SOS Lecture 2: Max Cut, Sparsest Cut, Small Set Expansion and some relations of Isoperimetry and Hypercontractivity / Boaz Barak

Lecture notes by Adrian Vladu and Henry Yuen

Suggested reading

- As I mentioned in the email, Spielman's Spectral Graph Theory lectures 1,2 and 6 [1] are good reading for the background to this lecture. Spielman's disclaimer (and in particular the warning that you should "Be skeptical of all statements in these notes that can be made mathematically rigorous") also applies to the lecture notes in this course.
- See Section 6.2 (pages 143-147 in electronic version) of [WS11] for an overview of the Goemans-Williamson Max-Cut algorithm.
- The Cheeger-Alon-Milman Inequality is covered in many places. One good source is Trevisan's CS359G Lecture Notes [2].
- The Feige-Schechtman graph is from their paper [FS02]. It is also described in several lecture notes. One good source is the lecture of Ryan O'Donnell from the CMU "Advanced Approximation Algorithms" course [3].
- The result that degree 4 SOS (or more accurately degree 2 + the squared triangle inequalities) solve max-cut on random geometric (i.e. Feige-Schechtman) graphs is from [BHHS11].
- Chapter 9 ("basic hypercontractivity") in Ryan O'Donnell's book is highly recommended reading. In particular what we show here is that what he calls "The Bonami Lemma" has a degree 4 SOS proof. This result, as well as other applications, and an equivalence between small set expansion and hypercontractive norm bounds is from [BBH$^+$12]. (The direction of the equivalence we described was known before, and in particular appears in O'Donnell's book, but you may be interested in looking at the proof of the other direction.)

  Both papers mentioned above are available on my home page.

Recap and some musings

The Sum of Squares algorithm is parameterized by a number ℓ, known as its *degree*, and its running time is $n^{O(\ell)}$. When $\ell = 1$ this corresponds to *linear programming*, when $\ell = 2$, it corresponds to *semi-definite programming*, and when $\ell = n$ it corresponds to the *brute force/exhaustive search* algorithm.

In this course, we are most interested in the range $2 < \ell \ll n$. To borrow an analogy from Avi Wigderson, this regime is a bit like the "dark matter" of the SOS algorithm. We know it exists, but we have surprisingly few examples of problems that cannot be solved by the degree 2 case, and can provably be solved by SOS of non-trivially small degree. This is related to the well known phenomenon that, while we know by Ladner's theorem that there are infinitely many "NP intermediate"' problems, most natural computational problems are either in **P** or **NP**-hard. In

[1] See `http://www.cs.yale.edu/homes/spielman/561/`

[2] See `http://theory.stanford.edu/~trevisan/cs359g/`, Lectures 3 and 4

[3] Available on `http://www.cs.cmu.edu/~anupamg/adv-approx/lecture16.pdf`.

fact, most natural problems either have a low-exponent polynomial-time algorithm (e.g., n^2 or n^3) or are exponentially hard (e.g., no $\exp(n^{0.99})$ algorithm is known). There are of course problems, such as k-SUM or k-CLIQUE, where the natural exhaustive search algorithm takes time $n^{O(k)}$ where k is some parameter of the problem; I would consider those as "exponential" in the sense that the best algorithm is still exhaustive search even if it runs in polynomial time.

One could play devil's advocate and suggest that maybe the only problems in this "dark matter" are artificial problems such as those constructed by Ladner, and so perhaps studying SOS for degree larger than 2 is a waste of time. In this course, starting with this lecture, we will see several of the few known examples where degree > 2 proofs help. I will leave to your judgment how natural they are, and, most importantly, I hope you manage to find some new ones!

(I should remark that, even if we restrict attention to domains where SOS is optimal, the "dark matter" region I am describing here is not identical to the set of problems that are not in **P** but not **NP** complete. There are very few examples, perhaps the most notable ones arising from Lattice-based cryptography, of natural problems in **NP** that are believed to be exponentially hard but not **NP**-complete.)

Here are some exercises to make sure that you are comfortable with pseudo-distributions:

Exercise 1: Prove the pseudo-distribution Cauchy-Schwarz condition: If μ is a pseudo-distribution of degree at most $2d$, and P, Q are polynomials of degree d then

$$\tilde{\mathbb{E}}_\mu PQ \leq \sqrt{\tilde{\mathbb{E}}_\mu P^2}\sqrt{\tilde{\mathbb{E}}_\mu Q^2}$$

Exercise 2: Recall that we say that a degree-d pseudo-distribution μ satisfies the constraint $\{p = 0\}$ if $\tilde{\mathbb{E}}_\mu pq = 0$ for every q for which this expectation makes sense (i.e. of degree at most $d - \deg p$). Give an example of a pseudo-distribution μ that satisfies $\tilde{\mathbb{E}}p(x) = 0$, but does not satisfy the constraint $\{p = 0\}$. How high can you make the degree of μ?

Exercise 3: Show that if $\tilde{\mathbb{E}}_\mu p^2 = 0$ then $\tilde{\mathbb{E}}_\mu pq = 0$ for every q of degree at most $d/2 - \deg P$.

Exercise 4: (Hölder's inequality) Prove that if μ is a degree 4 distribution over variables $u_1, \ldots, u_n, w_1, \ldots, w_n \in \mathbb{R}^n$ then $\tilde{\mathbb{E}}_\mu \sum_i u(i)^3 w(i) \leq \left(\tilde{\mathbb{E}}_\mu \|u\|_4^4\right)^{3/4} \left(\|w\|_4^4\right)^{1/4}$.

A tale of two problems

Let $G = (V, E)$ be a d-regular graph on n vertices with normalized adjacency matrix A ($A_{i,j} = 1/d$ if $(i, j) \in E$ and $A_{i,j} = 0$ otherwise). Let $L = I - A$ be its normalized *Laplacian matrix*. For a set $S \subseteq V$, we consider the following quantities:

1. The *expansion* or *conductance* of S, denoted by $\phi(S)$, is defined as

$$\phi(S) = \frac{E(S, \overline{S})}{d \cdot \min\{|S|, n - |S|\}}$$

   Note that this number is always in $[0, 1]$; sometimes this number is defined as

$$\phi(S) = \frac{nE(S, \overline{S})}{d|S||\overline{S}|}$$

   (one can see that this is equivalent up to a factor of 2). The conductance of the graph G is defined as $\phi(G) = \min_S \phi(S)$.

2. The (fractional) *cut size* of S is the number

$$\mathsf{cut}(S) = \frac{E(S, \overline{S})}{|E|}$$

Note that for $S = \Theta(n)$, $\mathsf{cut}(S) = \Theta(\phi(S))$. The *maximum cut value* of G, denoted by $\mathsf{maxcut}(G)$ is $\max_S \mathsf{cut}(S)$.

3. The (uniform) *sparsest cut* problem is the task of computing $\phi(G)$ (or possibly also finding the set S such that $\phi(G) = \phi(S)$ — we will not distinguish between these two problems). The *max-cut* problem is the task of computing $\mathsf{maxcut}(G)$ (or finding the set S such that $\mathsf{cut}(S) = \mathsf{maxcut}(G)$).

