Semidefinite Programming
Lecture 1

scribe: Reuben Settergren

January 23 1995

1 Introduction

The standard form Semidefinite Program problem (SDP) is defined as follows:

$$\begin{array}{rl} \min & C \bullet X \\ \text{st} & A_i \bullet X = b_i, \quad i = 1 \dots m \\ & X \succeq 0 \end{array}$$

where X, C, A_i are all real, symmetric, square matrices, $A \bullet B = \text{trace}\,(A^T B)$ is the inner product of two matrices (note that in case of symmetric matrices off-diagonal entries are included twice), and $A \succeq 0$ means that matrix A is positive semidefinite (PSD), i.e. $\forall \mathbf{x}$, $\mathbf{x}^T A \mathbf{x} \geq 0$. In general $A \succeq B$ means $A - B$ is positive semidefinite and $A \succ B$ means $A - B$ is positive definite. The partial order relation $\succeq$ is sometimes referred as Löwner partial order.

2 An Example

Here is an example of how SDP's could arise; the example is from Graph theory. For a simple, undirected graph G, let A be the following matrix, similar to an adjacency matrix: A has 1's on it's entire diagonal. For edge $\{i, j\} \in G$, $A_{ij} = A_{ji} = 1$. For non-edge $\{i, j\} \notin G, A_{ij} = A_{ji} = x_{ij}$, a variable. So we have $A(\mathbf{x})$, where $\mathbf{x}$ is the vector of the variables. Now define the function

$$\theta(G) = \min_{\mathbf{x}} \lambda_1(A(\mathbf{x})),$$

where $\lambda_1(A)$ is the largest eigenvalue of A (on down to $\lambda_n(A)$ is the smallest). $\theta(G)$ is called the Lovàsz number of the graph G. $\theta(G)$ has the nice property that, if $\omega(G)$ the maximum clique-size in G,

THEOREM 2.1. *For graph G, $\theta(G) \geq \omega(G)$.*

Proof. If $\omega(G) = k$, permute the rows and columns of $A(G)$ so that the first k vertices $v_1 \dots v_k$ represented in A are those of a maximum clique. Since $v_1 \dots v_k$ are a clique, the principal submatrix formed by those rows and columns has a 1 in every position; call this submatrix J_k. Recall from matrix theory that the largest eigenvalue of a matrix is no smaller than the largest eigenvalue of any principal submatrix. Thus, by the definition of θ, for any value of $\mathbf{x}$,

$$\theta(G) \geq \lambda_1(A(\mathbf{x})) \geq \lambda_1(J_k) = k = \omega(G).$$

■

3 A Short Review of Matrices

Before we continue with this example, let us digress and review some useful facts and definitions from matrix theory.

If $A\mathbf{x} = \lambda\mathbf{x}$, we say that $\mathbf{x}$ is an **eigenvector** and λ is an **eigenvalue** of A. For a (possibly complex) matrix A, A^* is A's conjugate transpose; if $A = A^*$, A is called **Hermitian**. In the real case, $A = A^* = A^T$ is just symmetric.

LEMMA 3.1. *Hermitian matrices have only real eigenvalues.*

Proof. If A is Hermitian, $A\mathbf{x} = \lambda\mathbf{x} \Rightarrow \mathbf{x}^*A^* = \overline{\lambda}x^* \Rightarrow \mathbf{x}^*A = \overline{\lambda}\mathbf{x}^* \Rightarrow \mathbf{x}^*A\mathbf{x} = \overline{\lambda}\mathbf{x}^*\mathbf{x}$. Since $\mathbf{x}^*A\mathbf{x}$ is a scalar, $\overline{(\mathbf{x}^*A\mathbf{x})} = (\mathbf{x}^*A\mathbf{x})^* = \mathbf{x}^*A^*\mathbf{x} = \mathbf{x}^*A\mathbf{x}$, so $\mathbf{x}^*A\mathbf{x}$ is real, and similarly, so is $\mathbf{x}^*\mathbf{x} = \mathbf{x}^*I\mathbf{x}$, so $\overline{\lambda}$ is real. ■

If, for a matrix Q, $Q^TQ = QQ^T = I$, Q is said to be **orthogonal**.

