Higher-degree integrality gaps: from computational hardness to limitations of sum-of-squares

We've seen integrality gaps for the degree 2 sum-of-squares algorithm, but so far have not seen any limitations for, say, degree 10 or even degree $\sqrt{n}$ sum-of-squares. However, since we believe certain NP problems (e.g., SAT) require super-polynomial, and in fact even *exponential* time (this is known as the *exponential time hypothesis*), we certainly think there should be polynomials $f : \{0,1\}^n \to \mathbb{R}$ such that there is a gap between $\min_{x\in\{0,1\}^n} f(x)$ and the minimum of $\tilde{\mathbb{E}}_\mu f$ over all degree $o(n)$ pseudo-distributions.

It turns out that the computational hardness results can be used as a guide to obtain sum of squares integrality gaps.[1] In particular, a canonical result by Håstad shows strong hardness for the Max-3XOR problem: given a collection of linear equations modulo 2 in n variables, with 3 variables in every equation (i.e. equations of the form $x_i + x_j + x_k = a_{i,j,k} \bmod 2$), find x that satisfies the largest possible fraction (called the *(optimal) value* of the instance) of equations. This result can be stated as follows:

[1] More surprisingly, Raghavendra [2008] showed that one can also do this in the other direction—use integrality gaps to obtain certain types of hardness results. We will talk about this result later.

1. Theorem (Håstad's 3XOR hardness). *For arbitrarily small constants $\varepsilon, \delta > 0$, it is NP-hard to distinguish whether a given instance of Max-3XOR has value at least $(1-\varepsilon)$ or at most $\frac{1}{2}+\delta$.*

As in the case of Max cut that we saw before, it is easy to express Max 3XOR as a polynomial optimization problem. A constraint $x_i + x_j + x_k = a_{i,j,k}$ is satisfied if and only if the polynomial $(1-2x_i)(1-2x_j)(1-2x_k)(1-2a_{i,j,k})$ (whose value on a binary input is either $+1$ or -1) is identical to $+1$ over $x \in \{0,1\}^n$. We can thus express fraction of constraints of ψ (seen as a collection of triples $\{i,j,k\}$ with an associated label $a_{i,j,k}$) satisfied by any $x \in \{0,1\}^n$ to be the cubic polynomial

$$f_\psi(x) = \tfrac{1}{2} + \tfrac{1}{2|\psi|} \sum_{\{i,j,k\}\in\psi} (1-2a_{i,j,k})(1-2x_i)(1-2x_j)(1-2x_k)\,. \quad (1)$$

Thus, the task of computing the value of ψ is equivalent to the task of maximizing f_ψ over all $x \in \{0,1\}^n$.

Observe that for any Max-3XOR instance , one of the all-1s or the all-0s assignment always satisfies at least $\frac{1}{2}$ of the equations. On the other hand, one can always efficiently check if a given system of linear equations is *exactly* satisfiable by using Gaussian Elimination. Thus Håstad's result shows that beating the trivial algorithms above is hard in the worst-case. As has become common since, Håstad's

proof of this result shows a reduction from a problem known as *Label Cover* to Max-3XOR. This reduction has only a linear blow-up - a Label Cover instance on n variables is reduced to a Max-3XOR instance on $O(n)$ variables. Moshkovitz and Raz [2008] showed that there is a reduction from 3SAT to Label Cover with only a quasi-linear blow-up. Gluing the above two reduction thus gives a quasi-linear reduction from 3SAT to Max-3XOR. Thus, if we make the reasonable assumption that 3SAT doesn't have a $2^{o(n)}$-time algorithm (i.e. the *exponential time hypothesis*) then, the sum-of-squares algorithm of degree $n^{0.99}$ (say) should not be able to distinguish between $(1-\varepsilon)$ and $\frac{1}{2}+\delta$ satisfiable instances of Max-3XOR.

Working in the sum-of-squares framework allows us the benefit of verifying this prediction *unconditionally* as the next theorem shows.[2]

[2] The "sum-of-squares hardness" of Max 3XOR was proven by Grigoriev [2001b] who phrased it as a lower bound in the *Positivstellensatz* proof system. It was later rediscovered by Schoenebeck [2008] who also observed that it immediately implies a similar lower bound for Max 3SAT as we will see later on in this section.

2. Theorem (Grigoriev's 3XOR sos hardness). *For every constant $\varepsilon > 0$, and large enough n, there is an instance ψ of Max-3XOR over n variables such that:*

1. *Every assignment $x \in \{0,1\}^n$ satisfies at most $\frac{1}{2}+\varepsilon$ fraction of the equations in ψ.*
2. *There exists a pseudodistribution of degree $\Omega(n)$ that is consistent with the constraints $\{x_i^2 - x_i = 0\}$ for every $i \in [n]$ and $\{(1-2x_i)(1-2x_j)(1-2x_k) = 1-2a_{i,j,k}\}$ for every constraint $x_i + x_j + x_k = a_{i,j,k}$ mod 2 in ψ.*

Notice that this result is stronger than what is predicted by the NP-hardness result of Håstad and asserts that sos cannot distinguish between a perfectly satisfiable instance and an instance in which only $\frac{1}{2}+\varepsilon$ fraction of the equations are satisfiable. Indeed, an analogous NP hardness result cannot hold (unless $P = NP$) since finding out whether an instance of linear equation modulo 2 is perfectly satisfiable can of course be done in polynomial time using Gaussian Elimination. This is rather disappointing - how can an allegedly strong algorithm like sos not simulate a simple efficient procedure such as Gaussian Elimination? One possible answer is that as a continuous relaxation, unlike Gaussian elimination, the sos algorithm cannot really distinguish between a perfectly satisfiable and one that is $1-o(1)$ satisfiable, and hence cannot solve the 1 vs $1/2+\varepsilon$ problem any better than the $1-\varepsilon$ vs $1/2+\varepsilon$ variant.

One corrollary of this theorem (or, more accurately, its proof) is that there is no generalization of the quadratic sampling lemma we saw to matching degree 3 moments, even if we significantly strengthen our assumption on the pseudo-distribution to an $\Omega(n)$ degree bound. We will work this out in Exercise 11 below.

Proving Grigoriev's theorem

In order to prove Theorem 2, we need to provide an instance ψ of 3XOR on n variables satisfying:

- **Soudnness:** No assignment $x \in \{0,1\}^n$ satisfies more than $1/2+\varepsilon$ fraction of ψ's constraints.
- **Completeness:** There exists a degree $\Omega(n)$ pseudo-distribution μ such that $\tilde{\mathbb{E}}_\mu f_\psi = 1$.

We will construct ψ by simply choosing it at random: for $C = C(\varepsilon)$ (that does not depend on n), we choose a random set of triplets $\{i,j,k\}$ by including every triplet in ψ independently with probability C/n^2. For each $\{i,j,k\}$ in ψ, we choose $a_{i,j,k}$ uniformly at random from $\{0,1\}$. It is useful to to describe ψ as (G,b) where G is the bipartite graph with m vertices on the left (corresponding to the randomly chosen triplets) and n vertices on the right (corresponding to the variables) and b is a vector in $\{0,1\}^m$. Every left vertex of G has degree three and if the ℓ^{th} triple in ψ is $\{i,j,k\}$ then the ℓ^{th} left vertex in G will be connected to i,j,and k and we set $b_\ell = a_{i,j,k}$.

