems. They were the focus of intense research during WWII, and the period after that. An LP is simply an optimization problem:

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

where $c \in \mathbb{R}^d, A \in \mathbb{R}^{m \times d}, b \in \mathbb{R}^m, G \in \mathbb{R}^{r \times d}, h \in \mathbb{R}^r$.

10.1.1 Duality in LPs

The idea of duality will seem a bit strange at first. We're going to develop a different optimization program (the dual) whose value lower bounds the value of this linear program (which will now be called the primal).

We notice that, for any vector $u \in \mathbb{R}^m, v \in \mathbb{R}^r, v \geq 0$, and for any x which is feasible for the primal, we have that,

$$u^T(Ax - b) + v^T(Gx - h) \leq 0.$$

This can be re-written as:

$$(-A^T u - G^T v)^T x \geq -u^T b - v^T h.$$

Consequently, if we set $-A^T u - G^T v = c$, then we obtain that,

$$c^T x \geq -u^T b - v^T h.$$

So every u, v which satisfies the constraints that $v \geq 0$ and $-A^T u - G^T v = c$ gives us a lower bound $u^T b + v^T h$ on our primal objective value. So we could imagine trying to find the largest possible lower bound, i.e. we could solve the program:

$$\begin{aligned} \max_{u,v} \quad & -u^T b - v^T h \\ \text{subject to} \quad & -A^T u - G^T v = c \\ & v \geq 0. \end{aligned}$$

This program is called the dual of our original linear program. Lets make some quick observations:

1. The dual is also a linear program. It is a maximization program (in contrast to the primal which was a minimization program).

2. Each constraint in the primal, yields a variable in the dual. Conversely, each variable in the primal will yield a constraint in the dual (and typically we'll additionally have some non-negativity constraints).

3. By construction, if we denote the primal optimal value by p^*, and the dual optimal value by d^* then it is the case that, $p^* \geq d^*$. This is known as *weak duality.* It will turn out that under some additional conditions (say if the primal and dual problems are feasible) it will be the case that these two values are in fact equal – this is known as *strong duality* and we will revisit this later.

4. A useful exercise, is to rewrite the dual as a minimization LP, and then take its dual (can be done mechanically). What you will observe is that you will end up back at the primal (up to eliminating some variables, and switching signs again). Concisely, the dual of the dual LP is the primal LP. This fact also turns out to be true in more generality.

5. We will say that $p^* = \infty$ if the primal is infeasible (i.e. no x satisfies the constraints), and that $d^* = -\infty$ if the dual is infeasible. We will say that the primal is *unbounded* if $p^* = -\infty$ and the dual is unbounded if $d^* = \infty$.

   Weak duality then tells us the following facts: if the dual is unbounded, then the primal must be infeasible. Similarly, if the primal is unbounded then the dual must be infeasible.

10.1.2 Lagrangian Duality in LPs

Our eventual goal will be to derive dual optimization programs for a broader class of primal programs. The previous approach was tailored very specifically to linear objective functions (and linear constraints), and we won't in general be able to re-express the objective exactly as a combination of constraints.

The idea of Lagrange duality is a powerful generalization – it will look very similar to what we just did, but will be different in a useful way. Notice that, for any u and $v \geq 0$ and feasible x, we could always write:

$$c^T x \geq c^T x + u^T(Ax - b) + v^T(Gx - h).$$

This is true because for a feasible x the second term is 0 and the third term is negative. We will call this function (the RHS) the *Lagrangian* and denote it $L(x, u, v)$.

Now, given the above inequality, we could minimize both sides over feasible x, i.e. we could write:

$$p^* \geq \min_{x \text{ feasible}} L(x, u, v) \geq \min_x L(x, u, v) := g(u, v).$$

where we simply drop the constraints that ensure feasibility of x. This is convenient for us since we can explicitly minimize with respect to x. We can see that:

$$g(u, v) = \begin{cases} -b^T u - h^T v & \text{if } c = -A^T u - G^T v \\ -\infty & \text{otherwise.} \end{cases}$$

So we could simply define the dual problem to be:

$$\max_{u,v} \; g(u, v)$$
$$\text{subject to } v \geq 0.$$

This would be equivalent to our earlier LP dual. Notice, that we didn't explicitly use the linearity of our objective function anywhere (in contrast to our previous approach). As before, notice that by construction we have that weak duality holds, i.e. $p^* \geq d^*$ (where p^* and d^* are primal and dual optimal values).

