Sum of Squares and Lasserre hierarchy

Arka Ray   Rameesh Paul

May 12, 2022

Abstract

In this survey we aim to study hierarchies, which are general procedures to write strengthened convex relaxations (LP/SDP) for optimization problems by introducing additional variables and generating corresponding constraints in a mechanical fashion. We discuss the the Sherali-Adams hierarchy [SA90] and the Lasserre hierarchy [Las01] and exhibit how they shrink the polytope of the feasible region. We then show the equivalence of the primal (Lasserre moment based approach) and the dual (sum of squares of polynomials approach). We will then discuss the MAX-CUT SDP formulation as an application of Lasserre hierarchies. Our presentation of these topics is primarily based on survey by Thomas Rothvoß [Rot13] and survey by Eden Chlamtac and Madhur Tulsiani [CT12]. The view point of hierarchies being generated from lift and project framework is based on [Lau03]. The presentation of sum of squares is based on the monograph by Kothari et al. [FKP19]

1 Introduction

A generic technique in solving combinatorial optimization problem is to formulate them as integer programs and consider convex relaxations of such integer programs; most common of these convex relaxations being linear programming (LP) and semi definite programming (SDP) relaxations. If the relaxation is exact the optimal solution to the relaxation turns out to be integral, but typically this is not the case and the LP/SDP relaxation has an integrality gap. A natural idea is to then examine such gap instances and consider additional constraints which are violated. A bulk of research in approximation algorithms attempts to find better relaxations with smaller integrality gaps. As an example, the canonical SDP relaxation obtained by adding triangle inequality constraints gives improved integrality gaps and approximation algorithms for sparsest cut problem [ARV04] and MAX-2-SAT [LLZ02] respectively.

However generally the process of adding such constraints is problem specific and heuristic, which begs the question whether there can be a systematic way to add such constraints. This is the question that hierarchies of convex programs addresses and various such LP/SDP hierarchies that have been proposed in the literature namely, the Lovász-Schrijver hierarchy [LS91], the Sherali-Adams hierarchy [SA90] and the most powerful of them, the Lasserre hierarchy [Las01]. In this survey we focus mostly on the Lasserre hierarchy or equivalently the sum of squares, however we give a brief overview of other hierarchies.

2 Hierarchies of convex programs

A hierarchy is a sequence of strengthened relaxation of some basic convex program. A level-t relaxation is produced in a mechanical fashion by adding $n^{\mathcal{O}(t)}$ decision variables $x_I, \forall I \subseteq [n], |I| \leqslant t$ and some local consistency constraints over these decision variables. Hence, a level-t hierarchy can be solved in $n^{\mathcal{O}(t)}$ and we are guaranteed that any possible local constraints upto level t hold true. This also implies that a level-n hierarchy corresponds to the integral program, however the number of variables introduced is already exponential. These hierarchies use the method of lift and project to generate polyhedral and spectrahedral lifts to obtain LP and SDP respectively and then project them back onto initial set of decision variables to obtain a polytope which is tighter. For this survey we consider a LP in inequality form and a SDP in matrix form as

$$\begin{aligned} \text{Minimize:} \quad & \langle \mathbf{c}, \mathbf{x} \rangle \\ \text{Subject to:} \quad & A\mathbf{x} \leqslant \boldsymbol{b} \\ & \mathbf{x} \geqslant 0 \end{aligned} \qquad (1)$$

$$\begin{aligned} \text{Minimize:} \quad & \langle C, X \rangle \\ \text{Subject to:} \quad & X \succeq \mathbf{0}, \\ & \langle A_i, X \rangle \leqslant \boldsymbol{b}_i, \quad \forall i = 1, \ldots, m. \end{aligned} \qquad (2)$$

Since the LP's we will be interested in are relaxations of integer programs we can add $x_i \in [0, 1]$ to (1). We denote the polytope of the feasible set by $\mathcal{P}$ and polytope at level-t of some hierarchy as $\mathcal{P}^t$. As note that for any hierarchy of the LP in (1),

$$\mathcal{P}^n (= \mathcal{I}) = \mathsf{conv}(\{\mathbf{x} \in \{0,1\}^n, A\mathbf{x} \leqslant \mathbf{b}\})$$

The above is a restatement of the fact that for a level-n hierarchy the problem is equivalent to the ILP version. We can write it in the above form using the fact that the optimum of LP occurs at a vertex [1] and hence we can even relax to consider the convex hull of integral solutions (integral hull, denoted by $\mathcal{I}$). Next we describe a general framework using lift and project methods for mechanically generating constraints corresponding to these additional variables.

2.1 Lift and project framework

The general lift and project framework can be described as a three step process,

[1] Bauer maximum principle states it more generally for convex functions over compact convex sets

1. Extending - Introduce variables for products of x_i and multiply $A\mathbf{x} \leqslant \mathbf{b}$ by these new variables to obtain new constraints which are polynomials in $\mathbf{x}$.
2. Lifting - Introduce decision variables $\mathbf{y}_I$ for $\prod_{i\in I} x_i$ and using $x_i^2 = x_i$ we convert the polynomial in $\mathbf{x}$ to a lifted linear system in $\mathbf{y}$.
3. Project - We project back on the the original space of decision variables $\mathbf{x} \in \mathbb{R}^n$.

We add constraints for variables of size atmost t to obtain a level-t hierarchy.

2.2 Sherali-Adams hierarchy

We now introduce the Sherali-Adams hierarchy as a tightening to the LP given in (1) by applying the lift and project framework discussed above. The level-t lifting is done by introducing new variables for degree-t terms which are called non-negative juntas.

Definition 2.1. A non-negative junta of degree-t (denoted by $J_{S,T}(\mathbf{x})$) is a multilinear polynomial defined as

$$J_{S,T}(\mathbf{x}) = \prod_{i\in S} x_i \prod_{j\in T}(1 - x_j) \text{ for } S \cap T = \phi \text{ and } |S \cup T| \leqslant t.$$

We multiply the constraint set $A\mathbf{x} \leqslant \mathbf{b}$ by these non-negative junta polynomials to impose new constraints as,

$$\prod_{i\in S} x_i \prod_{j\in T}(1 - x_j)(b_l - \mathbf{a}_l^T \mathbf{x}) \geqslant 0, \forall l \in [m]$$

The non-negativity of junta polynomials is also added as a constraint. We then apply the lifting procedure by multilinearizing the objective and constraints in y_I variables to obtain a lifted $\mathbf{y} \in \mathsf{SA}^t(\mathcal{P})$. We then solve the lifted LP given by,

LP 2.2.

$$\min \ \mathbf{c}^T\mathbf{y}$$

subject to

$$\sum_{T'\subseteq T} (-1)^{|T'|} y_{S\cup T'} \geqslant 0 \qquad \forall S \cap T = \phi, |S \cup T| \leqslant t \tag{3}$$

$$\sum_{T'\subseteq T} (-1)^{|T'|} \left(b_l y_{S\cup T'} - \sum_{i=1}^{n} a_{li} y_{S\cup T\cup\{i\}} \right) \geqslant 0 \qquad \forall S \cap T = \phi, |S \cup T| \leqslant t - 1, \forall l \in [m] \tag{4}$$

$$y_\phi = 1 \tag{5}$$

The LP above is exact for n variables i.e includes any possible constraint that can be added for integral solutions. However if we restrict to polynomial time procedures we can only consider $\mathsf{SA}^t(\mathcal{P})$ for a constant t so that we have a polynomial number of variables $n^{\mathcal{O}(t)}$ and constraints $m.n^{\mathcal{O}(t)}$. We will then project the solution $\mathbf{y}$ onto our original decision variables $y_1, \ldots, y_n$ via some rounding to obtain an approximate solution. We observe that the constraints at level t are retained at level $t+1$, thus we have that $\mathsf{SA}^{t+1}_{\mathsf{proj}}(\mathcal{P}) \subseteq \mathsf{SA}^{t}_{\mathsf{proj}}(\mathcal{P}) \subseteq \mathcal{P}$.

