CS 270 Combinatorial Algorithms and Data Structures, Spring 2015

Lecture 19: SDP duality

We extend the approach of proving LP duality to proving SDP duality, which is more complicated and not always true (but is true under some additional assumption).

Weak duality in semidefinite programming

We use the same idea to construct a dual SDP-program to provide an upper bound of the primal.

(primal)

$$\max \; C \bullet X$$
$$A_1 \bullet X \le b_1 \;\leftarrow y_1 \ge 0$$
$$\vdots$$
$$A_m \bullet X \le b_m \;\leftarrow y_m \ge 0$$
$$X \succeq 0$$

(dual)

$$\min \; b^T y$$
$$\sum_{i=1}^m y_i A_i \succeq C \quad \leftarrow \text{(dominate the objective)}$$
$$y_i \ge 0 \quad \leftarrow \text{(not to reverse the sign)}$$

To show that weak duality holds, we need to prove that if $A \succeq B$ (meaning that $A - B \succeq 0$) then $A \bullet X \ge B \bullet X$. Once we established it, then the weak duality follows because

$$b^T y = \sum_{i=1}^m y_i b_i \underset{\substack{\uparrow \\ \text{by primal constraints} \\ \text{and } y \ge 0}}{\ge} \sum_{i=1}^m y_i (A_i \bullet X) \underset{\substack{\uparrow \\ \text{linearity of} \\ \text{inner product}}}{=} \left(\sum_{i=1}^m y_i A_i\right) \bullet X \underset{\substack{\uparrow \\ \text{by domination and} \\ \text{the unproven claim}}}{\ge} C \bullet X$$

, and this is weak duality in SDP.

It remains to prove the following fact to complete the proof.

<u>Fact 1</u> $M \succeq 0$ if and only if $M \bullet X \ge 0$ for all $X \succeq 0$.

<u>proof</u> First, note that $vv^T \succeq 0$ for any v.

($\Leftarrow$) So, if $M \bullet X \ge 0$ for all $X \succeq 0$, then $M \bullet (vv^T) \ge 0$ for all v, but this is the same as saying $v^T M v \ge 0$ for all v (since $M \bullet (vv^T) = v^T M v$), and thus $M \succeq 0$.

($\Rightarrow$) Since $X \succeq 0$, we can write $X = UU^T = \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} - & u_1 & - \\ - & u_2 & - \\ & \vdots & \\ - & u_n & - \end{pmatrix} = \sum_{i=1}^n u_i u_i^T$.

Thus, $M \bullet X = M \bullet \left(\sum_{i=1}^n u_i u_i^T\right) = \sum_{i=1}^n \left(M \bullet u_i u_i^T\right) = \sum_{i=1}^m u_i^T M u_i \ge 0$ since $M \succeq 0$. $\square$

<u>Corollary</u> $\left(\sum y_i A_i\right) \bullet X \ge C \bullet X$

<u>proof</u> $\sum_i y_i A_i \succeq C \Rightarrow \left(\sum_i y_i A_i - C\right) \succeq 0 \Rightarrow \left(\sum_i y_i A_i - C\right) \bullet X \ge 0$ for $X \succeq 0$ $\Rightarrow \left(\sum_i y_i A_i\right) \bullet X \ge C \bullet X$ $\square$

One may expect that strong duality holds in SDP as well. Unfortunately, it is not always true. Fortunately, it is true most of the time. We will discuss this in detail.

To prove strong duality, it is more convenient to consider the equational form.

$$\text{(primal)}\quad \begin{aligned} &\max\ C\cdot X \\ &A_i\cdot X = b_i \quad \text{for } 1\le i\le m \\ &X \succeq 0 \end{aligned} \qquad\qquad \text{(dual)}\quad \begin{aligned} &\min\ b^T y \\ &\sum_{i=1}^m y_i A_i \succeq C \end{aligned}$$

Note that the constraint $y \geq 0$ is dropped in the dual.

<u>Exercise</u>: Show that this primal-dual pair in the equational form is equivalent to the primal-dual pair in inequality form above. That is, show that you can derive one form from another.

<u>SDP duality ?</u>

Let's try to apply the same idea for LP-duality to SDP-duality.

We will skip the two-step approach (first Farkas, then duality) and try to do it directly.

Consider the primal-dual pair:

$$\text{(primal)}\quad \begin{aligned} &\max\ C\cdot X \\ &A_i\cdot X = b_i \quad \text{for } 1\le i\le m \\ &X \succeq 0 \end{aligned} \qquad\qquad \text{(dual)}\quad \begin{aligned} &\min\ b^T y \\ &\sum_{i=1}^m y_i A_i \succeq C \end{aligned}$$

Again, we want to show that if the primal objective value is less than μ, then there is a dual feasible solution with objective value less than μ.

