Semidefinite and Second Order Cone Programming Seminar
Fall 2012
A Comparison of the Sherali-Adams, Lovàsz-Schrijver and Lasserre Relaxations for 0 - 1 Programming

Instructor: Farid Alizadeh
Scribe: Deniz S.Eskandani

12/10/2012

1 Overview

First we introduce three different methods of finding the linear inequality description for the polytope $P := conv(F)$, where $F \subseteq \{0,1\}^n$. A common feature of these methods is the construction of a hierarchy $K \supseteq K^1 \supseteq K^2 \supseteq \cdots \supseteq P$ of relaxations of P which finds the exact convex hull in n steps; that is, $K^n = P$. Then we compare these three methods. At the end we describe applications to the stable set polytope and to the cut polytope.

2 The Lovàsz-Schrijver hierarchy

Let K be a convex body contained in the cube $[0,1]^n$and let

$$P := conv(K \cap \{0,1\}^n)$$

be the $0-1$ polytope to be described. For convenience, define

$$\tilde{K} := \left\{\lambda \begin{pmatrix}1\\x\end{pmatrix} | x \in K, \lambda \geqslant 0\right\}, \tag{1}$$

the homogenization of K. Let $M(K)$ denote the set of symmetric matrices $Y = (y_{ij})_{i,j=0}^n$ satisfying

$$y_{j,j} = y_{0,j} \quad \text{for} \quad j = 1, \cdots, n \tag{2}$$

$$Ye_j, Y(e_0 - e_j) \in \tilde{K} \quad \text{for} \quad j = 1, \cdots, n \tag{3}$$

and define

$$N(K) := \left\{ x \in \mathbb{R}^n | \begin{pmatrix} 1 \\ x \end{pmatrix} = Ye_0 \quad \text{for some} \quad Y \in M(K) \right\},$$

where $e_0, e_1, \cdots, e_n$ denote the standard unit vector in $\mathbb{R}^{n+1}$. Define iteratively $N^1(K) := N(K)$ and, for $t \geq 2, N^t(K) := N(N^{t-1}(K))$. Then,

$$P = N^n(K) \subseteq \cdots \subseteq N^{t+1}(K) \subseteq N^t(K) \subseteq \cdots \subseteq N(K) \subseteq K$$

3 The Sherali-Adams and Lasserre hierarchies

Given $V := \{1, \cdots, n\}$, for $1 \leq t \leq n$, $\mathcal{P}_t(V)$ denotes the collection of subsets of cardinality $\leq t$. The component of a vector $y \in \mathbb{R}^{\mathcal{P}(V)}$ are denoted as y_I or $y(I)$.

Definition 1 (moment matrix) *Given* $y \in \mathbb{R}^{\mathcal{P}(V)}$*, an integer* $1 \leq t \leq n$*, and a subset* $U \subset V$*, define the matrices*

$$M_t(y) := (y(I \cup J))_{|I|,|J| \leq t}, M_U(y) := (y(I \cup J))_{I,J \subset U}. \tag{4}$$

The matrix $M_V(y)$ *is known as the* moment matrix *of y.*

Definition 2 (shift operator) *For* $x, y \in \mathbb{R}^{\mathcal{P}(V)}$*, define the vector* $x * y \in \mathbb{R}^{\mathcal{P}(V)}$ *by*

$$x * y := M_V(y)x; \quad \textit{that is}\ , x * y(I) = \sum_{K \subset V} x_K y_{I \cup K} \quad \text{for} \quad I \subset V. \tag{5}$$

Let

$$K := \{x \in [0, 1]^n | g_l(x) \geq 0 \quad \text{for} \quad l = 1, \cdots, m\} \tag{6}$$

where $g_l, (l = 1, \cdots, m)$, are polynomials in $x_1, \cdots, x_n$ that can be written as

$$\sum_{I \subseteq V} g_l(I) \prod^{i \in I} x_i.$$

Then the same symbol g_l is used for denoting the vector in $\mathbb{R}^{\mathcal{P}(V)}$ with components $g_l(I)(I \subseteq V)$.

