Lecture 11: "random 3-XOR and Res(k)" - part I

In this lecture and the next one we will prove an exponential size lower bound for random 3-XOR formulas for Res(k).
So what are random 3-XOR formulas?
A random 3-XOR formula in n variables and Δn equations is a formula of the following form

$$F = \bigwedge_{i=1}^{\Delta n} L_i$$ where each L_i is a parity of 3 variables chosen accordingly to the following process: for each i independently choose a set $S_i \subseteq \binom{n}{3}$ and a bit $b_i \in \{0,1\}$ and set

$$L_i = \text{“}\sum_{j \in S_i} x_j \equiv b_i \pmod 2\text{”} \quad (*)$$

Clearly (*) can be written as a CNF (how?) and hence random 3-XOR formulas are CNF in the end.

OBS: Lower bounds for random 3-XOR formulas in n vars. and Δn lin. eq. imply lower bounds for random 3-CNF formulas in n vars. and Δn clauses (how?)

We prove the following result:

Theorem 1: for any constant Δ sufficiently large with prob. $1-o(1)$ every Res(k) refutation of F, a random 3-XOR formula in n variables and Δn lin. equations, will require size $\geq 2^{n^{1-o(1)}}$ where $k < \sqrt{\log n}$

N.B.: Unlike PHP_n^{2n} this thm we prove is the best, i.e. a l.b. in the strongest propositional proof system, we know for random 3-XOR!

the proof scheme is similar to the one we used for $PHP^{\Delta n}_n$:

- we will have a notion of lines of Res(k) in normal-form analogous of the notion of lines in pigeon-normal-form for $PHP^{\Delta n}_n$.
- we will have a switching lemma working for lines in normal-form.
- we will need a "cleaning step" analogous as the step from $PHP^{\Delta n}_n$ to PHP(G) of bounded degree.
- we will need a width l.b. for resolution (obtained by other means).

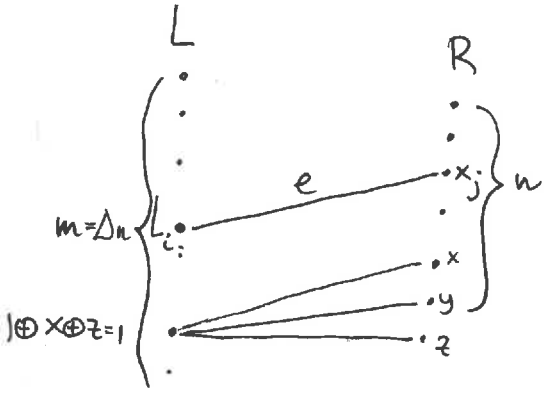
o——o

In this lecture we prove <u>theorem 1</u> modulo the proof of the switching lemma and the "cleaning step".

So, let's go with order, for $PHP^{\Delta n}_n$ we had naturally associated a bipartite graph (the pigeon-holes graph). Here we need a bipartite graph, the adjacency graph, capturing some properties of a 3-XOR formula $F = \bigwedge_{i=1}^{\Delta n} L_i$ where L_i's are XOR of 3 variables.

$G_F = (L \cup R, E)$ L "left" part also representing lines of $F = \bigwedge_{i=1}^{\Delta n} L_i$

R "right" part representing variables $\{x_1, \dots, x_n\}$



$e = (i, j) \in E$ iff x_j appears in L_i

the left degree of G_F is 3, the right degree might be large but don't worry about it for the moment.

- given $A \subseteq L$, $\partial A = \{ v \in R : \exists! \, w \in A \;\; (w, v) \in E \}$ "boundary of A"

  $N(A) = \{ v \in R : \exists \, w \in A \;\; (w, v) \in E \}$ "neighbors of A"
- $G = (L \cup R, E)$ is an (r, c)-∂-expander iff

  $\forall A \subseteq L \quad |A| \leq r \rightarrow |\partial A| \geq c|A|$

Some useful properties of G_F:

① Given $F = \bigwedge_{i \in [m]} L_i$ with L_i lin. equations, if G_F is an (r,c)-boundary expander with $c > 0$ then every subset $A \subseteq [m]$ s.t. $|A| \leq r$ has the property that $\bigwedge_{i \in A} L_i$ is satisfiable.

