COMS E6998: Proof Complexity and Applications (Spring '25) March 27, 2025

Lecture 8: Semi-Algebraic Proof Systems

Instructor: *Toniann Pitassi* Scribes: *Kashvi Gupta and Yizhi Huang*

1 Introduction

Today we discuss semi-algebraic proof systems, including Sherali-Adams (SA) and Sum-of-squares (SOS). For a first impression on the relative strength of these two proof systems, it is known that SA poly-simulates Resolution, SOS poly-simulates SA, and Frege poly-simulates SOS. Besides, lower bounds are known for SA and SOS.

2 Sherali-Adams

Sherali-Adams is a sound and complete proof system for refuting a family of polynomial inequalities over the reals.

2.1 Definition

For an unsatisfiable CNF formula, we first transform it into a system of inequalities. For example, suppose $f = C_1 \wedge C_2 \wedge \cdots \wedge C_m$, and suppose $C_i = (\bigvee_{j \in S_i} x_j) \vee (\bigvee_{j \in T_i} \neg x_j)$. Let

$$\widetilde{C_i} = \sum_{j \in S_i} x_j + \sum_{j \in T_i} (1 - x_j).$$

For example, if $C_1 = x_1 \vee \neg x_2 \vee x_3$, then $\widetilde{C_i} = x_1 + (1 - x_2) + x_3$.

The system of inequalities corresponding to $f = C_1 \wedge C_2 \wedge \cdots \wedge C_m$ is then:

$$\widetilde{C_i} - 1 \geq 0, \forall i \in [m]$$
$$x_j^2 - x_j = 0, \forall j \in [n]$$

For a conjunction $D = (\bigwedge_{j \in S} x_j) \wedge (\bigwedge_{j \in T} \neg x_j)$, we can write D equivalently as

$$D = \prod_{j \in S} x_j \prod_{j \in T} (1 - x_j).$$

Definition 1 (Conical juntas)**.** A conical junta is a non-negative linear combination of juntas $J = \sum_i \lambda_i D_i$, where $\lambda_i \geq 0$ and D_i are conjunctions.

Definition 2 (Sherali-Adams refutation for CNF)**.** Let $f = C_1 \wedge C_2 \wedge \cdots \wedge C_m$ be an unsatisfiable CNF over $x_1, \ldots, x_n$. A Sherali-Adams refutation of f is given by conical juntas $J_0, J_1, \ldots, J_m$ and polynomials $q_1, q_2, \ldots, q_n$ such that

$$J_0 + \sum_{i=1}^{m} J_i(\widetilde{C_i} - 1) + \sum_{j=1}^{n} q_i(x_j^2 - x_j) = -1.$$

Remark 3 (Multilinearization)**.** Equivalently, we can drop the part $\sum_{j=1}^{n} q_i(x_j^2 - x_j)$, and a Sherali-Adams refutation is simply

$$J_0 + \sum_{i=1}^{m} J_i(\widetilde{C_i} - 1) = -1,$$

where arithmetic is done as multilinear polynomials, by replacing $x_j^d, d \geq 1$ by x_j.

2.2 Soundness and completeness

It is not hard to see that Sherali-Adams is sound.

Theorem 4 (Soundness of Sherali-Adams)**.** If a CNF f has a Sherali-Adams refutation, then f is unsatisfiable.

Proof. Assume the contrary that there is an assignment $\alpha \in \{0,1\}^n$ that satisfies f. Suppose

$$\Pi = J_0 + \sum_{i=1}^{m} J_i(\widetilde{C_i} - 1) + \sum_{j=1}^{n} q_j(x_j^2 - x_j) = -1$$

is a Sherali-Adams refutation for f. Then,

$$\begin{aligned} q_j(\alpha_j^2 - \alpha_j) = 0, \forall j \in [n] \\ \widetilde{C_i}(\alpha) - 1 \geq 0 \\ J_i(\alpha) \geq 0, \forall i \in [m]. \end{aligned}$$

These imply that $\Pi(\alpha) \geq 0$, a contradiction. □

For completeness of Sherali-Adams, we give a constructive proof.

