Lecture 24: Goldreich-Levin Hardcore Predicate

Goldreich-Levin Hardcore Predicate: Intuition

- A One-way Function: A function that is easy to compute but hard to invert (efficiently)
- Hardcore-Predicate: A secret bit that is hard to compute

Theorem (Goldreich-Levin)

If $f : \{0,1\}^n \to \{0,1\}^n$ is a one-way function then it is hard to predict $b = r \cdot x$ given $(r, f(x))$, where $r, x \sim \mathbb{U}_n$

We will prove the contrapositive of this statement: If we can predict b given $(r, f(x))$, then we can efficiently invert f.

Prediction Experiment

- This is a game between two parties: honest challenger $\mathcal{H}$ and an adversary $\mathcal{A}$
- The honest challenger picks $x \sim \mathbb{U}_n$ and $r \sim \mathbb{U}_n$, computes $b = r \cdot x$, computes $y = f(x)$, and sends (r, y) to the adversary $\mathcal{A}$
- The adversary $\mathcal{A}$ replies back with a bit $\widetilde{b}$ (this is the guess of the adversary $\mathcal{A}$ of the hidden bit b with the honest challenger $\mathcal{H}$)
- The honest challenger outputs $z = 1$ if and only if $b = \widetilde{b}$; otherwise $z = 0$ ($z = 1$ represents the case that $\mathcal{A}$ has successfully predicted the bit b)

- Note that it is very easy to predict any bit with probability $1/2$ (the adversary $\mathcal{A}$ can always reply with a uniformly random bit $\widetilde{b}$, and we will have $b = \widetilde{b}$ with probability $1/2$)
- So, the adversary *actually* wins only when it can predict the bit with probability more than $1/2$

Definition (Advantage)

We say that an adversary $\mathcal{A}$ has advantage $\varepsilon > 0$ in predicting the bit if $\mathbb{P}[z = 1] \geqslant 1/2 + \varepsilon$

Prediction Experiment III

- So, we have the technical mechanism to formulate the statement "If we can predict b" in the contrapositive of the Goldreich-Levin result
- We will say that: Suppose there exists an adversary $\mathcal{A}$ that has advantage $\varepsilon > 0$ in predicting the bit b in the prediction experiment

One-way Function Experiment

- The experiment is between an honest challenger $\mathcal{H}$ and an adversary $\mathcal{B}$
- The honest challenger samples $x \sim \mathbb{U}_n$ and sends $y = f(x)$ to $\mathcal{B}$
- The adversary $\mathcal{B}$ replies with $\widetilde{x}$ (the adversary's guess of the pre-image of y)
- The honest challenger $\mathcal{H}$ outputs $z = 1$ if $f(\widetilde{x}) = y$; otherwise $z = 0$

Note that an adversary $\mathcal{B}$ wins if it predicts *any* pre-image of y (this need not necessarily be x)

One-way Function Experiment II

- To show that a function f is easy to invert, we need to demonstrate the existence of an adversary $\mathcal{B}$ who can invert f with significant probability, i.e. $\mathbb{P}[z = 1]$ is significant
- In the contrapositive of Goldreich-Levin result, we will construct a $\mathcal{B}$ such that $\mathbb{P}[z = 1] = \mathrm{poly}(\varepsilon)$

Contrapositive of Goldreich-Levin

- Suppose there exists $\mathcal{A}$ such that the advantage of $\mathcal{A}$ in the prediction experiment is ε,
- Then there exists $\mathcal{B}$ (with running time $\mathrm{poly}(t(\mathcal{A}), \varepsilon^{-1})$) such that the probability $\mathcal{B}$ successfully inverts f is at least $\mathrm{poly}(\varepsilon)$

