CS 880: Pseudorandomness and Derandomization 4/1/2013

Lecture 18: Construction of Extractors

Instructors: Holger Dell and Dieter van Melkebeek Scribe: Kevin Kowalski

DRAFT

In the previous lecture, we introduced randomness extractors, which take random bits from a weakly random source and output bits that are close to uniform. We showed that deterministic extractors cannot exist, but that seeded extractors do exist. In this lecture, we present a construction of one such seeded extractor.

1 Preliminaries

We begin by recapping some of the concepts we introduced in the last lecture. Recall that Shannon entropy was deemed inappropriate as a measure for the amount of randomness contained in a weak source, so we instead used the min- entropy H_∞, defined as

$$H_\infty(X) = \log \frac{1}{\max_x \Pr[X = x]}.$$

A random source X is then called a k-source if and only if $H_\infty(X) \geq k$, or equivalently, $\max_x \Pr[X = x] \leq 2^{-k}$.

We also defined seeded extractors $E : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^r$, where by convention $\{0,1\}^n$ comes from a weak random source and $\{0,1\}^d$ comes from a uniform random source. The function E is a (k, ϵ)-extractor if for all k-sources X,

$$d_{\text{stat}}(E(X, U_d), U_r) \leq \epsilon.$$

Last time, we showed that a random function E would be a good extractor, but we haven't yet addressed the question of whether we can both efficiently and deterministically construct an extractor. This is the subject we will tackle in this lecture.

2 Construction of the Extractor

The main theorem of the section is as follows:

Theorem 1. *For all $\alpha > 0$, n, $k \leq n$, and $\epsilon > 0$, there exists an explicit (k, ϵ)-extractor with $r \geq (1 - \alpha)k$ and $d \in O(\log \frac{n}{\epsilon})$ (where d also depends on α).*

This theorem essentially states that we can construct an extractor that extracts almost all the randomness from a k-source with a seed that is only logarithmic in n and $\frac{1}{\epsilon}$. Additionally, though it's not explicitly stated in the theorem, r has to be at most $k + d$ because an extractor cannot output more random bits than it takes in. It is currently an open question whether we can tighten the lower bound on r to $r \geq k + d - O(1)$ while d remains logarithmic in $\frac{n}{\epsilon}$, though we can achieve this bound if we allow d to be polylogarithmic in $\frac{n}{\epsilon}$.

Overview of the construction The shape our final construction will take is shown below:

$$\underbrace{\{0,1\}^n}_{X} \times \underbrace{\{0,1\}^d}_{U_d} \xrightarrow[\text{condenser}]{\delta > 0} \underbrace{\{0,1\}^{(1+\delta)(k+d)}}_{Y} \xrightarrow[\text{mild extractor}]{} \underbrace{\{0,1\}^{(1-\alpha)k}}_{Z}$$

$$H_\infty(X \circ U_d) \geq k + d \qquad H_\infty(Y) \geq k + d \qquad d_{\text{stat}}(Z, U_{(1-\alpha)k}) \leq \epsilon$$

First, we apply a condenser to the original two input strings, which combines them into a slightly longer string with the same lower bound on min-entropy. Then, we can use a simple construction of a known extractor on this string to get a distribution on slightly fewer than k bits that is ϵ-close to uniform. Technically, the extractor will need an additional seed of length logarithmic in $(1+\delta)(k+d)$, but this can be easily collapsed into the original seed. Though this method of construction might seem more convoluted than directly constructing an extractor that achieves the desired properties, no such method is known.

We now formalize what it means to be a condenser or a mild extractor, and construct usable examples of each.

Construction of the mild extractor We will frame our construction in the context of an existence proof.

Lemma 2 (Existence of a Mild Extractor). *For all $\epsilon > 0$, there exists a $\delta > 0$ and efficiently computable $E : \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^r$ such that*

$$|Y| \leq (1+\delta)H_\infty(Y) \implies d_{\text{stat}}(E(Y, U_\ell), U_r) \leq \epsilon.$$

The extractor whose existence is asserted by this lemma is "mild" in that it is only guaranteed to work for sources that already have a high degree of randomness to begin with. Informally, the construction works by interpreting Y as encoding a start vertex and sequence of edge labels in a random walk of length $t = 2^\ell$ on a constant-degree expander. Given a perfectly random seed $i \in \{0,1\}^\ell$ as input, the extractor outputs the i-th vertex of the walk specified by Y. In other words,

$$E(y, s) = \text{the } s\text{-th vertex in the walk encoded by } y.$$

The proof of the lemma is given below.

Proof. Let G be an explicit g-regular expander on $\{0,1\}^r$ with expansion γ, where g and γ are constants. On input $y \in \{0,1\}^n$ and $s \in \{0,1\}^\ell$, let y_0 be the first r bits of the weakly random input, which will specify the starting vertex of our walk, so $n-r$ bits are left to specify the remaining vertices. Before proceeding, we will calculate how many steps we can take in the random walk with the bits available.