Showing that the fraction of constraints satisfied by any assignment in ψ is close to $\frac{1}{2}$ is easy and only needs that the "right hand sides" of the equations, $a \in \{0,1\}^m$ are chosen at random.

3. Lemma (Soundness). *For any bipartite graph G with $m > 9n/\varepsilon^2$ vertices on the left, n on the right and all left-degrees 3, with probability at least $1-2^{-n}$ over the choice of $b \in \{0,1\}^m$, every assignment $x \in \{0,1\}^n$ satisfies at most $\frac{1}{2}+\varepsilon$ fraction of the constraints in $\psi = (G,b)$.*

Proof. Fix any $x \in \{0,1\}^n$ and $\ell \in [m]$ let Y_ℓ be the random variable (over the choice of $b \in \{0,1\}^m$ that equals 1 if $x_i + x_j + x_k = b_\ell \pmod 2$ and 0 otherwise, where $\{i,j,k\}$ are the neighbors of ℓ. The fraction of satisfied constraints is $\frac{1}{m}\sum_{\ell=1}^m Y_\ell$.

The random variables $Y_1,\ldots,Y_m$ are i.i.d Bernoulli variables each with expectation $1/2$ and so by the Chernoff bound

$$\mathop{\mathbb{P}}_b[\sum_{\ell=1}^m Y_i > (1/2+\varepsilon)m] < 2^{2^{-\varepsilon^2 m/3}}. \tag{2}$$

Thus by a union bound, the probability that there is an x that satisfies more than $\frac{1}{2}+\varepsilon$ fraction of constraints of ψ can thus be upper bounded by $2^{n+1}2^{-\varepsilon^2 m/3}$, which, for $m > 9n/\varepsilon^2$ is at most 2^{-n} as desired. □

While this proof is very simple, note that it does use the dreaded Chernoff + union bound combination (a.k.a the probabilistic method) and so we are outside the realm of the "Marley Corollary" and have no guarantee that we would be able to embed it in a low degree sum of squares argument. Indeed, Grigoriev's theorem exactly shows that this can't be done.

Constructing the pseudodistribution

The construction of a pseudodistribution μ that satisfies all the constraints of ψ is more interesting. Our goal is to build a pseudodistribution that "pretends" as if the instance ψ constructed above is perfectly satisfiable. Denote by $\chi_i(x)$ the function $1-2x_i$ and by $\chi_S(x)$ the function $\prod_{i\in S}\chi_i(x)$. Note that the functions $\{\chi_S\}_{|S|\leq d}$ form a basis for the subspsace of all polynomials of degree at most d, and hence to specify the pseudo-expectation operator corresponding to a degree d pseudo-distribution μ it is sufficient to specify $\tilde{\mathbb{E}}_\mu \chi_S$ for every $|S| \leq d$.

We start by observing that certain "hard" constraints are forced by the fact that our pseudo-distribution pretends to satisfy all the constraints of ψ. For example, if $x_i + x_j + x_k = a_{i,j,k} (\mod 2)$ is a constraint in ψ, then we have no choice but to set $\tilde{\mathbb{E}}_\mu \chi_{i,j,k} = 2a_{i,j,k} - 1$. As another example, suppose that ψ contains the constraints $x_1 + x_2 + x_3 = 1 (\mod 2)$ and $x_3 + x_4 + x_5 = 0 (\mod 2)$ then we can sume these two equations and conclude that $x_1 + x_2 + x_4 + x_5 = 1 (\mod 2)$ and hence we should expect our pseudo-distribution to satisfy $\tilde{\mathbb{E}}_\mu \chi_{1,2,4,5} = -1$.

To complete the definition of the pseudo-expectation, we need to to define $\tilde{\mathbb{E}}_\mu \chi_S$ for every $|S| \leq \varepsilon n$. This raises two questions:

- How should we set $\tilde{\mathbb{E}}_\mu \chi_S$ for any S such that $\tilde{\mathbb{E}}_\mu \chi_S$ is not set via the above *hard* constraints?
- Will we have enough "freedom" to force such constraints on $\tilde{\mathbb{E}}$ or would we run into contradictions, i.e., constraints as above that force us to give two differing values to $\tilde{\mathbb{E}}_\mu \chi_S$ for some S?

For the first question, we will use a general principle - our own home-brewed analog of Einstein's maxim.

Pseudodistributions should be made as random as possible but no randomer.

How does this help us answer the first question above? We will

pretend that our pseudodistribution looks like the uniform distribution over the hypercube subject to the hard parity constraints obtained above. This amounts to setting $\tilde{\mathbb{E}}_\mu \chi_S$ to 0 for every S for which the value is not already decided. Another way to justify this approach is through Jaynes' maximal entropy principle.[3] This states that if we are trying to guess some unknown $x \in \{0,1\}^n$, we should assume that x comes from the distribution of maximum entropy consistent with the observations we have made.[4] We will see that this "Bayesian view" will serve as a useful guide in problems (such as the planted clique) where we have to deal with a combination of "hard" and "soft" constraints, and where a simplistic strategy that pretends that random variables are either identical or independent will lead to a pseudodistribution that is "too random" and land us into trouble.

[3] See also this course of James Lee on entropy optimality as well as his blog posts on this topic.

[4] This is actually the same as assuming that our prior distribution was the uniform distribution, and then updating to the distribution minimizing the from it that satisfies our observations.

The second question is not at all trivial. After all, the instance ψ is in fact unsatisfiable which in particular means that by adding together linear equations we can in fact achieve a contradiction. Nevertheless, we will see that we have enough freedom whenever the bipartite graph G associated with the instance ψ satisfies a sufficiently strong notion of *expansion*.

4. Definition (expansion). A bipartite graph G with all left degrees 3 is said to be (t,β)-expanding if every subset T of vertices on the left of size $|T| \leq t$ satisfies $|\Gamma(T)| \geq \beta|T|$, where $\Gamma(T)$ denotes the set of neighbors of vertices in T in G.

As one might expect, random bipartite graphs enjoy excellent expansion:

5. Lemma (expansion of random graphs). *For G constructed above with $m = Cn$ vertices for some constant C and $\delta > 0$, there exist constant $\eta > 0$ (depending on C) such that with probability at least 0.9, the graph G is $(\eta n, 2-\delta)$-expanding.*

Proof. The proof is a simple application of the Union bound. Let $\beta = 2 - \delta$. We estimate from above, the probability that a collection of $s = \eta n$ constraints cover at most βs variables by:

$$\binom{n}{\beta s}\binom{\binom{\beta s}{3}}{s}(C/n^2)^s, \tag{3}$$

Using standard upper estimates on the binomial coefficients above, this is at most:

$$(\frac{ne}{\beta s})^{\beta s}(\frac{e\beta^3 s^2}{6})^s\frac{C^s}{n^{2s}}. \tag{4}$$

The expression above can be simplified to be at most:

$$\left((\frac{s}{n})^\delta 4Ce^2\right)^s, \tag{5}$$

which is at most 0.1 for $\eta < (1/4Ce^2)^{10}$. □

6. Exercise. Prove that for every $\eta > 0$, no 3-left-regular bipartite graph with n right vertices and $\Omega(n)$ left vertices is an $(\eta n, 2)$ expander.

