Semidefinite programming and computational aspects of entanglement **IHP Fall 2017**

Lecture 5: November 23

Lecturer: Anand Natarajan *Scribe: Aram and Anand*

Rest of this course is about SoS. In the remaining lectures, Aram will tell you about how to prove positive results on its performance (de Finetti theorems). In this lecture, you will see instances where it fails.

These instances will be based on CSPs. A special feature of CSPs: over n binary variables, degree-n SoS always converges exactly. This is because all polynomials can be reduced to degree $\leq n$ over Boolean variables. Or in the dual picture, deg-k SoS corresponds to searching over marginal probability distributions on subsets of k variables, so when $k = n$ this means searching over the full distribution on n variables.

5.1 Integrality Gaps for SoS

For constant degree, can cook up examples (e.g. MAX-CUT for degree 2) where SoS gives the wrong answer (although this may still be a good approximation). This is called an "integrality gap" following terminology from LP relaxations of integer programs. (If you're not familiar with this, then don't worry about where the term comes from.)

5.1.1 Grigoriev

Theorem 1. *For every $\epsilon > 0$ and every n there exists a 3XOR instance with $\Theta(n/\epsilon^2)$ questions s.t. $\mathrm{val}(\phi) \leq 1/2 + \epsilon$ but $\mathrm{val}_d(\phi) = 1$ for all $d < d^*$, where $d^* = \Omega(n)$.*

Proof. Pick a random instance. By Chernoff bound, for each fixed assignment x, the probability (over choice of instance) that x satisfies the instance is exponentially small, specifically let's say $\leq 2^{-2n}$. So by union bound, the probability that a random instance has no satisfying assignment can be made exponentially close to 1.

Next we need to argue that SoS estimates value 1 for this instance. Recall that the level-d SoS SDP is

$$\max_{\tilde{\mathbb{E}}} \tilde{\mathbb{E}}[\phi(x)] \tag{5.1a}$$

$$\text{s.t. } \tilde{\mathbb{E}}[p(x)(x_i^2 - 1)] = 0 \qquad \forall p \tag{5.1b}$$

$$\mathbb{E}[q^2(x)] \geq 0 \qquad \forall q(x) \tag{5.1c}$$

Saying that $\tilde{\mathbb{E}}[\phi(x)] = 1$ is equivalent to saying that for each clause $x_{i_1}x_{i_2}x_{i_3} = s_i$ we have

$$\tilde{\mathbb{E}}[x_{i_1}x_{i_2}x_{i_3}] = s_i. \tag{5.2}$$

We now use (5.2) to recursively build up assignments of $\tilde{\mathbb{E}}[x_S]$ for all $|S| \leq d$. Start with $\tilde{\mathbb{E}}[1] = 1$, with $\tilde{\mathbb{E}}[x_{i_1}x_{i_2}x_{i_3}] = s_i$ for each clause and with $\tilde{\mathbb{E}}[x_S]$ undefined other choices of S. (It suffices for now to consider only the action of $\tilde{\mathbb{E}}$ on monomials.) Then we repeatedly apply the following procedure. Choose $S, T \subseteq [n]$ s.t. $\tilde{\mathbb{E}}[x_S], \tilde{\mathbb{E}}[x_T]$ are both defined and such that $|S \oplus T| \leq d$. Here $\oplus$ means the symmetric difference. This is equivalent to demanding that $\deg(x_S x_T \mod \langle x_1^2 - 1, \ldots, x_n^2 - 1\rangle) \leq d$. To simplify notation we will implicitly work modulo the ideal $\langle x_1^2 - 1, \ldots, x_n^2 - 1\rangle$ so that we write (for example) $(x_1x_2x_3)(x_2x_4x_5) = x_1x_3x_4x_5$.

When we have such an S, T we assign $\tilde{\mathbb{E}}[x_S x_T] = \tilde{\mathbb{E}}[x_S]\,\tilde{\mathbb{E}}[x_T]$. Continue this process until there are no such pairs left. For any $S \subseteq [n], |S| \leq d$ that we haven't reach in this way, set $\tilde{\mathbb{E}}[x_S] = 0$.