3.1 The Sherali-Adams hierarchy

Let w_l denote the degree of the polynomial g_l and set

$$v_l := \left\lceil \frac{w_l}{2} \right\rceil, w := \max w_l, v := \max v_l. \tag{7}$$

and define

$$R_t(K) = \{y \in \mathbb{R}^{\mathcal{P}_{t+w}(V)} | M_U(g_l * y) \succeq 0 \text{ for all } U \subseteq V \text{ with } |U| = t \text{ and } l = 1, \cdots, m \\ M_W(y) \succeq 0 \text{ for all } W \subseteq V \; with \; |W| = \min(t+w, n)\}. \tag{8}$$

Let $S_t(K)$ denote the projection of $R_t(K) \cap \{y \mid y_\emptyset = 1\}$ on the subspace $\mathbb{R}^n$ indexed by the singletons. By the above, we deduce; that

$$P = S_n(K) \subseteq \cdots \subseteq S_{t+1}(K) \subseteq S_t(K) \subseteq \cdots \subseteq S_1(K).$$

Matrix reformulation Let $\mathcal{K}$ denote the linearization of K consisting of the vectors $y \in \mathbb{R}^{\mathcal{P}_w(V)}$ satisfying the linear system:

$$\begin{aligned} & g_l^T y \geq 0 \quad \text{for } l = 1, \cdots, m, \\ & \sum_{R \subseteq H \subseteq S} (-1)^{|H \setminus R|} y_H \geq 0 \quad \text{for all } R \subseteq S \subseteq V \; with \; |S| = w. \end{aligned} \tag{9}$$

Given $y \in \mathbb{R}^{\mathcal{P}_{t+w}(V)}$, consider the matrix Y whose rows and columns are indexed, respectively, by $\mathcal{P}_w(V)$ and $\mathcal{P}_t(V)$ and with entries $Y(K, H) := y(K \cup H)$ for $K \in \mathcal{P}_w(V)$ and $H \in \mathcal{P}_t(V)$. Denote by $e_H(H \in \mathcal{P}_t(V))$ the elementary unit vectors in $\mathbb{R}^{\mathcal{P}_t(V)}$; then $Ye_H$is the column of Y indexed by H. Then,

$$y \in R_t(K) \Leftrightarrow Y\Big(\sum_{I \subseteq H \subseteq U} (-1)^{|H \setminus I|} e_H\Big) \in \mathcal{K} \text{ for all } I \subseteq U \subseteq V \; with \; |U| = t \tag{10}$$

3.2 The Lasserre hierarchy

For $t \geq \nu - 1$, with ν defined as in (7), set

$$P_t(K) := \{y \in \mathbb{R}^{\mathcal{P}_{2t+2}(V)} \; | M_{t+1}(y) \succeq 0, M_{t+1-\nu_l}(g_l * y) \succeq o \text{ for } l = 1, \cdots, m\} \tag{11}$$

and define $Q_t(K)$ as the projection of $P_t(K) \cap \{y \; | y_{\emptyset=1}\}$ on the subspace $\mathbb{R}^n$ indexed by the singletons. Therefore,

$$P = Q_{n+\nu-1}(K) \subseteq \cdots \subseteq Q_\nu(K) \subseteq Q_{\nu-1}(K).$$

Proposition 3 *For any* $t = 1, \cdots, n, Q_{t+w-1}(K) \subseteq S_t(K)$.

4 Comparing the Lasserre, Sherali-Adams, and Lovász-Schrijver relaxations

Comparison between Sherali-Adams and Lovász-Schrijver The first steps in the Sherali-Adams and Lovász-Schrijver procedures coincide. The next steps are however distinct (although there is an inclusion relationship;) A main difference between the two methods is that the Lovász-Schrijver procedure constructs the successive relaxations recursively by applying t successive lift-and-project steps, each taking place in a space of dimension $O(n^2)$, whereas the Sherali-Adams procedure carries out only one direct lifting step, occurring in a space of dimension $O(n^{t+1})$. Moreover, the projection step is not mandatory in the Sherali-Adams procedure, as approximate solutions for optimization problems on P can be obtained by optimizing directly over the set $R_t(K)$.

