SOS Lecture 3: Lower Bounds — 3SAT/3XOR and Planted Clique / Boaz Barak

These notes are an expanded version of the notes for my summer course scribed by Akash Kumar

In this lecture we will see some *lower bounds* (or more accurately, negative results) for the Sum of Squares algorithm. Namely, we will see computational problems which the SOS algorithms *fails* to solve with a small degree. Another way to say this is that we will see cases that demonstrate the difference between pseudo-distributions and actual distributions. If we take the point of view that pseudo-distributions capture the knowledge of a computationally-bounded observer, then these are examples where being computationally bounded has very a significant effect on this knowledge. We have jokingly referred to "Marley's Corollary" as roughly saying that as long as you don't use the probabilistic method then "every little thing gonna be alright" and pseudo-distributions can be counted on to be similar to actual distributions. Thus it's not surprising that we will in fact use the probabilistic method in constructing these examples.

1 Lower bounds for random 3SAT/3XOR

In the Max-3XOR problem, we are given a set of linear equations $\pmod 2$ in n Boolean variables $x_1, \ldots, x_n$ such that each equation only involves three variables (i..e., has the form $x_i \oplus x_j \oplus x_k = a_{i,j,k}$), and we need to find the assignment x that satisfies the largest number of equations. (For simplicity of notation, we will actually think of x a string in $\{\pm 1\}^n$, meaning that the equations have the form $x_i x_j x_k = a_{i,j,k}$ for some $a_{i,j,k} \in \{\pm 1\}$.) Finding whether or not there exists an assignment that satisfies *all* equations can of course be done via Gaussian elimination, but Håstad proved in 1995 that for every $\epsilon > 0$ it is NP-hard to distinguish between the case that there is an assignment satisfying $1-\epsilon$ fraction of the equations, and the case that every assignment satisfies $1/2+\epsilon$ fraction of the equations. (There is always an assignment satisfying $1/2$ of the equations— can you see why?.) In fact, since Håstad's reduction only had linear blowup from the underlying PCP system (known as "Label Cover"), and thanks to the work Moshkovitz and Raz we now know of such PCP systems which themselves have a quasilinear blowup from 3SAT, if we assume that there is no $2^{n^{0.999}}$ algorithm for 3SAT (an extremely reasonable assumption which is in fact weaker than what's known as the "Exponential Time Hypothesis") then the SOS algorithm would require degree at least $n^{0.999}$ to do so. However, it is always good to verify these predictions by proving them unconditionally, which is what is achieved (in a very strong form) by the following theorem:

Theorem 1 (Grigoriev, 1999)**.** *For every constant $\epsilon > 0$ and large enough n, there exists an instance ψ of Max-3XOR over n variables such that:*

- *Every assignment $x \in \{\pm 1\}^n$ satisfies at most $1/2+\epsilon$ fraction of the equations of ψ.*
- *There exists a degree $\Omega(n)$ pseudo distribution $\{x\}$ that is consistent with the constraints $\{x_i^2 = 1\}$ for all $i \in [n]$ and the constraint $\{x_i x_j x_k = a_{i,j,k}\}$ for every i,j,k such that ψ contains the equation $a_{i,j,k} x_i x_j x_k = 1$.*

Moreover, there is $m = O(n)$ such that with constant positive probability, a random ψ with m equations will satisfy the properties above. (I believe this constant probability can actually be upgraded to $1-o(1)$ at the expense of a slight complication in the proof; ***Exercise 1:*** *verify this, see footnote for hint*[1]*)*

[1] **Hint:** The reason that a random instance is not an expander with high probability is that there may be a few pairs of 3-variable equations

Note that this theorem is stronger than what is predicted by the NP-hardness results, as it says that SOS cannot even distinguish between the case that the equations are completely satisfiable and the case that one can satisfy at most a $1/2+\epsilon$ fraction, which as we mentioned can in fact be done in polynomial time via Gaussian elimination. So how come this powerful algorithm does not solve this easy problem? One answer is that while the SOS algorithm may be optimal in some domains, it does not mean it's optimal for all problems.[2] In particular, it does not seem able to take advantage of algebraic structure such as the one present in linear equations. Another related observation is that, unlike some other algorithms, the SOS algorithm (at least when applied to natural simple systems of equations such as those arising from constraint satisfaction problems) doesn't seem to do "half measures" in the sense that it is *inherently robust to noise*. That is, because of its continuous nature, the SOS algorithm does not really distinguish between the case that x satisfies all the equations (i.e. $\sum_{i,j,k} a_{i,j,k}x_ix_jx_k = m$) and the case that it satisfies almost all of them (i.e. $\sum_{i,j,k} a_{i,j,k}x_ix_jx_k \geq (1-\epsilon)m$) . This is in stark contrast to algebraic algorithms such as Gaussian elimination that are very *brittle*, and completely break down even in the presence of very small amounts of noise. Therefore, if the SOS algorithm would have solved the "1 vs. $1/2+\epsilon$" Max-3XOR problem, then it would also have been able to solve the "$1-\epsilon$ vs. $1/2+\epsilon$" variant of the problem, but this latter variant is NP-hard.

