CS 830: Randomness and Complexity 03/03/05

Lecture 14: Impagliazzo-Zuckerman Generator

instructor: Jin-Yi Cai Scribe: Anand Kumar Sinha, Vinay Choudhary

1 Introduction

In the last lecture we stated the leftover hash lemma. In this lecture we prove the lemma and then use it to construct a pseudo random generator.

Lemma 1.1. *(Leftover Hash Lemma). Let H be a universal family of hash functions, with $h_s(y)$: $Y \to T, s \in S$, and $Y' \subseteq Y$. Let P be the distribution $s \circ h_s(y)$, where s and y are sampled uniformly from S and Y' respectively. Then*

$$c(P) \leq \frac{1 + |T|/|Y'|}{|S \times T|}$$

Proof. We need to estimate the collision probability $c(P)$.

$$\begin{aligned}
c(P) &= Pr_{s,s',y,y'}[s \circ h_s(y) = s' \circ h_{s'}(y')] \\
&= \frac{1}{|S|} Pr_{s,y,y'}[s \circ h_s(y) = s \circ h_s(y')] \\
&= \frac{1}{|S|}(Pr_{s,y,y'}[s \circ h_s(y) = s \circ h_s(y')|y = y']Pr_{y,y'}[y = y'] \\
&\quad + Pr_{s,y,y'}[s \circ h_s(y) = s \circ h_s(y')|y \neq y']Pr_{y,y'}[y \neq y']) \\
&= \frac{1}{|S|}(Pr_{s,y,y'}[h_s(y) = h_s(y')|y = y']Pr_{y,y'}[y = y'] + \\
&\quad Pr_{s,y,y'}[h_s(y) = h_s(y')|y \neq y']Pr_{y,y'}[y \neq y']) \\
&\leq \frac{1}{|S|}(\frac{1}{|Y'|} + \frac{1}{|T|}) \\
&= \frac{1 + |T|/|Y'|}{|S \times T|}
\end{aligned}$$

□

2 Impagliazzo-Zuckerman Generator

In this section we will outline the construction of a pseudo random generator which uses the above lemma. Observe that the distribution P obtained above is close to uniform. However we haven't used all the "randomness" in the value y which was sampled from a uniform distribution over Y'. In the information theoretic sense, it had $log_2(|Y'|)$ bits of randomness . If we are interested in derandomzing an RP algorithm, then the hashed value just indicates if it is in the witness set or not. In a vague sense we have extracted only one bit of randomness from the source. There are still $log_2(|Y'|) - 1$ bits of randomness left in the hashed value. We can extract this by re-hashing the value obtained in the first stage by another hash function chosen uniformly at random. Since the distribution after the first stage was very close to uniform, the new distribution will again be nearly uniform. The whole idea is to iterate this process many times. The hash function chosen at each stage is sampled from a uniform distribution over the hash family. This is the extra randomness that we will need apart from that initially used in sampling uniformly from the set Y'. Now we state and prove the main theorem due to Impagliazzo and Zuckerman.

THEOREM 2.1. *Let H be a universal family of hash functions, with $h_s(y) : \{0,1\}^n \rightarrow \{0,1\}^{n-l}, s \in \{0,1\}^l$, and $W_x' \subseteq \{0,1\}^n$. Let ϵ be $2^{-\frac{l}{2}}$. Define $G_k(y, z_1, z_2, \ldots, z_{k-1}) = \{y_1, y_2, \ldots, y_k\}$ where,*

$$\begin{aligned} y_1 &= y, \\ y_2 &= h_{z_1}(y_1) \circ z_1, \\ y_3 &= h_{z_2}(y_2) \circ z_2, \\ &\vdots \\ y_k &= h_{z_{k-1}}(y_{k-1}) \circ z_{k-1}, \end{aligned}$$

Then, $Pr_{y,z_1,z_2,\ldots,z_{k-1}}[G_k \subseteq \overline{W_x}] \leq \mu(\overline{W_x})^k + \frac{\epsilon}{2}$, where y and $z_1, z_2, \ldots z_{k-1}$ are sampled uniformly from $\{0,1\}^n$ and $\{0,1\}^l$ respectively.

