Semidefinite Programming
Lecture 2

scribe: Yuming Zhang

Feb. 6 , 1995

A lemma left over from last lecture

We want to show that

$$\lambda_1(A) = \max\{\mathbf{x}^T A\mathbf{x} : \ \|\mathbf{x}\| = 1\} = \max\{A \bullet U : \text{trace } U = 1 \ \text{ and } \ 0 \preceq U \preceq I\}$$

The first of these is the Rayleigh-Ritz theorem. To show the second we have already established that $\lambda_1(A) \leq \tilde{\lambda}(A)$, where $\tilde{\lambda}(A)$ is the the rightmost side in the equality above. To show the converse we use the following lemma:

LEMMA 0.1. *Let* $S_1 = \{\mathbf{y}\mathbf{y}^T : \mathbf{y} \in R^n, \|\mathbf{y}\|_2 = 1\}$ *and* $S_2 = \{Y : Y \succeq 0, \text{trace } (Y) = 1\}$, *then*

1. $S_2 = \text{conv } (S_1)$ *and*

2. $S_1 = \text{ext } (S_2)$, *where* $\text{ext } (S_2)$ *is the set of extreme points of* S_2.

Proof. By eigenvalue decomposition every symmetric matrix Y can be decomposed into:

$$Y = \lambda_1\mathbf{q}_1\mathbf{q}_1^T + ... + \lambda_n\mathbf{q}_n\mathbf{q}_n^T,$$

where λ_i are eigenvalues of Y $(i = 1, ..., n)$. By our assumption, we have $\lambda_1 + ... + \lambda_n = 1$, and $\lambda_i \geq 0, i = 1, ...n$. Thus, $S_2 \subseteq \text{conv } S_1$. In addition, any convex combination of the form ($\mathbf{y}_i$'s not necessarily orthogonal!)

$$Y = \lambda_1\mathbf{y}_1\mathbf{y}_1^T + \cdots + \lambda_n\mathbf{y}_n\mathbf{y}_n^T$$

for $\lambda_i \geq 0$ and $\sum \lambda_i = 1$ is positive semidefinite with trace equal to 1, and with $I - Y$ also positive semidefinite. Thus we've shown $S_2 = \text{conv}(S_1)$. To show 2, notice that no extreme point of S_2 can have rank larger than one, that is all extreme points of S_2 are in S_1. By reasons of symmetry, any matrix in S_1 is an extreme point of S_2, because notice that if for some $\mathbf{y}$, $\mathbf{y}\mathbf{y}^T$ is an extreme point and $\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T$ is another point, then, since one can rotate $\tilde{\mathbf{y}}$ to $\mathbf{y}$ by multiplying by a rotation matrix Q, the same rotation induces a rotation in S_2, which should map extreme points to extreme points, and thus $\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T$ must also be an extreme point. ∎

Using this lemma, we get

$$\lambda_1(A) = \max\{A \bullet Y : \text{ trace } Y = 1 \text{ and } 0 \preceq Y \preceq I\}.$$

Since the maximum is always attained at an extreme point which is a rank one matrix.

1 Strong Duality Theorem

1.1 Separation Theorem. The fundamental theorem that we need for duality theorem in the general case of cones is the following cone version of the **strong separation theorem**:

LEMMA 1.1. **(Strong Separation Theorem)**: *Let $\mathcal{K}$ be a closed convex pointed cone and $\mathbf{b} \notin \mathcal{K}$. Then there exists a hyperplane, specifically a linear space, $\mathbf{y}^T\mathbf{x} = 0$ such that $\mathbf{y}^T\mathbf{b} < 0$ and $\mathbf{y}^T\mathbf{x} \geq 0$, for all $\mathbf{x} \in \mathcal{K}$.*

This theorem and its proof can be found in most convex analysis textbooks, such as Rockafellar's. The key point is that the cone $\mathcal{K}$ must be a closed set and therefore contain its boundary. Thus, if $\mathbf{b} \notin \mathcal{K}$ then $\mathbf{b}$ will be completely outside of the cone and can be separated from it in a strong sense which is indicated by the strict inequality $\mathbf{y}^T\mathbf{b} < 0$. If $\mathcal{K}$ is not required to be closed then $\mathbf{b} \notin \mathcal{K}$ is still true if $\mathbf{b}$ is on some points of the boundary of $\mathcal{K}$ not included in $\mathcal{K}$. In that case it is impossible to separate it from the cone in the strong sense.