Some bibliographical remarks: Theorem 1 was proven by Grigoriev in 2001, and later rediscovered by Schoenebeck in 2008. Schoenebeck also observed that it immediately implies a lower bound for 3SAT, since $a \oplus b \oplus c = 1$ implies that $a \vee b \vee c = 1$, and a random 3SAT instance is unsatisfiable. Tusliani extended this further by first showing that the same ideas can be used to show a lower bound for a very particular type of the "Label Cover" problem, which can be thought of as an "SOS PCP theorem". He then showed that SOS lower bounds are in general closed under "gadget reductions" and so managed to transform many of the NP-hardness results obtained from the PCP theorem into matching unconditional SOS lower bounds. Siu On Chan (2013) provided a very interesting result in the other direction, giving an actual PCP Theorem that exactly matches the parameters of the "SOS PCP Theorem". An excellent question asked in class is whether there is an algorithm that combines both SOS and Gaussian elimination in a natural way. I don't know of an algorithm for this, but there is a proof system, which simply uses the same rules of derivation $\{P \geq 0, Q \geq 0\} \models \{P+Q \geq 0, PQ \geq 0\}$ as the SOS system but tracks the *actual* degree as opposed to the *syntactic* degree, see this paper of Grigoriev, Hirsch, and Pasechnik `http://eccc.hpi-web.de/report/2001/103/`. A proof system is still very interesting since it can demonstrate that a problem lies in $NP \cap coNP$ and (as we discussed) there are very few examples for problems in this class that are not also known to be in P. We know very few lower bounds for this system, though the NP completeness results imply that there should be a 3XOR instance where one can satisfy at most, say, 0.51 fraction of the equations, but the best upper bound proven by this system with degree $\ll n$ would not be better than 0.99.

that have two common variables. The effect of this should be the same as adding a small number of equations of the form $x_i \oplus x_j = \sigma$ (or $x_ix_j = \sigma$ in $\{\pm 1\}$ notation) to the instance, and one should be able to show that the pseudo-distribution we construct can be modified to satisfy these constraints as well.

[2] Throughout this course, when saying that the SOS algorithm solves a problem X, I always assume that the representation of X as polynomial equations is fixed to some canonical form. If we allowed arbitrary polynomial-time computable representations then we could simulate any algorithm using the SOS algorithm as even linear programming is **P**-complete.

1.1 Proving Theorem 1

To prove Theorem 1 we need to (1) give a construction of some highly unsatisfiable 3XOR instance ψ and (2) construct a degree $O(n)$ pseudo-distribution $\{x\}$ that pretends to satisfy all the constraints of ψ . As mentioned, the construction of ψ is simple - we simply choose it as a random 3XOR instance with $m = cn$ constraints for some constant c (depending on ϵ).[3] Let us think of the choice of ψ as choosing a bipartite graph on G on $m + n$ vertices with left-degree 3 and a random $a \in \{\pm 1\}^m$ such that the ℓ^{th} equation is that $a_\ell x_i x_j x_k = 1$ where $\{i, j, k\}$ are the neighbors of ℓ. The theorem follows from the following lemmas:

Lemma 2. *For every G, with $1 - \exp(-\Omega(n))$ probability over the choice of $a \in \{\pm 1\}^{cn}$ it will hold that every assignment x satisfies at most $1/2 + \epsilon$ fraction of the equations where ϵ tends to zero as c grows.*

Lemma 3. *There is some $\delta, p > 0$ such that with probability at least $p > 0$ the graph G is a $(\delta n, 1.7)$-expander, where we say that a bipartite graph $G = (L, R, E)$ is a (s, α) expander if $|\Gamma(S)| \geq \alpha |S|$ for every $S \subseteq L$ with $|S| \leq s$, where $\Gamma(S)$ denotes the set of neighbors of S.*

Lemma 4. *For every $a \in \{\pm 1\}^m$, if G is a (k, α) expander for $\alpha > 1.5$ then there exists a degree $k/100$ pseudo-distribution $\{x\}$ consistent with the constraints $\tilde{\mathbb{E}} x_i x_j x_k = a_\ell$ for every ℓ and $\{i, j, k\} = \Gamma(\ell)$.*

We defer the proofs of Lemmas 2 and 3. The proofs use, as promised, the probabilistic method, but are not complicated and follow by the usual Chernoff+union bound argument. Despite being simple, the fact that they use the probabilistic method is in some sense the reason that they do not carry over to the SOS setting. This serves as a caution that we shouldn't equate a proof being "SOS'able" with it being "simple"— there can be highly complex proof in the SOS setting, and every simple proof that is not "SOS'able" with low degree. I believe that Lemma 3 actually can be derandomized using the paper of Capalbo, Reingold, Vadhan, Wigderson. One way to demonstrate that Marley's corollary is not a universal rule is to do **Exercise 2:** Prove that there is no SOS proof that the CRVW graph is a (k, α) expander for $\alpha > 1.5$. (I actually don't know if this exercise is true, and if it is I think it would actually be an interesting research result, showing a "robust" SOS lower bound with a deterministic construction.) Alekhnovich (2001) had a fascinating conjecture that as long as G is a sufficiently good expander, the instance (G, a) with a random a would be hard.

Note that despite having a very simple proof, I don't know of any derandomization for Lemma 2.

1.2 Proof of Lemma 4

To prove Lemma 4 we need to show the existence of a degree $\Omega(n)$ pseudo-distribution $\{x\}$ that pretends to range over $x \in \{\pm 1\}^n$ that satisfies the constraints of the system (G, a). Our philosophy is that a pseudo-distribution captures the knowledge of *computationally bounded* observer. Note that (actual) distributions are the standard way to model knowledge of a *computationally unbounded* observers that only have *partial information*— this is known as Bayesian reasoning. For example, suppose that Mickey is a computationally all-powerful observer. If Mickey was given no information at all then we model its knowledge by the uniform distribution over $\{\pm 1\}^n$. Note that this distribution satisfies that $\mathbb{E} x_S := \mathbb{E} \prod_{i \in S} x_i = 0$ for every non empty set S. (Note that since

[3] We will use the words equations, constraints and clauses interchangebly.

the distribution is over x's that satisfy $x_i^2 = 1$ then knowing $\mathbb{E}\prod_{i\in S} x_i$ for every set S suffices to deduce $\mathbb{E}p(x)$ for every polynomial $p(\cdot)$.)