Known results. We now survey what is known about those two problems, which turn out to be extremely similar in their computational status. In both cases, finding the exact solution is NP hard, and so we are looking for some form of approximation.

A random subset of measure 1/2 will cut half the edges, and in particular this gives an algorithm achieving a cut of value at least $\mathsf{cut}(G)/2$ for the max-cut problem. In fact, this algorithm for max-cut was suggested by Erdös in 1967, and is one of the first analyses of any approximation algorithm.

A priori, it is not so clear how to beat this. Let us consider the case of max-cut. In a random d-regular graph (which is an excellent expander), one cannot cut more than a $1/2 + \epsilon$ fraction of the edges (where ϵ goes to zero as n goes to infinity). But locally, it is hard to distinguish a random d-regular graph from a random d-regular "almost bipartite" graph, where we split the graph into two equal parts L and R and each edge is with probability ϵ inside one of those parts and with probability $1 - \epsilon$ between them. Such a graph G obviously has $\mathsf{maxcut}(G) \geq 1 - \epsilon$ but every neighborhood of it looks like a d-regular tree, just as in the case of a random d-regular graph. For this reason, "combinatorial" (or even linear programming) algorithms have a hard time getting an approximation factor better than 1/2 for max-cut. To support this claim, [CLRS13] shows that no polynomial-time extension of the max-cut linear program can beat the 1/2 factor. For a similar reason, a priori it is not clear how to find any set with $\phi(S) \ll 1/2$, even if $\phi(G) = o(1)$. However, in both those cases it turns out one can beat the "combinatorial" (or linear programming) algorithms.

The famous *Cheeger's Inequality* (or more accurately, its discrete variant by Alon, Alon-Milman, Dodziuk) implies that there is a polynomial time algorithm to find S with $\phi(S) = O(\sqrt{\phi(G)})$. Cheeger's Inequality can be viewed as the degree-2 SOS algorithm. The analogous algorithm for max-cut took more time, but was found eventually by Goemans and Williamson [GW95] who gave an algorithm, based on rounding the degree-2 SOS algorithm; on an input graph G with $\mathsf{maxcut}(G) \geq 1-\epsilon$, it find a set S with $\mathsf{cut}(S) \geq 1-f(\epsilon)$ for some $f(\epsilon) = O(\sqrt{\epsilon})$. This algorithm is often described in terms of its approximation ratio $\mathsf{cut}(S)/\mathsf{maxcut}(G)$, which is $\min_{\epsilon>0} \frac{1-f(\epsilon)}{1-\epsilon} \approx 0.878$. Leighton and Rao gave a polynomial-time algorithm to find S with $\phi(S) = O(\log n)\phi(G)$ [LR99] and in a breakthrough work, Arora, Rao and Vazirani improved this to an algorithm that outputs a set S with $O(\sqrt{\log n})\phi(G)$ [ARV09]. Their algorithm uses the degree 4 SOS algorithm, and we will see it in this course. Shortly thereafter, Agarwal, Charikar, Makarychev and Makarychev gave the analogous result for max-cut, namely an algorithm that given G with $\mathsf{maxcut}(G) = 1 - \epsilon$, outputs S with $\mathsf{cut}(S) \geq 1 - O(\sqrt{\log n})\epsilon$ [ACMM05].

Assuming the *Unique Games Conjecture* (or its close variant the *Small-Set Expansion Conjecture*), the algorithms of Cheeger and Goemans-Williamson are optimal. Namely, there is no poly-time algorithm to find S with $\phi(S) = o(\sqrt{\phi(G)})$, and no poly-time algorithm that given G with $\mathsf{maxcut}(G) = 1 - \epsilon$, finds a set S with $\mathsf{cut}(S) = 1 - o(\sqrt{\epsilon})$. One way to think of the Unique

Games Conjecture is that it is saying that *degree* 2 *SOS is special* in the sense that for great many problems, improving on it requires to going to degree $n^{\Omega(1)}$. (A priori you could perhaps think that degree 2 is very special, in the sense that improving on it would require degree $\Omega(n)$, but that has been refuted by the sub-exponential algorithm for unique games [ABS10] and its encapsulation within the SOS framework [BRS11, GS11]. Degree 2 SOS *is* special in the sense that it seems much easier to for us to analyze, but whether degree 4 offers no improvements, or is just more challenging for us to prove that it does, remains to be seen. We will see today and in the next lectures some examples where higher (but not super high) degree does help, but as of now these still fall short of disproving the UGC.

Remark: Isoperimetry, local testing, and extremal questions

The max-cut and sparsest cut problems are all special case of the task of determining isoperimetric properties of graphs.
The classical isoperimteric inequality states that the circle is the body with the most area for a given perimeter. More generally, the isoperimetric question in a particular space, is to determine the set in that space with the smallest surface area given some prescribed volume. (With the circle being the answer in two dimensional Euclidean space, and spheres in higher dimensions.) Such questions have a great many applications in a variety of areas, including geometry, graph theory, probability theory and mathematical physics. In particular isoperimetric inequalities are often used to analyze the convergence times of random walks. For a graph, $\phi(S)$ is a natural proxy for the surface area of a set S, and so one variant of the isoperimetric question is to find, given some number $\delta > 0$, the set S of size δn that minimizes $\phi(S)$. This is known as the "small set expansion" question and is intimately related to the unique games conjecture. We are often able to prove isoperimetric inequalities for particular families of graphs, and understanding whether or not these proofs can be "SOS'ed" with low degree is key to understanding the power of this algorithm.
Isoperimetric inequalities are a special case of a more general paradigm of extremal questions in mathematics. Such questions again arise in great many areas, and in particular not just in geometry but also in coding theory and additive combinatorics. In many cases we have some collection of objects Ω (e.g. all subsets of some space of a certain size, or maybe all strings of some length, or all subsets of some group) and some parameter or "test" $T : \Omega \to \mathbb{R}$ (e.g., $T(S)$ can be the surface area of some set S, or $T(S)$ might measure the probability that a certain local test fails on S, or the size of the set $S+S$ where $+$ is the group operation). The generalization of an isoperimetric inequality would be to show that there is some set $\mathcal{C}$ of "special" objects (e.g., the spheres, or codewords of some code, or subgroups) that minimize $T(\cdot)$. The next natural question is the "unique decoding" question, showing that if $T(S)$ is close to the minimum, then S itself is close to some member of $\mathcal{C}$. Another natural question is the "list decoding" question, showing that if $T(S)$ is much smaller than the average value of $T(\cdot)$, then S is at least somewhat correlated with an element of $\mathcal{C}$.

Linear algebra view

Let x be the $\{\pm 1\}$ characteristic vector of S (i.e., $x_i = +1$ if $i \in S$, and $x_i = -1$ otherwise).
Then

$$\langle x, Lx \rangle = \langle x, \left(I - \tfrac{1}{d}A\right) x \rangle = \sum_i x_i^2 - \tfrac{1}{2d} \sum_{i \sim j} x_i x_j = \tfrac{1}{2d} \sum_{i \sim j} (x_i - x_j)^2 = 4E(S, \overline{S})/d = 2n \cdot \mathsf{cut}(S)$$

Observe that each edge is counted twice in the sum over $i \sim j$, and contributes 4 if it belongs to the cut.