Next, we will prove the strong duality theorem with respect to a closed convex pointed cone. The primal problem is:

$$\begin{array}{lll} \mathbf{P}: & \min & \mathbf{c}^T\mathbf{x} \\ & \text{s.t.} & A\mathbf{x} = \mathbf{b} \\ & & \mathbf{x} \geq_{\mathcal{K}} 0 \end{array}$$

The dual problem is :

$$\begin{array}{lll} \mathbf{D}: & \max & \mathbf{b}^T\mathbf{y} \\ & \text{s.t.} & A^T\mathbf{y} + \mathbf{z} = \mathbf{c} \\ & & \mathbf{z} \geq_{\mathcal{K}^*} 0 \end{array}$$

1.2 Generalized Farkas Lemma First we derive an extension of Farkas lemma with respect to a closed convex pointed cone $\mathcal{K}$:

LEMMA 1.2. **(Generalized Farkas Lemma)** *Assume $\mathcal{P} = \{A\mathbf{x} : \mathbf{x} \geq_{\mathcal{K}} 0\}$ is closed, then*

1. Either: *There exists $\mathbf{x} \in R^n$ such that $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq_{\mathcal{K}} 0$,*
2. Or : *There exists $\mathbf{y} \in R^m$ such that $A^T\mathbf{y} \geq^*_{\mathcal{K}} 0$ and $\mathbf{b}^T\mathbf{y} < 0$.*

But not both.

Proof. Suppose (1) is true, then $\mathbf{b}^T\mathbf{y} = (A\mathbf{x})^T\mathbf{y} \geq 0$, which means (2) cannot hold. If (1) is false, then $\mathbf{b}$ is not in $\mathcal{P}$. By separation theorem, there is at least one $\mathbf{y}$ such that $\mathbf{y}^T\mathbf{b} < 0$ and $\mathbf{y}^T\mathbf{z} \geq 0$, for any $\mathbf{z} \in \mathcal{P}$. Since any such $\mathbf{z}$ can be written as $\mathbf{z} = A\mathbf{x}$ for some $\mathbf{x} \in \mathcal{K}$, we have for any $\mathbf{x} \in \mathcal{K}$, $\mathbf{y}^T(A\mathbf{x}) = (A^T\mathbf{y})^T\mathbf{x} \geq 0$ which implies $A^T\mathbf{y} \geq_{\mathcal{K}^*} 0$ by definition of polar. ■

Remark. The Generalized Farkas lemma does not necessarily hold if $\mathcal{P}$ is not closed. For instance, let $\mathcal{K}$ be the cone of 2×2 symmetric positive semidefinite matrices. Let also that E_{ij} be a matrix whose all entries are zero except the i, j entry which is one. Then the "either" part of Farkas lemma says

$$\begin{array}{ll} \exists X \in R^{2\times 2} & \\ \text{s.t.} & 1/2(E_{12} + E_{12}) \bullet X = 1 \\ & E_{11} \bullet X = 0 \\ & X \succeq 0 \end{array}$$

and The "or" part says

$$\begin{array}{ll} \exists \mathbf{y} \in R^{2} & \\ \text{s.t.} & 1/2\mathbf{y}_1(E_{12} + E_{21}) + \mathbf{y}E_{11} \succeq 0 \\ & \mathbf{y}_1 < 0 \end{array}$$

But the "either" part is equivalent to

$$\begin{pmatrix} 0 & 1 \\ 1 & x \end{pmatrix} \succeq 0$$

has a solution in x which is clearly impossible. And the the "or" part says that

$$\begin{pmatrix} u & v \\ v & 0 \end{pmatrix} \text{ and } v < 0$$

has a solution, which is also impossible.