We have now defined $\tilde{\mathbb{E}}$ for all monomials so we can extend it by linearity to all polynomials. By construction it satisfies (5.1b) and (5.2). The nontrivial part is to show that it satisfies (5.1c). To do this, first we have to argue that the process we have defined will not run into "contradictions," that is, if we can reach x_S in two different ways, those must have involved the same clauses in a different order. This fact is nontrivial but follows from the variable-clause constraint graph being a sufficiently good bipartite expander. Random graphs have this property.

Another way to see it is that the original 3-XOR game could be thought of as the $\mathbb{F}_2$-linear equations

$$\exists u \in \mathbb{F}_2^n \text{ s.t. } Au = s \tag{5.3}$$

where $s \in \mathbb{F}_2^m, A \in \mathbb{F}_2^{m \times n}$ are given. A "refutation" is a proof that no such u exists. Specifically we will consider refutations that are vectors $v \in \mathbb{F}_2^m$ such that

$$v^T A = 0 \quad \text{and} \quad v^T s = 1. \tag{5.4}$$

It is not hard to see that (5.3) is satisfiable iff (5.4) is unsatisfiable. Also, finding a contradiction in our construction of $\tilde{\mathbb{E}}$ is equivalent to finding vectors v_1, v_2 such that $v = v_1 + v_2$ satisfies (5.4) and $|A^T v_1|, |A^T v_2| \leq d$. This implies that $|A^T v| \leq 2d$ and $v \neq 0$. However, a standard result in error-correcting codes due to Sipser and Spielman states that for random degree-3 graphs $\ker A^T$ contains no vectors of weight $o(n)$ except the 0 vector. (Note that this proof also relies on bipartite expansion, so is essentially a rephrasing of the above proof sketch.)

Once we have no contradiction we can assign the monomials to equivalence classes. Say that $S \sim T$ if $\tilde{\mathbb{E}}[x_S x_T] \neq 0$. (Note that for all S, $\tilde{\mathbb{E}}[x_S] \in \{-1, 0, 1\}$.) When this occurs we also have

$$\tilde{\mathbb{E}}[x_S x_T] = \tilde{\mathbb{E}}[x_S]\,\tilde{\mathbb{E}}[x_T]. \tag{5.5}$$

Given a polynomial q, group the terms by equivalence class so that $q = \sum_k q_k$ where each q_k has monomials in a single distinct equivalence class. Then $\tilde{\mathbb{E}}[q^2] = \sum_k \tilde{\mathbb{E}}[q_k^2]$ since the cross terms vanish. Consider now a single $q_k = \sum_S \alpha_S x_S$, so that

$$\tilde{\mathbb{E}}[q_k^2] = \sum_{S,T} \alpha_S \alpha_T\, \tilde{\mathbb{E}}[x_S x_T] \overset{(5.5)}{=} \sum_{S,T} \alpha_S \alpha_T\, \tilde{\mathbb{E}}[x_S]\,\tilde{\mathbb{E}}[x_T] = \left(\sum_S \alpha_S\, \tilde{\mathbb{E}}[x_S]\right)^2 \geq 0. \tag{5.6}$$

□

Implications for h_{Sep}. Given a $\tilde{\mathbb{E}}$ of degree k we can construct a state on $(\mathbb{C}^n)^{\otimes k/2}$

$$\rho \propto \sum_{S,T \text{ types of deg } \leq k/2} \tilde{\mathbb{E}}[x_S s_T]\, |S\rangle \langle T|\,. \tag{5.7}$$

Here a type S is a subset of $[n]$ of size $k/2$ and $|S\rangle$ is the superposition of all strings with this empirical distribution. The symmetry means $\rho \succeq 0$ and has support in the symmetric subspace. It is also PPT. Thus it passes the SoS-based tests for separable states presented in the previous lectures.

