Lectures 9 and 10

Lecturer: Michel X. Goemans *Scribe: Chiheon Kim and Michel X. Goemans*

Today, we are going to present how to use semidefinite programming to optimize a linear function over the theta body $TH(G)$ which was discussed in last lecture.

1 Semidefinite programming

Let $\mathcal{S}^{n\times n}$ be the set of n by n real symmetric matrices.

Definition 1 *$A \in \mathcal{S}^{n\times n}$ is called* positive semidefinite*, denoted $A \succeq 0$, if $x^T Ax \geq 0$ for any $x \in \mathbb{R}^n$.*

There are several well-known equivalent ways to state positive semidefiniteness.

Proposition 1 *The following are equivalent:*

(i) A is positive semidefinite.

(ii) Every eigenvalue of A is nonnegative.

(iii) There is a matrix V such that $A = V^T V$.

We define the Frobenius inner product for two matrices $A, B \in \mathcal{S}^{n\times n}$ by

$$\langle A, B\rangle = \sum_{i,j} A_{ij}B_{ij} = tr(A^T B).$$

Typically, semidefinite programs are described as optimization problems of the following form:

$$\sup \langle C, X\rangle \quad \text{subject to} \quad \langle A_i, X\rangle = b_i \text{ for } i = 1, \dots, m,\ X \succeq 0 \tag{1}$$

where C and A_i are n by n symmetric matrices and b_i are real numbers. Semidefinite programming generalizes linear programming because linear program is a special instance of semidefinite program that all C, A_i and X are diagonal matrices.

For the given semidefinite program (1), we define the dual program of it as

$$\inf \sum_{i=1}^{m} b_i y_i \quad \text{subject to} \quad \sum_{i=1}^{m} y_i A_i \succeq C. \tag{2}$$

Here we use the convention that $A \succeq B$ if $A - B$ is positive semidefinite.

Theorem 2 (Weak duality) *For any feasible solution X for the primal program (1) and a feasible solution y for the dual (2), $\sum_{i=1}^{m} b_i y_i \geq \langle C, X\rangle$.*

To prove the weak duality, we need the following useful lemma.

Lemma 3 *Let A and B be n by n positive semidefinite matrices. Let $C = A \circ B$ be an n by n matrix such that $C_{ij} = A_{ij}B_{ij}$. Then C is positive semidefinite. Moreover, $\langle A, B\rangle \geq 0$.*

Proof: Note that $\langle A, B\rangle = e^T C e$ where e is all-ones vector in $\mathbb{R}^n$. So $\langle A, B\rangle \geq 0$ if C is positive semidefinite.

Let A and B be positive semidefinite matrices. By the previous proposition, $A = V^T V$ and $B = W^T W$ for some matrices V and W. Let v_i (and w_i) be the i-th column of V (W, resp.) for $i = 1, \ldots, n$. Let $z_i = v_i \otimes w_i$, where component (k, l) of z_i is equal to component k of v_i times component l of w_i. Define Z be the matrix whose columns are $z_1, \ldots, z_n$. We have $z_i^T z_j = (v_i^T v_j)(w_i^T w_j) = A_{ij} B_{ij} = C_{ij}$. Hence $C = Z^T Z$ and C is positive semidefinite. $\square$

This lemma also implies the result that the positive semidefinite cone is self-dual, i.e. $\langle A, B\rangle \geq 0$ for all $B \succeq 0$ if and only if $A \succeq 0$. (For the only if part, if $A \not\succeq 0$ then any eigenvector v corresponding to a negative eigenvalue satisfies $\langle A, vv^T\rangle \geq 0$.

Proof of Theorem 2: $\sum_{i=1}^m b_i y_i - \langle C, X\rangle = \sum_{i=1}^m y_i \langle A_i, X\rangle - \langle C, X\rangle = \langle X, \sum_{i=1}^m y_i A_i - C\rangle \geq 0$ because $X \succeq 0$ and $\sum_{i=1}^m y_i A_i - C \succeq 0$. $\square$

In comparison to linear programming, strong duality does not always hold for semidefinite programming. Also, an optimum is not always attained by a feasible solution (which is the reason that we used inf instead of min).

