In this lecture we will finish constructing extractors for random variables. Recall that we divided this process into two tasks:

1. Condensing our random variable into a random variables whose length is close to its min-entropy. For this task we will an object called condenser which will be discussed today.

2. Extractors for sources with high min-entropy in relation to their length. This was discussed in lecture 7, imagining the input as describing a random walk on an expanding graph, and the random seed as a single vertex in this random walk which will be the output.

**Notation 1.** Let $X_1, X_2$ be two discrete random variables whose support is contained in $S$. We say that $X_1, X_2$ are $\epsilon$ close and denote $X_1 \approx_{\epsilon} X_2$ if the statistical distance between them is at most $\epsilon$, that is

$$\frac{1}{2} \|X_1 - X_2\|_1 \overset{\text{def}}{=} \frac{1}{2} \sum_{i \in S} \left| \Pr[X_1 = i] - \Pr[X_2 = i] \right| \leq \epsilon$$

Let us begin with a formal definition for condenser

**Definition 1.** A function $\text{con}: \{0, 1\}^n \times \{0, 1\}^d \rightarrow \{0, 1\}^m$ is called a $k \rightarrow \epsilon k'$ condenser if for every source $X$ for which $H_\infty(X) \geq k$ it holds that

$$\text{con}(X, U_d) \approx_{\epsilon} X'$$

and

$$H_\infty(X') \geq k'$$

The condenser will be called lossless if $k' = k + d$.

**Explicit Condensers**

**Definition 2.** We say that a random variable $X$ is $k$-flat if $X$’s distribution is uniform over a set of size $2^k$.

**Fact 1.** Every $X$ with $H_\infty(X) = k$ can be written as a convex combination of $k$-flat sources.

That is, there exist $X_1, \ldots, X_t$ $k$-flat and $\alpha_1, \ldots, \alpha_t \in [0, 1]$ such that $\sum_{i=1}^t \alpha_i = 1$ and

$$X = \sum_{i=1}^t \alpha_i X_i$$
**Fact 2.** Any condenser/extractor which works for every \(k\)-flat sources works also for any random variable with \(H_{\infty}(X) = k\).

**Proof.** We show for extractors. The proof is the same for condensers.

Let \(X\) have min-entropy \(k\). From the previous fact we can write \(X\) as a convex combination of \(k\)-flat sources

\[
X = \sum_{i=1}^{t} \alpha_i X_i
\]

It holds (as distributions) that

\[
Ext(X, U_d) = \sum_{i=1}^{t} \alpha_i Ext(X_i, U_d)
\]

Since the extractor works for \(k\)-flat sources, \(Ext(X_i, U_d) \approx_{\epsilon} U_m\).

\[
\frac{1}{2} \|Ext(X, U_d) - U_m\|_1 = \frac{1}{2} \| \sum_{i=1}^{t} \alpha_i Ext(X_i, U_d) - \sum_{i=1}^{t} \alpha_i U_m \|_1 \leq \frac{1}{2} \sum_{i=1}^{t} \alpha_i \|Ext(X_i, U_d) - U_m\|_1 \leq \sum_{i=1}^{t} \alpha_i \epsilon = \epsilon
\]

It remains to construct condensers for \(k\)-flat sources. Given a candidate condenser we define a related graph to it.

**Definition 3.** For \(\text{con}: \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m\) we define the graph \(G_{\text{con}} = (L \cup R, E)\) to be the bipartite graph with sides \(L = \{0,1\}^n\), \(L = \{0,1\}^m\) and the edges are

\[
E = \left\{ (x, \text{con}(x, i)) \mid x \in L, i \in \{0,1\}^d \right\}
\]

We remark that the vertices on the left side all have the same degree \(D = 2^d\).

With this related graph in hand, we show that a function is a condenser if and only if its related graph has good expansion properties. More precisely:

**Theorem 0.1.** A function \(\text{con}: \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m\) is \(k \rightarrow \epsilon\) \(k + d\) condenser if and only if the following property holds for \(G_{\text{con}}:\)

For every \(S \subseteq L\) of size \(|S| = 2^k\), \(\Gamma(S) \geq (1 - \epsilon)D2^k\).

**Proof.** \(\Rightarrow\) Fix \(S \subseteq L\) of size \(2^k\), and think about it as a \(k\)-flat source. Since \(\text{con}\) is a condenser there \(X'\) with \(H_{\infty}(X') \geq k + d\) such that

\[
\text{con}(S, U_d) \approx_{\epsilon} X'
\]

9-2
Since the support of $X'$ has at least $2^{k+d}$ elements, and the probability of each element is at most $2^{-(k+d)}$, the support of $\text{con}(S,U_d)$ must be at least of size $(1-\epsilon)2^{k+d} = (1-\epsilon)D2^k$. We are done since $|\text{Supp}(\text{con}(S,U_d))| = |\Gamma(S)|$.