If Mickey later learns that $x_7x_{13}x_{22} = -1$ and $x_{22}x_{44}x_{60} = +1$ then we model his knowledge by the distribution that is uniform over the the strings that satisfy these conditions— namely $\mathbb{E}x_S = 0$ unless S is either empty or one of $\{7, 13, 22\}, \{22, 44, 60\}$ and $\{7, 13, 44, 60\}$.

Now if Mickey was given all the equations, then, being computationally unbounded, then if the equations come from Lemma 2 he would be to figure out that there exists no x that satisfies all the equations (or even a 0.6 fraction of them), and hence that there simply exists no such distribution. However, if these equations are given to the Donald that can only do n^s-time computation, then he might not be able to do that. Specifically, Donald could deduce from $x_7x_{13}x_{22} = -1$ and $x_{22}x_{44}x_{60} = +1$ that $x_7x_{13}x_{44}x_{60} = -1$ etc.. but not able to draw all possible logical inferences and hence figure out that there is no solution x to these equations. Thus, just in the case of Mickey, we design the pseudo-distribution to be as random as possible subject to being consistent with the deductions Donald is able to make, in other words we follow Einstein's maxim that

Pseudo-distributions should be as random as possible but not randomer

More concretely we will assume that Donald's knowledge only applies to terms of the form $\prod_{i\in S} x_i$ for $|S| \leq s$, and that given subsets S, T such that $|S|, |T| \leq s$, if $U = S \oplus T$ has size at most s, then he can deduce that $x_U = x_S x_T$, but these are all the deductions he can make. Specifically we define the pseudo-distribution $\{x\}$ by the following iterative process:

- Input to the process: graph $G = ([m] \cup [n], E)$ and string $a \in \{\pm 1\}^m$.
- For every $\ell \in [m]$, define $\tilde{\mathbb{E}}x_{\Gamma(\ell)} = a_\ell$.
- Apply the following rule until we can't apply it any further: for every subsets S, T of size at most s such that $\tilde{\mathbb{E}}x_S$ and $\tilde{\mathbb{E}}x_T$ are defined, $|S \oplus T| \leq s$, define $\tilde{\mathbb{E}}x_{S\oplus T} = \tilde{\mathbb{E}}x_S\tilde{\mathbb{E}}x_T$. (If $\tilde{\mathbb{E}}x_{S\otimes T}$ was already defined before with a different value, the process fails and halts.)
- When done, for every nonempty $|S| \leq s$ such that $\tilde{\mathbb{E}}x_S$ is undefined, define $\tilde{\mathbb{E}}x_S = 0$. For every monomoial $m(x)$ of degree at most s that contains square terms define $\tilde{\mathbb{E}}m(x) = \tilde{\mathbb{E}}x_S$ where S is the set of i's such that x_i appears with an odd degree in $m(x)$.

We need to prove that $\{x\}$ defined above is a valid pseudo-distribution. For starters, we need to verify that it never halts:

Lemma 5. *If the G is a $(10s, 1.7)$ expander then the process above never halts.*

Before proving the lemma, lets see why it implies that $\{x\}$ is a valid pseudo-distribution for degree at most $s/2$. First note (**Exercise 3:** check this) that by definition, we get that for every polynomial p of degree at most $s - 3$ and every $S = \Gamma(\ell)$, $\tilde{\mathbb{E}}p(x)x_S = a_\ell\tilde{\mathbb{E}}p(x)$. Thus we need to prove that for every polynomial q of degree at most $s/2$, $\tilde{\mathbb{E}}q^2 \geq 0$. **Exercise 4:** Prove that it suffices to do so for the case that q is *multilinear* namely $q(x) = \sum_{|S|\leq s/2} \alpha_S x_S$ for some real numbers $\{\alpha_S\}_{|S|\leq s/2}$.

For two subsets S, T of size at most $s/2$, we say that $S \equiv T$ if $\tilde{\mathbb{E}}x_{S\oplus T} \neq 0$. Note that this is indeed an equivalence relation— it is symmetric, reflexive (since $\tilde{\mathbb{E}}x_\emptyset = \tilde{\mathbb{E}}1 = 1$), and transitive, since if $S \equiv T$ and $T \equiv U$ then $\tilde{\mathbb{E}}x_{S\oplus U} = \tilde{\mathbb{E}}x_{(S\oplus T)\oplus(T\oplus U} = \tilde{\mathbb{E}}x_{S\oplus T}\tilde{\mathbb{E}}x_{T\oplus U}$. Separate the monomials of q into the equivalence classes and so write $q = \sum_{i=1}^m q_i$ where all monomials in q_i belong to a particular equivalence class. Note that if S and T are not equivalent then $\tilde{\mathbb{E}}x_S x_T = \tilde{\mathbb{E}}x_{S\oplus T}$ equals zero, and hence

$$\tilde{\mathbb{E}}q^2 = \tilde{\mathbb{E}}(\sum_i q_i)^2 = \sum \tilde{\mathbb{E}}q_i^2$$

Thus it suffices to show that $\tilde{\mathbb{E}}q^2 \geq 0$ where $q(x) = \sum_{S\in\mathcal{C}} \alpha_S x_S$ where $\mathcal{C}$ is some fixed equivalence class. Let S_0 be a member of that class. Thus for every $S, T \in \mathcal{C}$, $\tilde{\mathbb{E}}x_{S\oplus S_0} \neq 0$ and $\tilde{\mathbb{E}}x_{T\oplus S_0} \neq 0$ and so (by our rule) $\tilde{\mathbb{E}}x_{S\oplus T} = \tilde{\mathbb{E}}x_{S\oplus S_0}\tilde{\mathbb{E}}x_{T\oplus S_0}$. Therefore,

$$\tilde{\mathbb{E}}q^2 = \sum_{S,T\in\mathcal{C}} \alpha_S\alpha_T\tilde{\mathbb{E}}x_S x_T = \sum_{S,T\in\mathcal{C}} \alpha_S\alpha_T(\tilde{\mathbb{E}}x_{S\oplus S_0})(\tilde{\mathbb{E}}x_{T\oplus S_0}) = \Big(\sum \alpha_S\tilde{\mathbb{E}}x_{S\oplus S_0}\Big)^2 \geq 0$$