Since it takes $\lceil \log_2 d \rceil$ bits to specify each step in the walk, we obtain the inequality

$$\begin{aligned} t &\leq \frac{n-r}{\log_2 d} \\ \iff 2^\ell &\leq \frac{n-r}{\log_2 d} \\ \iff \ell &\leq \log_2(n-r) - \log_2 \log_2 d, \end{aligned}$$

so we can set $\ell = \lfloor \log_2(n-r) - \log_2 \log_2 d \rfloor$. In particular, this means that $\ell \in O(\log n)$.

Now, we need to find a $\delta > 0$ so that our output distribution is ϵ-close to uniform. Let $A \subseteq \{0,1\}^n$ be an event, so it suffices to show that

$$|\Pr[E(Y, U_\ell) \in A] - \Pr[U_r \in A]| \leq \epsilon.$$

By Theorem 3 in Lecture 11 (i.e., the Chernoff bound for expander walks), we have that

$$\Pr_{y \leftarrow U_n}\left[\left|\frac{1}{2^\ell}\sum_{i \in [2^\ell]} X_i - \mu(A)\right| \geq \frac{\epsilon}{2}\right] \leq \exp\left(-b\gamma \cdot \frac{\epsilon^2}{4} \cdot 2^\ell\right),$$

where X_i indicates whether the i-th vertex in the random walk encoded by y is in A, and $b > 0$ is a universal constant.

This allows us to bound $|\Pr[E(Y, U_\ell) \in A] - \Pr[U_r \in A]|$ by splitting the difference in probabilities into two cases: one where the proportion of random walk vertices in A is close to $\mu(A)$, and another where the proportion of such vertices is not close to $\mu(A)$. In particular, let B be the event that $\left|\frac{1}{2^\ell}\sum_{i \in [2^\ell]} X_i - \mu(A)\right| \geq \frac{\epsilon}{2}$, so

$$|\Pr[E(Y, U_\ell) \in A] - \Pr[U_r \in A]| \leq \Pr_{y \leftarrow Y}[\neg B] \cdot \frac{\epsilon}{2} + \Pr_{y \leftarrow Y}[B] \cdot 1.$$

The first probability $\Pr_{y \leftarrow Y}[\neg B]$ is at most 1, and the second probability $\Pr_{y \leftarrow Y}[B]$ is at most $2^{n - H_\infty(Y)} \Pr_{y \leftarrow U_n}[B]$ because $2^n \Pr_{y \leftarrow U_n}[B]$ is the number of $y \in \{0,1\}^n$ for which B holds, and $2^{-H_\infty(Y)}$ is the maximum probability that any particular $y \in \{0,1\}^n$ can be selected from the distribution Y.

Putting these together and setting $H_\infty(Y) = \frac{n}{1+\delta}$, we get that

$$|\Pr[E(Y, U_\ell) \in A] - \Pr[U_r \in A]| \leq \frac{\epsilon}{2} + 2^{\frac{\delta}{1+\delta}n} \cdot \exp\left(-b\gamma \cdot \frac{\epsilon^2}{4} \cdot 2^\ell\right).$$

Since $\ell \in O(\log n)$, we can write the second term in the sum as $2^{\frac{\delta}{1+\delta}n} \cdot 2^{-\beta n}$ for some constant $\beta > 0$, and it becomes immediately apparent that for a sufficiently small choice of δ, this quantity will be less than $\frac{\epsilon}{2}$ and $d_{\text{stat}}(E(Y, U_\ell), U_r) \leq \epsilon$, as desired. □

Note that the last step in this lemma only works if ϵ is a constant. If we require that $\epsilon \leq n^{-\tau}$ for any $\tau > 0$, then the second term in the sum would be $2^{\frac{\delta}{1+\delta}n} \cdot 2^{-\beta n^{1-2\tau}}$ for some constant $\beta > 0$. If $\beta \leq n^{2\tau}$, it wouldn't be possible to choose δ so that this quantity is less than $\frac{1}{2}$.

Construction of the condenser We begin with the formal definition of a condenser.

Definition 1 (Condenser). *A function $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k, ϵ)-condenser if for all k-sources X, there exists a $(k+d)$-source Y such that $d_{\text{stat}}(C(X, U_d), Y) \leq \epsilon$.*

In particular, note that if $m = k + d$, then the condenser is actually an extractor because any $(k+d)$-source on $\{0,1\}^{k+d}$ is uniform. In this way, we can view condensers as "incomplete" extractors that take a weakly random source and "condense" the randomness to a smaller number of bits.

To prove our original theorem, it then suffices to show that these exists a (k, ϵ)-condenser with $m \leq (1+\delta)k + O(d)$, so that the distribution on its output satisfies the source condition on the mild extractor. It is an open question whether we can find a similar condenser with $m \leq k + d + O(1)$ with logarithmic d, though we can do it with polylogarithmic d.