$\Leftarrow$ Suffices to prove to $k$-flat sources. Let $X$ be some $k$-flat source, and define $S = \text{Supp}(X)$. We know that $|\Gamma(S)| \geq (1-\epsilon)D2^k$. There are exactly $D2^k$ outgoing edges from $S$, meaning there are $D2^k - |\Gamma(S)| \leq \epsilon D2^k$ vertices in $\Gamma(S)$ which have more than a single incoming edge from $S$. Let those vertices be $B$. Take some $B' \subseteq \{0,1\}^m$ disjoint of $\Gamma(S)$ of size $D2^k - |\Gamma(S)|$, and define $X'$ to be uniform over $\Gamma(S) \cup B'$. Then

$$\frac{1}{2}\|\text{con}(X,U_d) - X'\|_1 \leq \epsilon$$

Since those distributions only differ on $B, B'$ and the weight of each one of those sets is at most $\epsilon$.

This theorem means that in order to have a condenser we will need to construct a graph with those expanding properties. We note that unlike other explicit constructions\(^1\), we will need the degree to be logarithmic. We will now proceed to explicitly construct such expanding graph.

**Construction of an Expanding Graph**

Let $n, m, q, h \in \mathbb{N}$, for $q$ some power of a prime number (we will need to work with the field $\mathbb{F}_q$), and let $E(Y)$ be irreducible polynomial of degree $n$ above $\mathbb{F}_q$.

Consider the following bi-partite $G = (\mathbb{F}_q^n \cup \mathbb{F}_q^m, E)$. We associate

$$\tilde{\alpha} = (\alpha_0, ..., \alpha_{n-1}) \mapsto f(Y) = \sum_{i=0}^{n-1} \alpha_i Y^i$$

and define $f_i(Y) = f_i^k \pmod{E(Y)}$ for $i = 0, ..., m-1$.

The edge set of the graph is $E = \{(y, (f_0(y), ..., f_{m-1}(y))) \mid y \in \mathbb{F}_q, f \in \mathbb{F}_q^n\}$.

**Theorem 0.2.** Let $G = (L \cup R, E)$ be the graph defined above. Then for every set $S \subseteq L$ of size at most $K \leq h^m$ it holds that $|\Gamma(S)| \geq |S| (q - nhm)$.

The idea of the proof will be similar to the idea in the proof for list decoding of Reed-Solomon Codes. Assuming it does not hold, We use a non-trivial polynomial of low degree with zeros on the set of neighbours. We will observe a related low degree polynomial and show that every point in $S$ contribute a divisor, giving us a lower bound on the degree of this polynomial and a contradiction.

\(^1\)there are explicit constructions of constant degree expanders based zigzag/replacement products of graphs, see [1]

9-3
Proof. We focus for the moment on $K = h^m$ and assume towards contradiction that $T = \Gamma(S) \subseteq R$ is of size at most $K(q - nmh) - 1$. We will show that there are at most $K - 1$ vertices on $L$ for which all their neighbours are in $T$. We seek for a polynomial $Q(Y, Z_0, ..., Z_{m-1})$ satisfying the following conditions:

1. $\deg_Y Q \leq q - nmh - 1$
2. For every $i$, $\deg_Z_i Q \leq h - 1$
3. $Q|_T \equiv 0$
4. $Q \not\equiv 0$

There dimension of the space of coefficients of polynomials defined by the first two properties is $(q - nmh)h^m$, and each point $a \in T$ defines a homogenous constraint on those coefficients. Since $|T| < (q - nmh)h^m$ there exists a nontrivial polynomial satisfying all of the above conditions. We assume without loss of generality that $E(Y) \nmid Q$, otherwise we just take $Q(Y, Z_0, ..., Z_{m-1})$ (or by a power of $E(Y)$) and maintain all the conditions since $E(Y)$ does not have any zeroes.

Define a two-variable polynomial

$$Q^*(Y, Z) \overset{\text{def}}{=} Q(Y, Z, Z^h, ..., Z^{h^{m-1}}) \pmod{E(Y)}$$

For every polynomial $f(Y)$ it holds that

$$Q^*(Y, f(Y)) \pmod{E(Y)} = Q(Y, f(Y), f(Y)^h, ..., f(Y)^{h^{m-1}}) \pmod{E(Y)} = Q(Y, f_0(Y), f_1(Y), ..., f_{m-1}(Y)) \pmod{E(Y)}$$

Let $f \in S$. Since $Q(y, f_0(y), f_1(y), ..., f_{m-1}(y)) = 0$ for every $y \in \mathbb{F}_q$ and

$$\deg Q(Y, f_0(Y), f_1(Y), ..., f_{m-1}(Y)) \leq (q - nmh - 1) + m(h - 1)(n - 1) < q$$

$Q(Y, f_0(Y), f_1(Y), ..., f_{m-1}(Y))$ is the zero polynomial.

