Semidefinite Programming	**Lecture 10**
OR 6327 Spring 2012	**February 23, 2012**
Scribe: Alex Fix	

1 Concluding discussion of the Lovász ϑ-Function

We will use the insights of the previous lectures to end up with a general way to use SDP for combinatorial optimization problems. But first, we'll we will further explore the relationship between the ϑ-function and the stability and clique covering numbers of a graph.

Recall the two problem formulations for the ϑ function:

$$(P)\quad \begin{array}{rrcll} \max & e^T X e & & & \\ & I \bullet X & = & 1, & \\ & x_{ij} & = & 0, & ij \in E, \\ & X & \succeq & 0; & \end{array} \qquad\qquad (D)\quad \begin{array}{rl} \min & \lambda_{\max}(ee^T + F) \\ & f_{ij} = 0, \quad ij \notin E. \end{array}$$

We showed last time that the optimal solution to both problems is equal to $\vartheta(G)$. Since we have characterized ϑ by both a minimization problem and a minimization problem, we can prove the following *Sandwich Theorem*:

Theorem 1 *For all graphs G, if $\alpha(G)$ is the size of the largest stable set of G, and $\overline{\chi}(G)$ is the size of the smallest clique cover of G, then we have*

$$\alpha(G) \leq \vartheta(G) \leq \overline{\chi}(G).$$

Proof:

- For the left hand inequality: We will find a feasible solution to (P) with objective value at least $\alpha(G)$. If S is any stable set of G, then we claim that $X := \frac{\chi^S (\chi^S)^T}{|S|}$ is a feasible solution to (P) with $e^T X e \geq |S|$.

  Here, χ^S is the characteristic vector of S, i.e., $(\chi^S)_i = \begin{cases} 1 & i \in S \\ 0 & i \notin S \end{cases}$. So, χ^S has 1s corresponding to the stable set.

  Because of the fact that S is stable, for any $ij \in E$, we have either $(\chi^S)_i$ or $(\chi^S)_j$ is 0, so $x_{ij} = 0$ whenever $ij \in E$. Then, it's clear that $I \bullet X = 1$, since X has a $\frac{1}{|S|}$ on the diagonal for each x_{ii} with $i \in S$. Finally, since X is an outer product, we have that $X \succeq 0$, so X is feasible for (P) with value $|S|$.

- For the right hand inequality, we'll use the second formulation and construct a feasible solution to (D).

  Suppose $C_1, \ldots, C_p$ form a clique covering of G. Without loss, we can assume they partition V. For concreteness, renumber the vertices so that

  $$C_1 = \{1, \ldots, n_1\}, C_2 = \{n_1 + 1, \ldots, n_1 + n_2\}, \ldots, C_p = \{n_{p-1} + 1, \ldots, n_1 + \cdots + n_p = n\}.$$

We want to end up with some feasible F with $\lambda_{\max}(ee^T + F) \leq p$. That is, we want $pI - ee^T - F \succeq 0$.

Our proposed solution is to somehow choose F so that we end up with

$$pI - ee^T - F = \begin{bmatrix} (p-1)e_{n_1}e_{n_1}^T & -1 & \cdots & -1 \\ -1 & (p-1)e_{n_2}e_{n_2}^T & & -1 \\ \vdots & & \ddots & \vdots \\ -1 & -1 & \cdots & (p-1)e_{n_p}e_{n_p}^T \end{bmatrix}.$$

We can choose such an F as follows: let F be 0 along the diagonal, and inside each $n_i \times n_i$ block, set F to be $-p$ off the diagonal. Outside the blocks, set F to be 0.

Then, we have that F is feasible for (D) since for every $ij \notin E$ we have that i and j are not in the same clique, so the ijth entry is outside a block, so $f_{ij} = 0$.

Furthermore, we can write this same matrix as the following sum:

$$pI - ee^T - F = \sum_{i<j} (\chi^{C_i} - \chi^{C_j})(\chi^{C_i} - \chi^{C_j})^T.$$

This equality holds, since within a clique we get $p-1$ cases of 1 times 1, and for anything that's an edge not in a clique, we will get a single instance of 1 times -1 giving -1 for all entries outside the blocks.