Similarly, if x is the $\{0,1\}$ characteristic vector instead (i.e., $x_i = 1$ if $i \in S$ and $x_i = 0$ otherwise) and $|S| \leq n/2$ then

$$\langle x, Lx\rangle = \tfrac{1}{d}E(S,\overline{S}) = \phi(S)\|x\|_2^2$$

(since $\|x\|_2^2 = |S|$).

Exercise 5: Show that if x is the mean 0 characteristic vector of S (i.e., $x_i = n - |S|$ if $i \in S$ and $x_i = -|S|$ if $i \notin S$) then $\langle x, Lx\rangle/\|x\|^2 = \frac{nE(S,\overline{S})}{d|S||\overline{S}|}$. Show that this also equals $\phi(S)$ up to a multiplicative factor of two.

1 The Goemans Williamson Algorithm

The Goemans-Williamson algorithm takes as input a graph G with $\mathsf{maxcut}(G) \geq 1-\epsilon$ and outputs a set S such that $\mathsf{cut}(S) \geq 1 - O(\sqrt{\epsilon})$. It follows from the theorem below:

Theorem 1 (Goemans-Williamson)**.** *There is a polynomial-time algorithm R which, given*

- *an n-vertex d-regular graph $G = (V,E)$*
- *a degree 2 pseudo-distribution μ, such that $\tilde{\mathbb{E}}_{x\sim\mu} x_i^2 = 1$ and $\tilde{\mathbb{E}}\langle x, Lx\rangle \geq 2n(1-\epsilon)$*

outputs a vector $z \in \{\pm 1\}^n$, such that $\langle z, Lz\rangle \geq 2n(1 - f_{GW}(\epsilon))$ where $f_{GW}(\epsilon) \leq 10\sqrt{\epsilon}$.

This immediately implies the algorithm, because given a graph with $\mathsf{maxcut}(G) \geq 1-\epsilon$, running the degree-2 SOS algorithm on $\{x_i^2 = 1\ \forall i, \langle x, Lx\rangle = 2n(1-\epsilon)\}$ yields a valid pseudo-distribution satisfying the requirements for Theorem 1. Running algorithm R, then gives an integral cut of value at least $1 - f_{GW}(\epsilon) = 1 - O(\sqrt{\epsilon})$.

At the heart of the proof of the Goemans-Williamson algorithm is the following very useful lemma:

Lemma 2 (Quadratic Sampling Lemma)**.** *Let μ be a degree-2 pseudo-distribution over $\mathbb{R}^n$. Then there is a poly-time that algorithm can sample from a Gaussian distribution y over $\mathbb{R}^n$ such that $\mathbb{E}p(y) = \tilde{\mathbb{E}}_{x\sim\mu}p(x)$ for every polynomial p of degree at most 2.*

Note on notation: In the rest of this course, in cases where there is little chance of confusion, we will often denote a pseudo-distribution as $\{x\}$ rather than μ and then use notation such as $\tilde{\mathbb{E}}p(x)$ instead of $\tilde{\mathbb{E}}_{x\sim\mu}p(x)$. So, we could also write this lemma as:

Lemma 2 (Quadratic Sampling Lemma)**.** Let $\{x\}$ be a degree-2 pseudo-distribution over $\mathbb{R}^n$. Then there is a poly-time that algorithm can sample from a Gaussian distribution y over $\mathbb{R}^n$ such that $\mathbb{E}p(y) = \tilde{\mathbb{E}}p(x)$, for every polynomial p of degree at most 2.

Why does Lemma 2 imply Theorem 1? The theorem follows by simply letting $z_i = \mathsf{sign}(y_i)$. Note that under our assumptions

$$\tfrac{1}{2n}\tilde{\mathbb{E}}\langle x, Lx\rangle = \tfrac{1}{2n}\tilde{\mathbb{E}}\left(\tfrac{1}{2d}\sum_{i\sim j}(x_i - x_j)^2\right) = \tfrac{1}{4dn}\sum_{i\sim j}\tilde{\mathbb{E}}(x_i - x_j)^2 \geq 1-\epsilon$$

and hence

$$\tfrac{1}{4dn} \sum_{i \sim j} \mathbb{E}(y_i - y_j)^2 \geq 1 - \epsilon$$

We need to show that $\tfrac{1}{4dn} \sum_{i \sim j} \mathbb{E}(z_i - z_j)^2 \geq 1 - O(\sqrt{\epsilon})$. Using convexity, this follows from the following claim:

Claim: Let y, y' be Gaussian variables with $\mathbb{E}y^2 = \mathbb{E}y'^2 = 1$ such that $\mathbb{E}(y - y')^2 \geq 1 - \delta$. Then $\Pr[\mathsf{sign}(y) = \mathsf{sign}(y')] \leq 10\sqrt{\delta}$.

This can be proven via a standard calculation (**Exercise 6:** do this), and essentially follows from the fact that two unit vectors of inner product $1 - \delta$ have distance roughly $\sqrt{\delta}$. Thus all that remains is to prove the quadratic sampling lemma:

Proof of the quadratic sampling lemma. By shifting, it suffices to consider the case that $\tilde{\mathbb{E}} x_i = 0$ for all i (**Exercise 7:** show this). Let M be the $n \times n$ matrix such that $M_{i,j} = \tilde{\mathbb{E}} x_i x_j$. Since $\tilde{\mathbb{E}} f(x)^2 \geq 0$ for every linear function f, it must holds that M is positive semidefinite and hence $M = VV^\top$ for some matrix V. Another way to say this is that $M_{i,j} = \langle v^i, v^j \rangle$ for some N, and for some vectors $v^1, \ldots, v^n \in \mathbb{R}^N$. Let g be a standard Gaussian vector in $\mathbb{R}^N$ and define $y_i = \mathbb{E}\langle v^i, g \rangle$. (Note that, as desired, $\mathbb{E}\langle v^i, g \rangle = 0$ for all i.) For every i, j we have that

$$\mathbb{E} y_i y_j = \mathbb{E}\langle v^i, g \rangle \langle v^j, g \rangle = \sum_{k,\ell} \mathbb{E} v^i_k g_k v^j_\ell g_\ell$$

but since we have that $\mathbb{E} g_k g_\ell = \begin{cases} 1, & \text{if } k = \ell \\ 0, & \text{if } k \neq \ell \end{cases}$, this equals

$$\sum_k v^i_k v^j_k = \langle v^i, v^j \rangle = M_{i,j}$$

Hence y agrees with x on all quadratic monomials and hence on all quadratic polynomials as well. $\square$

Exercise 8: Prove that if x is an actual distribution over $\mathbb{R}^n$ with mean 0^n, taking the value $x^{(\alpha)} \in \mathbb{R}^n$ with probability p_α, where α ranges over some finite index set I, we would get an equivalent distribution to the one above by letting $y = \sum_{\alpha \in I} g_\alpha x^{(\alpha)}$, where g_α is a Gaussian with mean 0 and variance p_α.