□

1.3 Proof of Lemma 5

[BOAZ: this proof is copy pasted from previous lecture notes and so uses somewhat inconsistent notation. On a very high level, the proof goes as follows: for a set E of equations, let $\underline{\Gamma}(E)$ denote $\oplus_{\ell\in E}\Gamma(\ell)$, i.e. $\overline{\Gamma}(E)$ is the set of variables that appear an odd number of times in the equations in E. If we have a derivation $U_1, \ldots, U_t$ of sets of size at most d such that either U_i is the variables of an equation (i.e., $U_i = \Gamma(\ell)$) or the value $\tilde{\mathbb{E}}U_i$ is always derived from prior values $\tilde{\mathbb{E}}U_j$, $\tilde{\mathbb{E}}U_k$ for $j, k < i$, then we can keep track of the sets of equations E_i that correspond to every U_i (e.g., if $U_i = \Gamma(\ell)$ then $E_i = \{\ell\}$ and in the other case $E_i = E_j \oplus E_k$. Note that $U_i = \underline{\Gamma}(E_i)$ and the value derived to $\tilde{\mathbb{E}}U_i$ is equal to $\prod_{\ell\in E_i} a_{\Gamma(\ell)}$. We then use the expansion property to argue that for every i, $|E_i| \leq 10s$, and $|\underline{\Gamma}(E)| \geq |E_i|/10$ for all E with $|E| \leq 100s$. But this implies that U_i uniquely determines E_i, since if we had $U_i = U_j$ but $E_i \neq E_j$ then we would get that $\emptyset = U_i \oplus U_j = \underline{\Gamma}(E_i) \oplus \underline{\Gamma}(E_j) = \underline{\Gamma}(E_i \oplus E_j)$. But since $|\underline{\Gamma}(E_i \oplus E_j)| \geq |E_i \oplus E_j|/10$, if $E_i \neq E_j$ then the set $\underline{E_i \oplus E_j}$ can't be empty. Thus the value for a set U that can be derived in some way is always uniquely defined as $\oplus_{\ell\in E} x_{\Gamma(\ell)}$ no matter how we derived it.]

Observe that if there is a derivation that gives different values to some monomial U of degree at most d, then there is also another derivation that shows $\tilde{\mathbb{E}}\left[x_\phi\right] = -1$. We will show that this leads to a contradiction. Let the derivation of $\tilde{\mathbb{E}}\left[x_\phi\right] = -1$ be described by a sequence of set $U_1, U_2, \cdots U_t$ with $U_t = \phi$. Notice that each U_i in this derivation is either some constraint in $\mathcal{E}$ or a product of some of the constraints from $\mathcal{E}$. Let $E_1, E_2 \cdots E_m$ be the sets corresponding to the equations in $\mathcal{E}$ and $\sigma_1, \sigma_2, \cdots \sigma_m$ be the corresponding values. For every U_i that is derived, we see that there is some $S_i \subseteq L$ for which $U_i = \cup_{\ell in \mathbb{S}_i} E_\ell$.

Now, we will see that the assumption of there being two different derivations for $\tilde{\mathbb{E}}\left[x_\phi\right]$ with opposite values leads to the conclusion that the above algorithmic process must assign some value to $\tilde{\mathbb{E}}\left[x_U\right]$ for some $|U| > d$ which is not possible. This conflicts with the assumption that there are 2 different derivations for the same set with opposite values. The plan is roughly the following. First we observe that every derived monomial with short derivation spans many variables (by expansion property) (has large degree). The idea is to show that two derivations with opposite values means that there is some monomial with degree greater than d that the algorithm defines.

Claim 6. *For every $S \subseteq L$ with $|S| \leq 100d$, $|\oplus_{\ell\in S} E_\ell| \geq |S|/10$.*

Proof. Suppose not. Let $T = \oplus_{\ell \in S} E_\ell$. Observe that the monomial in T may not all of the variables from $\Gamma(S)$ (that is, those which appear in $\Gamma(S) \setminus T$) . So, these variable nodes must have at least 2 neighbors in S as they need to cancel out. This gives (by expansion property)

$$\begin{aligned}
&\#\text{edges leaving } |S| \geq \#\text{ edges entering } S \text{ from the ``omitted'' variables in } \Gamma(S) \setminus T \\
&\qquad \implies 3|S| \geq 2(1.7|S| - |T|) \\
&\qquad \implies |T| \geq 0.2|S|
\end{aligned}$$

□

As observed already the above claim asserts that whatever you derive by XORing a small subset of clauses on the left gives rise to a monomial with decent degree.

Corollary 7. *Notice that this means that every set S_i in the derivation must have size no larger than $10d$.*

Proof. A quick proof. Suppose not – then the first such violating set $S_i = S_j \oplus S_k$ where S_j and S_k both describe short derivations and S_i is a bigger derivation with length in the interval $(10d, 20d]$. And by the above observation, this would mean $\oplus_{\ell \in S_i} E_\ell > d$. □

Thus, the existence of 2 sets with different values implies that there is a set $S = S_t$ with size at most $10d$ such that $\oplus_{\ell \in S} E_\ell = \phi$ contradicting the observation above.

For $P : \{\pm 1\}^k \to \pm$, we define a Max-P instance ψ to be a collection of equations of the form $P(a_{i_1} x_{i_1}, \ldots, a_{i_k} x_{i_k}) = a$, and the goal is again to find an assignment x satisfying as many of the equations as possible. We say that P is a *nice subspace predicate* if there is some subspace $V \subseteq GF(2)^k$ such that $P = 1_V$ (using the identification $GF(2) \leftrightarrow \{\pm 1\}$ using the map $n \leftrightarrow (-1)^b$) and such that every $u \in V^\perp$ satisfies $|u| \geq 3$.