From the above equality we conclude that $Q^*(Y, f(Y)) \pmod{E(Y)}$ is also the zero polynomial. From now on think about $Q^*(Y, Z)$ as a field element of $\mathbb{F}_q[y]/(E(Y))$. Then it holds that $(Z - f(Y))|Q^*(Y, Z)$. This means that

$$\deg Z Q^* \geq |S| = K$$

On the other hand by considering the maximal individual contribution from $Z, Z^h, ..., Z^{h^{m-1}}$ separately

$$\deg Z Q^* \leq (h - 1) + (h - 1)h + ... + (h - 1)h^{m-1} = h^m - 1 < K$$

Contradiction. This means $|\Gamma(S)| \geq K(q - nmh)$.  

9-4
Remark 1. We only proved for $K = h^m$. It is not too difficult to adjust the proof presented above by replacing the first two conditions on $Q$ with the condition that it only has monomials of the form $Y^\alpha Z_0^{\beta_0} \cdots Z_{m-1}^{\beta_{m-1}}$ when $0 \leq \beta_i \leq h - 1$ and 
\[ \sum_{i=0}^{m-1} \beta_i h^i < K \]
\[ \square \]

Corollary 0.1. For every $\epsilon, \alpha > 0$, $n \geq k$ there exists $k \to \epsilon k + d$ lossless condenser with
\[ d = O(\log n + \log \frac{1}{\epsilon}) \]
\[ m = (1 + \alpha)k + O(\log \frac{n}{\epsilon}) \]

Proof. Take $h = \lceil (\frac{2nk}{\epsilon})^\frac{1}{\alpha} \rceil$, $q$ a power of two such that $\frac{1}{2}h^{1+\alpha} < q \leq h^{1+\alpha}$. \[ \square \]

Combinatorial Nullstellensatz

We begin by quoting Hilbert’s Nullstellensatz theorem.

**Theorem 0.3** (Hilbert’s Nullstellensatz thm). Suppose $f, g_1, \ldots, g_m \in \overline{F}(x_1, \ldots, x_m)$ such that if $g_i(x) = 0$ for every $i$, then $f(x) = 0$. Then there exist $r > 0$ and polynomials $h_1, \ldots, h_m$ such that
\[ f^r = \sum_{i=1}^{m} h_i g_i \]

We will prove a combinatorial analogs of this theorem. The case we will be dealing with is that $g_i$’s are of the form $g_i(x) = \prod_{s \in S_i} (x_i - s)$.

**Theorem 0.4** (thm1). Suppose $F$ be a field, $S_1, \ldots, S_n \subseteq F$ non empty sets and $f \in F(x_1, \ldots, x_m)$. Define $g_i(x) = \prod_{s \in S_i} (x_i - s)$. If $f(x) = 0$ for every $x \in S_1 \times S_2 \times \ldots \times S_n$ then there exist $h_1, \ldots, h_n$ such that
1. $\deg(h_i) \leq \deg(f) - \deg(g_i)$ for every $i$.
2. $f = \sum_{i=1}^{m} h_i g_i$

This theorem will easily imply the following theorem.

**Theorem 0.5** (thm2). Suppose $F$ be a field and $f \in F(x_1, \ldots, x_m)$. Suppose $\deg(f) = \sum_{i=1}^{n} t_i$ and the coefficient of $x_1^{t_1} \cdots x_n^{t_n}$ in $f$ is not zero. Then every set $S_1 \times S_2 \times \ldots \times S_n$ such that $|S_i| > t_i$ for every $i$ contains a point $\alpha$ such that $f(\alpha) \neq 0$.

For the proof of 0.4 we will need the following lemma.

9-5
Lemma 0.2. Let $f \in \mathbb{F}(x_1, \ldots, x_n)$ and suppose $\deg_{x_i} f \leq t_i$. If there exist $S_1, \ldots, S_n \subseteq \mathbb{F}$ satisfying $|S_i| \geq t_i$ such that $f|_{S_1 \times \ldots \times S_n} \equiv 0$, then $f \equiv 0$.

We will present two proofs for this lemma, one by induction and the other by interpolation argument.