Theorem 5 (Completeness of Sherali-Adams)**.** If a CNF f is unsatisfiable, then f has a Sherali-Adams refutation.

Proof. We partition the set of all assignments $\alpha \in \{0,1\}^n$ into m groups $\mathcal{G}_1, \ldots, \mathcal{G}_m$, such that $\alpha \in \mathcal{G}_i$ if and only if (1) $C_i(\alpha) = 0$, and (2) for all $i' < i$, $C_{i'}(\alpha) = 1$. It is easy to see that this is indeed a partition, and let i_α denote the unique index such that $\alpha \in \mathcal{G}_{i_\alpha}$.

Let P_i be the conical junta corresponding to $\mathcal{G}_i$, that is, a conical junta such that

$$P_i(\alpha) = \begin{cases} 0 & \alpha \notin \mathcal{G}_i \\ 1 & \alpha \in \mathcal{G}_i \end{cases}.$$

For example, if $\mathcal{G}_i = \{(0,0,0),(0,1,1),(1,0,0)\}$, then $P_i = (1-x_1)(1-x_2)(1-x_3) + (1-x_1)x_2x_3 + x_1(1-x_2)(1-x_3)$.

Then, for all $\alpha \in \{0,1\}^n$,

$$\sum_{i=1}^{m} P_i(\alpha)(\widetilde{C_i}(\alpha) - 1) = (\widetilde{C_{i_\alpha}}(\alpha) - 1) + \sum_{i \neq i_\alpha} 0 = -1.$$

We can find polynomials $q_1, q_2, \ldots, q_n$ such that

$$\Pi :- \sum_{i=1}^{m} P_i(\widetilde{C_i} - 1) + \sum_{j=1} q_j(x_j^2 - x_j)$$

is a multilinear polynomial. Since $\Pi(\alpha) = -1$ for all $\alpha \in \{0,1\}^n$, $\Pi = -1$. □

2.3 Lower bounds for Sherali-Adams via pseudodistirbutions

The complexity measure we use for a Sherali-Adams refutation is the largest degree among $J_0, J_1, \ldots, J_m$ and $q_1, q_2, \ldots, q_n$ in the refutation. We will show lower bounds for Sherali-Adams by a duality between degree-d Sherali-Adams refutations and degree-d pseudoexpectations, which we will soon define.

Definition 6 (Pseudodistributions)**.** A degree-d pseudodistribution on $\{0,1\}^n$ is a family of probability distributions $\mathcal{D} = \{\mathcal{D}_S \mid S \subset [n], |s| \leq d\}$ satisfying:

1. For all $S \subset [n]$, $\mathcal{D}_S$ is supported on $\{0,1\}^S$. $\mathcal{D}_S$ is understood as a probability distribution of all variables x_j where $j \in S$.
2. Marginals property: For all $S, T \in [n]$ such that $|S|, |T| \leq d$, $\mathcal{D}_S|_{S\cap T} = \mathcal{D}_T|_{S\cap T} = \mathcal{D}_{S\cap T}$. Here $|_{S\cap T}$ means taking the marginal distribution of variables x_j where $j \in S \cap T$.

We sometimes abuse the notation and use $\mathcal{D}$ for $\mathcal{D}_{[n]}$.

Example 7. The following is a part of a degree-3 pseudodistribution on $\{0,1\}^3$. $S = \{1,2\}$ and $T = \{2,3\}$.

