Lecture 20: Lift and Project, SDP Duality

Lecturer: *Sanjeev Arora* Scribe: *Yury Makarychev*

Today we will study the Lift and Project method. Then we will prove the SDP duality theorem.

1 Lift and Project

Consider the following integer program for the Vertex Cover problem:

$$\begin{aligned} \min \sum_{i \in V} & x_i \\ & x_i + x_j \geq 1 && \forall \{i,j\} \in E \\ & x_i \in \{0,1\} && \forall i \in V \end{aligned}$$

Replacing the last constraint with the constraint

$$0 \leq x_i \leq 1,$$

we get an LP relaxation for the problem. Unfortunately, the integrality gap of this LP is 2. However, we can rewrite the integer program as an equivalent quadratic program:

$$\begin{aligned} \min \sum_{i \in V} & x_i \\ & (1-x_i)(1-x_j) = 0 && \forall \{i,j\} \in E \\ & (1-x_i)x_i = 0 && \forall i \in V \\ & 0 \leq x_i \leq 1 && \forall i \in V \end{aligned}$$

The idea of Lift and Project is to simulate quadratic programming (or degree k programming) with an LP/SDP. We introduce new variables Y_{ij} that represent product terms:

$$Y_{ij} \overset{\text{intended}}{=} x_i x_j.$$

For brevity, let $x_0 = 1$, and then $Y_{i0} = x_i$. The condition that $(1-x_i)(1-x_j) = 0$ can be now expressed as

$$1 - Y_{i0} - Y_{j0} + Y_{ij} = 0.$$

We also introduce new constraints on Y_{ij}:

1. $Y_{ij} = Y_{ji}$;

2. $Y_{i0} = Y_{ii}$;

3. the matrix Y is positive semidefinite;

4. If $a^T x \geq b$ was a linear constraint for the original LP then we add the following linear constraints on Y_{ij} (for all i):

$$\begin{aligned}(a^T x)x_i &\geq bx_i \\ (a^T x)(1-x_i) &\geq b(1-x_i)\end{aligned}$$

Fact: All SDP's we saw (or that have been useful) can be derived using this process from the obvious linear relaxations[1]. The following section borrowed from the paper "Towards strong nonapproximability results in the Lovász-Schrijver hierarchy" by Mikhail Alekhnovich, Sanjeev Arora and Iannis Tourlakis presents the details.

2 LS_+ derivation of popular SDP relaxations

To illustrate the power of the LS_+ procedure, we sketch how to use a few rounds of LS_+ to derive popular SDP relaxations used in famous approximation algorithms. (This was suggested by the reviewers, who pointed out that this is not very well-known.)

It will be more convenient to view LS_+ as a method for generating new inequalities. Given any relaxation

$$a_r^T x \geq b \qquad r = 1, 2, \ldots, m \tag{1}$$

(where the trivial constraints $0 \leq x_i \leq 1$ are assumed to be included), one round of LS_+ produces a system of inequalities in $(n+1)^2$ variables Y_{ij} for $i, j = 0, 1, \ldots, n$. As mentioned, the intended "meaning" is that $Y_{ij} = x_i x_j$ and $Y_{00} = 1, Y_{0i} = x_i = x_i x_0$, and $Y_{00} = 1$ so every quadratic expression in the x_i's can be viewed as a linear expression in the Y_{ij}'s. This is how the quadratic inequalities below should be interpreted.

$$\begin{aligned}(1-x_i)a_r^T x &\geq (1-x_i)b && \forall i = 1, \ldots, n, \quad \forall r = 1, \ldots, m \\ x_i a_r^T x &\geq x_i b && \forall i = 1, \ldots, n, \quad \forall r = 1, \ldots, m \\ x_i x_i &= x_i x_0 && \forall i = 1, 2, \ldots, n\end{aligned}$$

(The last constraint corresponds to the fact that $x_i^2 = x_i$ for 0/1 variables.) Finally, one imposes the condition that (Y_{ij}) is positive semidefinite. Obviously, any positive combination of the above inequalities is also implied, and the derivations below will use this fact.

