CS 6815 Pseudorandomness and Combinatorial Constructions **Fall 2019**

Lecture 11: October 3

Lecturer: Eshan Chattopadhyay *Scribe: Sloan Nietert*

11.1 Overview

This lecture will cover the following topics:

- Probabilistic existence of seeded extractors
- Simulation of randomized algorithms using defective sources
- Universal hash functions (with explicit construction)
- Construct seeded extractors from hash functions

11.2 Existence of seeded extractors

Recall that we defined a seeded extractor to be a set of functions

$$\{\mathrm{Ext}_i : \{0,1\}^n \to \{0,1\}^m\}_{i \in \{0,1\}^d}$$

such that $|\mathrm{Ext}_{U_d}(X) - U_m| < \varepsilon$ for any (n, k)-source X.

Alternatively, we can view an extractor as a single function.

Definition 11.1 *We say that*

$$\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$$

is a (k, ε)-extractor if $|\mathrm{Ext}(X, U_d) - U_m| < \varepsilon$ for any (n, k)-source X.

Sometimes, we will need an even stronger notion.

Definition 11.2 *We say that*

$$\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$$

is a strong-seeded extractor if $|\,\mathrm{Ext}(X, U_d), U_d - U_m, U_d| < \varepsilon$.

This is equivalent to requiring that

$$\mathop{\mathbb{E}}_{y \sim U_d} |\mathrm{Ext}(X, y) - U_m| < \varepsilon.$$

In what follows, we will use the notation $D_1 \approx_\varepsilon D_2$ to mean that $|D_1 - D_2| < \varepsilon$. Next, we show the existence of seeded extractors.

Proof:

Let Ext : $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a random function, and let X be a flat (n,k)-source. Take $T \subseteq \{0,1\}^m$ to be a "test" subset for our extractor. The main idea of what follows is to take a union bound over all tests and flat sources. Note that all uppercase letters represent the base 2 exponential of their lowercase counterparts.

For any $x \in \{0,1\}^n$ and $y \in \{0,1\}^d$, let

$$Z_{x,y} = \begin{cases} 1, & \text{if } \mathrm{Ext}(x,y) \in T \\ 0, & \text{o.w.} \end{cases}$$

Next, define

$$W_T = \Pr_{x \sim X, y \sim U_d}\left[\mathrm{Ext}(x,y) \in T\right] = \frac{1}{KD} \sum_{x \in \mathrm{supp}(X), y \in \{0,1\}^d} Z_{x,y}.$$

Clearly, $\mathbb{E}[W_T] = |T|/M$. Noting that the $Z_{x,y}$ variables are i.i.d. and uniform, we can use Chernoff's bound to find

$$\Pr\left[|W_T - \mathbb{E}(W_T)| > \varepsilon\right] \leq \exp(-\Omega(\varepsilon^2 KD)).$$

Thus, we can bound the probability of Ext performing poorly on any test set or flat source by

$$2^M \binom{N}{K} \exp(-\Omega(\varepsilon^2 KD)).$$

We choose $m = k + d - 2\log(1/\varepsilon) - O(1)$ and bound $\binom{N}{K}$ by $(Ne/K)^K$ to see that

$$d \geq \log(n-k) + 2\log(1/\varepsilon) + O(1)$$

suffices to bring the error probability below 1 and guarantee the existence we desire. ■

11.3 Simulation of randomized algorithms

Suppose $\mathcal{A}$ is a randomized algorithm such that $\mathcal{A}(\cdot, U_m)$ is incorrect with probability at most ε, and suppose we have access to an (n,k)-source X with $k > m$.

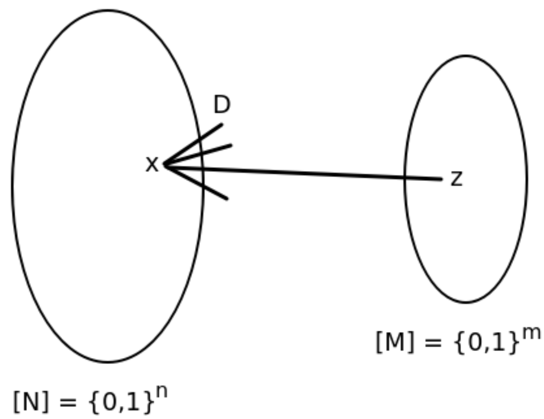
If Ext is a (k,δ)-extractor, we can define a new algorithm

$$\mathcal{A}'(\cdot, X) = \mathrm{maj}_{y \in \{0,1\}^d}\{\mathcal{A}(\cdot, \mathrm{Ext}(X,y))\}$$

Clearly, the time blow up is a factor of D.

