CS880: Approximation and Online Algorithms	**Scribe:** Kexin Li, Luke Swanson
Lecture 1: Hierarchies of Integer Programming Relaxations	**Date:** 2019/12/06

1.1 Preliminaries

We begin these notes by briefly reviewing a few definitions which will be used extensively in our discussion of convex programming hierarchies for integer programs.

The integer linear programming (ILP) which is NP-hard to solve is:

$$\begin{aligned} \text{maximize } & w^T x \\ \text{subject to } & Ax \le b \\ & x \in \{0,1\}^n \end{aligned} \tag{1.1.1}$$

Linear Program Relaxation is over the region

$$P := \{x \in \mathbb{R}^n | Ax \le b\} \tag{1.1.2}$$

Integral polytope we really want to find is relaxed convex hull of integral solution

$$P_I = conv(P \cap \{0,1\}^n) \tag{1.1.3}$$

Hierarchies are systematic ways of constructing tighter and tighter relaxtion between P and P_I

$$P \supseteq P^2 \supseteq \cdots \supseteq P^t \supseteq \cdots \supseteq P_I \tag{1.1.4}$$

Goal: hopefully it will converge in **finite number** of steps and also **easy to optimize** over P^t
Why are we interested in hierarchies?

1. **Polyhedral Combinatorics** How many rounds are required to reduce the feasible set to P_I? Which constraints are satisfied?
2. **Proof Systems** Complexity and lower bound
3. **Approximation Algorithm** Integrality gap after t rounds/levels, better approximation, conjecture on the hardness of the question

We will introduce three hierarchies for LP and SDP, give examples and compare them for tightness.

1. Lovasz-Schrijver hierarchy (LP/SDP) [1991]
2. Sherali-Adams hierarchy (LP) [1990]

3. Lasserre hierarchy (SDP) [2001]

$$P = P^1 \supseteq P^2 \supseteq \cdots \supseteq P^t \supseteq \cdots \supseteq P^n = P_I \tag{1.1.5}$$

They are all specific type of lift and project method. And they can all optimize over P^t in polynomial time for fixed t.

1.2 Lift and Project Method

Let's start with notation.

$$P := \{x \in \mathbb{R}^n | Ax \leq b)\}$$

We homogenize P to the cone in $n+1$ dimension

$$\begin{aligned} C(P) &:= \{(\lambda, \lambda x_1, \ldots, \lambda x_n) \in \mathbb{R}^{n+1} | \lambda \geq 0, x \in P\} \\ &= \{(x'_0, x'_1, \ldots, x'_n) \in \mathbb{R}^{n+1} | bx'_0 - Ax' \geq 0\} \\ &= \{y \in \mathbb{R}^{n+1} | g_l^T y \geq 0, l = 1, \ldots, m\} \end{aligned}$$

Where $Ax \leq b$ as $a_l^T x \leq b_l$ and $g_l = \begin{pmatrix} b_l \\ -a_{l1} \\ \vdots \\ -a_{ln} \end{pmatrix}$

This type of optimization is often called Conic Optimization: minimizing a convex function over the intersection of an affine subspace and a convex cone. The class of conic optimization problems includes some of the most well known classes of convex optimization problems, namely linear and semidefinite programming.

1.2.1 Framework

The lift and project method are carried out with 3 steps:

1. **Generate New Variables**
   Multiple the system $Ax \leq b$ by the product of constrains $x_i \geq 0$ and $1 - x_i \geq 0$
   $\rightarrow$ Polynomial system in x

2. **Linearize and lift**
   Introduce new variables y_I for the products $\prod_{i \in I} x_i$ and setting $x_i^2 = x_i$
   $\rightarrow$ Linear system in (x,y) denoted M
   Lift – higher dimension space (x,y)

3. **Project**
   Reduce it back on the x-variables space

$\rightarrow$ LP relaxation $P' = \{x|\exists y s.t.(x,y) \in M\}$ satisfying

$$P_I \subset P' \subset P$$

Intuitively, we cut P by some planes Y_I

The linear programming relaxation of an integer program is strengthened by lifting the problem into a higher dimensional space, where a more convenient formulation may give a tighter relaxation. One then has a choice between working with this tighter relaxation in the higher dimensional space,or projecting it back onto the original space. In this latter case,the whole procedure can be viewed as a method for generating cutting planes in the original space. The different hierarchy has different multiplier and linearization.

