Lecture 12

Semidefinite Duality*

In the past couple of lectures we have discussed semidefinite programs and some of their applications in solving computer science problems. In this lecture we introduce the concept of semidefinite duality and look at the relationship between a semidefinite program and its dual.

12.1 Semidefinite Matrices

Recall from Lecture 10 that a symmetric matrix A of size $n \times n$ is positive semidefinite (psd) if it meets any one of the following two criteria (or any of a host of others): the third criterion below is yet another useful way to characterize psd matrices.

1. $x^\top A x \geq 0$ for all $x \in \mathbb{R}^n$.

2. All eigenvalues of A are non-negative.

3. $A = PDP^\top$ where D is the diagonal matrix $\text{diag}(\lambda_1, \lambda_2, \ldots \lambda_n)$ where the λ_i are the eigenvalues of A and P's columns are the eigenvectors of A. In D we note that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.

   We define the function $\text{diag}(x_1, x_2, \ldots x_n)$ below.

$$\text{diag}(x_1, x_2, \ldots x_n) = \begin{pmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_n \end{pmatrix}$$

   We note from this property that $A^{1/2} = PD^{1/2}P^\top$ where $D^{1/2} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots)$. From this, due to the orthonormality of P, it is clear that $A = A^{1/2}A^{1/2}$.

We would like to include additional notation to simplify future computations. Recall that $A \succeq 0$ denotes that A is positive semidefinite; let us define $A \preceq B$ to mean $B - A \succeq 0$.

*Lecturer: Anupam Gupta. Scribe: Alex Beutel.

Definition 12.1. Given symmetric matrices A, B we define $A \bullet B = Tr(A^\top B) = \sum_{ij} A_{ij} B_{ij}$.

We can think of A and B as vector of length n^2, then $A \bullet B$ is just the usual inner product between vectors. Note that if $x \in \mathbb{R}^n$, then $(xx^\top)$ is an $n \times n$ matrix, where $(xx^\top)_{ij} = x_i x_j$.

Fact 12.2. $x^\top A x = \sum_{ij} x_i x_j A_{ij} = \sum_{ij} (xx^\top)_{ij} A_{ij} = (xx^\top) \bullet A$.

Based on these underlying principles of linear algebra and psd matrices, we can begin to derive some interesting facts that will be useful later in the lecture. The proofs were not all discussed in class, but are given here for completeness.

Fact 12.3. *For any two $n \times n$ matrices A, B, Tr(AB) = Tr(BA).*

Proof. $Tr(AB) = \sum_i (AB)_{ii} = \sum_i \sum_j A_{ij} B_{ji} = \sum_j \sum_i B_{ji} A_{ij} = Tr(BA)$. □

Lemma 12.4. *For a symmetric $n \times n$ matrix A, A is psd if and only if $A \bullet B \geq 0$ for all psd B.*

Proof. One direction is easy: if A is not psd, then there exists $x \in \mathbb{R}^n$ for which $A \bullet (xx^\top) = x^\top A x < 0$. But $xx^\top$ is psd, which shows that $A \bullet B < 0$ for some psd B.

In the other direction, let A, B be psd. We claim that C, defined by $C_{ij} = A_{ij} B_{ij}$, is also psd. (This matrix C is called the *Frobenius inner product* ~~Schur-Hadamard product~~ of A and B.) Then

$$\sum_{ij} A_{ij} B_{ij} = \sum_{ij} C_{ij} = \mathbf{1}^\top C \mathbf{1} \geq 0$$

by the definition of psd-ness of C. To see the claim: since $A \succeq 0$, there exist random variables $\{a_i\}_{i=1}^n$ such that $A_{ij} = E[a_i a_j]$. Similarly, let $B_{ij} = E[b_i b_j]$ for r.v.s $\{b_i\}$. Moreover, we can take the a's to be independent of the b's. So if we define the random variables $c_i = a_i b_i$, then

$$C_{ij} = E[a_i a_j] E[b_i b_j] = E[a_i a_j b_i b_j] = E[c_i c_j],$$

and we are done. (Note we used independence of the a's and b's to make the product of expectations the expectation of the product.) □

This clean random variable-based proof is from this blog post. One can also show the following claim. The linear-algebraic proof also gives an alternate proof of the above Lemma 12.4.