Claim 11.3 *$\mathcal{A}'$ has error at most $2(\varepsilon + \delta)$.*

Proof: We now introduce a graph interpretation of the seeded extractor Ext : $\{0,1\}^m \times \{0,1\}^n \to \{0,1\}^m$,



where the above represents a D left-regular bipartite graph with x and z joined if and only if there exists $y \in \{0,1\}^d$ such that $\mathrm{Ext}(x,y) = z$.

Define the set $\mathrm{Bad} = \{z \in \{0,1\}^m : \mathcal{A}(\cdot, z) \text{ is wrong}\}$. Clearly, $|\mathrm{Bad}| \leq \varepsilon M$. Next, define

$$\mathrm{Bad}_X = \{x \in \mathrm{supp}(X) : |\Gamma(x) \cap \mathrm{Bad}| > D/2\},$$

so that $\mathcal{A}'(\cdot, x)$ is wrong if and only if $x \in \mathrm{Bad}_X$. Then, because Ext is an extractor, we have that $|\mathrm{Bad}|/M \leq \varepsilon$, and if $x \in \mathrm{Bad}_X$, then

$$\Pr[\mathrm{Ext}(x, U_d) \in \mathrm{Bad}] > \frac{1}{2}.$$

Putting everything together, we have

$$\Pr_{x \sim X}[x \in \mathrm{Bad}_X] \leq 2 \Pr\big[\mathrm{Ext}(X, U_d) \in \mathrm{Bad}\big] \leq 2\left(\frac{|\mathrm{Bad}|}{M} + \delta\right) \leq 2(\varepsilon + \delta).$$

Thus, the error of $\mathcal{A}'$ is bounded by $2(\varepsilon + \delta)$, as desired. ■

11.4 Universal hash functions

Definition 11.4 *We say that $\mathcal{H} = \{h_i\}_{i \in I}$, $h_i : [N] \to [M]$, is a universal hash function if for all $x \neq y \in [N]$*

$$\Pr_{h \in H}[h(x) = h(y)] \leq \frac{1}{M}.$$

Let $\mathbb{F}_q$ be a finite field with $q = 2^n$, and define

$$h_a : x \mapsto a \cdot x \pmod{2^m}$$

for each $a \in \mathbb{F}_q$.

Claim 11.5 *$\mathcal{H} = \{h_a\}_{a \in \mathbb{F}_q}$ is a universal hash function.*

Proof: Take x, y distinct in $\mathbb{F}_q$. Then,

$$\Pr_{h \in H}[h(x) = h(y)] = \Pr_{a \in \mathbb{F}_q}\big[ax = ay \pmod{2^m}\big] = \Pr_{a \in \mathbb{F}_q}\big[az = 0 \pmod{2^m}\big] = 1/M,$$

where $z = x - y \neq 0$. ■

11.5 Constructing extractors from hash functions

Definition 11.6 *We define the collision probability*

$$\mathrm{cp}(D) = \Pr[D = D'],$$

where D and D' are independent copies of D.

Observe that

$$\mathrm{cp}(D) = \sum_{x \in \Omega} D(x)^2.$$

Lemma 11.7 *If D is a distribution on $\{0,1\}^n$ with $\mathrm{cp}(D) \leq \frac{1+\varepsilon}{N}$, then*

$$|D - U_n| \leq \sqrt{\varepsilon}$$

Proof: We have

$$2|D - U_n| = \|D - U_n\|_1 \leq \sqrt{N}\|D - U_n\|_2$$

by Cauchy-Schwartz, and

$$\|D - U_n\|_2^2 = \sum_{x \in \{0,1\}^n} (D(x) - U(x))^2 = \sum_{x \in \{0,1\}^n} D(x)^2 - 2 \sum_{x \in \{0,1\}^n} D(x)\frac{1}{N} + \frac{1}{N^2}N = cp(D) - \frac{1}{N},$$

from which the desired inequality follows. ■

(to be continued in next lecture)