2.2.1 Utility of SAierarchy

We consider the maximum independent set example on cycle graphs with the obvious LP relaxation as,

LP 2.3.

$$\max \sum_{i \in V} x_i$$

subject to

$$x_i + x_j \leqslant 1 \qquad \forall (i,j) \in E \qquad (6)$$

$$0 \leqslant x_i \leqslant 1 \qquad \forall i \in V \qquad (7)$$

If the length of the cycle is odd then the optimal solution to problem is a set of size $(|V|-1)/2$ but we can construct a feasible solution to the LP with value $|V|/2$ (by setting all $x_i = 0.5$). If we consider the LP obtained by $\mathsf{SA}^2(\mathcal{P})$ we can solve such instances of the problem exactly (Appendix B).

Next we wish to better characterize the points inside $\mathsf{SA}^t(\mathcal{P})$ and using the distribution viewpoint can express them as locally consistent expectation functions (called pseduo-expectations) in a one to one correspondence.

2.2.2 Locally consistent probability distribution

Let $\mathbf{z} \in \mathsf{SA}^t(\mathcal{P})$ be a point in the feasible polytope of level-t SA hierarchy then $\mathbf{z}$ defines a map on monomials,

$$\tilde{\mathbb{E}}\left[\prod_{i \in S} x_i\right] := \mathbf{z}_S, \forall S \subseteq [n], |S| \leqslant t.$$

which can be extended to polynomials in a natural way. The probabilistic view point is that these monomials $\prod_{i \in S} x_i$ can be thought of as event denoted by $\mathbb{1}_S$. We can then also interpret the junta polynomials $J_{S,T}(\mathbf{x}) = \prod_{i \in S} x_i \prod_{j \in T}(1 - x_j)$ as the event $\mathbb{1}_{S,T}$ which means $x_i = 1, \forall i \in S$ and $x_j = 0, \forall j \in T$. We can define a distribution over integral solutions for the points in $\mathcal{I}$ by observing that any feasible solution z can be written as a convex combination,

$$\mathbf{z} = \sum_{p_i \in \{0,1\}^n} \alpha_i p_i \text{ defines a probability distribution } \mu^z(p_i) : \{0,1\}^n \to \mathbb{R}^+ := z_i.$$

However if a point $\mathbf{z} \in \mathsf{SA}^t(\mathcal{P})$ and $\mathbf{z} \notin \mathcal{I}$ no such convex combination and hence no such probability distribution exists. However if we consider z_S to be the probability of event $\mathbb{1}_S$ in some underlying distribution and restrict to variables from some set S where $|S| \leqslant t$ the marginal distribution we obtain (denoted by μ_S^z) follows as,

$$z_S = \mathbb{P}\left[\mathbb{1}_S\right] = \sum_{p_i \in \{0,1\}^n : p_i = 1, \forall i \in S} \mathbb{P}\left[p_i\right] = \mu_S^z[\mathbb{1}_S]$$

Similarly we can show that for $T \subseteq S$ we can write distribution over events $\mathbb{1}_{T,K}$ where $T \cup K = S$ as,

$$\mu_S^z[\mathbb{1}_{T,K}] = \left(\sum_{K' \subseteq K} (-1)^{|K'|} z_{T \cup K'}\right)$$

The constraints in SA relaxation make the marginal distribution consistent.[2] The distribution is called pseudo-distribution and it can be shown that if we consider integral solutions (over S) or equivalently restrict ourselves to solutions in which only t variables become integral, then these behave like true distributions. Taking expectations over these distributions give rise to pseudo expectations for $\mathcal{P}$ i.e

$$\tilde{\mathbb{E}}\left[J_{S,T}(x)\right] = \mu_{S \cup T}(\mathbb{1}_{S,T})$$

We discuss these things more formally and in much more detail in Section 3.2.

[2]The constraints in SA relaxation ensure that probabilities are non-negative and sum to 1 and act like a true distribution for subsets upto size t.

2.3 Lasserre hierarchy

The Lasserre hierarchy which is attributed to Shor[Sho87], Parillo[Par03] and Lasserre [Las01] is an extension of the ideas in SA hierarchy in some SDP fashion. As before if we choose $\mathbf{x} \in \mathcal{I}$, we can write it as a convex combination of integral solutions and interpret x_i as the probability that $x_i = 1$ in an integral solution. However in general we are only guaranteed a fractional solution $\mathbf{x} \in \mathcal{P}$ which we can still interpret $x_i's$ as the marginal probabilities but they don't give much information about the joint events, e.g., all we can guarantee is $\mathbb{P}[x_i = 1 \wedge x_j = 1] \in [\max\{x_1 + x_j - 1, 0\}, \min\{x_i, x_j\}]$. Hence we introduce additional variables $y_I = \mathbb{P}[\wedge_{i \in I}(x_i = 1)]$ so that the fractional solution can be written as convex combination of integral solutions (locally for a level-t Lasserre) and we get the correct values for joint probabilities. Note that $y_\phi = 1$ as before (for homogenization) and the original variables are captured as $y_{\{i\}} = x_i$. However we need to do introduce such y_I for all $I \subseteq [n]$ and hence the vector $\mathbf{y}$ is a 2^n dimensional vector. To make the whole procedure tractable we introduce such variables only for sets of size t and hence we get $\binom{n}{t} = n^{\mathcal{O}(t)}$ variables. The lifting procedure add constraints on such variables in an SDP fashion (by arranging in a matrix where (i, j) entry is $\prod_{i \in I, j \in J} x_i x_j$) and then we linearize them and add the psd-ness constraint (instead of the non-negativity constraint in SA). This construction is referred to as a level-t Lasserre hierarchy denoted by $\mathsf{LAS}^t(\mathcal{P})$.

2.3.1 Lasserre Construction

As described above the variables are introduced for sets upto size t, to make the probability of joint events consistent (locally). Hence the additional constraints that impose this are called as consistency constraints and these are enforced by the moment matrix $M^t(\mathbf{y}) := [y_{I \cup J}]_{ij}\ \forall I, J \subseteq [n], |I|, |J| \leqslant t$ and 0 otherwise. To enforce the other constraints we define a slack matrix $M_l^t(\mathbf{y}) := \sum_{i=1}^n b_l y_{I \cup J} - A_{li} y_{I \cup J \cup \{i\}}, \forall l \in [m]$.

