LP Hierarchies for 0/1 Programming

Derek Modzelewski
dmodzel2@jhu.edu

David Li
dzli@jhu.edu

Abstract

In this final project, we will provide a brief introduction into linear programming (LP) and semi-definite programming (SDP) relaxations. While LP and SDP relaxations are powerful tools in approximation algorithms, the integrality gap inherent in the methods mostly dictate the approximation ratio. By using LP/SDP hierarchies, we can start with any relaxation of our choice and "lift" the relaxation such that the final relaxation gives exactly the integer polytope. Clearly this final relaxation takes exponential time, but the in-between relaxations are not well-constrained in running time and integrality gap. In many cases, these are actually open questions.

1 Introduction to LP/SDP relaxations

In class, we've seen LP relaxations in quite detail, since it transforms NP-complete problems (integer programming) into polynomial-time solvable problems (linear programming). However, when we do this relaxation, the optimal solution for the LP is at most the optimal solution to the integer program. This results in an integrality gap, which is the maximum ratio between the solution quality of the integer program and of the linear program relaxation.

Geometrically, the feasible region for the ILP problem is a polytope – it's a convex hull of all integer solutions to the integer program. The feasible region of the LP is also a polytope, but it's not constrained to integer solutions, and includes the entire integer polytope as a subset. As mentioned in the abstract, we can use the LP/SDP hierarchies along with a starting relaxation to generate a sequence of relaxations such that the final relaxation gives the integer polytope we were looking for. For linear objective functions, solving this program will give us the true integer optimal solution as all vertices of the integer hull are integral. Intuitively, such methods (sometimes called "lift-and-project" methods) attempt to simulate degree k programming with an LP or SDP.

However, nonlinear constraints are more powerful than linear constraints. For instance, if we require $x \in \{0, 1\}$, then in a quadratic program we can use the constraint $x(1 - x) = 0$ to enforce the integrality constraint.

We'll see specifically how the three major LP/SDP relaxations work in the following section. However, before we do that, we'll have to set up some notation first.

2 LP Hierarchies

2.1 Notation

For any given 0-1 programming problem, P, let $K \subseteq [0, 1]^n$ be a convex space containing all feasible solutions for the given problem, where

$$P = K \cap \{0, 1\}^n$$

For Lovász-Schrijver, we will denote $N^t(K)$ as the polytope described by the t-th level of the Lovász-Schrijver hierarchy, $S^t(K)$ as the polytope described by the t-th level of Sherali-Adams, and $Q^t(K)$ as the polytope described by the t-th level of the Lasserre hierarchy:

$$P = N^n(K) \subseteq N^0(K) = K$$
$$P = S^n(K) \subseteq S^0(K) = K$$
$$P = Q^n(K) \subseteq Q^0(K) = K$$

2.2 Lovász-Schrijver Hierarchy

Generally, the Lovász-Schrijver Hierarchy isn't used because it is weaker than either the Sherali-Adams or Lasserre Hierarchies.

As with the other methods we will discuss, we define the levels of the hierarchy recursively:

$$N^t(K) = N(N^{t-1}(K))$$

where we will define $N(V)$ as follows for any convex hull, V
First, define

$$C := \{\lambda \begin{pmatrix} 1 \\ x \end{pmatrix} | x \in V, \lambda \geq 0\}$$

(the cone extending from the origin in $\mathbb{R}^{n+1}$ space through the embedding of V into $\mathbb{R}^{n+1}$ where the first coordinate is one. Note that the added dimension is indexed by 0, and so e_0 is the elementary unit vector in the added dimension.