1.3 The Balas-Ceria-Cornuejols construction[1993]

Let's talk a look at a very simple example with Lift and Project method where we fix one variable to be integral value as a time.

1. **Multiple** the system $Ax \leq b$ by x_1 and $1 - x_1$

$$x_1(b - Ax) \geq 0, (1 - x_1)(b - Ax) \geq 0$$

2. **Linearize and lift**
   Introduce new variables y_i for the products $x_1 x_i$ and set $y_1 = x_1$
   $\rightarrow$ Linear system in (x,y)

$$M_1 = \{(x, y) : y_1 = x_1, bx_1 - Ay \geq 0, b(1 - x_1) - A(x - y) \geq 0\}$$

3. **Project**
   Reduce M_1 back on the x subspace and get P^1
   $\rightarrow$ LP relaxation P' satisfying

$$P_I \subset P^1 \subset P$$

4. **Iterate:** use variables x_2 starting from P^1 and get P^{12}

Lemma 1.3.1 $P^1 = conv(P \cap \{x : x_1 = 0, 1\})$

Proof: The forward direction: Let $x \in P^1$ and x can be written as $x = x_1 \frac{y}{x_1} + (1 - x_1)\frac{x-y}{1-x_1}$
Conversely, let $x \in (P \cap \{x : x_1 = 0, 1\})$, then $(x, x_1 x) \in M_1$. Therefore $x \in P^1$
If B is convex and $A \subset B$, then $conv(A) \subset B$ ∎

Corollary 1.3.2 $P_I = P^{12\ldots n}$ *after n iterations*

1.4 The Lovasz-Schrijver construction[1991]

The original construction of Lovasz-Schrijver (LS) starts with defination of N/N_+ operator.

Definition 1.4.1 *Let cone* $K = C(P) = (y \in \mathbb{R}^{n+1})$ *and* $N(P)$ *be the cone in* $\mathbb{R}^{n+1}$ *defined by* $y = (y_0, y_1, \ldots, y_n) \in N(K)$ *iff there exists a matrix* $Y \in \mathbb{R}^{(n+1)\times(n+1)}$ *such that*

1. Y *is symmetric*
2. $Y_{ii} = y_i$ *for* $i \in [n]$
3. $Y_{0i} = y_i$ *for* $i \in [n]$
4. $Y_i \in K$ *and* $Y_0 - Y_i \in K$ *for* $i \in [n]$ *where* Y_i *is the* i*-th row of* Y

In such case, Y is called protection matrix of y. In addition, if Y is positive semidefinite, then $y \in N_+(K)$. We define $N^0(K)$ and $N^0_+(K)$ as K and $N^t(K) = N(N^{t-1}(K))$ (respectively $N^t_+(K) = N_+(N^{t-1}_+(K))$).
Let's think of it in the Lift and Project framework first with LP where it introduces all the products of two variables at each level.

1. **Multiple** the system $Ax \le b$ by x_i and $1 - x_i$ for $i \in [n]$
   $(b_l - a_l^T x)x_i = g_l^T \begin{pmatrix} 1 \\ x \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix}^T e_i \ge 0$ for all l
   $(b_l - a_l^T x)(1 - x_i) = g_l^T \begin{pmatrix} 1 \\ x \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix}^T (e_0 - e_i) \ge 0$ for all l
2. **Linearize and lift**
   Introduce new matrix variables $Y = \begin{pmatrix} 1 \\ x \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix}^T$
   $\rightarrow$ Linear system in (x,y)
   $$M(P) = \{Y \in S_{n+1} : Y_{0i} = Y_{ii}, Y_{e_i}, Y(e_0 - e_i) \in P_I \forall i \in [n]\}$$
3. **Project**
   Reduce $M(P)$ back on the x subspace
   $$N(P) = \{x \in \mathbb{R}^n | \exists Y \in M(P) s.t. \begin{pmatrix} 1 \\ x \end{pmatrix} = Y_{e_0}\}$$
4. **Iterate** starting from $N(P)$ and get $N^2(P)$

