Information Theory & High-dimensional Statistics Lab **Prof. I-Hsiang Wang**

Sum-Of-Square (SOS) Method

November 4, 2015 **Chi-Ning Chou**

Contents

1 SOS

SOS method aims to answer the question:

When do equations and inequalities have real solutions?

SOS method answers the above question with two different directions: *pseudo distribution* and *Positivstellensatz*. Generally speaking, pseudo distribution gives a **close answer** that satisfies certain convex constraint and Positivstellensatz provides a **infeasible certificates** for real solutions to system of polynomial equations. The infeasible certificate is always correct and the pseudo distribution will have higher probability to be correct as long as the degree of SOS algorithm becomes higher.

In this section, we start with the motivation of SOS method, then provide a high-level overview. Next, we will focus on how to use SOS to conduct a lower bound proof. And in the next section, we will take planted clique problem as an example.

1.1 Motivation: Semidefinite Programming (LMIs)

Semidefinite programming (SDP), a.k.a. linear matrix inequality (LMI), is a general convex optimization problem:

$$\begin{array}{ll} \text{min} & \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(A_i X) = b_i,\ i = 1, ..., p \\ & X \geq 0 \end{array}$$

The importance of SDP is that it can be solved in **polynomial time!** (via interior point method) Moreover, with some modification, SDP can be written in an equivalent form as follow:

$$\begin{array}{ll} \text{min} & c^T x \\ \text{subject to} & A_0 + x_1 A_1 + ... + x_n A_n \geq 0 \end{array}$$

And the constraint is equivalent to

$$\exists x \forall y P(x, y) \geq 0$$

, where $P(x, y) = y^T(A_0 + x_1 A_1 + ... + x_n A_n)y$. We can see that this is actually a Σ_2 problem! Intuitively, we can somehow solve a class of Σ_2 problem in polynomial time. As a result, we want to characterize a general purpose method to keep the tractability to these problem.

Intuition (LMI)

Linear matrix inequalities (LMIs) are

quadratic forms that are nonnegative

Note that, although the problem is in Σ_2, they still can be solved in polynomial time.

From LMIs, we can generalize the idea to SOS. Formally, we define a multivariate **polynomial** $p(x)$ to be SOS if

$$p(x) = \sum_i q_i^2(x)$$

, and clearly that if a polynomial $p(x)$ is SOS, then $p(x) \geq 0,\ \forall x$. Thus, we define a SOS program to be:

$$\begin{array}{ll} \text{min} & c_1 u_1 + ... + c_n u_n \\ \text{subject to} & P_i(x, u) := A_{i0}(x) + A_{i_1}(x)u_1 + ... + A_{in}(x)u_n \text{ is SOS} \end{array}$$

Intuition (SOS)

SOS is a

affine families of polynomials that are sum of squares

Moreover, we can solve SOS program in polynomial time.

Note that just using **nonnegative polynomials** as constraints is a NP-hard problem and here as we turn to SOS program, the complexity is in P! Moreover, in several important cases (quadratic, univariate,...), nonnegativity and SOS are the same.

1.2 Introduction

To illustrate the main idea of SOS method, let's start with a imaginary game. Suppose there is a difficult problem $\mathcal{P}$ and there are two people want to find out whether $\mathcal{P}$ is correct or not. The optimist want to show the correctness of $\mathcal{P}$. So what he can do is to find an evidence to support his belief. On the other hands, the pessimist wants to refute the problem. His goal is to find a infeasible certificate to show that $\mathcal{P}$ is wrong.

In this traditional setting $\mathcal{P}$ is like all other decision problem. However, the main idea of SOS method here is to relax the requirement for the optimist. Namely, it's acceptable if the optimist can only provide a **close evidence to certain degree** r: *pseudo distribution*. Intuitively, this relaxation trade-off the soundness of the algorithm with time, since the algorithm will run in $O(n^r)$.