The primal objective value less than μ is equivalent to saying that the following system is infeasible:

$$\begin{aligned} &C\cdot X - s = \mu \\ &A_i\cdot X = b_i \quad \text{for } 1\le i\le m \\ &X\succeq 0 \\ &s\geq 0,\ s\in\mathbb{R} \end{aligned} \quad\Longleftrightarrow\quad \begin{aligned} &\begin{pmatrix} C & 0 \\ 0 & -1 \end{pmatrix}\cdot\begin{pmatrix} X & 0 \\ 0 & s \end{pmatrix} = \mu \\ &\begin{pmatrix} A_i & 0 \\ 0 & 0 \end{pmatrix}\cdot\begin{pmatrix} X & 0 \\ 0 & s \end{pmatrix} = b_i \quad \text{for } 1\le i\le m \\ &\begin{pmatrix} X & 0 \\ 0 & s \end{pmatrix}\succeq 0 \end{aligned}$$

Let $C' = \begin{pmatrix} C & 0 \\ 0 & -1 \end{pmatrix}$, $A_i' = \begin{pmatrix} A_i & 0 \\ 0 & 0 \end{pmatrix}$.

Consider the set $S := \left\{ (C'\cdot X', A_1'\cdot X', \ldots, A_m'\cdot X') \;\middle|\; X' = \begin{pmatrix} X & 0 \\ 0 & s \end{pmatrix},\ X'\succeq 0 \right\}$,

i.e. the set of $(m+1)$-dimensional vectors generated by PSD matrices of that form.

It is easy to check that S is convex.

The above system is infeasible is equivalent to saying that $(\mu, b_1, b_2, \ldots, b_m) \notin S$.

By the separation theorem, there exists y such that $\langle y, (\mu, b_1, \ldots, b_m)\rangle < 0$ but $\langle y, z\rangle \geq 0\ \forall z\in S$.

Let's write down what it means. Let $y = (y_0, y_1, \ldots, y_m)$.

Let's write down what it means. Let $y = (y_0, y_1, \dots, y_m)$.

The condition $\langle y, z \rangle \geq 0$ means $\langle (y_0, y_1, \dots, y_m), (C' \cdot X', A_1' \cdot X', \dots, A_m' \cdot X') \rangle \geq 0 \quad \forall X' = \begin{pmatrix} X & 0 \\ 0 & s \end{pmatrix}, X' \succcurlyeq 0$

$$\iff \quad y_0 C' \cdot X' + \sum_{i=1}^m y_i A_i' \cdot X' \geq 0 \qquad \forall X' = \begin{pmatrix} X & 0 \\ 0 & s \end{pmatrix}, X' \succcurlyeq 0$$

$$\iff \quad \left(y_0 C' + \sum_{i=1}^m y_i A_i' \right) \cdot X' \geq 0 \qquad \forall X' = \begin{pmatrix} X & 0 \\ 0 & s \end{pmatrix}, X' \succcurlyeq 0$$

$$\iff \quad y_0 C + \sum_{i=1}^m y_i A_i \succcurlyeq 0 \quad \text{and} \quad -y_0 \geq 0$$

The condition $\langle y, (\mu, b_1, \dots, b_m) \rangle < 0$ means $y_0 \mu + b^T y < 0$.

To summarize, we have $y_0 C + \sum_{i=1}^m y_i A_i \succcurlyeq 0$, $y_0 \leq 0$, $y_0 \mu + b^T y < 0$.

Consider two cases. Suppose $y_0 = 0$. Then, we have $\sum_{i=1}^m y_i A_i \succcurlyeq 0$ and $b^T y < 0$, but this implies that there is no solution X for $A_i \cdot X = b_i$ for $1 \leq i \leq m$ and $X \succcurlyeq 0$, and thus the primal program has no solution at all.

So assume $y_0 < 0$. Then we have $\sum_{i=1}^m \left(\frac{y_i}{-y_0} \right) A_i \succcurlyeq C$ and $b^T \left(\frac{y}{-y_0} \right) < \mu$.

Thus, $\left(-\frac{y_1}{y_0}, -\frac{y_2}{y_0}, \dots, -\frac{y_m}{y_0} \right)$ is a dual feasible solution with objective value less than μ.

It seems that we have proven the strong duality theorem for SDP. Is it true?