We define a matrix Y to be a *generator* for V iff

$$\begin{aligned} Y_{i,j} &= Y_{0,j} && \forall j \in [n] \\ Ye_j &\in C && \forall j \in [n] \\ Y(e_0 - e_j) &\in C && \forall j \in [n] \end{aligned}$$

Note that $\forall x \in V \cap \{0,1\}^n$, we have $\tilde{Y} = \binom{1}{x}\binom{1}{x}^T$ is a generator for V since

$$\begin{aligned} \tilde{Y}_{j,j} &= x_j * x_j = x_j = \tilde{Y}_{0,j} \\ \tilde{Y}e_j &= \begin{pmatrix} 1 \\ x \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix}^T e_j && x_j \in \{0,1\} \text{ valid } \lambda \\ &= \begin{pmatrix} 1 \\ x \end{pmatrix} x_j \\ \tilde{Y}(e_0 - e_j) &= \begin{pmatrix} 1 \\ x \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix}^T (e_0 - e_j) \\ &= \begin{pmatrix} 1 \\ x \end{pmatrix}(1 - x_j) && 1 - x_j \in \{0,1\} \text{ valid } \lambda \end{aligned}$$

And define

$$G(V) := \{Y : Y \text{ is a generator for } V\}$$
$$N(V) := \{x \in \mathbb{R}^n : \exists Y \in G(V) \textit{s.t.} \begin{pmatrix} 1 \\ x \end{pmatrix} = Ye_0\}$$

$\forall x \in V \cap \{0,1\}^n$:

$$\begin{aligned} \begin{pmatrix} 1 \\ x \end{pmatrix} &= \begin{pmatrix} 1 \\ x \end{pmatrix} * 1 \\ &= \begin{pmatrix} 1 \\ x \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix}^T e_0 \\ &= \tilde{Y}e_0 \end{aligned}$$

and by previous mention, $\tilde{Y} \in G(V) \Rightarrow V \cap \{0,1\}^n \subseteq N(V)$ (no integer solutions are removed from V when generating $N(V)$):
$\Rightarrow P \subseteq N^t(K)\ \forall t \geq 0$

2.3 Sherali-Adams Hierarchy

The formulation for Sherali-Adams we will give here is in terms of moment matrices – consult section 4 for a primer on the moment problem and the representation of positive polynomials as a sum of squares. This is the way that Lasserre's hierarchy is defined, but it turns out to be helpful to write Sherali-Adams in terms of this in order to compare Lasserre hierarchies directly. The original formulation of Sherali-Adams in 1990 [6] is in terms of the "reformulation-linearization technique" or RLT.

The construction of Sherali-Adams applies to semi-algebraic sets in the $[0,1]^n$ cube – that is sets

$$K = \{x \in [0,1]^n \ : \ g_\ell(x) \geq 0, \ell \in [m]\}$$

where $g_1, \dots, g_m$ are polynomials in $x_1, \dots, x_n$ where the degree of each variable is at most 1. Define $P = \text{conv}\{K \cap \{0,1\}^n\}$ be the convex hull that we want to describe. In terms of what we've seen in class, this is the same as the optimal solution to the 0/1 integer linear program. Because in an integer program, $x \in \{0,1\}^n$, $x_i^2 = x_i$ for $i \in [n]$, so we can assume each variable x_i appears in a polynomial g_ℓ with at most degree 1, and so we can rewrite $g_\ell(x) = \sum_{i \subseteq V} g_\ell(I) \prod_{i \in I} x_i$. The repetition of g_ℓ for this and the vector in $\mathbb{R}^{P(V)}$ with components $g_\ell(I)$ is intentional.

Now, we'll discuss the constructions of Sherali-Adams in terms of moment matrices (again, see section 4). For two $x, y \in \mathbb{R}^{P(V)}$, define $x * y \in \mathbb{R}^{P(V)}$ by

$$x * y := M_V(y)x; \text{ so that } (x * y)(I) = \sum_{K \subseteq V} x_K y_{I \cup K} \text{for all} I \subseteq V.$$

Furthermore, this operation $*$ obeys the following modified "commutativity" law

$$x * (y * z) = y * (x * z) \forall x, y, z \in \mathbb{R}^{P(V)}.$$

The next lemma provides the foundation for both the Sherali-Adams and Lasserre constructions:

