Lecture 21: Extractors (Leftover Hash Lemma)

2-Universal Hash Function Family

- Let $\mathcal{F}_{n,m}$ be the set of all function $f\colon \{0,1\}^n \to \{0,1\}^m$
- H is a distribution over the sample space $\mathcal{F}_{n,m}$

Definition (2-Universal Hash Function Family)

For every distinct $x_1, x_2 \in \{0,1\}^n$, we have:

$$\mathbb{P}_{h\sim H}[h(x_1) = h(x_2)] \leqslant \frac{1}{2^m}$$

- We want that the sampling $h \sim H$ can be efficiently performed by a randomized algorithm that takes a sample from U_d
- Intuitively, two separate inputs collide under h at the same probability that they collide under a random function from $\mathcal{F}_{n,m}$

Theorem (LHL)

Let H be a 2-universal Hash Function Family. For any X that is an (n, k)-source, the following is true:

$$2\mathrm{SD}\Big((H, H(X))\ ,\ (H, \mathbb{U}_{\{0,1\}^m}) \Big) \leqslant \sqrt{\frac{M-1}{K}}$$

- That is, H is a good extractor for (n, k)-sources
- So, we need to construct the family H that can be sampled using only d-bits of randomness, and we want d to be as small as possible
- Note about the proof: We will see a more general Fourier-based proof, because there is another result, namely "Lopsided-LHL," that (as far as I know) cannot be proven using elementary combinatorial techniques

- We will use $M = 2^m$ and $K = 2^k$
- We will use U_m to represent the distribution $\mathbb{U}_{\{0,1\}^m}$

- We bound the SD as follows:

$$
\begin{aligned}
& 2\mathrm{SD}\left((H, H(X)), (H, U_m)\right) \\
&= \mathbb{E}_{h\sim H}\left[2\mathrm{SD}\left(h(X), U_m\right)\right] \\
&= \mathbb{E}_{h\sim H}\left[\sum_{y\in\{0,1\}^m} \left|h(X)(y) - U_m(y)\right|\right] \\
&\leqslant \mathbb{E}_{h\sim H}\left[M^{1/2}\left(\sum_{y\in\{0,1\}^m} \left(h(X)(y) - U_m(y)\right)^2\right)^{1/2}\right], \quad \text{Cauchy-Schwartz} \\
&= M\mathbb{E}_{h\sim H}\left[\sqrt{\left\|h(X) - U_m\right\|_2^2}\right] \\
&\leqslant M\sqrt{\mathbb{E}_{h\sim H}\left[\left\|h(X) - U_m\right\|_2^2\right]}, \quad \text{Jensen's}
\end{aligned}
$$

- Let us upper bound $\left\| h(X) - U_m \right\|_2^2$

$$
\begin{aligned}
&\left\| h(X) - U_m \right\|_2^2 \\
&= \sum_{S \in \{0,1\}^m} \left(\widehat{h(X) - U_m}\right)(S)^2, && \text{Parseval's} \\
&= \sum_{S \in \{0,1\}^m \colon S \neq \emptyset} \widehat{h(X)}(S)^2 \\
&= \sum_{S \in \{0,1\}^m} \widehat{h(X)}(S)^2 - \widehat{h(X)}(S = \emptyset)^2 \\
&= \left\| h(X) \right\|_2^2 - 1/M^2
\end{aligned}
$$

- So, we have the bound:

$$
2\mathrm{SD}\left((H, H(X)), (H, U_m)\right) \leqslant M \sqrt{\mathbb{E}_{h \sim H}\left[\left\| h(X) \right\|_2^2 - M^{-2}\right]}
$$

- So, it suffices to upper bound $\mathbb{E}_{h\sim H}\left[\left\|h(X)\right\|_2^2\right]$

$$
\begin{aligned}
&= \mathbb{E}_{h\sim H}\left[\left\|h(X)\right\|_2^2\right] \\
&= \mathbb{E}_{h\sim H}\mathbb{E}_{y\sim U_m}\left[h(X)(y)^2\right] \\
&= \mathbb{E}_{h\sim H}\mathbb{E}_{y\sim U_m}\left[\mathbb{P}\left[h(X^{(1)}) = y \ \wedge \ h(X^{(2)}) = y\right]\right] \\
&= \mathbb{E}_{h\sim H}\mathbb{E}_{y\sim U_m}\left[\mathbb{P}\left[X^{(1)} = X^{(2)}\right]\mathbb{P}\left[h(X^{(1)}) = h(X^{(2)}) = y | X^{(1)} = X^{(2)}\right]\right] \\
&\quad + \mathbb{E}_{h\sim H}\mathbb{E}_{y\sim U_m}\left[\mathbb{P}\left[X^{(1)} \neq X^{(2)}\right]\mathbb{P}\left[h(X^{(1)}) = h(X^{(2)}) = y | X^{(1)} \neq X^{(2)}\right]\right]
\end{aligned}
$$

- The first term:

$$\mathbb{P}\left[X^{(1)} = X^{(2)}\right] \mathbb{E}_{h\sim H} \frac{1}{M} \sum_{y\in\{0,1\}^m} \mathbb{P}\left[h(X^{(1)}) = h(X^{(2)}) = y | X^{(1)} = X^{(2)}\right]$$

$$= \mathbb{P}\left[X^{(1)} = X^{(2)}\right] \mathbb{E}_{h\sim H} \frac{1}{M} \mathbb{P}\left[h(X^{(1)}) = h(X^{(2)}) | X^{(1)} = X^{(2)}\right]$$

$$= \mathbb{P}\left[X^{(1)} = X^{(2)}\right] \mathbb{E}_{h\sim H} \frac{1}{M} \cdot 1$$

$$= \frac{1}{M} \cdot \mathbb{P}\left[X^{(1)} = X^{(2)}\right]$$

- Second Term:

$$\frac{1}{M} \cdot \mathbb{P}\left[X^{(1)} \neq X^{(2)}\right] \mathbb{E}_{h \sim H} \mathbb{P}\left[h(X^{(1)}) = h(X^{(2)}) | X^{(1)} \neq X^{(2)}\right]$$
$$\leqslant \frac{1}{M^2} \mathbb{P}\left[X^{(1)} \neq X^{(2)}\right]$$
$$= \frac{1}{M^2}(1 - \mathbb{P}\left[X^{(1)} = X^{(2)}\right])$$

- So, we have:

$$E_{h \sim H}\left[\left\| h(X) \right\|_2^2\right] - \frac{1}{M^2}$$
$$\leqslant \mathbb{P}\left[X^{(1)} = X^{(2)}\right]\left(\frac{1}{M} - \frac{1}{M^2}\right)$$
$$\leqslant \frac{1}{K}\left(\frac{1}{M} - \frac{1}{M^2}\right)$$

- So, overall we have:

$$2\mathrm{SD}\left((H, H(X)), (H, U_m)\right) \leqslant \sqrt{\frac{M}{K} - \frac{1}{K}}$$

- Hence the result