10.2 Optimal Transport – Kantorovich

Usually most linear programming textbooks give an example of the duality between the maximum flow and minimum cut problems. Here is a different example that I think is quite fun.

The most classical example of LP duality comes from the work of Kantorovich in the context of optimal transport. Kantorovich invented all of these ideas (LPs, duality), in an infinite-dimensional context, to study the problem of optimal transport, and is usually considered the founder of the discipline of operations research (and of linear programming). We've seen a resurgence of interest in these ideas in ML (partly because of their connection to Wasserstein GANs).

Here is a simplified version of the problem of optimal transport. We have two distributions p and q, which are finite discrete measures supported on $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$, i.e. we can write:

$$
\begin{aligned}
p &= \sum_{i=1}^{n} \delta_{x_i} p_i, \\
q &= \sum_{j=1}^{m} \delta_{y_j} q_j.
\end{aligned}
$$

Our broad goal is to transport/re-arrange the mass from p to q.

We are given some cost matrix $C \in \mathbb{R}^{m \times n}$ where C_{ij} is the cost of moving a unit mass from x_i to y_j. (You can think of the cost as the distance between the points x_i and y_j).

Now, we would like to come up with a transport plan $M \in \mathbb{R}^{m \times n}$, where M_{ij} indicates the amount of mass we're moving from location x_i to location y_j. Ideally, we'd like our transport plan to have minimal cost. This corresponds to solving the following LP.

$$
\begin{aligned}
\min_{M} \quad & \sum_{ij} C_{ij} M_{ij}, \\
\text{subject to} \quad & \sum_{j=1}^{m} M_{ij} = p_i \text{ for all } i \in \{1, \ldots, n\} \\
& \sum_{i=1}^{n} M_{ij} = q_j \text{ for all } j \in \{1, \ldots, m\}, \\
& M_{ij} \geq 0, (i,j) \in \{1, \ldots, m\} \times \{1, \ldots, n\}.
\end{aligned}
$$

The constraints ensure that our transport plan actually moves all the mass of p to q.

Lets derive the dual of this problem. We first write the Lagrangian,

$$
L(M, u, v, w) = \sum_{ij} C_{ij} M_{ij} + \sum_{i=1}^{n} u_i \left(p_i - \sum_{j=1}^{m} M_{ij} \right) + \sum_{j=1}^{m} v_j \left(q_j - \sum_{i=1}^{n} M_{ij} \right) - \sum_{ij} w_{ij} M_{ij},
$$

where $w_{ij} \geq 0$. To derive the dual we simply minimize this function with respect to M, to obtain the dual function:

$$g(u,v,w) = \begin{cases} \sum_{i=1}^n u_i p_i + \sum_{j=1}^m v_j q_j, & \text{if } C_{ij} - u_i - v_j - w_{ij} = 0 \\ -\infty \ \ \text{otherwise.} \end{cases}$$

The dual LP is simply to maximize $g(u,v,w)$ with $w_{ij} \geq 0$. This yields the following dual LP:

$$\begin{aligned} \max_{u,v,w} \ & \sum_{i=1}^n u_i p_i + \sum_{j=1}^m v_j q_j \\ \text{subject to } & C_{ij} - u_i - v_j - w_{ij} = 0, \\ & w_{ij} \geq 0, \end{aligned}$$

or equivalently,

$$\begin{aligned} \max_{u,v} \ & \sum_{i=1}^n u_i p_i + \sum_{j=1}^m v_j q_j \\ \text{subject to } & u_i + v_j \leq C_{ij}. \end{aligned}$$

This dual is sometimes called a Shipper's problem. Say our original goal was to transport p to q. A shipper approaches us, and agrees to ship p to q for us, and we only have to pay the shipper costs for loading and unloading. The shipper tells us that the cost for loading a unit of mass at x_i is u_i and unloading at y_j is v_j.