Note that the algorithm resulting from the proof of Theorem 1 performs the following: given the vectors $v_1, \ldots, v_n$ that arise from the pseudo-distribution $\{x\}$, we obtain a cut z (identified with a vector in $\{\pm 1\}^n$) by choosing a random gaussian vector g and outputting $z_i = \mathsf{sign}\langle v^i, g \rangle$. That is, we cut the vertices based on which side of the hyperplane defined by g they fall on. For this reason, this rounding technique is often known as "random hyperplane rounding".

Also note that in analyzing this algorithm we didn't use "Marley's Corollary", since the quadratic sampling lemma is easy enough to prove even without assuming that $\{x\}$ was an actual distribution, but we will find that assumption useful as we analyze algorithms using higher degrees, including the Arora-Rao-Vazirani algorithm.

Remark: Vector view of SOS

In the proof above, we used the convenient fact that an $n \times n$ matrix M is psd if and only if there are vectors $v_1, \ldots, v_n \in \mathbb{R}^\Omega$ (for some Ω) such that $M_{i,j} = \langle v_i, v_j \rangle$. Thus a degree 2 pseudo-distribution can be completely characterized by such vectors (along with another vector which would correspond to the first moments / averages). This can be generalized to higher degrees as follows:

Exercise 9: Let $\mathcal{P}$ be some basis for the degree $\leq d/2$ polynomials (you can think of the monomial basis). Prove that a bilinear operator $M : \mathbb{R}^n_{d/2} \times \mathbb{R}^n_{d/2} \to \mathbb{R}$ is a degree d pseudo-expectation operator if and only if there exists vectors $\{v_p\}_{p \in \mathcal{P}}$ in $\mathbb{R}^N$ for some N such that $M(p,q) = \langle v_p, v_q \rangle$ and $\langle v_p, v_q \rangle = \langle v_r, v_s \rangle$ whenever $pq = rs$.

Note that if M corresponded to the expectation of an actual random variable X, which we can think of as a function from some probability space Ω to $\mathbb{R}^N$, then we could choose the vector v_p to have as its ω^{th} coordinate the value $p(X(\omega))$, where we will use the *expectation* inner product, i.e. $\langle\langle\langle u, v \rangle\rangle\rangle = \mathbb{E}_{\omega \in \Omega} u(\omega) v(\omega) = \sum_{\omega \in \Omega} p_\omega v_i(\omega) v_j(\omega)$ (p_ω is the probability of the element ω; the space Ω can even be infinite in which case the sum is replaced with an integral).

2 Can we do better with degree 2 SOS?

It is perhaps surprising that we can do better than the random cut algorithm, but knowing that we can bypass it whets our appetite for more. So far, we don't know if we can beat the Goemans-Williamson algorithm, but we do know that won't be possible with the degree-2 SOS program.

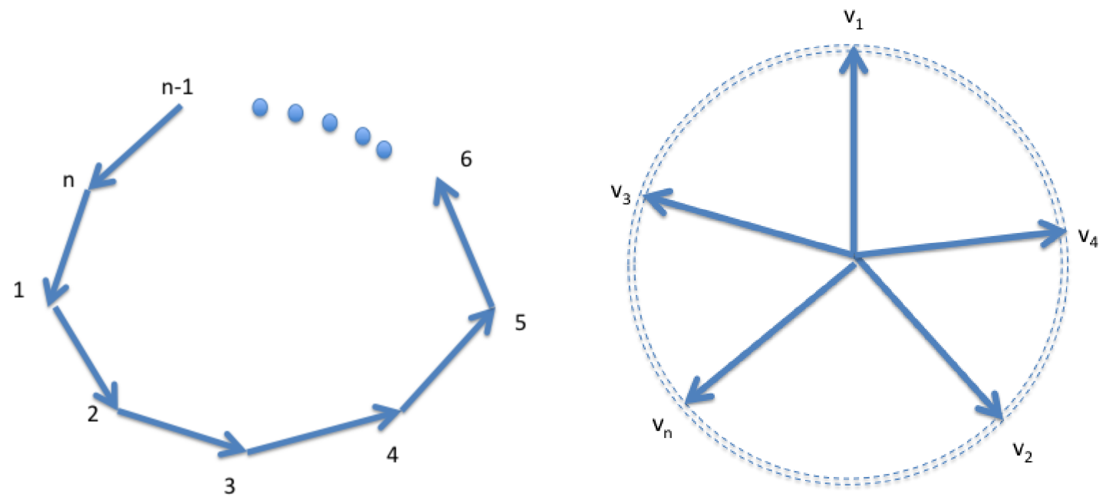
Theorem 3. *There is a graph G and $\epsilon > 0$ such that $\mathsf{maxcut}(G) \leq 1 - \sqrt{\epsilon}/10$ but there is a degree-2 pseudo-distribution $\{x\}$ such that $\tilde{\mathbb{E}} x_i^2 = 1$ for all i and $\tilde{\mathbb{E}} \langle x, Lx \rangle \geq 2n(1-\epsilon)$.*

Proof. The graph is simply the odd cycle on $n = 1/\sqrt{\epsilon}$ vertices. Since the graph is not bipartite, every cut must cut at least one edge, so $\mathsf{maxcut}(G) \leq 1 - 1/n = 1 - \sqrt{\epsilon}$. For the pseudo distribution, we arrange unit vectors $v_1, \ldots, v_n$ along the two dimensional circle such that $\langle v_i, v_j \rangle = -1 + \epsilon$ if $j = i + 1 \pmod{n}$. Let $\tilde{\mathbb{E}} x_i x_j = \langle v_i, v_j \rangle$. Then, for all i, $\tilde{\mathbb{E}} x_i^2 = \langle v_i, v_i \rangle = 1$, and $\tilde{\mathbb{E}} \langle x, Lx \rangle = \tilde{\mathbb{E}}(\sum_i x_i^2 - \frac{1}{2} \sum_{i \sim j} x_i x_j) = n(2-\epsilon) = 2n(1-\epsilon/2) \geq 2n(1-\epsilon)$. □

One issue with this result, apart from the fact that the odd cycle doesn't seem like a very hard instance, is that the value of ϵ here is $1/n^2$, which means that even finding a cut of $1 - 1/\sqrt{\epsilon}$ is pretty good (in fact, the best one can do, given that the graph isn't bipartite). However, this of course can be easily fixed by simply considering the disjoint union of many odd $1/\sqrt{\epsilon}$ cycles, hence yielding an instance where ϵ is independent of n. Another issue is to determine the right constant, and more generally, for every $\epsilon > 0$, come up with a graph that has a degree-2 pseudo-distribution pretending to range over cuts with value $1 - \epsilon$, but where the true max cut is at most $1 - f_{GW}(\epsilon) + o(1)$ where $f_{GW}(\cdot)$ is the function obtained by the proof of the Goemans-Williamson theorem (Theorem 1). This was achieved by Feige and Schechtman [FS02], who defined the following graph:

For $\epsilon > 0$, $\ell, n \in \mathbb{N}$, the Feige-Schechtman graph $FS(\epsilon, \ell, n)$ is obtained by sampling n random unit vectors $v_1, \ldots, v_n$ in $\mathbb{R}^\ell$ and letting $i \sim j$ if $\langle v_i, v_j \rangle \leq -1 + \epsilon$. We will typically think of the case that n is exponentially larger than ℓ and so this graph closely approximates the infinite graph where vertices are all vectors in the unit ℓ-dimensional sphere. In [FS02], Feige and Schechtman proved the following two results about this graph:

Figure 1: The odd-cycle graph, along with a depiction of a vector solution for the degree-2 SoS program.