Exercise 1. What is the analog of the set $\mathcal{P}$ in this example? Is it a closed set?

There are various sufficient conditions for making $\mathcal{P}$ being closed. One popular one which we will use is

LEMMA 1.3. *If* $\exists \mathbf{y}$ *such that* $A^T\mathbf{y} >_{\mathcal{K}^*} 0$, *then* P *is closed.*

This condition is sometimes referred to as the *Slater* constraint qualification.

Exercise 2. Prove the following variants of Farkas Lemma for the positive semidefinite cone. In each case a constraint qualification is missing. You should find this qualification analogous to the one given in the Generalized Farkas Lemma above.

1. **Either** there exists $x_1, \cdots, x_m$ such that $\sum x_i A_i \preceq B$, **Or** there exist $Y \succeq 0$ such that $A_i \bullet Y = 0$ and $B \bullet Y < 0$, but not both.

2. for $A \in \Re^{n\times n}$ not necessarily symmetric **Either** there exist $X \succeq 0$ and $AX + XA^T = B$, **Or** there exists symmetric matrix Y such that $A^TY + YA \succeq 0$ and $B \bullet Y < 0$.

1.3 Strong Duality

THEOREM 1.1. **(Strong duality theorem)**: *Suppose that A is full rank and $\mathcal{P} = \{A\mathbf{x} : \mathbf{x} \geq_{\mathcal{K}} 0\}$ is closed. If z_1 and z_2 are the finite optimal solution values to the primal problem and the dual problem, respectively, then $z_1 = z_2$.*

Proof. By weak duality, $z_1 \geq z_2$. Suppose $z_1 > z_2$, then the following system

$$\begin{aligned} \mathbf{c}^T\mathbf{x} &= z_2 \\ A\mathbf{x} &= \mathbf{b} \\ \mathbf{x} &\geq_{\mathcal{K}} 0 \end{aligned}$$

is infeasible. Then, by Farkas lemma, $\exists y_0, \mathbf{y}$ such that

$$A^T\mathbf{y} + \mathbf{c}^T y_0 \geq_{\mathcal{K}^*} 0 \text{ and } \mathbf{b}^T\mathbf{y} + z_2 y_0 < 0 \tag{1.1}$$

Exercise 3. Why is invoking Extended Farkas Lemma is justified here?

There are three possibilities.

1. $y_0 = 0$. Then, $A^T\mathbf{y} \geq_{\mathcal{K}^*} 0$, and $\mathbf{b}^T\mathbf{y} < 0$, which implies that the primal problem is infeasible. Therefore, z_1 is not finite, a contradiction.

2. $y_0 > 0$. In this case, dividing both sides of 1.1 by y_0 we get

$$\mathbf{c}^T - A^T\left(\frac{-\mathbf{y}}{y_0}\right) \geq_{\mathcal{K}^*} 0, \text{ and } z_2 - \mathbf{b}^T\left(\frac{-\mathbf{y}}{y_0}\right)$$

   which means z_2 is not an optimal solution of the Dual.

3. $y_0 < 0$. In this case, dividing both sides of 1.1 by y_0 we have

$$-\mathbf{c} + A^T\left(\frac{-\mathbf{y}}{y_0}\right) \geq_{\mathcal{K}^*} 0 \text{ and } -\mathbf{z_2} + \mathbf{b}^T\left(\frac{-\mathbf{y}}{y_0}\right) < -\epsilon < 0.$$

   for some $\epsilon > 0$. Also, optimality of z_2 for the Dual implies that $\exists\mathbf{y}^*$ such that

$$\mathbf{c} - A^T\mathbf{y}^* \geq_{\mathcal{K}^*} 0 \text{ and } \mathbf{z_2} - \mathbf{b}^T\mathbf{y}^* < \epsilon.$$

   Therefore, adding up, we get

$$A^T\left(\frac{-\mathbf{y}}{y_0} - \mathbf{y}^*\right) \geq_{\mathcal{K}^*} 0 \text{ and } \mathbf{b}^T\left(\frac{-\mathbf{y}}{y_0} - \mathbf{y}^*\right) < 0.$$

   And we get that the Primal problem is infeasible.