THEOREM 3.1. (SCHUR) *If A is a square matrix, then there exists an orthogonal matrix Q such that $QAQ^T = T$, where T is an upper triangular matrix.*

COROLLARY 3.1. *If A is symmetric, there is an orthogonal matrix Q such that $QAQ^T = D$ and D is a diagonal matrix.*

Proof. From Schur's theorem, there is orthogonal matrix Q such that $QAQ^T = T$ where T is triangular. Since A is symmetric so must be T; but the only symmetric triangular matrices are diagonal matrices. ■

Writing $A = Q^TDQ$, is referred to as the eigenvalue decomposition of A because the eigenvalues of A are the diagonal entries of D, and the eigenvectors are the columns of Q.

THEOREM 3.2. (RAYLEIGH-RITZ) *For A symmetric,*

$$\begin{aligned} \lambda_1(A) &= \max_{||\mathbf{x}||=1} \mathbf{x}^TA\mathbf{x} \\ \lambda_n(A) &= \min_{||\mathbf{x}||=1} \mathbf{x}^TA\mathbf{x} \end{aligned}$$

Proof. Since A is symmetric, decompose it as $A = Q^TDQ$. If we let $\mathbf{v} = Q\mathbf{x}$, then for any $\mathbf{x}$, $\mathbf{x}^TA\mathbf{x} = \mathbf{x}^TQ^TDQ\mathbf{x} = \mathbf{v}^TD\mathbf{v} = \sum v_i^2D_i$. Since Q is orthogonal, $||\mathbf{v}|| = ||\mathbf{x}|| = 1$, then $\lambda_n = \min D_i \leq \sum v_i^2D_i = \mathbf{x}^TA\mathbf{x} \leq \max D_i = \lambda_1$. The theorem is proved by realizing that equality holds when $\mathbf{x}$ is an eigenvector corresponding either to λ_1 or λ_n. ■

COROLLARY 3.2. *A symmetric matrix A is positive semidefinite if and only if its smallest eigenvalue $\lambda_n \geq 0$.*

Finally, since the set of positive semidefinite matrices is closed under addition and scalar multiplication, we can see that it is a convex set. Also, we observe that $\mathbf{x}^T A\mathbf{x} \geq 0$ for all $\mathbf{x} \Leftrightarrow (\mathbf{x}\mathbf{x}^T) \bullet A \geq 0 \;\; \forall \mathbf{x}$ defines infinitely many halfspaces. Since the PSD cone is the intersection of these halfspaces, it is convex.

Note now that $\mathbf{x}^T A\mathbf{x} = A \bullet (\mathbf{x}\mathbf{x}^T)$ (and in general, $A \bullet (BB^T) = \text{trace}\,(B^T AB)$), and $||\mathbf{x}|| = \text{trace}\,(\mathbf{x}\mathbf{x}^T)$. Consider, therefore, the following SDP:

$$\begin{aligned} \lambda' = \max \quad & A \bullet X \\ \text{st} \quad & I \bullet X = 1 \\ & X \succeq 0 \end{aligned}$$

We will show that $\lambda' = \lambda_1(A)$. To show that $\lambda' \geq \lambda_1$, let, for any unit-length vector $\mathbf{x}$, $X = \mathbf{x}\mathbf{x}^T$. First of all, $A \bullet X = \mathbf{x}^T A\mathbf{x}$. Furthermore, X is a rank one matrix, and it's only nonzero eigenvalue is 1 (since $(\mathbf{x}\mathbf{x}^T)\mathbf{x} = \mathbf{x}(||\mathbf{x}||^2)$), so X is PSD; also $I \bullet X = \text{trace}\,(X) = ||\mathbf{x}||^2 = 1$. Thus, X is feasible for the SDP, and $\lambda' \geq \lambda_1$. To prove $\lambda' \leq \lambda_1$ we need a lemma which will be proved in the next lecture.

Let us take another look at this problem. For any matrix A, the eigenvalues of $-A$ are the negatives of the eigenvalues of A, and the eigenvalues of $A + zI$ are z plus the eigenvalues of A. Thus $zI - A$ is positive semidefinite iff $z \geq \lambda_1(A)$, and the following semidefinite program (which happens to be the dual of the previous) also yields

$\lambda_1(A)$:

$$\begin{aligned} \min \quad & z \\ \text{st} \quad & zI - A \succeq 0 \end{aligned}$$

Returning now to our graph theory example, what are we to do about the fact that A is a function of it's variable entries? That would cause the objective function of the first SDP to become quadratic, but in the second SDP, we can allow the matrix to vary in the following way

$$\begin{aligned} \min \quad & z \\ \text{st} \quad & E_{ij} \bullet X = 1 \quad \{i,j\} \in E(G) \text{ or } i = j \\ & zI - X \succeq 0 \end{aligned}$$

The first constraint fixes to 1 diagonal entries of the matrix, and those corresponding to edges of the graph. The second constraint ensures that z is larger than or at least equal to the largest eigenvalues of X. The objective function, then ensures that z at the optimum is the minimal possible largest eigenvalue of X.

