Today
Hardness Amplification - I
Yao's XOR Lemma.

Lecture 29
Computational Complexity
(19 May 2020)
Instructor: Prahladh Harsha

Yao's XOR Lemma:

$b : \{0,1\}^n \to \{\pm 1\}$

b is mildly average-case hard
– $\forall$ poly sized ckts C

$$\Pr_{x \sim \{0,1\}^n}[C(x) = b(x)] \leq \frac{90}{100}$$

Qn: Can we construct another fn b' which is significantly harder?

Strongly average-case hard
ie, $\forall$ poly-sized ckts C

$$\Pr_{x \sim \{0,1\}^n}[C(x) = b'(x)] \leq \frac{51}{100}$$

Given $b : \{0,1\}^n \to \{\pm 1\}$

XOR fn: $b^{(k)} : \{0,1\}^{nk} \to \{\pm 1\}$ $\quad b^{(k)}(x_1, \ldots x_k) = \prod_{i=1}^{k} b(x_i)$

Understand: how much harder is $b^{(k)}$ compared to b?

Defn: $f, g : \{0,1\}^n \to \{\pm 1\}$ $\quad X$ - be r.v on $\{0,1\}^n$

$$\mathrm{Cor}_X(f,g) = \left|\mathbb{E}_X[f(X)g(X)]\right|$$

1

$\underset{X}{\mathrm{Cor}}(f,g) = p \iff \Pr_{x \leftarrow X}[f(x) = g(x)] = \begin{cases} \frac{1+p}{2} \\ \frac{1-p}{2} \end{cases}$

- X - distribution, -
  not mentioned - uniform distribution.

- **Defn: (Hardness).** b is (p, S)-hard if for all ckts C of size at most S:

$$\mathrm{corr}(b, C) \le p$$

Qn: If b is (p, S)-hard then $b^{(k)}$ is (p', S')-hard?

$$p' \approx p^k$$

- **Corr:** b is (p, ∞)-hard, then $b^{(k)}$ is (p^k, ∞)-hard.

Yao's XOR Lemma is a computational version of above stmt.

Yao's XOR Lemma:

$b_1, \ldots b_k$ is (p, S)-hard

$\Downarrow$

$\prod_{i=1}^{k} b_i$ is $(p^k + \varepsilon, \varepsilon^2 (1-p)^2 S - O(1))$-hard for all $\varepsilon \in (0,1)$.

- Today: Two function version

Lemma: b is (p, S)-hard, then $b^{(2)}$ is $(p^2 + \varepsilon, \varepsilon^2 S - O(1))$-hard

Proof: Levin's proof
Impagliazzo's proof (2 subproofs
Impagliazzo: MWUM
Nisan: Game theoretic)

Two fn
$b : \{0,1\}^n \to \{\pm 1\}$. (Assumption: b is (p,S)-hard
($\forall$ ckts C of size S, $E[b(x)C(x)] \le p$)
$C : \{0,1\}^n \times \{0,1\}^n \to \{\pm 1\}$

$$\mathrm{Corr}(C, b^{(2)}) = \underset{x,y}{E}\left[b(x)b(y)\,C(x,y)\right]$$
$$= \underset{x}{E}\Big[b(x)\,\underbrace{\underset{y}{E}\left[b(y)\,C(x,y)\right]}_{g(x)}\Big]$$
$$= \underset{x}{E}\left[b(x)\,g(x)\right]$$

If g "could be implemented" by ckt of size S, then $\mathrm{Corr}(b^{(2)}, C) \le p$.

- Bottleneck:
(1) g is not a small-sized ckt.
- it involves b - a hard fn
- it is an average of 2^n-quantities
(2) g is not a Boolean function
$g \in [-p, p]$ ($g = \mathrm{Corr}(C, b)$)

$$\mathrm{corr}(b^{(2)}, C) = \mathop{\mathbb{E}}_{x}[b(x) \cdot g(x)]$$

$$= p \cdot \mathop{\mathbb{E}}_{x}\left[b(x) \cdot \frac{g(x)}{p}\right]$$

<u>Claim</u>: Suppose there exists a randomized ckt $D(X, R)$ s.t.

(*) $\forall x$, $\mathop{\mathbb{E}}_{R}[D(x,R)] = g(x)/p$

(*) size $(D) \leq S$.

then, $\mathrm{corr}(b^{(2)}, C) \leq p^2$.

<u>Pf</u>: $\mathrm{Corr}(b^{(2)}, C) = p \cdot \mathbb{E}\left[b(x) \cdot \frac{g(x)}{p}\right]$

$$= p \cdot \mathop{\mathbb{E}}_{x}\left[b(x) \mathop{\mathbb{E}}_{R}[D(x,R)]\right]$$

$$= p \cdot \mathop{\mathbb{E}}_{x,R}[b(x) \cdot D(x,R)]$$

$$= p \cdot \mathop{\mathbb{E}}_{R}\left[\mathop{\mathbb{E}}_{x}[b(x) \cdot D(x,R)]\right]$$

$$\leq p \cdot \mathop{\mathbb{E}}_{R}[p] = p^2.$$

$\boxtimes$

Conclusion is stronger than what we promised, possibly because the hypothesis is too good to be true.