Similarly, we can construct it for SDP with

$$M_+(P) = \{Y \in S^+_{n+1} : Y_{0i} = Y_{ii}, Y_{e_i}, Y(e_0 - e_i) \in P_I \forall i \in [n]\}$$

and

$$N_{+}(P) = \{x \in \mathbb{R}^n | \exists Y \in M_{+}(P) s.t. \begin{pmatrix} 1 \\ x \end{pmatrix} = Y_{e_0}\}$$

As we can see, the higher dimensional space they use is obtained by multiplying every inequality by every 0-1 variable and its complement in turn, then linearizing the resulting system of quadratic inequalities and finally projecting back the system onto the original space.The lifting phase of this procedure involves a **squaring of the number of variables** and an even steeper increase in the number of constraints, but iterating the lifting/projecting step a number of times equal to the number of original 0-1 variables yields the convex hull of feasible 0-1 points. Comparing this to BCC construction, it only doubles number of variables.

1.4.1 Properties of LS/LS_+

Lemma 1.4.2 *1. Iterate* $N^t(P) = N(N^{t-1}(P))$, $N_+^t(P) = N_+(N_+^{t-1}(P))$

2. $P_I \subset N_+(P) \subset N(P) \subset P$

3. $N(P) \subset \cap_{i \in [n]} conv(P \cap \{x | x_i = 0, 1\})$

4. $N^n(P) = P_I$

5. if one can optimize in polynomial time over $C(P)$*, then the same holds for* $N^t(P)$ *and for* $N_+^t(P)$ *for any fixed* t

Example: Consider a l_1 ball centered at $e/2$
$P = \{x \in \mathbb{R}^V | \sum_{i \in I} x_i + \sum_{i \in V \setminus I} (1 - x_i) \geq 1/2 \forall I \subseteq V\}$
$P_I = \emptyset$ but $\frac{e}{2} \in N_+^{n-1}(P)$
Hence, n iterations of N_+ are need to find P_I ■

1.4.2 Distribution point of view

Alternatively, we can think of Y_{ij} as $E[y_i y_j] = Prob(y_i \& y_j)$. And at each level of the hierarchy, we inquire on the the value of a vertex.

1.4.3 Application of LS

$P = FR(G) = \{x \in \mathbb{R}_+^V | x_i + x_j \leq 1, (i, j) \in E\}$
$P_I = STAB(G)$: stable set polytope of G

1. $Y \in M(FR(G)) \rightarrow y_{ij} = 0$ for all edges $(i, j) \in E$
2. The **clique inequality**: $\sum_{i \in Q} x_i \leq 1$ is valid for $N_+(FR(G)$
3. The **odd cycle inequalities**: $\sum_{i \in V(C)} x_i \leq \frac{|C|-1}{2}$ are valid for $N(FR(G))$ and they determine it exactly

1.5 Sherali-Adams Hierarchy

We will motivate the Sherali-Adams (SA) relaxation by looking into Lovasz-Schrijver. Given a vector $y = (y_0, y_1, \ldots, y_n) \in N^2(K)$ and let Y be y's protection matrix. Let's consider $Y_i = (y_i y_0, y_i y_1, \ldots, y_i y_n)$ - ith row of Y. $Y_i \in N(K)$ must has a protection matrix Y'. The (jk) entry in Y' - $Y'_{jk} = y_i y_j y_i y_k = y_i y_j y_k$ if y is integral solution. This triple depends on the choice of i. If we pick j-th row in Y and ik-th entry in its protection matrix Y'', then $Y''_{ik} = y_j y_i y_k$. There is no guarantee that $Y'_{jk} = Y''_{ik}$.
Sherali Adams relaxation addresses this by adding constrains to enforce that all products evaluate to the same quantity. The main idea of Sherali-Adams is to introduce variables $Y_S = \prod_{i \in S} y_i$ for each $S \subset [n]$ of $t+1$. One may use these variables to define locally consistent family of distributions. Let $a^T y \leq b$ be one of the constrains define the convex polytope $\tilde{P}$. Sherali-Adams adds constrains that would be equivalent to the following in the case of boolean vector y, $\forall S, Y \subset [n]$ such that $|S| + |T| \leq t$

