Extended Formulations for Combinatorial Optimization

Samuel Fiorini

Université libre de Bruxelles (ULB, Brussels)

ISCO 2016 Summer School
Day 1 (afternoon)

The PSD Cone

Definition (PSD matrix)

A $k \times k$ matrix A is *positive semidefinite* (or *PSD*) if

$$\forall x \in \mathbb{R}^k : x^\intercal A x \geqslant 0$$

We consider only symmetric matrices (A is PSD $\iff$ $A + A^\intercal$ is PSD)

- $\mathbb{S}^k :=$ space of all $k \times k$ (real) symmetric matrices
- *Frobenius inner product*: for $A, B \in \mathbb{S}^k$

$$\langle A, B \rangle := \sum_{i,j} A_{ij} B_{ij} = \mathsf{Tr}(AB^\intercal) = \mathsf{Tr}(AB)$$

Remark

$$\langle A, xx^\intercal \rangle = \sum_{i,j} A_{ij} x_i x_j = x^\intercal A x$$

Definition (PSD cone)

$\mathbb{S}^k_+ := \{A \in \mathbb{S}^k \mid A \text{ is PSD}\}$

For $A, B \in \mathbb{S}^k$, we write $A \succcurlyeq B$ whenever $A - B \in \mathbb{S}^k_+$. So

$$A \succcurlyeq 0 \quad \text{simply means:} \quad \text{"}A \text{ is PSD"}$$

Theorem

The following are equivalent for $A \in \mathbb{S}^k$:

(i) $A \in \mathbb{S}^k_+$

(ii) *all the eigenvalues of A are nonnegative*

(iii) $A = BB^\intercal$ *for some* $B \in \mathbb{R}^{k \times \ell}$

(iv) A *is a sum of matrices of the form* $xx^\intercal$ *for* $x \in \mathbb{R}^k$

(v) $\langle A, C \rangle \geqslant 0$ *for all* $C \in \mathbb{S}^k_+$

Proof. (i) $\implies$ (ii): Let $v \in \mathbb{R}^k \smallsetminus \{0\}$ be an eigenvector of A with eigenvalue λ. We have $0 \leqslant v^\intercal A v = v^\intercal \lambda v = \lambda v^\intercal v$, so that $\lambda \geqslant 0$.

(ii) $\implies$ (iii): We have $A = M^{-1} D M = M^\intercal D M$ where M is a $k \times k$ orthogonal matrix and $D = \mathrm{Diag}(\lambda_1, \dots, \lambda_k)$ is a diagonal matrix with $\lambda_i \geqslant 0$ for all i. Letting $\sqrt{D} := \mathrm{Diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_k})$, we see that $A = M^\intercal \sqrt{D}\sqrt{D} M = (M^\intercal \sqrt{D})(M^\intercal \sqrt{D})^\intercal = BB^\intercal$ where $B := M^\intercal \sqrt{D}$.

(iii) $\implies$ (iv): For all $i, j \in [k]$ we have $A_{ij} = \sum_{t=1}^{\ell} B_{it} B_{jt}$. Hence $A = \sum_{t=1}^{\ell} B_{\star t} (B_{\star t})^\intercal$, where $B_{\star t} \in \mathbb{R}^k$ denotes the tth column of B. $\square$

Exercise

Prove (iv) $\implies$ (v) and (v) $\implies$ (i)

Proof (end). (iv) $\implies$ (v). By hypothesis, there exists a finite set $X \subseteq \mathbb{R}^k$ such that $A = \sum_{x \in X} xx^\intercal$. Thus

$$\langle A, C \rangle = \langle \sum_{x \in X} xx^\intercal, C \rangle = \sum_{x \in X} \langle xx^\intercal, C \rangle = \sum_{x \in X} \underbrace{x^\intercal C x}_{\geqslant 0} \geqslant 0$$

(v) $\implies$ (i). In particular, we have $\langle A, xx^\intercal \rangle \geqslant 0$ for all $x \in \mathbb{R}^k$ because $xx^\intercal \in \mathbb{S}^k_+$ for all $x \in \mathbb{R}^k$. So we get $x^\intercal A x = \langle A, xx^\intercal \rangle \geqslant 0$ for all $x \in \mathbb{R}^k$, that is, $A \in \mathbb{S}^k_+$. $\square$

The PSD Cone

Semidefinite Programs and Their Duals

Let $C, A_1, \ldots, A_m \in \mathbb{S}^k$, $b_1, \ldots, b_m \in \mathbb{R}$

SDP:
$$\begin{aligned} \min \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \langle A_i, X \rangle = b_i \quad \forall i \\ & X \succcurlyeq 0 \end{aligned}$$

Dual-SDP:
$$\begin{aligned} \max \quad & \textstyle\sum_i b_i y_i \\ \text{s.t.} \quad & C - \textstyle\sum_i y_i A_i \succcurlyeq 0 \end{aligned}$$

Exercise

Check that weak duality holds.

Remark

Strong duality holds when one of the two programs has a strictly feasible (positive definite) solution, but does not always hold.

Let G be a n-vertex graph. Often: $V(G) = [n] := \{1, \ldots, n\}$.

Definition (stable sets, cliques)

- A *stable set* of G is a set of mutually nonadjacent vertices.
- A *clique* of G is a set of mutually adjacent vertices.