Now, since we've added up a bunch of psd matrices, this matrix is psd. Therefore, $\vartheta(G) \leq \overline{\chi}(G)$.

□

Remark: *Perfect Graphs* are graphs where $\alpha(G) = \overline{\chi}(G)$, and where the same equality holds for all induced subgraphs of G. So, if G is perfect, by computing $\vartheta(G)$ we know both $\alpha(G)$ and $\overline{\chi}(G)$. However, just knowing the size of the maximum stable set doesn't directly give an actual stable set of size $\alpha(G)$.

The old method for computing such a stable set is as follows: compute $\vartheta(G)$. This will give an integer, since G is perfect. Now, remove a vertex v and compute $\vartheta(G)$ again. If it went down, then v is in every stable set, so we can remove it and its neighbors from the graph and recur. Similarly, if $\vartheta(G)$ does not go down, there is at least one stable set not containing v, so we can remove it from the graph and recur again.

The inefficiency of this method will motivate the next section: trying to find formulations of ϑ where the optimal solution gives us information about the associated combinatorial problems.

2 Programs for $\vartheta(G)$ that give combinatorially meaningful solutions

Here is another characterization of $\vartheta(G)$ (for more information see Knuth's paper on the website).

$$(\overline{P}) \quad \begin{array}{rrcl} \max_{x,W\in\mathbb{M}^n} & & e^T x & \\ & \begin{pmatrix} 1 & x^T \\ x & W \end{pmatrix} & \succeq & 0, \\ & \operatorname{diag}(W) & = & x, \\ & w_{ij} & = & 0, \quad ij \in E. \end{array}$$

The reason for considering this program is that it has a vector that's "like" the characteristic vector of a stable set.

[As a side-note, looking at the submatrix $\begin{pmatrix} 1 & x_i \\ x_i & x_i \end{pmatrix}$ which must be psd, we have that $0 \le x_i \le 1$. Thus, we can at least interpret the x_i as fractional values on the vertices of G.]

Theorem 2 $v(\overline{P}) = \vartheta(G)$.

Proof: Note first that $v(\overline{P}) > 0$, since we can choose $W = \frac{1}{n}I$ and $x = \frac{1}{n}e$. Similarly, $v(P) > 0$ since we already know it's between 1 and n.

So, suppose (x, W) is feasible in $(\overline{P})$ with $e^T x > 0$. Then, set $X = \frac{1}{e^T x}W$. We want to show that this X is feasible for (P) with at least the same value. First, we have

$$I \bullet X = \frac{I \bullet W}{e^T x} = \frac{e^T \operatorname{diag}(W)}{e^T x} = \frac{e^T x}{e^T x} = 1.$$

We also have $X \succeq 0$ since $W \succeq 0$, and finally, $x_{ij} = 0$ for $ij \in E$ since the same holds for W. Thus, X is feasible.

As for the objective value, we have that $W - xx^T \succeq 0$ by Schur complements, so pre- and post-multiplying by e gives $e^T W e \ge (e^T x)^2$, and hence

$$e^T X e = \frac{(e^T x)^2}{e^T x} \ge e^T x.$$

Therefore, $v(\overline{P}) \le v(P) = \vartheta(G)$.

Now, suppose conversely that X is a feasible solution to (P) with $e^T X e > 0$. We're going to factor X as $Y^T Y$ with $Y = [y_1, \ldots, y_n]$.

Let T be the set of indices i where $y_i \neq 0$. Define $v_i = \frac{y_i}{||y_i||}$ for $i \in T$ and let $\{v_j, j \notin T\}$ be any set of $n - |T|$ orthonormal vectors orthogonal to the v_i for $i \in T$.

Also, let $d = \frac{Ye}{||Ye||} = \frac{Ye}{\sqrt{e^T X e}}$.

Note that $(d, v_1, \ldots, v_n)$ is an orthonormal representation of $\overline{G}$.