$\mathcal{D}_S$:

x_1	x_2	
0	0	0.2
0	1	0.1
1	0	0.1
1	1	0.6

$\mathcal{D}_T$:

x_2	x_3	
0	0	0.3
1	0	0
1	0	0.5
1	1	0.2

$\mathcal{D}_{S\cap T}$:

x_2	
0	0.3
1	0.7

$\mathcal{D}_S|_{S\cap T}$:

x_2	
0	0.3
1	0.7

$\mathcal{D}_T|_{S\cap T}$:

x_2	
0	0.3
1	0.7

Definition 8 (Pseudoexpectation operators)**.** A degree-d pseudoexpectation operator $\widetilde{\mathbb{E}}$ is a functional that maps all multilinear polynomials over $x_1, \ldots, x_n$ of degree at most d to R and satisfies:

1. $\widetilde{\mathbb{E}}[1] = 1$.
2. $\widetilde{\mathbb{E}}$ is linear. That is, if p, q are polynomials of degree at most d and $\alpha, \beta \in \mathbb{R}$, then

$$\widetilde{\mathbb{E}}[\alpha p + \beta q] = \alpha\widetilde{\mathbb{E}}[p] + \beta\widetilde{\mathbb{E}}[q].$$

We have the following two lemmata, which roughly state that degree-d pseudoexpectations take expectations over degree-d pseudodistributions. They follow directly from the definitions.

Lemma 9. Let $\mathcal{D}$ be a degree-d pseudodistribution. Then, the functional $\widetilde{\mathbb{E}}$ defined by

$$\widetilde{\mathbb{E}}\left[\sum_S \alpha_S \prod_{j\in S} x_j\right] :- \sum_S \alpha_S \mathbb{E}_{x\sim\mathcal{D}}\left[\prod_{j\in S} x_j\right]$$

is a degree-d pseudoexpectation.

Lemma 10. Let $\widetilde{\mathbb{E}}$ be a degree-d pseudoexpectation. Then, the distribution $\{\mathcal{D}_S\}$ defined by

$$\mathcal{D}_S(y) = \Pr_{x\sim\mathcal{D}_S}[x = y] :- \widetilde{\mathbb{E}}\left[\prod_{j \text{ s.t. } y_j=1} x_j \prod_{j \text{ s.t. } y_j=0} (1 - x_j)\right] \text{ for } y \in \{0,1\}^S$$

is a degree-d pseudodistribution.

We define pseudoexpectation for a CNF.

Definition 11 (Pseudoexpectations for CNF)**.** Let $f = C_1 \wedge C_2 \wedge \cdots \wedge C_m$ be an unsatisfiable CNF. A degree-d pseudoexpectation is a pseudoexpectation for f if for every conjunction D of degree at most $d-1$, and for every $i \in [m]$, $\widetilde{\mathbb{E}}[D(\widetilde{C_i} - 1)] \geq 0$.

Now we state the duality between degree-d Sherali-Adams refutation and degree-d pseudoexpectations.

Theorem 12 (Sherali-Adams duality)**.** Let $f = C_1 \wedge \cdots \wedge C_m$ be an unsatisfiable CNF. Then, f has no degree-d Sherali-Adams refutation if and only if there exists a degree-d pseudoexpectation for f.

Proof for sufficiency. Assume f has a degree-d Sherali-Adams refutation

$$J_0 + \sum_{i=1}^{m} J_i(\widetilde{C_i} - 1) = -1$$

and assume $\widetilde{\mathbb{E}}$ is a degree-d pseudoexpectation for f. Applying $\widetilde{\mathbb{E}}$ to both sides of the above equation, we get

$$\widetilde{\mathbb{E}}[J_0] + \sum_{i=1}^{m} \widetilde{\mathbb{E}}[J_i(\widetilde{C_i} - 1)] = \widetilde{\mathbb{E}}[-1].$$

The left side is at least 0 since J_0 is a conical junta and $\widetilde{\mathbb{E}}[J_i(\widetilde{C_i} - 1)] \geq 0$ for every $i \in [m]$ by definition of pseudoexpectations for f. However, the right side equals -1, a contradiction. □

For the proof of necessity, please refer to [FKP19].

2.4 Linear programming tightening

We have seen that Sherali-Adams can be viewed as a proof system and as pseudodistributions. It can also be viewed as linear programming tightening.