Lemma 12.5. *For $A \succ 0$ (i.e., it is positive definite), $A \bullet B > 0$ for all psd B, $B \neq 0$.*

Proof. Let's write A as $PDP^\top$ where P is orthonormal, and D is the diagonal matrix containing A's eigenvalues (which are all positive, because $A \succ 0$.

Let $\widehat{B} = P^\top B P$, and hence $B = P \widehat{B} P^\top$. Note that $\widehat{B}$ is psd: indeed, $x^\top \widehat{B} x = (Px)^\top B (Px) \geq 0$. So all of $\widehat{B}$'s diagonal entries are non-negative. Moreover, since $B \neq 0$, not all of $\widehat{B}$'s diagonal entries can be zero (else, by $\widehat{B}$'s psd-ness, it would be zero). Finally,

$$Tr(AB) = Tr((PDP^\top)(P\widehat{B}P^\top)) = Tr(PD\widehat{B}P^\top) = Tr(D\widehat{B}P^\top P) = Tr(D\widehat{B}) = \sum_i D_{ii} \widehat{B}_{ii}.$$

Since $D_{ii} > 0$ and $\widehat{B}_{ii} \geq 0$ for all i, and $\widehat{B}_{ii} > 0$ for some i, this sum is strictly positive. □

Lemma 12.6. *For psd matrices A, B, $A \bullet B = 0$ if and only if $AB = 0$.*

Proof. Clearly if $AB = 0$ then $A \bullet B = tr(A^\top B) = tr(AB) = 0$. For the other direction, we use the ideas (and notation) from Lemma 12.4. Again take the Frobenius product C defined by $C_{ij} = A_{ij}B_{ij}$. Then C is also psd, and hence $C_{ij} = E[c_i c_j]$ for random variables $\{c_i\}_{i=1}^n$. Then

$$A \bullet B = \sum_{ij} C_{ij} = \sum_{ij} E[c_i c_j] = E[\sum_{ij} c_i c_j] = E[(\sum_i c_i)^2].$$

If this quantity is zero, then the random variable $\sum_i c_i = \sum_i a_i b_i$ must be zero with probability 1. Now

$$(AB)_{ij} = \sum_k E[a_i a_k]E[b_k b_j] = \sum_k E[a_i b_j (a_k b_k)] = E[a_i b_j (\sum_k c_k)] = 0,$$

so AB is identically zero. □

12.2 Semidefinite Programs and their Duals

Given this understanding of psd matrices, we can now look at semidefinite programs (SDPs), and define their duals. Let us describe two common forms of writing SDPs. Consider symmetric matrices $A_1, A_2, \dots A_m, C$, and reals $b_1, b_2, \dots b_m$. The first form is the following one.

$$\begin{aligned} \min\, & C \bullet X \\ \text{s.t.} \quad & A_i \cdot X = b_i \quad i = 1 \dots m \\ & X \succeq 0 \end{aligned} \tag{12.1}$$

Another common form for writing SDPs is the following.

$$\begin{aligned} \max & \sum_{i=1}^m b_i y_i = b^\top y \\ \text{s.t.} \quad & \sum_{i=1}^m A_i y_i \preceq C \end{aligned} \tag{12.2}$$

This of course means that $C - \sum A_i y_i \succeq 0$. If we set $S = C - \sum A_i y_i$ and thus $S \succeq 0$ then it is clear that this constraint can be rewritten as $\sum y_i A_i + S = C$ for $S \succeq 0$.

$$\begin{aligned} & \max b^\top y \\ \text{s.t.} \quad & \sum_{i=1}^m A_i y_i + S = C \\ & S \succeq 0 \end{aligned} \tag{12.3}$$

Given an SDP in the form (12.1), we can convert it into an SDP in the form (12.3), and vice versa—this requires about a page of basic linear algebra.