$$(a^T y - b) \prod_{i \in S} y_i \prod_{j \in T} (1 - y_i) \leq 0 \tag{1.5.6}$$

Carrying out the multiplication, we can express these constrains using our variables $\{Y_S\}_{|S| \leq t+1}$

$$\sum_{T' \subset T} (-1)^{|T'|} \sum_{i=1}^{n} a_i Y_{S \cup T' \cup \{i\}} - b Y_{S \cup T'} \leq 0 \tag{1.5.7}$$

The number of new variables and constraints added are $O(n^{O(t)})$ and resulting LP can be solved in $O(n^{O(t)})$ time. Some works has been devoted to show one ca noptimize over t-th level of SA using only weak separation oracle and in certain case can solve such problems in $n2^{O(t)}$

1.5.1 Example of Max Independent Set

The t-th level of SA relaxation for MIS is:

$$\text{maximize} \sum_{i=1}^{n} Y_{\{i\}}$$

subject to

$$\sum_{T' \subset T} (-1)^{|T'|} (Y_{S \cup T' \cup \{i\}} + Y_{S \cup T' \cup \{j\}} - Y_{S \cup T'}) \leq 0, |S| + |T| \leq t, (i,j) \in E \tag{1.5.8}$$

$$0 \leq \sum_{T' \subset T} (-1)^{|T'|} Y_{S \cup T' \cup \{i\}} \leq \sum_{T' \subset T} (-1)^{|T'|} Y_{S \cup T'}$$

$$Y_{\emptyset} = 1$$

Let's think of it in the Lift and Project framework.

1. **New polynomial constraints**

$$x^I(1-x)^{W\setminus I}(b-Ax)\geq 0 \qquad \text{for } I\subseteq W \text{ with } |W|=t \tag{1.5.9}$$

$$x^I(1-x)^{W\setminus I}\geq 0 \qquad \text{for } I\subseteq U \text{ with } |U|=t+1 \tag{1.5.10}$$

2. **Linearize & lift** Introduce new variables y_U for all $U\in P_{t+1}(V)$, setting $y_i=x_i$

3. **Project** back on x-variable space and get $SA^t(P)$

Lemma 1.5.1 *1.* $SA^1(P)=N(P)$

2. $SA^t(P)\subseteq N^t(P)$

Let's consider the last step of SA at full lifting.

$$\begin{aligned} x\in\{0,1\}^n \to y^x &= \prod_{i\in I} x_i \in \{0,1\}^{\mathcal{P}(V)} \\ &= (1,x_1,x_2,\ldots,x_n,x_1x_2,\ldots,x_{n-1}x_n,\ldots,\prod_{i\in V}x_i) \\ \to Y = y^x(y^x)^T &= (\prod_{i\in I}x_i\prod_{j\in J}x_j)_{I,J\subseteq V} \end{aligned}$$

If $x\in P\cap\{0,1\}^n$, then $Y=(y^x)(y^x)^T$ satisfies:

1. $Y(\emptyset,\emptyset)=1$

2. Y(I,J) depends only on $I\cap J$ **moment matrix**

3. $Y\succeq 0$

4. $g_l(x)Y\succeq 0$ **localizing moment matrix**

These conditions characterize $conv(y^x : x\in P\cap\{0,1\}^n)$, thus P_I

1.5.2 Moment matrix and localizing moment matrices

We can define these $Y(I,J)$ as moment matrix.