On the other hands, the pessimist has a general strategy too: *Positivstellensatz*. It is a very strong semidefinite program that generates a degree-r SOS infeasible proof as long as long as there's such proof! Note that both pseudo distribution and Positivstellensatz are convex optimization program (semidefinite programming), which is solvable in polynomial time. And the key idea here is that they are actually **dual** to each other![Las01],[Par00] The following table summarize this imaginary game:

Optimist	Close evidence	*Pseudo distribution*
Pessimist	Valid refutation	*Positivstellensatz*

Table 1: Imaginary game of SOS method.

There are two usage of SOS method: SOS algorithm and SOS lower bound proof. The former care about how to use a pseudo distribution to approximate a real solution. The latter focus on showing that there is no Positivstellensatz proof on certain degree, or equivalently showing that there's a pseudo distribution. Basically, the SOS algorithm and SOS lower bound proof have the following concepts:

1. For a problem, we can form a set of polynomial **axioms**.

   $$\mathcal{E} = f_1(x) = 0,\ f_2(x) = 0,\ ...,\ f_m(x) = 0$$

   and then we can use the following *rules of inference* to derive some inequalities

   - $p \geq 0, q \geq 0 \vdash p + q \geq 0$
   - $p \geq 0, q \geq 0 \vdash pq \geq 0$
   - $p \geq 0 \vdash p^2 \geq 0$

   Remark: We can always replace inequality $P(x) \geq 0$ by equation $P(x) - y^2 = 0$ for some slack variable y.

2. If we can use these axioms to derive some contradiction through the above rules of inference, then we can certify infeasibility of the problem. Most of the time, we derive the contradiction in the following form:

   $$\text{positivstellensatz refutation } (\mathbf{PS}(\text{r})):\ \sum_{i=1}^{m} f_i g_i = 1 + \sum_{i=1}^{N} h_i^2$$

, where $g_1, ..., g_m$ and $h_1, ..., h_N$ are some arbitrary polynomials and $deg(f_i g_i) \leq 2r$.

In other words, the above process shows the **infeasibility** of a system or the failure of a problem. As a result, it turns out to be a general algorithm to decline a problem. We call this a degree$-r$ SOS proof.

However, what's the importance of such idea?

Actually, it is because that SOS hierarchy has some connection with the above proof concept and can be stated as the following theorem:

Theorem 1 (SOS) *Under some mild conditions, there is a $n^{O(r)}$ time algorithm that given a set of polynomial axioms $\mathcal{E}$ and either output*

- *A degree$-r$ pseudo-distribution μ consistent with $\mathcal{E}$*

or

- *A degree$-r$ SOS proof that $\mathcal{E}$ is unsatisfiable.*

Intuition (degree)

When $r = 1$, this is actually a *linear programming*. When $r = 2$, this is a *semi-definite programming*. And when $r = n$, this is a *brute force/exhaustive search* algorithm!
As a result, what we are interested in is r in the range $2 < r \ll n$.

Here, we don't explain the concept of **pseudo-distribution** too deep. Intuitively, it is an object that closely related to the problem instance and obey the axioms $\mathcal{E}$. However, we cannot use a pseudo-distribution to certify the correctness of the problem.

Example: Now, let's take Max-Clique as an example:
Let $G = (V, E)$ be a graph and $V = [n]$. Let $x_i = \mathbf{1}_{\{i \in k-clique\}}$, we can construct the following axioms:

(Max-Clique) :	$x_i^2 - x_i,\ \forall i \in V$	$x_i = 0, 1$
	$x_i x_j,\ \forall (i,j) \notin E$	if $(i,j) \notin E$, either i in k-clique or j
	$k - \sum_{i \in V} x_i$	there are k vertices in the k-clique

[MW13] showed that with high probability there's no such infeasible proof for $k \leq \frac{\sqrt{n}}{(C \log n)^{r^2}}$. And [DM15] sharpened the result to $k \leq \frac{Cn^{1/3}}{\log n}$ for degree-4 SOS relaxation. In other words, Deshpande and Montanari showed that there's no $O(n^4)$ SOS algorithm to solve Max-Clique as $k \leq \frac{n^{1/3}}{\log n}$.

