SOS Lecture 6: The SOS approach to refuting the UGC / Boaz Barak

Morally speaking, the Unique Games Conjecture (UGC) asserts that a simple algorithm— namely the degree 2 SOS program— is the *optimal efficient algorithm* for a wide range of optimization problems. Thus if the UGC is true one might expect that the SOS hierarchy is very often *useless*— there is no point in going beyond the degree two case unless you go for a large degree that would amount to the exponential-time brute force algorithm. We do not know whether the UGC as stated is true, but we will see that this strong version of it is false— we have already seen examples where we can get non-trivial improvements by the SOS algorithm with moderate degree, and as we will see today, we can get such guarantees even for the Unique Games problem itself. Namely, the SOS algorithm yields a sub-exponential (2^{n^ϵ}) time algorithm for the Unique Games problem. While falling short of disproving the UGC, this algorithm does significantly outperform the trivial brute-force algorithm.

The SOS hierarchy represents the most promising approach I know of towards refuting the UGC. Progress towards this goal has been somewhat slow, with papers that show the algorithm works for particular instances, or work for general instances but with parameters that are far from what's needed to refute the UGC. But it has also been steady, and an algorithm-optimistic (or complexity-pessimistic) view towards it is that the problem might eventually succumb by the Grothendieck "Rising Sea" method:

> I can illustrate the second approach [to solving a problem] with the same image of a nut to be opened. The first analogy that came to my mind is of immersing the nut in some softening liquid, and why not simply water? From time to time you rub so the liquid penetrates better, and otherwise you let time pass. The shell becomes more flexible through weeks and months— when the time is ripe, hand pressure is enough, the shell opens like a perfectly ripened avocado!
>
> A different image came to me a few weeks ago. The unknown thing to be known appeared to me as some stretch of earth or hard marl, resisting penetration. . . the sea advances insensibly in silence, nothing seems to happen, nothing moves, the water is so far off you hardly hear it. . . yet it finally surrounds the resistant substance.

Another, perhaps slightly less classy metaphor that comes to my mind is the Austin Powers steamroller, though at this point it's still unclear if the SOS steamroller will stop short of running over the UGC...

I should note that the SOS hierarchy also plays an important role in the most promising approach I know of to *prove* the conjecture. This is the work of Khot and Moshkovitz which Dana discussed in our reading group and I will refer to briefly today. I do believe in the "rising sea" approach in the sense that, true or false, settling the UGC will eventually be only a minor part of a much larger theory that will give us a broad understanding of the power of the SOS algorithm, and of efficient algorithms in general, for many classes of optimization problems.

1 The Small Set Expansion Hypothesis

There is actually a family of problems related to the Unique Games Conjecture. All these problems share the feature that the best known approximation algorithm for them is the degree 2 SOS and that we have no proof one can't do better. The three most prominent problems in this class are:

- $SSE(\epsilon)$: Distinguishing, given a d-regular graph $G = (V, E)$ of n vertices, between the YES case where there exists a set S of $n/\log n$ vertices such that $|E(S, \overline{S})| \leq \epsilon d|S|$ and the NO case where every set S of at most $n/\log n$ vertices satisfies $|E(S, \overline{S})| \geq (1-\epsilon)d|S|$.

- $UG(\epsilon)$: Distinguishing, given a set of linear equations over n variables taking values in $\mathbb{Z}_{\log n}$ such that each equation only involves two variables, between the YES case where there exists an assignment to the variables that satisfies a $1-\epsilon$ fraction of the equations, and the NO case where every assignment satisfies at most ϵ fraction of them.[1]

- $2LIN(\epsilon)$: Distinguishing, given a set of linear equations over n variables taking values in $\mathbb{Z}_2$ such that each equation only involves two variables, between the YES case where there exists an assignment to the variables that satisfies a $1-\epsilon$ fraction of the equations, and the NO case where every assignment satisfies a $1-\sqrt{\epsilon}/10$ fraction of them.

In all cases we think of ϵ as a small constant tending to zero (e.g., think $\epsilon = 0.01$ or $\epsilon = 0.001$). We have the following relation between these problems

$$SSE(\epsilon) \preceq UG(\epsilon) \preceq 2LIN(\epsilon)$$

by which we mean that there is a polynomial-time reduction from $SSE(\epsilon)$ to $UG(\epsilon')$ (for some ϵ' depending on ϵ and tending to 0 with ϵ) and from $UG(\epsilon)$ to $2LIN(\epsilon')$. Reductions in the other directions are not known, but current knowledge suggests that all three problems are likely to be computationally equivalent. In particular all known algorithmic and hardness results apply equally well to the $SSE(\epsilon)$ and $UG(\epsilon)$. The best polynomial-time algorithm known for all three problems is the degree 2 SOS algorithm. In particular for SSE this algorithm corresponds to a generalization of Cheeger's Inequality, while for $2LIN$ it corresponds to a (small) generalization of the Goemans-Williamson Max-Cut algorithm. The version of $2LIN(\epsilon)$ of distinguishing between value $1-\epsilon$ vs. $1-1.01\epsilon$ is known to be NP-hard (for some value of 1.01) and in fact this reduction has quasilinear blowup and so under standard assumptions this problem cannot be solved in $2^{n^{0.999}}$ time.