Definition 1.5.2 *Given* $y\in\mathbb{R}^{\mathcal{P}(V)}$ *define:*

1. *The* ***moment matrix*** $M_V(y)=(y_{I\cup J})_{I,J\in\mathcal{P}(V)}$

2. *The* ***shift vector*** $g*y=(\sum_{J\in\mathcal{P}(V)} g_i y_{I\cup J})_{I\in\mathcal{P}(V)}$ *[linearize* $g(x)y^x=(g(x)x^I)_I$*]*

3. *The* ***localizing moment matrix****:* $M_V(g*y)$

Theorem 1.5.3 *1.* $conv(y^x(y^x)^T : x\in P\cap\{0,1\})$ *is equal to* $\Delta_P=\{y\in\mathbb{R}^{\mathcal{P}(V)}|y_\emptyset=1, M_V(y)\succeq 0, M_V(g_l*y)\succeq 0\}$

2. P_I *is the projection of* Δ_P

3. Δ_P *is a polytope*

Since SA^t is a convex relaxation, then any convex combination of 0/1 solution is a feasible for SA. The converse is only true locally. Any feasible solution to t-th level Sherali-Adams relaxation is equivalent to a family of distribution $\{D(S)\}_{|S|\leq t+1}$ such that they are locally consistent.
Specifically, we have the following lemma:

Lemma 1.5.4 *Consider a family of distribution* $\{D(S)\}_{|S|\leq t+2, S\subset[n]}$ *where* $D(S)$ *is defined over* $0,1^S$. *If the distribution satisfies:*

1. *For all* $(i,j)\in E$ *and* $S\supseteq\{i,j\}$, $P_{D(S)}[(y_i=1)\wedge(y_j=1)]=0$

2. *For all* $S'\subseteq S\subseteq[n]$ *with* $|S|\leq t+1$, *the distribution* $D(S)$ *and* $D(S')$ *agrees on* S'

Then $Y_S = P_{D(S)}[\wedge_{i\in S}(y_i=1)]$ *is a feasible solution for the above level-t SA relaxation. Conversely, for any feasible solution* $\{Y'_S\}$ *for the level* $t+1$, *there exists a family of distributions satisfying the above properties as well as* $P_{D(S)} = \wedge_{i\in S'}(y_i=1) = Y'_S$ *for all* $S'\subseteq S\subset[n]$ *s.t.* $|S|\leq t+1$

Extending further the intuition of the variables Y_S as probabilities, we can also define variables for arbitrary events over a set S of size at most t. A basic event is given by a partial assignment $\alpha\in\{0,1\}^S$ which assigns value 0 to some variables in S which we denote by $a^{-1}(0)$ and 1 to the others denoted by $a^{-1}(1)$. We can define variables $X_{S,\alpha}$ when $S\leq t$ and $\alpha\in\{0,1\}^S$ as

$$X_{S,\alpha} := \sum_{T\subset} \tag{1.5.11}$$

Now, let's motivate the idea of Lasserre construction by comparing the moment matrices.
First, we consider moment matrices in greater detail, so we can use them to show the relation between Sherali-Adams and Lasserre hierarchies.

1.5.3 Moment Matrices

The underlying idea in our use of moment matrices is to index each matrix by subsets of some other set. In general, if V is some set of variables, and $I,J\subseteq V$, we can form a moment matrix as $M_V(y)_{I,J} = f(I,J)$. In particular, we use $f(I,J) = y_{I\cup J}$ for moment matrices, and $f(I,J) = \sum_{i=1}^{n} A_{li}y_{I\cup J\cup\{i\}} - b_l y_{I\cup J}$ for localizing moment matrices corresponding to constraint l.
We can obtain this equation for a localizing moment matrix by defining a *shift operator* on vectors x,y as follows:

$$x*y \text{ is a vector indexed by } I\subseteq V \text{ such that } (x*y)_I = \sum_{J\subseteq V} x_J y_{I\cup J} \tag{1.5.12}$$

If a_l is the vector of coefficients for constraint l, then the localizing moment matrix for constraint l is $M^l_V(a*y)$.