Definition (stable set polytope of G)

$$\mathsf{STAB}(G) := \mathrm{conv}\{\chi^S \in \{0,1\}^n \mid S \text{ stable set of } G\}$$

Basic valid inequalities for $\mathsf{STAB}(G)$:

- *nonnegativity inequality*: $x_i \geqslant 0$ for $i \in V(G)$
- *edge inequality*: $x_i + x_j \leqslant 1$ for $ij \in E(G)$
- *clique inequality*: $x(K) \leqslant 1$ for K clique of G

where $x(K) := \sum_{i \in K} x_i$

Definition (clique relaxation)

$$\mathsf{QSTAB}(G) := \{x \in \mathbb{R}^n_+ \mid \forall K \text{ clique of } G : x(K) \leqslant 1\}$$

Since nonnegativity, clique inequalities are valid for $\mathsf{STAB}(G)$:

$$\mathsf{STAB}(G) \subseteq \mathsf{QSTAB}(G)$$

In general, $\mathsf{STAB}(G) \subsetneq \mathsf{QSTAB}(G)$. For instance take $G = C_5$.

If $x := \chi^S$ where S is a stable set, then the matrix

$$X := \begin{pmatrix} 1 \\ x \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix}^{\mathsf{T}}$$

has the following properties:

- $X \in \mathbb{S}_+^{n+1}$
- $X_{00} = 1$
- $X_{ii} = X_{0i} = x_i$ for $i \in [n]$
- $X_{ij} = 0$ for $ij \in E(G)$

Definition (theta body of G, Grötschel-Lovász-Schrijver)

$$\mathsf{TH}(G) := \{x \in \mathbb{R}^n \mid \begin{array}{l} \exists X \in \mathbb{S}_+^{n+1} : \\ X_{00} = 1 \\ \forall i \in V(G) : X_{ii} = X_{0i} = x_i \\ \forall ij \in E(G) : X_{ij} = 0\} \end{array}$$

- $\mathsf{TH}(G)$ is a closed convex set (a projected spectrahedron)
- $\mathsf{TH}(G)$ given via a *semidefinite ext. formulation* of *size* $n+1$

Let:

- $x \in \mathsf{TH}(G)$
- $X \in \mathbb{S}^{n+1}$ be any PSD matrix witnessing that $x \in \mathsf{TH}(G)$
- K clique and $y := \chi^K$

Since X is PSD, we have

$$\begin{aligned}\begin{pmatrix}-1\\ y\end{pmatrix}^{\mathsf{T}} X \begin{pmatrix}-1\\ y\end{pmatrix} \geqslant 0 &\iff 1 - 2\sum_{i \in K} x_i + \sum_{i \in K} x_i \geqslant 0\\ &\iff \sum_{i \in K} x_i \leqslant 1\\ &\iff x(K) \leqslant 1\end{aligned}$$

Theorem (Grötschel-Lovász-Schrijver)

STAB$(G) \subseteq$ *TH*$(G) \subseteq$ *QSTAB*(G)

Definition (perfect graph)

A graph G is said to be *perfect* if $\chi(H) = \omega(H)$ for every induced subgraph H of G, where $\chi(H)$ denotes the chromatic number of H, and $\omega(H)$ the clique number of H.

Theorem (Grötschel-Lovász-Schrijver, Chvátal)

The following are equivalent for a graph G:

- **(i)** *G is perfect*
- **(ii)** *$STAB(G) = QSTAB(G)$*
- **(iii)** *$TH(G)$ is a polytope*
- **(iv)** *the complement $\overline{G}$ of G is perfect*

If G is perfect then $\alpha(G) = \max\{\sum_{i \in V(G)} x_i \mid x \in \mathrm{TH}(G)\}$.

Yields the only polytime algorithm we know for computing α on perfect graphs!

For x vertex of STAB(G), and K clique of G:

$$\left(1 - \sum_{i \in K} x_i\right)^2 \geqslant 0$$

Using $x_i^2 = x_i$ for $i \in V(G)$ and $x_i x_j = 0$ for $ij \in E(G)$:

$$\begin{aligned}
0 &\leqslant 1 - 2\sum_{i \in K} x_i + \left(\sum_{i \in K} x_i\right)^2 \\
&= 1 - 2\sum_{i \in K} x_i + \sum_{i,j \in K} \underbrace{x_i x_j}_{=0 \text{ if } i \neq j} \\
&= 1 - 2\sum_{i \in K} x_i + \sum_{i \in K} \underbrace{x_i^2}_{=x_i} \\
&= 1 - \sum_{i \in K} x_i
\end{aligned}$$

Let I_G denote the ideal of $\mathbb{R}[x_1, \ldots, x_n]$ generated by the polynomials $x_i^2 - x_i$ for $i \in V(G)$ and $x_i x_j$ for $ij \in E(G)$.

We have just shown

$$\left(1 - \sum_{i \in K} x_i\right) \equiv \left(1 - \sum_{i \in K} x_i\right)^2 \pmod{I_G}$$

More generally, suppose that $f(x)$ is an affine function (degree $\leqslant 1$ polynomial) such that

$$f(x) \equiv \sum_{t=1}^{r} (g_t(x))^2 \pmod{I_G}$$

where $g_t(x) \in \mathbb{R}[x_1, \ldots, x_n]$. Then $f(x) \geqslant 0$ is valid for STAB(G)!

If $\deg(g_t(x)) \leqslant k$ we say that *$f(x)$ is SOS$(1, k)$ modulo I_G*

Theorem (Lovász-Schrijver)

For every affine function, $f(x) \geqslant 0$ is valid for TH(G) if and only if $f(x)$ is SOS$(1,1)$ modulo I_G.

Proof.($\Longrightarrow$) Suppose $c_0 + \sum_{i=1}^{n} c_i x_i \geqslant 0$ is valid for TH(G)

For $C :=$ Diag$(c_0, c_1, \ldots, c_n)$, have that $\langle C, X \rangle \geqslant 0$ is valid for

$$\{X \in \mathbb{S}_+^n \mid \forall i \in V(G) : X_{ii} = X_{0i},\ \forall ij \in E(G) : X_{ij} = 0\}$$

Write this spectrahedron as

$$\{X \in \mathbb{S}_+^n \mid \forall i \in V(G) : \langle A_i, X \rangle = 0,\ \forall ij \in E(G) : \langle B_{ij}, X \rangle = 0\}$$

where $A_i := E_{i0} + E_{0i} - 2E_{ii}$ for $i \in V(G)$ and $B_{ij} := E_{ij} + E_{ji}$ for $ij \in E(G)$.