The *Unique Games Conjecture* (UGC) asserts that for every $\epsilon > 0$, $UG(\epsilon)$ is NP hard. The *Small Set Expansion Hypothesis* (SSEH) asserts that for every $\epsilon > 0$, $SSE(\epsilon)$ is NP hard. One can also phrase a *$2LIN$ hypothesis* (2LINH) that for every $\epsilon > 0$, $2LIN(\epsilon)$ is NP-hard. Given the discussion above, the SSEH implies the UGC and is very likely to be equivalent to it. Since it also implies all consequences of the UGC (including the 2LINH), the SSEH is a natural anchor for the problems in the "Unique Games Sphere" and so a natural object of study towards refuting the UGC. Conversely, the 2LINH is a natural object of study towards proving the UGC (which is indeed the Khot-Moshkovitz approach).

Note: All these problems are typically stated with more parameters than ϵ, and we fixed the other parameter to be $\log n$ in SSE and UG for simplicity. This version can be shown to be equivalent to the more standard version by known reductions.

2 2 to q norm and small set expansion

The main current approach to the attacking the small-set expansion problem via SOS goes through *hyper-contractive norms*. Specifically we will use the following result. Informally, we call a d-regular graph $G = (V, E)$ a *small set expander* if subsets S of size $o(|V|)$ satisfy $|E(S, \overline{S})| \geq (1 - o(1))d|S|$.

[1] For every $c < 1/2$, the version of the $UG(\epsilon)$ where we compare $1-\epsilon$ vs $1-\epsilon^c$ is known to be equivalent to $UG(\epsilon')$ (for ϵ' related to ϵ) using Rao's parallel repetition theorem.

Theorem 1 (Informal)**.** *For every even $q > 2$, a graph G is a* small set expander *if and only if every vector w in G's top eigenspace satisfies*

$$\mathbb{E}_i w_i^q \leq O(\mathbb{E} w_i^2)^{q/2} \tag{1}$$

Thus a sufficiently good approximation algorithm for (1) would yield an algorithm for the small set expansion problem. This approach seems to be extremely ambitious in the sense that we try to approximate the ratio of the q and 2 norms over an *arbitrary* subspace W, forgetting any additional structure W may have had since it is the top eigenspace of some graph. However, this makes the problem also cleaner and presumably, if there are hard instances for it, then it would be easier to find them.

Keeping to a rough, informal level, the proof that $O(1)$ levels of the SOS algorithm solve all previously known hard instances rely on the fact that it can certify (1) for the subspace of low degree polynomials over $\{\pm 1\}^{\log n}$. With Kelner and Steurer, we showed that $O(1)$ rounds of SOS yield an $\dim(W)^{1/3}$ approximation (in some precise sense) of the 2 to 4 problem. If this result could be improved to a constant or even *polylog*(n) (perhaps even $n^{o(1)}$), even at the expense of using *polylog*(n) (or perhaps even $n^{o(1)}$) rounds this would refute the SSEH and likely can be extended to refute the UGC as well.

We will now give an informal intuition behind the theorem, and sketch how it implies a sub-exponential algorithm for small-set expansion (that can be extended to the Unique Games problem as well), and then describe in full the $\dim(W)^{1/3}$ algorithm. (Depending on time, we may or may not cover the full proof of Theorem 1.)

2.1 The relation between the 2 to 4 norm and small set expansion

We now give some intuition on the relation between the 2 to q norm and small set expansion. For simplicity, we focus on the case $q = 4$ although it is not hard to see that the same intuition holds for every even $q > 2$. In the second lecture we saw the following results that says that if $\mathbb{E} w_i^4 \leq O(\mathbb{E} w_i^2)^2$ for every w in the top eigenspace of G then G is a small set expander:

Lemma 2. *Let $G = (V, E)$ be regular graph, $\lambda \in (0, 1)$ and W be the span of eigenvectors of G's normalized adjacency matrix corresponding to eigenvalue at least $1 - \lambda$. If every $w \in W$ satisfies:*

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

then for every set S of measure δ set,

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

The proof of the lemma (which we saw) is a fairly straightforward contrapositive argument— if S is a set of $o(1)$ measure that doesn't expand, then the projection of 1_S to the top eigenspace will still have large 4 norm compared to its 2 norm.

The other direction— transforming a vector w in the top eigenspace with large 4 to 2 norm ratio into a small set that doesn't expand— is trickier. It is instructive to compare this with Cheeger's Inequality. The difficult direction of Cheeger's Inequality transforms a vector w in the top eigenspace (but orthogonal to the all 1's vector) into a set of measure at most 1/2 that does not expand. In fact, Cheeger doesn't need the vector w to be an eigenvector at all. As long as

$$w^\top G w \geq (1 - \epsilon)\|w\|^2 \tag{3}$$

the transformation of Cheeger (which involves choosing a random threshold τ with probability proportional to τ^2 and taking the set of all coordinates of w that are larger than τ) will yield such a set.

One could hope that if w satisfies $\mathbb{E}w_i^4 \gg (\mathbb{E}w_i^2)^2$ then by using the same or a similar transformation we can get a *small* set that doesn't expand. However, this is a bit tricky— in particular we cannot do so by only assuming (3), since it is trivial to modify every vector w to have high 4 norm without hurting (3) too much. For example if w satisfies $\mathbb{E}w_i^2 = 1$, we can add $n^{0.3}e_1$ to w. This change will make the 4-norm of w huge, but will be negligible in the 2 norm and hence will not hurt (3). Therefore, to make the proof go through we must use the fact that w is completely contained in the top eigenspace, as opposed to merely satisfying (3).

