

Lecture 20: 15-10-2024

Scribe: Priyanshu Singh

Lecturer: Rohit Gurjar

Definition of PRG

Let g be a function from $\{0, 1\}^l \rightarrow \{0, 1\}^m$. Then, g is said to be an $m(l)$ -PRG if the following conditions hold for all l :

1. For $r \in \{0, 1\}^l$, $g(r)$ can be computed in $2^{O(l)}$ time.
2. For all circuits C of size $\leq m(l)^3$ with m inputs,

$$|\Pr[x \in U_m : C(x) = 1] - \Pr[y \in U_l : C(g(y)) = 1]| \leq \frac{1}{10}$$

In other words, an algorithm that takes m -random inputs can be "fooled" by g .

Remark

If we do not care about the first condition, then a counting argument shows that a function g satisfying the second condition exists. The proof for this statement is left as homework.

Average Case Hardness (H_{avg})

A function $f : \{0, 1\}^k \rightarrow \{0, 1\}$ has $H_{\text{avg}}(f) = S$ if for any circuit C of size $\leq S$,

$$\Pr[x \in \{0, 1\}^k : f(x) = C(x)] \leq \frac{1}{2} + \frac{1}{S}$$

[NW96]

The result from [NW96] shows that if there exists a function $f : \{0, 1\}^k \rightarrow \{0, 1\}$ such that $H_{\text{avg}}(f) \geq S$, then there exists a pseudorandom generator $g : \{0, 1\}^l \rightarrow \{0, 1\}^m$ with stretch $m = \Omega(l)$, which is secure against circuits of size S .

Specifically, the generator g can be constructed from f , and the pseudorandomness property holds under the assumption that no circuit of size S can compute f correctly on more than $\frac{1}{2} + \frac{1}{S}$ fraction of inputs.

Moreover, if $f : \{0, 1\}^k \rightarrow \{0, 1\}$ is computable in $2^{O(k)}$ time and $H_{\text{avg}}(f) \geq 2^{2k/3}$, then there exists a $2^{l/45}$ -PRG.

Exponential Increase in Randomness (Yao's Construction)

We need to get an exponential increase in the amount of randomness. Yao showed that to go from l to $l+1$ random bits, we can construct the bits as follows:

$$r_1, r_2, r_3, \dots, r_l \rightarrow r_1, r_2, \dots, r_l, f(r_1, \dots, r_l)$$

Yao's Another Result

Yao also showed the following result:

Let D be a distribution on $\{0, 1\}^m$. Suppose that for all $1 \leq i \leq m - 1$ and for all circuits of size $\leq 2S$,

$$\Pr_{y \sim D} [C(y_1, \dots, y_i) = y_{i+1}] < \frac{1}{2} + \epsilon$$

Then, for all circuits B of size $\leq S$,

$$\left| \Pr_{y \sim D} [B(y) = 1] - \Pr_{y \sim U_m} [B(y) = 1] \right| < m\epsilon$$

Define intermediate distributions D_0, D_1, \dots, D_m between U_m and D as follows:

$$D_i : (y_1, \dots, y_i, z_{i+1}, \dots, z_m)$$

where y_1, \dots, y_i represent the first i bits sampled from D , and z_{i+1}, \dots, z_m represent the remaining $m - i$ bits sampled uniformly from U_m .

Now, suppose there exists a circuit B of size $\leq S$, such that:

$$\Pr_{y \sim D} [B(y) = 1] - \Pr_{y \sim U_m} [B(y) = 1] > m\epsilon$$

Then, we have:

$$\sum_{i=0}^{m-1} \left(\Pr_{y \sim D_{i+1}} [B(y) = 1] - \Pr_{y \sim D_i} [B(y) = 1] \right) > m\epsilon$$

By the averaging argument, there must exist some i such that:

$$\Pr_{y \sim D_{i+1}} [B(y) = 1] - \Pr_{y \sim D_i} [B(y) = 1] > \epsilon$$

(l,k,d)-Design and Related Claims

We define an (l, k, d) -Design as follows:

Let I_1, I_2, \dots, I_r be subsets of $\{1, 2, \dots, l\}$, such that for all j , $|I_j| = k$, and for all $j \neq j'$, we have $|I_j \cap I_{j'}| \leq d$.