In order to facilitate the construction of our condenser C, we will envision it as a bipartite graph $G_C = (V_1, V_2, E)$, where $V_1 = \{0,1\}^n$, $V_2 = \{0,1\}^m$, and there is an edge between $x \in V_1$ and $y \in V_2$ for every $s \in \{0,1\}^d$ such that $C(x, s) = y$. The graph G_C then has constant degree 2^d. The following lemma clarifies the relationship between C and its corresponding graph G_C, and provides us with a graph-based route to constructing C.

Lemma 3. *The following are equivalent, where $\Gamma(S)$ denotes the set of vertices neighboring S in G_C.*

(a) C is a (k, ϵ)-condenser.

(b) For all $S \subseteq V_1$ such that $|S| = 2^k$, $|\Gamma(S)| \geq A \cdot |S|$ where $A = (1 - \epsilon)2^d$. In other words, G_C is a $(2^k, A)$- vertex expander.

(c) For all $T \subseteq V_2$ such that $|T| < A \cdot 2^k$, $|\{x \in V_1 \mid \Gamma(x) \subseteq T\}| < 2^k$.

In particular, statement (b) can be considered as a weaker analogue of vertex expansion for bipartite graphs, where we only require $|\Gamma(S)|$ to be large for sets S of size exactly 2^k, rather than sets of size at most 2^k.

Proof. **(a) ⇒ (b).** Let $S \subseteq V_1$ such that $|S| = 2^k$, and let X_S be a random variable that is uniformly distributed on S, i.e., X_S is a flat k-source on S. Then, $\Gamma(S) = \text{Support}(C(X_S, U_d))$. Since C is a condenser and X_S is a k-source, its output must be ϵ-close to a $(k + d)$-source Y, which by definition must satisfy

$$|\,\text{Support}(Y)| \geq 2^{H_\infty(Y)} \geq 2^{k+d}.$$

This gives us that

$$|\Gamma(S)| \geq (1 - \epsilon)2^{k+d} = A \cdot |S|,$$

since only a $1 - \epsilon$ fraction of the probability mass of Y can be located in a $(1 - \epsilon)2^{k+d}$-subset of its support. If $|\Gamma(S)|$ was any smaller, the statistical distance between $C(X_S, U_d)$ and Y would be greater than ϵ.

(b) ⇒ (c). Let $T \subseteq V_2$ such that $|T| < A \cdot 2^k$, and let $S = \{x \in v_1 \mid \Gamma(x) \subseteq T\}$. Trivially, $\Gamma(S) \subseteq T$ so $|\Gamma(S)| < A \cdot 2^k$. If $|S| = 2^k$, then by (b) we must have that $|\Gamma(S)| \geq A \cdot 2^k$, contradicting our assumption that $|T| < A \cdot 2^k$. If $|S| > 2^k$, then we obtain the same contradiction by noting that S must contain a subset of size k. Thus, $|S| < 2^k$, as desired.

(c) ⇒ (a). Let X_S be a flat k-source on a subset $S \subseteq V_1$ of size 2^k. From last lecture, it suffices to show that $C(X_S, U_d)$ is ϵ-close to a $(k + d)$-source, since any k-source can be written as a convex combination of flat k-sources. Since $|S| = 2^k$, by (c) we must have that $|\Gamma(S)| \geq A \cdot 2^k = (1 - \epsilon)2^{k+d}$ since S is trivially contained in $\{x \in V_1 \mid \Gamma(x) \subseteq \Gamma(S)\}$.

Now, note that since the degree of any vertex in G_C is exactly 2^d, there are exactly 2^{k+d} edges between S and $\Gamma(S)$, so if we fix an ordering of the vertices in S, at most a ϵ fraction of these edges will lead to a vertex that is in the neighborhood of a previous element of S. Thus, any uniform distribution Y on 2^{k+d} vertices that includes $\Gamma(S)$ will be a $(k + d)$-source that is ϵ-close to $C(X_S, U_d)$, as desired. □

Thus, to construct our condenser, it suffices to construct a bipartite graph that satisfies property (c). While our final goal is to construct $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$, we will first construct $C : \mathbb{F}_q^a \times \mathbb{F}_q \to \mathbb{F}_q^{b+1}$ instead.

Each element of $\mathbb{F}_q^a$ can be viewed as specifying the coefficients of a polynomial $f \in \mathbb{F}_q[y]$ of degree at most $a - 1$. Then, consider C defined as

$$C(f, s) = (s, f_0(s), f_1(s), \ldots, f_{b-1}(s))$$

where $f_i(y) = (f(y))^{h^i} \bmod p(y)$, $h \in \mathbb{N}$, and $p(y)$ is some irreducible polynomial of degree a.

Though we do not have the time to complete the construction in this lecture, we will prove the following claim in the next.

Claim. G_C is a (k, A)-vertex expander where $k = h^b$ and $A = q - ahb$.