Lecture Notes for Week 3

3.5 Dual problem

Duality is a very powerful tool for problem solving. For example, the dual problem may have nicer structure than the primal problem ('primal' basically means 'original') and can be solved more efficiently. Sometimes, as in the case of interior-point algorithm, solving primal and dual problems simultaneously in a symmetric way is more efficient than solving either primal or dual by themselves. Dual solution provides a certificate of optimality. Duality gap provides an effective termination criterion for iterative methods.

To motivate the dual problem, let us associate a Lagrange multiplier $\mu_i \in \mathbb{R}$ with the ith constraint in (18) and introduce the corresponding Lagrangian function:

$$L(x, \mu) = \langle c, x\rangle + \sum_{i=1}^{m} (b_i - \langle a_i, x\rangle)\mu_i.$$

Notice that, for a fixed $x \in K$, we have

$$\min_{\mu} L(x, \mu) = \begin{cases} \langle c, x\rangle & \text{if } \langle a_i, x\rangle = b_i \ \forall i \\ -\infty & \text{else} \end{cases}$$

so that (18) is equivalent to $\max_{x\in K} \min_{\mu} L(x, \mu)$. Switching the order of 'max' and 'min' yields the dual problem: $\min_{\mu} \max_{x\in K} L(x, \mu)$. (This has the interpretation of a 2-person game, with primal or dual problem corresponding to the primal player choosing its strategy first or second.) For a fixed $\mu \in \mathbb{R}^m$, we have

$$\begin{aligned} \max_{x\in K} L(x, \mu) &= \max_{x\in K} \left\langle c - \sum_{i=1}^{m} a_i\mu_i, x \right\rangle + \sum_{i=1}^{m} b_i\mu_i \\ &= \begin{cases} \sum_{i=1}^{m} b_i\mu_i & \text{if } c - \sum_{i=1}^{m} a_i\mu_i \in K^\circ \\ \infty & \text{else} \end{cases} \end{aligned}$$

Thus the dual of (18) is

$$\begin{array}{ll} \min_{\mu} & \sum_{i=1}^{m} b_i\mu_i \\ \text{s.t.} & c - \sum_{i=1}^{m} a_i\mu_i \in K^\circ. \end{array} \tag{19}$$

For those familiar with LP duality, this should look very familiar! By introducing $z = c - \sum_{i=1}^{m} a_i\mu_i$, this dual problem may be rewritten as

$$\begin{array}{ll} \min_{z,\mu} & \sum_{i=1}^{m} b_i\mu_i \\ \text{s.t.} & z + \sum_{i=1}^{m} a_i\mu_i = c \\ & z \in K^\circ. \end{array} \tag{20}$$

In fact, the primal problem (18) and the dual problem (20) can be rewritten to look nearly identical. Specifically, suppose there exists $d \in \mathbb{H}$ satisfying $\langle a_i, d\rangle = b_i$ for all i. (If no such d

exists, then (18) is clearly infeasible.) Then $\langle a_i, x\rangle = b_i$ iff $\langle a_i, x-d\rangle = 0$, so that (18) is equivalent to

$$\begin{array}{ll} \max_x & \langle c, x\rangle \\ \text{s.t.} & x - d \in \text{Null}(a_1,\ldots,a_m) \\ & x \in K \end{array}$$

where $\text{Null}(a_1,\ldots,a_m) = \{u \in H \mid \langle a_i, u\rangle = 0,\ i = 1,\ldots,m\}$. Also, $z = c - \sum_{i=1}^m a_i\mu_i$ satisfies $\sum_{i=1}^m b_i\mu_i = \sum_{i=1}^m \langle a_i, d\rangle\mu_i = \langle \sum_{i=1}^m a_i, d\rangle = \langle c - z, d\rangle = \langle c, d\rangle - \langle z, d\rangle$, so that (20) is equivalent to

$$\begin{array}{ll} \max_z & \langle z, d\rangle \\ \text{s.t.} & z - c \in \text{Span}(a_1,\ldots,a_m) \\ & z \in K^\circ \end{array}$$

where $\text{Span}(a_1,\ldots,a_m) = \{\sum_{i=1}^m a_i\mu_i \in H \mid \mu_i \in \mathbb{R},\ i = 1,\ldots,m\}$. Since $\text{Null}(a_1,\ldots,a_m)$ is the orthogonal complement of $\text{Span}(a_1,\ldots,a_m)$, we see that the rewritten primal and dual problems have the same form, namely, maximizing a linear function over the intersection of a translated subspace and a closed convex cone. They differ only in that their subspaces are orthogonal complements of each other and their cones are polars of each other. On the other hand, the algebraic representations of the subspaces, which are not unique, impact the numerical linear algebra involved in solving these problems. Thus, although the above rewrites are elegant, from a practical standpoint, the original primal and dual problems (18), (20) are more efficient to work with.

Let $\mathcal{F}(P)$ and $\mathcal{F}(D)$ denote the feasible sets of (18) and (20), i.e.,

$$\begin{aligned} \mathcal{F}(P) &= \{x \in K \mid \langle a_i, x\rangle = b_i,\ i = 1,\ldots,m\} \\ \mathcal{F}(D) &= \left\{(z,\mu) \in K^\circ \times \mathbb{R}^m \mid z + \sum_{i=1}^m a_i\mu_i = c\right\}. \end{aligned}$$

Two key relationship between the primal and dual problems are the following.