Intuition for the actual proof. To get some intuition for the proof, lets assume that the graph G is "nice" in the following sense: for every vector w, if w is in the eigenspace of G corresponding to eigenvalues larger than $1-\epsilon$ then the vector $w^{\odot q}$, defined as $w_i^{\odot q} = w_i^q$, is in the eigenspace corresponding to eigenvalues larger than $1 - q\epsilon$. For example, Cayley graphs over the Boolean cube are "nice":

Exercise 1: Prove that if G is a Cayley graph over $GF(2)^\ell$ (i.e., G's vertices are elements of $GF(2)^\ell$ and x is connected to y if $x \oplus y \in S$ for some subset $S \subseteq GF(2)^\ell$) then it is nice. See footnote for hint[2] Can you generalize this to other Cayley graphs?

Now this means that if there is a vector w in G's top eigenspace satisfying $\mathbb{E}w_i^4 \gg (\mathbb{E}w_i^2)^2$ then the vector $v = w^{\odot 2}$ is also in G's top eigenspace (for a slightly looser definition of "top") and satisfies $\mathbb{E}v_i^2 \gg (\mathbb{E}|v_i|)^2$. However it turns out that the Cheeger transformation does actually produce a set of measure at most $O(\delta)$ if you apply it to a vector v satisfying $\delta\mathbb{E}v_i^2 > (\mathbb{E}|v_i|)^2$. Indeed, without loss of generality we can normalize so that $\sum |v_i| = 1$, and so we can think of $|v_i|$ as a probability distirbution and this condition means that it has collision probability at least $1/(\delta n)$. Let S be the set of i's such that $|v_i| > 1/(100\delta n)$. Note that $|S| \leq 100\delta n$. Now one can show that if we do the Cheeger transformation then with high probability we will output a subset of S. (For a formal argument see the paper of Dimitriou and Impagliazzo (1998) or the appendix of Arora,Barak Steurer (2010).)

The argument for "non nice" graphs is substantially more complicated. Specifically, rather than giving a simple transformation that takes any w satisfying the conditions into a set S that does not expand, the argument needs to assume that w is (close to) the *optimal* vector in the subspace in terms of the relation between its q and 2 norm. The full proof is enclosed below and we will cover it in class based on time constraints.

Exercise 2: (Open) Find a simpler proof for Theorem 1

Theorem 1 immediately implies a *sub-exponential time* algorithm for small set expansion using the following exercise (a version of which for $q = 4$ we've already seen in a previous lecture):

Exercise 3: Let q be any even constant and W be a subspace of R^n with dimension $\gg n^{2/q}$. Then there exists $w \in W$ such that $(\mathbb{E}w_i^q) \gg (\mathbb{E}w_i^2)^2$. See footnote for hint[3]

This yields a sub-exponential algorithm since for $SSE(\epsilon)$ we can take consider the subspace corresponding to eigenvalues larger than $1-\epsilon$ and so q to be roughly $1/\epsilon$. This means that

[2] **Hint:** The eigenvectors for such a graph are always the functions $\{\chi_\alpha\}_{\alpha \in GF(2)^\ell}$ where $\chi_\alpha(x) = -1^{\langle x,\alpha \rangle}$. Note that $\chi_\alpha \odot \chi_\beta = \chi_{\alpha\oplus\beta}$.

[3] **Hint:** Given an orthonormal basis $w_1, \ldots, w_d$ for W, we want to find some signs $\sigma_1, \ldots, \sigma_d \in \{\pm 1\}$ so that some coordinate i of the vector $w = \sum \sigma_i w_i$ will satisfy $|w_i| \geq \Omega(d/\sqrt{n})$ which would imply $\mathbb{E}w_i^q \geq d^q/n^{q/2+1}$, while of course $\|w\|^2 = d$ and so $\mathbb{E}w_i^2 = d/n$.

either G's top eigenspace has dimension at most $O(n^{2/q})$, in which case we can enumerate over (a sufficiently fine net of) it in $\exp(O(n^{2\epsilon}))$ time, or if the dimension of the subspace is higher and then we know that it can't be a small set expander by the combination of this exercise and Theorem 1. This algorithm can be extended for the Unique Games problem as well (see Arora, Barak and Steurer, 2010). Note that this means that if we take $\epsilon = 0.01$, we would need to look at graphs of size roughly 2^{50} before this algorithm is slower than a quadratic time one (for $\epsilon = 0.001$ this would be 2^{500}). So, even if the SSEH/UGC are true, they do not seem to tell us very much on inputs that actually fit in the world's storage capacity. This is in contrast to problems such as SAT where, despite progress in SAT solvers, its exponential behavior is in fact quite observable even on relatively modest sized inputs of a few thousand variables or so (not to mention variants of SAT arising from private-key cryptography, where we can see the exponential behaviour even on inputs as small as a few dozen variables— e.g. we still don't know of a much better than brute force algorithm to break the 56-bit cipher DES.) That said, even if the SSEH hypothesis is false, it would be still very interesting to know if a *linear time* algorithm exists for the $SSE(\epsilon)$ problem.

3 Using SOS for the 2 to 4 problem

In a previous lecture we saw that SOS can certify that the span of low degree polynomials over the Bolean cube has bounded 4 to 2 norm ratio. We did not show how this implies that the SOS algorithm solves the UG/SSE/Max-Cut problems on previously suggested candidate hard instances (see our STOC 2012 paper for that), but in any case this work was only for *specific* instances. We now describe an approach for *general* instances of the problem.