Intuition (Positivstellensatz refutation and SOS lower bound)

- SOS algorithm:
  - Positivstellensatz refutation provides a general way to decline a problem.
  - SOS can certify the infeasibility in $O(n^{\Theta(r)})$ time.
- SOS lower bound:
  - We can show that there's **no PS**(r) refutation and thus there's no r round SOS algorithm to show the indistinguishability and thus provide a lower bound.
  - The lower bound is in the sense that SOS algorithm fails to solve the problem in some small degree.

Remark: SOS algorithm and SOS lower bound are working in the different directions but share the same core idea: Positivstellensatz refutation.

1.3 Proof structure

Simply speaking, here we want to present a lower bound for **PS**(r). And first we need to define the polynomials set that we work on. Let $\mathcal{P}(n, 2r) : \{f : \mathbb{R}^n \to \mathbb{R}\}$ being a set of n-variate polynomials with degree at most $2r$. Next, we would like to have a definition for mapping that captures the non-negativity of square polynomial. Thus, we define PSD mapping as

Definition 2 (PSD mapping) *A linear mapping $\mathcal{M} : \mathcal{P}(n, 2r) \to \mathbb{R}$ is a PSD mapping if $\mathcal{M}(P^2) \geq 0$ for all n-variate polynomial with degree at most r.*

Moreover, we need a mapping to quickly check whether a polynomial is nonnegative. Thus, we call a PSD mapping that preserves the zeroness as dual certificate

Definition 3 (dual certificate) *Given a set of axioms $\{f_1, f_2, ..., f_m\}$, a dual certificate for these axioms is a PSD mapping $\mathcal{M}$ such that $\mathcal{M}(f_i g) = 0$ for all f_i in the axiom and polynomial g where $deg(f_i g) \leq 2r$.*

Remark: Actually, the pseudo-distribution mentioned in Theorem 1 is just a general name for dual certificate!

And there's a lemma that connects the existence of dual certificate and **PS**(r) refutation:

Lemma 4 *Given a set of axioms $\{f_1, f_2, ..., f_m\}$, there does not exist a **PS**(r) refutation if there exists a $(n, 2r)$-dual certificate.*

The reason why we can find a dual-certificate as long as the **PS**(r) refutation exists is because of the **dual property** semidefinite programming. Intuitively, finding a dual certificate is equivalent to show a lower bound for **PS**(r) refutation. That is, as long as we can find a dual certificate, we can never turn down a infeasible problem!

Conclusion

To sum up, the proof flow can be simplified as follow:

1. Write down axioms of the given problem $\rightarrow$ Polynomial constraints $\{f_i\}$
2. Turn into constraints for dual certificate $\rightarrow$ Guess a natural solution to the dual certificate $\mathcal{M}$.
3. Show that $\mathcal{M}$ is PSD w.h.p $\rightarrow$ Show the PSDness of M is sufficient.
   (a) $E[M]$ is PSD
   (b) Reduction to PSDness of M'_r, the principle minor of M and bound the mean and variance of M'_r.

1.4 Pseudo-distribution

Pseudo-distribution is a crucial concept in SOS proof. It can be regarded as a **computationally -bounded** observer. Concretely, the generation of a pseudo-distribution does not consume every possible computational resource. Moreover, the the results must satisfy problem constraints and SOS nonnegativity. Formally speaking, the expectation of pseudo distribution on the axiom polynomials must be 0 and the expectation of all square polynomial must be nonnegative.

The algorithm only search on a low degree of possibility. Somehow, this captures our limited computational abilities. For convenience, we can simply think of a pseudo-distribution as our "computational knowledge".