As was noted by Tulsiani, Chan, the proof of Theorem 1 generalizes to the following statement: **Exercise 5:** Prove the following theorem.

Theorem 8. *For every nice subspace predicate $P = 1_V$, constant $\epsilon > 0$ and large enough n, there exists an instance ψ of Max-P over n variables such that:*

- *Every assignment $x \in \{\pm 1\}^n$ satisfies at most $|V|/2^k + \epsilon$ fraction of the equations of ψ.*
- *There exists a degree $\Omega(n)$ pseudo distribution $\{x\}$ that is consistent with the constraints $\{x_i^2 = 1\}$ for all $i \in [n]$ and the constraint corresponding to every equation in ψ.*

Moreover, there is $m = O(n)$ such that with constant positive probability, a random ψ with m equations will satisfy the properties above.

1.4 Proof of Lemmas 3 and 2

[BOAZ — these two proofs are also at the moment copy-pasted from previous notes and use inconsistent notation.]

Proof. We will prove the result for the graphs G_ϕ obtained from a random k-xor instance with n variables and γn constraints for some large constant Γ. In fact, the same proof holds for a random instance of any *Max-K-CSP* over any alphabet. Let us begin by trying to understand what is the probability that a set fails to have large expansion. Let us say that in fact, there is a small set of variables $T \subseteq R$ that is the vertex boundary of some set S of size $s = \epsilon n$ where ϵ is some constant to be determined later. Let $|T| \le cs = (k-1-\delta)s$ denote the size of this small set of variables that appear in this set S of the random k-xor instance (where $\delta \in (0, \frac{1}{2})$). Observe that the probability p that this indeed happens – that a set S of size ϵn happens to have small vertex boundary can be upper bounded by

$$p \le \binom{n}{cs} \cdot \binom{\binom{cs}{k}}{s} \cdot s! \binom{\gamma n}{s} \cdot \binom{n}{k}^{-s}$$

Here,

- $\binom{n}{cs}$ denotes which cs variables to use.
- $\binom{\binom{cs}{k}}{s}$ denotes the number of ways cs variables can be put together to get s clauses each of arity k.
- $s!\binom{\gamma n}{s}$ counts the choices for where to put such clauses in our ordered sequence of γn clauses.
- And the last term, $\binom{n}{k}^{-s}$ denotes the probability that the clauses were generated using the described method.

Now, using Stirling's which says $\frac{a}{b}^b \le \frac{a}{b} \le \frac{ae}{b}^b$ and $s! \le s^s$, we obtain by collecting terms we get

$$\begin{aligned} p &\le \left(\frac{s}{n}\right)^{2\delta \cdot s/2} \left(e^{2k+1-\delta} k^{1+\delta} \gamma\right)^s \\ &\le \left(\frac{s}{n}\right)^{\delta s} \cdot \left(\gamma^5\right)^s \\ &= \left(\frac{s\gamma^{5/\delta}}{n}\right)^{\delta s} \end{aligned}$$

We need to show that the probability that any set of at most $s \le \epsilon n$ constraints contains less than cs variables is $o(1)$. To do this, consider the following.

$$\begin{aligned} \sum_{s=1}^{\epsilon n} \left(\frac{s\gamma^{5/\delta}}{n}\right)^{\delta s} &= \sum_{s=1}^{\ln^2 n} \left(\frac{s\gamma^{5/\delta}}{n}\right)^{\delta s} + \sum_{s=\ln^2 n+1}^{\epsilon n} \left(\frac{s\gamma^{5/\delta}}{n}\right)^{\delta s} \\ &\le O\left(\frac{\gamma^5}{n^\delta} \cdot \ln^2 n\right) + O\left(\epsilon \gamma^{5/\delta}\right)^{\delta \ln^2 n} \end{aligned}$$

Now, we are almost done. Observe that the first term is clearly $o(1)$ and that the second term is $o(1)$ for $\epsilon = O(\frac{1}{100\gamma^{5/\delta}})$. This gives us that indeed small sets of size ϵn fail to have a large boundary with large probability which is what we wanted. □

2 SOS Lower bounds for planted clique

The planted clique problem is one of the most classical computational problems, whose roots come from a 1976 question of Karp of whether we can find the largest clique in a random graph. In the early 1990's, Jerrum and Kucera suggested the easier *planted* model, whereby the goal is to find a ω-clique that has been added to a random graph. Note that if $\omega \gg \log n$ then this would be the unique maximum ω-clique in the graph. (**Exercise 6:** prove that with probability at least 0.99 a random $G(n, 1/2)$ graph has the maximum clique size of at most $c \log n$ for some constant c; how small can you make c? .) The larger ω is, the easier this problem. Another variant which seems to have equivalent difficulty is to distinguish between a random graph and a graph to which a random ω-clique was added. One easy bound on ω for this problem arises from the following exercises:

Exercise 7: Let A be the adjacency matrix of a random $G(n, 1/2)$ graph and $B = 2A - J$ where J is the all 1's matrix. **(1)** Prove that for every t, $\mathbb{E}\mathrm{Tr}(B^t) \leq 2^{O(t)} n^{t/2}$. **(2)** Conclude that with probability at least 0.99, $\|B\| \leq O(\sqrt{n})$.

Exercise 8: Let A be the adjacency matrix of a n-vertex graph with average degree $n/2$ that contains a ω clique and $B = 2A - J$. Prove that $\|B\| \geq \Omega(\omega)$.

