SOS Lecture 8: SOS Lower bounds for planted clique / Boaz Barak

Reading This lecture described work in an unpublished manuscript of Meka, Potechin and Wigderson, and also mentioned some discussions/work in progress with Chan, Kelner, Meka, Potechin, Steurer and Wigderson.

Despite containing a fatal error, the paper of Meka and Wigderson (see version 1 on the arxiv `http://arxiv.org/abs/1307.7615v1`) has some beautiful ideas and is still very worthwhile reading.

Their approach for the planted clique extends the knapsack lower bound of Grigoriev, for which a proof can be found in Section 8 of Grigoriev, Hirsch, Pasechnink (`http://eccc.hpi-web.de/report/2001/103/`). See also Section 6 of the Meka-Wigderson paper for a simplified proof.

Planted Clique The planted clique problem is one of the most natural average case problems, and has been studied for many years by many people. The setting is that we are given a random graph $G = (V, E)$ drawn from $G(n, 1/2)$ in which a clique of size ω was "planted" — i.e., we chose a random subset $S \subseteq V$ and added to E all edges in $\binom{S}{2}$ to obtain the planted graph G'. There are three natural goals, and the algorithmic status of all of them is the same:

- *Search:* Given the planted graph G', find S.
- *Refutation:* Given the random graph G, certify that it has no clique of size ω.
- *Decision:* Give an algorithm that can distinguish with probability, say, 0.9, between a random graph G drawn from $G(n, 1/2)$ and a graph G' obtained by this planting process.

As usual, the search and refutation variants are incomparable, while decision is easier than both. While the search problem is probably the most natural for applications, the refutation question is the natural one to consider when thinking of SOS lower bounds.

Algorithmic results It's a good exercise to show that with high probability, a random graph will have no clique of size larger than $c \log n$ for some c (in fact $c = 2$). Hence the refutation and decision problems can be obviously solved in $n^{O(\log n)}$ time as long as $\omega \gg \log n$. A moment's thought shows that this is true for the search problem as well. Given the graph G', we will find in time $n^{O(\log n)}$ a clique S' of size $10 \log n$, and then one can show that it will be a subset of S and in fact S will simply be the set of joint neighbors of all vertices in S'.

The decision version can be seen to be easy to solve if $\omega > c\sqrt{n}$ for some constant. The number of edges in a random graph is a random variable with expectation $m = \frac{1}{2}\binom{n}{2}$ and standard deviation $\sqrt{m}/4$. When we plant a clique we add about $\frac{1}{2}\binom{\omega}{2}$ edges and so if $\omega = c\sqrt{n}$ for a large c this will be many standard deviations and easily distinguishable from the random case.

The search and refutation versions are a bit more subtle, but can still be efficiently done in that case. There are algorithms for the search version that are based on the above observation, looking for the vertices with the highest degrees as potential members of the clique. These problems can also be solved using degree 2 SOS or eigenvalue methods, using the observation that in a random graph, the value of the second largest eigenvalue certifies the non-existence of such a clique.

Hardness results While the planted clique problem is not as hard as, say, 3SAT, given that it can be solved in quasipolynomial time, it has still been an object of much interest, and the basis for various reductions.

Some support for the potential difficulty of the problem stems from various worst-case hardness results suggesting that the time to find a k-clique scales more like $\binom{n}{k}$ than like $2^k poly(n)$. (The constant in the exponent can be improved a bit using fast matrix multiplication, but still behaves like $n^{\Omega(k)}$ for some $c > 0$.)

However, this does not suggest why $\sqrt{n}$ should be the right bound. The best known algorithms can detect a clique of size $\sqrt{n}/t$ in time roughly $n^{\log t}$ and there are some lower bounds for weaker convex programs than the SOS method.

However, there was no negative result known for the SOS method, which is a very natural approach to try on this problem.

Meka-Wigderson paper About a year ago, Meka and Wigderson posted a paper claiming that for every constant degree d the SOS method cannot refute the existence of an $\epsilon\sqrt{n}$-clique in a random graph for some $\epsilon > 0$. This paper turned out to have a fatal flaw (which we will discuss), but in a very recent (yet unpublished) manuscript with Potechin, they have managed to salvage part of their approach and show that the degree d SOS method cannot refute the existence of an $\Omega(n^{1/d})$ clique in a random graph. I am also involved in discussions with them, as well as Chan, Kelner and Steurer on how to potentially improve this further.