2.1 Deriving the GW relaxation

The Goemans-Williamson relaxation for MAX-CUT [GW'94] involves finding unit vectors $u_1, u_2, \ldots, u_n$ so as to minimize

$$\sum_{\{i,j\} \in E} \frac{1}{4} |u_i - u_j|^2 .$$

This SDP relaxation can be derived by one round of LS_+ on the trivial linear relaxation. This relaxation has 0/1 variables x_i and d_{ij}. In the integer solution, x_i indicates which side of the cut

[1] this process may involve several rounds of Lift and Project

vertex i is on, and d_{ij} is 1 iff i, j are on opposite sides of the cut.

$$\max_{\{i,j\} \in E} d_{ij} \tag{2}$$

$$d_{ij} \geq x_i - x_j \qquad \forall i, j = 1, 2, \ldots, n \tag{3}$$

$$d_{ij} \leq x_i + x_j \qquad \forall i, j = 1, 2, \ldots, n \tag{4}$$

$$d_{ij} \leq 2 - (x_i + x_j) \qquad \forall i, j = 1, 2, \ldots, n \tag{5}$$

Then one round of LS_+ generates the following inequalities on d_{ij}:

$$x_i d_{ij} \geq x_i(x_i - x_j) \tag{6}$$

$$(1 - x_i)d_{ij} \geq (1 - x_i)(x_j - x_i). \tag{7}$$

Adding these and simplifying using the fact that $x_i^2 = x_i$ for 0/1 variables, one obtains $d_{ij} \geq (x_i - x_j)^2$. Similarly one can obtain $d_{ij} \leq (x_i - x_j)^2$ whereby it follows $d_{ij} = (x_i - x_j)^2 = Y_{ii} + Y_{jj} - 2Y_{ij}$.

Now if (Y_{ij}) is any feasible solution then its Cholesky decomposition $v_0, v_1, \ldots, v_n \in \Re^{n+1}$ are vectors such that $Y_{ij} =< v_i, v_j >$. Then $d_{ij} = |v_i - v_j|^2$. Now define the set of vectors $u_1, u_2, \ldots, u_n$ as $u_i = v_0 - 2v_i$. These satisfy

$$d_{ij} = \frac{1}{4}|u_i - u_j|^2 \tag{8}$$

$$|u_i|^2 = |v_0|^2 - 4 < v_0, v_i > +4\,|v_i|^2 = 1. \tag{9}$$

Thus the u_i's are a feasible solution to the GW relaxation. We conclude that one round of LS_+ produces a relaxation at least as tight as the GW relaxation (and in fact one can show that the two relaxations are the same).

2.2 Deriving the ARV relaxation

Arora, Rao, and Vazirani [ARV'04] derive their $\sqrt{\log n}$-approximation for SPARSEST CUT using a similar SDP relaxation whose salient feature is the *triangle inequality*:

$$|u_i - u_j|^2 + |u_j - u_k|^2 \geq |u_i - u_k|^2 \qquad \forall i, j, k.$$

(In other words, $d_{ij} = |u_i - u_k|^2$ forms a metric space.) This relaxation minus the triangle inequality is derived similarly to the GW relaxation above (details omitted). The claim is that the triangle inequality is implied after three rounds of LS_+. As shown in [LS'91], r rounds of LS_+ imply all inequalities on subsets of size r that are true for the integer solution. In other words, the induced solution on subsets of size r lies in the convex hull of integer solutions. Thus after three rounds the d_{ij} variables restricted to sets of size three lie in the *cut cone*. Since the cut cone is just the set of ℓ_1 (pseudo)metrics, it follows that the d_{ij} variables form a (pseudo)metric. Thus three rounds of LS_+ give a relaxation that is at least as strong as the ARV relaxation.

3 Sherali–Adams Relaxations

Now let us generalize the Lift and Project method to k-ary programming [Sherali-Adams'91]. For every subset F of indices of size at most k, we introduce a new variable Y_S. The intended value of Y_S is

$$Y_S = \prod_{i \in S} x_i.$$

First we require that for every set S (s.t. $|S| \leq k-2$), the matrix $(Y_{S\cup\{i\}\cup\{j\}})_{ij}$ is positive semidefinite. Then we add another family of generic constraints:

$$\forall S,T : S\cap T=\emptyset, \quad |S|+|T|\leq k \qquad \prod_{i\in S}(1-x_i)\prod_{j\in T}x_j \geq 0.$$

As it often happens in theoretical CS the case $k>2$ is much more complicated than the case $k=2$.