Linear programming as a proof system. We first demonstrate that linear programming can be viewed as a proof system for linear inequalities. Consider a linear programming:

$$\begin{aligned} &\text{maximize } c^T x \\ &\text{s.t. } Ax \leq b \\ &\text{and } x \geq 0 \end{aligned}$$

The decision version of the linear programming is, whether there is a value of x satisfying $Ax \leq b$.

We have the following lemma, which essentially states that linear programming is a sound and complete proof system for satisfiability of linear inequalities over $\mathbb{R}$.

Lemma 13 (Farka's lemma)**.** A set $\{Ax - b \geq 0\}$ of linear inequalities is unsatisfiable over $\mathbb{R}$ if and only if there exists a vector y whose entries are all non-negative such that $y^T A = 0$ and $y^T b = -1$.

Recall the duality of linear programming:

(P) Primal:	(D) Dual:
maximize $c^T x$	minimize $b^T y$
s.t. $Ax \leq b$	s.t. $A^T y \leq c$
and $x \geq 0$	and $y \geq 0$

The following duality theorem is a consequence of Farka's lemma.

Theorem 14. Exactly one of the following holds:

1. Neither primal nor dual has a feasible solution.
2. Primal has solutions with arbitrarily large values while dual is unsatisfiable.
3. Primal is unsatisfiable while dual has solutions with arbitrarily large values.
4. Both primal and dual have optimal solutions, respectively x^* and y^*. Then, $c^T x^* = b^T y^*$.

A linear programming refutation of $\{Ax \leq b, x \geq 0\}$ is a nonnegative linear combination of these inequalities that equals -1. It is complete and sound by Farka's lemma.

A linear programming derivation of $c^T x \leq c_0$ from $\{Ax \leq b, x \geq 0\}$ is a nonnegative y^* such that $(y^*)^T b = c_0$. The soundness is because $c^T x \leq (y^*)^T Ax \leq (y^*)^T b = c_0$ and the completeness follows from Theorem 14.

Sherali-Adams as linear programming tightening. For an unsatisfiable $f = C_1 \wedge C_2 \wedge \cdots \wedge C_m$, there is a corresponding linear programming (or more precisely, linear inequalities, as there's no optimization objective) $\{\widetilde{C_i} - 1 \geq 0, 0 \leq x_j \leq 1\}$. The property that is not preserved in the linear programming is that x_j is an integer.

We call a tightening of the linear programming in the following way *Sherali-Adams degree-d tightening*:

- Suppose the original linear programming is

$$\begin{aligned} &\text{(ignore maximize } c^T x\text{)} \\ &Ax \leq b \\ &0 \leq x \leq 1 \end{aligned}$$

- We add new variables y_S for all $S \subset [n], |S| \leq d$. The meaning of y_S is $\prod_{j \in S} x_j$.
- We impose new constraints

$$\prod_{j \in S} x_j \prod_{j \in T} (1 - x_j) \cdot (a^T x - b) \geq 0$$

  for all rows a in A and all $S, T \subset [n]$ such that $S \cap T = \emptyset$ and $|S \cup T| \leq d$. The part $\prod_{j \in S} x_j \prod_{j \in T} (1 - x_j)$ corresponds to a conical junta in Sherali-Adams.

- We rewrite and add new constraints with the new variables y_S with multilinearization ($x_j^2 - x_j = 0$):

$$\begin{aligned} &y_\emptyset = 1 \\ &y_{\{j\}} = x_j \\ &0 \leq y_S \leq 1 \\ &\sum_{T' \subset T} (-1)^{|T'|} \left(\sum_{j=1}^{n} a_j y_{S \cup T' \cup \{j\}} - b y_{S \cup T'} \right) \geq 0 \end{aligned}$$

  where a is any row of A and a_j is the j-th entry of a.

The last constraint is because

$$\prod_{j\in T}(1-x_j)=\sum_{T'\subset T}(-1)^{|T'|}\prod_{j\in T'}x_j$$

and multilinearization. Because of multilinearization, the new constrains are a tightening of the original linear programming, while preserving all Boolean solutions. The following is an example.