Weak duality: For any $x \in \mathcal{F}(P)$ and $(z,\mu) \in \mathcal{F}(D)$, we have $\langle c, x\rangle \le b^T\mu$.[8]

Strong duality: For an LP ($K = [0,\infty)^n$), it is known that if $\mathcal{F}(P) \ne \emptyset$ and $\mathcal{F}(D) \ne \emptyset$, then both (18) and (19) have optimal solutions x^* and (z^*,μ^*), and $\langle c, x^*\rangle = b^T\mu^*$. Also, if (18) has an optimal solution, then so does (19), and vice versa. (This in fact holds for any polyhedral cone K, and can be shown by either applying the simplex method or using a separating hyperplane argument.) However, both these facts are false for SOCP and SDP! For example, the SDP

$$\begin{array}{ll} \max_x & x_{12} \\ \text{s.t.} & x_{11} = 0 \\ & \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{pmatrix} \succeq 0 \end{array}$$

has the primal form (18) with $\mathbb{H} = \mathbb{S}^2$, $C = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$, and $A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. It's not hard to see that

[8] This follows from switching the order of "max" and "min". Specifically, $\langle c, x\rangle = \langle z + \sum_i a_i\mu_i, x\rangle = \langle z, x\rangle + \sum_i \langle a_i, x\rangle\mu_i = \langle z, x\rangle + \sum_i b_i\mu_i \le b^T\mu$.

it has as optimal solutions $X^* = \begin{pmatrix} 0 & 0 \\ 0 & \alpha \end{pmatrix}$ with $\alpha \geq 0$. Its dual problem (19) is

$$\begin{array}{ll} \min_\mu & 0 \\ \text{s.t.} & \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \mu \preceq 0. \end{array}$$

Dual feasibility is equivalent to $\begin{pmatrix} \mu & -\frac{1}{2} \\ -\frac{1}{2} & 0 \end{pmatrix} \succeq 0$, but this can never be satisfied since the determinant can never be nonnegative. Thus the primal problem has an optimal solution, and yet the dual problem is infeasible. Here the set of optimal solutions of the primal problem is unbounded. It can be shown that if the set of optimal solutions is bounded, then the dual problem must be feasible and their optimal values are equal (though the minimum in the dual may or may not be attained).

On the other hand, it can be shown that if both the primal and dual have feasible solutions in the interior of their cones, i.e.,

$$\mathcal{F}(P) \cap \text{int}K \neq \emptyset \quad \text{and} \quad \mathcal{F}(D) \cap (\text{int}K^\circ \times \mathbb{R}^m) \neq \emptyset, \tag{21}$$

then both have optimal solutions x^* and (z^*, μ^*) and there is zero duality gap, i.e., $\langle c, x^* \rangle = b^T \mu^*$. (It can also be shown that the sets of optimal x^* and z^* are convex compact.) The existence and compactness of optimal solutions are not hard to show by a contradiction argument. Zero duality gap can be proved nonconstructively by a separating hyperplane argument. However, we will prove this (for LP, SOCP, and SDP anyway) algorithmically, as a byproduct of the convergence analysis for interior-point methods.

In general, some work may be needed to put a problem in the primal or dual conic form. Consider the robust version of LP we saw in week 1:

$$\begin{array}{ll} \min_x & c^T x \\ \text{s.t.} & \bar{a}_i^T x + \|A_i^T x\|_2 \leq b_i, \quad i = 1, \ldots, m, \end{array} \tag{22}$$

where $c \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, $\bar{a}_i \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n \times p_i}$ are given. Rewriting the ith constraints as

$$\|A_i^T x\|_2 \leq b_i - \bar{a}_i^T x \quad \Longleftrightarrow \quad \begin{pmatrix} A_i^T x \\ b_i - \bar{a}_i^T x \end{pmatrix} \in \text{SOC}^{p_i+1} \quad \Longleftrightarrow \quad \begin{pmatrix} 0 \\ -b_i \end{pmatrix} - \begin{pmatrix} A_i^T \\ -\bar{a}_i^T \end{pmatrix} x \in -\text{SOC}^{p_i+1}$$

and letting $K = \text{SOC}^{p_1+1} \times \cdots \times \text{SOC}^{p_m+1}$, we can rewrite (22) as

$$\begin{array}{ll} \min_x & c^T x \\ \text{s.t.} & \begin{pmatrix} 0 \\ -b_1 \\ \vdots \\ 0 \\ -b_m \end{pmatrix} - \begin{pmatrix} A_1 & -\bar{a}_1 & \cdots & A_m & -\bar{a}_m \end{pmatrix}^T x \in -K, \end{array}$$

which has the dual conic form (19) (since $-K = K^\circ$). Moreover, $\mathcal{F}(D) \cap (\text{int}K^\circ \times \mathbb{R}^m) \neq \emptyset$ if and only if there exists an $\bar{x} \in \mathbb{R}^n$ satisfying all constraints in (22) strictly. By writing down the primal problem, conditions for $\mathcal{F}(P) \cap \text{int}K \neq \emptyset$ can be similarly derived.