CS 6815: Pseudorandomness and Combinatorial Constructions **Fall 2022**

Lecture 17: Oct 20, 2022

Lecturer: Eshan Chattopadhyay *Scribe: Ricky Shapley*

1 Recap: Seeded Extractors

We say that $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k, ϵ)-seeded extractor if for all (n, k)-sources X, $\mathrm{Ext}(X, U_d) \approx_\epsilon U_m$. We saw with using a random construction, we can achieve such extractors with

$$d = \log(n-k) + 2\log(1/\epsilon) + O(1)$$
$$m = d + k - 2\log(1/\epsilon) - O(1).$$

2 Randomized Algorithms with Weak Sources

Consider some language $L \in \mathbf{BPP}$ with some algorithm $\mathcal{A}$. Recall that this means for all inputs x,

$$\Pr_{r \sim U_m}[\mathcal{A}(x, r) = L(x)] \geq \frac{9}{10}.$$

But what if $\mathcal{A}$ only has access to (n, k)-sources? If Y is an (n, k)-source, we want to be able to construct some algorithm $\mathcal{A}'$ so that for all inputs x,

$$\Pr_{y \sim Y}[\mathcal{A}'(x, y) = L(x)] \geq \frac{2}{3}.$$

Our idea is to try all possible seeds. We will take a seeded extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$, where the first input is from our weak source Y. Let $r_i = \mathrm{Ext}(y, s_i)$ for all $i \in [D]$ where s_i is the ith element in $\{0,1\}^d$ under some fixed ordering. Then for each seed, we calculate $z_i = \mathcal{A}(x, r_i)$. Let z be the concatenation $z_1 z_2 \ldots z_D$ and output $\mathrm{Maj}(z)$.

Here, $\mathcal{A}'$ runs in $poly(n) \cdot D$ as long as we can compute Ext in polynomial time. We have D total seeds, and for each seed we need to run $\mathcal{A}$ which is $poly(n)$ and our extractor, which we assume to also be $poly(n)$.

Theorem 2.1. *$\mathcal{A}'$ as defined above satisfies*

$$\Pr_{y \sim Y}[\mathcal{A}'(x, y) = L(x)] \geq \frac{2}{3}.$$

Proof. Fix some input x. Let $\mathsf{Bad} = \{r \in \{0,1\}^m : \mathcal{A}(x, r) \neq L(x)\}$. Then by definition of $\mathcal{A}$, $\frac{|\mathsf{Bad}|}{M} \leq \frac{1}{10}$.

Which $y \in Y$ are bad? Each y can be mapped to D elements in $\{0,1\}^m$ when considering all possible seeds. So bad choices for y are those that map a majority of outputs to Bad. We will describe these as

$$\mathsf{Bad}_y = \{y \in \mathrm{supp}(Y) : |N(y) \cap \mathsf{Bad}| \geq D/2\}$$

where $N(y) = \{\mathrm{Ext}(y, s_1), \ldots, \mathrm{Ext}(y, s_D)\}$, the set of all possible outputs of y.

Then

$$\Pr[\mathcal{A}' \text{ fails on } x] = \Pr_{y \sim Y}[y \in \mathsf{Bad}_y] \leq \frac{|\mathsf{Bad}_y|}{2^k}$$

because Y is a (n, k)-source.

Now we wish to bound $|\mathsf{Bad}_y|$. Suppose that our extractor Ext is a (k', ϵ)-seeded extractor. We claim that $|\mathsf{Bad}_y| < 2^{k'}$.

