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

Lecture 16: Oct 19, 2022

*Lecturer: Eshan Chattopadhyay* — *Scribe: Atul Ganju*

# 1 Randomness Extractors

## 1.1 Deterministic Randomness Extractors

In our definition of a randomness extractor, the extractor must work for every source in some family of sources $\mathcal{X}$. For this reason, we were able to show that there exists no deterministic extractor for even $(n, n-1)$ sources. However, if we were to instead only require the randomness extractor to work for a specific source, then we can show that a random function will work as an extractor for that source with high probability.

Interestingly, flat $(n, k)$-sources, which are uniform distributions over a set $S \subseteq \{0,1\}^n$ with $|S| = 2^k$, are really representative of general $(n, k)$ sources:

**Claim 1.1.** *Any $(n, k)$-source is a convex combination of flat $(n, k)$ sources.*

*Proof.* Let $X$ be an $(n, k)$-source. Then, since we can view any random variable taking values in $[2^n]$ as a unique vector of dimension $2^n$ where the $i$-th coordinate is the probability the random variable takes the value $i$, we know that $X$ can be uniquely represented as some vector $v$ of dimension $2^n$ where, for each $i \in [2^n]$, $v_i \in [0, 2^{-k}]$, and $\sum_i v_i = 1$. The set of vectors that satisfy these constraints, and therefore uniquely represent some $(n, k)$-source, form the convex polytope in $\mathbb{R}^{2^n}$ that has the set of corners $\{\sum_{i \in S} 2^{-k} e_i : S \subseteq n, |S| = 2^k\}$. As these corners are the set of flat sources, by the convexity of the polytope, we have that $X$, and therefore any $n, k$-source, is a convex combination of flat $(n, k)$-sources. □

As a result, if we define for all $S \subseteq [2^n]$ of cardinality $2^k$ the flat source $X_S$ over the subset $S$, then to sample from an arbitrary $(n, k)$ source $X = \sum_S \lambda_S X_S$ where each $\lambda_S \in [0, 1]$, we can sample from $X_S$ with probability $\lambda_S$. Therefore, if we can create probabilistic algorithms to work for flat $(n, k)$-sources, then we can make them work for any $(n, k)$-source.

Now back to showing that for any $(n, k)$-source, a random function will work as an extractor with high probability. As shown above, it suffices to show this for flat $(n, k)$-sources.

**Theorem 1.2.** *For every $n, m, k \in \mathbb{N}$, every $\epsilon > 0$, and every flat $(n, k)$-source $X$, if we choose a random function $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}^m$ with $m = k - 2\log(1/\epsilon) - O(1)$, then with probability $1 - 2^{-\Omega(2^k \epsilon^2)}$, we have:*

$$|\mathrm{Ext}(X) - U_m| \leq \epsilon,$$

*where $U_m$ is a uniform random variable on $\{0,1\}^d$.*

*Proof.* As stated above, it suffices to show that there exists an extractor for the family of flat $(n, k)$-sources. Take a flat $(n, k)$-source $X$ and denote $S \subseteq \{0,1\}^n$ as its support. If we randomly chose Ext, then for any $x \in S$ and $T \subseteq \{0,1\}^m$, we have that the probability that $\mathrm{Ext}(x) \in T$ is

$|T| \cdot 2^{-m}$, where these events are independent. Therefore, we have:

$$\Pr[\mathrm{Ext}(X) \in T] = \frac{1}{2^k} \sum_{x \in S} \mathbb{1}\{\mathrm{Ext}(x) \in T\},$$

where by the Chernoff bound, we know that:

$$\Pr\left[\left|\frac{1}{2^k} \sum_{x \in S} \mathbb{1}\{\mathrm{Ext}(x) \in T\} - \frac{|T|}{2^m}\right| > \epsilon\right] \leq 2^{-\Omega(2^k \epsilon^2)}.$$

As this is for a and specific $T$, we can union bound over all $2^{2^m}$ possible $T$ to get that:

$$\Pr\left[|\mathrm{Ext}(X) - U_m| > \epsilon\right] \leq 2^{2^m} 2^{-\Omega(2^k \epsilon^2)},$$

which, for $m = k - 2\log(1/\epsilon) - O(1)$, gives us our desired result. We leave showing this as an exercise to the reader. □

One would hope that e could get an extractor that was good for all flat $(n, k)$-sources with another union bound; however, since the number of flat $(n, k)$-sources is $\binom{2^n}{2^k} \approx 2^{n2^k}$, this would fail atrociously. Therefore, in order to overcome the shortcomings of deterministic randomness extractors, we look towards seeded randomness extractors.