4 Cones and Cone Duality

A **cone** is a set A such that if $\mathbf{x} \in A$ then $\forall \alpha \geq 0$, $\alpha\mathbf{x} \in A$. A cone is **convex** if $x, y \in A$ implies $x + y \in A$. A cone is **pointed** if $A \cap (-A) = \{0\}$. A **closed** cone is a closed set; in particular, it contains its boundary.

Here are some examples of closed, pointed, convex cones:

1. Positive orthant: $\{\mathbf{x} : x_i \geq 0, i = 1, \ldots, n\}$
2. Ice cream cone: $\left\{(x_0, x_1, \ldots x_n) : x_0 \geq \sqrt{x_1^2 + \ldots + x_n^2}\right\}$
3. p-cones: $\left\{(x_0, x_1, \ldots x_n) : x_0 \geq (x_1^p + \ldots + x_n^p)^{1/p}\right\}$ (Note: p cones are convex only for $p \geq 1$.)
4. Positive semidefinite matrices (They only exist in spaces of dimension $n = p(p+1)/2$)

If K is a cone, then the cone **polar** to K is the cone $K^* = \{\mathbf{z} : \mathbf{z}^T\mathbf{x} \geq 0, \forall \mathbf{x} \in K\}$.

Exercise 1. *Show that the positive orthant, the ice cream cone and the cone of PSD matrices are all self-polar, i.e.* $K^* = K$.

Of course, not all cones are self-polar. The p-cone, which is a generalization of the ice cream cone, is polar to the q-cone, where $\frac{1}{p} + \frac{1}{q} = 1$ (which shows why the ice cream cone, with $p = 2$, is self-polar).

Exercise 2. *Show that p-cones are not convex when* $p < 1$.

A pointed, convex cone K yields a partial order $\geq_K$, defined by

$$\mathbf{a} \geq_K \mathbf{b} \Leftrightarrow \mathbf{a} - \mathbf{b} \in K.$$

Since $\mathbf{0} \in K$, reflexivity is assured, pointed-ness gives antisymmetry, and convexity causes transitivity. With this in mind, we can formulate a general optimization problem:

$$\begin{array}{ll} \textbf{Primal}: & \\ \min & \mathbf{c}^T\mathbf{x} \\ \text{st} & \mathbf{a}_i^T\mathbf{x} = \mathbf{b}_i \\ & \mathbf{x} \geq_K 0 \end{array}$$

We can also define (and then justify) a dual of this optimization problem:

$$\begin{array}{ll} \textbf{Dual}: & \\ \max & \mathbf{b}^T\mathbf{y} \\ \text{st} & A^T\mathbf{y} + \mathbf{s} = \mathbf{c} \\ & \mathbf{s} \geq_{K^*} 0 \end{array}$$

LEMMA 4.1. (WEAK DUALITY) *If $\mathbf{x}$ is feasible for the "primal" problem, and $\mathbf{y}$ and $\mathbf{s}$ are feasible for the "dual" problem, then* $\mathbf{c}^T\mathbf{x} \geq \mathbf{b}^T\mathbf{y}$.

Proof.

$$\begin{aligned} \mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y} &= \mathbf{c}^T\mathbf{x} - (A\mathbf{x})^T\mathbf{y} \\ &= \mathbf{c}^T\mathbf{x} - \mathbf{y}^T A\mathbf{x} \\ &= (\mathbf{c} - \mathbf{y}^T A)\mathbf{x} \\ &= \mathbf{s}^T\mathbf{x} \\ &\geq 0 \end{aligned}$$

the last inequality is by definition of polar cones. ■

For any "primal" feasible $\mathbf{x}$ and "dual" feasible $\mathbf{y}$ and $\mathbf{s}$, the nonnegative quantity $\mathbf{x}^T\mathbf{s}$ is called the **duality gap**. And just as in Linear Programming, if the duality gap is 0, i.e. if $\mathbf{c}^T\mathbf{x} = \mathbf{b}^T\mathbf{y}$, then $\mathbf{x}$ and $(\mathbf{y}, \mathbf{s})$ are optimal solutions to their respective problems.