Lemma 4. *There is a degree-2 pseudo-distribution $\{x\}$ such that $\tilde{\mathbb{E}}x_i^2 = 1$ for all i and $\tilde{\mathbb{E}}(x_i - x_j)^2 \geq 2(1-\epsilon)$ for all $i \sim j$.*

Proof. This is essentially by construction. Define $\tilde{\mathbb{E}}x_i x_j = \langle v_i, v_j \rangle$ where $v_1, \ldots, v_n$ are the vectors used to define the graph. Note that this is PSD and hence is a valid degree-2 pseudo-expectation operator, and that it satisfies the conditions of the Lemma. □

Lemma 5. *For every $\delta > 0$, if n is large enough, with high probability* $\mathsf{maxcut}(FS(\epsilon, \ell, n)) \leq 1 - f_{GW}(\epsilon) + \delta$

This is the heart of their proof. To show this, one needs to show that the maximum cut in the Feige-Schechtman graph is obtained by a *hyperplane cut*, namely by a set S of the form $\{i : \langle v_i, a \rangle > 0\}$ for some vector a. This turns out to be related to classical isoperimetric results of Borell. We will prove a slightly weaker result. Namely, that the Feige-Schechtman graph is at least not worse than the cycle:

Lemma 6. *There is some constant $c > 0$ such that for every $\delta > 0$, if n is large enough,* $\mathsf{maxcut}(FS(\epsilon, \ell, n)) \leq 1 - c\sqrt{\epsilon}$.

Proof. We will assume that $1/\sqrt{2\epsilon - \epsilon^2} = k$ for some odd integer k (this assumption can be waived with a bit more work, at the cost of having a suboptimal value for the constant c). For sufficiently large n, we can imagine that the FS graph is simply on the continuous sphere. Pick a random edge $i \sim j$ of the graph, and consider the intersection of the sphere with the plane spanned by v_i and v_j. Inside that plane we can find a k-cycle subgraph of the graph that contains the edge $i \sim j$. By following this approach we obtain a set $C_1, \ldots, C_N$ of k-cycles that uniformly covers the edges of the Feige-Schechtman graph (i.e., each edge of the Feige-Schechtman graph is contained in about the same numbe of cycles). Now, for every possible cut S, it must miss at least one edge from every one of those C_i's. But by the uniformity condition, if a cut misses a α fraction of the edges in the FS graph, on average it should miss an α fraction (or $\alpha \pm o(1)$, to account for our finite approximations) of the edges in the cycles. □

This is a weaker result because the constant c will not match that in f_{GW}.

Similar results hold for *sparsest cut*— the cycle (here it doesn't matter if it's even or odd) yields an example of a graph with eigenvalue gap of ϵ but where the best cut has conductance $\sqrt{\epsilon}$, and the analog of the Feige-Schechtman graph can give better constant dependence.

3 Can we do better with higher degree?

This is a question many people are interested in, and we don't know the answer. As mentioned above, Khot's *Unique Games Conjecture* implies that no polynomial-time algorithm (or even an $\exp(n^{o(1)})$-time one) can beat the Goemans-Williamson algorithm. So, in particular it should mean that using SOS algorithm with degree $n^{o(1)}$ will not yield improved performance over degree 2. However, we don't even know if degree 4 SOS doesn't do better than degree 2. Indeed, the known hard instances for degree 2, including the odd cycle and the Feige-Schechtman graph, can in fact be solved via degree 4 SOS (for some other instances the best known bound is 16 or so, but we have no evidence that degree 4 doesn't work as well).

Lemma 7. *Let n be odd. There is no degree-4 pseudo-distribution $\{x\}$ over $\mathbb{R}^n$ consistent with the constraints $\{x_i^2 = 1\}_{i=1..n}$ such that $\tilde{\mathbb{E}} \sum_{i=1}^n (x_i - x_{i+1})^2 > 4(n-1)$ (identifying x_{n+1} with x_1).*

The proof of this lemma follows from the following exercise

Exercise 10: (Squared Triangle Inequality) Let $\{x\}$ be a degree-4 pseudo-distribution over $\mathbb{R}^n$ consistent with the constraints $\{x_i^2 = 1\}$. Then for all $i, j, k \in [n]$, $\tilde{\mathbb{E}}(x_i - x_k)^2 \le \tilde{\mathbb{E}}(x_i - x_j)^2 + \tilde{\mathbb{E}}(x_j - x_k)^2$.

Note that if $\{x\}$ is an *actual* distribution consistent with these constraints, then it means that it is supported on $\{\pm 1\}^n$, and hence $\mathbb{E}(x_i - x_j)^2 = 4 \Pr[x_i \neq x_j]$ which immediately implies the inequality. Therefore, the 3-variate polynomial $(x_i - x_j)^2 + (x_j - x_k)^2 - (x_i - x_k)^2$ is non-negative on $\{\pm 1\}^3$ which immediately implies the result for degree-6 pseudo-distributions. (Can you see why?) This is just as good for our purposes, but working out the degree 4 case should be a nice exercise.

We can now prove the lemma:

Proof. Using the equation $(a+b)^2 = 2a^2 + 2b^2 - (a-b)^2$ we get that

$$\tilde{\mathbb{E}} \sum_{i=1}^n (x_i + x_{i+1})^2 < 4n - 4(n-1) = 4$$

By the triangle inequality applied to the variables x_i, x_{i+2} and $-x_{i+1}$, we get that

$$\tilde{\mathbb{E}}(x_i - x_{i+2})^2 \le \tilde{\mathbb{E}}(x_i + x_{i+1})^2 + (x_{i+1} + x_{i+2})^2$$

which repeating $(n-1)/2$ times we get that

$$\tilde{\mathbb{E}}(x_i - x_{i+1})^2 \le \sum_{j \in (i+1,..,n,1,..,i-1)} \tilde{\mathbb{E}}(x_j + x_{j+1})^2$$

If we sum this over all i's, then on the RHS every term $\tilde{\mathbb{E}}(x_j + x_{j+1})^2$ gets counted $n-1$ times and so we get

$$\sum_i \tilde{\mathbb{E}}(x_i - x_{i+1})^2 \le (n-1) \sum_j \tilde{\mathbb{E}}(x_j + x_{j+1})^2 < (n-1)4$$

contradicting our assumptions. □

This results immediately implies that the degree 4 SOS algorithm can also certify that the max-cut of the $FS(\epsilon, \ell, n)$ graph is $1 - \Omega(\sqrt{\epsilon})$ if n is large enough (can you see why?). Interestingly, it can be shown that if we choose a significantly smaller n (though still exponential in ℓ) so that almost all short cycles (and in particular the odd ones) disapper, then as long as the average degree remains large enough, the value of the maximum cut remains $1 - \Omega(\sqrt{\epsilon})$. However, the proof that the degree 4 SOS algorithm certifies this breaks down. Nonetheless, it turns out that the degree-4 SOS algorithm still gives a value of $1 - \Omega(\sqrt{\epsilon})$ even in this regime (see Barak, Hardt, Holenstein, Steurer [BHHS11]).