Thus $z_1 > z_2$ resulted in contradiction. Since $z_1 \geq z_2$ by weak duality, we must have $z_1 = z_2$. ■

Exercise 4. In the standard primal and dual SDP pair

$$\begin{array}{llll} \min & C \bullet X & & \max \;\; \sum_i b_i y_i \\ \text{s.t.} & A_i \bullet X = b_i & \text{for } i = 1, \cdots, m & \text{s.t.} \;\; \sum_i y_i A_i + S = C \\ & X \succeq 0 & & \phantom{\text{s.t.}} \;\; S \succeq 0 \end{array}$$

Let $m = n = 4$, $\mathbf{b} = (0, 0, 0, 10)^T$, and

$$C = E_{12} + E_{21}, \quad A_1 = E_{22}, \quad A_2 = E_{13} + E_{31}, \quad A_3 = E_{23} + E_{32} \;\; \text{and} \; A_4 = -E_{12} - E_{21}$$

a. Find the optimal value for the primal problem and optimal value for the dual problem.

b. Why is there a duality gap here, even when both primal and dual have finite solutions that are attained at finite points?

2 A Review of Kronecker Product

Kronecker product will play an important role in the upcoming lectures, and once mastered will greatly simplify several algebraic manipulations for the semidefinite programming problem. Here we just review the definition and some properties.

Definition 2.1. *(Kronecker Product). Given two matrices $A \in R^{m \times n}$ and $B \in R^{p \times q}$. The Kronecker product is defined as follows:*

$$A = \begin{pmatrix} a_{11} & . & . & . & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & . & . & . & a_{mn} \end{pmatrix} \; \textit{and } B = \begin{pmatrix} b_{11} & . & . & . & b_{1q} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{p1} & . & . & . & b_{pq} \end{pmatrix}$$

Then

$$A \otimes B = \begin{pmatrix} a_{11}B & . & . & . & a_{1n}B \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1}B & . & . & . & a_{mn}B \end{pmatrix}$$

Also sometimes we need to think of matrices, not as linear operations, but themselves as vectors to be operated on by other linear transformations. We define two operation **vec** and Mat which map a matrix to a vector and vice versa.

Definition 2.2. *If $A \in \Re^{m \times n}$, then $\mathbf{vec}(A)$ is a column vector whose entries come from A by stacking up columns of A, from the fist to n'th column, on top of each other.*

Example. if

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \text{ Then } \mathbf{vec}(A) = \begin{pmatrix} 1 \\ 3 \\ 2 \\ 4 \end{pmatrix}$$

The most important and useful properties of the Kronecker product are summarized below:

1. $\mathbf{vec}(ABC) = (C^T \otimes A)\mathbf{vec}(B)$
2. $(AB) \otimes (CD) = (A \otimes C)(B \otimes D)$
3. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
4. If A is a $m \times m$ matrix and B a $n \times n$ matrix then the eigenvalues of $A \otimes B$ are $\lambda_i(A)\lambda_j(B)$ for $i = 1, \cdots, m$ and $j = 1, \cdots, n$. The eigenvectors of $A \otimes B$ corresponding to eigenvalue $\lambda_i(A)\lambda_j(B)$ is $q_i(A) \otimes q_j(B)$, where $q_i(A)$ is eigenvector of A corresponding to $\lambda_i(A)$ and $q_j(B)$ is the eigenvector of B corresponding to $\lambda_j(B)$.