Lemma 1. *Fix $x \in K \cap \{0,1\}^n$. Then the vector $y \in \mathbb{R}^{P(V)}$ with entries $y(I) := \prod_{i \in I} x_i$ satisfies*

$$M_V(y) \succeq 0, M_V(g_\ell * y) \succeq 0 \textit{ for } \ell \in [m].$$

Proof. As constructed, $M_V(y) = yy^\top$ and $M_V(g_\ell * y) = g_\ell(x) yy^\top$ because $y(I \cup J) = y(I)y(J)$ for all $I, J \subseteq V$ by definition. □

If we relax the condition in Lemma 1 and require positive semidefiniteness of specific submatrices, then we arrive at the Lasserre and Sherali conditions. Particularly, the Lasserre hierarchy requires

$$M_{t+1}(y) \succeq 0, M_{t-v_\ell+1}(g_\ell * y) \succeq, \ell \in [m],$$

for an integer $t \geq v_\ell - 1$ and $v_\ell = \lceil \frac{\deg g_\ell}{2} \rceil$ and the Sherali-Adams hierarchy requires

$$M_W(y) \succeq 0 \text{ for } w \subseteq V \text{ and } |W| = \min(t + w, n) \tag{1}$$

$$M_U(g_\ell * y) \succeq 0 \text{ for } U \subseteq V \text{ and } |U| = t, \ell \in [m] \tag{2}$$

In both the Sherali-Adams construction and Lasserre construction, the polytope P is found after n iterations. As it turns out, the Sherali-Adams relaxations turn out to be linear relaxations since the condition in 1,2 can be reformulated as a linear system in y.

The original formulation of Sherali-Adams is in terms of what's called a "lift-and-project" system [4]. Given a polytope P, defined by $\alpha_\ell^\top x - b \leq 0$, $\ell \in [m]$, $x \in [0,1]^n$, we define $S_A^t(P) \subseteq \mathbb{R}^{P_t^n - 1}$ to be the set of $z \in \mathbb{R}^{P_t^n - 1}$ that satisfy the following lifted linear system (P_t^n are all subsets of $[n]$ of size at most t):

For each constraint $c(x) \leq 0$ of P, for each $U \in P_t^n$, and for each $W \subseteq U$, we have a valid constraint

$$c(x) \prod_{s \in W} x_s \prod_{s \in U - W} (1 - x_s) \leq 0,$$

and for each $I \in P^n_{|U|+1}$, replace $\prod_{s \in I} x_s$ by z_I. Then, we add this linear inequality to the set of constraints determining $S^t_A(K)$.

Then, the t-th Sherali-Adams tightening of P is $S^t(K)$ is the projection of z onto singleton indices. The Sherali-Adams rank of a polytope P is the smallest integer t for which $S^t(K)$ is the integral hull of K.

2.4 Lasserre Hierarchy

The Lasserre system extends the Sherali-Adams system. In order to define the Lasserre system, we use the terminology from the lift-and-project definition of Sherali-Adams.

Consider a vector $z \in S^t_A(P)$. We say z is *of PSD form* if we can write a correspondence between every set I and vectors $v \in \mathbb{R}^{P^n_{t+1}}$ such that each pair of sets I, J that satisfy $|I \cup J| \leq t+1$ also satisfies $v_I \cdot v_J = z_{I \cup J}$.

Then, if z satisfies the conditions for Sherali-Adams, and is also of PSD form, then $z \in L^t_A(P)$. Intuitively, this means z is a lifted solution for t Lasserre tightenings of P. Then, the t-th Lasserre tightening of P is $L^t(K)$, and is the projection of z onto singleton indices.

3 Applications of Hierarchies

3.1 Stable-Set Polytope Problem

The Stable-Set problem (also known as Maximum Independent Set) is, given a graph G, to pick the set of nodes with highest total weight such that no two selected nodes are adjacent.

Linear Relaxation:

$$\begin{aligned} \max_x \quad & \sum_{i \in [n]} x_i w(v_i) & \\ \textit{s.t.} \quad & x_i + x_j \leq 1 & \forall (i,j) \in E(G) \end{aligned}$$

Quadratic Relaxation:

$$\begin{aligned} \max_x \quad & \sum_{i \in [n]} ||x_i||^2 w(v_i) & \\ \textit{s.t.} \quad & x_i \cdot x_j \leq 0 & \forall (i,j) \in E(G) \end{aligned}$$

Stephen and Tuncel (1999) showed that $N_+^{\frac{n-1}{2}}(K)$, the $\frac{n-1}{2}$ level of the Lovász-Schrijver Hierarchy on the Quadratic Relaxation, is the optimal polytope.