However corresponding to the Grigoriev 3XOR instance ϕ there is a measurement M_ϕ such that $\max_x \phi(x) = h_{\text{Sep}}(M_\phi)$. Then

$$\text{SoS}_k(M_\phi) = 1, \qquad h_{\text{Sep}}(M_\phi) \leq \frac{1}{2} + \epsilon. \tag{5.8}$$

The relevant dimensions come the hardness constructions earlier which maps CSPs (there 3-SAT but could also be 3-XOR) into measurements. Thus deciding if $h_{\text{Sep}(d,2)}$ is $= 1$ or $\leq 1/2$ needs $k = \tilde{O}(\log^2 d)$ of SoS and deciding if $h_{\text{Sep}(d,2)}$ is $= 1$ or $\leq 1 - 1/d$ needs $k = \tilde{O}(d)$ levels. The full details are in our paper with Xiaodi Wu [2].

5.2 Extension Complexity

SoS is an instance of an SDP relaxation: approximating a convex body by a projection of a simpler body. The study of when this is possible is called extension complexity.

5.2.1 Polytopes and nonegative rank

To build intuition, we'll start with the case of polytopes and linear programming relaxations. For polytopes Q, P, say that Q is an extension of P if $P = T(Q)$ for some linear map T. The complexity of Q is the dimensionality of Q + the number of facets. The *extension complexity* of P (denoted $\text{xc}(P)$) is the minimum complexity of Q over all extensions.

This is relevant because if Q extends P, can optimize over P by solving an LP over Q, i.e.

$$\max\{y^T x : x \in P\} = \max\{(y^T T)z : z \in Q\}. \tag{5.9}$$

Any polytope can be represented by its slack matrix M_P with rows corresponding to facets and columns to vertices. Suppose $P = \{x : Ax = b, Cx \leq d\}$ has f faces (each of the form $c_j x = d_j$) and v vertices $x_1, \ldots, x_v$. Then $(M_P)_{i,j} := d_j - c_j^T x_i$. This measures how much "slack" there is in the constraint. By construction the entries are nonnegative.

Definition 2. Given a matrix $A \in \mathbb{R}^{m,n}_{\geq 0}$, the nonnegative rank $\text{rank}_+(A)$ is the minimum r such that $A = BC$ for nonnegative matrices B and C with dimensions $m \times r$ and $r \times n$ respectively.

Theorem 3 (Yannakakis)**.** *Extension complexity of $P \subseteq \mathbb{R}^n$ is $\Theta(n + \text{rank}_+(M_P))$.*

Proof. First we show that nonnegative rank is an upper bound on extension complexity. Let the polytope $P = \{x : Ax = b, Cx \leq d\}$ with f faces and v vertices, and suppose that its slack matrix has a decomposition $M_P = FV$, with dimensions $f \times r$ and $r \times v$. Then we claim that the polytope

$$Q = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} : Ax = b, Cx + Fy = d, y \geq 0 \right\} \tag{5.10}$$

is an extension of P. (This appears to have many equality constraints but you can just pick a linearly independent subset of them.) To show this, take T to be the map that projects onto the x component, i.e. $T \begin{pmatrix} x \\ y \end{pmatrix} = x$. Then clearly $T(Q) \subseteq P$. Moreover, every vertex of P lies in $T(Q)$ by taking y to be the corresponding column of V.

Now we show that extension complexity is an upper bound on nonnegative rank. By rotating and introducing slack variables (at most doubling the number of vars and constraints), we assume that the map T is just a projection on to the first n coordinates, and the Q has the form $\left\{ \begin{pmatrix} x \\ y \end{pmatrix} : Rx + Sy = t, y \geq 0 \right\}$. Then for each vertex x_i, there exists y_i such that $(x_i, y_i) \in Q$. Moreover, each vertex satisfies $c_j^T x_i \leq d_j$ for each facet inequality j, and hence the optimum of the LP

$$\begin{aligned} \max_{x,y} \quad & c_j^T x_i \\ \text{s.t.} \quad & \begin{pmatrix} R & S \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = t \\ & y \geq 0 \end{aligned}$$

is at most d_j. The dual to this LP is

$$\begin{aligned} \min_{u,f} \quad & u^T t \\ \text{s.t.} \quad & \begin{pmatrix} R^T \\ S^T \end{pmatrix} u = \begin{pmatrix} c_j \\ f \end{pmatrix}. \\ & f \geq 0 \end{aligned}$$