## 2 Seeded Extractors

Seeded extractors are more powerful than regular extractors because, instead of using one deterministic function as a randomness extractor for a family of sources $\mathcal{X}$, a seeded extractor is a randomized function which extracts randomness from $\mathcal{X}$. One can think of this randomized function as a family of deterministic functions that are randomly chosen depending on a sequence of coin flips. Formally, a seeded extractor and a strong seeded extractor are defined as follows:

**Definition 2.1** (seeded extractor)**.** *Take some family of distributions $\mathcal{X}$ on $\{0,1\}^n$. Then the function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a $(k, \epsilon)$-seeded extractor for $\mathcal{X}$, if for any $X \in \mathcal{X}$ we have that:*

$$|\mathrm{Ext}(X, U_d) - U_m| \leq \epsilon,$$

*where $U_d$ and $U_m$ are uniform random variable on $\{0,1\}^d$ and $\{0,1\}^m$ respectively.*

**Definition 2.2** (strong seeded extractor)**.** *Take some family of distributions $\mathcal{X}$ on $\{0,1\}^n$. Then the function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a $(k, \epsilon)$-strong seeded extractor for $\mathcal{X}$, if for any $X \in \mathcal{X}$ we have that:*

$$|(\mathrm{Ext}(X, U_d), U_d) - (U_m, U_d)| \leq \epsilon,$$

*where $(a, b)$ denotes concatenation of $b$ onto $a$.*

Just by providing a little randomness, we will be able to show that seeded extractors exist for many interesting families of sources. Specifically, we have the following existence theorem for seeded extractors:

**Theorem 2.3.** *For every $n \in \mathbb{N}, k \in \{0, \ldots, n\}, \epsilon > 0$, there exists a $(k, \epsilon)$-seeded extractor* $\mathrm{Ext}$ : $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *with $m = k + d - 2\log(1/\epsilon) - O(1)$ and $d = \log(n-k) + 2\log(1/\epsilon) + O(1)$.*

*Proof.* As stated above, it suffices to show that there exists an extractor for the family of flat $(n, k)$-sources. Take a flat $(n, k)$-source $X$ and denote $S \subseteq \{0,1\}^n$ as its support. If we follow the probabilistic method proof strategy in Theorem 1.2, we have that if we randomly chose Ext, then for any $x \in S, y \in \{0,1\}^d$, and $T \subseteq \{0,1\}^m$, the probability that $\text{Ext}(x, y) \in T$ is $|T| \cdot 2^{-m}$, where these events are independent. Therefore, we have:

$$\Pr[\text{Ext}(X, U_d) \in T] = \frac{1}{2^k 2^d} \sum_{x,y} \mathbb{1}\{\text{Ext}(x, y) \in T\},$$

where by the Chernoff bound, we know that:

$$\Pr\left[\left|\frac{1}{2^k 2^d} \sum_{x,y} \mathbb{1}\{\text{Ext}(x, y) \in T\} - \frac{|T|}{2^m}\right| > \epsilon\right] \leq 2^{-\Omega(2^k 2^d \epsilon^2)}.$$

As this is for a and specific $T$, we can union bound over all $2^{2^m}$ possible $T$ to get that:

$$\Pr\left[|\text{Ext}(X) - U_m| > \epsilon\right] \leq 2^{2^m} 2^{-\Omega(2^k 2^d \epsilon^2)},$$

where by setting $m = k + d - 2\log(1/\epsilon) - O(1)$, we get that the failure probability of Ext on $X$ is at most $2^{-\Omega(2^k 2^d \epsilon^2)}$. Notice that in comparison to what we saw in Theorem 1.2, we have an additional dependence on a double exponential in the length of the seed $d$which gives us room to again union bound over all flat sources to get an upper bound on the probability that a random function is an extractor for all flat, and therefore all, $(n, k)$-sources. Taking the union bound over all flat sources, we have that since there are $\binom{2^n}{2^k}$ flat sources, the probability Ext fails on some flat source is upper bounded by:

$$\binom{2^n}{2^k} \cdot 2^{-\Omega(2^k 2^d \epsilon^2)} \leq \left(\frac{2^n e}{2^k}\right)^{2^k} \cdot 2^{-\Omega(2^k 2^d \epsilon^2)},$$

where the latter expression is less than 1 if $d \geq \log(n - k) + 2\log(1/\epsilon) + O(1)$. We leave showing this as an exercise to the reader. □