12.2.1 Examples of SDPs

The Max-Cut Problem

An example, which we've already seen, is the semidefinite program for the maxcut problem. Given a graph $G = (V, E)$, with edge weights w_{ij},

$$\frac{1}{2}\max \sum_{(i,j)\in E} w_{ij}(1 - \langle v_i, v_j\rangle)$$
$$\text{s.t.} \quad \langle v_i, v_i\rangle = 1 \quad \forall i \in V$$

This is equivalent to

$$\frac{1}{2}\max \sum_{(i,j)\in E} w_{ij}(1 - X_{ij})$$
$$\text{s.t.} \quad X_{ii} = 1 \quad \forall i \in V$$
$$X \succeq 0$$

where we used the fact that $X \succeq 0$ iff there are vectors v_i such that $X_{ij} = \langle v_i, v_j\rangle$. For i, j such that $\{i, j\} \notin E$, define $w_{ij} = 0$, and for $\{i, j\} \in E$ define $w_{ji} = w_{ij}$; hence the objective function can now be written as

$$\frac{1}{4}\max \sum_{i,j\in V} w_{ij}(1 - X_{ij}).$$

(The extra factor of 1/2 is because we count each edge $\{u, v\}$ twice now.) We can write this even more compactly, once we introduce the idea of the Laplacian matrix of the weighted graph.

$$L_{ij} = L(w)_{ij} = \begin{cases} \sum_k w_{ik} & \text{if } i = j, \\ -w_{ij} & \text{if } i \neq j. \end{cases}$$

Again, the objective function of the above SDP (ignoring the factor of $\frac{1}{4}$ for now) is

$$\begin{aligned}
\sum_{i,j} w_{ij}(1 - X_{ij}) &= \sum_{i,j} w_{ij} - \sum_{i\neq j} w_{ij}X_{ij} \\
&= \sum_i \left(\sum_j w_{ij}\right) + \sum_{i\neq j} L_{ij}X_{ij} \\
&= \sum_i L_{ii}\, X_{ii} + \sum_{i\neq j} L_{ij}\, X_{ij} \\
&= L \bullet X
\end{aligned}$$

Finally rewriting $X_{ii} = X \bullet (e_i e_i^\top)$, the SDP is

$$\begin{aligned}
\max \frac{1}{4} L \bullet X \\
X \bullet (e_i e_i^T) = 1 \quad \forall i \\
X \succeq 0
\end{aligned} \tag{12.4}$$

Note that this SDP is in the form (12.1).

Maximum Eigenvalue of Symmetric Matrices

Another simple example is using an SDP to find the maximum eigenvalue for a symmetric matrix A. Suppose A has eigenvalues $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_n$. Then the matrix $tI - A$ has eigenvalues $t-\lambda_1, t-\lambda_2, \ldots, t-\lambda_n$. Note that $tI - A$ is psd exactly when all these eigenvalues are non-negative, and this happens for values $t \geq \lambda_1$. This immediately gives us that

$$\lambda_1 = \min\{t \mid \text{s.t. } tI - A \succeq 0\} \tag{12.5}$$

We will use $\lambda_1(A) = \lambda_{\max}(A)$ to denote the maximum eigenvalue of the matrix A. Note that this SDP uses the form (12.3) given above.

> **Note:** John pointed out that one could also write the maximum eigenvalue computation as the following SDP:
>
> $$\begin{aligned} &\max A \bullet X \\ \text{s.t.}\quad & X \bullet I = 1 \\ & X \succeq 0 \end{aligned} \tag{12.6}$$
>
> Indeed, we will soon show that (12.6) is precisely the SDP dual of (12.5).

12.2.2 Weak Duality

Given the two SDP forms above, namely (12.1) and (12.3), let's first note that one can move purely syntactically from one to the other. Next, one can show that these form a primal-dual pair. Let us consider the minimization problem in (12.1) to be the primal, and the maximization problem (12.3) to be the dual form.

