**Recent Works on Time, and Error Trade-off**

**October 15, 2015** **Chi-Ning Chou**

# Contents



# 1 Introduction

This survey focus on the time, error trade-off problem of two problems: *hidden submatrix* and *hidden cluster*. Moreover, there also two kinds of different approaches to study these problems: *detection* and *localization*.

The goal here is to find the fundamental limit (w.r.t. size, sparsity) to achieve certain criteria. Traditional aspects consider the **feasible/infeasible** issues. That is, in what region can guarantee us to detect/localize the submatrix/cluster w.h.p. Furthermore, we want to find out in what region can we do so in a reasonable time or using a particle method to achieve the goals. Namely, where's the **tractable/intractable** region?

This week, I've surveyed the goals, motivations, and results of the following paper: [MW$^+$15], [CX14], [DM15], [CLR15], [HWX14], [BR13]. They focused on the time-error trade-off of hidden submatrix and cluster problem ([BR13] discussed the sparse PCA detection) with slightly different settings. I will show the difference and present their results.

### 1.1 Hidden submatrix problem (HSP)

There's a (symmetric) $n \times n$ matrix $A$ where each of the entry $A_{i,j}$ follows the distribution (Gaussian, sub-Gaussian, etc.) with mean $\theta_{i,j}$. We say that $A$ has a submatrix with size $k \times k$ (for simplicity, I consider the square submatrix case) if there is a index set $C, R \subset [n],\ |C| = |R| = k$ such that

- If $i \in R,\ j \in C$, then $\theta_{i,j} \geq \lambda$.
- Else, $\theta_{i,j} = 0$.

, where $\lambda$ can be thought of as the strength of signal.

Clearly that the detection problem is a hypothesis testing of whether there's a submatrix with size $k \times k$ and the localization version is to find the position of the submatrix.

Note that there are basically three parameters for us to tune with:

- Matrix size $n$
- Submatrix size (sparsity) $k$
- Signal strength $\lambda$

Intuitively, as $\frac{k}{n}$ becomes small, the problem becomes more difficult. And as $\lambda$ grows large, the problem becomes easier.

HSP is discussed in [DM15], [CLR15], [MW$^+$15], [CX14].

### 1.2 Hidden clique/clustering problem (HCP)

Most of the HCP settings are based on the Erdos-Renyi random graph. Here, I introduce a general (multi-cluster) version of HCP[CX14]. Suppose there are $n$ nodes and $r$ cluster with $k$ nodes in each cluster respectively. The connection rule is as follow:

- If $u, v$ are in the same cluster, then they connect to each other with probability $p$.
- If $u, v$ are **not** in the same cluster, then they connect to each other with probability $q$.

There are lots of variation by modifying the setting of $p$ and $q$. For example:

- Planted $r$-Disjoint Clique: $p = 1,\ 0 < q < 1$
- Planted Densest Subgraph: $0 < q < p < 1$
- Planted Partition (Stochastic block model): $n = rk,\ p, q \in (0, 1)$

- Planted $r$-Coloring: $n = rk,\ 0 = p < q < 1$

The detection problem is to find out whether there is such a clique/cluster and the location problem is to find where it is.

There are basically four parameters for us to tune with:

- Number of nodes $n$
- Size of the clique $k$
- Number of cliques $r$
- Connection probability $p, q$

Intuitively, $\frac{k}{n}$ makes the problem harder as it becomes smaller, and the difference of $p, q$ influences the difficulty of detection/localization.

HCP is an important model in communication network since it can describe the relationship among each node. It's discussed in [CX14], [DM15].

**Remark**: The best known polynomial time algorithm is $\Theta(\sqrt{n})$ and current state-of-the-art lower bound is $\Omega(\frac{n^{1/3}}{\log n})$ with SOS techniques in [DM15].