Now we assume in this section that K is a polytope; that is, K is defined by (6) where all the polynomials g_l have degree 1(thus $v = w = 1$,or $v = w = 0$ if $K = [0, 1]^n$).

Theorem 4 *If K is a polytope, then $S_t(K) \subseteq N(S_{t-1}(K))$ for all $t = 1, \cdots, n$ (setting $S_0(K) := K$).*

Corollary 5 $S_t(K) \subseteq N_t(K)$ *for all* $t = 1, \cdots, n$.

Proposition 3 and Corollary 5 imply that

$$Q_t(K) \subseteq S_t(K) \subseteq N^t(K)$$

for $t = 1, \cdots, n$.

5 Application to the stable set polytope

Given a graph $G = (V = \{1, \cdots, n\}, E)$, let $ST(G)$ denote the stable set polytope of G, let

$$FR(G) := \{x \in \mathbb{R}^n_+ \mid x_i + x_j \leq 1 \text{ for all } ij \in E\}$$

be its basic linear relaxation defined by nonnegativity and the edge inequalities, and let

$$TH(G) := \left\{ x \in \mathbb{R}^n \mid \begin{pmatrix} 1 \\ x \end{pmatrix} = Ye_0 \text{ for some positive semidefinite matrix } Y = (Y_{ij})^n_{i,j=0} \right.$$
$$\left. \text{satisfying } Y_{ii} = Y_{0i} \ (i \in V), Y_{ij} = 0 \ (ij \in E) \right\} \tag{12}$$

be the basic semidefinite relaxation of $ST(G)$. Let us compare how the various methods apply to the pair $P := ST(G)$, $K := FR(G)$.

Define the *N-rank* of $FR(G)$as the smallest integer t for which $N^t(FR(G)) = ST(G)$; define similarly *SA-rank* and the *Lasserre rank* of $FR(G)$.

The following results are shown by Lovás and Schrijver. The polytope $N(FR(G))$ is defined by the non-negativity and edge constraints together with the odd hold inequalities: $\sum_{i \in V(C)} x_i \leq \frac{|C|-1}{2}$ for C odd hole in G. If G has n nodes and stability number $\alpha(G)$, then its *N-rank* t satisfies:

$$\frac{n}{\alpha(G)} - 2 \leq t \leq n - \alpha(G) - 1; \tag{13}$$

the *N-rank* t of an inequality $a^T x \leq \beta$ valid for *ST(G)* (with integer coefficients and distinct from the nonnegativity constraints) satisfies:

$$\frac{1}{\beta}(\sum_{i \in V} a_i - 2\beta) \leq t \leq \sum_{i \in V} a_i - 2\beta. \tag{14}$$

The Sherali-Adams hierarchy does not seem to yield a significant improvement with respect to the sequence $N^t(FR(G))$. Indeed, the lower bounds from (13) and (14) remain valid for the *SA-rank* of $FR(G)$.

The Lasserre hierarchy does improve on the sequence $N^t(FR(G))$. For an edge $ab \in E$, let $g_{ab}(x) := 1 - x_a - x_b$ be the polynomial corresponding to the edge inequality $x_a + x_b \leq 1$. Then the positive semidefinite constraint $M_t(g_{ab} * y) \succeq 0$ can be replaced by the linear equation: $y_{ab} = 0$. It can be shown that the Lasserre hierarchy already finds $ST(G)$ at step $\alpha(G) - 1$.

Proposition 6 $ST(G) = Q_{\alpha(G)-1}(FR(G))$ *for a graph with stability number* $\alpha(G) \geq 2$.

By the definition (12), $TH(G)$ can be seen as the projection on $\mathbb{R}^n$ of the set of vectors $y \in \mathbb{R}^{\mathcal{P}_2(V)}$ satisfying $y_\emptyset = 1$ and

$$M_1(y) \succeq 0, y_{ab} = 0 (ab \in E).$$

6 Application to the max-cut problem

Given a graph $G = (V = \{1, \cdots, n\}, Ev$, the max-cut problem asks for a partition $(S, V \setminus S)$ maximizing the total cardinality (or weight) of the edges ij cut by the partition (i.e., such that $|S \cap \{i, j\}| = 1$). Hence it can be formulated as an unconstrained quadratic ± 1-problem:

$$\max (x^T A x \mid x \in \{\pm 1\}^n), \tag{15}$$

where A is a (suitably defined) symmetric matrix, but the treatment below remains valid for A arbitrary.

