6.842 Randomness and Computation April 16, 2014

Lecture 19

Lecturer: Ronitt Rubinfeld *Scribe: Maryam Aliakbarpour*

The goal of this lecture is to amplify the hardness of problems. In complexity theory, it has been shown that many problems are "hard" based on worst case analysis. i.e. there is no computationally efficient way to solve all instances. However, in cryptography we are looking for the problems that are "hard" on average. In this lecture, we will show that if a problem is hard on at least a constant fraction of its input, we can turn it into an average case hard problem based on Yao's Xor Lemma.

1 Preliminaries

Assume we are given a δ-biased coin where $\delta \leq \frac{1}{2}$ (i.e. $\Pr[head] = \delta$). If we simply predict "tail" for one toss of the coin, then we will be right with probability $1 - \delta$. We can predict the parity of k tosses with probability $\approx (1-2\delta)^k$. So, when k goes to the infinity, the probability of predicting correctly is one half (just like answering randomly).

So, it seems that solving k independent instances of a problem is harder than solving one. However, this is not true in general. Assume $f(x) = Ax$ where A is a $n \times n$ matrix and x is a vector of size n. Clearly, we can compute f in $\theta(n^2)$. Hence we'd expect that solving n instances would take $\theta(n^3)$. However, computing n instances of this function is exactly equivalent to multiplying two $n \times n$ matrices and we know that there are algorithms for matrix multiplication with running time of $o(n^3) < \theta(n^3)$.

Today all functions are from $\{+1,-1\}^n$ into $\{+1,-1\}$. Here is a definition that we need in this lecture.

Definition 1 *A function $f : \{+1,-1\}^n \to \{+1,-1\}$ is* δ-hard *on distribution $\mathcal{D}$ for size g, if for any Boolean circuit C with at most g gates, $\Pr_{x\in_{\mathcal{D}}\{+1,-1\}^n}[C(x) = f(x)] \leq 1-\delta$ i.e. any circuit of size at most g makes mistakes on at least δ fraction of the domain.*

Assume f is a δ-hard function; C is a circuit of size at most g; and $\mathcal{D}$ is the uniform distribution. If $\delta = 2^{-n}$ then $C(x) \neq f(x)$ for at least one input x. If $\delta = \frac{1}{2}$ then no circuit does better than random guessing. Obviously, we can always get $\delta \leq \frac{1}{2}$ with $C \equiv 1$ or $C \equiv -1$.

Recall that in the previous lecture, we had:

$$R_C(x) = \begin{cases} +1 & \text{if } C(x) = f(x) \\ -1 & \text{otherwise.} \end{cases}$$

Assume M is a measure on the domain $\{+1,-1\}^n$. Let $|M|$ denote $\sum_x M(x)$. We define $\mu(M) = \frac{|M|}{2^n}$. Moreover, $\mathcal{D}_M$ is a distribution such that for any x, $\mathcal{D}(x) = \frac{M(x)}{|M|}$. The advantage of M is denoted by $Adv_c(M)$ and it is equal to $\sum_x R_C(x)M(x)$.

Definition 2 *Consider a function $f : \{+1,-1\}^n \to \{+1,-1\}$ and a measure M. If for any circuit C of size at most g, $Adv_C(M) < \epsilon|M|$ (or equivalently $\Pr_{x\in_{\mathcal{D}_M}\{+1,-1\}^n}[C(x) = f(x)] \leq \frac{1}{2} + \frac{\epsilon}{2}$), then f is* ϵ-hardcore *on M for size g.*

Suppose the measure M in the above definition is a characteristic function of a set S i.e.

$$M(x) = \begin{cases} 1 & \text{if } x \in S, \\ 0 & \text{otherwise.} \end{cases}$$

Or equivalently, for an $x \in S$, $\mathcal{D}_M(x) = \frac{1}{|S|}$. So, we can define ϵ-hardcore based on set S as follows:

Definition 3 *A function $f : \{+1,-1\}^n \rightarrow \{+1,-1\}$ is* ϵ-hardcore on a set S for size g, *iff for any circuit C of size at most g*

$$\Pr_{x \in_U S}[C(x) = f(x)] \leq \frac{1}{2} + \frac{\epsilon}{2}$$

.