A variant of Theorem 1 shows that for any constants $\beta > \alpha$ and $\delta > 0$, an algorithm for the following problem is sufficient to solve $SSE(\epsilon)$ (with ϵ related to the parameters below— in the application we will let W be the subspace corresponding to eigenvalues of G larger than $1 - O(\epsilon)$): given a subspace $W \subseteq \mathbb{R}^n$, distinguish between the YES case when there is a set S with $|S| \leq n/\beta$ such that the projection of 1_S to W has norm $(1-\delta)\|1_S\|$ and the NO case where for every $w \in W$ such that $\mathbb{E}w_i^4 \leq \alpha(\mathbb{E}w_i^2)^2$. For simplicity lets consider the version with $\delta = 0$. (This seems potentially easier, but we don't know of any better algorithm than the one for $\delta > 0$.)

Therefore we consider the following problem: given a d-dimensional subspace $W \subseteq \mathbb{R}^n$, distinguish between the YES case: there is a set S with $|S| \leq n/\beta$ such that $1_S \in W$ and the NO case: $\mathbb{E}w_i^4 \leq \alpha(\mathbb{E}w_i^2)^2$ for every $w \in W$. We identify the approximation factor of this problem with β/α. Ideally we would like an algorithm solving the problem for $O(1)$ approximation factor, but what we would show is an $O(d^{1/3})$ approximation algorithm (taken from the work with Kelner and Steurer which also handled the case of $\delta > 0$).

Theorem 3. *There is some constant c such that if $\beta \geq cd^{1/3}\alpha$, then if there exists a degree 20 psuedo-distribution $\{w\}$ over $\mathbb{R}^n$ satisfying the constraints $w \in W$, $\|w\|^2 = 1$, $w_i^2 = w_i$ for all i, $\mathbb{E}_i w_i^4 \geq \beta(\mathbb{E}w_i^2)^2$, then there exists some $v \in W$ satisfying $\mathbb{E}v_i^4 \geq \alpha(\mathbb{E}v_i^2)^2$. Moreover, we can efficiently find such a v from the moments of $\{w\}$.*

Proof. Let Π be the projector to W and let $\delta^i = \Pi e^i$ where e_i is the i^{th} standard basis vector.

The algorithm will use the combination of the following steps:

Random vector rounding: pick a random $w \in W$.

Projection rounding: try all vectors of the form δ^i.

Conditioning: find $i_1, \ldots, i_4$ and change $\{w\}$ to the distribution where we re-weigh the probability of every vector by a factor of $w_{i_1}^2 \cdots w_{i_4}^4 = w_{i_1} \cdots w_{i_4}$.

Quadratic sampling: sample a gaussian distribution $\{v\}$ that matches the first two moments of $\{w\}$.

We will show that if the first two methods fail, then after conditioning, quadratic sampling will yield a good solution.

If there exists an i such that $\mathbb{E}_j(\delta_j^i)^4 \geq \alpha(\mathbb{E}_j(\delta_j^i)^2)^2$ then we're done, so we can assume that $\mathbb{E}_j(\delta_j^i)^4 \leq \alpha(\mathbb{E}_j(\delta_j^i)^2)^2$ for every $i \in [n]$.

Note that $w_i = \langle w, e^i \rangle = \langle w, \delta^i \rangle$ for every $w \in W$.

Note that since the w's are characteristic vectors of sets of size n/β, $\mathbb{E}_i w_i^4 = 1/\beta$. Thus, by Cauchy-Schwarz

$$\tfrac{1}{\beta} = \tilde{\mathbb{E}}_w \mathbb{E}_i \langle w, \delta^i \rangle^4 \leq \sqrt{\tilde{\mathbb{E}} \langle w, w' \rangle^4 \mathbb{E}_{i,j} \langle \delta^i, \delta^j \rangle^4} \tag{4}$$

(**Exercise 4:** verify this.)

Under our assumption for every i,

$$\mathbb{E}_j \langle \delta^i, \delta^j \rangle^4 = \mathbb{E}_j(\delta_j^i)^4 \leq \alpha \left(\mathbb{E}(\delta_j^i)^2 \right)^2$$

Lets assume the RHS is the same up to a factor of α for every j (**Exercise 5:** show that otherwise random vector rounding succeeds). Then we get that

$$\mathbb{E}_i \mathbb{E}_j \langle \delta^i, \delta^j \rangle^4 \leq \alpha^2 \left(\mathbb{E}_{i,j}(\delta_j^i)^2 \right)^2$$

but $\delta_j^i = \langle e^j, \Pi e^i \rangle$ and so $\mathbb{E}_{i,j}(\delta_j^i)^2$ is simply $1/n^2$ times the Frobenius norm squared of Π which is d. Hence we get that

$$\mathbb{E}_i \mathbb{E}_j \langle \delta^i, \delta^j \rangle^4 \leq \alpha^2 d^2 / n^4$$

Plugging this into (4), squaring and dividing both sides by $\alpha^2 d^2 / n^4$ we get that

$$\tilde{\mathbb{E}} \langle w, w' \rangle^4 \geq \tfrac{n^4}{\alpha^2 \beta^2 d^2}$$

since $d = \beta^3 / (c^3 \alpha^3)$ we get

$$\tilde{\mathbb{E}} \langle w, w' \rangle^4 \geq \tfrac{n^4 c^6 \alpha^4}{\beta^8}$$

We now make the following claim, which crucially depends on the fact that $w_i^2 = w_i$ for all i:

Claim: There exists $i_1, \ldots, i_4$ such that if we modify the distribution $\{w\}$ by multiplying the probability of every w with $w_{i_1}^2 \cdots w_{i_4}^2$ then

$$\tilde{\mathbb{E}}_{new} \langle w, w' \rangle \geq \left(\tilde{\mathbb{E}}_{old} \langle w, w' \rangle^4 \right)^{1/4}$$

Exercise 6: prove this claim

The claim implies that

$$\tilde{\mathbb{E}} \langle w, w' \rangle \geq \tfrac{nc\alpha}{\beta^2}$$

Now if we pick v to be a random vector matching the first two moments of $\{w\}$, then we claim that $\mathbb{E}_i v_i^4 \geq \frac{c\alpha}{\beta^2}$ (**Exercise 7:** verify this.) but on the other hand $\mathbb{E} v_i^2 = \mathbb{E} w_i^2 = 1/\beta$, hence concluding the proof. $\square$

Theorem 3 and the Khot-Moshkovitz construction. The factor $d^{1/3}$ seems rather arbitrary and it is natural to ask whether by using more rounds we can improve it further, perhaps getting a factor of $d^{\Omega(1/r)}$ for degree-r SOS. This should be sufficient to refute the SSEH and quite possible extended to refute the UGC and 2LINH as well.

In contrast, Khot and Moshkovitz gave a candidate integrality gap for the $2LIN$ problem. Specifically for arbitrarily large constants c, r they construct an instance I of $2LIN$ (mod 2) for which there is a degree-r pseudo-distribution consistent with satisfying $1 - \epsilon$ fraction of I's constraint, but they conjecture that in fact one cannot satisfy more than $1 - c\epsilon$ fraction of I's constraints. Lets call this conjecture the Khot-Moshkovitz conjecture. Even if true, the Khot-Moshkovitz conjecture does not contradict the conjecture that r-rounds of SOS yield an $n^{O(1/r)}$ approximation algorithm for the SSE,UG and 2LIN problems, since one would need to take $r = \Omega(\log n)$ for the latter algorithm to reach the range of parameters of the UGC and its ilk.

However, the Khot-Moshkovitz construction is only "step zero" in their plan to eventually prove the UGC/SSEH and hence contradict the conjecture that r-SOS rounds yield an $n^{O(1/r)}$ approximation for this problem. Specifically, to obtain a proof of the UGC, one would need to improve the KM paper in the following aspects: Step 1 would be to prove the Khot-Moshkovitz conjecture that no assignment can satisfy more than a $1 - c\epsilon$ fraction of the equations in their instance. Step 2 would be to improve the gap from $1 - \epsilon$ vs $1 - c\epsilon$ to $1 - \epsilon$ vs $1 - \Omega(\sqrt{\epsilon})$ and improve the number of rounds from a constant to $n^{\Omega(1)}$. Step 3 is to extend the result from $2LIN(\epsilon)$ to $UG(\epsilon)$ (and maybe also to $SSE(\epsilon)$). Step 4 would be to to extend the result from a lower bound on SOS to an NP-hardness proof. In my (personal and quite possibly wrong) opinion, the significance of these hurdles is in the order listed. In particular, if the first and second step are completed then this would be sufficient to rule out what seems to be the most natural scenario under the assumption that the UGC is false— that SOS solves $SSE(\epsilon)$, $UG(\epsilon)$ and $2LIN(\epsilon)$ in polynomial or quasipolynomial time, and I believe that achieving Steps 3 and 4 in this case should not be that much harder. Thus, despite the fact that Steps 3 and 4 seem more *qualitative* in nature than Step 2, I actually view Steps 1 and 2 as the most significant hurdles to overcome (or the more likely to be false if the UGC is false).

4 Formal statement and proof of Theorem 1

In this section (which is more or less copied from (Barak, Brandao, Harrow, Kelner, Steurer, and Zhou 2012) we show that a graph is a *small-set expander* if and only if the projector to the subspace of its adjacency matrix's top eigenvalues has a bounded $2 \to q$ norm for even $q \geq 4$. While the "if" part was known before, the "only if" part is novel. This characterization of small-set expanders is of general interest, and also leads to a reduction from the SSE problem to the problem of obtaining a good approximation for the $2 \to q$ norms. For simplicity of notation, throughout this section we use *expectation norms* — i.e. $\|w\|_p = (\mathbb{E}_i |w_i|^p)^{1/p}$.

Notation For a regular graph $G = (V, E)$ and a subset $S \subseteq V$, we define the *measure* of S to be $\mu(S) = |S|/|V|$ and we define $G(S)$ to be the distribution obtained by picking a random $x \in S$ and then outputting a random neighbor y of x. We define the *expansion* of S, to be $\Phi_G(S) = \Pr_{y \in G(S)}[y \notin S]$, where y is a random neighbor of x. For $\delta \in (0, 1)$, we define $\Phi_G(\delta) = \min_{S \subseteq V : \mu(S) \leq \delta} \Phi_G(S)$. We often drop the subscript G from Φ_G when it is clear from context. We identify G with its normalized adjacency (i.e., random walk) matrix. For every $\lambda \in [-1, 1]$, we denote by $V_{\geq \lambda}(G)$ the subspace spanned by the eigenvectors of G with eigenvalue at least λ. The projector into this subspace is denoted $P_{\geq\lambda}(G)$. For a distribution D, we let $\mathrm{cp}(D)$ denote the collision probability of D (the probability that two independent samples from D are identical).