We note that it is known how to construct *deterministically* left-degree $d = O(1)$ bipartite graphs expansion $(1-\epsilon)d > d/2$ (Capalbo et al. [2002]) (see also chapter 10 in (Hoory et al. [2006]) and chapter 6 in (Vadhan [2012])). Such graphs can be used to construct integrality gaps for d-XOR instances, though we still don't know how to avoid the step of choosing the "right hand side" at random (i.e., Lemma 3).

We now go on to the most interesting step—defining the pseudodistribution following our intuitive discussion based on Einstein's maxim above. We will work with $\psi = (G, b)$ such that G is $(\eta n, 1.7)$-expanding[5] and construct a degree $d = \eta n/10$-degree pseudodistribution. This construction doesn't depend on the "right hand sides" of the equations and only needs expansion. To emphasize this, we record the construction as the following lemma.

[5] Any expansion greater than 1.5 will be good enough for us

7. Lemma (pseudodistribution-from-expansion). *Let $\psi(G, b)$ be an $(\eta n, 1.7)$-expanding instance of Max 3-XOR. There exists a pseudodistribution of degree $d \geq \eta n/10$ that satisfies the constraints $x_i^2 = x_i$ and $\chi_{S_\ell} = b_\ell$ for every ℓ and S_ℓ being the set of neighbors of the ℓth left vertex in G and b_ℓ the corresponding RHS of the XOR constraint.*

We will build the pseudodistribution by describing the associated pseudoexpectation $\tilde{\mathbb{E}}$. As mentioned above, it is enough to specify the values of $\tilde{\mathbb{E}}_\mu \chi_S$ for $S \subseteq [n], |S| \leq d$. Let $d = \eta n/10$, $\Gamma(\ell)$ be the 3 neighbors of any left vertex ℓ in G and $S \oplus T$ denote the symmetric difference of sets $S, T \subseteq [n]$. We will use the following algorithm to set values of $\tilde{\mathbb{E}}_\mu \chi_S$ for $|S| \leq d$.

- Set $\tilde{\mathbb{E}}\, 1 = 1$.
- For every $\ell \in [m]$, let $\tilde{\mathbb{E}}[\chi_{\Gamma(\ell)}(x)] = 1 - 2b_\ell$.
- Repeat the following until impossible: Choose S, T such that $\tilde{\mathbb{E}}\, \chi_S$ and $\tilde{\mathbb{E}}\, \chi_T$ have already been set and $|S \Delta T| \leq d$. If $\tilde{\mathbb{E}}_\mu \chi_{S\Delta T}$ is already defined and doesn't equal $(\tilde{\mathbb{E}}\, \chi_S)\,(\tilde{\mathbb{E}}\, \chi_T)$, halt and declare failure. Otherwise, set $\tilde{\mathbb{E}}\, \chi_{S\Delta T} = (\tilde{\mathbb{E}}\, \chi_S)\,(\tilde{\mathbb{E}}\, \chi_T)$.
- For every S, $|S| \leq d$ such that $\tilde{\mathbb{E}}\, \chi_S$ was not set in the first two steps, set $\tilde{\mathbb{E}}\, \chi_S = 0$.

We will show that $\tilde{\mathbb{E}}$ defined above is a valid pseudoexpectation. The next lemma shows that the defining algorithm above never returns failure if G has good enough expansion.

8. Definition (Degree-d derivation). Let $G = ([m], [n], E)$ be a bipartite graph as above. For every set S with $|S| \leq d$, we define a *degree d derivation* of S to be a sequence $T_0, \ldots, T_t$ of subsets of $[m]$ such that

- $T_0 = \emptyset$,
- $\Delta_{\ell \in T} \Gamma(\ell) = S$ where as above for every left vertex $\ell \in [m]$, $\Gamma(\ell) \subseteq [n]$ denotes the set of neighbors of ℓ,
- for every $i \in \{1, \ldots, t\}$, $|\Delta_{\ell \in T_i} \Gamma(\ell)| \leq d$,
- for every $i \in \{1, \ldots, t\}$, there exist $j, k \in \{0, \ldots, i-1\}$ such that $T_i = T_j \Delta T_k$.

For every $S \subseteq [n]$ and $\sigma \in \{0,1\}$, we say that the equation $\sum_{i \in S} x_i = \sigma (\mod 2)$ is *d-derivable* from the instance (G, a) if there exists a degree-d derivation of S such that $\sum_{\ell \in T_t} a_\ell = \sigma (\mod 2)$.

Note that deriving an equation in degree d means that there is a way to add up the basic equations $\sum_{i \in \Gamma(\ell)} x_i = a_i (\mod 2)$ to obtain the resulting equation in a way that no intermediate equation ever involves more than d variables. We say that an instance (G, a) contains a *degree d contradiction* if there exists some $S \subseteq [n]$ with $|S| \leq d$ such that we can d-derive both the equation $\sum_{i \in S} x_i = 0 (\mod 2)$ and the equation $\sum_{i \in S} x_i = 1 (\mod 2)$ from (G, a).[6]

[6] Note that since the instances we are interested in are unsatisfiable, it is certainly possible to derive the equations, say, $x_1 = 1$ and $x_1 = 0$, from the original constraints, but the question is whether it is possible to do so while never requiring an intermediate equation involving more than $d = \varepsilon n$ variables.

9. Lemma (Pseudo expectation is well defined). *Suppose G is $(100d, 1.7)$-expanding. Then for every $a \in \{0,1\}^m$, (G, a) does not contain a degree-d contradiction. Consequently, the algorithm above never halts and returns failure.*

Before proving Lemma 9, we first show why it implies that $\tilde{\mathbb{E}}$ defined by the algorithm is indeed a valid pseudoexpectation and hence complete the proof of Theorem 2.

Proof of Theorem 2. Let (G, a) be the instance corresponding to ψ and $\tilde{\mathbb{E}} = \tilde{\mathbb{E}}_\mu$ be the pseudo-expectation operator we defined above. By construction, $\tilde{\mathbb{E}}$ is a linear operator on degree $\leq d$ polynomials satisfying $\tilde{\mathbb{E}}\, 1 = 1$. Also, since every $\ell \in [m]$, our procedure defines $\tilde{\mathbb{E}}\, \chi_{\Gamma(\ell)} = 1 - 2a_i$, it is an easy calculation to show that via linearity of expectation the expected fraction of satisfied constraint $\tilde{\mathbb{E}}\, f_\psi$ equals 1. Therefore all that's left is to show that the positivity condition, which as usual is the most challenging. Namely, we need to show that $\tilde{\mathbb{E}}\, p^2 \geq 0$ for any polynomial p of degree at most $d/2$.

Define an equivalence relation on $\{S \subseteq [n] \mid |S| \leq d/2\}$ where $S \sim T$ if $\tilde{\mathbb{E}}[\chi_{S \Delta T}] \neq 0$ (i.e, $\tilde{\mathbb{E}}[\chi_{S \Delta T}]$ is defined in the process above). It

is easy to verify that this is indeed an equivalence relation (and thus reflexive, i.e. $S \sim S$ and transitive: if $S \sim T$, $T \sim U$ then $S \sim U$) and thus partitions the set $\{S \subseteq [n] \mid |S| \leq d/2\}$ into equivalence classes $C_1, C_2, \ldots, C_N$, and choose for every C_i we choose $S_i \in C_i$ to be some representative member.