1.5 Resources

1. **Lectures:**
   - Barak SOS lecture: link
   - Laurent Semidefinite optimization lecture: link
   - Parrilo Algebraic Techniques and Semidefinite Optimization lecture: link
2. **Videos**:
   - TCS+ talk - Boaz Barak: link
   - TCS+ talk - Aaron Potechin: link
   - ICM2014 - Boaz Barak: link
3. **Papers**:
   - Complexity of Null-and Positivstellensatz proofs: [GV01]
   - Global optimization with polynomials and the problem of moments: [Las01]
   - Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization: [Par00]

- Association schemes, non-commutative polynomial concentration, and sum-of-squares lower bounds for planted clique: [MW13]
- Approximability and proof complexity: [OZ13]

2 Example: SOS Lower Bound For Planted Clique Problem

In this section, we see a SOS lower bound example in planted clique problem. The proof is provided by [MW13]. Meka and Wigderson constructed a special moment matrix as a dual certificate to show the impossibility of refutation in some low degree.

2.1 Problem settings & Goal

First, let's recall the problem setting:

$$\begin{array}{lll} (\textsc{Max-Clique}): & x_i^2 - x_i,\ \forall i \in V & x_i = 0,1 \\ & x_i x_j,\ \forall (i,j) \notin E & \text{if } (i,j) \notin E, \text{ either } i \text{ in } k\text{-clique or } j \\ & k - \sum_{i \in V} x_i & \text{there are } k \text{ vertices in the } k\text{-clique} \end{array}$$

1. $M_{I,J} = \mathcal{M}(\prod_{s \in I \cap J} x_s)$, where I, J are subsets with size no larger than r.

   $$M \text{ is PSD} \Leftrightarrow \mathcal{M} \text{ is PSD}$$

   Here, M is the association scheme for $\mathcal{M}$. We can utilize some good properties and results of it from combinatorial theory, which will be introduced in latter section.

2. $X_I = \prod_{s \in I} x_s$, where $I \subseteq [n]$.

We want to show the following two results: ([MW13])

Theorem 5 (planted clique lower bound)

1. *(Maximum clique) W.h.p., for $G \leftarrow G(n, 1/2)$ the natural r-round SOS relaxation of the maximum clique problem has an integrallity gap of at least $\sqrt{n}/(C \log n)^{r^2}$.*
2. *(Planted clique) W.h.p., for $G \leftarrow G(n, 1/2, t)$ the natural r-round SOS relaxation of the maximum clique problem has an integrallity gap of at least $\sqrt{n}/t(C \log n)^{r^2}$.*

2.2 Meka-Wigderson's candidate for dual certificate

With the above axioms, for any suitable dual certificate $\mathcal{M}$ for graph G should satisfies

1. $\mathcal{M}(X_I) = 0,\ \forall I,\ |I| \leq 2r$, I is not a clique in G.
2. $\mathcal{M}((\sum_i x_i - k)\, X_I) = 0,\ \forall I,\ |I| \leq 2r$.

With these two constraints, Meka and Wigderson then guess the following candidate for dual certificate:

Proposition 6 *Define a candidate of dual certificate for graph G as $\mathcal{M}$. For all $I \subseteq [n]$ and $|I| \leq 2r$, let*

$$\mathcal{M}(\prod_{i \in I} x_i) = deg_G(I) \cdot \frac{\begin{pmatrix} k \\ |I| \end{pmatrix}}{\begin{pmatrix} 2r \\ |I| \end{pmatrix}}$$

Remark:

1. $deg_G(I)$ = the number of size $2r$ clique in G that contains I. Or equivalently, $deg_G(I) = |\{S \subseteq [n] : I \subseteq S, |S| = 2r, S \text{ is a cliquie in } G\}|$.
2. Since $\mathcal{M}$ is a mapping from polynomial of degree at most $2r$ to a real number, it's sufficient to specify all the possible monomial with degree no more than $2r$.
3. We can check that this $\mathcal{M}$ satisfies the above constraints. (page 16 in [MW13])

Now, to prove Theorem 5, it's sufficient to show the PSD'ness of $\mathcal{M}$. And from the previous section, we also know that it's equivalent to show the following matrix is PSD:

$$M(I, J) = deg_G(I \cup J) \cdot \frac{\begin{pmatrix} k \\ |I \cup J| \end{pmatrix}}{\begin{pmatrix} 2r \\ |I \cup J| \end{pmatrix}}$$