Hint: A has a matching in G_F.

② If $F = \bigwedge_i L_i$ is unsatisfiable and G_F is an (r,c)-boundary expander with $c > 0$ then $\text{width}(F \vdash) \geq \frac{rc}{2}$.

Hint: write down a suitable complexity measure on clauses.

③ If $\overbrace{\bigwedge_{i \in A} L_i}^{F} \wedge t$ is minimally unsatisfiable, where the L_is are lin. eq. and t is a term, then (i) G_F is connected

(ii) $\partial A \subseteq \text{var}(t)$ or more precisely

$$\partial A \subseteq \{ j : x_j \in \text{var}(t) \}$$

Support and normal forms

Def: (support) Let $G=(L\cup R, E)$ be a bipartite graph that is an (r,c)-boundary expander. Let $J\subseteq R$, we say that a set $A\subseteq L$ is J-valid if $|A|\leq r$ and $\partial A\subseteq J$. The support of J, $supp(J)$, is the union of all sets $A\subseteq L$ that are J-valid.

obs 1: $supp(\{v\})=\emptyset$ if $c>\frac{1}{2}$ and the L-degree of G is >1.

pf: Suppose by contr. that there exists some set $A\subseteq L$ $\{v\}$-valid and non-empty

CASE 1: $|A|=1$ then $|\partial A|=|N(A)|>1$ by hyp so it can't hold that $\partial A\subseteq\{v\}$.

CASE 2: $|A|\geq 2$ then by expansion $|\partial A|\geq c|A|\geq 2c>1$ and again it can't hold that $\partial A\subseteq\{v\}$.

■

obs 2: if $A\subseteq L$ is J-valid then $|A|\leq\frac{|J|}{c}$.

pf: $c|A|\leq|\partial A|\leq|J|$. ■

obs 3: if $|J|\leq\frac{cr}{2}$ then $|supp(J)|\leq\frac{|J|}{c}$.

pf: we prove that every union $\bigcup_{i=1}^{m}A_i$ of J-valid sets is J-valid, since $supp(J)$ is a union of J-valid sets and by the previous obs we have the bound. By ind on m:

$m=1$ ✓

$m\to m+1$: $\partial\left(\bigcup_{i=1}^{m+1}A_i\right)\subseteq J$ clearly and $\left|\bigcup_{i=1}^{m+1}A_i\right|\leq\left|\bigcup_{i=1}^{m}A_i\right|+|A_m|\leq\frac{2|J|}{c}\leq r$ (obs 2 & ind. hyp.)

so $\bigcup_{i=1}^{m+1}A_i$ is J-valid.

■

obs 4: $J\subseteq J'\Rightarrow supp(J)\subseteq supp(J')$

Def: (normal form) Let $F = \bigwedge_i L_i$ the usual 3-XOR formula we are considering in this lecture and let t be a term, we say that t is locally consistent, if

wrt F $\quad t \wedge \bigwedge_{i \in \mathrm{supp}(t)} L_i$ is satisfiable,

where $\mathrm{supp}(t) = \mathrm{supp}(\{j : x_j \in \mathrm{var}(t)\})$.

A DNF φ is in **normal form** if all its terms are locally consistent.

obs 5: If t is locally consistent wrt $F = \bigwedge_{i \in [m]} L_i$ then for each $A \subseteq [m], |A| \leq r$ $\quad t \wedge \bigwedge_{i \in A} L_i$ is satisfiable.

pf: By contradiction let $t \wedge \bigwedge_{i \in A} L_i$ be unsatisfiable and let $t' \subseteq t$, $A' \subseteq A$ s.t.

$t' \wedge \bigwedge_{i \in A'} L_i$ minimally unsatisfiable.

