SOS Lecture 4: SOS Lower bound for random 3XOR / Boaz Barak

Debt from last time Gaussian Sampling Lemma - didn't forget but will return to it when we need it

Limitations of SOS The SOS method is quite powerful and we have very few (in some sense only one*) lower bounds for it. We will see the central one now. In some sense this is not surprising, since it merely says that the SOS method cannot solve an NP hard problem, but "beggars can't be choosers" and we should be happy with what we are able to prove. This lower bound was first proven by Grigoriev in 2001 (interestingly, about the same time that Parrilo and Lasserre gave the SOS algorithm, which Grigoriev wasn't aware of) and rediscovered by Schoenebeck in 2008.

Theorem There exists $\epsilon > 0$ such that if $\mathcal{E}$ is a random set of $m = 100n$ 3XOR equations over $\mathrm{GF}(2)^n$, then with constant probability there is a degree ϵn distribution $\{x\}$ that satisfies $\mathcal{E}$.

(To formulate this as a polynomial program we naturally identify GF(2) with $\{\pm 1\}$ in which case addition modulo 2 becomes multiplication and every linear equation becomes a polynomial equation of the form $x_i x_j x_k = \sigma$ for some $\sigma \in \{\pm 1\}$.)

Notes This is non-trivial since (as you'll show in an exercise) with probability $1 - o(1)$, $\mathcal{E}$ will not be satisfiable and in fact there will not be an assignment satisfying $1/2 + \delta$ fraction of the equations (where δ tends to zero with 100..).

The problem of determining whether a set of linear equations modulo 2 is satisfiable is of course not NP hard. As a rule of thumb it seems that the SOS algorithm is often "inherently robust" in the sense that if it can determine whether some family of equations is satisfiable, it can also determine whether it is *almost satisfiable*. In contrast the Gaussian elimination algorithm is extremely "brittle" and works only for perfectly (or exponentially close to perfectly) satisfiable equations. So, the hardness for SOS can be "explained" by the NP hardness (due to Håstad) of the task of distinguishing between a set of linear equations that can be 0.999 satisfied and a set for which no assignment satifies more than 0.51 fraction of the equations.

The proof Let $G = (L \cup R, E)$ be the $m+n$ vertex bipartite graph where we connect a left vertex $\ell \in L$ to the right vertices $i, j, k \in R$ if the equation $x_i x_j x_k = \sigma$ exists in $\mathcal{E}$. We make the following claim:

Lemma 1. *There exists some $\epsilon > 0$ such that with constant probability, for every subset $S \subseteq L$ with $|S| \leq \epsilon n$,$|\Gamma(S) \geq 1.7|S|$, where $\Gamma(S) = \{j \in R : \exists i \in S s.t. (i,j) \in R\}$.*

We say that G is a "lossless expander" if it satisfies the conditions of Lemma 1. (The name comes from the extractor literature and generally applies to d-regular graphs where sufficiently small sets have expansion larger than $d/2$; interestingly we do not know how to efficiently verify lossless expansion, and also have very few explicit constructions of such graphs.)

The proof of the lemma is left as an exercise. Note that the qualification of constant probability (as opposed to probability close to 1) is needed since with probability at least 0.9 or so we will have two equations that share two variables (for every particular pair this will happen with probability n^{-2} and there are m^2 pairs). However, it is actually true that with very high probability for random equations there will be a pseudo-distribution of $\Omega(n)$ degree that satisfies $m - o(m)$ of the constraints.

Given this the main result follows from the following lemma:

Lemma 2. *Suppose that G is a lossless expander as in Lemma 1. Then there exists a degree $d = \epsilon n/1000$ pseudo-distribution $\{x\}$ satisfying all constraints in $\mathcal{E}$.*

Constructing the pseudo-distribution The pseudo-distribution $\{x\}$ is constructed by following Einstein's method who said

> Pseudo-distributions should be constructed to be as random as possible but not randomer.