A slight variation on these definitions is to take set V, and choose some constant $t \leq |V|$. Then $M_t(y)$ only takes the entries of $M_V(y)$ where the corresponding I, J have sizes each at most t. This leads us to the following definitions of hierarchies in terms of moment matrices:
Sherali Adams relaxation(Local) $SA^t(P)$
Consider $M_U(y) = (y_{I\cup J})_{I,J\subseteq U}$, indexed by $\mathcal{P}(U)$ for $U \subseteq V$
$M_U(y) \succeq 0, M_W(g_l * y) \succeq 0, \forall U \in \mathcal{P}_{t+1}(V), W \in \mathcal{P}_t(V)$
LP with variables y_I for all $I \in \mathcal{P}_{t+1}(V)$
Lasserre relaxation(Global) $Las^t(P)$
Consider $M_t(y) = (y_{I\cup J})_{|I|,|J|\leq t}$, indexed by $\mathcal{P}_t(V)$ for some $t \leq n$
$M_t(y) \succeq 0, M_{t-1}(g_l * y) \succeq 0$
SDP with variables y_I for all $I \in \mathcal{P}_{2t}(V)$

From the moment matrices, we can see $L^t(P) \subseteq SA_{t-1}(P)$

1.6 Lasserre Construction [2001]

Suppose, as in previous hierarchies, that we wish to solve some integer linear program with constraints $Ax \leq b$ and $x \in 0, 1^n$. We relax this to a general linear program with $x \in \mathbb{R}$, and wish to introduce new constraints to tighten the relaxation. Previously, we saw the Lovász-Schrijver and Sherali-Adams hierarchies, which applied new constraints for up to n rounds, at which point we were left with $P = \text{conv}(K \cap 0, 1^n)$, where K was the set of feasible solutions to the linear program with $x \in [0, 1]^n$. Now, the elements of P are clearly either themselves solutions to the integer program, or a convex combination of integer program solutions. We can interpret the fractional solutions as probabilities for a value of '1' in an integer solution, or the probability a variable is "included" in a set, as in the case of Max Independent Set. In the Lasserre hierarchy, as in the Sherali-Adams hierarchy, we attempt to tighten the solution set more quickly by introducing constraints on joint events, relative to the Lovàsz-Schrijver hierarchy that only constrains individual events. We then define the Lasserre hierarchy as in the notes by Rothvoß [1]:
Assume $K = \{x \in \mathbb{R}^n | Ax \geq b\}$, and $y \in \mathbb{R}^{2^{[n]}}$. Let $\text{LA}_t(K)$ denote the t^{th} level of the Lasserre hierarchy, $I, J \subseteq [n]$ with $|I|, |J| \leq t$, and let m be the number of constraints in the problem. Further, for a vector y, define $M_t(y) = (y_{I\cup J})_{I,J}$ the *moment matrix* of y, and $M_t^l(y) = \left(\sum_{i=1}^{n} A_{li} y_{I\cup J\cup\{i\}} - b_l y_{I\cup J}\right)_{I,J}$ $\forall l \in [m]$ the *moment matrix of slacks.*
Then $y \in \text{LA}_t(K)$ if $M_t(y) \succeq 0$ and $M_t^l(y) \succeq 0$. Returning to the idea of probabilities of joint events, we then interpret y_S as the probability that all $i \in S \subseteq [n]$ are included in the solution.

1.6.1 Example of Max Independent Set

The Lasserre hierarchy produces semidefinite programs. Adopting the SDP notation used by Chlamtac and Tulsiani [2], the Maximum Independent Set SDP produced by the t^{th} level of the Lasserre hierarchy is:

$$
\begin{aligned}
\text{maximize} \quad & \sum_{i\in V} ||U_{\{}i\}||^2 \\
\text{subject to} \quad & \langle U_{\{i\}}, U_{\{j\}}\rangle = 0 \\
& \langle U_{S_1}, U_{S_2}\rangle = \langle U_{S_3}, U_{S_4}\rangle \quad \forall S_1 \cup S_2 = S_3 \cup S_4 \\
& ||U_{\emptyset}||^2 = 1
\end{aligned}
\tag{1.6.13}
$$

1.7 Comparison on 3 hierarchies

It can be shown that the t^{th} Lasserre hierarchy level is a subset of the same hierarchy levels for Sherali-Adams and Lovasj-Schrijver hierarchies. First, it is clear that $LS_+^t \subseteq LS^t$, since LS_+ imposes an additional constraint on LS. Now, because Lasserre constrains not only products of pairs of variables, but products of triples, quadruples, quintuples, etc. we have $L^t \subseteq LS_+^t$. Finally, as we saw in the discussion of moment matrices, the Lasserre hierarchy uses moment matrices based on all subsets up to a given size, while Sherali-Adams has moment matrices based on subsets of a specific subset, and its subsets. Thus, at the same hierarchy level, Lasserre has more constraints, and $L^t \subseteq SA^t$.
However, it is important to note that while the Lasserre hierarchy has the best performance in terms of quickly reducing the size of the polytope, it introduces many new variables. Thus, there may be cases where another hierarchy is preferable to Lasserre.