We say that a vector $\mathbf{y} \in \mathsf{LAS}^t(\mathcal{P})$ is a solution to the SDP relaxation given by ,

SDP 2.4.

$$\min\ \mathbf{c}^T \mathbf{y}$$

subject to

$$M^t(\mathbf{y}) \succeq 0 \tag{8}$$

$$M_l^t(\mathbf{y}) \succeq 0 \qquad \forall l \in [m] \tag{9}$$

$$y_\phi = 1 \tag{10}$$

where $\mathbf{c}$ is defined to be a vector in $\mathbb{R}^{2^n}$ (so that we have a valid inner product) but it is non-zero only in the first $n+1$ entries corresponding to $y_{\{i\}}$. We note that although $\mathbf{y} \in \mathbb{R}^{2^n}$ we know that $\mathbf{y}_I = 0, \forall |I| \geqslant 2t + 1$. After we obtain a $\mathbf{y} \in \mathsf{LAS}^t(\mathcal{P})$ we project back onto original decision variables $x_i = y_i$ as,

$$\mathsf{LAS}^t_{\mathsf{proj}}(\mathcal{P}) = \left\{ \left(y_{\{1\}}, y_{\{2\}}, \ldots, y_{\{n\}}\right) : \mathbf{y} \in \mathsf{LAS}^t(\mathcal{P}) \right\}$$

We note that in our discussion we start with an LP and then we described the SA/Lasserre relaxation; however it can be done more generally for polynomial optimization problems (we refer to [FKP19] for more details).

2.3.2 Properties of Lasserre

We now prove a few basic yet important properties of the solution $\mathbf{y} \in \mathsf{LAS}^t(\mathcal{P})$ returned by SDP 2.4 First we show that the additional constraints (8)-(10) hold true for integral solutions and that SDP 2.4 is a valid relaxation.

Claim 2.5. $\mathcal{I} \subseteq \mathsf{LAS}^t_{\mathsf{proj}}$

Proof. We show that given any feasible solution $\mathbf{x} \in \mathcal{I}$ for the original LP we can construct a feasible solution $\mathbf{y}$ to the SDP 2.4. We let $\mathbf{y}$ such that $y_I = \prod_{i \in I} x_i$ and consider $M^n(\mathbf{y})$ then,

$$[M^n(\mathbf{y})]_{ij} = y_{I \cup J} = \prod_{i \in I \cup J} x_i = \prod_{i \in I \Delta J} x_i \prod_{i \in I \cap J} x_i = \prod_{i \in I \Delta J} x_i \left(\prod_{i \in I \cap J} x_i \right)^2 = \prod_{i \in I} x_i \prod_{j \in J} x_j = y_I y_J$$

where we have used the fact that $x_i = x_i^2$ since $\mathbf{x} \in \mathcal{I}$ and hence $x_i \in \{0, 1\}$. This shows that $M^n(\mathbf{y})$ takes the form $\mathbf{y}\mathbf{y}^T$ and hence $M^n(\mathbf{y}) \succeq 0$. Now since $M^t(\mathbf{y})$ is a principal submatrix of $M^n(\mathbf{y})$, we have that $M^t(\mathbf{y}) \succeq 0$ and constraint (8) holds.

We similarly show that $M_l^n(\mathbf{y}) \succeq 0$ as,

$$[M_l^n(\mathbf{y})]_{ij} = \sum_{i=1}^{n} b_l y_{I \cup J} - A_{li}.y_{I \cup J \cup \{i\}} = \sum_{i=1}^{n} b_l y_I y_J - A_{li}.y_I y_J y_{\{i\}} = y_I y_J (b_l - A_l \mathbf{x})$$

where again we use $x_i = x_i^2$ and using the fact that $b_l - A_l \mathbf{x} \geqslant 0$ we show that $M_l^n(\mathbf{y}) \succeq 0$ and hence $M_l^t(\mathbf{y}) \succeq 0$. □

Next we show that SDP 2.4 is truly a relaxation of the original LP as,

Claim 2.6. $\mathsf{LAS}^t_{\mathsf{proj}}(\mathcal{P}) \subseteq \mathcal{P}$

Proof. Let $\mathbf{y} \in \mathsf{LAS}^t(\mathcal{P})$, then using the constraint (9) with $I = J = \phi$ we get,

$$\sum_{i=1}^{n} b_l y_\phi - A_{li} y_{\phi \cup \{i\}} \geqslant 0$$

Now since $y_\phi = 1$ we have that $\sum_{i=1}^{n} b_l - A_{li} y_i \geqslant 0$ and hence $\mathsf{LAS}^t_{\mathsf{proj}}(\mathcal{P}) = (y_1, \ldots, y_n) \in \mathcal{P}$. □

Claim 2.7. *a)* $0 \leqslant y_I \leqslant 1, \forall I \subseteq [n], |I| \leqslant t$ *b)* $y_I \leqslant y_J, \forall I \subseteq J \subseteq [n], |I|, |J| \leqslant t$

Proof. a) Consider the submatrix indexed by $\{I, J\}$, $\begin{bmatrix} y_\phi & y_I \\ y_I & y_I \end{bmatrix}$. Now since the submatrix is a principal submatrix of $M^t(\mathbf{y})$ it must be PSD and have $|.| \geqslant 0$. Hence $\begin{vmatrix} y_\phi & y_I \\ y_I & y_I \end{vmatrix} \geqslant 0$.

b) Consider the submatrix indexed by $\{I, J\}$, $\begin{bmatrix} y_I & y_J \\ y_J & y_J \end{bmatrix}$ where we have used the fact that $I \cup J = J$ since $I \subseteq J$. Now since the submatrix is a principal submatrix of $M^t(\mathbf{y})$ it must be PSD and have $|.| \geqslant 0$. Hence $\begin{vmatrix} y_I & y_J \\ y_J & y_J \end{vmatrix} \geqslant 0$. Hence $y_I y_J - y_J y_J = y_J(y_I - y_J) \geqslant 0$. Since we already know form (a) that $y_J \geqslant 0$ we have that $y_I \geqslant y_J$.

□

We can rewrite the claim above for a degree-1 SOS as $y_{\{i\}} \geqslant 0$ and $1 - y_{\{i\}} \geqslant 0$ which were our constraints due to junta polynomials in level-1 SA hierarchy. One can show that this is true in general and a level-t Lasserre relaxation is atleast as expressive as level-t SA relaxation.

2.3.3 Evolution of Lasserre hierarchy

We know that for any $\mathbf{x} \in \mathcal{I}$, we can write $\mathbf{x}$ as a convex combination of integral solutions. This is not true for a level-t Lasserre relaxation. However we can still show (Corollary 2.10) that for a level-t Lasserre we can write $\mathbf{y} \in \mathsf{LAS}^t(\mathcal{P})$ as a convex combination of solutions of $\mathcal{P}$ in which t variables become integral. To this effect we first define a one step integral vector $\mathbf{z}$ from $\mathbf{y}$ as,

Definition 2.8. We define $\mathbf{z}^{(0)}$ to be a vector $\mathbf{y}$ where the i^{th} variable is 0 i.e $\mathbf{z}_i^{(0)} = 0$ and $\mathbf{z}^{(1)}$ to be a vector $\mathbf{y}$ where the i^{th} variable is 1 i.e $\mathbf{z}_i^{(1)} = 1$. We can then write this from a vector $\mathbf{y}$ as,

$$\mathbf{z}_I^{(0)} = y_I \frac{(1-y_i)}{(1-y_i)} = \frac{y_I - y_{I\cup\{i\}}}{1-y_i} \qquad \mathbf{z}_I^{(1)} = y_I \frac{(y_i)}{(y_i)} = \frac{y_{I\cup\{i\}}}{y_i}$$