Example 15. Consider the clause $x_1\vee x_2\vee\neg x_3$. The corresponding linear inequality is $x_1+x_2+1-x_3-1\geq 0$, i.e.

$$x_1+x_2-x_3\geq 0.$$

Note that $(x_1,x_2,x_3)=(1/2,1/2,1/2)$ is a feasible solution (but not a Boolean solution). However, in degree-2 Sherali-Adams (viewed as linear programming tightening), we have the following constraint:

$$(1-x_1)x_3(x_1+x_2-x_3)\geq 0.$$

After multilinearization, we get

$$x_1x_3+x_2x_3-x_3-x_1x_2x_3\geq 0,$$

or with y_S,

$$y_{\{1,3\}}+y_{\{2,3\}}-y_{\{3\}}-y_{\{1,2,3\}}\geq 0.$$

We can see that $(x_1,x_2,x_3)=(1/2,1/2,1/2)$ is no longer a feasible solution after tightening.

In the way described above, Sherali-Adams can be viewed as a successive tightening of linear programming. Sherali-Adams degree-0 tightening just corresponds to the original linear programming. A degree-$(d+1)$ tightening tightens a degree-d tightening, and a degree-n tightening is a "programming" where all feasible solutions are Boolean. Moreover, all Boolean solutions for the original linear programming is still a solution in the degree-n tightening. This intuition is formalized by the following lemma:

Lemma 16. Let $\mathcal{H}:-\{Ax-b\geq 0, 0\leq x\leq 1\}$ be a linear programming. Then, the Sherali-Adams degree-d tightening of $\mathcal{H}$ has no feasible solution if and only if there is a degree-d Sherali-Adams refutation of $\mathcal{H}$.

Proof sketch. For sufficiency, assume the contrary, suppose there is a feasible solution. We plug in the feasible solution into the degree-d Sherali-Adams refutation. Since the solution is feasible, each part in the refutation in the form of a conical junta multiplying the left-hand-side of a linear inequality, plus necessary multilinearization, should be nonnegative when plugging in the solution. Thus, the left-hand-side is nonnegative when plugging in the solution, a contradiction.

For necessity, since there is no feasible solution for the degree-d tightening, by Farka's lemma, there is a nonnegative linear combination of the linear inequalities in the degree-d tightening that equals -1. This linear combination is itself a degree-d Sherali-Adams refutation after necessary multilinearization. □

Automatizability of linear programming and Sherali-Adams. It is easy to see that satisfiability of linear inequalities over $\mathbb{R}$ is in NP. By Farka's Lemma, it is also in coNP. Thus, it is in NP $\cap$ coNP. However, actually it is known that linear programming is in P (e.g. using ellipsoid algorithm [Kha79]). We can also find a linear programming refutation in P.

By Lemma 16, deciding whether there is a degree-d Sherali-Adams refutation is equivalent to deciding whether the corresponding Sherali-Adams degree-d tightening has a feasible solution. Since the degree-d tightening has $n^{O(d)}$ linear constraints, deciding whether there is a degree-d Sherali-Adams refutation takes $n^{O(d)}$ time.

Theorem 17 (Degree-automatizability of Sherali-Adams)**.** A degree-d Sherali-Adams refutation can be found in time $n^{O(d)}$ when there is one.

Remark 18. For size-automatizability, for Sherali-Adams refutation of size s, the best-known upper bound for the time to find one is $2^{\sqrt{n \log s}}$. For s that is polynomial in n, this is exponential time.

3 Sum-of-Squares (SOS)

Sum-of-Squares is a refutation system (like Sherali-Adams) for certifying that a system of polynomial inequalities is unsolvable.