2 Yao's Xor Lemma

Theorem 4 *Assume $f : \{+1,-1\}^n \rightarrow \{+1,-1\}$ is δ-hard for size g on the uniform distribution. Then for an $\epsilon \in (0,1)$, there exists an M such that $\mu(M) \geq \delta$ and f is ϵ-hardcore on M for size $g' = \frac{\epsilon^2 \delta^2 g}{4}$.*

Proof Intuitively, the theorem says that if we have a function whom any circuit of size at most g computes wrongly some of the time, then for some distribution, any circuit of a bit smaller size than g will make mistakes almost half of the time (for good setting of ϵ).

To prove this theorem we follow the boosting outline. Assume there is no such M. Thus, for all M with $\mu(m) \geq \delta$, f is not ϵ-hardcore for size g'. By definition, there exists a circuit C of size g' where $\Pr_{x \in_{\mathcal{D}_M} \{+1,-1\}^n}[C(x) = f(x)] > \frac{1}{2} + \frac{\epsilon}{2}$ or equivalently, $Adv_C \geq \epsilon|M|$. This means that we have a weak learner for f on any input distribution $\mathcal{D}_{\mathcal{M}}$. By what we had in previous lecture, the majority function of $\frac{1}{\epsilon^2 \delta^2}$ circuits of size at most g' will predict f with error at most δ. The total size of these circuits is at most $\frac{g}{4}$ and we can construct the majority function with the $\frac{3g}{4}$ remaining gates. So, there is a circuit of size at most g that predicts f well and this contradicts the fact that f is δ hard. ■

Theorem 5 *If M is ϵ-hardcore measure for size $2n < g' < \frac{\epsilon^2 \delta^2}{8} \frac{2^n}{n}$, then there exists a 2ϵ-hardcore set S for f for size g' with $|S| \geq \delta 2^n$.*

Proof It has been shown in literature that the number of circuits of size g' is significantly less than $\frac{1}{4} e^{2^n . \epsilon^2 \delta^2}$. Pick S randomly according to $\mathcal{D}_{\mathcal{M}}$. Let M_S denote the characteristic measure of S. Since S is drawn form $\mathcal{D}_M$, $\text{Exp}_S[Adv_C(M_S)] = Adv_C(M) \leq \epsilon|M|$. It suffices to show that Pr[any C of size g' has $2\epsilon|M|$ advantage] is small. We know

$$\Pr[\text{any } C \text{ of size } g' \text{ has } 2\epsilon|M| \text{ advantage}] \quad \leq \#\mathit{circuits} \Pr[\text{a circuit of size } g' \text{ has } 2\epsilon|M| \text{ advantage}]$$

But we know that the expected $Adv_C(M_S)$ is less than $\epsilon|M|$. Thus, we can prove this by Chernoff bound. ■

Theorem 6 *[**Yao's XOR Lemma**] Given a function f, we can define $f^{\oplus k}(x_1, \ldots, x_k) \equiv f(x_1) \oplus \cdots \oplus f(x_k)$. If f is ϵ-hardcore for a set H of size at least $\delta 2^n$ for size $g+1$, then $f^{\oplus k}$ is $\epsilon + 2(1-\delta)^k$-hardcore for size g on all inputs.*

Proof Assume not. Then there exists a circuit C of size at most g gates such that

$$\Pr_{x_1,\ldots,x_k}[C(x_1,\ldots,x_k) = f^{\oplus k}(x_1,\ldots,x_k)] \geq \frac{1}{2} + \frac{\epsilon}{2} + (1-\delta)^k.$$

to prove the above theorem, we want to show that for any H such that $|H| \geq \delta 2^n$, we will get a circuit C' with at most $g+1$ gates which predicts f with probability greater than $\frac{1}{2} + \frac{\epsilon}{2}$ on H. So, f is not ϵ-hardcore.