Musings To quote some comments in prior lecture:

- In statistical learning problems (and economics) we often capture our knowledge (or lack thereof) by a distribution. If an unknown quantity X is selected and we are given the observations y about it, we often describe our knowledge of by a the distribution $X|y$. In computational problems, often the observations y completely determine the value X, but pseudo-distribution can still capture our "computational knowledge".
- The proof system view can also be considered as a way to capture our limited computational abilities. In the example above, a computationally unbounded observer can deduce from the observations y all the true facts it implies and hence completely determine X. One way to capture the limits of a computationally bounded observer is that it can only deduce facts using a more limited, sound but not complete, proof system.

Here we we can think of a more concrete instantiation of this. Think of X is a "planted" solution to the linear system (perhaps with some noise) and y as the actual equations themselves. Pseudo-distributions can be thought of as capturing the knowledge of a computationally bounded observer about the potential satisfying assignment for the equations $\mathcal{E}$. A priori the assignment could be completely random, but of course once we see the equations they determine some constraints on it. An unbounded observer would see that the equations determine so many constraints that in the planted case, determine it completely and in the random case there simply doesn't exist such an assignment. But a bounded observer might only be able to deduce short linear relations that are obtained by combining few of the equations. This is what is meant to be captured by the quote above. Lets see this in action.

Definition of $\{x\}$ For every $U \subseteq [n]$ of size at most d, we need to define $\tilde{\mathbb{E}} x_U = \tilde{\mathbb{E}} \prod_{i \in U} x_i$. We start by having $\tilde{\mathbb{E}} x_U$ undefined for all U. Clearly if $U = \{i, j, k\}$ where the equation $x_i x_j x_k = \sigma$ appears in $\mathcal{E}$ then we must define $\tilde{\mathbb{E}} x_U = \sigma$. We then apply the following rule until we can't do it anymore: if U, V are subsets of $[n]$ of size at most d for which $\tilde{\mathbb{E}} x_U$ and $\tilde{\mathbb{E}} x_V$ are defined and $|U \oplus V| \leq d$ then we define $\tilde{\mathbb{E}} x_{U \oplus V} = (\tilde{\mathbb{E}} x_U)(\tilde{\mathbb{E}} x_V)$. After we finish applying this rule, everything that is not defined is defined to be zero.

Lemma 3. *We never attempt to give two different values to $\tilde{\mathbb{E}} x_U$.*

Proof. If there is U that is assigned two different values, then it means that there is a derivation that shows $\tilde{\mathbb{E}} x_\emptyset = -1$. We will now show this leads to a contradiction. Such a derivation can be described by a sequence of sets $U_1, \ldots, U_t$ such that $U_t = \emptyset$ and for every i, either U_i is one of the basic sets in $\mathcal{E}$ or $U_i = U_j \oplus U_k$ where $j, k < i$. Let $E_1, \ldots, E_m$ be the sets corresponding to the equations in $\mathcal{E}$ and $\sigma_1, \ldots, \sigma_m$ the corresponding values. For every i in this derivation there is a set $S_i \subseteq L$ such that $U_i = \oplus_{\ell \in S_i} E_\ell$ and the value assigned to $\tilde{\mathbb{E}} x_{U_i}$ is $\prod_{\ell \in S_i} \sigma_\ell$.

We claim that for every set $S \subseteq L$ of size at most $100d$, $|\oplus_{\ell \in S}| \geq |S|/10$. Indeed, suppose otherwise, and let $T = \oplus_{\ell \in S}$. Then, we get that every vertex in $\Gamma(S) \setminus T$ must have at least two neighbors in S (since they need to be canceled). Since (by expansion) $|\Gamma(S)| \geq 1.7|S|$ and there $3|S|$ total edges leaving S, we get that

$$3|S| \geq 2(1.7|S| - |T|)$$

or

$$|T| \geq 1.7|S| - 1.5|S| = 0.2|S|$$

By induction, this means that every set S_i in the derivation must have size at most $10d$ (otherwise, the first set S_i violating this will equal $S_j \oplus S_k$ each of size at most $10d$ and hence would have size in $(10d, 20d]$ but then it can't be the case that $|\oplus_{\ell \in S_i} E_\ell| \leq d$. In particular we get that there is a set $S = S_t$ of size at most $10d$ such that

$$\oplus_{\ell \in S} S_i = \emptyset$$

contradicting the claim. □

Positivity Let $P = \sum c_U x_U$ be a polynomial of degree at most $d/10$. We need to show that $\tilde{\mathbb{E}} P^2 \geq 0$.