Lemma 2.9. *For $\mathbf{y} \in \mathsf{LAS}^t(\mathcal{P})$ we can write as, $\mathbf{y} = y_i\mathbf{z}^{(1)} + (1-y_1)\mathbf{z}^{(0)}$ where $\mathbf{z}^{(0)}, \mathbf{z}^{(1)} \in \mathsf{LAS}^{t-1}(\mathcal{P})$.*

Proof. Using the definition of $\mathbf{z}$ in Definition 2.8 it is straightforward to see that $\mathbf{y} = y_i\mathbf{z}^{(1)} + (1-y_1)\mathbf{z}^{(0)}$. The non-trivial part is showing that $\mathbf{z}^{(0)}, \mathbf{z}^{(1)} \in \mathsf{LAS}^{t-1}(\mathcal{P})$. We start from the fact that if $\mathbf{y} \in \mathsf{LAS}^t(\mathcal{P})$ we have that $M^t(\mathbf{y}) \succeq 0$ and by Cholesky factorization there exists vectors $\{v_I\}_{I\subseteq[n],|I|\leqslant t}$ such that $y_{I\cup J} = \langle v_I, v_J\rangle$. We need to show that $M^{t-1}(\mathbf{z}^{(1)}) \succeq 0$ and if we let $v_I^{(1)} = \frac{v_{I\cup\{i\}}}{\sqrt{y_i}}, \forall I \subseteq [n], |I| \leqslant t-1$, we have that,

$$z_{I\cup J}^{(1)} = \frac{y_{I\cup J\cup\{i\}}}{y_i} = \frac{\left\langle v_{I\cup\{i\}}, v_{J\cup\{j\}}\right\rangle}{y_i} = \left\langle v_I^{(1)}, v_J^{(0)}\right\rangle$$

. Similarly we can set $v_I^{(0)} = \frac{v_I - v_{I\cup\{i\}}}{\sqrt{1-y_i}}, \forall I \subseteq [n], |I| \leqslant t-1$ and we can show that $M^{t-1}(\mathbf{z}^{(0)}) \succeq 0$. □

Corollary 2.10. *For any $\mathbf{y} \in \mathsf{LAS}^t(\mathcal{P})$ and $S \subseteq [n], |S| \leqslant t$ we can write,*

$$\mathbf{y} \in \mathsf{conv}\left\{\mathbf{z} \in \mathsf{LAS}^{t-|S|}(\mathcal{P}) : z_i \in \{0,1\}\ \forall i \in S\right\}$$

Proof. Apply Lemma 2.9 inductively $t - |S|$ times. □

This again confirms the fact that after n rounds of hierarchy we get the exact ingeral solution, $\mathsf{LAS}^n_{\mathsf{proj}} = \mathcal{I}$.

2.3.4 Utility of Lasserre hierarchy

We consider the problem of finding an independent set and can write the LAS^t SDP relaxation as,

SDP 2.11.

$$\max \ \|\mathbf{y_i}\|^2$$

subject to

$$\|\mathbf{y}_e\|^2 = 0 \qquad \forall e \in E \qquad (11)$$

$$\langle \mathbf{y}_I, \mathbf{y}_J\rangle = \langle \mathbf{y'_I}, \mathbf{y'_J}\rangle \qquad \forall I, J, I', J' \subseteq [n], |I|, |I'|, |J|, |J'| \leqslant t, I\cup J = I' \cup J' \qquad (12)$$

The SDP 2.11 was analyzed for $t = 3$ by Eden Chlamtac in the work [Chl07] where they round this to give $\Omega\big(n^{1/2-\gamma}\big)$ sized independent set for 3-uniform hypergraphs. This is an improvement over the basic SDP based algorithm by Krivelevich, Nathaniel and Sudakov [KNS01] that finds an independent set of size $\tilde{\Omega}\big(\min\big(n, n^{6\gamma-3}\big)\big)$ for $\gamma \geqslant 1/2$ in 3-uniform hypergraphs.

Infact the work by Prasad Rraghavendra [Rag08] shows that a level-2 Lasserre hierarchy gives the optimal SDP (also commonly called the canonical SDP) for all CSP's under Unique Games conjecture.

3 Sum of Squares on a Hypercube

This section and the following sections are based on [BS16].

3.1 Sum of Squares Proofs

In this section we shift our attention to a seemingly unrelated question of showing a function is non-negative. The sum of squares paradigm attempt to achieve this feat by (as its name implies) writing the function as sum of square of polynomials. For simplicity we shall restrict ourselves to $\mathbb{R}^{\{0,1\}^n} = \mathcal{B}$ which are sometimes called Boolean functions.

Before we start we note a few things about Boolean functions. Firstly, in this domain $x_i^k = x_i$ for any $k \geqslant 1$. Secondly, any $f \in \mathcal{B}$ can be uniquely represented as[3] $f = \sum_{S \subset [n]} c_S x_S$ where $x_S = \prod_{i \in S} x_i$. These two facts will be extensively used to obtain the "right form" for the function in any given context.

Definition 3.1. A degree-d sum of squares (sos) proof for $f \geqslant 0$ is a collection of $d/2$ degree polynomial $p_1, p_2, ..., p_r$ for some r such that $f = \sum_{i \in [r]} p_i^2$. We write $\vdash_d f \geqslant 0$ if such a proof exists.

We start by showing an interesting (though not quite useful) fact about sos proofs,

Lemma 3.2. *If $f \geqslant 0$ then $\vdash_{2n} f \geqslant 0$.*

Proof. Define $g \in \mathcal{B}$ by

$$g(x) = \sum_{y \in \{0,1\}^n} \sqrt{f(y)} \prod_{i | y_i = 1} x_i \prod_{j | y_j = 0} (1 - x_j)$$

Then g is degree $d/2$ polynomial such that $g = \sqrt{f}$, i.e, $f = g^2$ as $\mathbf{1}_y(x) = \prod_{i|y_i=1} x_i \prod_{j|y_j=0}(1 - x_j)$. □

Though this shows sos proofs are possible for any non-negative function this does not make the task of constructing sos proofs easy as the above method evaluates the function at all points which could be done anyway. In fact we would like a constant degree proof with a small number of polynomials. This will allow us to describe the proof with a few numbers and possible get a fast algorithm.

For remainder of the section let $v_d(x)$ denote an $\binom{n}{\leqslant d}$ dimensional vector with $v_d(x)_S = \prod_{i \in S} x_i$. With this we can represent any degree d polynomial as a $\langle v, v_d \rangle$ for some vector v.

Lemma 3.3. *$\vdash_d f \geqslant 0$ if and only if there exists a psd matrix A such that $\langle v_{d/2}(x), A v_{d/2}(x) \rangle$.*

Proof. Suppose $\vdash_d f \geqslant 0$ then there are $d/2$ degree polynomials $\{p_l\}_{l \in [r]}$ such that $f = \sum_{l \in [r]} p_l^2$. Let $v_l \in \mathbb{R}^{\binom{n}{\leqslant \frac{d}{2}}}$ be vectors such that $p_l(x) = \langle v_l, v_{d/2}(x) \rangle$. Taking $A = \sum_{l \in [r]} v_l v_l^T$ we have,

$$f(x) = \sum_{l \in [r]} p_l^2(x) = \sum_{l \in [r]} \langle v_l, v_{d/2}(x) \rangle^2 = \sum_{l \in [r]} v_{d/2}(x)^T v_l v_l^T v_{d/2}(x) = v_{d/2}(x)^T A v_{d/2}(x)$$

as required.