Thus we can easily distinguish between a random $G(n, 1/2)$ graph and an n vertex graph containing a $\omega \gg \sqrt{n}$ clique, and in fact these ideas can be used to actually find the clique in the latter case (**Exercise 9:** Show this; see footnote for hint[4]). Many people have thought of improving this $\sqrt{n}$ bound but with no success, often proving that certain methods *won't* work. In fact by now the difficulty of this question has been conjectured in several works that have connected this problem to questions in machine learning, compressed sensing, computing equilibrium and more.

The algorithm that works for $\omega \sim \sqrt{n}$ can be thought of as an instantiation of the degree 2 SOS algorithm and thus we come again to the question of whether degree $d > 2$ SOS can do better than degree 2. As in the case of the Unique Games, Small-Set Expansion, Max-Cut, Cheeger etc.., the answer is that we don't know. But, given that this is an average case problem (such as the random 3XOR problem discussed above), one could perhaps hope that we will be able to prove some SOS lower bounds in this case. Indeed last year Meka and Wigderson claimed that for every constant d there is some $\epsilon > 0$ such that the degree d SOS algorithm cannot certify that a random graph doesn't contain an $\epsilon\sqrt{n}$ clique. However (as we will see) their proof was flawed. Nevertheless in a very new result, Meka, Potechin and Wigderson were able to prove a weaker result. Namely that the degree d SOS cannot certify that a random graph doesn't contain an $\tilde{\Omega}(n^{1/d})$ clique. We will see a (slight weakening of a) special case of their result, namely

Theorem 9. *Let $G = G(n, 1/2)$ be a random graph. With probability at least 0.9, there exists a degree 4 pseudo-distribution $\{x\}$ over $\mathbb{R}^n$ satisfying $\{x_i^2 = x_i\}$ for all i, $\{x_i x_j = 0\}$ for all i and j that are not neighbors in G, and*

$$\tilde{\mathbb{E}} \sum x_i \geq \Omega(n^{1/8})$$

2.1 Proof of Theorem 9

Once again, to construct a pseudo distribution, we think of a computationally bounded observer that is told that there is a planted clique in the graph, and needs to form beliefs about what is the probability that some set S with $|S| \leq 4$ is contained in the clique (or equivalently, what should be the expectation $\tilde{\mathbb{E}} x_S$). Clearly if S is not itself a clique, then this probability is zero. Otherwise, since both clique and surrounding graph are random, we would expect the probability

[4] **Hint:** Show that we can get a set S with large correlation with the clique by looking at the largest eigenvector of B, and then show that we can use to actually find the clique by looking at the set of vertices that have very large degree into S.

to be the roughly the same for every clique. This motivates defining our pseudo-distribution: for every set S of size at most 4, if S is not a clique then $\tilde{\mathbb{E}} x_S = 0$, and otherwise $\tilde{\mathbb{E}} x_S = 2^{\binom{|S|}{2}}(\omega/n)^{|S|}$ (where $\binom{1}{2} = 0$).[5] Analogous to what we did for 3XOR, we use the constraints that $x_i^2 = x_i$ to reduce every monomial $m(x)$ to a multilinear monomial of the form x_S. Note that we get that $\tilde{\mathbb{E}} \sum x_i = n(\omega/n) = \omega$.

It turns out that this simplistic pseudo-distribution is already sufficient to prove a weak lower bound on the clique size

Lemma 10. *If $\omega \ll n^{1/8}$ then the pseudo-distribution above is a valid degree 4 pseudo-distribution satisfying the conditions of Theorem 9*

On the other hand, we (and Meka-Wigderson) have violated Einstein's maxim and made this pseudo-distribution "randomer than possible" if we want to reach the $\omega \sim \sqrt{n}$ bound.

Lemma 11. *If $\omega \gg n^{1/3}$ then there exists a quadratic polynomial Q such that $\tilde{\mathbb{E}} Q^2 < 0$*

2.2 Proof of Lemma 10

Let $M_{a,b,c,d} = \tilde{\mathbb{E}} x_a x_b x_c x_d$. We need to prove that the matrix M is positive semidefinite. Let us focus our attention to the rows $\{a, b\}$ of M where $a \neq b$ and columns $\{c, d\}$ where $c \neq d$ (this turns out to be the crux of the proof). Since the entire row corresponding to $\{a, b\}$ will be zero if (a, b) is not an edge of$G$, we can think of M as an $E \times E$ matrix where E is the edge set of G. Recall that for every a, b, c, d $M_{a,b,c,d}$ depends solely on $|\{a, b, c, d\}$. Thus we can write $M = M^2 + M^3 + M^4$ where $M^s_{a,b,c,d} = M_{a,b,c,d}$ if $|\{a, b, c, d\}| = s$ and equals zero otherwise. Let us ignore M^3 for now (this is not the main issue) and so focus on proving that $M^2 + M^4$ is positive semidefinite. Note that M^2 simply contains all diagonal elements and each has magnitude $\tilde{\mathbb{E}} x_a x_b = 2\omega^2/n^2$. Therefore, we can scale by this number and reduce showing that $M^2 + M^4$ is psd to showing that $I + M'$ is p.s.d where I is the $E \times E$ identity, and M' is defined as follows:

$$M'_{a,b,c,d} = \begin{cases} 0 & |\{a, b, c, d\}| \neq 4 \\ 3\omega^2/n^2(1 - 1/16) & \{a, b, c, d\} \text{ is 4-clique} \\ -(3/16)\omega^2/n^2 & \{a, b, c, d\} \text{ is not a 4-clique} \end{cases}$$

Note that M' is simply M^4, scaled by $n^2/(2\omega^2)$ and subtracting from each entry $\{a, b, c, d\}$ with $|\{a, b, c, d\}|$ a constant so that the expected entry of M^4 is zero. The reason that this suffices follows from the following exercise:

Exercise 10: Let E be the $n^2 \times n^2$ *expectation* matrix of M. Namely, for every a, b, c, d, $E_{a,b,c,d}$ is the expected value of $M_{a,b,c,d}$ which is $2^{-\binom{s}{2}} 2^{\binom{s}{2}} (\omega/n)^s$ where $s = |\{a, b, c, d\}|$. Prove that E is p.s.d and its smallest nonzero eigenvalue is $\Omega(\omega^2/n^2)$.