This is insufficient to derive a poly-time algorithm, but it reduces the exponent of the runtime.

3.2 Max-Cut Problem

Recall that max-cut can be formulated as an unconstrained quadratic ± 1-problem by

$$\max x^\top A x | x \in \{\pm 1\}^n$$

and A is the adjacency matrix, where $A_{i,j} = c(i,j)$, the cost of the edge from vertex i to vertex j

The most straightforward way is to formulate the problem as a linear problem

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

where CUT_n is the cut polytope

$$CUT_n := \text{conv}(xx^\top | x \in \{\pm 1\}^n).$$

We'll also define MET_n as the metric polytope containing symmetric matrices X with ones on the diagonal and satisfying $X_ij + X_ik + X_jk \geq -1, X_ij - X_ik - X_jk \geq -1$.

Here, we would try to apply the constructions to CUT, represented as an LP. The most practical LP formulation of the cut polytope of a graph $G = (V, E)$. Define $CUT(G), MET(G)$ by the projections of CUT, MET onto the subspace $\mathbb{R}^E$ indexed by the edges of G. Barahona and Mahjoub in 1986 showed that $CUT(G) \subseteq MET(G)$ and equality holds iff G does not have K_5 as a subgraph (that is, the connected graph on five vertices).

If we apply the Lovász-Schrijver construction on $K = MET(G)$, then we find that $N(MET(G)) = CUT(G)$. we can also apply Lovász-Schrijver on $K = MET(K_n)$ and project back into $\mathbb{R}^E$ – this instead will give us $N(G) = \text{proj}_E(N(MET(K_n)))$, a relaxation of $CUT(G)$. As of the publication of [5], it is known that $N(G) \subseteq N(MET(G))$, but general equality is an open question.

Laurent (2001) showed that $N^t(MET(G)) = CUT(G)$ if G has t edges whose contraction produces a graph with no K_5 minor. Additionally, $N^{n-4}(K_n) = CUT(K_n)$ for $n \geq 4$. Finally, the inclusion $N_+(MET(G)) \subseteq N(MET(G))$ is generally strict, there has not been an instance of a graph for which the iterations needed to find $CUT(G)$ is smaller for N_+ than for N.

If we instead apply the Sherali-Adams or Lasserre constructions to $K = MET(G)$, the resulting relaxations $Q^t(MET(G)) \subseteq S^t(MET(G)) \cap N_+^t(MET(G))$. However, the definition of Q^t has a SDP with (worst-case) exponentially many constraints – this can be fixed by projecting onto $\mathbb{R}^E$ – this gives $O(n^3)$ semidefinite constraints.

According to Lasserre's paper [8] specifically pertaining to max-cut, the Lasserre hierarchy has very good results (ALG $\geq$ 0.87OPT for Q with nonzeroes entries, where Q is the matrix associated with the quadratic formulation, where the algorithm is only to use the first iteration of the Lasserre hierarchy. Of course, deeper levels of the hierarchy will yield better results, but as yet no layer reachable in polynomial time has been proven to provide asymptotically stronger results.

4 Moment Theory

As mentioned earlier, we can formulate both the constructions of Lasserre and Sherali-Adams in terms of moment matrices – matrices indexed by subsets of the set $V = \{1, \ldots, n\}$ such that their (I, J)-entry depends only on the union $I \cup J$. The following survey of the moment problem and moment matrices is due to [5], but some additional clarifying details are provided.

In order to introduce the moment problem, we need a little bit of machinery. First, a semigroup is a mathematical structure composed of a set S and an operation $\cdot$ on the set. The set must be closed under $\cdot$, and for any 3 elements $a, b, c \in S$, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$. Note that $a \cdot b \neq b \cdot a$ in general. However, if such an identity holds for all pairs $a, b \in S$, we call S abelian.