Let us think how to construct $\mathcal{B}$

- $\mathcal{B}$ will participate in the one-way function experiment
- $\mathcal{B}$ will be given y as input
- We can use the adversary $\mathcal{A}$ to construct our adversary $\mathcal{B}$
- A Simplifying Assumption: Suppose that for all y, the adversary $\mathcal{A}$ takes two inputs (r, y) and it will correctly predict $r \cdot x$ with probability $1/2 + \varepsilon$
  - Think of $\mathcal{A}(\cdot, y)$ as a function that takes r as input and its output agrees with $\chi_x(r)$ for $1/2 + \varepsilon$ fraction of the total possible values of r
  - Recall: This is identical to the list decoding of the Hadamard Code. We are given a function H that agrees with $1/2 + \varepsilon$ fraction of the inputs with some χ_S. And, we are interested in recovering S. We will think of $\mathcal{A}(\cdot, y) \equiv H$ and $x \equiv S$. Now, list decoding of the Hadamard Code gives us a list L such that $S \in L$ with probability $1/2$.
  - So, $\mathcal{B}$ can run the list decoding algorithm for Hadamard Code with oracle $\mathcal{A}(\cdot, y)$ and output a random element of L. With probability $1/2|L|$ the output will be identical to x.

- So, we have the following problem. We are guaranteed that the winning probability of $\mathcal{A}$ in the prediction experiment is $1/2+\varepsilon$ when $x \sim \mathbb{U}_n$. The adversary $\mathcal{A}$ is *not* guaranteed to have winning probability $1/2+\varepsilon$ for *every* x
- We will show that there are a small fraction of inputs for whom this holds

Reduction to the Simplifying Assumption II

- Let p_x denote the probability that $\mathcal{A}$ successfully predicts $r \cdot x$ in the prediction experiment, over $r \sim \mathbb{U}_n$
- We have:
$$\mathop{\mathbb{E}}_{x \sim \mathbb{U}_n} [p_x] \geqslant 1/2 + \varepsilon$$
And, we want to say that p_x is high with some probability.
- So, we consider:
$$\mathop{\mathbb{E}}_{x \sim \mathbb{U}_n} [1 - p_x] \leqslant 1/2 - \varepsilon$$
- By Markov inequality, we have:
$$\mathop{\mathbb{P}}_{x \sim \mathbb{U}_n} [1 - p_x \geqslant t] \leqslant \frac{\mathbb{E}_{x \sim \mathbb{U}_n} [1 - p_x]}{t} \leqslant \frac{1/2 - \varepsilon}{t}$$
- Choose $t = 1/2 - \varepsilon/2$

- So, we get

$$\underset{x\sim\mathbb{U}_n}{\mathbb{P}}\left[1-p_x \geqslant 1/2-\varepsilon/2\right] \leqslant \frac{1/2-\varepsilon}{1/2-\varepsilon/2} = \frac{1-2\varepsilon}{1-\varepsilon}$$

- Equivalently,

$$\underset{x\sim\mathbb{U}_n}{\mathbb{P}}\left[p_x \leqslant 1/2+\varepsilon/2\right] \leqslant \frac{1-2\varepsilon}{1-\varepsilon} \leqslant 1-\varepsilon$$

- Equivalently,

$$\underset{x\sim\mathbb{U}_n}{\mathbb{P}}\left[p_x \geqslant 1/2+\varepsilon/2\right] \geqslant \varepsilon$$

- So, for ε fraction of the inputs, the success probability p_x of the adversary $\mathcal{A}$ is at least $\varepsilon/2$
- So, for these ε fraction of input, our strategy of constructing $\mathcal{B}$ will recover x with probability $1/2|L|$

Final One-way Function Inversion Adversary

- Our adversary $\mathcal{B}$ on input y runs the Hadamard Code list decoding algorithm with the oracle $\mathcal{A}(\cdot, y)$
- Let L be the list output by the list decoding algorithm
- Return a random element in L

Note that we successfully invert x with probability $\varepsilon/2|L|$. And we will see that the size of the list L is $\mathrm{poly}(n, 1/\varepsilon)$