Theorem 12.7 (Weak Duality)**.** *If X is feasible for the primal SDP and (y, S) are feasible for the dual SDP, then $C \bullet X \geq b^\top y$.*

Proof. Suppose (y, S) and X are feasible, then:

$$C \bullet X = (\sum y_i A_i + S) \bullet X \tag{12.7}$$

$$= \sum y_i (A_i \bullet X) + (S \bullet X) \tag{12.8}$$

$$= \sum y_i b_i + (S \bullet X) \tag{12.9}$$

Since S and X are psd, Lemma 12.4 implies $S \bullet X \geq 0$. Therefore, $C \bullet X \geq b^\top y$. □

Note that the transformation between the primal SDP form and the dual form was syntactic, much like in the case of LPs. And we have weak duality, like LPs. However, in Section 12.3 we will see that strong duality does not always hold (there may be a gap between the primal and dual values), but will also give some natural conditions under which strong SDP duality does hold.

12.2.3 General Cone Programs

Before we move on, let us actually place semidefinite duality (and LP duality) in a slightly broader context, that of duality in general cone programming. Suppose we consider a convex cone $K \in \mathbb{R}^n$ (i.e., it is convex, and for $x \in K$ and $\alpha \geq 0$, $\alpha x \in K$). We can now define the *dual cone* $K^* = \{y \mid x^\top y \geq 0 \ \forall x \in K\}$. E.g., here is an example of a cone in $\mathbb{R}^2$, and its dual cone.

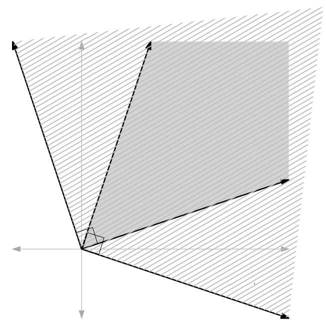

Figure 12.1: The cone K (in dark grey) and its dual cone K^* (shaded).

Moreover, here are some examples of K and the corresponding K^*.

$$
\begin{aligned}
K &= \mathbb{R}^n & K^* &= \{0\} \\
K &= \mathbb{R}^n_{\geq 0} & K^* &= R^n_{\geq 0} \\
K &= PSD_n & K^* &= PSD_n
\end{aligned}
$$

Let us write two optimization problems over cones, the primal and the dual. Given vectors $a_1, a_2, \ldots, a_m, c \in \mathbb{R}^n$ and scalars $b_1, b_2, \ldots, b_m \in \mathbb{R}$, the primal cone program (P') is

$$
\begin{aligned}
&\min c^\top x \\
\text{s.t.} \quad & a_i^\top x = b_i \quad i = 1 \ldots m \\
& x \in K
\end{aligned}
$$

The dual cone program (D') is written below:

$$
\begin{aligned}
&\max b^\top y \\
&\sum_{i=1}^m y_i a_i + s = c \\
&s \in K^*, y \in \mathbb{R}^m
\end{aligned}
$$

Claim 12.8 (Weak Duality for Cone Programs)**.** *If x is feasible for the primal (P') and (y, s) feasible for the dual (D'), then $c^\top x \geq b^\top y$.*

Proof. $c^\top x = (\sum y_i a_i + s)^\top x = \sum y_i a_i^\top + s^\top x \geq \sum y_i b_i + 0 = b^\top y$, where we use the fact that if $x \in K$ and $s \in K^*$ then $s^\top x \geq 0$. $\square$

Now instantiating K with the suitable cones we can get LPs and SDPs: e.g., considering $K = \mathbb{R}^n_{\geq 0} = K^*$ gives us

$$\begin{aligned} &\min c^\top x \\ \text{s.t.} \quad & a_i^\top x = b_i \quad i = 1 \dots m \\ & x \geq 0 \end{aligned} \qquad \begin{aligned} &\max b^\top y \\ &\textstyle\sum_{i=1}^m y_i a_i + s = c \\ & s \geq 0, y \in \mathbb{R}^m \end{aligned}$$

which is equivalent to the standard primal-dual pair of LPs

$$\begin{aligned} &\min c^\top x \\ \text{s.t.} \quad & a_i^\top x = b_i \quad i = 1 \dots m \\ & x \geq 0 \end{aligned} \qquad \begin{aligned} &\max b^\top y \\ &\textstyle\sum_{i=1}^m y_i a_i \leq c \\ & y \in \mathbb{R}^m \end{aligned}$$

And setting $K = K^* = PSD_n$ gives us the primal-dual pair of SDPs (12.1) and (12.3).