A counterexample

Consider the following primal-dual pair:

(primal)

$$\max \ -Y_{33}$$
$$Y_{12} + Y_{21} + Y_{33} = 1$$
$$Y_{22} = 0$$
$$Y \succcurlyeq 0$$

(dual)

$$\min \ X_1$$
$$\begin{pmatrix} 0 & X_1 & 0 \\ X_1 & X_2 & 0 \\ 0 & 0 & X_1 + 1 \end{pmatrix} \succcurlyeq 0$$

Recall that if $A = \begin{pmatrix} 0 & q^T \\ q & N \end{pmatrix} \succcurlyeq 0$, then we must have $q = 0$ (see the section on Cholesky factorization).

Therefore, the primal solutions must be of the form $\begin{pmatrix} a & 0 & b \\ 0 & 0 & 0 \\ b & 0 & 1 \end{pmatrix}$ because $Y_{22} = 0$, and dual solutions must have $X_1 = 0$.

Hence, the primal optimal value is -1 but the dual optimal value is 0.

So, there is a gap. In fact, the gap could be unbounded by changing the RHS of the first equality.

What's wrong?

What's wrong?

Closed convex sets

The error in the proof is that the set $S = \left\{ (C'\cdot X', A_1'\cdot X', \dots, A_m'\cdot X') \mid X' = \begin{pmatrix} X & 0 \\ 0 & s \end{pmatrix}, X' \succeq 0 \right\}$ is not necessarily closed.

And the separation theorem is not necessarily true for a set that is not closed.

Limit Value

To fix the proof, we consider the closure $\bar{S}$ of the set S. That is, $\bar{S}$ consists of the points in S and also the limit points of S. More formally, $y \in \bar{S}$ if and only if there exists a sequence $(y_k)_{k\in\mathbb{N}}$ such that $y_k \in S$ for all k and $\lim_{k\to\infty} y_k = y$.

In other words, we slightly enlarge the set S by including all the points on the boundary of S.

From the definition of S in our setting, the boundary points correspond to the limit points defined below.

Definition A sequence $(X_k)_{k\in\mathbb{N}}$ is called limit-feasible if $X_k \succeq 0$ for all k and

$$\lim_{k\to\infty} (A_1\cdot X_k, A_2\cdot X_k, \dots, A_m\cdot X_k) = (b_1, b_2, \dots, b_m).$$

Definition The value of a sequence $(X_k)_{k\in\mathbb{N}}$ is defined as $\langle c, (X_k)_{k\in\mathbb{N}} \rangle := \limsup C\cdot X_k$

The limit value of the primal SDP is defined as $\sup\{ \langle c, (X_k)^{k\to\infty}_{k\in\mathbb{N}} \rangle \mid (X_k)_{k\in\mathbb{N}} \text{ limit-feasible} \}$.

It is counterintuitive that the value of an SDP could be very different from its limit value, but it could happen. Let us look more closely at the counterexample again.

It is more convenient to consider the dual of the counterexample (which could be rewritten as the primal form):

$$\min \; x_1$$

$$\begin{pmatrix} 0 & x_1 & 0 \\ x_1 & x_2 & 0 \\ 0 & 0 & x_1+1 \end{pmatrix} \succeq 0$$

Let $Z_k = \begin{pmatrix} \frac{1}{k} & -1 & 0 \\ -1 & k & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Then $Z_k \succeq 0$ for all k.

Clearly, the value of this sequence $(Z_k)_{k\in\mathbb{N}}$ is -1, while the actual value is 0 as x_1 must be 0.

This shows the subtle difference between feasibility and arbitrarily close to feasibility.

Regular duality

After we consider the closure $\bar{S}$ of S, the proof would work, and it translates to:

Theorem The primal limit value is equal to the dual optimal value.

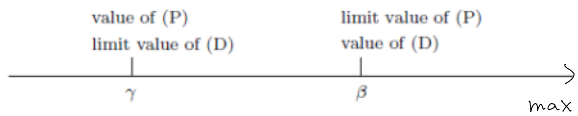
Proof When we consider $\bar{S}$, then the argument above showed that if the primal limit value is less than μ, then there is a dual feasible solution with objective value is less than μ.

This shows that dual value is at most primal limit value.

On the other hand, check that weak duality still applies, because $X_k \succeq 0$ for all k.

This implies that primal limit is at most the dual value, and hence the equality. $\square$

The following is a picture from Gartner-Matousek describing the situation



Strong duality

After we proved regular duality, the question of whether strong duality holds becomes the question of whether the primal limit value can be achieved by a primal feasible solution.

Of course, if we could show that S is closed, then strong duality holds, but this is usually difficult to check.

One well-known sufficient condition that is often satisfied is the following condition by Slater.

A real symmetric matrix M is called positive definite, written as $M \succ 0$, if all of its eigenvalues are positive.