This exercise is actually a special case follows from the theory of *Johnson Association Schemes* that is used in proving more general bounds. In particular the following is true (and may be useful for the previous exercise):

Exercise 11: Let $\ell \in \mathbb{N}$ and J be a matrix indexed by all subsets $S \subseteq [n]$ of size at most s such that $J_{S,T} = \binom{|S \cap T|}{\ell}$. Then J is psd. (If you get stuck, take a look at the Meka-Wigderson paper.)

[5] This is because conditioned on S being a clique, the probability that it is contained in the random ω-clique would be roughly $\binom{w}{|S|}$ divided by the the number of $|S|$-clique in the graph which is $\binom{n}{|S|} 2^{-\binom{|S|}{2}}$; we get the above probability using the approximation $\binom{n}{k} \sim \left(\frac{en}{k}\right)^k$.

The Frobenius norm squared of M', namely $\sum_{a,b,c,d}(M'_{a,b,c,d})^2$ equals $O(n^4\omega^4/n^4) = O(\omega^4)$. Note that the Frobenius norm squared is the sum of the eigenvalues squared, and thus if M' was "generic" or "pseudorandom" in the sense that it would have $\Theta(n^2)$ eigenvalues with roughly the same magnitude λ, then λ will satisfy $n^2\lambda^2 = O(\omega^4)$ or $\lambda \leq O(\omega^2/n)$. Thus in this case, as long as $\omega \ll \sqrt{n}$ the matrix $I + M'$ (and hence M) will be positive semidefinite. A priori you might hope that, sicne M' arises from a random graph then it will in fact be sufficiently "psuedorandom" to satisfy this, but as we will see in Lemma 11, this turns out to be false. Nevertheless we are able to prove the following claim:

Claim: w.h.p. $\mathrm{Tr}(M'^4) \leq O(\omega^8/n)$

The theorem follows from the claim since we get that $\|M'\| \leq \mathrm{Tr}(M'^4)^{1/4} \leq \omega^2/n^{1/4}$. Hence if $\omega^2 \ll n^{1/4}$ (or $\omega^8 \ll n$) then $I + M'$ will be psd.

Proof of Claim: By Markov we simply need to show that $\mathbb{E}\mathrm{Tr}(M'^4) \leq O(\omega^8/n)$. (Note that this is an actual expectation, taken over the random choice of the graph G; since the set E of edges depends on this randomness, it might be more convenient to think of M' as an $n^2 \times n^2$ matrix for this argument.) The expectation of the trace is the sum over all 4-tuples of edges e_1, e_2, e_3, e_4 of

$$\mathbb{E}M'_{e_1,e_2}M'_{e_2,e_3}M'_{e_3,e_4}M'_{e_4,e_1} \tag{1}$$

Now if e_1, e_2, e_3, e_4 are all disjoint, then, conditioned on e_1, e_2, e_3, e_4 being edges, the events "$e_1 \cup e_2$ is a 4-clique", "$e_2 \cup e_4$ is a 4-clique", etc.. are independent. Thus we get that (1) equals

$$\mathbb{E}M'_{e_1,e_2}\mathbb{E}M'_{e_2,e_3}\mathbb{E}M'_{e_3,e_4}\mathbb{E}M'_{e_4,e_1}$$

but $\mathbb{E}M'_{e,f} = 0$ for every pair of edges e, f by the construction of M'. Therefore, the contribution to the trace must come from 4-tuples of edges that are not all disjoint. Since each such 4-tuple involves at most 7 vertices, there are $O(n^7)$ such tuples, and since every entry as magnitude $O(\omega^2/n^2)$, the expectation of the trace is at most

$$O(n^7) \cdot O(\omega^2/n^2)^4 = O(\omega^8/n)$$

Exercise 12: Prove that in fact $\mathbb{E}\mathrm{Tr}(M'^4) \leq O(\omega^8/n^2)$, hence concluding that the pseudo-distribution is psd as long as $\omega \ll n^{1/4}$. See footnote for hint[6]

2.3 Proof of Lemma 11

Intuitively, one may hope that the pseudo-distribution above remains valid even for a larger value of ω, as long as $\omega \ll \sqrt{n}$. However, Lemma 11 shows this is not the case. To understand the reason, let's see that there is a simple observation available to the computationally-bounded Donald that would yield different results in this pseudo-distribution than it would have if it was truly a distribution over planted cliques. Specifically, for a n-vertex graph G, let $r_1, \ldots, r_n \in \mathbb{R}^n$ be the vectors such that

$$r_i(j) = \begin{cases} +1 & (i,j) \text{ is an edge} \\ 0 & i = j \\ -1 & \text{otherwise} \end{cases}$$

Let $P(x)$ be the polynomial $\sum\langle r_i, x\rangle^4$. Note the following facts about $P(x)$: (**Exercise 13:** verify those)

[6] **Hint:**Show that in fact the only nonzero contribution to the trace come from 4-tuples of edges e_1, e_2, e_3, e_4 such that $|e_1 \cup e_2 \cup e_3 \cup e_4| \leq 6$.

1. For every fixed $x \in \mathbb{R}^n$, if we choose the graph at random then $\mathbb{E}P(x) = O(n\|x\|^4)$.
2. If x is the 0/1 characteristic vector of an ω-clique (and hence $\|x\|^2 = \omega$) then $P(x) \geq \omega(\omega-1)^4 = \Omega(\omega^5)$.