4 SOS'ing proofs of isoperimetric inequalities

Moving beyond max cut, another important problem is the *small set expansion* problem. This is the task, given some graph $G = (V, E)$ and δ, of computing $\min_{|S| \leq \delta |V|} \phi(S)$. Once again, the question is about approximating it, and the *small set expansion conjecture* posits that it is NP hard to determine if this quantity is at most ϵ or at least $1-\epsilon$, where $\epsilon = 1/O(\log(1/\delta))$. (The conjecture was stated by Raghavendra and Steurer [RS10] in a slightly different form, and its equivalence to this form was shown by Raghavendra, Steurer and Tulsiani [RST10].)

The *Boolean cube* (i.e., the graph on 2^ℓ vertices identified with $\{\pm 1\}^\ell$ such that $x \sim y$ if $\sum |x_i - y_i| = 2$) is a canonical example of a small set expander. That is, even though the graph is not a great expander, since for example $1 - 1/\ell$ fraction of the edges touching the set $S = \{(+1, x) : x \in \{\pm 1\}^{\ell-1}\}$ stay inside it, for smaller set a much larger fraction of the edges go out. In fact, one can show that for every k, the sets S of measure 2^{-k} that minimize $\phi(S)$ have the form $S = \{(\alpha, x) : x \in \{\pm 1\}^{\ell-k}\}$ for some $\alpha \in \{\pm 1\}^k$. (Note that these sets have $1 - k/\ell$ of their edges staying inside them.)

How do you prove such a thing (at least approximately)? The key here is again linear algebra. Recall that for every set S of measure less than 1/2, $\langle x, Lx \rangle / \|x\|^2 = \phi(S)$ where x is the $\{0, 1\}$ characteristic vector of S. Therefore, to prove that if $|S|$ is small then $\phi(S)$ is large, it is enough to show that sparse vectors are not close to the low eigenspace of the operator L. Specifically, we have the following result:

Lemma 8. *Let $G = (V, E)$ be regular graph, $\lambda \in (0, 1)$ and W be the span of eigenvectors of $L(G)$ corresponding to eigenvalue at most λ. If every $w \in W$ satisfies:*

$$\mathbb{E}_i w_i^4 \leq C \left(\mathbb{E}_i w_i^2 \right)^2 \tag{1}$$

then for every set S of measure δ set,

$$\phi(S) \geq \lambda(1 - \sqrt{C\delta})$$

To understand the lemma, let us try to parse what (1) means in the case that w is the 0/1 characteristic vector of some set S of measure μ. In this case the LHS equals μ and the RHS equals $C\mu^2$, and so we get that $\mu \geq 1/C$. Thus in particular (1) implies that the space W does not contain the characteristic vector of any set of measure $< 1/C$, and the conclusion of the Lemma is that if S has measure $\ll 1/C$ then it has expansion at least $\lambda - o(1)$, which means that its characteristic vector has almost all of its mass outside W.

Proof. Throughout this proof, it will be convenient for us to use the *expectation* norms and inner product, and so we denote $\|\!|x|\!\|_p = (\mathbb{E}_i |x_i|^p)^{1/p}$ and $\langle\!\langle\!\langle x, y \rangle\!\rangle\!\rangle = \mathbb{E} x_i y_i$. Thus (1) translates to

$\|w\|_4^4 \leq C\|w\|_2^4$ for every $w \in W$. Let S be the set and x its 0/1 characteristic vector. Note that we still have

$$\phi(S) = \langle\langle\langle x, Lx\rangle\rangle\rangle / \|x\|^2 .$$

Write $x = x' + x''$ where $x' \in W$ and $x'' \in W^{\perp}$. Our main claim is

CLAIM: $\|x'\|_2 \leq C^{1/4}\|x\|_{4/3}$

PROOF OF CLAIM:

$$\|x'\|_2^2 = \langle\langle\langle x', x'\rangle\rangle\rangle = \langle\langle\langle x', x\rangle\rangle\rangle \leq \|x'\|_4\|x\|_{4/3}$$

where the second equality is because x' is a projection of x, and the last inequality is an application of Holder's inequality. The proof then follows using $\|x'\|_4 \leq C^{1/4}\|x'\|_2$.

Given this, since $\|x\|_{4/3} = \delta^{3/4}$ we can use the eigenvector decomposition $(v_1, \ldots, v_n)$ and $(\lambda_1, \ldots, \lambda_n)$ of L to write

$$\langle\langle\langle x, Lx\rangle\rangle\rangle = \sum \lambda_i \langle\langle\langle x, v_i\rangle\rangle\rangle^2$$

bunching together all the vectors with eigenvalue smaller than λ (whose contribution to the sum is non-negative), and all the vectors with eigenvalues larger than λ (who contribute at least $\lambda\|x''\|_2^2$) we get

$$\langle\langle\langle x, Lx\rangle\rangle\rangle \geq \lambda\|x''\|_2^2 = \lambda(\|x\|_2^2 - \|x'\|_2^2) \geq \lambda(\delta - C^{1/2}\delta^{3/2})$$

and thus (using $\|x\|_2^2 = \delta$) we get

$$\phi(S) = \tfrac{\langle\langle\langle x, Lx\rangle\rangle\rangle}{\|x\|_2^2} \geq \lambda(1 - \sqrt{C\delta})$$

□

Let us now see how we apply this result to the Boolean cube. First, we need the following characterization of the Boolean cube

Exercise 11: Let G be the Boolean cube on $\{\pm 1\}^{\ell}$. Prove that the eigenvectors of G are $\{\chi_S\}_{S\subseteq[\ell]}$ where for every $x \in \{\pm 1\}^{\ell}$, $\chi_S(x) = \prod_{i\in S} x_i$ and the eigenvalue corresponding to χ_S is $|S|/\ell$.

Thus, for every λ, the subspace spanned by eigenvectors of eigenvalue at most λ is the set of $f : \{\pm 1\}^{\ell} \to \mathbb{R}$ that are spanned by the functions χ_S with $|S| \leq \lambda\ell$. Thus the following result shows that sufficiently small sets in the hypercube expand a lot

Theorem 9 ($(2,4)$ hypercontractivity)**.** *Let $f = \sum f_\alpha \chi_\alpha$ with $|\alpha| \leq d$. Then*

$$\mathbb{E}_{x\in\{\pm 1\}^{\ell}} f(x)^4 \leq 9^d \left(\mathbb{E}_{x\in\{\pm 1\}^{\ell}} f(x)^2\right)^2 \tag{2}$$

Theorem 9 has a simple proof but underlies many results used in hardness of approximation, social choice theory, and more (see Ryan's book [O'D14]). In particular, as we mentioned, by combining it with Lemma 8 it implies some isoperimetric results on the Boolean cube.