Since we are now working with ± 1 variables in place of $0 - 1$ variables, we should adapt some of the definitions given earlier in the note. In particular, given $K \subseteq [-1, 1]^n$, the conditions (2) and (3) defining the matrix set $M(K)$ in the Lovász-Schrijver procedure read now: $y_{jj} = y_{00}$ and $Y(e_0 \pm e_j) \in \tilde{K}$ $\texttt{for}$ $j = 1, \cdots, n$. When defining the moment matrices in (4), one should consider the semigroup $\mathcal{P}(V)$ with the symmetric difference as semigroup operation in place of the union. Namely, the $(I, J)-$th entry of a moment matrix $M_V(y)$ is now $y(I \triangle J)$ instead of $y(I \cup J)$.

Now we formulate relaxations for the problem (15). In this strategy we formulate (15) as a linear problem

$$\max(\langle A, X \rangle \mid X \in CUT_n)$$

over the *cut polytope*

$$\mathsf{CUT}_n := \mathrm{conv}(xx^T \mid x \in \{\pm 1\}^n).$$

Consider MET_n as the *metric polytope* which consists the symmetric matrices X with an all one diagonal and satisfying the *triangle inequalities*:

$$X_{ij} + X_{ik} + X_{jk} \geq -1, X_{ij} - X_{ik} - X_{jk} \geq -1$$

for all distinct $i, j, k \in V$.

Let $\mathsf{CUT}(G)$ and $\mathsf{MET}(G)$ denote the projections of CUT_n and MET_n, respectively, on the subspace $\mathbb{R}^E$ indexed by the edge the edge set of G. Then, $\mathsf{CUT}(G) \subseteq \mathsf{MET}(G)$ with equality if and only if G has no $K_5 - minor$.

By considering $K := \mathsf{MET}(G)$, $N(\mathsf{MET}(G))$is the relaxation of $\mathsf{CUT}(G)$ in the LS construction. Also we can first apply the LS construction to $K := \mathsf{MET}(K_n)$ and then project back on the edge space $\mathbb{R}^E$ by π_E,(the projection from the space indexed by the edge set of K_n to the space indexed by the edge set of G):

$$N(G) \subseteq B(\mathsf{MET}(G))$$

but it is not known whether equality holds in general. Laurent proved that $N^t(\mathsf{MET}(G)) = \mathsf{CUT}(G)$ holds if G has t edges whose contraction produces a graph with no $K_5 - minor$. In particular, $N^{n-\alpha(G)-3(G)} = \mathsf{CUT}(G)$; moreover, $N^{n-\alpha(G)-3}(\mathsf{MET}(G)) = \mathsf{CUT}(G)$if G has a maximum stable set whose complement induces a graph with at most three connected components. In particular, $N^{n4}(K_n) = \mathsf{CUT}(K_n)$ for $n \geq 4$.

When applying the Sherali-Adams and Lasserre constructions to $K = \mathsf{MET}(G)$, one finds the relaxations $S_t(\mathsf{MET}(G))$ and $Q_t(\mathsf{MET}(G))$ satisfying: $Q_t(\mathsf{MET}(G)) \subseteq S_t(\mathsf{MET}(G))$. The definition of $Q_t(\mathsf{MET}(G))$ involves a semidefinite program contain- ing possibly exponentially many constraints (at least as many as the number of circuits in G). In order to decrease the number of constraints, one can consider instead the set $\pi_E(Q_t(\mathsf{MET}(K_n)))$ whose definition involves $O(n^3)$ semidefinite constraints.

References

[1] Laurent, Monique. "A comparison of the Sherali-Adams, Lovsz-Schrijver, and Lasserre relaxations for 01 programming." Mathematics of Operations Research 28.3 (2003): 470-496.