1.8 Dual Approach of SDP

First let's review SDP and its dual: The standard form of SDP is analogous to the standard form of LP.
For C, X, A_i be $n \times n$ matrices for $i \in [m]$ in the primal (P), and for y is $n \times 1$ vector and S is $n \times n$ matrix in the dual (D)

$$
\begin{aligned}
&\min_X && \langle C, X\rangle && \max_y && b^T y \\
&\text{s.t.} && \langle A_i, X\rangle = b_i, \quad i \in [m] && \text{s.t.} && \sum_{i=1}^m y_i A_i + S = C \\
& && X \succeq 0, && && S \succeq 0
\end{aligned}
$$

where the Frobenius product $\langle C, X\rangle = \sum_{i,j} c_{i,j} x_{ij}$

Dual Approach of SDP Recall the definition of positive semdefinte matrices: let $X \in \mathbb{R}^{n\times n}$ be symmetric, we say X is positive semidefinite (PSD or $X \succeq 0$) if one of the following condition is met:

1. $\forall a \in \mathbb{R}^n, a^T X a \geq 0$
2. $X = BB^T$ for some B

3. all of X's eigenvalues are non-negative

Facts from linear algebra

1. $\langle A, B\rangle = Tr(A^T B)$ by the definition of matrix multiplication
2. $Tr(AB) = Tr(BA)$. This implies any cyclic permutation of matrices has the same trace.

Lemma 1.8.1 *Let X be symmetric, then $X \succeq 0 \iff \langle A, X\rangle, \forall A \succeq 0$*

This is LP if all the matrices are diagonal. The feasible region is $\{X | \langle A_i, X\rangle = b, \forall i, a^T X a \geq 0, \forall a\}$. The positive semidefinite constrains efficiently creates infinite number of linear constrains on X
Proof:

1. **Backward direction:** Suppose $X \not\succeq 0$, then $\exists$ vector a such that $a^t X a < 0$. Let $A = aa^T$, then $A \succeq 0$. Therefore $\langle A, X\rangle = Tr(A^T X) = Tr((aa^T)^T X) = Tr(aa^T X) = Tr(a^T X a) < 0$
2. **Forward direction:** Suppose $X \succeq 0$, Let $A \succeq 0$. Then $A = BB^T = \sum_i b_i b_i^T$ for some matrix B with columns $b_1, b_2, \ldots, b_n$. Then $\langle A, X\rangle = Tr(A, X) = \sum_i b_i^T X b_i \geq 0$ by similiar proof of first part.

■

Lemma 1.8.2 *If x/y are feasible for $(P)/(D)$, then $b^T y \leq \langle C, X\rangle$*

Proof: By the feasibility of y, $\langle C, X\rangle = \langle \sum_i y_i A_i, X\rangle + \langle S, X\rangle$
By the feasibility of x, $\langle \sum_i y_i A_i, X\rangle = b^T y$
By lemma (previous page), $\langle S, X\rangle \geq 0$
Thus $b^T y = \langle C, X\rangle - \langle S, X\rangle \leq \langle C, X\rangle$ Therefore any solution for the dual lower bounds the minimum of the primal. ■

Strong duality for SDP only holds under Slater's condition.