Claim: For $k = \frac{l}{30}$ and $d = \frac{k}{3} = \frac{l}{90}$, there exists an (l, k, d) -Design with at least $2^{d/10} = 2^{l/900}$ subsets. Thus, we obtain polynomially many subsets.

Application to Pseudorandom Generator Construction

From [NW96], we know the following result:

For $l = 900 \log n$ (the number of true random bits) and for f satisfying $H_{\text{avg}}(f) \geq 2^{2k/3}$, we can define a function NW_l^f as:

$$NW_l^f = (f(Z_{I_1}), f(Z_{I_2}), \dots, f(Z_{I_r})),$$

where $f : \{0, 1\}^k \rightarrow \{0, 1\}$ and Z_{I_j} represents the input bits indexed by the set I_j .

Using this construction, we can obtain a pseudorandom generator that achieves stretch n^{10} , i.e., an $(n^{20/2})$ -PRG.

Theorem: For any circuit C of size $\leq n^{20}$,

$$\Pr [C(f(Z_{I_1}), f(Z_{I_2}), \dots, f(Z_{I_{i-1}})) = f(Z_{I_i})] \leq \frac{1}{2} + \frac{\epsilon}{n}$$

Proof: We prove this by contradiction. Assume there exists a circuit C of size $\leq S$ such that:

$$\Pr [C(f(Z_{I_1}), f(Z_{I_2}), \dots, f(Z_{I_{i-1}})) = f(Z_{I_i})] > \frac{1}{2} + \frac{\epsilon}{n}$$

Now, we will exploit the overlaps between different sets I_j 's to arrive at a contradiction.

Step 1: Overlap Structure and Fixing

By the structure of the (l, k, d) -Design, we know that for any $j \neq j'$, $|I_j \cap I_{j'}| \leq d$, where $d = \frac{l}{90}$.

This overlap condition allows us to "fix" some of the variables. Specifically, we will argue that there exists a fixing of some of the variables in Z such that the remaining free variables are contained in at most d indices.

Step 2: Reducing to a Circuit Finding f

By fixing the variables as described, we reduce the number of free variables that C depends on. This fixing reduces the complexity of the problem to a setting where C effectively becomes a circuit for finding f .

Let us denote the fixed subset of variables by Z' , which corresponds to the free variables in the sets I_j that remain after the fixing. Since the number of free variables is bounded by d , and $d = \frac{l}{90}$, the size of the circuit needed to compute f on these inputs becomes much smaller.

Step 3: Contradiction by Circuit Size

We now construct a new circuit C' that simulates the behavior of C on the fixed inputs. Since C was assumed to be of size $\leq n^{20}$, the circuit C' , which operates on a smaller input space (of size $d = \frac{l}{90}$), will also be small.

However, this contradicts the assumption that f is hard on average for circuits of size $\leq S$, where $S \geq 2^{2k/3}$. By assumption, no such circuit can predict f with probability significantly better than $\frac{1}{2}$.

Thus, the initial assumption that:

$$\Pr [C(f(Z_{I_1}), f(Z_{I_2}), \dots, f(Z_{I_{i-1}})) = f(Z_{I_i})] > \frac{1}{2} + \frac{\epsilon}{n}$$

must be false.

Thus, we conclude:

$$\Pr [C(f(Z_{I_1}), f(Z_{I_2}), \dots, f(Z_{I_{i-1}})) = f(Z_{I_i})] \leq \frac{1}{2} + \frac{\epsilon}{n}$$

This completes the proof.