Then $\exists A' \subseteq \mathrm{var}(t') \subseteq \mathrm{var}(t)$ so $A' \subseteq \mathrm{supp}(t)$ but t was locally consistent. ↯

□

exercise : Let $F = \bigwedge L_i$ have a Res(k) refutation of size S.
Suppose that F is a 3-XOR, G_F is an (r, c)-boundary expander with $c > \frac{1}{2}$
and $k \leq \log n$, then
there exists another Res(k) refutation of F of size
$S \cdot n^{O(1)}$ but having just lines in normal form.

exercise : Let π be a Res(k) proof of F and ρ a part. assignment,
then if π is in normal form also $\pi|_\rho$ is in normal form
and as usual $\pi|_\rho$ is also a Res(k) refutation of $F|_\rho$.

In the next lecture we will construct a distribution $\mathcal{D}$ on partial assignments s.t. the following holds:

Lemma: Let F be a 3-XOR formula s.t. G_F is an (r,c)-∂-expander with $r = \Omega\left(\frac{n}{d}\right)$ where $n = \#$ vars of F

$d =$ max degree of G_F.

Let φ be any K-DNF in normal form, then

$$\Pr_{\alpha \in \mathcal{D}}\left[\varphi|_\alpha \neq 1\right] \leq 3 \cdot e^{-\frac{cov(\varphi)}{d^{O(k)}}},$$

where $cov(\varphi)$ is the covering number of φ and $\mathcal{D}$ is a distribution we have to build.

So let's finish this lecture with the proof of theorem 1. This proof is not exactly self-contained (modulo the proof of Lemma above) but there will be just two claims we will see next lecture.

Proof of theorem 1: Let π be a Res(k) refutation of F, w.l.o.g we can suppose that π is in normal form (why?). Whp G_F is an $(\delta n, .8)$-boundary expander for some constant $\delta \gamma < 1$.

Claim 1: there exists a restriction ρ s.t. $G_{F|_\rho}$ is an $(\frac{\delta n}{2}, .4)$-boundary expander and $d =$ max degree of $G_{F|_\rho}$ is at most $c\Delta$ for some constant c, i.e. d is a constant (since Δ is a constant).

Using the switching lemma and the lemma above we can estimate the probability that after a restriction $\sigma \in \mathcal{D}$ the lines φ in $\pi|_{\rho\sigma}$ have large height. Let "large" for the moment mean $\geq \frac{W}{k}$ for some parameter W we have yet to choose.

$$\Pr_{\sigma\in\mathcal{D}}\left[\exists \varphi\in \Pi\restriction_\rho \; h(\varphi\restriction_\sigma) > \frac{W}{K}\right] \leq |\Pi| \cdot \Pr_{\sigma\in\mathcal{D}}\left[h(\varphi\restriction_\sigma) > \frac{W}{K}\right]$$

union bound & $|\Pi\restriction_\rho| \leq |\Pi|$

$$\leq |\Pi| \cdot 3K\, 2^{-\delta' \frac{W}{2K}}, \qquad (*)$$

switching lemma

where $\delta' = 2\left(\frac{\log_2 e}{4\Delta^{O(K)}}\right)^K \approx \frac{1}{2^{O(K^2)}}$

if $|\Pi| < \frac{2^{\delta' W/2K}}{3K}$ then the probability in $(*)$ is <1, so the complementary event has probability >0 and hence there exists some $\sigma\in\mathcal{D}$ s.t. $\forall \varphi\in\Pi\restriction_\rho \; h(\varphi\restriction_\sigma) \leq \frac{W}{K}$

But $\Pi\restriction_{\rho\circ\sigma}$ is a valid $\mathrm{Res}(K)$ proof of $F\restriction_{\rho\circ\sigma}$ and each line ψ in $\Pi\restriction_{\rho\circ\sigma}$ is s.t. $h(\psi) \leq \frac{W}{K}$ so $\mathrm{width}(F\restriction_{\rho\circ\sigma}\vdash) \leq W$ - eq. 1

Claim 2: for every $\sigma\in\mathcal{D}$, $G_{F\restriction_{\rho\circ\sigma}}$ is an $(\frac{\gamma n}{4}, .2)$-boundary expander.