Our main theorem of this section is the following:

Theorem 4. *For every regular graph G, $\lambda > 0$ and even q,*

1. (Norm bound implies expansion) *For all $\delta > 0, \epsilon > 0$, $\|P_{\geq\lambda}(G)\|_{2\to q} \leq \epsilon/\delta^{(q-2)/2q}$ implies that $\Phi_G(\delta) \geq 1 - \lambda - \epsilon^2$.*
2. (Expansion implies norm bound) *There is a constant c such that for all $\delta > 0$, $\Phi_G(\delta) > 1 - \lambda 2^{-cq}$ implies $\|P_{\geq\lambda}(G)\|_{2\to q} \leq 2/\sqrt{\delta}$.*

One corollary of Theorem 4 is that a good approximation to the $2 \to q$ norm implies an approximation of $\Phi_\delta(G)$.

Corollary 5. *If there is a polynomial-time computable relaxation $\mathcal{R}$ yielding good approximation for the $2 \to q$, then the* Small-Set Expansion Hypothesis *is false.*

(Note: here I use the standard notion of the SSEH with δ being an arbitrary small constant as opposed to equaling $1/\log n$; I didn't verify that the Raghvandra-Steurer-Tulsiani reduction can be used for the setting of $\delta = 1/\log n$ as well.)

Proof. Using (Raghavendra, Steurer, Tulsiani), to refute the small-set expansion hypothesis it is enough to come up with an efficient algorithm that given an input graph G and sufficiently small $\delta > 0$, can distinguish between the *Yes* case: $\Phi_G(\delta) < 0.1$ and the *No* case $\Phi_G(\delta') > 1 - 2^{-c\log(1/\delta')}$ for any $\delta' \geq \delta$ and some constant c. In particular for all $\eta > 0$ and constant d, if δ is small enough then in the *No* case $\Phi_G(\delta^{0.4}) > 1 - \eta$. Using the first part of Theorem 4, in the *Yes* case we know $\|V_{1/2}(G)\|_{2\to q} \geq 1/(10\delta^{1/4})$, while in the *No* case, we can choose η to be sufficiently small so that the condition $\Phi_G(\delta^{0.2}) \geq 1 - \eta$ implies (via the second part of Theorem 5) that $\|V_{1/2}(G)\|_{2\to q} \leq 2/\delta^{0.1}$. Thus an $O(\delta^{-0.15})$ approximation for the $2 \to q$ norm will refute the SSEH. □

The first part of Theorem 4 follows from previous work (e.g., see [**?**]). For completeness, we include a proof in Appendix **??**. The second part will follow from the following lemma:

Lemma 6. *Set $e = e(\lambda, q) := 2^{cq}/\lambda$, with a constant $c \leq 100$. Then for every $\lambda > 0$ and $1 \geq \delta \geq 0$, if G is a graph that satisfies $\mathrm{cp}(G(S)) \leq 1/(e|S|)$ for all S with $\mu(S) \leq \delta$, then $\|f\|_q \leq 2\|f\|_2/\sqrt{\delta}$ for all $f \in V_{\geq\lambda}(G)$.*

Proving the second part of Theorem 4 from Lemma 6 We use the variant of the local Cheeger bound obtained in [**?**, Theorem 2.1], stating that if $\Phi_G(\delta) \geq 1-\eta$ then for every $f \in Ł2(V)$ satisfying $\|f\|_1^2 \leq \delta\|f\|_2^2$, $\|Gf\|_2^2 \leq c\sqrt{\eta}\|f\|_2^2$. The proof follows by noting that for every set S, if f is the characteristic function of S then $\|f\|_1 = \|f\|_2^2 = \mu(S)$, and $\mathrm{cp}(G(S)) = \|Gf\|_2^2/(\mu(S)|S|)$. □

Proof of Lemma 6. Fix $\lambda > 0$. We assume that the graph satisfies the condition of the Lemma with $e = 2^{cq}/\lambda$, for a constant c that we'll set later. Let $G = (V, E)$ be such a graph, and f be function in $V_{\geq\lambda}(G)$ with $\|f\|_2 = 1$ that maximizes $\|f\|_q$. We write $f = \sum_{i=1}^m \alpha_i\chi_i$ where $\chi_1, \ldots, \chi_m$ denote the eigenfunctions of G with values $\lambda_1, \ldots, \lambda_m$ that are at least λ. Assume towards a contradiction that $\|f\|_q > 2/\sqrt{\delta}$. We'll prove that $g = \sum_{i=1}^m(\alpha_i/\lambda_i)\chi_i$ satisfies $\|g\|_q \geq 5\|f\|_q/\lambda$. This is a contradiction since (using $\lambda_i \in [\lambda, 1]$) $\|g\|_2 \leq \|f\|_2/\lambda$, and we assumed f is a function in $V_{\geq\lambda}(G)$ with a maximal ratio of $\|f\|_q/\|f\|_2$.

Let $U \subseteq V$ be the set of vertices such that $|f(x)| \geq 1/\sqrt{\delta}$ for all $x \in U$. Using Markov and the fact that $\mathbb{E}_{x\in V}[f(x)^2] = 1$, we know that $\mu(U) = |U|/|V| \leq \delta$, meaning that under our assumptions any subset $S \subseteq U$ satisfies $\mathrm{cp}(G(S)) \leq 1/(e|S|)$. On the other hand, because $\|f\|_q^q \geq 2^q/\delta^{q/2}$, we know that U contributes at least half (in fact $1 - 2^{-q}$) of the term $\|f\|_q^q = \mathbb{E}_{x\in V} f(x)^q$. That is, if we define α to be $\mu(U)\mathbb{E}_{x\in U} f(x)^q$ then $\alpha \geq \|f\|_q^q/2$. We'll prove the lemma by showing that $\|g\|_q^q \geq 10\alpha/\lambda$.