Definition 19. A polynomial q over $x_1, \ldots, x_n$ is a sum-of-squares polynomial if $q = \sum_i p_i^2$ for some polynomials p_i

Definition 20. Let $\phi = C_1 \wedge \cdots \wedge C_n$ be some unsatisfiable CNF formula. A sum-of-squares (SOS) refutation of ϕ is a set of sum-of-squares polynomials $\{q_0, q_1, \ldots, q_m\}$ such that $\sum_{i=1}^{m} q_i * (\widetilde{C}_i - 1) + q_o = -1$
**where we assume multi linear arithmetic

3.1 Soundness and Completeness

Soundness: See proof of soundness for Sherali Adams
Completeness: Follows from completeness of SA as any non-negative junta can be written as a sum-of-squares:

$$\text{Let } D = \Pi_{i \in S} x_i \ \Pi_{j \in T}(1 - x_j)$$

$$\text{Then } D^2 = (\Pi_{i \in S} x_i \ \Pi_{j \in T}(1 - x_j))^2 = \Pi_{i \in S} x_i^2 \ \Pi_{j \in T}(1 - x_j)^2$$

3.2 SOS Equivalent Views

Similar to SA, we have the following properties of SOS refutations. See [FKP19] for details.

- Like SA, we can define a suitable notion of pseudo-distribution and pseudo-expectation so that $\not\exists$ degree-d SOS refutation of ϕ iff $\exists$ a degree-d SOS pseudo-expectation for ϕ. This gives us a complete method for proving SOS degree lower bounds.
- SOS can be viewed as tightening of SDP (semi-definite program)
- Efficient algorithms for SDP imply degree-d SOS refutations are automatizable (ignoring coefficient size) in time $n^{O(d)}$

The following theorem shows that SOS proofs can have significantly smaller degree than SA proofs.

Definition 21. Let $n = 2^d$, with variables x_{ij} such that $i \in [0, \ldots, n-1]$ and $j \in [0, \ldots, d-1]$. x_{ij} is the jth bit of the binary representation showing where pigeon i is mapped. We write $\langle x_i \rangle = j$ to denote the width-d conjunction, which is true if and only if pigeon i maps to hole j. A BPHP$_n$ (binary pigeon hole principle) instance, therefore, is a set of clauses such that

1. No pigeon maps to hole 0: $\bigvee_{j=1}^{d} x_{ij}$
2. No two pigeons map to the same hole: $\neg(\langle x_i \rangle = j) \vee \neg(\langle x_{i^1} \rangle = j) \quad \forall i \neq i^1 \in [0, \ldots, n-1]$

Theorem 22. There are polysize, degree-3 SOS refutations of BPHP_n, but SA refutations require $\Omega(n)$ degree

Upper bounds can automatically generate efficient algorithms! Since SOS is automatizable with respect to degree, constant-degree SOS proofs can be found in polynomial time. This has been used in order to obtain some state-of-the-art algorithms for solving some learning and distributional tasks. Some examples are:

1. Dictionary Learning [BKS' 15]
2. Tensor Completion [BM16, PS17]
3. Tensor Decomposition [MSS16]
4. Robust moment estimation [KS17]
5. Clustering [HL18][KS17]
6. Robust linear regression [KKM18]

Lower bounds imply lower bounds for a broad class of algorithms. On the other hand, SOS lower bounds have also had a variety of applications. Using the lifting machinery, SOS lower bounds have been used to prove obtain superpolynomial lower bounds on the size of SDP Extended Formulations required in order to solve (exactly and even approximately) some NP-hard optimization problems [LRS15,CLRS16].

Most of these lower bounds (for SDP Extended Formulation size) reduce to the following SOS degree lower bound. (See [FKP19] for a proof.)

Theorem 23. There exist UNSAT kCNFs (Tseitin over constant-degree expander graphs) that require $\Omega(n)$ degree refutations in SOS.

References

[FKP19] Noah Fleming, Pravesh Kothari, and Toniann Pitassi. *Semialgebraic Proofs and Efficient Algorithm Design*. Foundations and trends in theoretical computer science. Now Publishers, 2019.

[Kha79] Leonid Genrikhovich Khachiyan. A polynomial algorithm in linear programming. In *Doklady Akademii Nauk*, volume 244, pages 1093–1096. Russian Academy of Sciences, 1979.