Next, we claim that if $S, T \in C_i$, then $\tilde{\mathbb{E}}\, \chi_S \chi_T = \left(\tilde{\mathbb{E}}\, \chi_{S \Delta S_i}\right)\left(\tilde{\mathbb{E}}\, \chi_{T \Delta S_i}\right)$. Indeed, since $S, T \in C_i$, both $\tilde{\mathbb{E}}\, \chi_{S \Delta S_i}$ and $\tilde{\mathbb{E}}\, \chi_{T \Delta S_i}$ are nonzero, which means that under our definition of the pseudo-distribution

$$\tilde{\mathbb{E}}\, \chi_{S \Delta T} = \tilde{\mathbb{E}}\, \chi_{S \Delta S_i} \chi_{T \Delta S_i} = \left(\tilde{\mathbb{E}}\, \chi_{S \Delta S_i}\right)\left(\tilde{\mathbb{E}}\, \chi_{T \Delta S_i}\right) \ . \tag{6}$$

Hence, if we let p be a polynomial of degree at most d, we can write $p = p_1 + \ldots + p_N$ where p_i consists of only monomials in the equivalence class C_i. By the way we defined our equivalence relation, is not hard to see that $\tilde{\mathbb{E}}\, p_i p_j = 0$ if $i \neq j$. Hence

$$\tilde{\mathbb{E}}\, p^2 = \tilde{\mathbb{E}}(\sum p_i)^2 = \sum_{i,j} \tilde{\mathbb{E}}\, p_i p_j = \sum_i \tilde{\mathbb{E}}\, p_i^2 \ . \tag{7}$$

But every polynomial p_i can be written as $\sum_{S \in C_i} p_S \chi_S$, and then one can see that

$$\tilde{\mathbb{E}}\, p_i^2 = \tilde{\mathbb{E}}(\sum_{S \in C_i} p_S \chi_S)^2 = \sum_{S,T \in C_i} p_S p_T\, \tilde{\mathbb{E}}\, \chi_S \chi_T \ . \tag{8}$$

But by what we claimed above, the RHS of Eq. (8) equals

$$\sum_{S,T \in C_i} p_S p_T \left(\tilde{\mathbb{E}}\, \chi_{S \Delta S_i}\right)\left(\tilde{\mathbb{E}}\, \chi_{T \Delta S_i}\right) = \left(\sum_S p_S\, \tilde{\mathbb{E}}\, \chi_{S \Delta S_i}\right)^2 \geq 0 \tag{9}$$

□

We now prove Lemma 9.

Proof of Lemma 9. Suppose, towards a contradiction, that there exists some S such that we can derive both $\sum_{i \in S} x_i = 0 \pmod 2$ and $\sum_{i \in S} x_i = 1 \pmod 2$ using degree d derivations. Then by combining these two together, we can derive using a degree $2d$ derivation that $\sum_{i \in \emptyset} x_i = \sum_{i \in (S \Delta S)} x_i = 1 \pmod 2$. Let $T_1, \ldots, T_t$ be this derivation, satisfying $\Delta_{\ell \in T_t} \Gamma(\ell) = \emptyset$. This means that that every neighbor of T_t has an even number of (and in particular at least two) edges from T_t to it. Since there are $3|T_t|$ edges exiting T_t, we get that $|\Gamma(T_t)| \leq 1.5|T_t|$ and hence by the expansion property of G this means that $|T_t| \geq 100d$.

Thus to get a contradiction, it is enough to prove that for every i, $|T_i| \leq 10d$. Indeed, suppose otherwise and let i be the smallest i such

that $|T_i| > 10d$, then $T_i = T_j \Delta T_k$ for some $k, j < i$ and in particular $|T_j|, |T_k| \leq 10d$. But this means that $|T_i| \leq 20d$ which contradicts the fact that, as part of a $2d$-derivation, it must satisfy $|\Delta_{\ell \in T_i} \Gamma(\ell)| \leq 2d$. Indeed, by the expansion of the graph, the $3|T_i|$ edges leaving T_i touch at least $1.7|T_i|$ vertices, which means that there is a set $S' \subseteq [n]$ of at least $0.4|T_i| \geq 4d$ right vertices that have only a single neighbor in T_i.[7] But such a set S' is contained in $\Delta_{\ell \in T_i} \Gamma(\ell)$ (can you see why?), contradicting the fact that the latter set has at most $2d$ elements □

With this we have completed the proof of Theorem 2.

[7] Indeed, if out of T_i's more than $1.7|T_i|$ neighbors there there are more than $1.3|T_i|$ vertices with two or more edges going to T_i then we would get that T_i has more than $3|T_i|$ edges leaving it, which is of course a contradiction.

Lower Bounds for Refuting Random 3XOR

Our proof of Grigoriev's theorem in fact ends up proving something stronger - a lower bound for refuting random 3XOR instances. In the *refutation* problem, we are given an instance defined by randomly generated 3-variable linear equations modulo 2 (i.e. both the set of variables and the "right hand sides" of each equation are chosen uniformly at random) and are asked to *certify* that the resulting instance is not satisfiable (*weak* refutation) or better yet, to certify that no more than some fixed constant fraction of the equations can be satisfied simultaneously (*strong* refutation) with high probability. The proof of Grigoriev's theorem shows that random 3XOR instances with $\Theta(n)$ equations cannot even be weakly refuted by the SoS algorithm. The next exercise shows a simple generalization of this analysis to get a more general lower bound for the refutation problem.

10. Exercise (Expansion of Random Bipartite Graphs with super-linear vertices). Fix $\epsilon > 0$. Let G be a bipartite graph with the number of left vertices L is $n^{1.5-\epsilon}$ and the right vertices R of size n with each left-degree being 3. Show that G satisfies $(n^{-1+\Theta(\epsilon)}, 0.1)$-expansion with probability at least $1 - \frac{1}{n}$.

Conclude that with high probability for a random 3 XOR instance with $n^{1.5-\epsilon}$ equations, there's a degree d (for $d = n^{\Theta(\epsilon)}$) pseudodistribution that is consistent with all the equations and $\{x_i^2 = x_i\}$ for every i.

Ruling out a "cubic sampling lemma"

As mentioned before, the *quadratic sampling lemma*, giving a distribution (albeit over $\mathbb{R}^n$ instead of $\{0,1\}^n$) matching specified degree two moments, is one of the most useful tools in rounding sos-based

algorithms, as well as in many other areas, including finance and forecasting, where one needs to generate probabilistic models matching some given parameters. It turns out that the proof of Theorem 2 establishes that we cannot generalize this to higher moments:

11. Exercise (No cubic sampling lemma). Prove that there exists some $\delta > 0$ such that for every n, there is a degree δn pseudo-distribution μ over $\{0,1\}^n$ such that there does not exist an actual distribution ρ over $\mathbb{R}^n$ satisfying $\left|\mathbb{E}_\rho x_i x_j x_k - \tilde{\mathbb{E}}_\mu x_i x_j x_k\right| < 0.001$ for all i, j, k. You can do so by following the steps below:

1. Prove that if there exists a distribution ρ as above then it satisfies the following condition:

$$\left|\mathop{\mathbb{E}}_{\rho} \chi_S - \mathop{\tilde{\mathbb{E}}}_{\mu} \chi_S\right| < 0.1 \text{ for all } |S| \leq 3. \tag{10}$$

2. Prove that if (G, a) and μ are a 3XOR instance and a pseudo-distribution as in Grigoriev's proof, then if there is an actual distribution ρ satisfying Eq. (10) then there exists a vector $x \in \mathbb{R}^n$ satisfying:

$$\|x\| \leq 10 \text{ and } \frac{1}{m}\sum_{\ell} \chi_{\Gamma(\ell)}(x)(1 - 2a_\ell) \geq 0.01\,. \tag{11}$$

3. Prove that for every set $N \subseteq \mathbb{R}^n$ such that $\|x\| \leq 10$ for every $x \in N$, if we select at random a 3XOR instance with $C > 100 \log |N|/\epsilon^2$ constraints then with probability at least 0.9 there will not exist an $x \in N$ satisfying Eq. (11).

