Lecture 22: Basic Applications of Fourier Analysis (Extractors and Leftover Hash Lemma)

Imperfect Randomness Sources

- A probability distribution X has min-entropy at least k if $\Pr[X = x] \leqslant 2^{-k}$, for all x in the sample space

Definition ((n,k)-Source)

A source over sample space $\{0,1\}^n$ with min-entropy at least k is known as an (n, k)-source.

- There are other specialized imperfect randomness sources like, bit-fixing sources, Santha-Vazirani sources
- Goal: Design an *extractor* to extract pure randomness from any min-entropy source from a class of sources
- For example, design an extractor that extracts pure randomness from any (n, k)-source

Definition ($(\mathcal{C}_n, m, \varepsilon)$-Extractor)

Let $\mathcal{C}_n$ be a class of imperfect randomness sources over the sample space $\{0,1\}^n$. A $(\mathcal{C}_n, m, \varepsilon)$-extractor is a function $\mathrm{Ext}\colon \{0,1\}^n \to \{0,1\}^m$ such that, for all $X \in \mathcal{C}_n$, we have $\mathrm{SD}\left(\mathrm{Ext}(X), U_m\right) \leqslant \varepsilon$.

- Such a function is also known as a deterministic extractor

Impossibility Result

Lemma (Negative Result)

Let $\mathcal{C}_n$ be the set of all $(n, n-1)$-sources. For any $\varepsilon < 1/2$, there does not exist a $(\mathcal{C}_n, 1, \varepsilon)$-extractor.

- This result is extremely strong. Even if the sources have $(n-1)$ min-entropy, we cannot extract even one bit that is close to uniform!
- If possible let there exists such an extractor Ext
- Let $P_b = \mathrm{Ext}^{-1}(b)$, for $b \in \{0, 1\}$
- Note that at least one of P_0 or P_1 is of size 2^{n-1}. Suppose $|P_{b^*}| \geqslant 2^{n-1}$
- Let X be the uniform distribution over the set P_{b^*}, represented by $U(P_{b^*})$, and $\Pr[X = x] \leqslant 2^{-(n-1)}$, for all $x \in \{0, 1\}^n$
- Note that $\mathrm{SD}\left(\mathrm{Ext}(X), U_1\right) = 1/2$

Anything to Salvage?

- Note that computing the distribution $U(P_{b^*})$ might be computationally inefficient. What if we restrict to distributions that are easy (or, efficient) to sample?

Lemma (Efficient Negative Result)

Let $\mathcal{C}_n$ be the sources that are samplable in time T (given uniform random bits as input) and have min-entropy at least $k = (n-1) - lg(3/2)$. Then, for all $\varepsilon < 1/4$ there does not exist any $(\mathcal{C}_n, 1, \varepsilon)$-extractor that has time complexity T', such that $T' \leqslant T - 2n - \Theta(1)$.

- Let P_b be the distribution that takes as input two uniform random strings $(r, r') \in \{0,1\}^{2n}$. If $\mathrm{Ext}(r) = b$, output r; otherwise output r'.

Proof

- This technique is known as *rejection sampling*, i.e. "keep rejecting the samples till you get something you desire, or (after a threshold number of sample draws) give up and output the final sample"
- The time complexity T to sample P_b is $T' + 2n + \Theta(1)$, hence the bound $T \geqslant T' + 2n + \Theta(1)$ is satisfied
- Let p_b be the probability of $\mathsf{Ext}(U_n) = b$
- Then, we have $\Pr[\mathsf{Ext}(P_b) = b] = p_b \cdot 1 + (1 - p_b) \cdot p_b = p_b(2 - p_b)$ and, similarly, $\Pr[\mathsf{Ext}(P_{\overline{b}}) = \overline{b}] = p_{\overline{b}}(2 - p_{\overline{b}})$
- Maximum of these two probabilities is at least $3/4$
- So, the statistical distance from U_1 of one of these two distributions is at least $1/4$
- That distribution will have maximum probability $2^{-k} \leqslant 2^{-(n-1)} + 2^{-n}$, and this satisfies the min-entropy bound

Take-away Message

- It is still possible to have the *complexity of the extractor* to be *significantly larger* than the *sampling complexity of the sources*
- There are positive results where good deterministic extractors exist when the class of sources are simple, for example, bit-fixing sources, affine sources, sources samplable by small-depth circuits
- In the computational setting, if *hard to invert functions* exist then we can construct an efficient extractor for sources samplable in time $p(n)$, where $p(\cdot)$ is a fixed polynomial
- A more general version of the above statement is considered by Nisan-Wigderson

Seeded Extractors

- These extractors take as inputs a uniform random string $s \sim U_d$ known as the seed
- Goal: Given this initial investment of pure d bits, we are interested in obtaining m pure random bits as output from k imperfect bits. We want $m \approx n + d$ and d to be as small as possible.

Definition (Strong Extractor)

A $(\mathcal{C}_n, d, m, \varepsilon)$-strong-extractor $\mathrm{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a function such that, for any $X \in \mathcal{C}_n$, we have:

$$\mathrm{SD}\left((U_d, \mathrm{Ext}(X, U_d)), (U_d, U_m)\right) \leqslant \varepsilon$$

- For $\mathcal{C}_n = (n,k)$-sources, our aim is to get $m \approx k$ and d as small as possible

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:

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

- 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 $\mathcal{F}_{n,m}$

Leftover Hash Lemma

Theorem (LHL)

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

$$\mathrm{SD}\left((H, H(X)), (H, U_m)\right) \leqslant \varepsilon,$$

where $2\varepsilon = \sqrt{2^{-(k-m)} - 2^{-k}}$

- 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

Proof

- We will use $M = 2^m$ and $K = 2^k$
- We bound the SD as follows:

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

Proof

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

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

- So, we have the bound:

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

Proof

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

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

Proof

- The first term:

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

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

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

$$\leqslant \frac{1}{M} \cdot \Pr[X^{(1)} = X^{(2)}]$$

Proof

- Second Term:

$$
\begin{aligned}
&\frac{1}{M} \cdot \Pr[X^{(1)} \neq X^{(2)}] \mathop{\mathbb{E}}_{h \sim H} \Pr[h(X^{(1)}) = h(X^{(2)}) | X^{(1)} \neq X^{(2)}] \\
&\leqslant \frac{1}{M^2} \Pr[X^{(1)} \neq X^{(2)}] \\
&= \frac{1}{M^2}(1 - \Pr[X^{(1)} = X^{(2)}])
\end{aligned}
$$

Proof

- So, we have:

$$\begin{aligned}
&E_{h\sim H}\left[\|h(X)\|_2^2\right]-\frac{1}{M^2}\\
&\leqslant \Pr[X^{(1)}=X^{(2)}]\left(\frac{1}{M}-\frac{1}{M^2}\right)\\
&\leqslant \frac{1}{K}\left(\frac{1}{M}-\frac{1}{M^2}\right)
\end{aligned}$$

- 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