12.2.4 Examples: The Maximum Eigenvalue Problem

For the maximum eigenvalue problem, we wrote the SDP (12.5). Since it of the "dual" form, we can convert it into the "primal" form in a purely mechanical fashion to get

$$\begin{aligned} &\max A \bullet X \\ \text{s.t.} \quad & X \bullet I = 1 \\ & X \succeq 0 \end{aligned}$$

We did not cover this in lecture, but the dual can be reinterpreted further. recall that $X \succeq 0$ means we can find reals $p_i \geq 0$ and unit vectors $x_i \in \mathbb{R}^n$ such that $X = \sum_i p_i(x_i x_i^\top)$. By the fact that x_i's are unit vectors, $Tr(x_i x_i^\top) = 1$, and the trace of this matrix is then $\sum_i p_i$. But by our constraints, $X \bullet I = Tr(X) = 1$, so $\sum_i p_i = 1$.

Rewriting in this language, $\lambda_{\max}$ is the maximum of

$$\sum_i p_i \left(A \bullet x_i x_i^\top\right)$$

such that the x_i's are unit vectors, and $\sum_i p_i = 1$. But for any such solution, just choose the vector x_{i^*} among these that maximizes $A \bullet (x_{i^*} x_{i^*}^\top)$; that is at least as good as the average, right? Hence,

$$\lambda_{\max} = \max_{x \in \mathbb{R}^n : \|x\|_2 = 1} A \bullet (xx^\top) = \max_{x \in \mathbb{R}^n} \frac{x^\top A x}{x^\top x}$$

which is the standard variational definition of the maximum eigenvalue of A.

12.2.5 Examples: The Maxcut SDP Dual

Now let's revisit the maxcut SDP. We had last formulated the SDP as being of the form

$$\begin{aligned} &\max \frac{1}{4} L \bullet X \\ & X \bullet (e_i e_i^T) = 1 \quad \forall i \\ & X \succeq 0 \end{aligned}$$

It is in "primal" form, so we can mechanically convert it into the "dual" form:

$$\begin{aligned} \min \frac{1}{4} \sum_{i=1}^{n} y_i \\ \text{s.t.} \quad \sum_i y_i (e_i e_i^\top) \succeq L \end{aligned}$$

For a vector $v \in \mathbb{R}^n$, we define the matrix $\operatorname{diag}(v)$ to be the diagonal matrix D with $D_{ii} = v_i$. Hence we can rewrite the dual SDP as

$$\begin{aligned} \min \tfrac{1}{4} \mathbf{1}^\top y \\ \text{s.t.} \quad \operatorname{diag}(y) - L \succeq 0 \end{aligned}$$

Let us write $y = t\mathbf{1} - u$ for some real $t \in \mathbb{R}$ and vector $u \in \mathbb{R}^n$ such that $\mathbf{1}^\top u = 0$: it must be the case that $\mathbf{1}^\top y = n \cdot t$. Moreover, $\operatorname{diag}(t\mathbf{1} - u) = tI - \operatorname{diag}(u)$, so the SDP is now

$$\begin{aligned} \min \tfrac{1}{4} n \cdot t \\ \text{s.t.} \quad tI - (L + \operatorname{diag}(u)) \succeq 0 \\ \mathbf{1}^\top u = 0. \end{aligned}$$

Hey, this looks like the maximum eigenvalue SDP from (12.5): indeed, we can finally write the SDP as

$$\frac{n}{4} \cdot \min_{u : \mathbf{1}^\top u = 0} \lambda_{\max}(L + \operatorname{diag}(u)) \tag{12.10}$$

What is this saying? We're taking the Laplacian of the graph, adding in some "correction" values u to the diagonal (which add up to zero) so as the make the maximum eigenvalue as small as possible. The optimal value of the dual SDP is this eigenvalue scaled up by $n/4$. (And since we will soon show that strong duality holds for this SDP, this is also the value of the max-cut SDP.) This is precisely the bound on max-cut that was studied by Delorme and Poljak [DP93].