Intuition (Meka-Wigderson's moment matrix)

Consider the moment of pseudo distribution:

$$\tilde{\mathbb{E}}(X_I) = \mathcal{M}(I)/deg_G(\emptyset) = \frac{\begin{pmatrix} k \\ |I| \end{pmatrix}}{\begin{pmatrix} 2r \\ |I| \end{pmatrix}} \times \frac{deg_G(I)}{deg_G(\emptyset)}$$

Observe that if I is a **clique**

- $deg_G(I) \sim n^{d-|I|}$
- $deg_G(\emptyset) \sim n^d$
- $\begin{pmatrix} k \\ |I| \end{pmatrix} \sim k^{|I|}$
- $\begin{pmatrix} 2r \\ |I| \end{pmatrix} \sim constant$

Thus,

$$\tilde{\mathbb{E}}(X_I) \sim \frac{k^{|I|}}{n^{|I|}}$$

The Meka-Wigderson moment matrix will look like follow:

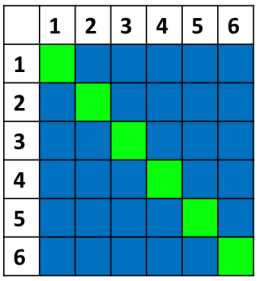

Figure 1: $r = 1$

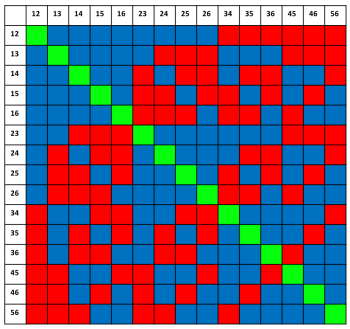

Figure 2: $r = 2$

Figure 1:

- ■: $M(i,i) \approx \frac{k}{n}$
- ■: If there's an edge between i and j, then $M(i,j) \approx \frac{k^2}{n^2}$.Else, $M(i,j) = 0$.

Figure 2:

- ■: If there's an edge between i and j, then $M(\{i,j\},\{i,j\}) \approx \frac{k^2}{n^2}$. Else, $M(\{i,j\},\{i,j\}) = 0$.
- ■: If there's a clique among i, j and k, then $M(\{i,j\},\{i,k\}) \approx \frac{k^3}{n^3}$. Else, $M(\{i,j\},\{i,k\}) = 0$.
- ■: If there's a clique among i, j, k and l, then $M(\{i,j\},\{k,l\}) \approx \frac{k^4}{n^4}$. Else, $M(\{i,j\},\{k,l\}) = 0$.

2.3 Proof flow

To prove the lower bound of maximum clique and planted clique problems through SOS lower bound techniques, we simply need to do two things:

1. Design a dual certificate $\mathcal{M}$ based on clique axioms.
2. Show that $\mathcal{M}$ is PSD.

With these two statements, we can thus claim that the problem does not have a degree $2r$ SOS algorithm.

However, the second step is not easy, most of the technical parts lie in there. In the following example of [MPW15], they prove the PSD'ness of $\mathcal{M}$ with the following sub-steps:

1. Relate $\mathcal{M}$ to a matrix representation M such that $\mathcal{M}$ is PSD $\Leftrightarrow$ M is PSD.
2. Add small value to the nonzero entries of M and get M'.
   - If M' is PSD, then M is PSD.

- Write $M' = E + L + \Delta$, where $E = \mathbb{E}[M']$, L is a special local random matrix and Δ is a small global random matrix.

3. Show M' is PSD:
   - $E \succ 0$ and $\lambda_{\min}(E) = \Omega(k^r n^r)$
   - $||L|| < Ck^{2r} n^{r-0.5} \log n$
   - $||\Delta|| < Ck^{2r} n^{r-0.5} \log n$

2.4 Appendix: Johnson scheme

Johnson scheme is a special matrix whose columns and rows refer to a subset of set S. More interestingly, the value of the entry say $M(I, J)$ only depends on the size of $|I \cup J|$. This beautiful structure immediately leads to some good properties especially in the upper bound of eigenvalue. As we want to show the PSDness of matrices, it's sufficient for us to bound their least eigenvalue. And we are going to utilize Johnson scheme to help us do so.