Now, if A is psd matrix such that $f(x) = \langle v_{d/2}(x), A v_{d/2}(x) \rangle$ then taking $g_S(x) = \langle e_S, A^{1/2} v_{d/2}(x) \rangle$ gives us,

$$\begin{aligned} f(x) = \langle v_{d/2}(x), A v_{d/2}(x) \rangle &= \langle A^{1/2} v_{d/2}(x), A^{1/2} v_{d/2}(x) \rangle \\ &= \sum_{\substack{S \subset [n] \\ |S| \leqslant d/2}} \langle A^{1/2} v_{d/2}(x), e_S \rangle \langle e_S, A^{1/2} v_{d/2}(x) \rangle \\ &= \sum_{\substack{S \subset [n] \\ |S| \leqslant d/2}} g_l^2(x) \end{aligned}$$

□

The most important consequence of the above lemma is, if $f \geqslant 0$ is a polynomial of degree d (or can be simplified to such using $x_i^k = x_i$) then we can get a sos proof of degree d by solving the following SDP:

[3]This is commonly known as Fourier expansion

SDP 3.4.

$$\sum_{S_1 \cup S_2 = S} A_{S_1 S_2} = c_S \qquad \forall S \subseteq [n], |S| \leqslant d/2$$
$$A \succeq 0$$

We also get another consequence by noticing that if we have $\vdash_d f \geqslant 0$ then we can get a psd matrix as above from which we can write f as sum of squares of $\binom{n}{\leqslant d/2}$ polynomial. In essence we have the following corollary,

Corollary 3.5. *If* $\vdash_d f \geqslant 0$ *then there are* $\binom{n}{\leqslant d/2}$ $d/2$ *degree polynomials* $\{p_S\}_{\substack{S \subset [n] \\ |S| \leqslant d/2}}$ *such that* $f = \sum_{\substack{S \subset [n] \\ |S| \leqslant d/2}} p_S^2$.

Proof. Since $\vdash_d f \geqslant 0$, therefore there is psd matrix such that,

$$f(x) = \langle v_{d/2}, A v_{d/2} \rangle = \sum_{\substack{S \subset [n] \\ |S| \leqslant d/2}} \langle A^{\frac{1}{2}} v_{d/2}, e_S \rangle \langle e_S, A^{\frac{1}{2}} v_{d/2} \rangle = \sum_{\substack{S \subset [n] \\ |S| \leqslant d/2}} \langle e_S, A^{\frac{1}{2}} v_{d/2} \rangle^2$$

Thus, taking $p_S = \langle e_S, A^{\frac{1}{2}} v_{d/2} \rangle$ we get $f = \sum_{\substack{S \subset [n] \\ |S| \leqslant d/2}} p_S^2$. □

In this unconstrained setting we can in fact (try to) find the maximum (minimum) value for a function f by minimizing (maximizing) c such that $\vdash_d c - f \geqslant 0$ ($\vdash_d f - c \geqslant 0$). In essence we will be solving,

SDP 3.6.

$$\min c$$

subject to

$$\sum_{S_1 \cup S_2 = S} A_{S_1 S_2} = -c_S \qquad \forall S \in 2^{[n]} - \phi, |S| \leqslant d/2$$
$$A_{\phi\phi} = c - c_\phi$$
$$A \succeq 0$$

Now as a prelude to the next part of this section we observe that functions with degree d proof form a convex cone (set).

Lemma 3.7. $K_d = \{f \mid \vdash_d f \geqslant 0\}$ *is a convex cone.*

Proof. For any $f_1, f_2 \in K_d$ and $\alpha, \beta \geqslant 0$, we can write $f_1 = \sum_i p_i^2$ and $f_2 = \sum_i q_i^2$. Therefore,

$$\alpha f_1 + \beta f_2 = \sum_i \alpha p_i^2 + \sum_i \beta q_i^2 = \sum_i (\sqrt{\alpha} p_i)^2 + \sum_i (\sqrt{\beta} q_i)^2$$

□

Finally we end this discussion on sos proofs with the following result,

Lemma 3.8. *If* $f \in \mathcal{B}$ *has a degree* d *polynomial representation then,* $\exists M_d$ *such that* $\vdash_{2d} M_d - f \geqslant 0$

Proof. f can be written as $f = \sum_{\substack{S \subset [n] \\ |S| \leqslant d}} c_S x_S$ (see Appendix A). Taking $M_d = \sum_{\substack{S \subset [n] \\ |S| \leqslant d}} c_S m_S$ where,

$$m_S = \begin{cases} 1 & c_S > 0 \\ 0 & c_S \leqslant 0 \end{cases}$$

So,

$$M_d - f = \sum_{\substack{S\subset[n]\\|S|\leqslant d}} c_S m_S - c_S x_S = \sum_{S|c_S<0} -c_S x_S + \sum_{S|c_S>0} c_S(1-x_S)$$
$$= \sum_{S|c_S<0} (\sqrt{-c_S}x_S)^2 + \sum_{S|c_S>0} c_S(1-2x_S+x_S)$$
$$= \sum_{S|c_S<0} (\sqrt{-c_S}x_S)^2 + \sum_{S|c_S>0} (\sqrt{c_S}(1-x_S))^2$$

□

3.2 Pseudo-Distribution and Pseudo-Expectation

In the previous part of to this section we noted that the set of functions having a degree d sos proofs form a convex cone (set). This leads us to an interesting observation (using hyperplane separation theorem) when $\not\vdash_d f \geqslant 0$, which is, $\exists \mu \in \mathcal{B}$ such that $\langle \mu, f\rangle = \sum_x \mu(x)f(x) < 0$ while $\langle \mu, p\rangle \geqslant 0$ for every $p \in K_d$. Roughly speaking the function is called a pseudodistribution as it behaves like an probability distribution w.r.t a quantity called pseudoexpectation.

Let us now form these notions of pseudodistribution and pseudoexpectation more carefully. Firstly, we call the following formal expectation w.r.t $\mu \in \mathcal{B}$ is,

$$\tilde{\mathbb{E}}_{\mu} f = \langle \mu, f\rangle$$

With this we can define pseudodistribution as,

Definition 3.9. $\mu \in \mathcal{B}$ is called a degree d pseudodistribution if

1. $\tilde{\mathbb{E}}_\mu 1 = 1$
2. $\tilde{\mathbb{E}}_\mu f^2 \geqslant 0$ for polynomial of degree at most $d/2$.

and $\tilde{\mathbb{E}}_\mu$ is called a pseudoexpectation.