In fact, by weak duality alone, any setting of the vector u would give us an upper bound on the max-cut SDP (and hence on the max-cut). For example, one setting of these correction values would be to take $u = 0$, we get that

$$maxcut(G) \le SDP(G) \le \frac{n}{4} \lambda_{\max}(L(G)), \tag{12.11}$$

where $L(G)$ is the Laplacian matrix of G. This bound was given even earlier by Mohar and Poljak [MP90].

Some Examples

Sometimes just the upper bound (12.11) is pretty good: e.g., for the case of cycles C_n, one can show that the zero vector is an optimal correction vector $u \in \mathbb{R}^n$, and hence the max-cut SDP value equals the $n/4\lambda_{\max}(L(C_n))$.

To see this, consider the function $f(u) = \frac{n}{4}\lambda_{\max}(L + \mathrm{diag}(u))$. This function is convex (see, e.g. [DP93]), and hence $f(\frac{1}{2}(u + u')) \leq \frac{1}{2}(f(u) + f(u'))$. Now if $f(u)$ is minimized for some non-zero vector u^* such that $\mathbf{1}^\top u = 0$. Then by the symmetry of the cycle, the vector $u^{(i)} = (u^*_i, u^*_{i+1}, \ldots, u^*_n, u^*_1, \ldots, u^*_{i-1})^\top$ is also a minimizer. But look: each coordinate of $\sum_{i=1}^n u^{(i)}$ is itself just $\sum_i u^*_i = 0$. On the other hand, $f(0) = f(\frac{1}{n}\sum_i u^{(i)}) \leq \frac{1}{n}\sum_i f(u^{(i)}) = f(u)$ by the convexity of $f()$. Hence the zero vector is a minimizer for $f()$.

Note: Among other graphs, Delorme and Poljak considered the gap between the integer max-cut and the SDP value for the cycle. The eigenvalues for $L(C_n)$ are $2(1 - \cos(2\pi t/n)$ for $t = 0, 1, \ldots, n-1$. For $n = 2k$, the maximum eigenvalue is 4, and hence the max-cut dual (and primal) value is $\frac{n}{4} \cdot 4 = n$, which is precisely the max-cut value. The SDP gives us the right answer in this case.

What about odd cycles? E.g., for $n = 3$, say, the maximum eigenvalue is $3/2$, which means the dual (=primal) equals $9/8$. For $n = 5$, $\lambda_{\max}$ is $\frac{1}{2}(5 + \sqrt{5})$, and hence the SDP value is 4.52. This is ≈ 1.1306 times the actual integer optimum. (Note that the best current gap is $1/0.87856 \approx 1.1382$, so this is pretty close to the best possible.)

On the other hand, for the star graph, the presence of the correction vector makes a big difference. We'll see more of this in the next HW.

12.3 Strong Duality

Unfortunately, for SDPs, strong duality does not always hold. Consider the following example (from Lovász):

$$\begin{aligned} &\min y_1 \\ &\text{s.t.} \quad \begin{pmatrix} 0 & y_1 & 0 \\ y_1 & y_2 & 0 \\ 0 & 0 & y_1 + 1 \end{pmatrix} \succeq 0 \end{aligned} \tag{12.12}$$