### 1.3 Statistical and computational concerns

**Statistical point of view** In the statistical context, we care whether the problem is feasible. That is, we want to know whether the problem can be solved w.h.p. under a given parameter setting. In other words, we want to find the region where there is a method for us to detect/localize the submatrix/clique with vanished error.

Note that here we not necessary consider the asymptotic scheme. Namely, some papers provide a condition for any finite parameter setting.

**Computational point of view** Here, we not only care about whether we can solve the problem or not, we care whether we can/cannot solve the problem **efficiently** (*i.e.,* polynomial time). Note that there are two slightly different philosophies:

- Achievable: provide a polynomial time algorithm to show the achievability.
- Non-achievable: use computational theoretic lower bound techniques to show the impossibility.

### 1.4 Common approaches

Here we don't focus on the statistical point of view since most of the results and techniques are well-studied. This section introduce the main methods used in proving the fundamental limit in a computational point of view.

**Achievable** [CX14] used convex relaxation MLE to achieve polynomial time localization for both HSP and HCP.

**Non-achievable** [BR13], [HWX14], [CLR15], and [MW+15] used *hidden clique hypothesis* to show the impossibility of detect/localize sparse PCA, HSP, and HCP respectively.

On the other hand, [DM15] used *sum of square* (SOS) semidefinite hierarchy to improve the lower bound of HSP and HCP.

# 2 Recent results

In this section, I listed the results of the related papers without detail informations. The focus will be there methodologies, assumptions, and implications.

## 2.1 Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems [DM15]

This paper use SOS to provide a better lower bound for the detection HSP and HCP. They showed that SOS fails to detect

- **HCP**: $k \lesssim \frac{n^{1/3}}{\log n}$
- **HSP**: $k \lesssim \frac{\lambda^{-1} n^{1/3}}{\log n}$

**Remark** There's no poly-time algorithm known to detect HCP and HSP when $k = o(n^{1/2})$ but no one yet to prove it. The above lower bound is the state of the art.

We will discuss more on SOS later.

## 2.2 Computational and statistical boundaries for submatrix localization in a large noisy matrix [CLR15]

This paper focus on the submatrix localization problem in a noisy setting. As a result, here we consider the signal to noise ratio SNR $:= \frac{\lambda}{\sigma}$. And they provided the following results in Figure 1 The two thresholds are:

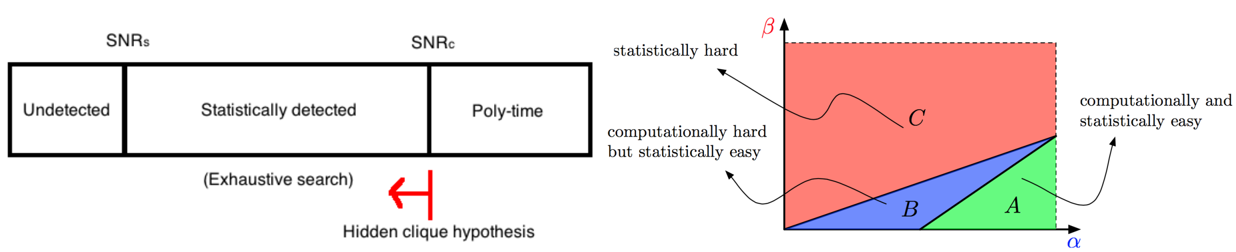


Figure 1: [CLR15]

- $\mathrm{SNR}_s \asymp \sqrt{\frac{\log n}{k}}$
- $\mathrm{SNR}_c \asymp \sqrt{\frac{n}{k^2}} + \sqrt{\frac{\log n}{k}}$

We can parametrize the model with $n$: $k = \Theta(n^\alpha)$, $\frac{\lambda}{\sigma} = \Theta(n^{-\beta})$

### 2.3 Computational barriers in minimax submatrix detection [MW+15]

This paper discusses the submatrix detection problem. They parametrized with $n$: $k = \Theta(n^{\alpha})$, $\lambda = \Theta(n^{-\beta})$ and got the following results:

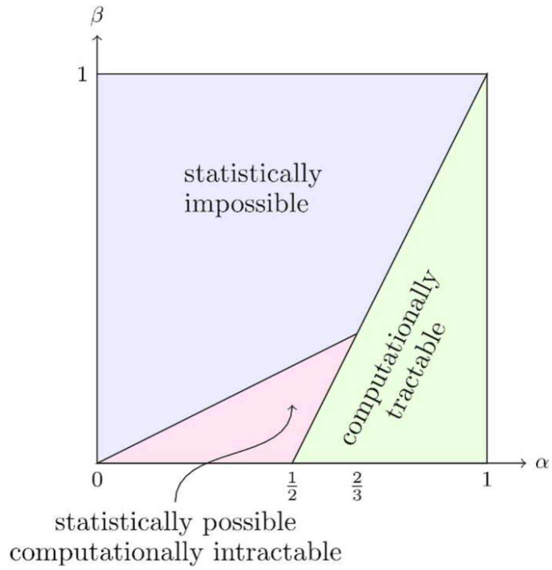


Figure 2: [MW+15]

They used discretized Gaussian model to approximate the continuous setting. Also, the computational lower bound is provided by hidden clique hypothesis. Also, the authors pointed out that the results will be different when using different loss function.

**Remark** There are known necessary and sufficient condition for the possibility of detection (feasibility):

- Error $\to 0$ if $\frac{\lambda}{n/k^2} \to \infty$
- Error $\to 1$ if $\frac{\lambda}{n/k^2} \to 0$

### 2.4 Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices [CX14]

This paper defined four regions:

- **Impossible region**: Infeasible any algorithm.
- **Hard region**: Feasible for statistical algorithm.
- **Easy region**: Feasible for polynomial time algorithm.
- **Simple region**: Feasible for counting/threshold algorithm.

The following consider HCP and HSP respectively.

**HCP** They parametrized with $p$ and $q$: $p = 2q = \Theta(n^{-\alpha})$, $k = \Theta(n^{\beta})$. And the quantity that captures the hardness is SNR $:= \frac{(p-q)^2}{q(1-q)}$.

**HSP** They parametrized with $n$: $\lambda^2 = \Theta(n^{-\alpha})$, $k = \Theta(n^{\beta})$. And the quantity that captures the hardness is SNR $:= \lambda^2$.

The four regions for both problems are characterized with SNR in Figure 3.

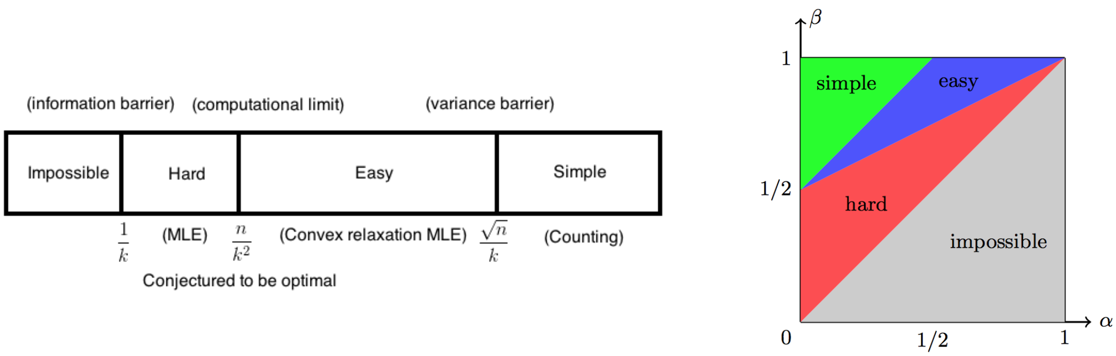


Figure 3: [CX14]

# 3 SOS

## 3.1 Resources

- Barak SOS lecture: link
- Laurent Semidefinite optimization lecture: link
- Parrilo Algebraic Techniques and Semidefinite Optimization lecture: link
- 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]

Highly recommend Barak's lecture notes and videos.