For us to accept the deal, it seems reasonable to want that $u_i + v_j \leq C_{ij}$ (i.e. the cost we'd pay the shipper to load and unload should be less than the cost we'd pay to ship ourselves).

The shipper in turn will try to maximize his/her profit (the total loading, unloading price he/she can charge you) subject to you accepting the deal. So the shipper will attempt to solve the dual to decide loading/unloading costs.

Weak duality tells us that this will always be a good deal for us (i.e. the total amount of money we pay the shipper will be less than what it would have cost us to ship things ourselves). In this case, strong duality will tell you that a clever shipper (one who solves the dual) can make us pay him/her the same amount as we would have paid to ship things ourselves.

10.3 Lagrangian Dual in General

We will now start working with a broader class of optimization problems. Suppose we are interested in understanding a problem of the form:

$$\begin{aligned} \min_x \ & f(x) \\ \text{subject to } & h_i(x) \leq 0 \quad i \in \{1, \dots, m\} \\ & \ell_j(x) = 0, \quad j \in \{1, \dots, r\}. \end{aligned}$$

For now, nothing needs to be convex. We can proceed as before, and notice that for any u, $v \geq 0$, and feasible x, we have that,

$$f(x) \geq f(x) + \sum_{j=1}^{r} u_j \ell_j(x) + \sum_{i=1}^{m} v_j h_j(x) := L(x, u, v),$$

where $v \geq 0$. We can alternatively, define $L(x, u, v) = -\infty$ if any component of $v < 0$. Often the variables u, v are either referred to as dual variables or *Lagrange multipliers*.

We then have that,

$$p^* = \min_{x \text{ feasible}} f(x) \geq \min_x L(x, u, v) := g(u, v).$$

So as before, we can define our (Lagrange) dual problem as:

$$\begin{aligned} & \max_{u,v} g(u, v) \\ & \text{subject to } v \geq 0. \end{aligned}$$

Notice that defining this problem (and observing that weak duality holds) made no mention of convexity. These basic properties hold in general.

10.3.1 Dual is always concave maximization

We have already observed above that the dual problem could be defined, and lower bounds the primal problem, in general. Now, we'll additionally note that even if our primal constraints and objective are arbitrary (i.e. not convex) the dual function $g(u, v)$ is always a concave function. Consequently, the dual program is always "nice", i.e. involves maximizing a concave function over a convex set.

To see this we observe that,

$$g(u, v) = \min_x \left[f(x) + \sum_{j=1}^{r} u_j \ell_j(x) + \sum_{i=1}^{m} v_j h_j(x) \right],$$

is the pointwise minimum of a set of affine functions which is always concave.

10.3.2 Interpreting the Dual

There are lots of ways to think about what we're doing. One is to think of first re-writing the constraints as part of the objective. In this case, we would have that the primal is equivalent to:

$$\min_x f(x) + \sum_{j=1}^{r} \mathbb{I}(\ell_j(x) = 0) + \sum_{i=1}^{m} \mathbb{I}(h_j(x) \leq 0),$$

where the indicators are 0 if their condition is satisfied and ∞ otherwise. This function penalizes us infinitely for violating the constraints. We can view our Lagrange dual as similar in spirit but the penalty is softer, and depends on the magnitude of the Lagrange multipliers.

The Lagrange dual function is for $v \geq 0$,

$$g(u, v) = \min_x f(x) + \sum_{j=1}^{r} u_j \ell_j(x) + \sum_{i=1}^{m} v_j h_j(x).$$

If we satisfy the constraints, then the penalty is 0 for the equality constraints (i.e. the Lagrange multipliers have no effect). We are in fact "encouraged" to strictly satisfy the inequality constraints. On the other hand when we violate the constraints we pay a "linear" penalty (depending on the sign and magnitude of the Lagrange multipliers). The linear function can be quite a bad approximation of the indicator function (but not if we're judicious in our choice of the Lagrange multipliers). At the very least however we can observe that, $u_j \ell_j(x) \leq \mathbb{I}(\ell_j = 0)$, and $v_j h_j(x) \leq \mathbb{I}(h_j(x) \leq 0)$, so our linear penalty is at least an underestimate of the indicator penalty. This is just a different way of seeing that the dual function $g(u, v)$ lower bounds the primal.