So Claim 2 and the observation between δ-expansion and width

$$\Rightarrow \mathrm{width}(F\restriction_{\rho\circ\sigma}\vdash) \geq \frac{\gamma n}{8}\cdot 0.2 = \frac{\gamma n}{40} \qquad \text{eq. 2}$$

Let $W = \frac{\gamma n}{40} - 1$. With this choice of W from eq. 1 and eq. 2 we have a contradiction. So it must be that

$$|\Pi| \geq \frac{2^{\delta'\gamma n/80K}}{3K} \approx 2^{2^{n/O(K^2)}} \text{ which is of the form } 2^{n^\varepsilon}$$

whenever $K < \sqrt{\log n}$.

. In the proof above we used Δ constant, what if Δ is allowed to grow as $n\to\infty$, say e.g $\Delta = \log n$. How far we can push theorem 1 w.r.t. to the parameter Δ so that it still hold? (Maybe with some loss in the parameter K...)

□

- The supp* -

We are going to define supp*() for generic expander graphs although we are going to use it just to the graph G_F (or restrictions of it).

Def: (support*) Let $G = (L \cup R, E)$ be a (finite) bipartite graph that is an (r,c)-boundary expander. Let $J \subseteq R$ and consider a sequence $I_0 \subsetneq I_1 \subsetneq I_2 \subsetneq \dots$ sequence of subsets of L s.t. ($I_0 = \emptyset$)

(i) $|I_{k+1} \setminus I_k| \leq \frac{r}{2}$

(ii) $\left|\partial(I_{k+1} \setminus I_k) \setminus (J \cup N(I_k))\right| < \frac{c}{2} |I_{k+1} \setminus I_k|$.

For shortness let's call <u>valid</u> any sequence of subsets (I_k) satisfying the properties above.

Let $\text{supp}^*(J) = \bigcup_{k \leq \bar{u}} I_k = I_{\bar{u}}$ where $(I_k)_{k \leq \bar{u}}$ is a valid sequence of maximal length.

With similar proofs as for the support we have that:

<u>exercise ①</u>: for each $J \subseteq R$, $\text{supp}^*(J)$ is well-defined.

<u>exercise ④</u>: given any valid sequence $(I_k)_{k \in \mathbb{N}}$ defining $\text{supp}^*(J)$, $|\partial I_k \setminus J| < \frac{c}{2} |I_k|$ for any k

<u>exercise ②</u>: if $|J| \leq \frac{cr}{4}$ then $|\text{supp}^*(J)| < \frac{2}{c} |J|$.

pf of ②: by contr. $|\text{supp}^*(J)| \geq \frac{2}{c}|J|$

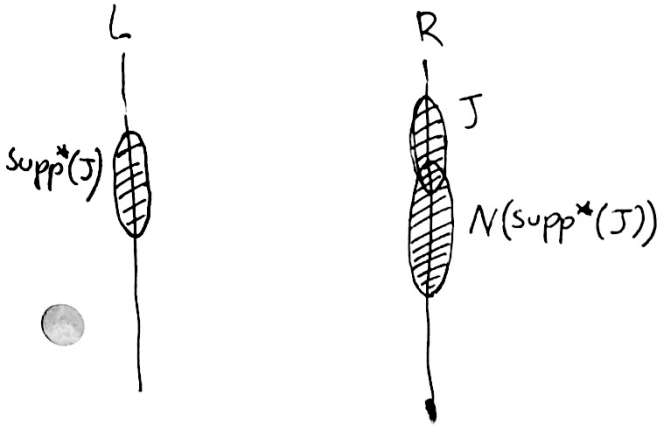
let $I_0 = \emptyset \subsetneq I_1 \subsetneq I_2 \subsetneq \dots \subsetneq I_{\bar{u}} = \text{supp}^*(J)$, there exists some $i \leq \bar{u}$ s.t. $|I_i| \geq \frac{2}{c}|J|$ so

$$|\partial I_i \setminus J| \geq |\partial I_i| - |J| \geq c|I_i| - |J| \geq c|I_i| - \frac{c}{2}|I_i| = \frac{c}{2}|I_i|$$

contradicting ex. 1. ↯

□

observation 1: Let G'' be an (r,c)-boundary expander, let G'_J be the subgraph of G induced on the vertex set $V(G) \setminus (J \cup supp^*(J) \cup N(supp^*(J)))$, where $J \subseteq R$. Then for each $J \subseteq R$, G'_J is an $(\frac{r}{2}, \frac{c}{2})$-boundary expander.