4 Discrepancy

4.1 Two Party Case

We are now interested in applying the Lift and Project method to analyze the discrepancy. Recall the definition of the discrepancy for two parties.

DEFINITION 1 *(2 party case) Discrepancy of a matrix M is*

$$\|M\|_C = \max_{\substack{x_1,\ldots,x_n\in\{0,1\}\\ y_1,\ldots,y_n\in\{0,1\}}} |\sum_{i,j} M_{ij}x_iy_j|.$$

We saw in the last lecture that the cut norm $\|M\|_C$ is well approximated by the norm $\|\cdot\|_{\infty\mapsto 1}$:

$$\|M\|_{\infty\mapsto 1} = \max_{\substack{x_1,\ldots,x_n\in\{-1,1\}\\ y_1,\ldots,y_n\in\{-1,1\}}} |\sum_{i,j} M_{ij}x_iy_j|.$$

Alon and Naor showed that this norm is approximated within a constant factor by the following SDP relaxation:

$$\|M\|_C = \max_{\substack{u_1,\ldots,u_n\in S^{2n-1}\\ v_1,\ldots,v_n\in S^{2n-1}}} |\sum_{i,j} M_{ij}\langle u_iv_j\rangle|.$$

4.2 Multiparty Case

We are interested in extending this method to multiparty communication complexity. Recall that in the multiparty case we have a "multidimensional matrix" $M_{ij\ldots k}$ (i.e. a covariant tensor); our objective function is to maximize the sum of entries in a set S over all sets S that are cylinder intersections.

DEFINITION 2 *(3 party case) In the 3 party case, each cylinder intersection is determined by its projections on the planes O_{ij}, O_{ik}, and O_{jk}. Let x_{ij*}, y_{i*k}, and z_{*jk} be indicator functions of these projections. Then*

$$\textit{cylinder intersection} = \{(i,j,k) : x_{ij*}=y_{i*k}=z_{*jk}=1\} = \{(i,j,k) : 1-x_{ij*}y_{i*k}z_{*jk}=0\}.$$

So the discrepency is equal to

$$Disc(M) = \max_{\substack{x_{ij*}\in\{0,1\}\\ y_{i*k}\in\{0,1\}\\ z_{*jk}\in\{0,1\}}} |\sum_{i,j,k} M_{i,j,k}\,x_{ij*}\,y_{i*k}\,z_{*jk}|.$$

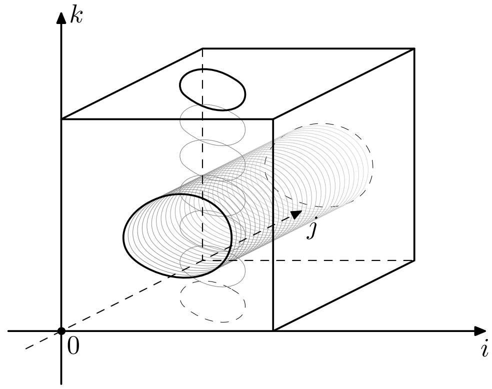

Figure 1: We maximize $|\sum_{i,j,k} M_{i,j,k}\, x_{ij*}\, y_{i*k}\, z_{*jk}|$ over all cylinder intersections.

We now represent this optimization problem as a Sherali-Adams relaxation. Denote the optimum value of this relaxation by $SDP_{opt}(M)$. Clearly, $SDP_{opt}(M) \leq Disc(M)$. How well does $SDP_{opt}(M)$ approximate $Disc(M)$?

Question 1: Is $SDP_{opt} = O(\text{“small”} \cdot Disc(M))$?

Question 2: Show for some natural function (e.g. clique) that SDP_{opt} is small.

Guess: Perhaps, the answer to the first question is NO.

How to solve these problems? The natural way is to use duality.