Definition 1.8.3 *Slater's Condition Feasible region has an interior point.*

Primal (moment)

$$\max \sum_{i=1}^{n} c_i y_i \text{ s.t. } y_\emptyset = 1, M_t(y) \succeq 0, M_{t-1}(g_\ell * y) \succeq 0 (\ell \in [m])$$

Dual (sums of squares)

$$\min \lambda \text{ s.t. } \lambda - \sum_{i=1}^{n} c_i x_i \in \Sigma_{2t} + \sum_{\ell=1}^{m} g_\ell \Sigma_{2t-2} \quad \bmod \left\langle x_i^2 - x_i : i \in [n]\right\rangle$$

Example of Stable Set (Independent Set) Primal $\max \sum_{i \in V} y_i$ s.t. $y_\emptyset = 1, M_t(y) \succeq 0, y_{ij} = 0 (ij \in E)$
Dual
$\min \lambda$ s.t. $\lambda - \sum_{i \in V} x_i \in \Sigma_{2t} + I$

Ideal $I = \left\langle x_i^2 - x_i (i \in V), x_i x_j (ij \in E)\right\rangle$

1.8.1 Application to MAX CUT

Let's revisit the famous Goemans-Williamson SDP relaxation in 1995. Lasserre relaxation of order 1 and ovasz-Schrijver of order 1 can both capture the product of two variables. Therefore it can achieve the same approximation factor as 0.878. Max Cut max $\sum_{ij \in E} w_{ij}(1 - x_i x_j)/2 s.t. x \in \{\pm 1\}^n$
Cut polytope: $CUT_n = conv(xx^T | x \in \{\pm 1\}^n)$
Lasserre relaxation of order 1:

$$L_1 = \{X \in S^n : X \succeq 0, X_{ii} = 1(i \in V)\}$$

1.9 Literature summary on some positive and negative results

Below are a list of literatures in which the authors explores on the bound of hierarchies. There are some promising results in MAX CUT case because one round of SDP Hierarchy can achieve a non-trivial approximation. However we cannot approaching the optimum arbitrarily with small number of rounding. Some people studies the hirarchy in some specific graph such as dense graph. Also other problems such as vertex cover, set cover, 3-SAT has been looked at using this technique.

1. MAX CUT problem: The integral gap for trivial relaxation is 1/2
2. MAX CUT problem: The integral gap for metric relaxation is also 1/2 [Poljak-Tuza 1994]
3. MAX CUT problem: Lovasz-Schrijver Hierarchies: [Schoenebeck-Trevisan-Tulsiani 2006]
   (a) The integrality gap remains $1/2 + \epsilon$ after $c_\epsilon n$ rounds of the N operator
   (b) 0.878 after one round of the N_+ operator
4. MAX CUT problem: Sherali Adams Hierarchies:
   (a) The integrality gap remains $1/2+\epsilon$ after n^{γ_ϵ} iterations[Charikar-Makarychev-Makarychev 2009]
   (b) PTAS for Max cut in dense graph [De la Vega and Kenyon-Mathieu 2007]
5. $\theta(\log n)$ rounds of LS_+ gives tight relaxation for MIS (Quasi-poly time improvement) [Feige Hrauthgamer 2003]
6. Approximation guarantee improves with higher order relaxation for MIS in 3-uniform hypergraph [Chlamtac Singh 2008]
7. Lasserre relaxation of order $O(r/\epsilon^2)$ to get $\frac{1+\epsilon}{\min\{1,\lambda_r\}}$ [Guruswami Sinp 2011]
8. Strong nonapproximability results in the LS hierarchy for max-3sat, hypergraph vertex cover and set cover. $\omega(n)$ rounds of the LS+ procedure do not allow nontrivial approximation[Alekhnovich, Arora, Tourlakis 2005]
9. $2 - o(1)$ integrality gap after $\Omega(\sqrt{\frac{\log n}{\log \log n}})$[Georgiou, Magen, Pitassi, Tourlakis 2010]

10. n^δ round of SA doesn't yield a better than $2-\epsilon$ approximation for MC and VC (Unique game hardness)[Charikar, Makarychev, Makarychev]

References

[1] T. Rothvoß. The Lasserre hierarchy in Approximation algorithms – Lecture Notes for the MAPSP 2013 Tutorial. 2013.

[2] E. Chlamtac, M. Tulsiani. Convex Relaxations and Integrality Gaps In Handbook on Semidefinite, Conic and Polynomial Optimization