Proof. We prove the result by induction on d and ℓ. (The case $\ell = 0$ or $d = 0$ is trivial.) Separate f to the parts that do and don't depend on ℓ and write

$$f(x) = f_0(x_1, \ldots, x_{\ell-1}) + x_\ell f_1(x_1, \ldots, x_{\ell-1})$$

note that the degree of f_1 is at most $d-1$. Now let us expand $\mathbb{E}f(x)^4$ and note that the expectation of odd powers of x_ℓ vanish (since it is independent from the other variables) and so we get that

$$\mathbb{E}f^4 = \mathbb{E}f_0^4 + f_1^4 + 6f_0^2 f_1^2 \tag{3}$$

By Cauchy Schwarz we can bound the last term by $6\sqrt{\left(\mathbb{E}f_0^4\right)\left(\mathbb{E}f_1^4\right)}$. By induction we can assume $\mathbb{E}f_b^4 \leq 9^d \left(\mathbb{E}f_b^2\right)^2$ for $b = 0, 1$ and so plugging this into (3) we get

$$\mathbb{E}f^4 \leq 9^d \left(\mathbb{E}f_0^2\right)^2 + 9^{d-1}\left(\mathbb{E}f_1^2\right)^2 + 6 \cdot 9^{d-1/2}\mathbb{E}f_0^2\mathbb{E}f_1^2 \leq$$

$$9^d \left(\left(\mathbb{E}f_0^2\right)^2 + \left(\mathbb{E}f_1^2\right)^2 + 2\mathbb{E}f_0^2\mathbb{E}f_1^2\right) =$$

$$9^d \left(\mathbb{E}f_0^2 + \mathbb{E}f_1^2\right)^2$$

but this equals $9^d \left(\mathbb{E}f^2\right)^2$, since $f = f_0 + x_\ell f_1$ and $\mathbb{E}x_\ell f_0 f_1 = 0$. □

Remark: Max-cut, sparsest-cut and small-set expansion as finding non-Gaussian vectors in subspaces

The sparsest-cut, max-cut, and small-set expansion can all be thought of as the problem of finding

$$\min_x p(x)$$

where $p : \mathbb{R}^n \to \mathbb{R}$ is a quadratic polynomial (e.g., $p(x) = \langle x, Lx\rangle$ or $p(x) = -\langle x, Lx\rangle$) and subject to x satisfying certain constraints (e.g. $x \in \{0,1\}^n$ (or sometimes $\{\pm 1\}^n$) in the case of max-cut/sparsest cut, or x is the characteristic vector of a sparse set in the case of small set expansion). By scaling appropriately, we can also assume x is restricted to have unit norm.

Since a vector x of unit norm minimizes a quadratic form $p(\cdot)$ if and only if it resides in the linear subspace correspondings to the small eigenvalues of $p(\cdot)$ (thought of as a linear operator), the problem essentially reduces to finding a vector in (or close to) W that satisfies these constraints. If W was a "generic" subspace of not too high a dimension, then all the vectors w inside it would be rather "smooth" and in particular every $w \in W$ will satisfy that the distribution $X(w)$ which is obtained by sampling a random coordinate i and outputting w_i is close to the Gaussian distribution. So, in some sense all of these problems are about finding non-Gaussian vectors in a subspace. Note that for every w there is some Gaussian distribution that matches the first two moments of $X(w)$. Thus being "non Gaussian" inherently implies looking at higher moments. For example, one can verify that if w is a sparse vector shifted and scaled to have $\mathbb{E}X(w) = 0$ and $\mathbb{E}X(w)^2 = 1$ then $\mathbb{E}X(w)^4$ is much larger than $\mathbb{E}N(0,1)^4$ (indeed this is the underlying reasoning behind Lemma 8, and as I mentioned, some kind of a reverse direction holds as well). So, one can hope that degree > 2 SOS would help with that.

Making this into an SOS proof

Lemma 8 reduces the question of certifying whether a graph is a small set expander to certifying a polynomial equation, and so to understand if the degree 4 SOS algorithm can certify the expansion properties of the Boolean cube, we need to come up with an SOS proof for the $(2,4)$ hypercontractivity theorem (Theorem 9). Eyeballing the proof, we see that it doesn't use the probabilistic method, and so by Marley's Corollary it should have an SOS proof. However, as far as we know, Marley didn't publish his proof in a peer-reviewed journal, and so we'd better doublecheck the case. (I should note that joking aside, it is definitely *not* a universal statement that all known interesting low degree polynomial inequalities that have a non-probabilistic method proof are known to have a low degree SOS proof, in fact Ryan O'Donnell has several interesting open questions along these lines. However so far in my experience, though it took some work, we typically were always able to find such proofs for the statements that arise in analyzing SOS algorithms (or some close-enough approximation of them), and so the main hurdle was to actually phrase the statement as low degree polynomial inequality in the first place.)

Indeed, this turns out to be the case as shown by the following lemma giving an SOS proof for Theorem 9. For two polynomials p, q we write $p \preceq q$ if $p = q - \sum_{i=1}^{m} r_i^2$ for some polynomials $r_1, .., r_m$.

Lemma 10. *Let $d, \ell \in \mathbb{N}$. For every $x \in \{\pm 1\}^n$, let $L_x(\cdot)$ be the linear function in the variables $\{f_\alpha\}_{|\alpha| \le d}$ such that*

$$L_x(f) = \sum_{\alpha} \left(\prod_{i \in \alpha} x_i \right) f_\alpha$$

then

$$\mathbb{E}_{x \in \{\pm 1\}^\ell} L_x(f)^4 \preceq 9^d \left(\mathbb{E}_{x \in \{\pm 1\}^\ell} L_x(f)^2 \right)^2 \tag{4}$$

The notation in this lemma are a bit subtle, and so it's worth taking the time and make sure we parse it. First, note that the lemma immediately implies Theorem 9. Indeed, since $L_x(f) = f(x)$, plugging this in (5) yields (2). However, we use the second notation to emphasize that x is *not* a variable of L_x, nor of the polynomials implicitly defined in (5). These are polynomials in the coefficients of f. In particular, L_x is linear, no matter what the number d is, and both the LHS and RHS of (2) are degree 4 polynomials. The lemma also implies that if $\{f\}$ is a degree-4 pseudo-distribution then

$$\tilde{\mathbb{E}}_f \mathbb{E}_{x \in \{\pm 1\}^\ell} f(x)^4 \le 9^d \tilde{\mathbb{E}}_f \left(\mathbb{E}_{x \in \{\pm 1\}^\ell} f(x)^2 \right)^2$$

We now turn to prove the lemma. The proof is a close variant of the proof of Theorem 9 but we have to be a bit more careful and make a stronger induction hypothesis. In particular, we will prove the following stronger result