G'_J is obtained from G removing all the vertices in $J \cup supp^*(J) \cup N(supp^*(J))$ and all the edjes adjacent to them.

In particular G'_J is yes a reasonably good expander wrt G but the left degree might decrease.

Suppose you take σ a partial assignment of domain $J \cup N(supp^*(J))$ satisfying all the lin. eq. of $F = \bigwedge_i L_i$ for $i \in supp^*(J)$ and F was a 3-XOR, then $F|_\sigma$ is such that $G_{F|_\sigma}$ is exactly the graph G'_J constructed above but now some of the lin. eq. of $F|_\sigma$ might have just one or two variables.

proof of observation 1: let $A \subseteq L \setminus supp^*(J)$ s.t. $|A| \leq r/2$ and let $(I_k)_{k \leq \bar{u}}$ a maximal lenght chain defining $supp^*(J)$.

Then consider the chain

$$\phi \subsetneq I_1 \subsetneq I_2 \subsetneq \dots \subsetneq \underbrace{I_{\bar{u}}}_{= supp^*(J)} \subsetneq I_{\bar{u}} \cup A \qquad (*)$$

If $|\partial A \setminus (J \cup N(supp^*(J)))| < \frac{c}{2}|A|$ then $(*)$ is a valid chain that is longer contradicting the maximality of $(I_k)_{k \leq \bar{u}}$. So it must be that $|\partial A \setminus (J \cup N(supp^*(J)))| \geq \frac{c}{2}|A|$. That is G'_J is an $(\frac{r}{2}, \frac{c}{2})$-boundary expander

□

Moreover if in observation ① we choose as $J \subseteq R$ the set consisting of the $\frac{cr}{4}$ largest degree vertexes of R and, as in our application, we suppose that the left degree of G is 3 and $|L| = \Delta n$, then G'_J has degree at most $\frac{3\Delta w}{\frac{cr}{4}+1}$.

From this it follows immeadely Claim 1 -

proof of Claim ①: We are going to apply the previous observation to G_F where $F = \bigwedge_i L_i$ is a random 3-XOR s.t. G_F is a $(\gamma n, .8)$-boundary expander. Consider J the set of $0.2\gamma n$ variables of largest degree in G_F Let ρ' be a minimal size partial assignment satisfying $supp^*(J)$ - Why ρ' exists? since $|J| \leq \frac{cr}{4} = 0.2\gamma n$ then, by exercise ②, $|supp^*(J)| \leq \frac{r}{2}$ So $\bigwedge_{i \in supp^*(J)} L_i$ is satisfiable - By minimality of ρ' we have $= \frac{\gamma n}{2}$ that $dom\,\rho' \subseteq N(supp^*(J))$ - Take as ρ some extension of ρ' to $N(supp^*(J)) \cup J$ - Then clearly $G_{F|_\rho} = G'_J$ and then $G_{F|_\rho}$ is an $(\frac{\gamma n}{2}, .4)$-boundary expander and with maximum degree at most $\frac{3\Delta w}{0.2\gamma n+1} < c'\Delta$ for some constant c' - ∎

Regarding Claim ② it is enough to build our distribution $\mathcal{D}$ in such a way that each $\sigma \in \mathcal{D}$ will be of the form

$$\sigma \models supp^*(J) \quad \text{and} \quad dom\,\sigma = J \cup N(supp^*(J))$$

for some not too large J -

We are going now to define the distribution $\mathcal{D}$, then the check that Claim 2 holds is left as an exercise -