Note that if $\mu \geqslant 0$ then it is a probability distribution and also by Lemma 3.2 degree $2n$ pseudodistribution are probability distributions. As a further justification of the name we state the Cauchy-Schwartz inequality for pseudodistribution,

Lemma 3.10 (Cauchy-Schwartz Inequality). *If μ is a degree d pseudodistribution, then for any degree $d/2$ polynomials p, q,*

$$\left(\tilde{\mathbb{E}}_{\mu} pq\right)^2 \leqslant \tilde{\mathbb{E}}_{\mu} p^2 \tilde{\mathbb{E}}_{\mu} q^2$$

and we also note,

Lemma 3.11. *If μ is a degree $2n$ distribution then $\mu \geqslant 0$, i.e, it is a probability distribution.*

The main take away here is pseudoexpectations behave exactly like a true expectation when fed a low degree polynomial.

Now we prove the theorem which was the initial motivation for the Definition 3.9,

Theorem 3.12 (duality of sos and pseudodistribution). *For every function f of degree at most d, $\vdash_d f \geqslant 0$ if and only if $\tilde{\mathbb{E}}_\mu f \geqslant 0$ for all degree d pseudodistribution.*

Proof. Suppose $\vdash_d f \geqslant 0$ then there are degree $d/2$ polynomials g_i such that $f = \sum_i g_i^2$. Therefore,

$$\tilde{\mathbb{E}}_{\mu} f = \tilde{\mathbb{E}}_{\mu} \sum_i g_i^2 = \sum_i \tilde{\mathbb{E}}_{\mu} g_i^2 \geqslant 0$$

Now, suppose $\not\vdash_d f \geqslant 0$ then $f \notin K_d$ (cone of d-sos provable functions) which means by Hyperplane Separation Theorem, there is a $\mu \in \mathcal{B}$ such that,

1. $K_d \subset \left\{ g | \tilde{\mathbb{E}}_\mu g = \langle \mu, f \rangle \geqslant 0 \right\}$, i.e, K_d is contained in one side of the hyperplane.
2. $\tilde{\mathbb{E}}_\mu f = \langle \mu, f \rangle < 0$, i.e, f is in the other side of the hyperplane.

Since, f has degree d by Lemma 3.8 we must have some M such that $\vdash_d M + f \geqslant 0$. Therefore,

$$\tilde{\mathbb{E}}_{\mu} 1 = \frac{1}{M} \tilde{\mathbb{E}}_{\mu} M = \frac{1}{M} \left[\tilde{\mathbb{E}}_{\mu}(M+f) - \tilde{\mathbb{E}}_{\mu} f \right] > 0$$

Hence, by appropriate re-scaling of μ we can obtain a pseudodistribution. □

We now turn to showing a result which in spirit mirrors Lemma 3.3,

Lemma 3.13. *μ is a degree d pseudodistribution if and only if*

1. $\tilde{\mathbb{E}}_\mu 1 = 1$
2. $\tilde{\mathbb{E}}_\mu v_{d/2}(x) v_{d/2}(x)^T \succeq 0$

Proof. Suppose μ is a degree d pseudo-distribution, $\tilde{\mathbb{E}}_\mu 1 = 1$ by definition. For any $v \in \mathbb{R}^{\binom{n}{\leqslant d/2}}$ we have a corresponding degree $d/2$ polynomial $p(x) = \langle v, v_{d/2}(x) \rangle$. Again, by definition $\tilde{\mathbb{E}}_\mu p^2 \geqslant 0$. But,

$$p^2(x) = \langle v, v_{d/2}(x) \rangle^2 = v^T v_{d/2}(x) v_{d/2}(x)^T v$$

. So, $v^T \tilde{\mathbb{E}}_\mu v_{d/2}(x) v_{d/2}(x)^T v \geqslant 0$. Which means $\tilde{\mathbb{E}}_\mu v_{d/2}(x) v_{d/2}(x)^T \succeq 0$.

Similarly if $\tilde{\mathbb{E}}_\mu v_{d/2}(x) v_{d/2}(x)^T \succeq 0$ then for any v we get $\tilde{\mathbb{E}}_\mu \langle v, v_{d/2}(x) \rangle^2$. As any degree $d/2$ polynomial is equal to some multilinear degree $d/2$ polynomial we have for any degree $d/2$ polynomial p, $\tilde{\mathbb{E}}_\mu p^2 \geqslant 0$. Therefore along with the condition $\tilde{\mathbb{E}}_\mu 1 = 1$, μ is a degree d pseudo-distribution. □

As we noted earlier only the first d moments of a pseudo-distribution are relevant(See Theorem 3.12). Therefore we can again use the following SDP to characterize pseudo-distributions.

SDP 3.14.

$$A_{\phi\phi} = 1$$

$$A \succeq 0 \qquad\qquad A \in \mathbb{R}^{\binom{n}{\leqslant d/2} \times \binom{n}{\leqslant d/2}}$$

where each entry $A_{S_1 S_2}$ gives the $|S_1 \cup S_2|$ moment $\tilde{\mathbb{E}}_\mu x_{S_1} x_{S_2} = \tilde{\mathbb{E}}_\mu x_{S_1 \cup S_2}$ for some degree d pseudo-distribution. Before looking at the optimization variant consider the following fact with M_d as the set of degree d pseudo-distribution and $\mathcal{D}$ the set of actual distributions.

Observation 3.15. *$M_d \supset M_{d+2}$ and $M_{2n} = \mathcal{D}$, i.e, $M_2 \supset M_4 \supset \ldots \supset M_{2n} = \mathcal{D}$*

Proof. The containment is by definition and the equality is a restatement of Lemma 3.11. □

Due to the above observation we see $\max_x f \leqslant \max_{\mu \in \mathcal{D}} \mathbb{E}_\mu f \leqslant \max_{\mu \in M_d} \tilde{\mathbb{E}}_\mu f$ (similarly for minimization). As we will see it will be worthwhile to consider the following optimization

SDP 3.16.

$$\max \sum_S A_S c_S$$

subject to

$$A_{\phi\phi} = 1$$

$$A \succeq 0 \qquad A \in \mathbb{R}^{\binom{n}{\leqslant d/2} \times \binom{n}{\leqslant d/2}}$$

Since $A_{S_1 S_2} = \tilde{\mathbb{E}}_\mu x_{S_1 \cup S_2}$ we can refer to $A_{S_1 S_2}$ by $A_{S_1 \cup S_2}$.

To find a pseudo-distribution with the given moments we consider the following statement,

Lemma 3.17. *Let μ be a degree d pseudo distribution then there exists a multilinear polynomial μ' of degree at most d such that,*

$$\tilde{\mathbb{E}}_{\mu} p = \tilde{\mathbb{E}}_{\mu_d} p$$

for every polynomial p of degree at most d.

Proof. Let $\mathcal{B}_d$ be the subspace of all possible degree d polynomial. We know $\mathcal{B}_d$ is spanned by the set of multilinear polynomial (See *Appendix A*). Write μ as $\mu = \mu_d + \mu_\perp$ where $\mu_d \in \mathcal{B}_d$ and $\mu_\perp \in \mathcal{B}_d^\perp$. So, if $p \in \mathcal{B}_d$ then,

$$\tilde{\mathbb{E}}_{\mu} p = \langle \mu, p \rangle = \langle \mu_d + \mu_\perp, p \rangle = \langle \mu_d, p \rangle = \tilde{\mathbb{E}}_{\mu_d} p$$

□

Using this lemma we know that degree d polynomial μ with required moments. Say, $\mu = \sum_{|S| \leqslant d} c_S x_S$ then we need to solve for the equations $\sum_{|S'| \leqslant d} c_{S'} |S \cap S'| = \langle \mu, x_S \rangle = \tilde{\mathbb{E}}_\mu x_S = A_S$.