To define Johnson scheme, we first need to introduce one concept: *set symmetry.*

Definition 7 (set symmetry) *We say a matrix* $M \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ *is set symmetric if for every subset* $I, J \in \binom{[n]}{r}$, $M(I, J)$ *only depends on* $|I \cap J|$.

Intuitively, the matrix depends only on the size of intersection.

Definition 8 (Johnson scheme) *For* n *and* $r \leq n/2$, *let* $\mathcal{J} \equiv \mathcal{J}_{n,r} \subseteq \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ *be the subspace of all set symmetry matrix.* $\mathcal{J}$ *is called the Johnson scheme.*

Clearly, we can see that the rank of Johnson scheme is $r + 1$ because the size of intersection can only be ranged from 0 r. Now, we might wonder is there a good basis for Johnson scheme. The following shows two common bases:

- (**D-Basis**) For $0 \leq l \leq r$,
$$D_l(I, J) = \mathbf{1}_{|I \cap J| = l}$$
That is, the indicator of the size of intersection.

- (**P-Basis**) For $0 \leq t \leq r$.
$$P_t(I, J) = \binom{|I \cap J|}{t}$$
That is, the number of possible subset of the intersection of I and J.

And the transformation of D-Basis and P-Basis are:

- $P_t = \sum_l \binom{l}{t} D_l$

- $D_l = \sum_t (-1)^{t-l} \begin{pmatrix} t \\ l \end{pmatrix} P_t$

Now, with these two bases, we want to further characterize the eigenvalues of matrices in Johnson scheme. And here we first present some well-known facts:

Lemma 9 *The matrices in Johnson scheme commute to each other. In other words, they are* **simultaneously diagonalizable**.

Lemma 10 *There are subspaces* $V_0, ..., V_r \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ *such that*

1. *$V_0, ..., V_r$ are eigenspaces of $\mathcal{J}$ and orthogonal to each other.*
2. *$dim(V_j) = \begin{pmatrix} n \\ j \end{pmatrix} - \begin{pmatrix} n \\ j-1 \end{pmatrix}$. Thus, $\sum_j dim(V_j) = \begin{pmatrix} n \\ r \end{pmatrix}$.*
3. *The eigenvalues of P_t corresponds to eigenspace V_j is*

$$\lambda_j(P_t) = \mathbf{1}_{j \le t} \begin{pmatrix} n-t-j \\ r-t \end{pmatrix} \cdot \begin{pmatrix} r-j \\ t-j \end{pmatrix}$$

4. *For arbitrary matrix $Q \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$, $Q = \sum_l \alpha_l D_l$, and $\beta = \sum_{l \le t} \binom{l}{t} \alpha_l$, where $\alpha_l \ge 0$. Then,*

$$\lambda_j(Q) \le \sum_{t \ge j} \beta_t \cdot \begin{pmatrix} n-t-j \\ r-t \end{pmatrix} \cdot \begin{pmatrix} r-j \\ t-j \end{pmatrix}$$

References

[DM15] Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. *arXiv preprint arXiv:1502.06590*, 2015.

[GV01] Dima Grigoriev and Nicolai Vorobjov. Complexity of null-and positivstellensatz proofs. *Annals of Pure and Applied Logic*, 113(1):153–160, 2001.

[Las01] Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[MPW15] Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for planted clique. *arXiv preprint arXiv:1503.06447*, 2015.

[MW13] Raghu Meka and Avi Wigderson. Association schemes, non-commutative polynomial concentration, and sum-of-squares lower bounds for planted clique. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 20, page 105, 2013.

[OZ13] Ryan O'Donnell and Yuan Zhou. Approximability and proof complexity. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1537–1556. SIAM, 2013.

[Par00] Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, Citeseer, 2000.