Let c be a sufficiently large constant ($c = 100$ will do). We define U_i to be the set $\{x \in U : f(x) \in [c^i/\sqrt{\delta}, c^{i+1}/\sqrt{\delta})\}$, and let I be the maximal i such that U_i is non-empty. Thus, the sets $U_0, \ldots, U_I$ form a partition of U (where some of these sets may be empty). We let α_i be the contribution of U_i to α. That is, $\alpha_i = \mu_i \mathbb{E}_{x\in U_i} f(x)^q$, where $\mu_i = \mu(U_i)$. Note that $\alpha = \alpha_0 + \cdots + \alpha_I$. We'll show that there are some indices $i_1, \ldots, i_J$ such that:

(i) $\alpha_{i_1} + \cdots + \alpha_{i_J} \geq \alpha/(2c^{10})$.

(ii) For all $j \in [J]$, there is a non-negative function $g_j : V \to \mathbb{R}$ such that $\mathbb{E}_{x\in V} g_j(x)^q \geq e\alpha_{i_j}/(10c^2)^{q/2}$.

(iii) For every $x \in V$, $g_1(x) + \cdots + g_J(x) \leq |g(x)|$.

Showing these will complete the proof, since it is easy to see that for two non-negative functions and even q, g', g'', $\mathbb{E}(g'(x) + g''(x))^q \geq \mathbb{E}g'(x)^q + \mathbb{E}g''(x)^q$, and hence **(ii)** and **(iii)** imply that

$$\|g\|_4^4 = \mathbb{E}g(x)^4 \geq (e/(10c^2)^{q/2}) \sum_j \alpha_{i_j} . \tag{5}$$

Using **(i)** we conclude that for $e \geq (10c)^q/\lambda$, the right-hand side of (5) will be larger than $10\alpha/\lambda$.

We find the indices $i_1, \ldots, i_J$ iteratively. We let $\mathcal{I}$ be initially the set $\{0..I\}$ of all indices. For $j = 1, 2, ...$ we do the following as long as $\mathcal{I}$ is not empty:

1. Let i_j be the largest index in $\mathcal{I}$.
2. Remove from $\mathcal{I}$ every index i such that $\alpha_i \leq c^{10}\alpha_{i_j}/2^{i-i_j}$.

We let J denote the step when we stop. Note that our indices $i_1, \ldots, i_J$ are sorted in descending order. For every step j, the total of the α_i's for all indices we removed is less than $c^{10}\alpha_{i_j}$ and hence we satisfy **(i)**. The crux of our argument will be to show **(ii)** and **(iii)**. They will follow from the following claim:

Claim 7. *Let $S \subseteq V$ and $\beta > 0$ be such that $|S| \leq \delta$ and $|f(x)| \geq \beta$ for all $x \in S$. Then there is a set T of size at least $e|S|$ such that $\mathbb{E}_{x\in T} g(x)^2 \geq \beta^2/4$.*

The claim will follow from the following lemma:

Lemma 8. *Let D be a distribution with $\mathrm{cp}(D) \leq 1/N$ and g be some function. Then there is a set T of size N such that $\mathbb{E}_{x\in T} g(x)^2 \geq (\mathbb{E}g(D))^2/4$.*

Proof. Identify the support of D with the set $[M]$ for some M, we let p_i denote the probability that D outputs i, and sort the p_i's such that $p_1 \geq p_2 \cdots p_M$. We let β' denote $\mathbb{E}g(D)$; that is, $\beta' = \sum_{i=1}^M p_i g(i)$. We separate to two cases. If $\sum_{i>N} p_i g(i) \geq \beta'/2$, we define the distribution D' as follows: we set $\Pr[D' = i]$ to be p_i for $i > N$, and we let all $i \leq N$ be equiprobable (that is be output with probability $(\sum_{i=1}^N p_i)/N$). Clearly, $\mathbb{E}|g(D')| \geq \sum_{i>N} p_i g(i) \geq \beta'/2$, but on the other hand, since the maximum probability of any element in D' is at most $1/N$, it can be expressed as a convex combination of flat distributions over sets of size N, implying that one of these sets T satisfies $\mathbb{E}_{x\in T}|g(x)| \geq \beta'/2$, and hence $\mathbb{E}_{x\in T}g(x)^2 \geq \beta'^2/4$.

The other case is that $\sum_{i=1}^N p_i g(i) \geq \beta'/2$. In this case we use Cauchy-Schwarz and argue that

$$\beta'^2/4 \leq \left(\sum_{i=1}^N p_i^2\right)\left(\sum_{i=1}^N g(i)^2\right) . \tag{6}$$

But using our bound on the collision probability, the right-hand side of (6) is upper bounded by $\frac{1}{N}\sum_{i=1}^N g(i)^2 = \mathbb{E}_{x\in[N]}g(x)^2$. □

Proof of Claim 7 from Lemma 8. By construction $f = Gg$, and hence we know that for every x, $f(x) = \mathbb{E}_{y\sim x}g(y)$. This means that if we let D be the distribution $G(S)$ then

$$\mathbb{E}|g(D)| = \mathbb{E}_{x\in S}\mathbb{E}_{y\sim x}|g(y)| \geq \mathbb{E}_{x\in S}|\mathbb{E}_{y\sim x}g(y)| = \mathbb{E}_{x\in S}|f(x)| \geq \beta .$$

By the expansion property of G, $\mathrm{cp}(D) \leq 1/(e|S|)$ and thus by Lemma 8 there is a set T of size $e|S|$ satisfying $\mathbb{E}_{x\in T}g(x)^2 \geq \beta^2/4$. □

We will construct the functions $g_1, \ldots, g_J$ by applying iteratively Claim 7. We do the following for $j = 1, \ldots, J$:

1. Let T_j be the set of size $e|U_{i_j}|$ that is obtained by applying Claim 7 to the function f and the set U_{i_j}. Note that $\mathbb{E}_{x\in T_j}g(x)^2 \geq \beta_{i_j}^2/4$, where we let $\beta_i = c^i/\sqrt{\delta}$ (and hence for every $x \in U_i$, $\beta_i \leq |f(x)| \leq c\beta_i$).