We end this section by noting that the SDP 3.16 is also called the Lassere SDP [4]. Furthermore, SOS in its full generality can use additional non-negativity constraints.

4 Applications of SOS

In this section we will look at application of SOS in finding approximation algorithms for NP-complete problems. In these algorithms instead finding the optimum solution we seek to find some solution which is guaranteed to be close the optimum. More formally,

Definition 4.1. An α-approximation algorithm for an optimization is a polynomial-time algorithm that (for all instances of the problem) produces a solution whose value is within a factor of α of the optimum value. In essence, the obtained value for the solution to maximization (minimization) problem is a lower (upper) bound for the maximum (minimum).

In most cases we can formulate such problems as optimization of a class of function on the 0-1 hypercube. The general approach to solving a maximization problem in this framework to show that for any function f in the given class we have a degree d sos proof for $c - f \geqslant 0$ where $\alpha c \leqslant \max_x f$, i.e, if solve *SDP* 3.6 we get a lower bound αc for the maximum value.

Alternately, we can use the dual approach, i.e, we can show that for any degree d pseudo-distribution μ we can (try to) find an actual distribution ρ such that,

$$\mathbb{E}_{\rho} f \geqslant \alpha \tilde{\mathbb{E}}_{\mu} f$$

then by finding a degree d pseudo-distribution μ which maximizes $\tilde{\mathbb{E}}_\mu f$ (by solving SDP 3.16) we can find an actual distribution whose expected value is bounded by $\alpha \max_x f$. Therefore we can obtain a solution

[4] in our initial discussion about Lassere we had considered only linear objectives but linearity is not necessary

which is fairly close to the optimum by sampling the distribution. This approach is commonly referred to as “rounding”.

For a concrete example we look at the MAX-CUT problem for we will show how to obtain the best known approximation ratio (and best possible under a fairly plausible conjecture called Unique Games Conjecture). For the following tool which is in itself quite powerful,

Lemma 4.2 (quadratic sampling lemma)**.** *For every degree-2 pseudo distribution μ over $\{0,1\}^n$, there is a multivariate Gaussian ρ with the same first two moments, i.e,*

$$\tilde{\mathbb{E}}_{\mu} v_2(x) = \mathbb{E}_{\rho} v_2(x)$$

which can be sampled in polynomial time.

Proof. Let $v = \tilde{\mathbb{E}}_{\mu} x$ and $\Sigma = \mathbb{E}_{\mu}(x-v)(x-v)^T$ be the mean vector and covariance matrix for μ. By taking any $u \in \mathbb{R}^n$ we can see that,

$$\langle u, \Sigma u\rangle = \tilde{\mathbb{E}}_{\mu} \langle u, v-x\rangle^2 \geqslant 0$$

i.e, Σ is p.s.d. Generate a Gaussian ρ with mean v and covariance Σ using the following algorithm,

1. sample a standard Gaussian vector g, i.e, a vector whose component is chosen independently from $\mathcal{N}(0,1)$.
2. output $y = v + \Sigma^{1/2} g$

Indeed, as $\mathbb{E}\, g = 0$ and $\mathbb{E}\, gg^T = I$ therefore, $\mathbb{E}_{x\sim\rho}\, x = \mathbb{E}\, y = v$ and

$$\mathbb{E}_{x\sim\rho}(x-v)(x-v)^T = \mathbb{E}(y-v)(y-v)^T = \mathbb{E}\,\Sigma^{1/2} gg^T (\Sigma^{1/2}) = \Sigma$$

as required. □

Now, before proceeding we need to formulate the MAX-CUT problem in terms of optimization over $\{0,1\}^n$. This is done by considering the function $f_G(x) = \sum_{\{i,j\}\in E}(x_i - x_j)^2$ which gives $E(S, \overline{S})$ at $\mathbf{1}_S$.

Theorem 4.3. *Let μ be a degree 2 pseudo distribution then there is randomized polynomial time procedure to sample a distribution μ' such that for a given graph G,*

$$\mathbb{E}_{\mu'} f_G \geqslant 0.878\, \tilde{\mathbb{E}}_{\mu} f_G$$

Proof. Assume $\tilde{\mathbb{E}}_{\mu}\mathbf{1} = \frac{1}{2}\mathbf{1}$, otherwise $\frac{1}{2}(\mu(x) + \mu(\mathbf{1}-x))$ can be considered instead. Now generate μ' as follows,

1. choose a standard Gaussian g with matching first and second moments.
2. output $x' \in \{0,1\}^n$ where,

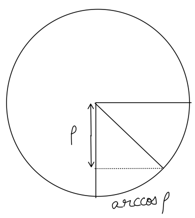
$$x'_i = \begin{cases} 0 & g_i < 1/2 \\ 1 & g_i \geqslant 1/2 \end{cases}$$

Since the expectation operators are linear so showing the following is sufficient,

$$\mathbb{E}_{x\sim\mu'}(x_i - x_j)^2 \geqslant 0.878 \tilde{\mathbb{E}}_{x\sim\mu}(x_i - x_j)^2$$

Now, note that the variance of x_i and x_j are $\mathbb{E}_\mu x_i^2 - \frac{1}{4} = \mathbb{E}_\mu x_j^2 - \frac{1}{4} = \frac{1}{4}$ and let $\rho = 4\,\mathbb{E}_\mu x_i x_j - 1$. Since, g has the same first two moments as x therefore we can write $g_j = \rho g_i + \sqrt{1-\rho^2} g_i^\perp$ where $g_i^\perp \sim \mathcal{N}(0,1)$ is a Gaussian independent of g_i. This gives,

$$\begin{aligned}\mathbb{E}(x_i' - x_j')^2 &= \mathbb{P}\left[\operatorname{sign}(2g_i - 1) \neq \operatorname{sign}(2g_j - 1)\right] \\ &= \underset{s,t\sim\mathcal{N}(0,1)}{\mathbb{P}}\left[\operatorname{sign}(s) \neq \operatorname{sign}(\rho s + \sqrt{1-\rho^2}t)\right]\end{aligned}$$



The last probability can be calculated by considering the fact that (s,t) is an uniform point on the circle of radius 1. As the figure shows it turns out to be $\frac{\arccos\rho}{\pi}$ Now observe that,

$$\underset{\mu}{\tilde{\mathbb{E}}}(x_i - x_j)^2 = \frac{1}{2}(1-\rho)$$

One can easily verify that,

$$\inf_\rho \frac{2\arccos\rho}{\pi(1-\rho)} = 0.878$$

$\square$

References

[ARV04] Sanjeev Arora, Satish Rao, and Umesh Vazirani, *Expander flows, geometric embeddings and graph partitioning*, Proceedings of the 36th Annual ACM Symposium on Theory of Computing, ACM, New York, 2004, pp. 222–231. MR 2121604 1

[BS16] Boaz Barak and David Steurer, *Proofs, beliefs, and algorithms through the lens of sum-of-squares*, 2016. 6

[Chl07] E. Chlamtac, *Approximation algorithms using hierarchies of semidefinite programming relaxations*, 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), 2007, pp. 691–701. 6