4. Prove that if there is some C such that if we select at random a 3XOR instance with Cn constraints then with probability at least 0.9 there will not exist $x \in \mathbb{R}^n$ satisfying Eq. (11) (perhaps with 0.01 replaced by 0.02).[8]

5. Use this to complete the proof of the exercise.

[8] **Hint:** Show that there exists a set N of size at most $2^{O(n)}$ such that for every $x \in \mathbb{R}^n$ with $\|x\| \leq 10$ there exists $x' \in N$ with $\|x - x'\| < 0.001$. Then show that if there is no $x' \in N$ satisfying Eq. (11) then there is no $x' \in \mathbb{R}^n$ satisfying the relaxed version of Eq. (11).

The probabilistic method and the "Marley Corollary"

We have previously humorously referred to the "Marley Corollary" as roughly saying that "if you proved a statement X without using the probabilistic method, then you should be able to prove it with a low degree sos proof". This is not meant as a "law of nature" that always needs to hold, but rather as a heuristic rule of thumb that is helpful when you look at some particular statement that you know (or have strong reasons to believe) that is true, and want to

understand how likely is it to be provable within the low degree sos framework. Nevertheless, now that we have seen Theorem 2, this might be a good point to pause and reflect on how well this heuristic matches reality.

A priori Theorem 2 seems to directly contradict Marley's Corollary. If we generate some random 3XOR instance (G, a), then with very high probability it will be unsatisfiable, and in fact we can easily certify this fact via Gaussian elimination and hence prove it without any appeal to the probabilistic method. Yet, by Theorem 2, there will be no $o(n)$ degree sos proof for this fact.

In fact, using similar ideas to those he used to prove Theorem 2, Grigoriev [2001a] showed that for any odd n there is no $o(n)$ degree sos proof for the following simple statement $\min_{x\in\{0,1\}^n}(\sum_{i=1}^n x_i - \frac{n}{2})^2 \geq 1/4$.

That is, sos cannot even succinctly certify the fact that you can't divide equally divide eleven toys between two children.

The main reason we appeal to the probabilistic method for sos hardness is to generate *robust* integrality gaps. For example, do not at the moment have a deterministic construction of a 3XOR instance such that every assignment satisfies at most 0.9 fraction of the equations, but there is no sos proof that one can't satisfy 0.99 fraction. In this sense, the "Marley Corollary" should be rephrased as "if you proved a statement X without using the probabilistic method, then you should be able to prove a slightly quantitatively weaker statement X' with a low degree sos proof".[9] Yet, even this corollary should be taken with a grain of salt. Just like we have deterministic constructions of lossless expanders, we may eventually get deterministic constructions of such robust integrality gaps.[10] Yet the fact that it is so hard to come up with such examples is a hopeful sign and justifies the methodology of pretending that pseudo-distributions are actual distributions in algorithm design.

Another critique of the "Marley Corollary" is that it may be too *weak*. After all, when we use the probabilistic method to show that with high probability an object I from a distribution $\mathcal{D}$ has some property P it does not give us a mathematical proof that a particular I sampled from $\mathcal{D}$ has P. For example, it is reasonable to conjecture that there does not exist *any polynomial sized proof*, no matter what in proof system, for a random 3XOR instance (G, a) that is as sampled in the proof of Theorem 2 (see (Feige [2002])). This can be thought of as an average case version of the conjecture that $NP \neq coNP$. So, one could claim that the corollary should be phrased as "if you

[9] For example, if you use Gaussian elimination to prove that for some instance of m equations, every assignment satisfies at most $m-1$ of them, then you can of course via sos the trivial statement that every assignment satisfies at most m equations.

[10] Indeed perhaps we already do: one nice open question is to find out whether for the lossless ledt-degree-d expanders of (Capalbo et al. [2002]) there exists a degree $\Omega(n)$ distribution that pretends to be over sets with expansion at most $0.51d$. This would give a deterministic construction of a robust integrality gap.

have any way to certify a statement X then you should be able to certify a weaker statement X'''.[11] Nevertheless, since the probabilistic method is the most typical way one generates statements that we believe are true but have no certificate for, the current phrasing of the corollary seems like a useful rule of thumb. Note also that while the use of the probabilistic method is a sign one should be careful, it does not automatically rule out the existence of a low degree sos proof. Often the same ideas that can be used to derandomize probabilistic constructions, such as replacing the uniform distribution by a distribution with limited independence, can be used to find short sos proofs to statements that are originally proven (or, more accurately, given evidence for) by the probabilistic method.

[11] As remarked above, we would not expect this to be universally true since it would mean that sos is an *instance optimal* proof system. Indeed, it is quite possible that one can deterministically generate robust integrality gap instances using some "un-natural" tools from pseudo-randomness, such as those used in the construction of (Capalbo et al. [2002]) or constructions of Ramsey graphs, multiple source extractors and related objects (e.g., see (Chattopadhyay and Zuckerman [2016]) and the references therein).

Reductions within the sum-of-squares framework

We just saw how NP-hardness inspires strong sos hardness for Max 3XOR. One extremely successful way of obtaining new NP-hardness results is by using reductions. One could hope to import this technology to the sos framework to obtain new hardness results. Luckily, this can indeed be done in a fairly generic way, as we show in this section.

We begin with an immediate corollary of Grigoriev's Theorem that shows sos hardness for the Max 3SAT problem, where the constraints have the form $y_i \vee y_j \vee y_k$ where each y_i (known as a *literal*) is either some variable x_i or its negation.

12. Theorem (sos hardness for Max 3SAT). *For every constant $\varepsilon > 0$, and large enough n, there is an instance ψ of Max-3SAT over n variables such that: 1. Every assignment $x \in \{0,1\}^n$ satisfies at most $\frac{7}{8} + \varepsilon$ fraction of the constraints in ψ. 2. There exists a pseudodistribution of degree $\Omega(n)$ over $\{0,1\}^n$ that that satisfies in expectation all of the constraints of ψ.*

Proof. We construct the instance by generating a random bipartite graph G as in Grigoriev's theorem. For every left vertex ℓ with edges to $\{i,j,k\}$ on the right, we choose $a_{\ell,i}, a_{\ell,j}$ and $a_{\ell,k}$ uniformly at random and independently from $\{0,1\}^n$. The 3SAT instance ψ is then defined so that the ℓth constraint is given by $y_i \vee y_j \vee y_k = 1$ where $a_{\ell,i}$ (and similarly for j and k) decides whether $y_i = x_i$(when $a_{\ell,i} = 0$) or $\neg(x_i)$ (when $a_{\ell,i} = 1$). By essentially the same argument as for Lemma 3, we can argue that for any bipartite graph G as above with $m > 9n/\varepsilon^2$ left vertices and all left degrees 3, choosing $a_{\ell,i}$s uniformly at random ensures that with probability $1 - 2^{-n}$, every assignment $x \in \{0,1\}^n$ satisfies at most $\frac{7}{8} + \varepsilon$ fraction of the constraints in ψ.