By strong duality, therefore, there exists a dual feasible point u_j, f_j s.t. $u_j^T t = d_j$ and $c_j x_i + f_j y_i = u_j^T \begin{pmatrix} R & S \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = d_j$. Hence putting together the ys and fs into a matrix one gets a factorization for M_P. The dimension r of the matrices in this factorization are given by the dimension of y. □

Yannakakis used this to show that there exist no sub-exponential size symmetric LPs for matching/TSP polytopes.

5.2.2 SDPs and PSD-rank

The preceeding discussion can be generalized to SDP relaxations. Define the semidefinite extension complexity of a convex body K (denoted also $\mathrm{xc}_{\mathrm{sdp}}(K)$) as the minimum dimension + number of constraints of an SDP expressing K. It turns out that for a polytope P we have again

Theorem 4. $xc_{sdp}(P) = \mathrm{rank}_{psd}(M_P)$

Here $\mathrm{rank}_{\mathrm{psd}}(M)$ is the lowest number r for which $M_i = \langle A_i, B_j \rangle$ for some r-dimensional psd matrices $\{A_i\}, \{B_j\}$, and where $\langle A, B\rangle = \mathrm{tr}\, A^\dagger B$.

This is about representing a polytope exactly as a spectrahedral lift. But we want to approximately solve some approximation problem. For instance, suppose we want to obtain a (c, s)-aproximation to a CSP using semidefinite programming, i.e. we want to know whether $\mathrm{val}(\phi)$ is $\geq c$ or $\leq s$ given the promise that at least one of these holds.

Turns out, there is a "slack matrix" for this problem. Let $\Pi = \{\phi : \mathrm{val}(\phi) \leq s\}$, and define the matrix M with rows indexed by instances in Π and columns indexed by assignments, such that $M(\phi, x) = c - \phi(x)$. Then $\mathrm{rank}_{\mathrm{psd}}(M)$ is the size of the smallest SDP to obtain a (c, s) approximation to the problem.

Here we consider only SDPs where the constraints depend on the type of CSP (e.g. 3-XOR, k-coloring, etc.), but not the particular instance. The choice of instance should enter only via the objective function.

5.2.3 LRS: connection SoS degree to PSD rank

LRS shows that SoS is in some sense the optimal SDP relaxation for CSPs.

Theorem 5. *Suppose SoS needs at least degree d to achieve a (c, s) approximation for some CSP on instances of size m. Then for all $n \geq 2m$, $\mathrm{rank}_{psd}(M) \geq C_m \left(\frac{n}{\log n}\right)^{d/4}$, where M is the slack matrix for instances of size n as defined above.*

Important technical note: C_m is allowed to depend on m! So the above theorem is not as good as it looks. There are some bounds on the dependence of C_m which are only nontrivial when d is small ($\log(n)$). We end up with extension complexity lower bounds for e.g. 3XOR of $n^{\tilde{\Omega}(\log n)}$.

Proof. The proof is quite technical so we will just sketch the main steps.

First, the claim that we need degree d to achieve a (c, s) approximation for the CSP can be reformulated: it is saying that for some instance ϕ of size m with $\mathrm{val}(\phi) < s$, function $f(x) = c - \phi(x)$ has SOS degree d. This implies that there exists some pseudoexpectation $\tilde{\mathbb{E}}[\cdot]$ defined up to degree $d - 2$ such that $\tilde{\mathbb{E}}[f] < 0$.

We will show the contrapositive of the theorem; that is, we will show that if M has psd rank that is too small, then $\tilde{\mathbb{E}}[f] \geq 0$ which violates our assumption.