[CT12] Eden Chlamtac and Madhur Tulsiani, *Convex relaxations and integrality gaps*, Handbook on semidefinite, conic and polynomial optimization, Internat. Ser. Oper. Res. Management Sci., vol. 166, Springer, New York, 2012, pp. 139–169. MR 2894694 1

[FKP19] Noah Fleming, Pravesh Kothari, and Toniann Pitassi, *Semialgebraic proofs and efficient algorithm design*, Foundations and Trends® in Theoretical Computer Science **14** (2019), no. 1-2, 1–221. 1, 4

[KNS01] Michael Krivelevich, Ram Nathaniel, and Benny Sudakov, *Approximating coloring and maximum independent sets in 3-uniform hypergraphs*, J. Algorithms **41** (2001), no. 1, 99–113. MR 1855351 6

[Las01] Jean B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM J. Optim. **11** (2000/01), no. 3, 796–817. MR 1814045 1, 4

[Lau03] Monique Laurent, *A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming*, Math. Oper. Res. **28** (2003), no. 3, 470–496. MR 1997246 1

[LLZ02] Michael Lewin, Dror Livnat, and Uri Zwick, *Improved rounding techniques for the MAX 2-SAT and MAX DI-CUT problems*, Integer programming and combinatorial optimization, Lecture Notes in Comput. Sci., vol. 2337, Springer, Berlin, 2002, pp. 67–82. MR 2061046 1

[LS91] L. Lovász and A. Schrijver, *Cones of matrices and set-functions and* 0-1 *optimization*, SIAM J. Optim. **1** (1991), no. 2, 166–190. MR 1098425 1

[Par03] Pablo A Parrilo, *Semidefinite programming relaxations for semialgebraic problems*, Mathematical programming **96** (2003), no. 2, 293–320. 4

[Rag08] Prasad Raghavendra, *Optimal algorithms and inapproximability results for every CSP? [extended abstract]*, STOC'08, ACM, New York, 2008, pp. 245–254. MR 2582901 6

[Rot13] Thomas Rothvoß, *The lasserre hierarchy in approximation algorithms – Lecture Notes for the MAPSP Tutorial*, 2013, `https://sites.math.washington.edu/~rothvoss/lecturenotes/lasserresurvey.pdf`. 1

[SA90] Hanif D. Sherali and Warren P. Adams, *A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems*, SIAM J. Discrete Math. **3** (1990), no. 3, 411–430. MR 1061981 1

[Sho87] NZ Shor, *An approach to obtaining global extremums in polynomial mathematical programming problems. kibernetika 5 102–106.. 1998*, Nondifferentiable Optimization and Polynomial Problems (1987). 4

A Polynomials in Hypercube

In the hypercube we can represent any function by a multilinear linear as $x_S(x) = \prod_{i \in S} x_i$ form a basis for the hypercube.

Lemma A.1. *$\{x_S\}_{S \subset [n]}$ forms a basis for $\mathcal{B}$, where $\mathcal{B} = \mathbb{R}^{\{0,1\}^n}$ is the set of Boolean function. Hence, if $f \in \mathbf{B}$, then $f = \sum_{S \subset [n]} c_S x_S$ for some c_S.*

Proof. Consider any linear combination summing to 0, i.e, $\sum_{S' \subset [n]} c_{S'} x_{S'} = 0$. Evaluating the same at 1_S gives us $\sum_{S' \subset S} c_S = 0$. Now, using induction on $|S|$, we can get $c_S = 0$. Hence x_S is linearly independent.

Since there are 2^n such functions they must form a basis. □

Using the fact that $x_i^k = x_i$ in the hypercube we have $\prod_i x_{S_i} = x_S$ where $S = \bigcup_i S_i$ we get the following corollary,

Corollary A.2. *$\{x_S | S \subset [n], |S| \leqslant d\}$ forms a basis for degree d polynomials in the hypercube.*

Proof (Sketch). For any d degree polynomial simplify each term to multilinear term using $x^k = x$. As multilinear terms span the space and are independent they form a basis. □

B $\mathsf{SA}^2(\mathcal{P})$elaxation for independent set in cycle graphs

We will now compute a $\mathsf{SA}^2(\mathcal{P})$ relaxation of the LP given in LP 2.3. As we discussed in SA^t hierarchy we introduce non-negative junta variables and multiply by constraints to obtain new constraints which are polynomials of degree at most t. Here we introduce family of non negative junta polynomials of degree at most 2 as,

$$x_k \geqslant 0 \qquad 1 - x_k \geqslant 0 \qquad \forall k \in [n] \tag{13}$$

$$x_k x_l \geqslant 0 \qquad (1 - x_k)(1 - x_l) \geqslant 0 \qquad \forall k, l \in [n], k \neq l \tag{14}$$

Multiplying the original constraints in LP 2.3 and the constraints above we obtain constraints as

$$x_i x_k + x_j x_k \leqslant x_k \qquad x_i - x_i x_k + x_j - x_j x_k \leqslant 1 - x_k \qquad \forall k \in [n] \tag{15}$$

$$0 \leqslant x_i x_k \leqslant x_k \qquad 0 \leqslant x_i - x_i x_k \leqslant 1 - x_k \qquad \forall k \in [n] \tag{16}$$

We can linearize the above to obtain the $SA^2(\mathcal{P})$ LP relaxation equvalent to LP 2.2 as,

LP B.1.

$$\max \sum_{i \in V} y_i$$

subject to

$$y_{ik} + y_{jk} \leqslant y_k \qquad \forall (i,j) \in E, \forall k \in [n] \tag{17}$$

$$y_i - y_{ik} + y_j - y_{jk} \leqslant 1 - y_k \qquad \forall (i,j) \in E, \forall k \in [n] \tag{18}$$

$$0 \leqslant y_{ik} \leqslant y_k \qquad \forall i, k \in [n], i \neq k \tag{19}$$

$$0 \leqslant y_i - y_{ik} \leqslant 1 - y_k \qquad \forall i, k \in [n], i \neq k \tag{20}$$

$$y_i + y_j \leqslant 1 \qquad \forall (i,j) \in E \tag{21}$$

$$0 \leqslant y_i \leqslant 1 \qquad \forall i \in V \tag{22}$$

We can solve the LP explicitly to exhibit that the LP value is $n/2$. However using the Sherali-Adams hierarchy as a refutation system we can show that the objective value cannot exceed $n/2$. Here we show how to do this for a specific case of $n = 7$ (we can easily extend it to any C_n).

Derive	By
$y_{12} \leqslant 0$	$k = 1, (i, j) = (1, 2)$ in (17)
$y_2 - y_{12} + y_3 - y_{13} \leqslant 1 - y_1$	$k = 1, (i, j) = (2, 3)$ in (18)
$y_{13} + y_{14} \leqslant y_1$	$k = 1, (i, j) = (3, 4)$ in (17)
$y_4 - y_{14} + y_5 - y_{15} \leqslant 1 - y_1$	$k = 1, (i, j) = (4, 5)$ in (18)
$y_{15} + y_{16} \leqslant y_1$	$k = 1, (i, j) = (5, 6)$ in (17)
$y_6 - y_{16} + y_7 - y_{17} \leqslant 1 - y_1$	$k = 1, (i, j) = (6, 7)$ in (18)
$y_{17} \leqslant 0$	$k = 1, (i, j) = (1, 7)$ in (17)

Add all to obtain $y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 \leqslant 3$.