Main Theorem We will demonstrate this with the $d = 4$ case, and so prove the following result: let $G = (V, E)$ be a random graph drawn from $G(n, 1/2)$. Then there is $\epsilon > 0$ such that if $\omega < \epsilon n^{1/8}$ then with high probability there exists a degree 4 pseudo-distribution $\{x\}$ satisfying the constraints $\{x_i^2 = x_i\}$ for all i, $\{x_i x_j = 0\}$ for every $(i, j) \notin E$, and $\tilde{\mathbb{E}} \sum x_i \geq \omega$.

Note that an actual distribution satisfying these constraints would be distributed over characteristic vectors of cliques and has expectation ω, and hence in particular its existence would imply that the graph has an ω-sized clique. (In the Meka-Wigderson and Meka-Potechin-Wigderson papers they consider some slightly modified distributions that satisfy $\{\sum x_i = \omega\}$ as a constraint; see the former paper for details.) The bound can be easily improved to $n^{1/4}$ (exercise), and with more work probably can reach $n^{1/3}$ using similar ideas. Beyond that (as we will see) we would need different moments, though it may well still be possible.

The pseudo-distribution Per Einstein's maxim, Pseudo-distributions should be "as random as possible but not randomer". Nevertheless, the pseudo-distribution we will use will in some sense be "too random", which is the reason we cannot reach ω close to $\sqrt{n}$. But, it would still be good enough for $\omega \sim n^{1/8}$.

The most random distribution would emulate a random ω-sized set, and so we would set $\tilde{\mathbb{E}} x_S = \tilde{\mathbb{E}} \prod_{i \in S} x_i = c_{|S|} (\omega/n)^{|S|}$, for every clique S of at most 4, where c_t is some constant, set to ensure that

$$\tilde{\mathbb{E}} (\sum x_i)^t = \omega^t$$

for $t = 1, 2, 3, 4$. (Since we expect about $N_t = 2^{-\binom{t}{2}} \binom{n}{t}$ t-cliques in the graph, c_t would be roughly n^t / N_t.) We also simplify every monomial to a multilinear one using the $x_i^2 = x_i$ substitution.

If S is not a clique, then we set $\mathbb{E} x_S = 0$.

Satisfying the constraints We satisfy $x_i^2 = x_i$ by definition. Also, since we set $\mathbb{E}x_S = 0$ for every S containing a non-edge, the constraint $\{x_i x_j = 0\}$ for $(i,j) \notin E$ is satisfied as well. Finally, for every i, we set $\tilde{\mathbb{E}}x_i = \omega/n$ and hence $\tilde{\mathbb{E}} \sum x_i = \omega$.

Proving PSD-ness We will prove that $\tilde{\mathbb{E}}P^2 \geq 0$ for every P by directly proving that the corresponding matrix is p.s.d. We will focus on the main part of the matrix, that corresponds to degree 4 monomials, leaving the other parts for the exercise.

So, we will think of the following matrix $N \times N$ matrix matrix M' where $N \sim \frac{1}{2}\binom{n}{2}$ is the number of edges in G:

For every $\{i,j\},\{k,\ell\} \in E(G)$, if $\{i,j\} = \{k,\ell\}$ then $M'_{\{i,j\},\{k,\ell\}} = c_2(\omega/n)^2$, if $\{i,j,k,\ell\}$ is a 4-clique then $M'_{\{i,j\},\{k,\ell\}} = c_4(\omega/n)^4$, and otherwise $M'_{\{i,j\},\{k,\ell\}} = 0$.

Why is this the right moment matrix to look at? The actual pseudo expectation matrix would contain a row and column for every monomial of degree at most 2. However, it can be simplified as follows:

- Since it satisfies the constraint $x_i^2 = x_i$, we can assume the degree of each variable is at most 1, and so need to have rows and columns for subsets of $[n]$ with size 1 and 2. We ignore the rows and columns for singletons in this discussion, and leave it as an exercise to complete those.
- Since it satisfies the constraint $x_i x_j = 0$ for every $(i,j) \notin E$, the rows corresponding to non-edges would be identically zero and can be removed.
- If $\{i,j,k,\ell\}$ is a 3-clique (i.e., $\{i,j\}$ and $\{k,\ell\}$ are two edges of a triangle) then the corresponding entry should be $x_3(\omega/n)^3$. However, we zero them out for now, and hint how to handle them later.