To construct a pseudodistribution that satisfies all the constraints of ψ, we work with a 3XOR instance where corresponding to the ℓth triple (i,j,k) and the corresponding $a_{\ell,i}, a_{\ell,j}$ and $a_{\ell,k}$ we include the 3XOR constraint $x_i + x_j + x_k = a_{\ell,i} + a_{\ell,j} + a_{\ell,k} (\mod 2)$.

Observe that if this equation is satisfied then in particular the corresponding OR constraint is satisfied as well (this follows from the simple equation $0+0+0 = 0 (\mod 2)$). Moreover, thethe bipartite graph associated with the 3XOR instance is the same as that for the 3SAT instance- i.e., random - and thus enjoys expansion properties as in Lemma 5. Thus the same pseudo-distribution we used in the proof of Grigoriev's Theorem to works for the 3SAT instance. □

The sum-of-squares PCP Theorem

The starting point of most reductions that prove NP-hardness of approximation is the famous PCP Theorem (standing for *probabilistically checkable proofs*). This theorem gives an NP-hardness of approximation result for what is known as the *Max-P* or *Constraint Satisfaction* problem (CSP). Given $P\colon \{0,1\}^k \to \{0,1\}$ (which is known as a *predicate*), we define a Max-P instance ψ to be a collection of constraints of the form $P(y_{i_1}, y_{i_2}, \ldots, y_{i_k}) = 1$ for some literals $y_{i_1}, \ldots, y_{i_k}$. The goal is to find an $x \in \{0,1\}^n$ that satisfies as many of the constraints as possible.

An analog for the sos framework was first developed by Tulsiani who observed that the proof of Theorem 2 generalizes to Max-P problem for any *nice subspace predicate* P. We say that $V \subseteq GF(2)^k$ is a nice subspace if every non-zero $u \in V^\perp$ has Hamming weight at least 3. We say that P is a nice subspace predicate if there is a nice subspace V such that $P(x) = 1$ iff $x \in V$. By choosing V to be subspace of $GF(2)^k$ spanned by the codewords of the Hamming code, Tulsiani [2009] proved the following:

13. Theorem (sos hardness for Nice-Subspace CSP, aka sos PCP). *Let $k, \varepsilon > 0$ be given. There exist $\beta = O(2^k/\varepsilon^2)$, $c = \Omega(1/\beta^{25})$ such that there's an instance ψ of Max-P problem for a k-variate predicate P with at most $2k$ satisfying assignments on $n \gg 1/c$ variables and $m = \beta n$ constraints such that: * Every assignment $x \in \{0,1\}^n$ satisfies at most $\frac{2k}{2^k} + \varepsilon$ fraction of the constraints of ψ. * There exists a degree cn pseudodistribution $\{x\}$ that satisfies in expectation all of the constraints of ψ.*

We omit the proof of Theorem 13 here though it follows from the same argument as above. The key property used is that if $V \subseteq$

$GF(2)^k$ is a nice subspace then the uniform distribution over V is a *pairwise independent* distribution over $GF(2)^k$ (or, equivalently, $\{0,1\}^k$). We will see in a future lecture that this is a necessary and sufficient condition[12] to obtaining such an integrality gap result.

[12] The necessary part is true for random instances but not for worst case instances in general, right? I think there are non-pairwise-independent predicates for which we can construct degree 2 (and also higher degree, but this may not have been written down somewhere) gap instances.

Using Reductions: sos hardness of Max Independent Set

We next consider the case of the *Max Independent Set* problem as an illustrative example of how NP-hardness reductions can be composed with the above result for the Max-P problem to yield new sos hardness results.

Let us first discuss the broad outline of how such a hardness result would work. The following discussion, though specialized to the case of Max Independent Set here, is entirely generalizable to any situation where there is a known reduction from Max-P problem to the problem of interest.

The idea of the construction is to analyze the standard NP-hardness reduction from Max-P problem to the independent set problem. In obtaining NP-hardness for independent set using this methodology, we show two claims:

1. **Soundness:** If the starting instance of Max-P has a low value, then the Max Independent Set instance output by the reduction also has a low value.
2. **Completeness:** If the starting instance of Max-P has a high-value, then the Max Independent Set instance output by the reduction also has a high value.

Ultimately, the gap between the values of the indepedent set instances in the completeness and soundness claims decides the final inapproximability ratio obtained.

Suppose now, instead of starting from the NP-hardness, we start from the sos hardness of approximating the Max-P problem and obtain an Max Independent Set instance by applying the reduction to the sos-hard instance of Max-P. That is, we take an *integrality gap* instance I of Max-P, which in actuality has low value, but has a pseudo-distribution μ that pretends that it has high value, and use the NP-hardness reduction to map it to an instance I' of Max Independent Set. We want to show that I' is an integrality gap for Max Independenet Set. To do this, we again need to argue two claims:

1. **Soundness:** The resulting Max Independent Set instance I' output by the reduction has a low value. This actually is a direct corollary of the soundness property of the reduction: since I had low value as a Max-P instance, I' will have low value as a Max Independent Set instance.

2. **Completeness:** In the sos setting, it is the (usually trivial) *completeness* case requires some work. We must show that starting from a pseudodistribution that pretends that the Max-P instance is perfectly satisfiable, we can produce a pseudodistribution that satisfies the constraints of the Max Independent Set polynomial optimization program and pretends that the size of the Max Independent Set in it is higher than the actual value.

The next three exercises develop simple tools to analyze such reductions.

14. Exercise (Porting low-degree reductions). Let $\mathcal{R}$ be a Karp reduction from a polynomial optimization problem Q_1 to a polynomial optimization problem Q_2 over the Boolean hypercube. Suppose further that for any instance ψ of Q_1 and any x that is a feasible for ψ, there's a y that is feasible for $\mathcal{R}(\psi)$ and that for each j, $y_j = f_j(x)$ for a degree $\leq t$ polynomial f_j. Show that if there's a degree d pseudodistribution x feasible for the instance ψ of Q_1 then there's a degree d/t pseudodistribution over y feasible for instance $\mathcal{R}(\psi)$ of Q_2.

15. Exercise (Composing independent pseudodistributions). Suppose $\mathcal{D}_x$ and $\mathcal{D}_y$ are degree d_1 and d_2 pseudodistributions in variables $x \in [q]^{n_1}$ and $y \in [q]^{n_2}$ respectively. Let $\mathcal{D} = \mathcal{D}_x \times \mathcal{D}_y$ be a pseudodistribution over (x, y) defined by $\tilde{\mathbb{E}}_{\mathcal{D}}[f(x)g(y)] = \tilde{\mathbb{E}}_{\mathcal{D}_x}[f(x)]\,\tilde{\mathbb{E}}_{\mathcal{D}_y}[f(y)]$ and linearly extending to any $h(x, y)$ for polynomials f, g, h. Show that $\tilde{\mathbb{E}}_{\mathcal{D}}$ is a degree $\min\{d_1, d_2\}$-pseudodistribution.