So, let $(S, +)$ be a an abelian semi-group, and let S^* be the set of multiplicative nonzero mappings $f : S \to \mathbb{R}$ i.e. nonzero functions that satisfy $f(a + b) = f(a)f(b)$ for all $a, b \in S$. For a sequence $y = \{y_\alpha\}_{\alpha \in S}$ indexed by S, its *moment matrix* $M(y)$ is defined as the $S \times S$ matrix such that $M(y)_{\alpha\beta} = y_{\alpha+\beta}$ for any two $\alpha, \beta \in S$.

If S is the semigroup $P(V)$ (the power set of V), with union operation, the moment matrix $M_V(y)$ is exactly the one we saw in section 2.3. Specifically, we index by subsets of V, and generate new subsets by unions. If, instead, S is the semigroup $\mathbb{Z}_+^n$ with the addition operation, then we denote the moment matrix as $M^{\mathbb{Z}}$ for the moment matrix of $y \in \mathbb{R}^{\mathbb{Z}_+^n}$, and $M_t^{\mathbb{Z}}(y)$ is the principal submatrix indexed by all sequences $\alpha \in \mathbb{Z}_+^n$ with $|\alpha| \leq t$.

Then, we can also characterize these sequences like we characterize matrices – specifically there is the notion of a *positive semi-definite* sequence $y \in \mathbb{R}^S$. Following Berg's notation in 1979[2] and 1984[3], such a sequence is positive semi-definite if every finite principal submatrix of its moment matrix $M(y)$ is positive semi-definite. Recall from lecture that a symmetric matrix M is positive semi-definite if $z^\top Mz$ is nonnegative for every nonzero column vector z in $\mathbb{R}^n$.

For a subset $F \subseteq S^*$, a sequence $y \in \mathbb{R}^S$ is called an F-moment sequence if there exists a positive Radon measure μ on S^* supported by F such that

$$y_\alpha = \int_{S^*} f_\alpha d\mu(f) \qquad \forall \alpha \in S.$$

In particular, in measure theory, a measure m is called *inner regular* or *tight* if, for any Borel set B, $m(B)$ is the supremum of $m(K)$ over all compact subsets K of B. A measure m is called *locally finite* if every point of X has a neighborhood U for which $m(U)$ is finite. Finally, a measure m is called a *Radon measure* if it is both inner regular and locally finite.

Finally, since we now know what moment matrices and moment sequences are, the moment problem is just the problem of characterizing moment sequences. It's been studied extensively, especially with the semigroup $\mathbb{Z}_+^n$ since that results in $S^* = \mathbb{R}^n$ and the moment sequence condition is simply

$$y_\alpha = \int x^\alpha d\mu(x).$$

References

[1] Arora, S. *Lift and Project Methods.* http://www.cs.princeton.edu/courses/archive/spr05/cos598B/liftproj.pdf

[2] Berg, C., Christensen, J. P. Reuss, and Jensen, C.U. "A Remark on the Multidimensional Moment Problem." Mathematische Annalen 243 (1979): 163-170.

[3] Berg, C., Christensen, J. P. Reuss, and Ressel, P. "Harmonic Analysis on Semigroups" Springer-Verlag (1984).

[4] K. Georgiou, A. Magen, Limitations of the Sherali-Adams Lift and Project System: Compromising Local and Global Arguments, 2008.

[5] Laurent, M. *A Comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre Relaxations for 0 - 1 Programming.*

[6] Sherali, H., Adams, W. *A Hierarchy of Relaxations between the Continuous and Convex Hull Representations for Zero-One Programming Problems* SIAM J. Discrete Math., 3(3), 411430.

[7] Stephen, T., Tuncel, L. 1999. On a representation of the matching polytope via semidefinite liftings. Mathematics of Operations Research 24 17.

[8] Lasserre, Jean-Bernard. "A Max-cut Formulation of 0/1 Programs." ArXiv: https://arxiv.org/pdf/1505.06840.pdf.