Actually, instead of considering the full matrix M, consider a submatrix called the "pattern matrix" M_n^f, with rows indexed by subsets $S \subset [n]$ of size m, and columns indexed by assignments $x \in \{0, 1\}^n$, such that

$$M_n^f(S, x) = f(x_S).$$

If the psd rank of M is small, then so is that of M_n^f. Write our factorization

$$M_n^f(S, x) = \mathrm{Tr}(P(S)Q(x)) = f(x_S).$$

If $Q(x)^{1/2}$ were a low-degree polynomial in x, we'd be done, since it would imply that $f(y)$ has low SOS degree.

What LRS show is that given an arbitrary $Q(x)$, can find a low-degree $R(x)$ such that

$$|\mathop{\mathbb{E}}_x \mathrm{Tr}\big(\Lambda(x)(Q(x) - R(x)^2)\big)| \leq \epsilon,$$

for all "simple" functions $\Lambda(x)$ (i.e. low degree polynomials with bounded norm). If we choose ϵ to be small enough, then applying this to our factorization, we get

$$\mathop{\mathbb{E}}_S \tilde{\mathbb{E}}_x \mathrm{Tr}\big(P(S)R(x)^2\big) < 0$$

which is impossible since $\mathbb{E}$ is a valid pseudo-expectation covering polynomials of the degree we encounter here.

Why can we restrict to $\Lambda(x)$ that are themselves low degree? This uses $n \geq 2m$ and some new facts about random restrictions. □

5.2.4 Applications to h_{Sep} and the no-disentangler conjecture

Everything we said above was for CSPs. How do we apply it to h_{Sep}? Everything in this section is from [2].

As we have seen with the hardness results, we can embed a CSP into h_{Sep} by mapping $\phi \mapsto M_\phi$ such that $\phi(x) \approx h_{\mathrm{Sep}}(M_\phi)$. The way this works is that there is also a map from solutions x to "honest witness states" ρ_x such that

$$h_{\mathrm{Sep}}(M_\phi) = \max_{\sigma \in \mathrm{Sep}} \mathrm{tr}[M_\phi \sigma] \approx \max_x \mathrm{tr}[M_\phi \rho_x] = \max_x \phi(x). \tag{5.11}$$

Following LRS we can get results like this:

Corollary 6. *To achieve a $(1, 1 - 1/d^2)$ approximation to h_{Sep}, we need an SDP of size $d^{\log d / \log \log d}$.*

This is the worst of both worlds; we have not only $1 - 1/\mathrm{poly}(d)$ soundness but only a quasipolynomial lower bound on dimension. This is because of the various losses mostly from LRS. There was a big improvement in efficiency for LPs, due to [3].

One neat application of Corollary 6 is a weak version of the "no approximate disentangler" conjecture of Watrous, first reported in [1]. An exact disentangler is a surjective linear map from $D(\mathbb{C}^{d'})$ to $\mathrm{Sep}(d, 2)$. An ϵ-approximate disentangler is a map from $D(\mathbb{C}^{d'})$ whose image is ϵ-close in Hausdorff distance (see wikipedia for def) to $\mathrm{Sep}(d, 2)$.

If Φ is an ϵ-approximate disentangler then $\max_{\rho \in D(\mathbb{C}^{d'})} \mathrm{tr}[\Phi(\rho) M]$ gives an ϵ-approximation to $h_{\mathrm{Sep}}(M)$. So if $\epsilon \leq 1/d^2$ then we can show that $d' \lesssim d^{\log d / \log \log d}$ is impossible.

References

[1] S. Aaronson, S. Beigi, A. Drucker, B. Fefferman, and P. Shor. The power of unentanglement. *Annual IEEE Conference on Computational Complexity*, 0:223–236, 2008, arXiv:0804.0802.

[2] A. W. Harrow, A. Natarajan, and X. Wu. Limitations of semidefinite programs for separable states and entangled games, 2016, arXiv:1612.09306.

[3] P. K. Kothari, R. Meka, and P. Raghavendra. Approximating rectangles by juntas and weakly-exponential lower bounds for lp relaxations of csps. In *Proc. of STOC*, pages 590–603. ACM, 2017, arXiv:1610.02704.