Suppose E denotes the event that $C(x_1,\ldots,x_k) = f^{\oplus k}(x_1,\ldots,x_k)$. Let A_m denote the event that exactly m of $x_1,\ldots,x_k$ are in H. Since $|H| \geq \delta 2^n$, $\Pr_{x_1,\ldots,x_k}[A_0] \leq (1-\delta)^k$. Also, we know:

$$\begin{aligned}
\mathrm{Pr}_{x_1,\ldots,x_k}[E] \quad &= \mathrm{Pr}_{x_1,\ldots,x_k}[E \text{ and } A_0] + \mathrm{Pr}_{x_1,\ldots,x_k}[E \text{ and } \overline{A_0}] \\
&\leq \mathrm{Pr}_{x_1,\ldots,x_k}[A_0] + \mathrm{Pr}_{x_1,\ldots,x_k}[E \text{ and } \textstyle\bigcup_{m>0} A_m] \\
&\leq (1-\delta)^k + \mathrm{Pr}_{x_1,\ldots,x_k}[E \text{ and } \textstyle\bigcup_{m>0} A_m] \\
&\leq (1-\delta)^k + \sum_{m>0} \mathrm{Pr}_{x_1,\ldots,x_k}[E \text{ and } A_m] \\
&\leq (1-\delta)^k + \sum_{m>0} \mathrm{Pr}_{x_1,\ldots,x_k}[E|A_m]\, \mathrm{Pr}_{x_1,\ldots,x_k}[A_m]
\end{aligned}$$

Recall that $\mathrm{Pr}_{x_1,\ldots,x_k}[E] \geq \frac{1}{2} + \frac{\epsilon}{2} + (1-\delta)^k$, we have $\sum\limits_{m>0} \mathrm{Pr}_{x_1,\ldots,x_k}[E|A_m]\, \mathrm{Pr}_{x_1,\ldots,x_k}[A_m] \geq \frac{1}{2} + \frac{\epsilon}{2}$. Consider that $\sum\limits_{m>0} \mathrm{Pr}_{x_1,\ldots,x_k}[A_m] = 1 - \mathrm{Pr}_{x_1,\ldots,x_k}[A_o] \leq 1$. Thus, by averaging, there exists an m such that

$$\Pr_{x_1,\ldots,x_k}[C(x_1,\ldots,x_k) = f^{\oplus k}(x_1,\ldots,x_k)|A_m] \geq \frac{1}{2} + \frac{\epsilon}{2}.$$

Now, we can construct an idealized circuit to compute $C(x)$ for x drawn from uniform distribution on H.

Idealized circuit:

1. Pick $x_1, \ldots, x_{m-1} \in_R H$
2. Pick $y_{m+1}, \ldots, y_k \in_R \overline{H}$
3. Permute $x_1, \ldots, x_{m-1}, x, y_{m+1}, \ldots, y_k$ via random permutation π. Let $z = \pi(x_i'$s, y_i's, $x)$.
4. Call C on z.
5. Output $C(z) \oplus b$ where $b \in \{+1, -1\}$.

Note that if $x \in H$, exactly m elements in z are in H i.e. A_m is happening. Thus,

$$\Pr_z[C(z) = f^{\oplus k}(z)] = \Pr_{x_1,\ldots,x_k}[C(x_1,\ldots,x_k) = f^{\oplus k}(x_1,\ldots,x_k)|A_m] \geq \frac{1}{2} + \frac{\epsilon}{2}.$$

By averaging again, there exist $x_1, \ldots, x_{m-1}, y_{m+1}, \ldots, y_k$, and π such that

$$\Pr_x[C(z) = f^{\oplus k}(z)] \geq \frac{1}{2} + \frac{\epsilon}{2}.$$

Now, let $b = f(x_1) \oplus \cdots \oplus f(x_{m-1}) \oplus f(y_{m+1}) \oplus \cdots \oplus f(y_k)$. So, if $C(z) = f(z)$, then $f(x) = C(z) \oplus b$. Note that we may not know what b is. However, we know that for fixed $x_1, \ldots, x_{m-1}, y_{m+1}, \ldots, y_k$, there is a $b \in \{+1, -1\}$ such that this equality holds. So, let C' be the circuit that always outputs $C(z) \oplus b$. Consider b as its internal parameters. As we showed above,

$$\Pr_x[C'(x) = f(x)] \geq \frac{1}{2} + \frac{\epsilon}{2}$$

. Also, C' has only one gate more than C (because of XORing with a constant b). Therefore, C' is of size $g+1$ which contradicts the fact that f is ϵ-hardcore for H. Hence, the proof is complete. ■