16. Exercise (Projections of Pseudodistributions). Show that if $\mathcal{D}$ is a degree d pseudodistribution over $\{0,1\}^n$ then the distribution corresponding to restricting $x \in \{0,1\}^n$ to any subset of the n bits is also a degree d pseudodistribution.

We now formally discuss the sos-Hardness for Max Independent Set by Tulsiani [2009] .

17. Theorem (sos hardness for Independent Set). *Fix any positive integer r and k such that there's a nice subspace predicate with at most $2k$ satisfying assignments on $\{0,1\}^k$. For large enough n, there exist constant $c_1, c_2 > 0$ and family on graphs G on $N = 100nr(2k)^r$ vertices such that:*

- *Every independent set in G is of size at most $200nr$.*

- *There exists a degree $n/2^{\Theta(k)}$-degree pseudodistribution $\mathcal{D}$ consistent with the constraints $\{x_i^2 = x_i\}$ for every $i \in [n]$, $\{x_i x_j = 0\}$ for every $\{i,j\}$ that is an edge in G and $\tilde{\mathbb{E}}_{\mathcal{D}}[\sum_{i=1}^n x_i] = 2^{\Theta(r)} \left(\frac{k}{2^k}\right)^r$.*

Consequently, by choosing $r = \frac{\log(n)}{\log\log(n)}$ and $k = \Theta(\log(n))$, the integrality gap of degree $2^{\Theta(\sqrt{\log(N)\log\log(N)})}$ sos algorithm for Max Independent Set problem is at least $\frac{N}{2^{\Theta(\sqrt{\log(N)\log\log N})}}$.

The classical proof of inapproximability of the independent set problem involves constructing the well-known FGLSS graph. Let ψ be an instance on n variables of Max-P problem with m constraints where P is a k-variate nice-subspace predicate. Let G be the bipartite graph of the instance ψ. Generate a new graph H with vertex set given by (ℓ, α) for every left vertex ℓ of G and every possible satisfying assignment α in $\{0,1\}^k$ for the ℓth constraint of ψ. Add an edge between any two (ℓ_1, α_1) and (ℓ_2, α_2) of H if ℓ_1 and ℓ_2 share at least one neighbor in G and α_1 and α_2 differ in the assignment to the shared neighbor (in other words, if α_1, α_2 are conflicting partial assignments for the instance ψ). H is called as the FGLSS graph of the Max-P instance ψ. The next exercise relates the size of independent set in H to the maximum number of satisfiable constraints in ψ.

18. Exercise (Polynomial Feasibility Formulation of Independent Set). Let H be a graph on vertex set $[n]$. Consider the following polynomial feasibility formulation in variables $x \in \mathbb{R}^n$: $\{x_i^2 - x_i = 0\}$ for each $i \in [n]$, $\{\sum_{i \le n} x_i = q\}$ and $\{x_i x_j = 0\}$ for every $\{i,j\}$ that is an edge in H. Verify that x is feasible for the program if and only if it is the 0-1 indicator of a subset of vertices that forms an independent set of size q in H.

19. Exercise (FGLSS Graph). Show that the maximum independent set in the FGLSS graph H above is of size at most m. Further, if ψ is satisfiable, show that H has an independent set of size m. Finally, show that if there's an independent set of size sm in H then, there's an assignment that satisfies at least s fraction of the constraints in ψ.

Let ψ be the sos hard instance of Max-P problem on n variables and m constraints given by Tulsiani's theorem. The above exercise shows that the FGLSS graph of ψ has an independent set of size at most sm for $s \approx 2k/2^k$.

The next exercise shows how to construct a pseudodistribution that pretends that there's a m-size independent set in the FGLSS graph H of ψ using the exercise on porting low-degree reductions above.

20. Exercise (Pseudodistribution for FGLSS Graph). Let $\tilde{\mathbb{E}}_x$ be the degree d pseudodistribution given by Tulsiani's theorem for the instance ψ of Max-P problem.

Use the exercise on porting low-degree reductions above to conclude that there's a degree d/k pseudodistribution $\tilde{\mathbb{E}}_y$ consistent with the constraints of the independent set polynomial feasibility formulation for h above.

Next, consider $H^{\otimes r}$ obtained by taking r-fold product of the graph H (the vertices are r-tuples of vertices from H and with edges between any pair of r-tuples if one of the r constituent pairs have an edge between them in H) obtained from ψ above.

The next exercise computes the size of the independent set in $H^{\otimes r}$.

21. Exercise (Soundness-of-product-instance). Suppose no assignment satisfies more than s fraction of the constraints of ψ. Show that there's no independent set of size greater than $s^r m$ in $H^{\otimes r}$.

Next, use the exercise on composing independent distributions to conclude there's a pseudodistribution of degree d/k that is consistent with the independent set polynomial feasibility formulation in $H^{\otimes r}$.

What we have so far shows a degree d/k pseudodistribution with an integrality gap of s^r for the independent set problem on a graph with $(2k)^r m^r$ vertices. For r that is superconstant, the number of vertices in this constructed graph are too large. Fortunately, we can just sub-sample appropriate vertices from $H^{\otimes r}$ and select a graph on $100nr$ vertices that has an independent set of size at most $2s^r m$. It is not hard to fill in the details for constructing a pseudodistribution for this sub-sampled instance using the third tool above that deals with projections of pseudodistributions.

Lower Bound for the Parity Problem

As an application of the general strategy for reductions within the SoS framework, we show the following somewhat surprising lower bound for the SoS algorithm.

The parity problem is simple: it asks for dividing a set of n objects into two equal parts. When n is odd, this is obviously impossible. However, we can show that the SoS algorithm "thinks" this is possible - that is, we will construct a pseudodistribution that pretends to be supported on points $x \in \{0,1\}^n$ such that $\sum_{i=1}^n x_i = n/2$. This

result was first shown by Grigoriev via a reduction from instances of r-XOR problem traditionally referred to as *Tseitin Tautologies* in algebraic proof complexity.

22. Theorem (Grigoriev's Hardness for the Parity Problem). *There exists a pseudodistribution of degree $\Omega(n)$ that satisfies the constraints 1) $x_i^2 = x_i$ for every $i \in [n]$ 2) $\sum_{i=1}^n x_i = n/2$ for any n large enough.*

In particular, the above theorem shows the existence of an $\Omega(n)$-degree pseudodistribution even for odd n establishing the lower bound for the parity problem.

The proof is via a reduction from the lower bound for a special kind of XOR instance known in proof complexity literature as *Tseitin Tautology*. The construction and proof is same as in the proof of Grigoriev's theorem above - we describe it next. The idea for constructing the instance is very simple. We start with an expander graph G on vertex set $[n]$. The following existence of expander graphs is easy to show (by analyzing the expansion of random d-regular graphs, for example):

23. Lemma (good expanders exist). *For every n large enough and even $d \geq 4$, there exist graphs d-regular graphs G such that for every $S \subseteq [n]$, $|S| \leq n/2$, $|\Gamma(S)| \geq (1+\epsilon)|S|$ where $\Gamma(S)$ is the set of vertices in $[n]$ adjacent to some vertex in S in G and $\epsilon > 0$ is some fixed constant.*

For each edge e in the graph G, we introduce a variable x_e. The graph G yields a d-XOR system with the constraints $\sum_{e:e\ni i} x_e = 1 \mod 2$ for every $i \in [n]$. Notice that these are n equations over $dn/2$ variables. For any x, $\sum_{i\in[n]} \sum_{e:e\ni i} x_e = \sum_{e\in G} 2x_e = 0 mod 2$. On the other hand, if n is odd, then, $\sum_{i\in[n]} \sum_{e:e\ni i} x_e = \sum_{i\in[n]} +1 = 1 \mod 2$. Thus, if n is odd, there cannot exist an x satisying the given system linear equations modulo 2. > Nevertheless, it is easy to construct a pseudodistribution satisfying $x_e^2 = x_e$ and all the constraints in the instance constructed above using a strategy similar to the one employed in the proof of Grigoriev's theorem above - we record this in the lemma below.