By SDP duality, there exist $y_i \in \mathbb{R}$ for $i \in V(G)$ and $z_{ij} \in \mathbb{R}$ for $ij \in E(G)$ such that

$$C - \sum_{i \in V(G)} y_i A_i - \sum_{ij \in E(G)} z_{ij} B_{ij} \succcurlyeq 0$$

So we can write

$$C - \sum_{i \in V(G)} y_i A_i - \sum_{ij \in E(G)} z_{ij} B_{ij} = \sum_{t=1}^{r} \begin{pmatrix} c_{t,0} \\ c_{t,1} \\ \vdots \\ c_{t,n} \end{pmatrix} \begin{pmatrix} c_{t,0} \\ c_{t,1} \\ \vdots \\ c_{t,n} \end{pmatrix}^{\mathsf{T}}$$

for some vectors $c_1, \ldots, c_r \in \mathbb{R}^{n+1}$. Each of these vectors yields a polynomial $g_t(x) := \sum_{i=0}^{n} c_{t,i} x_i$.

Now let's compute $\sum_{t=1}^{r}(g_t(x))^2$ modulo I_G. We have

$$(g_t(x))^2 = \sum_{i,j} c_{t,i}c_{t,j}x_i x_j = \left\langle \begin{pmatrix} c_{t,0} \\ c_{t,1} \\ \vdots \\ c_{t,n} \end{pmatrix} \begin{pmatrix} c_{t,0} \\ c_{t,1} \\ \vdots \\ c_{t,n} \end{pmatrix}^{\mathsf{T}}, \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}^{\mathsf{T}} \right\rangle$$

so that,

$$\begin{aligned} \sum_{t=1}^{r}(g_t(x))^2 &= \left\langle C - \sum_{i \in V(G)} y_i A_i - \sum_{ij \in E(G)} z_{ij} B_{ij}, \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}^{\mathsf{T}} \right\rangle \\ &\equiv \left\langle C, \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}^{\mathsf{T}} \right\rangle \pmod{I_G} \end{aligned}$$

$\square$

- $V \subseteq \mathbb{R}^n$ any finite set
- $I := \mathcal{I}(V)$ set of polynomials vanishing on V (*vanishing ideal* of V): $I := \{g \in \mathbb{R}[x_1, \ldots, x_n] \mid \forall x \in V : g(x) = 0\}$

Definition (SOS$(1,k)$ modulo I)

An affine function $f(x)$ is $SOS(1,k)$ *modulo* I if there are polynomials $g_t(x)$ ($t = 1, \ldots, r$) of degree $\leqslant k$ such that:

$$f(x) \equiv \sum_{t=1}^{r} (g_t(x))^2 \pmod{I}.$$

This is equivalent to asking that for evaluations of f,

$$\forall x \in V : f(x) = \sum_{t=1}^{r} (g_t(x))^2 .$$

Definition (Theta Body)

The k*-th theta body* $\mathsf{TH}_k(I)$ of V is the convex set defined by the inequalities $f(x) \geqslant 0$ where f is $\mathsf{SOS}(1,k)$ modulo I.

Theorem (Gouveia-Parrilo-Thomas)

For $V \subseteq \{0,1\}^n$,

$$\mathit{conv}(V) = \mathit{TH}_n(I) \subseteq \cdots \subseteq \mathit{TH}_2(I) \subseteq \mathit{TH}_1(I)\,.$$

Proof. It is clear that $\mathsf{TH}_{r+1}(I) \subseteq \mathsf{TH}_r(I)$.

Every function $f(x)$ that is nonnegative over $V \subseteq \{0,1\}^n$ can be expressed as sum of squares of degree $\leqslant n$ monomials of the form $\sqrt{f(a)}\prod_{i:a_i=1} x_i \prod_{i:a_i=1}(1-x_i)$ for $a \in V$. $\square$

Exercise

Prove that if G is a graph with stability number $\alpha(G)$ (that is, the maximum size of a stable set is $\alpha(G)$), then already $\mathsf{TH}_{\alpha(G)} = \mathsf{STAB}(G)$. In other words, the set of characteristic vectors of stable sets of G has theta rank at most $\alpha(G)$.

Theorem (Gouveia-Parrilo-Thomas)

The k-th theta body of $V \subseteq \{0,1\}^n$ has a size-$n^{O(k)}$ semidefinite extended formulation. When $V \subseteq \{0,1\}^n$ is down-monotone, the k-th theta body can be described as:

$$
\begin{aligned}
TH_k(I) = \{x \in \mathbb{R}^n \mid \quad & \exists M \in \mathbb{S}_+^{V_k \times V_k} \ s.t. \\
& M_{\varnothing\varnothing} = 1 \\
& \forall i \in [n] : M_{\varnothing\{i\}} = M_{\{i\}\varnothing} = M_{\{i\}\{i\}} = x_i \\
& \forall a, a' \in V_k \ s.t. \ a \cup a' \notin V_k : M_{aa'} = 0 \\
& \forall a, b, a', b' \in V_k \ s.t. \ a \cup a' = b \cup b' : M_{aa'} = M_{bb'} \}
\end{aligned}
$$

where V_k denotes the set of feasible solutions supported on at most k elements.

Definition (theta rank)

theta-rank of V is min k such that $\operatorname{conv}(V) = \mathrm{TH}_k(I)$.