Hence when $\omega^5 \gg n\omega^2$ (or $\omega \gg n^{1/3}$), $P(x)$ will distinguish between a "typical" vector x and a vector x that is the characteristic vector of an ω-clique. We claim that in the distribution $\{x\}$ above, $\tilde{\mathbb{E}}P(x)$ actually behaves as if x was "typical" and thus gives it too low a value:

Claim: With high probability $\tilde{\mathbb{E}}P(x) \leq O(n\omega^2)$.

Proof: Using Markov, it suffices to prove that $\mathbb{E}_G\tilde{\mathbb{E}}P(x) \leq O(n\omega^2)$ where the expectation is taken over the random choices in making the graph and the pseudo-expectation is as defined above. Lets open up the definition of $P(x)$ and write

$$\mathbb{E}\tilde{\mathbb{E}}P(x) = \sum_i \sum_{a,b,c,d \in [n]\setminus\{i\}} \mathbb{E}r_i(a)r_i(b)r_i(c)r_i(d)\tilde{\mathbb{E}}x_ax_bx_cx_d$$

(using the fact that $r_i(i) = 0$ for all i). Fix some $i \in [n]$, and now suppose that we fix the random choices of all neighbors in the graph except the neighbors of i. This means that $\tilde{\mathbb{E}}x_ax_bx_cx_d$ is determined for every $\{a,b,c,d\} \subseteq [n] \setminus \{i\}$ and the random $\{\pm 1\}$ variables $r_i(a), r_i(b), r_i(c), r_i(d)$ are independent of this choice. Thus we can write for every i and $a,b,c,d \in [n] \setminus \{i\}$, Therefore

$$\mathbb{E}r_i(a)r_i(b)r_i(c)r_i(d)\tilde{\mathbb{E}}x_ax_bx_cx_d = \tilde{\mathbb{E}}x_ax_bx_cx_d\mathbb{E}r_i(a)r_i(b)r_i(c)r_i(d)$$

Note that $\mathbb{E}r_i(a)r_i(b)r_i(c)r_i(d) = 0$ unless $a = b = c = d$ or $|\{a,b,c,d\}| = 2$. In the first case $\tilde{\mathbb{E}}x_ax_bx_cx_d = \omega/n$ and in the second case it equals $O(\omega^2/n^2)$. Hence for every i

$$\sum_{a,b,c,d \in [n]\setminus\{i\}} \mathbb{E}r_i(a)r_i(b)r_i(c)r_i(d)\tilde{\mathbb{E}}x_ax_bx_cx_d \leq n\omega/n + O(n^2\omega^2/n^2) = O(\omega^2)$$

which summing over all i implies that

$$\mathbb{E}\tilde{\mathbb{E}}P(x) = O(n\omega^2)$$

The above shows that $\{x\}$ is already very fishy as a pseudo-distribution, since (in the $\omega \gg n^{1/3}$ range) it gives $P(x)$ a value that is far too low to be consistent with being a distribution over ω-cliques. But we still haven't shown that it does in fact violate the constraints of being a valid pseudo-distribution. We now show this

Lemma 12 (Lemma 11,restated). *If $\omega \gg n^{1/3}$ then there exists a quadratic polynomial Q such that $\tilde{\mathbb{E}}Q^2 < 0$*

Proof. We let

$$Q(x) = (c(n/\omega)x_1 - n\langle r_1, x\rangle^2)^2$$

for some large enough constant c to be determined later. Now

$$\tilde{\mathbb{E}}Q^2 = \tfrac{c^2n^2}{\omega^2}\tilde{\mathbb{E}}x_1^2 + \tilde{\mathbb{E}}\langle r_1, x\rangle^4 - \tfrac{2cn}{\omega}\tilde{\mathbb{E}}\langle r_1, x\rangle^2 x_1^2 \qquad (2)$$

Let us compute each term of (2). First, clearly

$$\tilde{\mathbb{E}}x_1^2 = \tilde{\mathbb{E}}x_1 = \omega/n$$

and hence the first term equals c^2n/ω. We just computed the second term above as $O(\omega^2)$. To compute the third term, note that

$$\tilde{\mathbb{E}}\langle r_1, x\rangle^2 x_1^2 = \sum_{a,b\in[2..n]} r_1(a)r_1(b)\tilde{\mathbb{E}}x_a x_b x_1^2$$

which simply counts the number of triangles in the graph of the form $\{1, a, b\}$ (which is $\Omega(n^2)$) multiplied by $\Omega(\omega^3/n^3)$. (Indeed, note that if $\{1, a, b\}$ is a triangle then $r_1(a) = r_1(b) = +1$, and otherwise $\tilde{\mathbb{E}}x_a x_b x_1^2 = \tilde{\mathbb{E}}x_a x_b x_1 = 0$.) Thus the third term is $-\Omega(c(n/\omega)n^2(\omega^3/n^3) = -\Omega(c\omega^2)$. We see that if we want this expression to be negative then we need the third term to dominate the other two, and hence we need to satisfy $c \gg 1$ and $c\omega^2 \gg c^2n/\omega$, or $\omega^3 \gg cn$. Thus if $\omega \gg n^{1/3}$ then we can find a constant c that would make $\tilde{\mathbb{E}}Q^2 < 0$. □

3 Knapsack lower bound

One very nice SOS lower bound we did not show is the following theorem of Grigoriev

Theorem 13 (Grigoriev)**.** *For every n there is degree $\Omega(n)$ pseudo-distribution $\{x\}$ satisfying the constraints $\{x_i^2 = x_i\}$ and $\{\sum x_i = n/2\}$.*

Note that if n is odd, then there cannot be an actual distribution satisfying these constraints. The arxiv paper of Meka and Wigderson, despite having a fatal flaw, is still very much worth reading, and in particular is a good source for understanding the proof of this result.