We define a relation $U \sim V$ if $\tilde{\mathbb{E}} x_{U \oplus V}$ is defined. Note that this is indeed an equivalence relation, since if $\tilde{\mathbb{E}} x_{U \oplus V}$ is defined and $\tilde{\mathbb{E}} x_{V \oplus W}$ is defined then so is $\tilde{\mathbb{E}} x_{U \oplus W} = \tilde{\mathbb{E}} x_{(U \oplus V) \oplus (V \oplus W)}$.

Split $P = \sum P_i$ where each P_i contains the monomials that come from a particular equivalence class. Then

$$\tilde{\mathbb{E}} P^2 = \sum_{i,j} \tilde{\mathbb{E}} P_i P_j = \sum_i \tilde{\mathbb{E}} P_i^2$$

since for every U, V if $U \not\sim V$ then $\tilde{\mathbb{E}} x_U x_V = 0$.

Thus it suffices to consider the case when all the monomials of P come from the same equivalence class. In this case we can write $P = x_{U_0} \sum c_U x_U$ where for every U in this sum the value $\tilde{\mathbb{E}} x_U$ is defined (using the fact that in our distribution x_i^2 is identical to 1). Therefore,

$$\tilde{\mathbb{E}} P^2 = \tilde{\mathbb{E}} (x_{U_0})^2 \sum_{U,V} c_U c_V \tilde{\mathbb{E}} x_U x_V = \tilde{\mathbb{E}} (\sum c_U \tilde{\mathbb{E}} x_U)^2$$

where the last equality holds because $(x_{U_0})^2$ is identical to 1 and for every two defined values $\tilde{\mathbb{E}} x_U, \tilde{\mathbb{E}} x_V$ $\tilde{\mathbb{E}} x_U x_V = (\tilde{\mathbb{E}} x_U)(\tilde{\mathbb{E}} x_V)$.

Reductions The random 3XOR lower bound is the analog of Håstad's PCP, and just like the latter serves as a basis for many hardness of approximation results, the former serves as a basis for many SOS lower bounds. However, because this mimics the NP hardness results, it amounts to showing that the SOS algorithm cannot prove that P=NP (or that NP has a subexponential algorithm) which of course we didn't expect it (or any other algorithm) to do anyway.

The relation between these integrality gaps and NP hardness is not yet fully understood. As one example, the same methods as above can show an SOS lower bound for linear equations that are somewhat more structured - pick a subspace $V \subseteq \mathrm{GF}(2)^k$ that does not contain any

vector of Hamming weight 1 or 2, and pick random equations of the form $(x_{i_1}, \ldots, x_{i_k}) \oplus (\sigma_1, \ldots, \sigma_k) \in V$. Until recently there was no corresponding NP hardness result, but this was rectified in a recent breakthrough of Siu On Chan.

A very interesting open question is to show either an SOS lower bound or an NP hardness result for the same question but where V is not a subspace but rather an arbitrary subset of $\{0,1\}^k$ that supports a *pairwise independent distribution*. Perhaps the simplest example is when $(x_1, \ldots, x_5) \in V$ if $x_1 \oplus x_2 \oplus x_3 = x_4 \wedge x_5$. (This is known as the TSA predicate.) Only unique game hardness is known, but in contrast to problems such as Max-Cut, in this case the unique games conjecture seems less inherent and I would imagine that it could be replaced with NP hardness. In particular, no sub-exponential algorithm is known for this question.

In a paper with Kindler and Steurer we made the (somewhat bold) conjecture that in fact all these problems are hard even for random instances.