Theorem (Gouveia-Parrilo-Thomas)

Let $V \subseteq \mathbb{R}^n$ be a finite set. Suppose that conv(V) can be defined by linear inequalities of the form $f(x) \geqslant 0$, where $f(x)$ takes at most $k+1$ different values on V. Then the theta-rank of V is at most k.

Proof. Consider any affine function $f(x)$ that takes at most $k+1$ different values on V, say $0 = y_0, y_1, \ldots, y_k \in \mathbb{R}_+$. Let $g(z) \in \mathbb{R}[z]$ be any (univariate) degree $\leqslant k$ polynomial such that $g(y_i) = \sqrt{y_i}$ for all i. Then we have

$$f(x) = \sum_{i=0}^{k} (g(f(x)))^2$$

each $g(f(x))$ being a degree $\leqslant k$ polynomial. □

Theorem (Gouveia-Parrilo-Thomas)

A finite set $V \subseteq \mathbb{R}^n$ has theta rank 1 if and only if conv(V) can be expressed by linear inequalities of the form $f(x) \geqslant 0$, where $f(x)$ takes at most 2 different values on V.

Definition

A polytope P is said to be 2-*level* if each of its facet-defining hyperplanes H has a parallel hyperplane H' such that every vertex of P is either on H or on H'.

Remark

A polytope P is 2-level iff its vertex set $V :=$ vert(P) has theta-rank 1.

Exercise

STAB(G) is 2-level iff G is perfect. Find another family of examples of 2-level polytopes. Find all of them in dimension 3 (there are 5).

We know many examples of 2-level polytopes, some properties but do not know (research questions!):

- **what** these polytopes are exactly
- **how many** there are in dimension n (up to isomorphism, asymptotically)
- **what** is their linear extension complexity

Some data:

n	3	4	5	6	7
#(n-dim 2L-polytopes)	5	19	106	1150	27291

As a comparison (both counts up to isomorphism):

n	3	4	5	6	7
#(n-vertex graphs)	4	11	34	156	1044

The Goemans-Williamson Relaxation of Max-Cut

Definition (Max-Cut)

Given a graph G and a *nonnegative* weight w_{ij} for each edge $ij \in E(G)$, find a partition of $V(G) = V_+ \cup V_-$ into two disjoint sets such that the total weight $\sum_{i \in V_+, j \in V_-} w_{ij}$ across the corresponding cut is maximized.

Max-Cut can be formulated as a quadratic binary program:

$$
\begin{aligned}
\max \quad & \frac{1}{2} \sum_{ij \in E(G)} w_{ij}(1 - x_i x_j) \\
\text{s.t.} \quad & x_i \in \{-1, +1\} \quad \forall i \in V(G)
\end{aligned}
$$

By considering the rank-1 matrix $X := xx^\intercal$, we are led to the following relaxation of Max-Cut, defined by a size-n semidefinite extended formulation:

$$\begin{aligned} \max \quad & \frac{1}{2} \sum_{ij \in E(G)} w_{ij}(1 - X_{ij}) \\ \text{s.t.} \quad & X_{ii} = 1 \quad \forall i \in V(G) \\ & X \in \mathbb{S}^n_+ \end{aligned}$$

Theorem (Goemans-Williamson)

The above relaxation gives a 0.878-approximation of Max-Cut in the sense that

$$\mathit{OPT} \geqslant 0.878 \cdot \mathit{SDP}$$

Proof. Let X denote an optimal solution of the relaxation. Since $X \succcurlyeq 0$, we can find a vector u_i for each $i \in V(G)$ in some $\mathbb{R}^d$ such that $u_i^\intercal u_j = X_{ij}$ for each $i, j \in V(G)$. Because $X_{ii} = 1$, the u_i's are unit vectors.

Take a random vector r uniformly on the unit sphere in $\mathbb{R}^d$, and let

$$x_i := \mathsf{sign}(r^\intercal u_i) \in \{-1, 1\}$$

This is called *hyperplane rounding*.

If denote by W the random variable giving the value of the random cut produced, can decompose:

$$W = \sum_{ij \in E(G)} w_{ij} I_{[\text{edge } ij \text{ is cut}]}$$

Finally get:

$$\begin{aligned}
E[W] &= \sum_{ij \in E(G)} w_{ij} P[\text{edge } ij \text{ is cut}] \\
&= \sum_{ij \in E(G)} w_{ij} \frac{1}{\pi} \arccos(X_{ij}) \\
&\geqslant 0.878 \cdot \sum_{ij \in E(G)} w_{ij} \frac{1}{2}(1 - X_{ij}) \\
&\geqslant 0.878 \cdot \mathsf{SDP}
\end{aligned}$$

for the first inequality we use:

$$\forall t \in [-1, 1] : \frac{1}{\pi} \arccos(t) \geqslant 0.878 \cdot \frac{1}{2}(1 - t)$$

$\square$

$$P := \{x \in \mathbb{R}^n \mid Ax - b \geqslant 0\} \subseteq [0,1]^n$$

Definition (Sherali-Adams)

Let $0 \leqslant r \leqslant n$. The round-$r$ *Sherali-Adams* relaxation $\mathrm{SA}_r(P)$ is obtained by:

(i) multiplying each inequality of $Ax - b \geqslant 0$ by the monomial $\prod_{i \in I_0}(1 - x_i) \prod_{i \in I_1} x_i$ for each pair I_0, I_1 of disjoint subsets of $[n]$ with $|I_0| + |I_1| = r$;

(ii) expanding using $x_i^2 = x_i$;

(iii) replacing each monomial $\prod_{i \in S} x_i$ by a variable y_S where $S \subseteq [n]$, $|S| \leqslant r + 1$;

(iv) adding the constraint $y_\varnothing = 1$;

(v) projecting back to the x-variables by letting $x_i := y_{\{i\}}$.