<u>Claim</u>: Suppose C is a ckt of size S'. Then $\forall\, \delta \in (0,1)$, there is a randomized

ckt $D(x,R)$ st.

- ★ $\forall x, \left| \mathbb{E}_R[D(x,R)] - \frac{g(x)}{p} \right| \leq \frac{\delta}{p}$.
- ★ $Size(D) \leq \frac{S'}{\delta^2} + O\left(\frac{1}{\delta}\right)$

Proof of 2-fn XOR Lemma using above claim.

Let C be any ckt of size S'.

$$corr(b^{(2)}, C) = p \cdot \left| \mathbb{E}_x\left[b(x) \frac{g(x)}{p} \right] \right|$$

$$\leq p \cdot \left\{ \left| \mathbb{E}_{x,R}\left[b(x) \cdot D(x,R) \right] \right| + \left| \mathbb{E}_{x,R}\left[b(x) \left(D(x,R) - \frac{g(x)}{p} \right) \right] \right| \right\}$$

$$\leq p \cdot \left| \mathbb{E}_{x,R}\left[b(x) \cdot D(x,R) \right] \right| + p \cdot \delta/p$$

$$\leq p \cdot p + \delta \quad \text{if} \quad \frac{S'}{\delta^2} + O\left(\frac{1}{\delta}\right) \leq S.$$

$$= p^2 + \delta.$$

$Corr(b^{(2)}, C) \leq p^2 + \delta$ if $size(C) \leq \delta^2 S - O(1)$

Construction of Randomized Circuit D:

Throw the hardness of b into R.

D - sample ℓ inputs $y_1 \cdots y_\ell$ & take the average of $b(\cancel{x}) \cdot C(x,y)$ ℓ

$R \in (\{0,1\}^n \times \{0,1\})$

$\Pr[R = \langle\langle y_1, e_1\rangle, \langle y_2, e_2\rangle, \ldots \langle y_\ell, e_\ell\rangle\rangle]$

$$= \begin{cases} \prod_{i=1}^{\ell} \Pr[Y = y_i] & \text{if } e_i = b(y_i)\ \forall i \\ 0 & \text{otherwise} \end{cases}$$

Description of D: $\mathop{E}_{R}[D(x,R)] = \frac{1}{p} \mathop{E}_{Y}[b(Y) C(x,Y)]$

On input x.

1. Pick $R = \langle\langle y_1, e_1\rangle, \ldots \langle y_\ell, e_\ell\rangle\rangle$
2. Compute $C(x,y_1), \ldots C(x,y_\ell)$
3. Compute $v = \langle e_1 C(x,y_1), e_2 C(x,y_2) \ldots e_\ell C(x,y_\ell)\rangle$
4. Let k be the #1's in v
5. (We expect the #1's in v to be between $k_1 = \left(\frac{1-p}{2}\right)\ell$ & $k_2 = \left(\frac{1+p}{2}\right)\ell$)

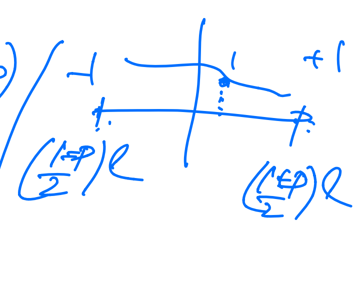
If $k \le \frac{(1-p)\ell}{2}$, output -1

$k \ge \frac{(1+p)}{2}\ell$, output $+1$

Else $k = \left(\frac{1+q}{2}\right)\ell$; $(-p < q < p)$

output 1 w/ prob $\frac{1}{2}\left(1 + \frac{p}{q}\right)$

-1 w/ prob $\frac{1}{2}\left(1 - \frac{p}{q}\right)$



Size of $D = \frac{1}{\delta^2} s' + O\left(\frac{1}{\delta}\right)$.

Approximation Guarantee:

$$\mathbb{E}_R\left[D(x,R)\right] = -\sum_{i=0}^{k_1} \binom{s}{i} \alpha^i (1-\alpha)^{s-i}$$
$$+ \sum_{i=k_1}^{s} \binom{s}{i} \alpha^i (1-\alpha)^{s-i}$$
$$+ \sum_{i=k_1}^{k_2} \binom{s}{i} \alpha^i (1-\alpha)^{s-i} \left(\frac{2i-s}{sp}\right)$$

$\alpha = \Pr_y\left[b(y)\,C(x,y)\right]$

<u>Exercise</u>: $\forall x$

$$\left| \mathbb{E}_R\left[D(x,R)\right] - \frac{\mathbb{E}_x\left[b(y)\,C(x,y)\right]}{p} \right| \le \delta/p.$$

Completes the construction of randomized ckt D & hence claim $\boxtimes$

<u>Two function Version</u>

$b_1 : \{0,1\}^n \to \{\pm 1\}$, $b_2 : \{0,1\}^n \to \{\pm 1\}$ are (p, s_1)-hard & (q, s_2)-hard then $b_1 \cdot b_2$ is $(pq+\varepsilon, s')$-hard where $s' = \min\{\varepsilon^2 s_1 - O(1), s_2\}$

Induction on this: k-function version