Now, we'll define $Z := [d, v_1, \ldots, v_n]^T [d, v_1, \ldots, v_n] \in \mathbb{M}^n$. We can see that

$$Z = \begin{pmatrix} 1 & d^T V \\ V^T d & V^T V \end{pmatrix} \qquad \text{where } V = [v_1, \ldots, v_n].$$

We do have that $\operatorname{diag}(V^T V) = e$, since all the v_i are unit vectors, but $V^T d$ might not be e. So, let $\Lambda = \operatorname{Diag}([1; V^T d])$. Then (writing $(V^T d)^2$ for the vector $((d^T v_i)^2))$)

$$\Lambda Z \Lambda = \begin{pmatrix} 1 & ((V^T d)^2)^T \\ (V^T d)^2 & W \end{pmatrix} \qquad \text{and } w_{ii} = (d^T v_i) \cdot 1 \cdot (d^T v_i) = (d^T v_i)^2.$$

So, this is feasible for $(\overline{P})$ with $x = (V^T d)^2, x_i = (d^T v_i)^2$.

For the objective value, we have:

$$\begin{aligned} e^T x &= \sum_i (d^T v_i)^2 \\ &= (I \bullet X) \sum_i (d^T v_i)^2 \\ &= \left(\sum_i ||y_i||^2 \right) \sum_i (d^T v_i)^2. \end{aligned}$$

And by Cauchy-Schwarz, this last line is at least

$$\left(\sum_i ||y_i|| d^T v_i \right)^2 = \left(d^T \sum_i ||y_i|| v_i \right)^2 = (d^T Y e)^2 = \left(\frac{e^T Y^T Y e}{\sqrt{e^T X e}} \right)^2 = e^T X e.$$

(For the second equality, recall that y_i is zero for $i \notin T$.) Therefore, $v(\overline{P}) \geq \vartheta(G)$. □

Question: What if we want to know what x satisfies $(\overline{P})$ for some W?

We could define $\text{TH}(G) = \{x \in \mathbb{R}^n_+ \mid \exists W, (x, W) \text{ feasible in } \overline{P}\}$, and then try to find $\max\{e^T x \mid x \in \text{TH}(G)\} = \vartheta(G)$.

In fact, Lovász and Schrijver have given another representation of TH(G):

$$\text{TH}(G) = \{x \in \mathbb{R}^n_+ \mid \sum_i (c^T u_i)^2 x \leq 1 \text{ for all orthonormal rep'ns. } (c, u_1, \ldots, u_n) \text{ of } G\}.$$

This is a semi-infinite formulation of the set, in that it has a finite number of variables, but infinitely many constraints. We'll prove just the one inclusion, $\subseteq$. To do so, we'll have to "reverse engineer" an orthonormal representation of $\overline{G}$ from x, W and then show that all inequalities hold.

Proof: Let $\Lambda = \text{Diag}\left([1; \sqrt{x_1}; \ldots, \sqrt{x_n}]\right)$, and $Z = \Lambda^{-1} \begin{pmatrix} 1 & x \\ x^T & W \end{pmatrix} \Lambda^{-1}$. Then $Z \succeq 0$ since we're pre- and post-multiplying a psd matrix by an invertible matrix. Note that we have $\text{diag}\,(Z) = \overline{e}$ so we can write it as

$$Z = [d, v_1, \ldots, v_n]^T [d, v_1, \ldots, v_n].$$

We now have an "orthonormal representation" of $\overline{G}$ (that is, all vectors have unit length, since the diagonal of Z is all 1s, and they have the appropriate dot products, the only problem being that $d, v_1, \ldots, v_n \in \mathbb{R}^{n+1}$).