Exercise

Find $\mathrm{SA}_1(P)$ for $P := \{x \in \mathbb{R}^2 \mid x_1 \geqslant 0,\ x_2 \geqslant 0,\ -2x_1 - 2x_2 + 1 \geqslant 0\}$.

Notice that $P_I = \{(0,0)\}$. After step (i), we obtain:

$$\left\{\begin{array}{l} x_1 x_1 \geqslant 0 \\ x_1(1-x_1) \geqslant 0 \\ x_1 x_2 \geqslant 0 \\ x_1(1-x_2) \geqslant 0 \\ x_2 x_1 \geqslant 0 \\ x_2(1-x_1) \geqslant 0 \\ x_2 x_2 \geqslant 0 \\ x_2(1-x_2) \geqslant 0 \\ (-2x_1 - 2x_2 + 1)x_1 \geqslant 0 \\ (-2x_1 - 2x_2 + 1)(1-x_1) \geqslant 0 \\ (-2x_1 - 2x_2 + 1)x_2 \geqslant 0 \\ (-2x_1 - 2x_2 + 1)(1-x_2) \geqslant 0 \end{array}\right. .$$

After steps (ii)–(iv), we obtain:

$$\begin{cases} y_\varnothing = 1 \\ y_{\{1\}} \geqslant 0 \\ 0 \geqslant 0 \\ y_{\{1,2\}} \geqslant 0 \\ y_{\{1\}} - y_{\{1,2\}} \geqslant 0 \\ y_{\{1,2\}} \geqslant 0 \\ y_{\{2\}} - y_{\{1,2\}} \geqslant 0 \\ y_{\{2\}} \geqslant 0 \\ 0 \geqslant 0 \\ -y_{\{1\}} - 2y_{\{1,2\}} \geqslant 0 \\ y_\varnothing - y_{\{1\}} - 2y_{\{2\}} + 2y_{\{1,2\}} \geqslant 0 \\ -y_{\{2\}} - 2y_{\{1,2\}} \geqslant 0 \\ y_\varnothing - 2y_{\{1\}} - y_{\{2\}} + 2y_{\{1,2\}} \geqslant 0 \end{cases} .$$

From these constraints, we derive $y_{\{1\}} \geqslant y_{\{1,2\}} \geqslant 0$ and at the same time $y_{\{1\}} \leqslant -2y_{\{1,2\}}$, which implies that $y_{\{1\}} = y_{\{1,2\}} = 0$. Similarly, we obtain $y_{\{2\}} = 0$. Thus the round-1 Sherali-Adams relaxation is exact.

Sherali-Adams is a hierarchy of increasingly stronger relaxations of the *integer hull* $P_I := \text{conv}(P \cap \mathbb{Z}^n)$, obtained from the constraints defining P.

Theorem

$P_I \subseteq SA_n(P) \subseteq \cdots \subseteq SA_1(P) \subseteq SA_0(P) = P$

Definition (linearization map)

$L : \mathbb{R}[x_1, \ldots, x_n] \to \mathbb{R}_{\leqslant 1}[y_S : S \subseteq [n]]$ linear map such that $L(x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}) := y_{\{i | a_i \neq 0\}}$ for $a \in \mathbb{Z}_+^n$

Remark

Every inequality defining the round-r Sherali-Adams relaxation if of the form

$$L\left((A_i x - b_i) \prod_{i \in I_0} (1 - x_i) \prod_{i \in I_1} x_i\right) \geqslant 0$$

for some $i \in [m]$ and some disjoint $I_0, I_1 \subseteq [n]$ such that $|I_0| + |I_1| = r$.

Proof. Let $k \in [n] \setminus (I_0 \cup I_1)$. Notice that

$$(A_i x - b_i) \prod_{j \in I_0 \cup \{k\}} (1 - x_j) \prod_{j \in I_1} x_j + (A_i x - b_i) \prod_{j \in I_0} (1 - x_j) \prod_{j \in I_1 \cup \{k\}} x_j$$
$$= (A_i x - b_i) \prod_{j \in I_0} (1 - x_j) \prod_{j \in I_1} x_j \,.$$

Together with linearity of the linearization map, this implies that $\mathsf{SA}_{r+1}(P) \subseteq \mathsf{SA}_r(P)$ since the constraints defining $\mathsf{SA}_{r+1}(P)$ (before projection) imply those defining $\mathsf{SA}_r(P)$ (before projection). $\square$

Lemma

Let $u^\intercal(Ax-b) \geqslant 0$ *where* $u \in \mathbb{R}^m_+$ *be any inequality that can be derived by taking some nonnegative combination of the inequalities of* $Ax-b \geqslant 0$*, and let* I_0 *and* I_1 *be two disjoint subsets of* $[n]$ *such that* $|I_0|+|I_1|=r$*. Then the linearization of*

$$u^\intercal(Ax-b)\prod_{j\in I_0}(1-x_j)\prod_{j\in I_1}x_j \geqslant 0$$

is a nonnegative combination of the linearizations of

$$(A_i x-b_i)\prod_{j\in I_0}(1-x_j)\prod_{j\in I_1}x_j \geqslant 0$$

for $i \in [m]$*.*

The Sherali-Adams Hierarchy

Consistent Local Distributions

Now, suppose $P = \{x \in \mathbb{R}^n \mid \forall i \in [n] : 0 \leqslant x_i \leqslant 1\}$. Then $\mathsf{SA}_r(P)$ is defined by the following extended formulation:

$$y_\varnothing = 1$$

$$\sum_{I_1 \subseteq S \subseteq I_0 \cup I_1} (-1)^{|S \cap I_0|} y_S \geqslant 0 \quad \forall \text{disjoint } I_0, I_1 \subseteq [n] \text{ s.t. } |I_0 \cup I_1| = r+1$$

Goal: interpret y_S variables as probabilities defining "consistent local distributions"

Definition

A family of *local distributions* $(\mathcal{D}(I))_{I \subseteq [n] : |I| \leqslant r+1}$ where each $\mathcal{D}(I)$ provides a distribution on $x \in \{0,1\}^I$ is said to be *consistent* if for each $I' \subseteq I$, the marginal distribution induced by $\mathcal{D}(I)$ on $\{0,1\}^{I'}$ coincides with the distribution $\mathcal{D}(I')$.