Since the top-left entry is zero, SDP-ness forces the first row and column to be zero, which means $y_1 = 0$ in any feaasible solution. The feasible solutions are $(y_1 = 0, y_2 \geq 0)$. So the primal optimum is 0. Now to take the dual, we write it in the form (12.2):

$$y_1 \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + y_2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \succeq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

to get the dual:

$$\begin{aligned} \max -X_{33} \\ \text{s.t.} \quad X_{12} + X_{21} + X_{33} = 1 \\ X_{22} = 0 \\ X \succeq 0 \end{aligned}$$

Since $X_{22} = 0$ and $X \succeq 0$, we get $X_{12} = X_{21} = 0$. This forces $X_{33} = 1$, and the optimal value for this dual SDP is -1. Even in this basic example, strong duality does not hold. So the strong duality theorem we present below will have to make some assumptions about the structure of the primal and dual, which is often called *regularization* or *constraint qualification*.

Before we move on, observe that the example is a fragile one: if one sets the top left entry of (12.12) to $\epsilon > 0$, suddenly the optimal primal value drops to -1. (Why?)

One more observation: consider the SDP below.

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

By PSD-ness we want $x_1 \geq 0$, $x_2 \geq 0$, and $x_1, x_2 \geq 1$. Hence for any $\epsilon > 0$, we can set $x_1 = \epsilon$ and $x_2 = 1/x_1$—the optimal value tends to zero, but this optimum value is never achieved. So in general we define the optimal value of SDPs using infimums and supremums (instead of just mins and maxs). Furthermore, it becomes a relevant question of whether the SDP achieves its optimal value or not, when this value is bounded. (This was not an issue with LPs: whenever the LP was feasible and its optimal value was bounded, there was a feasible point that achieved this value.)

12.3.1 The Strong Duality Theorem for SDPs

We say that a SDP is *strictly feasible* if it satisfies its positive semidefiniteness requirement strictly: i.e. with positive <u>definiteness</u>.

Theorem 12.9. *Assume both primal and dual have feasible solutions. Then $v_{\text{primal}} \geq v_{\text{dual}}$, where v_{primal} and v_{dual} are the optimal values to the primal and dual respectively. Moreover, if the primal has a strictly feasible solution (a solution x such that $x \succ 0$ or x is positive definite) then*

1. *The dual optimum is attained (which is not always the case for SDPs)*
2. $v_{\text{primal}} = v_{\text{dual}}$

Similarly, if the dual is strictly feasible, then the primal optimal value is achieved, and equals the dual optimal value. Hence, if both the primal and dual have strictly feasible solutions, then both v_{primal} and v_{dual} are attained.

Note: For both the SDP examples we've considered (max-cut, and finding maximum eigenvalues), you should check that strictly feasible solutions exist for both primal and dual programs, and hence there is no duality gap.

Strict feasibility is also a sufficient condition for avoiding a duality gap in more general convex programs: this is called the *Slater condition.* For more details see, e.g., the book by Boyd and Vanderberghe.

12.3.2 The Missing Proofs*

We did not get into details of the proof in lecture, but they are presented below for completeness. (The presentation is based on, and closely follows, that of Laci Lovász's notes.) We need a SDP version of the Farkas Lemma. First we present a homogeneous version, and use that to prove the general version.

Lemma 12.10. *Let A_i be symmetric matrices. Then $\sum_i y_i A_i \succ 0$ has no solution if and only if there exists $X \succeq 0$, $X \neq 0$ such that $A_i \bullet X = 0$ for all i.*

One direction of the proof (if $\sum_i y_i A_i \succ 0$ is infeasible, then such an X exists) is an easy application of the hyperplane separation theorem, and appears in Lovasz's notes. The other direction is easy: if there is such a X and $\sum_i y_i A_i \succ 0$ is feasible, then by Lemma 12.5 and the strict positive definiteness $(\sum_i y_i A_i) \bullet X > 0$, but all $A_i \bullet X = 0$, which is a contradiction.