Example 1 *Let the primal to be*

$$\sup \left\langle \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, X \right\rangle \quad \textit{subject to} \quad \left\langle \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, X \right\rangle = 1, \left\langle \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, X \right\rangle = 0, X \succeq 0.$$

Then dual is

$$\inf y_1 \quad \textit{subject to} \quad y_1 \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + y_2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \succeq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

This primal-dual pair doesn't satisfy strong duality since the primal has optimum value -1 but dual has the optimum value 0 (and in this case both primal and dual attain their optimum value).

Theorem 4 *Strong duality If the dual (primal, resp.) is feasible and the primal (dual, resp.) has a strictly feasible solution, i.e., there is positive definite $X^{(0)}$ with $\langle A_i, X^{(0)}\rangle = b_i$ for $i = 1, \ldots, m$, then there is no duality gap and dual (primal, resp.) has a feasible solution which attains the minimum.*

We will prove strong duality next time.

2 Theta body

Let $G = (V, E)$ be a graph, and $w : V \to \mathbb{R}_{\geq 0}$ be weights for vertices. Recall that a collection of unit vectors $\{v_i\}_{i \in V}$ and a unit vector c is called an orthonormal representation of G is $v_i^T v_j = 0$ for any $(i, j) \notin E$. The theta body of G is defined as

$$TH(G) = \left\{ x \in \mathbb{R}^{|V|} : \sum_{i \in V} (c^T u_i)^2 x_i \leq 1 \text{ for all orthonormal rep. of } G \right\}.$$

We define the weighted theta number of G as $\theta(G, w) = \max_{x \in TH(G)} w^T x$.

Theorem 5 *Let $W \in \mathcal{S}^{n\times n}$ with $W_{ij} = \sqrt{w_i w_j}$. Then,*

$$\begin{aligned} \theta(G,w) = \quad & \max \quad \langle W, X\rangle \\ & s.t. \quad \langle I, X\rangle = 1 \\ & \qquad\quad \langle E_{ij}, X\rangle = 0 \text{ for } (i,j)\in E \\ & \qquad\quad X \succeq 0, \end{aligned}$$

where E_{ij} is n by n symmetric matrix such that (i,j) and (j,i) entries are 1 but other entries are all zero.

Proof: The dual program of this SDP is

$$\min t \quad \text{subject to} \quad tI + \sum_{ij\in E} y_{ij}E_{ij} \succeq W.$$

Note that $X = \frac{1}{n}I$ is a strictly feasible solution for primal, and $t > \lambda_{max}(W)$ and $y_{ij} = 0$ gives a strictly feasible solution for dual, where $\lambda_{max}(W)$ is the maximum eigenvalue of W. Hence strong duality holds and there are optimal solutions to both primal and dual.

($\le$) We are going to prove that any feasible solution for dual gives an upper bound of $\theta(G,w)$. Let (t,y) be a feasible solution for the dual. Then $M = tI + \sum_{ij\in E} y_{ij}E_{ij} - W \succeq 0$, so there are vectors $s_i \in \mathbb{R}^n$ such that $M_{ij} = s_i^T s_j$ for all $i,j\in[n]$. In particular, $s_i^T s_i = t - w_i$ and $s_i^T s_j = -\sqrt{w_i w_j}$ if $(i,j)\notin E$.

Let c be a unit vector orthogonal to all of s_i (introduce an extra dimension if needed). Let $u_i = (\sqrt{w_i}c + s_i)/\sqrt{t}$. Check that $u_i^T u_i = 1$ and $u_i^T u_j = 0$ if $(i,j)\notin E$. So c and u_i is an orthonormal representation of G.