Lemma

A vector $(y_S)_{S \subseteq [n] : |S| \leqslant r+1}$ *is feasible for* $\mathsf{SA}_r(P)$ *where* $P = [0,1]^n$ *iff there exist consistent local distributions* $(\mathcal{D}(I))_{I \subseteq [n] : |I| \leqslant r+1}$ *s.t.*

$$P_{x \sim \mathcal{D}(I)}\left[\bigwedge_{i \in S} x_i = 1\right] = y_S$$

for each $S \subseteq I \subseteq [n]$ *with* $|I| \leqslant r+1$*.*

Proof. "Only if" part. Suppose that (y_S) is feasible for the r-round SA and define distribution $\mathcal{D}(I)$ by letting

$$\begin{aligned} P_{x \sim \mathcal{D}(I)}[x = a] &= P_{x \sim \mathcal{D}(I)}\left[\bigwedge_{i \in I} x_i = a_i\right] \\ &= P_{x \sim \mathcal{D}(I)}\left[\left\{\bigwedge_{i \in I_0} x_i = 0\right\} \wedge \left\{\bigwedge_{i \in I_1} x_i = 1\right\}\right] \\ &:= \sum_{I_1 \subseteq S \subseteq I_0 \cup I_1} (-1)^{|S \cap I_0|} y_S \geqslant 0 \end{aligned}$$

for each $I \subseteq [n]$ with $|I| \leqslant r+1$, and each $a \in \{0,1\}^I$, where $I_0 := \{i \in I \mid a_i = 0\}$ and $I_1 := \{i \in I \mid a_i = 1\}$.

Why do the probabilities sum up to 1? Consider two sets $I' \subseteq I \subseteq [n]$ with $|I| \leqslant r+1$ such that $I = I' \cup \{j\}$ for some $j \in [n]$. Then, for $a' \in \{0,1\}^{I'}$,

$$\begin{aligned} P_{x \sim \mathcal{D}(I')}\left[\bigwedge_{i \in I'} x_i = a'_i\right] &= \sum_{I'_1 \subseteq S \subseteq I'} (-1)^{|S \cap I'_0|} y_S \\ &= \sum_{I'_1 \subseteq S \subseteq I} (-1)^{|S \cap (I'_0 \cup \{j\})|} y_S + \sum_{I'_1 \cup \{j\} \subseteq S \subseteq I} (-1)^{|S \cap I'_0|} y_S \end{aligned}$$

since the terms y_S with $j \in S$ cancel out in the sum the last expression. So:

$$\begin{aligned} P_{x \sim \mathcal{D}(I')}\left[\bigwedge_{i \in I'} x_i = a'_i\right] &= P_{x \sim \mathcal{D}(I)}\left[\left\{\bigwedge_{i \in I'} x_i = a'_i\right\} \wedge x_j = 0\right] \\ &\quad + P_{x \sim \mathcal{D}(I)}\left[\left\{\bigwedge_{i \in I'} x_i = a'_i\right\} \wedge x_j = 1\right] \end{aligned}$$

We get that

$$\begin{aligned} \sum_{a \in \{0,1\}^I} P_{x \sim \mathcal{D}(I)}[x = a] &= \sum_{a' \in \{0,1\}^{I'}} P_{x \sim \mathcal{D}(I')}[x = a'] = \cdots \\ &= (y_\varnothing - y_{\{k\}}) + y_{\{k\}} = y_\varnothing = 1 \end{aligned}$$

Finally, check consistency. It suffices to consider two sets $I' \subseteq I \subseteq [n]$ with $|I| \leqslant r+1$ such that $I = I' \cup \{j\}$ for some $j \in [n]$. Then, for $a' \in \{0,1\}^{I'}$,

$$\begin{aligned} P_{x\sim\mathcal{D}(I)}\left[\bigwedge_{i\in I'} x_i = a'_i\right] &= P_{x\sim\mathcal{D}(I)}\left[\left\{\bigwedge_{i\in I'} x_i = a'_i\right\} \wedge x_j = 0\right] \\ &\quad + P_{x\sim\mathcal{D}(I)}\left[\left\{\bigwedge_{i\in I'} x_i = a'_i\right\} \wedge x_j = 1\right] \\ &= P_{x\sim\mathcal{D}(I')}\left[\bigwedge_{i\in I'} x_i = a'_i\right]. \end{aligned}$$

Notice that, by consistency,

$$P_{x\sim\mathcal{D}(I)}\left[\bigwedge_{i\in S} x_i = 1\right] = P_{x\sim\mathcal{D}(S)}\left[\bigwedge_{i\in S} x_i = 1\right] = \sum_{S\subseteq T\subseteq S} (-1)^{|S\cap\varnothing|} y_S = y_S$$

whenever $S \subseteq [n]$ and $|S| \leqslant r+1$.

"If" part.