2. Let g'_j be the function on input x that outputs $\gamma\cdot|g(x)|$ if $x \in T_j$ and 0 otherwise, where $\gamma \leq 1$ is a scaling factor that ensures that $\mathbb{E}_{x\in T_j}g'(x)^2$ equals exactly $\beta_{i_j}^2/4$.

3. We define $g_j(x) = \max\{0, g'_j(x) - \sum_{k<j} g_k(x)\}$.

Note that the second step ensures that $g'_j(x) \leq |g(x)|$, while the third step ensures that $g_1(x) + \cdots + g_j(x) \leq g'_j(x)$ for all j, and in particular $g_1(x) + \cdots + g_J(x) \leq |g(x)|$. Hence the only thing left to prove is the following:

Claim 9. $\mathbb{E}_{x\in V}g_j(x)^q \geq e\alpha_{i_j}/(10c)^{q/2}$

Proof. Recall that for every i, $\alpha_i = \mu_i\mathbb{E}_{x\in U_i}f(x)^q$, and hence (using $f(x) \in [\beta_i, c\beta_i)$ for $x \in U_i$):

$$\mu_i\beta_i^q \leq \alpha_i \leq \mu_i c^q\beta_i^q . \tag{7}$$

Now fix $T = T_j$. Since $\mathbb{E}_{x\in V}g_j(x)^q$ is at least (in fact equal) $\mu(T)\mathbb{E}_{x\in T}g_j(x)^q$ and $\mu(T) = e\mu(U_{i_j})$, we can use (7) and $\mathbb{E}_{x\in T}g_j(x)^q \geq (E_{x\in T}g_j(x)^2)^{q/2}$, to reduce proving the claim to showing the following:

$$\mathbb{E}_{x\in T}g_j(x)^2 \geq (c\beta_{i_j})^2/(10c^2) = \beta_{i_j}^2/10 . \tag{8}$$

We know that $\mathbb{E}_{x\in T}g'_j(x)^2 = \beta_{i_j}^2/4$. We claim that (8) will follow by showing that for every $k < j$,

$$\mathbb{E}_{x\in T}g'_k(x)^2 \leq 100^{-i'}\cdot\beta_{i_j}^2/4 , \tag{9}$$

where $i' = i_k - i_j$. (Note that $i' > 0$ since in our construction the indices $i_1, \ldots, i_J$ are sorted in descending order.)

Indeed, (9) means that if we let momentarily $\|g_j\|$ denote $\sqrt{\mathbb{E}_{x\in T} g_j(x)^2}$ then

$$\|g_j\| \geq \|g'_j\| - \| \textstyle\sum_{k<j} g_k\| \geq \|g'_j\| - \displaystyle\sum_{k<j} \|g_k\| \geq \|g'_j\|(1 - \sum_{i'=1}^{\infty} 10^{-i'}) \geq 0.8\|g'_j\| \ . \tag{10}$$

The first inequality holds because we can write g_j as $g'_j - h_j$, where $h_j = \min\{g'_j, \sum_{k<j} g_k\}$. Then, on the one hand, $\|g_j\| \geq \|g'_j\| - \|h_j\|$, and on the other hand, $\|h_j\| \leq \|\sum_{k<j} g_k\|$ since $g'_j \geq 0$. The second inequality holds because $\|g_k\| \leq \|g'_k\|$. By squaring (10) and plugging in the value of $\|g'_j\|^2$ we get (8).

Proof of (9) By our construction, it must hold that

$$c^{10}\alpha_{i_k}/2^{i'} \leq \alpha_{i_j} \ , \tag{11}$$

since otherwise the index i_j would have been removed from the $\mathcal{I}$ at the k^{th} step. Since $\beta_{i_k} = \beta_{i_j} c^{i'}$, we can plug (7) in (11) to get

$$\mu_{i_k} c^{10+4i'}/2^{i'} \leq c^4 \mu_{i_j}$$

or

$$\mu_{i_k} \leq \mu_{i_j} (2/c)^{4i'} c^{-6} \ .$$

Since $|T_i| = e|U_i|$ for all i, it follows that $|T_k|/|T| \leq (2/c)^{4i'} c^{-6}$. On the other hand, we know that $\mathbb{E}_{x\in T_k} g'_k(x)^2 = \beta_{i_k}^2/4 = c^{2i'}\beta_{i_j}^2/4$. Thus,

$$\mathbb{E}_{x\in T} g'_k(x)^2 \leq 2^{4i'} c^{2i'-4i'-6} \beta_{i_j}^2/4 \leq (2^4/c^2)^{i'} \beta_{i_j}^2/4 \ ,$$

and now we just choose c sufficiently large so that $c^2/2^4 > 100$. □

□