We can assume the matrix M' is regular (every row sum equals the same number λ), and so it has the all 1's vector as an eigenvector with the corresponding eigenvalue being λ. Therefore M' is psd if and only if $M' - \lambda J$ is non-negative where J is the matrix with all entries equalling $1/N$. Note that M' has now both positive and negative entries that average to zero. For a given pair of edges $\{i,j\}$, $\{k,\ell\}$, the probability that the other 4 edges exist in the graph to form a 4-clique is $2^{-4} = 1/16$, and therefore in M' $1/(16)^{th}$ of the non-diagonal entries were positive and the rest zero.

After scaling by $16c_4^{-1}(n/\omega)^4$, we see that M' is psd if and only if the matrix $c(n/\omega)^2 I + M$ is p.s.d where $c > 0$ is some constant, I is the $N \times N$ identity, and M is the matrix defined as

$$M_{\{i,j\},\{k,\ell\}} = \begin{cases} +16 & \{i,j,k,\ell\} \text{ is a 4 clique} \\ -1 & |\{i,j,k,\ell\}| = 4 \text{ but it is not a 4 clique} \\ 0 & \text{otherwise} \end{cases}$$

for every pair of edges $\{i,j\}$, $\{k,\ell\}$.

Bottom line It is enough to prove the following lemma:

Lemma: If $\omega \ll n^{1/8}$ then (with high probability over the choice of the graph) $\|M'\| \ll (n/\omega)^2$.

Proving this lemma The obvious way to try to prove that a matrix is psd is via diagonal dominance. That is, we would want to prove that for every edge $\{i,j\}$,

$$\sum_{\{k,\ell\}\in E(G)} |M_{\{i,j\},\{k,\ell\}}| \ll (n/\omega)^2$$

Unfortunately this would not work— there are $\Omega(n^2)$ edges, and each entry of M is $\Omega(1)$, and so this sum has value $\Omega(n^2)$.

Second attempt We are going to use the following very useful inequality

Trace bound: For every matrix M and even integer t, $\|M\| \leq \mathrm{Tr}(M^t)^{1/t}$.

Proof: $\mathrm{Tr}(M^t) = \sum \lambda_i^t \geq \max_i \lambda_i^t$, where $\lambda_1, \lambda_2, \ldots$ are the eigenvalues of M.

Calculating the trace The simplest case to calculate is $\mathrm{Tr}(M^2)$ which is simply the sum of squares of all entries. However, since there are $\Omega(n^4)$ entries each with $\Omega(1)$ value, this would still only give us a useless bound of $\sqrt{\Omega(n^4)} = \Omega(n^2)$ on the norm.

The case $t = 4$ actually is useful. We will compute the expectation of $\mathrm{Tr}(M^4)$ over the choice of a random graph. Note that by Markov, with high probability this quantity won't be much higher than its expectation. Since we are trying to show that $\|M\| \ll (n/\omega)^2$, our goal is to show that $\mathbb{E}\mathrm{Tr}(M^4) \leq o(n^8/\omega^8)$ which in our setting of $\omega = o(n^{1/8}$ means that we need to show

$$\mathbb{E}\mathrm{Tr}(M^4) \leq O(n^7)$$

This is actually not that hard: let e_1, e_2, e_3, e_4 be 4 pairs of vertices. If all vertices in these pairs are distinct, then conditioned on the event that $e_1, \ldots, e_4$ are actually edges, the 4 events $E_1, \ldots, E_4$ are independent where E_i is the event that $e_i \cup e_{i+1}$ is a 4-clique (letting $e_5 = e_1$). The contribution of $e_1, \ldots, e_4$ to the trace of M^4 is $\sigma_1\sigma_2\sigma_3\sigma_4$ where σ_i equals $+15$ if E_i happens and -1 if E_i doesn't happen. Note that $\mathbb{E}\sigma_i = 0$ and since all these events are independent, in expectation the contribution of e_1, e_2, e_3, e_4 to the trace is zero. Thus, all the contribution to the trace must come from 4-tuples of edges that correspond to at most 7 distinct vertices, but there are $O(n^7)$ such tuples, and each one can contribute at most a constant to the trace, hence concluding the result. (A slightly more sophisticated argument can show that in fact the contribution comes from 4-tuples corresponding to at most 6 distinct vertices, yielding a bound of $O(n^6)$ that will translate to requiring merely $\omega \ll n^{1/4}$. Going beyond this requires some more effort, and as we see below, going beyond $\omega = n^{1/3}$ requires changing the moments.)