- Start from a consistent family $(\mathcal{D}(A))_{A\subseteq[n]:|A|\leqslant r+1}$ of local distributions
- define $y_S := P_{x\sim\mathcal{D}(I)}\left[\bigwedge_{i\in S} x_i = 1\right]$ for each $S \subseteq [n]$, where I is arbitrary with $S \subseteq I \subseteq [n]$ and $|I| \leqslant r+1$

Then (y_S) is feasible for the r-round Sherali-Adams formulation:

- have $y_\varnothing = 1$
- for $I_0, I_1 \subseteq [n]$ disjoint such that $|I_0| + |I_1| = r+1$,

$$\sum_{I_1\subseteq S\subseteq I_0\cup I_1} (-1)^{|S\cap I_0|} y_S \geqslant 0$$

since, by inclusion-exclusion, the left-hand side is the probability of the event $\left\{\bigwedge_{i\in I_0} x_i = 0\right\} \wedge \left\{\bigwedge_{i\in I_1} x_i = 1\right\}$ in $\mathcal{D}(I_0 \cup I_1)$

$\square$

Definition (local infeasibility)

Consider $I \subseteq [n]$ and $a \in \{0,1\}^I$. Say that *partial solution* $a \in \{0,1\}^I$ is *locally infeasible* if there exists some i such that $A_i x - b_i \geqslant 0$ has support contained in I and is violated by every point $x \in \mathbb{R}^n$ such that $x_i = a_i$ for $i \in I$

Lemma

If a vector $(y_S)_{S \subseteq [n] : |S| \leqslant r+1}$ is feasible for $SA_r(P)$ where $P = \{x \in \mathbb{R}^n \mid Ax \geqslant b\}$, then there exist consistent local distributions $(\mathcal{D}(I))_{I \subseteq [n] : |I| \leqslant r+1}$ such that

$$P_{x \sim \mathcal{D}(I)} \left[\bigwedge_{i \in S} x_i = 1 \right] = y_S$$

for each $S \subseteq I \subseteq [n]$ with $|I| \leqslant r+1$. Moreover, if $I \subseteq [n]$ and $|I| \leqslant r$ and $a \in \{0,1\}^I$ is locally infeasible, then $P_{x \sim \mathcal{D}(I)}[x = a] = 0$.

Proof. First part : **OK**

Second part. Suppose that $\sum_{i\in I_0} c_i(1-x_i)+\sum_{i\in I_1} c_i x_i-\delta \geqslant 0$ locally cuts $a\in\{0,1\}^I$. Then from the round-r Sherali-Adams relaxation we can infer the inequality

$$\begin{aligned}
0 &\leqslant L\left(\left(\sum_{i\in I_0} c_i(1-x_i)+\sum_{i\in I_1} c_i x_i-\delta\right)\prod_{i\in I_0}(1-x_i)\prod_{i\in I_1} x_i\right)\\
&= L\left(\left(\sum_{i\in I} c_i-\delta\right)\prod_{i\in I_0}(1-x_i)\prod_{i\in I_1} x_i\right)\\
&= \underbrace{\left(\sum_{i\in I} c_i-\delta\right)}_{<0}\cdot \underbrace{L\left(\prod_{i\in I_0}(1-x_i)\prod_{i\in I_1} x_i\right)}_{\geqslant 0}
\end{aligned}$$

which implies

$$L\left(\prod_{i\in I_0}(1-x_i)\prod_{i\in I_1} x_i\right)=0 \iff \sum_{I_1\subseteq S\subseteq I_0\cup I_1} (-1)^{|S\cap I_0|} y_S = 0\,.$$

□

A nice **corollary**:

Theorem

For every polytope $P \subseteq [0,1]^n$ we have $SA_n(P) = P_I$.

Partial converse to previous lemma:

Lemma

Suppose that each inequality in $Ax \geqslant b$ is supported on at most s coordinates. Then every consistent collection of local distributions $(\mathcal{D}(I))_{I \subseteq [n] : |I| \leqslant r+s+1}$ such that the probability of sampling locally infeasible partial solutions is 0 yields a feasible solution $(y_S)_{S \subseteq [n] : |S| \leqslant r+1}$ of the round-r Sherali-Adams relaxation of $P = \{x \in \mathbb{R}^n \mid Ax \geqslant b\}$.

Definition (Tree-decompositions, tree-width)

Let G be a graph, T a tree, and let $\mathcal{B} = (B_t)_{t \in T}$ be a family of vertex sets $B_t \subseteq V(G)$ indexed by the vertices t of T, called the *bags*. The pair $(T, \mathcal{B})$ is called a *tree-decomposition* of G if:

(i) $V(G) = \bigcup_{t \in T} B_t$;

(ii) for every edge $e \in E(G)$ there exists a $t \in T$ such that both ends of e lie in B_t;

(iii) $B_{t_1} \cap B_{t_3} \subseteq B_{t_2}$ whenever $t_1, t_2, t_3 \in V(T)$ are such that t_2 lies on the unique t_1–t_3 path in T.

The *width* of a tree-decomposition $(T, \mathcal{B})$ is defined as $\max_{t \in V(T)} |B_t| - 1$. The *tree-width* of G is the minimum width of a tree-decomposition of G. It is denoted by $\mathsf{tw}(G)$.

Wlog, work only with **nice** tree-decompositions:

Definition (Nice tree-decompositions)

A rooted tree decomposition $(T, \mathcal{B}, r)$ of G is *nice* if for every $u \in V(T)$:

- $|B_u| = 1$ (leaf), or
- u has one child v with $B_u \subseteq B_v$ and $|B_u| = |B_v| - 1$ (forget), or
- u has one child v with $B_v \subseteq B_u$ and $|B_u| = |B_v| + 1$ (introduce), or
- u has two children v and w with $B_u = B_v = B_w$ (join).