5 SDP Duality

A general semidefinite program is an optimization program of the following form:

$$\min C \circ X$$

subject to

$$\begin{aligned} A_1 \circ X &= b_1 \\ A_2 \circ X &= b_2 \\ &\vdots \\ A_m \circ X &= b_m \\ X &\succeq 0 \end{aligned}$$

where C, A_1,..., A_m are $n \times n$ symmetric matrices; the inner product $A \circ B$ is defined as

$$A \circ B \stackrel{\text{def}}{=} \sum_{i,j} A_{ij} B_{ij} = tr(AB).$$

Note that this program is a convex program since the set of all positive semidefinite matrices forms a convex cone. This follows from the following two observations:

- If A is a positive semidefinite, then so is λA for every $\lambda > 0$.
- If A and B are positive semidefinite matrices, then so is $A + B$. Indeed, for every vector v, we have
$$v^T(A+B)v = v^T Av + v^T Bv \geq 0 + 0.$$

THEOREM 1 (FACT AT THE ROOT OF DUALITY)
A matrix Y is positive semidefinite iff for every positive semidefinite matrix A, $A \circ Y \geq 0$.

PROOF: First, assume $Y \succeq 0$. Let A be an arbitrary semidefinite matrix. Consider Cholesky decompositions of matrices A and Y: $A = V^T V$, and $Y = Z^T Z$. Then

$$A \circ Y = tr(AY) = tr(V^T V Z^T Z) = tr(ZV^T V Z^T) = tr(ZV^T (ZV^T)^T) \geq 0.$$

Note that if A is a positive *definite* matrix and $Y \neq 0$, then $A \circ Y > 0$. Indeed, since A is nonsingular, V is also nonsingular, therefore $ZV^T \neq 0$, which implies $A \circ Y > 0$.

The other direction is trivial:
$$v^T Y v = Y \circ \underbrace{(vv^T)}_{\succeq 0} \geq 0.$$

Therefore Y is positive semidefinite. □

THEOREM 2
The following system of equations is infeasible

$$\begin{aligned} A_1 \circ X &= 0 \\ A_2 \circ X &= 0 \\ &\cdots \\ A_m \circ X &= 0 \\ X &\succeq 0 \\ X &\neq 0 \end{aligned}$$

iff $\exists x_1, \ldots, x_m \in \mathbf{R}$ *s.t.* $\sum_i x_i A_i \succ 0$.

PROOF: Suppose there are no $x_1, \ldots, x_m \in \mathbf{R}$ s.t. $\sum_i x_i A_i \succ 0$. This means that the linear subspace $\{\sum_i x_i A_i\}$ is disjoint from the cone of positive definite matrices (which equals the interior of the cone of positive semidefinite matrices).

By Farkas lemma, there exists a separating hyperplane that contains this linear space s.t. the semidefinite cone lies on one side of the hyperplane. Let Y be the normal to this hyperplane. Then $Y \circ A_i = 0$, and, by Theorem 1, $Y \succeq 0$. In other words, Y is a feasible solution of the system of the equations. We get a contradiction.

The proof of the other direction is straight forward. Assume that there exist $x_1, \ldots, x_n$ for which $\sum_i x_i A_i \succ 0$; but the system of equations has a feasible solution X. We have

$$\sum_i x_i A_i \circ X = \sum_i x_i (A_i \circ X) = \sum_i x_i \cdot 0 = 0;$$

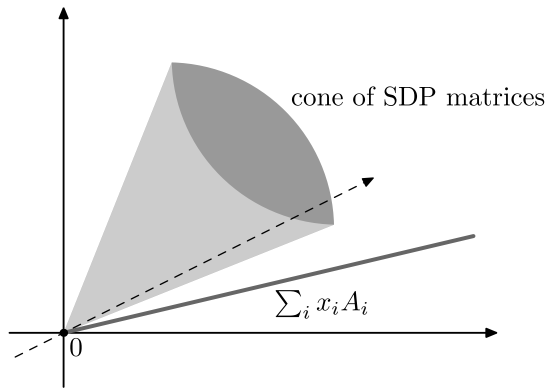

Figure 2: The SDP cone is disjoint from the set $\sum_i x_i A_i$.

on the other hand

$$\sum_i x_i A_i \circ X = \underbrace{(\sum_i x_i A_i)}_{\succ 0} \circ X > 0.$$

We get a contradiction. This concludes the proof. $\square$