## 3.2 Introduction

Intuitively, 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

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

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 $\mu$ consistent with $\mathcal{E}$*

*or*

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

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.

Now, let's take Max-Clique as an example:

$$\begin{aligned}(\text{Max-Clique}) : & x_i^2 - x_i,\ \forall i \in V \\ & x_i x_j,\ \forall (i,j) \in E \\ & k - \sum_{i \in V} x_i\end{aligned}$$

[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** $\mathbf{PS}(r)$ 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.

### 3.3 Proof structure

Simply speaking, here we want to present a lower bound for $\mathbf{PS}(r)$. And first we need to 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 fancy name for dual certificate!

And there's a lemma that connects the existence of dual certificate and $\mathbf{PS}(r)$ refutation:

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

Intuitively, finding a dual certificate is equivalent to show a lower bound for $\mathbf{PS}(r)$ 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 $M'_r$.

### 3.4 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. 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".

## 4 Example: SOS Lower Bound For Planted Clique Problem

### 4.1 Problem settings & Goal

First, let's recall the problem setting:

$$\begin{aligned}(\textsc{Max-Clique}) : & x_i^2 - x_i,\ \forall i \in V \\ & x_i x_j,\ \forall (i,j) \in E \\ & k - \sum_{i \in V} x_i\end{aligned}$$

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** 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 $\sqrt{n}/t(C \log n)^{r^2}$.*

## 4.2 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*

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

**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:

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

## 4.3 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 the following two, 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 [MW13], 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. Consider $E = \mathbb{E}(M)$ and $\Sigma^2 = Var[M]$, prove $M \geq 0$ via proving $E - M \leq E$
   (a) $E - M \leq \Delta\Sigma$ for small $\Delta$ w.h.p via large deviation theory.
   (b) $\Delta\Sigma \leq E$ with some matrix analysis.

There are two main difficulties:

1. $E, M$ will be **singular**.
   - Sow the kernel of them is PSD and with *interlacing eigenvalues theorem*, we can show that they are PSD.
2. $M$ has too high variance.
   - Work on another closely related matrix $M'$.
   - Bound the eigenvalues of $Var[M']$ via considering a variant of the principal minor of $M'$: $M'_r$.

## 4.4 Johnson scheme

We use association scheme and Johnson scheme from combinatorial theory. There are some classical results for bounding the eigenvalues of Johnson scheme, which is desirable for us.

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:

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

- $D_l = \sum_t (-1)^{t-l} \binom{t}{l} P_t$

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. *$dim(V_j) = \binom{n}{j} - \binom{n}{j-1}$. Thus, $\sum_j dim(V_j) = \binom{n}{r}$.*
3. *The eigenvalues of $P_t$ corresponds to eigenspace $V_j$ is*

$$\lambda_j(P_t) = \mathbf{1}_{j \leq t} \binom{n-t-j}{r-t} \cdot \binom{r-j}{t-j}$$

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

$$\lambda_j(Q) \leq \sum_{t \geq j} \beta_t \cdot \binom{n-t-j}{r-t} \cdot \binom{r-j}{t-j}$$

# References


[BR13] Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory*, pages 1046–1066, 2013.

[CLR15] T Tony Cai, Tengyuan Liang, and Alexander Rakhlin. Computational and statistical boundaries for submatrix localization in a large noisy matrix. *arXiv preprint arXiv:1502.01988*, 2015.

[CX14] Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *arXiv preprint arXiv:1402.1267*, 2014.

[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.

[HWX14] Bruce Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs. *arXiv preprint arXiv:1406.6625*, 2014.

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

[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.

[MW+15] Zongming Ma, Yihong Wu, et al. Computational barriers in minimax submatrix detection. *The Annals of Statistics*, 43(3):1089–1116, 2015.

[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.