Start with the *edge relaxation*

$$\mathsf{ESTAB}(G) := \{x \in \mathbb{R}_+^{V(G)} \mid \forall ij \in E(G) : x_i + x_j \leqslant 1\}$$

of $\mathsf{STAB}(G)$.

Notice that the $\mathsf{ESTAB}(G)_I = \mathsf{STAB}(G)$.

Theorem (Bienstock-Ozbay)

Let G be a graph and let H be a subgraph of G with tw$(H) = k$. Then every inequality $\sum_{i \in V(H)} w_i x_i \leqslant \alpha(H, w)$ that is valid for STAB(H) is valid for SA$_{k+1}$(ESTAB(G)).

Proof. Wlog, assume that $H = G$.

Want to prove: $\mathsf{SA}_{k+1}(\mathsf{ESTAB}(G)) = \mathsf{STAB}(G)$.

Start with a fractional point $x^* \in \mathsf{SA}_{k+1}(\mathsf{ESTAB}(G))$ and consistent local distributions, try to define a single distribution over the stable sets of G that has the right marginals x^*_i for $i \in V(G)$.

Take nice rooted tree-decomposition $(T, \mathcal{B}, r)$ of G with all the bags of size at most k. For $u \in V(T)$, G_u denote the subgraph of G induced on the union of the bags of the descendants of u, so that $G_r = G$. Define a distribution $\mathcal{D}(V(G_u))$ on $\{0,1\}^{V(G_u)}$ for each node u of T, starting with the leaves.

Key equation: for node $u \in V(T)$: Define $P_{x \sim \mathcal{D}(G_u)}[x = a]$ as

$$P_{x \sim \mathcal{D}(B_u)}[\forall i \in B_u : x_i = a_i]$$

$$\cdot \prod_{v \text{ child of } u} P_{x \sim \mathcal{D}(G_v)}[\forall i \in V(G_v) \setminus B_u : x_i = a_i \mid \forall i \in V(G_v) \cap B_u : x_i = a_i]$$

Then check that the marginals are correct. $\square$

Definition (metric polytope)

Let $P \subseteq \mathbb{R}^{\binom{n}{2}}$ be defined by

$$\begin{aligned} x_{ij} + x_{jk} - x_{ik} &\geqslant 0 \quad \forall i,j,k \text{ distinct} \\ 2 - x_{ij} - x_{jk} - x_{ik} &\geqslant 0 \quad \forall i,j,k \text{ distinct} \\ 1 - x_{ij} &\geqslant 0 \quad \forall i,j \text{ distinct} \\ x_{ij} &\geqslant 0 \quad \forall i,j \text{ distinct} \end{aligned}$$

Remark

P_I is the cut-polytope (of K_n)

Know: consistent local distributions of cuts on sets of size $\leqslant 2r+3$ define feasible solutions to $\mathsf{SA}_r(P)$.

Theorem (Charikar, Makarychev, Makarychev)

For "small" r, can find graph G with m edges with

$$\mathit{OPT}(G) \leqslant \left(\frac{1}{2} + \varepsilon\right) m$$

but at the same time

$$\max\left\{ \sum_{ij \in E(G)} x_{ij} \mid x \in \mathsf{SA}_r(P) \right\} \geqslant (1 - \varepsilon) m \,.$$

Can even take $r \approx n^{\delta}$ for some $\delta = \delta(\varepsilon) > 0$

Proof idea. Take G to be a random Δ-regular graph, where $\Delta = \Delta(\varepsilon)$ is large enough.

We need to show:

1. $\mathsf{OPT}(G) \leqslant \left(\frac{1}{2} + \varepsilon\right) m$
2. $\mathsf{SA}_r(P)$ "thinks" that most of the edges of G can be cut

For (1), use fact that almost surely, the eigenvalues of G $\Delta = \lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ of $G(n, \Delta)$ satisfy

$$\max\{|\lambda_2|, |\lambda_n|\} \leqslant 2\sqrt{2\Delta - 1} + o(1)$$

And the GW relaxation.

Exercise

Use this to check that $\mathsf{OPT}(G) \leqslant \left(\frac{1}{2} + \varepsilon\right) m$ for $G = G(n, \Delta)$.

Notice that the GW relaxation can be written

$$\max\left\{\frac{1}{4}\langle L_G, X\rangle \mid X \succcurlyeq 0,\ \forall i : X_{ii} = 1\right\}$$

where $L_G := \Delta I - A_G$ is the Laplacian

Then pick μ carefully so that

$$\mu I - \frac{1}{4} L_G \succcurlyeq 0$$

By weak duality, get

$$0 \leqslant \langle \mu I - \frac{1}{4} L_G, X\rangle = \mu n - \frac{1}{4}\langle L_G, X\rangle$$

so that

$$\mathsf{OPT}(G) \geqslant \frac{1}{4}\langle L_G, X\rangle \geqslant \mu n$$

Here can take: $\mu = \frac{\Delta + 2\sqrt{2\Delta - 1}}{4}$

For (2), use fact that G "locally looks like a forest", so that most edge can be cut in every induced subgraph on at most $k := 2r + 3$ vertices.

Difficulty. Define local distributions of cuts that are consistent.

Idea.

- define *metric* ν on the whole graph, based on vertex distances
- show that can embed ν restricted to every set of size at most k isometrically into the unit sphere of some $\mathbb{R}^d$, in such a way that $\nu(i, j) \geqslant 2 - \varepsilon'$ for *adjacent* vertices i, j
- use embedding to define local distributions by hyperplane rounding, with

  $$x_{ij} = \frac{1}{\pi} \arccos \left(1 - \frac{\nu^2(i,j)}{2} \right)$$

  so that $x_{ij} \geqslant 1 - \varepsilon$ for $ij \in E(G)$.
- consistency is automatic