Can these moments show $\omega \sim \sqrt{n}$? A natural question is whether our analysis is tight. First, perhaps we could have gotten a better bound on the norm by using a higher power of the trace. Second, perhaps the matrix could be psd even if we had $\|M\| \gg (n/\omega)^2$, as long as the reason for this high norm was that M has large *positive* (as opposed to negative) eigenvalues. Indeed, Meka and Wigderson original claim was that essentially the same moments can yield a proof for $\omega \sim \sqrt{n}$, and while their proof had a bug it was not clear that this claim is false.

Nevertheless, we will now see that it is in fact false when $\omega \gg n^{1/3}$. In some sense the moral is that those moments are "randomer than possible" and candidate moments should encode more of the information that is present in the graph. I should note that I am involved in ongoing discussions with Chan, Kelner, Meka, Potechin, Steurer, and Wigderson on trying to improve this lower bound and better understand SOS and planted clique in general, and the observation below arose out of these discussions.

Let $r \in \{\pm 1\}^n$ be the vector such that $r_1 = 0$ and for $i > 1$:

$$r_i = \begin{cases} +1 & \{i,j\} \in E(G) \\ -1 & \{i,j\} \notin E(G) \end{cases}$$

Let $\epsilon > 0$ be some constant to be determined later, and consider the polynomial

$$P(x) = \langle r, x\rangle^2 - \epsilon\omega^2 x_1$$

I claim that that the pseudo distribution defined above satisfies

$$\tilde{\mathbb{E}} P(x)^2 < 0$$

Indeed, lets write

$$\tilde{\mathbb{E}} P(x)^2 = \tilde{\mathbb{E}}\langle r, x\rangle^4 - 2\epsilon\omega^2 \tilde{\mathbb{E}}\langle r, x\rangle^2 x_1 + \epsilon^2\omega^4 \tilde{\mathbb{E}} x_1^2$$

Let us now compute the pseudo expectation of each of these terms separately:

The last term is easy to compute: since $\tilde{\mathbb{E}} x_1^2 = \tilde{\mathbb{E}} x_1$,

$$\epsilon^2\omega^4 \tilde{\mathbb{E}} x_1^2 = \epsilon^2\omega^4(\omega/n) = \epsilon^2\omega^5/n$$

For the first term, note that

$$\tilde{\mathbb{E}}\langle r, x\rangle^4 = \sum_{i,j,k,\ell} r_i r_j r_k r_\ell \tilde{\mathbb{E}} x_i x_j x_k x_l \sim 2c_2(\omega/n)^2 \sum_{\{i,j\}\in E(G)} r_i^2 r_j^2 c_4(\omega/n)^4 \sum_{\{i,j,k,\ell\}\text{ 4-clique}} r_i r_j r_k r_\ell$$

(I am ignoring the terms arising when $|\{i,j,k,\ell\}|$ is 1 or 3, since they will be dominated by the other terms)

If $1 \in \{i,j,k,\ell\}$ then $r_i r_j r_k r_\ell$ vanishes. Otherwise, if they are distinct conditioning on $\{i,j,k,\ell\}$ being a clique, the choice for r_i, r_j, r_k, r_ℓ and so in expectation the contribution is zero, and hence

$$\mathbb{E}_G \tilde{\mathbb{E}}_x \langle x, u\rangle^2 \leq O((\omega/n)^2 n^2) = O(n^2)$$

and by Markov this will happen with high probability.

For the middle term we see that

$$\tilde{\mathbb{E}}\langle r, x\rangle^2 x_1 = \sum_{i,j} r_i r_j \tilde{\mathbb{E}} x_i x_j x_k = c_3(\omega/n)^3 \sum_{i,j\text{ s.t. }\{1,i,j\}\text{ is triangle}} r_i r_j \ .$$

However, since $\{1,i,j\}$ being a triangle implies that $r_i r_j = +1$, we see that the last sum is simply the number of triangles that 1 is involved in, which in expectation is $n^2/8$ (and with high probability close to its expectation) and so we see that

$$\tilde{\mathbb{E}}\langle r, x\rangle^2 x_1 = \Omega(\omega^3/n)$$

This means that the middle term is smaller than

$$-\Omega(\epsilon\omega^5/n)$$

where the constant in the $\Omega(\cdot)$ notation are independent of ϵ and so for sufficiently small ϵ this will dominate the last term which was $+\epsilon^2\omega^5/n$.

But if $\omega^3 \gg n$ then $\omega^5/n \gg \omega^2$ and hence the middle term will dominate the first term as well, and the expectation would be negative.