For $x\in TH(G)$ such that $w^T x = \theta(G,w)$, we have $\sum_i (c^T u_i)^2 x_i \le 1$. Since $(c^T u_i)^2 = w_i/t$, it implies that $\theta(G,w) = w^T x \le t$.

($\ge$) We are going to prove that for each feasible solution X for primal, we can find $x\in TH(G)$ such that $w^T x \ge \langle W, X\rangle$. It implies that $\theta(G,w) = \max_{x\in TH(G)} w^T x \ge \langle W, X\rangle$.

Let X be a feasible solution for primal. Since X is positive semidefinite, there is a matrix R with $X = R^T R$. For each $i = 1,\ldots,n$, let $y_i = Re_i$ where e_i is the standard unit vector in which i-th coordinate is 1. Let $v_i = \frac{y_i}{\|y_i\|}$ and $d = \frac{R\sqrt{w}}{\|R\sqrt{w}\|}$. Here $\sqrt{w}$ is the vector with entries $\sqrt{w_i}$. Clearly v_i and d are unit vectors, and $v_i^T v_j = \frac{e_i^T R^T R e_j}{\|y_i\|\|y_j\|} = 0$ for $(i,j)\in E$ since $e_i^T R^T R e_j = X_{ij} = 0$ if $(i,j)\in E$. It implies that d and v_i is an orthonormal representation of $\bar{G}$, the complement of G.

Claim 6 *Let d and v_i be an orthonormal representation of $\bar{G}$. Let $x_i = (d^T v_i)^2$ for each $i\in V$. Then $x\in TH(G)$.*

Let c and u_i be an orthonormal representation of G. We need to show that $\sum_i (c^T u_i)^2 (d^T v_i)^2 \le 1$. For, let $a_i = u_i \otimes v_i$. Then $a_i^T a_i = (u_i^T u_i)(v_i^T v_i) = 1$ and $a_i^T a_j = (u_i^T u_j)(v_i^T v_j) = 0$ since either $u_i^T u_j = 0$ (if $(i,j)\notin E$) or $v_i^T v_j = 0$ (if $(i,j)\in E$). Hence the a_i's form an orthonormal set of vectors (not necessarily a basis). On the other hand, $\langle c\otimes d, a_i\rangle = (c^T u_i)(d^T v_i)$. Hence,

$$\sum_i (c^T u_i)^2 (d^T v_i)^2 = \sum_i \langle c\otimes d, a_i\rangle^2 \le 1$$

since the a_i's are orthonormal. $\diamond$

By the claim, we get a point $x\in TH(G)$ from d and v_i by letting

$$x_i = (d^T v_i)^2 = \frac{(\sqrt{w}^T R^T R e_i)^2}{\|Re_i\|^2 \|R\sqrt{w}\|^2}.$$

$\|Re_i\|^2 = e_i^T R^T R e_i = X_{ii}$ and $\|R\sqrt{w}\|^2 = \sqrt{w}^T R^T R \sqrt{w} = \langle W, X \rangle$ so

$$x_i = \frac{(\sqrt{w}^T X e_i)^2}{X_{ii}\,\langle W, X\rangle} = \frac{1}{X_{ii}\,\langle W, X\rangle}\left(\sum_{j=1}^{n}\sqrt{w_j}X_{ij}\right)^2$$

since $W = \sqrt{w}\sqrt{w}^T$. Hence

$$\begin{aligned} w^T x &= \frac{1}{\langle W, X\rangle}\sum_{i=1}^{n}\frac{w_i}{X_{ii}}\left(\sum_{j=1}^{n}\sqrt{w_j}X_{ij}\right)^2 \\ &= \frac{1}{\langle W, X\rangle}\sum_{i=1}^{n}(\sqrt{X_{ii}})^2 \cdot \sum_{i=1}^{n}\left(\frac{1}{\sqrt{X_{ii}}}\sum_{j=1}^{n}\sqrt{w_i w_j}X_{ij}\right)^2 \\ &\geq \frac{1}{\langle W, X\rangle}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sqrt{w_i w_j}X_{ij}\right)^2 = \langle W, X\rangle\,. \end{aligned}$$

Second equality comes from that $\sum_i X_{ii} = \langle I, X\rangle = 1$ and third inequality follows from Cauchy-Schwartz.