We could ask for a similar theorem of alternatives for $\sum_i y_i A_i \succeq 0$: since we're assuming more, the "only if" direction goes through just the same. But the "if" direction fails, since we cannot infer a contradiction just from Lemma 12.4. And this will be an important issue in the proof of the duality theorem. Anyways, the Farkas Lemma also extends to the non-homogeneous case:

Lemma 12.11. *Let A_i, C be symmetric matrices. Then $\sum_i y_i A_i \succ C$ has no solution if and only if there exists $X \succeq 0$, $X \neq 0$ such that $A_i \bullet X = 0$ for all i and $C \bullet X \geq 0$.*

Proof. Again, if such X exists then we cannot have a solution. For the other direction, the constraint $\sum_i y_i A_i - C \succ 0$ is equivalent to $\sum_i y_i A_i + t(-C) \succ 0$ for $t > 0$ (because then we can divide through by t). To add this side constraint on t, let us define

$$A'_i := \begin{pmatrix} A_i & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad C' := \begin{pmatrix} -C & 0 \\ 0 & 1 \end{pmatrix}.$$

Lemma 12.10 says if $\sum_i y_i A'_i + tC' \succ 0$ is infeasible then there exists psd $X' \neq 0$, with $X' \bullet A'_i = 0$ and $X' \bullet C' = 0$. If X is the top $n \times n$ part of X', we get $X \bullet A_i = 0$ and $(-C) \bullet X + x_{n+1,n+1} = 0$. Moreover, from $X' \succeq 0$, we get $X \succeq 0$ and $x_{n+1,n+1} \geq 0$—which gives us $C \bullet X \geq 0$. Finally, to check $X \neq 0$: in case $X' \neq 0$ but $X = 0$, we must have had $x_{n+1,n+1} > 0$, but then $C' \bullet X \neq 0$. □

Now, here's the strong duality theorem: here the primal is

$$\min\{b^\top y \mid \sum_i y_i A_i \succeq C\}$$

and the dual is

$$\max\{C \bullet X \mid A_i \bullet X = b_i \forall i, X \succeq 0\}$$

Theorem 12.12. *Assume both primal and dual are feasible. If the primal is strictly feasible, then (a) the dual optimum is achieved, and (b) the primal and dual optimal values are the same.*

Proof. Since $b^\top y < opt_p$ and $\sum_i y_i A_i \succeq C$ is not feasible, we can define

$$A_i' := \begin{pmatrix} -b_i & 0 \\ 0 & A_i \end{pmatrix} \qquad \textit{and} \qquad C' := \begin{pmatrix} -opt_p & 0 \\ 0 & C \end{pmatrix}$$

and use Lemma 12.11 to get psd $Y' \neq 0$ with $Y' \bullet A_i' = 0$ and $Y' \bullet C' \geq 0$. Say

$$Y' := \begin{pmatrix} y_0 & y \\ y & Y \end{pmatrix}$$

then $A_i \bullet Y = y_0 \, b_i$ and $C \bullet Y \geq y_0 \, opt_p$. By psd-ness, $y_0 \geq 0$. If $y_0 \neq 0$ then we can divide through by y_0 to get a feasible solution to the dual with value equal to opt_p.

What if $y_0 = 0$? This cannot happen. Indeed, then we get $A_i \bullet Y = 0$ and $C \bullet Y \geq 0$, which contradicts the *strict* feasibility of the primal. (Note that we are using the "if" direction of the Farkas Lemma here, whereas we used the "only if" direction in the previous step.) Again, we really need to assume strict feasibility, because there are examples otherwise. □

The notion of duality we've used here is Lagrangian duality. This is not the only notion possible, and in fact, there are papers that study other notions of duality that avoid this "duality gap" without constraint qualification. For example, see the paper of Ramana, Tüncel, and Wolkowicz (1997).

Bibliography

[DP93] C. Delorme and S. Poljak. Laplacian eigenvalues and the maximum cut problem. *Math. Programming*, 62(3, Ser. A):557–574, 1993. 12.2.5, 12.2.5

[MP90] Bojan Mohar and Svatopluk Poljak. Eigenvalues and max-cut problem. *Czechoslovak Math. J.*, 40(115)(2):343–352, 1990. 12.2.5