Suppose for a contradiction $|\mathsf{Bad}_y| \geq 2^{k'}$. Let W be a distribution flat on Bad_y. So W is a (n, k')-source. Then

$$\Pr[\mathrm{Ext}(w, U_d) \in \mathsf{Bad}] \geq \frac{1}{2}$$

for every $w \in W$ by the definition of Bad and so

$$\Pr[\mathrm{Ext}(W, U_d) \in \mathsf{Bad}] \geq \frac{1}{2}.$$

And we know

$$\Pr[U_m \in \mathsf{Bad}] \leq \frac{1}{10}.$$

But this is a contradiction! Our extractor should guarantee that $\mathrm{Ext}(W, U_d)$ is very close to U_m, but

$$|\mathrm{Ext}(W, U_d) - U_m| \geq |\Pr[\mathrm{Ext}(W, U_d) \in \mathsf{Bad}] - \Pr[U_m \in \mathsf{Bad}]| \geq \frac{2}{5}.$$

So if we choose an extractor with $\epsilon = 1/4$, then $|\mathsf{Bad}_y| < 2^{k'}$. This means

$$\Pr[\mathcal{A}' \text{ fails on } x] \leq \frac{|\mathsf{Bad}_y|}{2^k} < 2^{k'-k}$$

and we can easily choose our extractor such that the failure probability is sufficiently small.

□

Note that choosing our seed length to be $d = O(\log(n/\epsilon))$ suffices here - as this means the runtime of our algorithm $\mathcal{A}'$ is *poly*(n).

3 Sampling

Suppose we have some function $f : \{0,1\}^m \to [0,1]$. We wish to estimate $\mu = \mathbb{E}_{x \sim U_m} f(x)$.

The standard method to do this is simple: we take $x_1, \ldots, x_D$ from U_m i.i.d., then compute $\tilde{\mu} = \frac{1}{D}\sum_{i \in [D]} f(x_i)$.

A standard application of the Chernoff bound gives

$$\Pr[|\mu - \tilde{\mu}| > \epsilon] < \delta$$

where $\delta = 2^{-\Omega(\epsilon^2 D)}$, or equivalently $D = O(1/\epsilon^2 \log(1/\delta))$.

Definition 3.1 (Sampler)**.** *Samp* $: \{0,1\}^n \to [M]^D$ *is a (k, ϵ, δ)-sampler if for all functions f : $[M] \to [0,1]$ and for all (n, k)-sources X,*

$$\Pr\left[\left|\frac{1}{D}\sum_{i=1}^{D} f(y_i) - \mu(f)\right| > \epsilon\right] < \delta$$

where $(y_1, \ldots, y_D) = \mathrm{Samp}(x)$ for $x \sim X$.

3.1 Construction

We start with a (k', ϵ')-extractor $\mathrm{Ext} : [N] \times [D] \to [M]$. Consider the natural bipartite graph representation of the extractor. We have $[N]$ nodes on the left, and $[M]$ nodes on the right. We connect a left node $x \in [N]$ to a right node $y \in [M]$ if there is some seed $r \in [D]$ that maps (x, r) to y. This is a left-regular bipartite graph with degree D.

Then $\mathrm{Samp}(x) = N(x)$, the neighbors of x in our graph. Or equivalently, $N(x)$ is the set $\{\mathrm{Ext}(x, r) : r \in [D]\}$.

We will defer the proof, but prove a claim that will be useful.

Claim 3.2. *Let $z \approx_\epsilon U_m$, then $|\mathbb{E}[f(z)] - \mu(f)| \leq 2\epsilon$.*

Proof. Using the definition of expectation,

$$\begin{aligned}
|\mathbb{E}[f(z)] - \mu(f)| &= \left| \sum_{z \in [M]} f(z)(\Pr[Z = z] - \Pr[U_m = z]) \right| \\
&\leq \sum_{z \in [M]} f(z) \left| \Pr[Z = z] - \Pr[U_m = z] \right| \\
&\leq \sum_{z \in [M]} \left| \Pr[Z = z] - \Pr[U_m = z] \right| \\
&= 2|z - U_m| \leq 2\epsilon
\end{aligned}$$

where the inequalities follow from the triangle inequality, the boundedness of f, and the definition of statistical distance. $\square$