□

Now we can compute $\theta(G, w)$ by solving an equivalent semidefinite program (either primal or dual). Modulo some assumptions, semidefinite programs can be approximated within any tolerance in polynomial time. To be a bit more precise, there are interior-point algorithms for SDP that when being given an initial strictly primal feasible solution and an initial strictly dual feasible solution of duality gap δ will converge to primal and dual solutions of duality gap at most ϵ in time polynomial in n and in $\log(\Delta/\epsilon)$. In our case, we can take $X_0 = \frac{1}{n}I$ and $t_0 > \lambda_{max}(W)$ for an initial duality gap smaller than $2w(E)/n$. In particular, we can compute $\alpha(G)$ up to ϵ in time polynomial in the input size and $\log(1/\epsilon)$ (and similarly for $\omega(G) = \chi(G)$) when G is perfect. Once we can approximate $\alpha(G, w)$ for a perfect graph, we can easily find a stable set of that weight as the problem self-reduces. Indeed, for any $v \in V$, one can check whether $\alpha(G, w) = w_v + \alpha(G \setminus (\{v\} \cup N(v))$ and if so include v in the stable set and recurse (and otherwise delete v and recurse).

Finding the coloring (or the clique covering) for a perfect graph is slightly more tricky and was not covered in class.

3 Strong Duality for Semidefinite Programming

Strong duality does not always hold for semidefinite programming, although it typically does for combinatorial applications. We'll prove strong duality here as stated in Theorem 4, and explain when it holds and why it sometimes fails.

Suppose we would like to show that there is no primal solution of value $\geq \mu$ for some given μ. This means there is no solution to

$$\begin{cases} \langle C, X\rangle + s = \mu & \\ \langle A_i, X\rangle = b_i & \text{for all } i \\ X \succeq 0, & \\ s \geq 0. & \end{cases}$$

Let us reformulate this differently. Let

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

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

and

$$X' = \begin{pmatrix} X & 0 \\ 0 & s \end{pmatrix}.$$

Then letting

$$S = \left\{ (\langle C', X' \rangle, \langle A_1', X \rangle, \cdots, \langle A_m', X' \rangle) : X' = \begin{pmatrix} X & 0 \\ 0 & s \end{pmatrix} \succeq 0 \right\} \tag{3}$$

then no primal solution of value greater or equal to μ means that $v := (\mu, b_1, \cdots, b_m) \notin S$. At this point, we would like to use the separating hyperplane theorem (see below) to separate v from S since S is a nonempty convex set.

Theorem 7 (Separating hyperplane theorem) *Let $S \subset \mathbb{R}^n$ be a nonempty, closed, convex set and $v \notin S$. Then there exists $y \in \mathbb{R}^n : y^T v < y^T x$ for all $x \in S$.*

However, and here is where the trouble is, the set S given in (3) might not be a closed set. In the case of linear programming, S would be a closed set as it is a linear mapping of a polyhedral set. However, for semidefinite programming, a linear map of the positive semidefinite cone might not be closed. Here is an example.

Example 2 *Consider*

$$S = \left\{ \left(\langle \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, X \rangle, \langle \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, X \rangle \right) : X \succeq 0 \right\}.$$

Observe that $(\epsilon, 2) \in S$ for any $\epsilon > 0$ as it corresponds to $X = \begin{pmatrix} 1/\epsilon & 1 \\ 1 & \epsilon \end{pmatrix} \succeq 0$. However, $(0, 2) \notin S$, showing that S is not closed.

So, in order to use Theorem 7, we need to consider the closure $\bar{S}$ of S, where S is given by (3). So, there are 2 possibilities:

1. $v \notin \bar{S}$; in this case, we'll use the separating hyperplane theorem,
2. $v \in \bar{S} \setminus S$; in this case, we will need to show that we can get arbitrarily close to μ in the primal.