Lemma 11. *Let $d, e, \ell \in \mathbb{N}$. For every $x \in \{\pm 1\}^n$, let $L_x(\cdot)$ be the linear function in the variables $\{f_\alpha\}_{|\alpha| \le d}$ such that*

$$L_x(f) = \sum_{\alpha} \left(\prod_{i \in \alpha} x_i \right) f_\alpha$$

and let $L'_x(\cdot)$ be the linear function in the variables $\{g_\alpha\}_{|\alpha| \le e}$ such that

$$L'_x(f) = \sum_{\alpha} \left(\prod_{i \in \alpha} x_i \right) g_\alpha$$

then
$$\mathbb{E}_{x\in\{\pm1\}^\ell} L_x(f)^2 L'_x(g)^2 \preceq 9^{(d+e)/2}\left(\mathbb{E}_{x\in\{\pm1\}^\ell} L_x(f)^2\right)\left(\mathbb{E}_{x\in\{\pm1\}^\ell} L'_x(f)^2\right) \tag{5}$$

Proof. We prove the lemma by induction on ℓ, d, e, again, if any of those is zero then the result is trivial. Below we use the notation $f(x)$ for $L_x(f)$ and $g(x)$ for $L'_x(g)$, but you should remember that $f(x)$ is a linear polynomial in the variables $\{f_\alpha\}$ (rather than being a degree d polynomial in x).

Let f_0, f_1, g_0, g_1 be such that $f(x) = f_0(x) + x_\ell f_1(x)$ and $g(x) = g_0(x) + x_\ell g_1(x)$. Note that the coefficients of f_0, f_1, g_0, g_1 are a linear function of the coefficients of f, g (because $f_0(x) = \frac{1}{2}f(x_1,\ldots,x_{n-1},1)+\frac{1}{2}f(x_1,\ldots,x_{n-1},-1)$ and $f_1(x) = \frac{1}{2}f(x_1,\ldots,x_{n-1},1)-\frac{1}{2}f(x_1,\ldots,x_{n-1},-1)$). Moreover, the monomial-size of f_0, f_1, g_0, and g_1 are at most d, $d-1$, e, and $e-1$, respectively.

Since $\mathbb{E}x_\ell = \mathbb{E}x_\ell^3 = 0$, if we expand $\mathbb{E}f^2g^2 = \mathbb{E}(f_0 + x_\ell f_1)^2(g_0 + x_\ell g_1)^2$ then the terms where x_ℓ appears in an odd power vanish, and we obtain

$$\mathbb{E}f^2g^2 = \mathbb{E}f_0^2g_0^2 + f_1^2g_1^2 + f_0^2g_1^2 + f_1^2g_0^2 + 4f_0f_1g_0g_1$$

By expanding the square expression $2\mathbb{E}(f_0f_1 - g_0g_1)^2$, we get $4\mathbb{E}f_0f_1g_0g_1 \preceq 2\mathbb{E}f_0^2g_1^2 + f_1^2g_0^2$ and thus

$$\mathbb{E}f^2g^2 \preceq \mathbb{E}f_0^2g_0^2 + \mathbb{E}f_1^2g_1^2 + 3\mathbb{E}f_0^2g_1^2 + 3\mathbb{E}f_1^2g_0^2\ . \tag{6}$$

Applying the induction hypothesis to all four terms on the right-hand side of 6 (using for the last two terms that the monomial-size of f_1 and g_1 is at most $d-1$ and $e-1$),

$$\begin{aligned}\mathbb{E}f^2g^2 &\preceq 9^{\frac{d+e}{2}}\left(\mathbb{E}f_0^2\right)\left(\mathbb{E}g_0^2\right) + 9^{\frac{d+e}{2}}\left(\mathbb{E}f_1^2\right)\left(\mathbb{E}g_1^2\right)\\ &\quad + 3\cdot 9^{\frac{d+e}{2}-1/2}\left(\mathbb{E}f_0^2\right)\left(\mathbb{E}g_1^2\right) + 3\cdot 9^{\frac{d+e}{2}-1/2}\left(\mathbb{E}f_1^2\right)\left(\mathbb{E}g_0^2\right)\\ &= 9^{\frac{d+e}{2}}\left(\mathbb{E}f_0^2 + \mathbb{E}f_1^2\right)\left(\mathbb{E}g_0^2 + \mathbb{E}g_1^2\right)\ .\end{aligned}$$

Since $\mathbb{E}f_0^2 + \mathbb{E}f_1^2 = \mathbb{E}(f_0 + x_\ell f_1)^2 = \mathbb{E}f^2$ (using $\mathbb{E}x_\ell = 0$) and similarly $\mathbb{E}g_0^2 + \mathbb{E}g_1^2 = \mathbb{E}g^2$, we derive the desired relation $\mathbb{E}f^2g^2 \preceq 9^{\frac{d+e}{2}}\left(\mathbb{E}f^2\right)\left(\mathbb{E}g^2\right)$. □

References

[ABS10] S. Arora, B. Barak, and D. Steurer. Subexponential algorithms for unique games and related problems. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 563–572. IEEE, 2010. 4

[ACMM05] A. Agarwal, M. Charikar, K. Makarychev, and Y. Makarychev. $o(\sqrt{\log n})$ approximation algorithms for min uncut, min 2cnf deletion, and directed cut problems. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 573–581. ACM, 2005. 3

[ARV09] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56(2):5, 2009. 3

[BBH+12] B. Barak, F. G. Brandao, A. W. Harrow, J. Kelner, D. Steurer, and Y. Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326. ACM, 2012. 1

[BHHS11] B. Barak, M. Hardt, T. Holenstein, and D. Steurer. Subsampling mathematical relaxations and average-case complexity. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 512–531. SIAM, 2011. 1, 10

[BRS11] B. Barak, P. Raghavendra, and D. Steurer. Rounding semidefinite programming hierarchies via global correlation. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 472–481. IEEE, 2011. 4

[CLRS13] S. O. Chan, J. R. Lee, P. Raghavendra, and D. Steurer. Approximate constraint satisfaction requires large lp relaxations. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 350–359. IEEE, 2013. 3

[FS02] U. Feige and G. Schechtman. On the optimality of the random hyperplane rounding technique for max cut. *Random Structures & Algorithms*, 20(3):403–440, 2002. 1, 7

[GS11] V. Guruswami and A. K. Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with psd objectives. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 482–491. IEEE, 2011. 4

[GW95] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995. 3

[LR99] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM (JACM)*, 46(6):787–832, 1999. 3

[O'D14] R. O'Donnell. *Analysis of boolean functions.* Cambridge University Press, 2014. 11

[RS10] P. Raghavendra and D. Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 755–764. ACM, 2010. 10

[RST10] P. Raghavendra, D. Steurer, and M. Tulsiani. Reductions between expansion problems. *arXiv preprint arXiv:1011.2586*, 2010. 10

[WS11] D. P. Williamson and D. B. Shmoys. *The design of approximation algorithms.* Cambridge University Press, 2011. 1