First observe that G_k is indeed a pseudo-random generator for an RP type algorithm. Here W_x can be thought of as the witness set of an input x whose membership is being tested by some RP algorithm. Without loss of generality we can assume that $\mu(W_x) \geq \frac{1}{2}$.The theorem states that we can amplify the success probability by an exponent of k using $G_k(\cdot)$ as a pseudo-random generator. Also note that the total number of random bits we need is $n + (k-1)l$ for amplifying the success probability to $\frac{\epsilon}{2} + 2^{-k}$. If we use $k = log_2(\epsilon) - 1$ then we can reduce the one way error to ϵ. Total number of random bits needed is $n + O(log^2(\epsilon))$. Now, we give a prove the main theorem.

Proof. For the sake of notational simplicity we will drop the subscripts in all probability expressions. It will be clear from the context as to what we are taking the probability over. We need to estimate the probability of the event $G_k \subseteq \overline{W_x}$. We will prove the result inductively with the induction being on k. Base case is easy to see.

$$\begin{aligned} Pr[G_k \subseteq \overline{W_x}] &= Pr[\{y_1, y_2, \ldots, y_k\} \subseteq \overline{W_x}], \\ &= Pr[y_1 \in \overline{W_x}] Pr[\{y_2, y_3, \ldots, y_k\} \subseteq \overline{W_x} | y_1 \in \overline{W_x}], \\ &= \mu(W_x) Pr[\{y_2, y_3, \ldots, y_k\} \subseteq \overline{W_x} | y_1 \in \overline{W_x}]. \end{aligned}$$

At this point we appeal to the previous lemma to estimate the second term of the above expression. Observe that the conditioning can be seen as y_2 being sampled from the distribution P of the Leftover Hash Lemma with the set Y' being $\overline{W_x}$. This is so because after the conditioning y_1 is uniformly distributed over $\overline{W_x}$. Hence $y_2 = h_{z_1}(y_1) \circ z_1$ is $\frac{1}{2}\sqrt{\frac{2^{n-l}}{2^n}} = \frac{\epsilon}{2}$ close to uniform. This is same as saying that,

$$Pr[\{y_2, y_3, \ldots, y_k\} \subseteq \overline{W_x} | y_1 \in \overline{W_x}] \leq \frac{\epsilon}{2} + Pr[\{\tilde{y_2}, y_3, \ldots, y_k\} \subseteq \overline{W_x}],$$

where $\tilde{y_2}$ is assumed to be sampled uniformly from $\{0,1\}^n$. Now we can apply our induction hypothesis by observing that

$$Pr[\{\tilde{y_2}, y_3, \ldots, y_k\} \subseteq \overline{W_x}] \quad = \quad Pr[G_{k-1} \subseteq \overline{W_x}].$$

It is easy to verify that the induction works. Here we use that fact that $\mu(W_x) \leq \frac{1}{2}$. □

In the next few lectures we will look at random walks on exapander graphs. In preparation we prove the following theorem.

THEOREM 2.2. *Let $G(V,E)$ be a d-regular graph on n vertices. Let A and B be two randomly chosen subsets of V of fixed cardinalities. Then $Exp[E(A,B)] = \frac{d}{n}|A| \cdot |B|$, where $E(A,B)$ denotes the number of edges between the sets A and B.*

Proof. Let $X_{ij} : 1 \leq i \leq |A|, 1 \leq j \leq |B|$ be the $0-1$ random variable which takes value 1 if there is an edge between the i^{th} vertex of A and the j^{th} vertex of B. Clearly $E(A,B) = \sum_{i,j} X_{ij}$. We claim that $Exp[X_{ij}] = \frac{d}{n}$. To see this just note that $Exp[X_{ij}]$ is the probability that there is an edge between the i^{th} vertex of A and the j^{th} vertex of B. Since we are randomly choosing the vertices, this probability is same as the probability that two randomly chosen vertices share an edge. Clearly probability of the latter event is $\frac{d}{n}$. The result follows by the linearity of expectation. □