Note that $x_i = (d^T v_i)^2$ for all i. We have to show that x satisfies all the inequalities above, so let $(c, u_1, \ldots, u_n)$ be any orthonormal representation of G. We have to show that $\sum (c^T u_i)^2 x_i \leq 1$. So, using the fact that $(u \otimes v)^T (\overline{u} \otimes \overline{v}) = u^T \overline{u} v^T \overline{v}$, we have

$$\begin{aligned} \sum (c^T u_i)^2 x_i &= \sum (c^T u_i)^2 (d^T v_i)^2 \\ &= \sum \left[(c \otimes d)^T (u_i \otimes v_i) \right]^2. \end{aligned}$$

Then, we have by the same fact that $||c \otimes d|| = ||c||||d|| = 1$. Similarly, $(u_i \otimes v_i)^T(u_i \otimes v_i) = 1$, whereas for $i \neq j$, $(u_i \otimes v_i)^T(u_j \otimes v_j) = (u_i^T u_j)(v_i^T v_j)$ is equal to 0, since ij is either an edge in G or an edge in $\overline{G}$, so that one or the other of $(u_i^T u_j), (v_i^T v_j)$ is 0.

Therefore, we have that $\sum \left[(c \otimes d)^T(u_i \otimes v_i)\right]^2$ is the first several terms of the length squared of $(c \otimes d)$ measured by an orthonormal basis including the $u_i \otimes v_i$, and therefore the entire sum is at most $||c \otimes d||^2 = 1$. Overall, we get that $\sum (c^T u_i) x_i \leq 1$ for all orthonormal representations. □

3 Generalization of this approach to general 0-1 programming

Let P^I be the convex hull of an integer programming problem, i.e, $P^I = \text{conv}(\{x \in \{0,1\}^n \mid Ax \leq b\})$. We have $P^I \subseteq P := \{x \mid Ax \leq b, 0 \leq x \leq e\}$, its linear programming relaxation.

We want to get a *tighter* relaxation of P^I that is still tractable. The main idea is called "lift-and-project" — we'll get our relaxation as a *projection* of a simpler higher-dimensional object.

As an example, consider the unit ball in the 1-norm: $\{x \in \mathbb{R}^n \mid ||x||_1 \leq 1\}$. This has 2^n facets, so requires exponentially many constraints to describe as the feasible set of a linear programming problem. However, this set can also be written as

$$\{y - z \mid y, z \in \mathbb{R}^n_+, e^T y + e^T z \leq 1\}$$

which has $2n + 1$ inequalities in $2n$ dimensional space.

So now, given the general problem above, we'll get a tighter relaxation of P^I by "lifting" to $\mathbb{M}^{1+n}$ and get constraints by taking any constraint $c^T x - \delta \geq 0$ from the description of P and multiplying it by $x_j \geq 0$ or by $1 - x_j \geq 0$ for all j. These can be viewed as *linear* constraints in $\begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix}$.

Let $M(P)$ be the set of matrices $\overline{X}$ in $\mathbb{M}^{1+n}$ satisfying all these inequalities as well as $\overline{x}_{00} = 1$ and $\overline{X}\overline{e}_0 = \text{diag}\,(\overline{X})$.

Then, define $N(P) = \left\{ x \in \mathbb{R}^n_+ \mid \begin{pmatrix} 1 \\ x \end{pmatrix} = \overline{X}\overline{e}_0 \text{ for some } X \in M(P) \right\}$. This is our projection back into $\mathbb{R}^n$. The claim is that the following holds:

$$P^I \subseteq N(P) \subseteq P.$$

Furthermore, we can iterate this process to get $N^2(P), N^3(P), \ldots$. In fact we have the following theorem.

Theorem 3 *(Lovász-Schrijver)* $N^n(P) = P^I$.

Similarly, we can define $M_+(P)$ as the set of $\mathbb{M}^{1+n}_+$ that satisfy the same constraints, and $N_+(P)$ analogously, and we get

$$P^I \subseteq N_+(P) \subseteq N(P) \subseteq P.$$

For any fixed r, optimizing a linear function over $N^r(P)$ ($N_+^r(P)$) is a linear (semidefinite) programming problem of polynomial size, but this grows fast with r. Luckily, the relaxations are quite good even for small r.

Bringing this back to the stable set problem, we have $N_+(\mathrm{FRAC}(G)) \subseteq \mathrm{TH}(G)$, where $\mathrm{FRAC}(G)$ is the natural LP relaxation of the stable set polytope with constraints $x_i + x_j \leq 1$ for all $ij \in E$, $0 \leq x \leq e$.