We first consider case 1. By the separating hyperplane theorem, we get a vector $y = (y_0, \cdots, y_m)$ such that

$$y_0 \mu + \sum_{i=1}^{m} y_i b_i < y_0 \langle C', X' \rangle + \sum_{i=1}^{m} y_i \langle A_i', X' \rangle \tag{4}$$

for all $X' = \begin{pmatrix} X & 0 \\ 0 & s \end{pmatrix} \succeq 0$. The left-hand-side of (4) must be < 0 (by considering $X' = 0$), and the right-hand-side must be ≥ 0 for any X' (otherwise we could tak a multiple of X' and contradict (4). Thus,

$$y_0 \mu + \sum_{i=1}^{m} y_i b_i < 0, \tag{5}$$

and from $\langle y_0 C' + \sum_{i=1}^{m} y_i A_i, X' \rangle \geq 0$ for all $X' = \begin{pmatrix} X & 0 \\ 0 & s \end{pmatrix} \succeq 0$, we get $y_0 \leq 0$ and

$$y_0 C + \sum_{i=1}^{m} y_i A_i \succeq 0. \tag{6}$$

If $y_0 = 0$ then we get that $b^T y < 0$ and $\sum_i y_i A_i \succeq 0$ which implies that the dual is unbounded (indeed, if $\tilde{y}$ is dual feasible then $\tilde{y} + \lambda y$ is dual feasible for any $\lambda \geq 0$) contradicting the fact that the primal is feasible. Thus we must have that $y_0 < 0$. By dividing by $-y_0$, we get that $\hat{y} = \frac{1}{-y_0} y$ satisfies $\sum_i \hat{y}_i A_i \succeq C$ and $\hat{y}^T b < \mu$. Thus we have found a dual solution of value strictly smaller than μ.

In the second case, we assume that $v \in \bar{S} \setminus S$. Thus we have a sequence $v^{(k)} \in S$ such that $v^{(k)} \to v$. This means that we have a sequence $X^{(k)} \succeq 0$ such that $\langle A_i, X^{(k)} \rangle \to b_i$ for all $i = 1, \cdots, m$ and $\langle C, X^{(k)} \rangle \to \mu$. But the A_i's can be assumed to be linearly independent (otherwise just get rid of linearly dependent equations) and therefore, there exists Z_i for $i = 1, \cdots, m$ such that $\langle A_j, Z_i \rangle = \delta_{ij}$ (ie. 1 if $i = j$ and 0 otherwise). Now, consider

$$\hat{X}^{(k)} = X^{(k)} + \sum_{i=1}^{m} (b_i - \langle A_i, X^{(k)} \rangle) Z_i.$$

Observe that $\langle A_i, \hat{X}^{(k)} \rangle = b_i$ for all i, but of course $X^{(k)} \not\succeq 0$. Still it is close to be, as $X^{(k)}$ is and $b_i - \langle A_i, X^{(k)} \rangle \to 0$ as $k \to \infty$. We'll finally use the condition that there exists a strictly feasible solution $X^{(0)}$ to the primal. This means that for an appropriate $0 < \alpha_k < 1$, we have that $\bar{X}^{(k)} = \alpha_k X^{(0)} + (1 - \alpha_k) \hat{X}^{(k)}$ is primal feasible and, after calculations, one can show that we can choose α_k such that $\alpha_k \to 0$ as $k \to \infty$. Thus $\bar{X}^{(k)}$ is primal feasible and its primal value converges to μ.

Summarizing, for any μ, we have one of the following:

- a primal feasible solution of value $\geq \mu$,
- a dual feasible solution of value $< \mu$, or
- a sequence of primal feasible solutions converging to μ.

By choosing μ to be the infimum in the dual, we see that there is no duality gap. We haven't shown yet that the dual value is attained, but we leave this to the reader.