Slater's condition: If the primal SDP has a feasible solution $\tilde{X}$ such that $\tilde{X} \succ 0$, then the primal optimal value is equal to the primal limit value.

proof (optional): The idea is quite natural: use the interior point $\tilde{X}$ to "correct" the limit-feasible sequence $(X_k)_{k\in\mathbb{N}}$ to a (really) feasible sequence $(W_k)_{k\in\mathbb{N}}$, while the value of the sequence $(W_k$ is close to the value of $(X_k)_{k\in\mathbb{N}}$.

Let μ be the primal limit value, attained by some sequence $(X_k)_{k\in\mathbb{N}}$.

For any $\varepsilon>0$, we want to construct a feasible solution Y with $C\cdot Y \geq \mu-\varepsilon$.

To do this, we will construct a sequence $(W_k)_{k\in\mathbb{N}}$ such that $W_k \succeq 0$ for large enough k, W_k is feasible and $\|X_k - W_k\|$ tiny for large enough k.

Then, for large enough k, W_k is a feasible solution with $C\cdot W_k \geq C\cdot X_k - \varepsilon$.

For $\alpha>0$, let $W_k := (1-\alpha)(X_k + Y_k) + \alpha\tilde{X}$, where $\tilde{X}$ is the interior point, and Y_k is a "correcting" term to make sure that W_k is feasible and $\|Y_k\| \to 0$ when $k\to\infty$.

The correcting matrix Y_k should satisfy $(A_1\cdot Y_k, \ldots, A_m\cdot Y_k) = (b_1 - A_1\cdot X_k, \ldots, b_m - A_m\cdot X_k)$.

Note that we don't require $Y_k \succeq 0$ at all.

Since $\tilde{X}$ is also feasible, it follows that W_K is feasible for all k by our construction of Y_K.

First, we assume the sequence $(Y_k)_{k\in\mathbb{N}}$ satisfies $\|Y_k\| \to 0$ and finish the proof, and then we will show how to construct such a sequence.

If $\|Y_k\| \to 0$, then by choosing α small enough, we have $\|X_K - W_K\|$ tiny for large enough k, and thus $C \cdot W_K \geq C \cdot X_K - \varepsilon$.

We also need to check that $W_K \succcurlyeq 0$. Note that $\tilde{X} \succ 0$, and since $X_K \succcurlyeq 0$, we have $(1-\alpha) X_K + \alpha \tilde{X} \succcurlyeq \rho I$ for some $\rho > 0$. Since $\|Y_k\| \to 0$, for large enough k, $\|Y_k\| < \rho$ and thus $W_K \succcurlyeq 0$.

Finally, we show how to construct Y_k with $\|Y_k\| \to 0$.

Without loss of generality, we assume that the constraints $A_i \cdot X$ are linearly independent.

Then, let $e_i \in \mathbb{R}^m$ be the i-th in the standard basis, we can find M_i so that $e_i^T = (A_1 \cdot M_i, \ldots, A_m \cdot M_i)$, because A_j are linearly independent.

Given $M_1, \ldots, M_m$, we can set $Y_k = \sum_{i=1}^{m} (b_i - A_i \cdot X_k) M_i$, and by construction we have $(A_1 \cdot Y_k, \ldots, A_m \cdot Y_k) = (b_1 - A_1 \cdot X_k, \ldots, b_m - A_m \cdot X_k)$, and thus Y_k is a correcting matrix.

Since $(X_k)_{k\in\mathbb{N}}$ is limit feasible, we have $b_i - A_i \cdot X_k \to 0$ for all i, and thus $\|Y_k\| \to 0$.

This is the construction of the correcting matrices with the desired property, completing the proof. □

To summarize, we have the following SDP duality theorem.

<u>Theorem</u> If the primal SDP has a feasible solution with $X \succ 0$, then strong duality holds.

<u>Concluding Remarks</u>

- Duality theorems are some of the most important theorems in optimization. Make sure you know them. They are very useful in giving bounds, e.g. in proving lower bounds in communication complexity.
- The material is extracted from chapter 4 of Gartner-Matousek, they formulate in a more general setting called cone programming. We specialize their content to SDP so that it is more concrete and easier to understand, and also the presentation is quite different. It would be very good to also read chapter 4 in detail to acquire a more general point of view, and pick up some concept that we didn't explicitly stated, e.g. dual cones.
- It is possible to prove the strong duality theorem (with Slater's condition) directly without going through the concept of limit value (Lovász survey), and the proof is shorter (like the length of our "wrong" proof).

However, then it is not very clear why the Sleater condition is needed, and not clear why there is a duality gap. The current presentation is longer, and the part on limit value is a bit technical, but I think it is conceptually clearer and provides a better understanding when there is a duality gap.