24. Lemma (Grigoriev's Hardness for Tseitin Tautology). *Let G be a $(1+\epsilon)$-expanding d-regular graph on $[n]$ and ψ be the corresponding d-XOR instance. Then, there's a degree $n/4$ pseudodistribution satifying $x_e^2 = x_e$ and all the constraints in ψ.*

Let $E(i)$ be the edges incident to vertex i in G. As before, we use the χ-basis to specify the associated pseudoexpectation by the same strategy as in the proof of Grigoriev's theorem.

For every set $S \subseteq [n]$, $|S| \leq n/4$, let $\tilde{\mathbb{E}}[\chi_{\Delta_{i\in S}E(i)}] = (-1)^{|S|}$.

The following claim will be useful to show that the pseudoexpectation above is well-defined.

25. Lemma (small sets cannot cancel out). *For every $S \subseteq [n]$, $\Delta_{i\in S}E(i) \neq \emptyset$.*

The proof is simple.

Proof. Observe that $\Delta_{i\in S}E(i) = \emptyset$ if and only if every edge $e \in E(i)$ appears twice in $E(i)$. This can happen only if $N(S) = S$ which is impossible since $|S| \leq n/2$ and G is $(1+\epsilon)$-expanding. □

Proof of Lemma 24. To see why the above lemma implies that $\tilde{\mathbb{E}}$ above is well-defined, observe that for any two sets $S_1, S_2 \subseteq [n]$ of size at most $n/4$ if $\Delta_{i\in S_1}E(i) = \Delta_{i\in S_2}E(i)$, then, $S_1 \Delta S_2$ is a set of size at most $n/2$ and $\Delta_{i\in S_1\Delta S_2}E(i) = \emptyset$ - such a set however cannot exist by the Lemma above. Thus, the $\tilde{\mathbb{E}}$ defined above cannot give inconsistent values.

The fact that $\tilde{\mathbb{E}}$ satisfies $\tilde{\mathbb{E}}[1] = 1$ and all the constraints of the instance corresponding to G are immediate from above. The proof of PSDness of $\tilde{\mathbb{E}}$ is entirely analogous to the one in Grigoriev's theorem and we skip it here. □

We now present the reduction from the Tseitin Tautology instance on $dn/2$ variables corresponding to a d-regular graph to an instance of the parity problem on $m = 2dn + n$ variables. Observe that for odd n, m is odd.

Proof of Theorem 22. Let G be the d-regular expander graph defining the Tseitin Tautology instance above. For each undirected edge $\{i, j\}$ in the graph G, construct 4 variables $y_{i,j,0}, y_{i,j,1}, y_{j,i,0}, y_{j,i,1}$. In addition, for every vertex i of G, we have a variable y_i. We let I_G be the instance of the parity problem defined on y variables by the constraints: $y_\alpha^2 = y_\alpha$ for every index α for the y variables and $\sum_\alpha y_\alpha = (2dn + n)/2$. □

We will show the following:

26. Lemma (pseudodistribution-for-parity-problem). *Suppose there exists a degree t pseudodistribution for the Tseitin Tautology instance on G. Then, there exists a degree t/d pseudodistribution for instance I_G of the parity problem.*

Proof. From Exercise **??**, it is enough to show that for any x that satisfies the constraints in the Tseitin Tautology instance corresponding to G, there is a y such that every y_α is a degree r function of x and y satisfies the constraints of the parity problem instance I_G.

We will construct an assignment to y variables such that for every i, j, a, $y_{i,j,a}$ is a function of $x_{\{i,\ell\}}$ for $\{i,\ell\}$ edges in G. Since there are d edges incident to i in G, $y_{i,j,a}$ is a degree d function of x. We describe this construction next.

If $x_{i,j} = 1$ and $i < j$, let $y_{i,j,1} = 1$, $y_{j,i,1} = 0$ and $y_{i,j,0} = (-1)^{\sum_{\ell \in \Gamma(i), \ell \leq j} x_{\{i,j\}}}$. If $x_{i,j} = 0$, let $y_{i,j,0} = 0$ and $y_{i,j,1} = 1$. Finally, set $y_i = 1$.

By construction, y_α are functions of at most d values of x and thus are at most degree d functions of x (being multilinear). To verify that they satisfy the parity constraint we will partition y_α into groups of two such that in each group, the included y_α have differing assignments. This will complete the proof.

Let $j_1, j_2, \ldots, j_k$ be the all the neighbors if i in G such that $x_{i,j} = 1$ and $j_{k+1}, \ldots, j_d$ be the neighbors of i such that $x_{i,j} = 0$. Notice that k must be odd since $\sum_{e : e \ni i} x_e = 1$.

For every group $y_{i,j_t,1}$ and $y_{j_t,i,1}$ together for $t \leq k$. Group $y_{i,j_{2t-1},0}$ and $y_{i,j_{2t},0}$ for $t < k/2$. Group $y_{i,k,0}$ and y_i. It is easy to verify from above that every group created has y_αs of differing values. □

The following simple exercise verifies that in fact the above reduction shows a lower bound of $\Omega(n)$-degree for solving the perfect matching problem.

27. Exercise (Lower Bound for Perfect Matching). Show that the pseudodistribution constructed via reduction from the Tseitin Tautology instance above in fact satisfies the constraints of being supported on a perfect matching on $n + 2dn$ vertex complete graph. Conclude a lower bound of $\Omega(n)$-degree for the SoS algorithm to detect that there's no perfect matching in an odd-vertex complete graph.

References

Michael R. Capalbo, Omer Reingold, Salil P. Vadhan, and Avi Wigderson. Randomness conductors and constant-degree lossless expanders. In *STOC*, pages 659–668. ACM, 2002.

Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *STOC*, pages 670–683. ACM, 2016.

Uriel Feige. Relations between average case complexity and approximation complexity. In *STOC*, pages 534–543. ACM, 2002.

Dima Grigoriev. Complexity of positivstellensatz proofs for the knapsack. *Computational Complexity*, 10(2):139–154, 2001a.

Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.*, 259(1-2):613–622, 2001b.

Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561 (electronic), 2006. ISSN 0273-0979. doi: 10.1090/S0273-0979-06-01126-8. URL http://dx.doi.org/10.1090/S0273-0979-06-01126-8.

Dana Moshkovitz and Ran Raz. Two query PCP with sub-constant error. In *FOCS*, pages 314–323. IEEE Computer Society, 2008.

Prasad Raghavendra. Optimal algorithms and inapproximability results for every csp? In *STOC*, pages 245–254. ACM, 2008.

Grant Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *FOCS*, pages 593–602. IEEE Computer Society, 2008.

Madhur Tulsiani. CSP gaps and reductions in the lasserre hierarchy. In *STOC*, pages 303–312. ACM, 2009.

Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.