Proof. For $n=1$: $f$ is univariate polynomial with $|S_1| > \deg(f)$ roots, and hence must be the zero polynomial. Suppose the lemma holds for polynomials of at most $n-1 \geq 1$ variables, prove for $n$. View $f$ as polynomial in $x_n$ with coefficients in $\mathbb{F}(x_1, \ldots, x_{n-1})$

$$f(\alpha, x_n) = \sum_{j=0}^{t_n} f_j(\alpha)x_n^j$$

If for every $j$ it holds that $f_j|_{S_1 \times \ldots \times S_{n-1}} \equiv 0$ then by the induction hypothesis $f_j \equiv 0$, and then $f \equiv 0$. Otherwise there exists $\alpha \in S_1 \times \ldots \times S_{n-1}$ such that $f(\alpha, x_n)$ is a polynomial in $x_n$ with at least one non zero coefficient. It is of degree $t_n < |S_n|$ so there exists $\alpha_n \in S_n$ such that $f(\alpha, \alpha_n) \neq 0$, and contradiction. \hfill \Box

Proof. We present a basis for the space $\{g: S_1 \times \ldots \times S_n \rightarrow \mathbb{F}\}$. For every $\alpha \in S_1 \times \ldots \times S_n$ define

$$f_\alpha(x_1, \ldots, x_n) = \prod_{i=1}^{\beta \in S_1 \setminus \{\alpha\}} (x_i - \beta)$$

Then $f_\alpha(x) = 1$ if $x = \alpha$ and otherwise it is zero. This easily implies that all these functions are linearly independent. From dimension argument we conclude this is a basis for functions $g$ such that $\deg_{x_i} g \leq |S_i| - 1$. This implies that every such $G$ has a unique representation in this basis, in particular $f$. Since $f|_{S_1 \times \ldots \times S_n} \equiv 0$, the coefficients in the representation of $f$ must all be zeroes. \hfill \Box

Proof of theorem 0.4. Define $t_i = |S_i| - 1$, and write $g_i(x_i) = x_i^{t_i+1} - g_i'(x_i)$ when $\deg_{x_i}(g_i') \leq t_i$. We will use the fact that on $S \overset{\text{def}}{=} S_1 \times \ldots \times S_n$ it holds that $x_i^{t_i+1} = g_i'(x_i)$, and replace high degree appearances of $x_i$ in $f$ with lower degree appearances by subtracting a multiplicative of $g_i$. This process will lead us to a polynomial $f'$ which is identical to $f$ on $S$ and $\deg_{x_i}(f') \leq t_i$, and hence the zero polynomial by the lemma. On the other hand $f' = f - \sum h_ig_i$, and the theorem will follow.

Let us describe the process in more details. We go in the order $i = n, \ldots, 1$, and fix $f'$ such that the degree of $x_i$ will be at most $t_i$. For example for $i = n$, we replace each monomial of the form

$$x_1^{\alpha_1} \cdots x_{n-1}^{\alpha_{n-1}} x_n^{t_n+1+\ell}$$

for $\ell \geq 0$ by

$$x_1^{\alpha_1} \cdots x_{n-1}^{\alpha_{n-1}} x_n^{\ell}g_n(x_n)$$

This corresponds to subtracting $x_1^{\alpha_1} \cdots x_{n-1}^{\alpha_{n-1}} g_i(x_i)$ from the current function. We notice that the degree of $x_1^{\alpha_1} \cdots x_{n-1}^{\alpha_{n-1}}$ is at most $t_1 + \ldots + t_{n-1} \leq \deg(f) - \deg(g_i)$, and hence the $h_i$ we will get will satisfy the first property. Remains to argue that this process will
terminate. This is clear since each step the degree of \( x_n \) is lowered by at least one, until it reaches to be at most \( t_n \).

Next we will prove 0.5, which will be useful for us next lecture to prove some results.

**Proof of theorem 0.5.** Let \( S_1, \ldots, S_n \) be as in the theorem, and assume towards contradiction that \( f|_{S_1 \times \cdots \times S_n} \equiv 0 \). By 0.4 there exist \( h_i \) such that

\[
f = \sum_{i=1}^{n} h_i g_i
\]

We know that \( f \) has a monomial of maximal degree \( x_1^{t_1} \cdots x_n^{t_n} \), therefore there is some \( i \) for which \( h_i g_i \) has a non-zero coefficient for this monomial. Since \( \deg(h_i g_i) \leq \deg(f) \), it follows that this monomial is of maximal degree in \( h_i g_i \) as well. But since \( \deg(h_i) \leq \deg(f) - \deg(g_i) \) maximal degree monomials for it must be in the form of \( x_i^{\deg(h_i)} p(x) \) when \( p(x) \) is some maximal monomial of \( h_i(x) \). This is contradiction because the degree of \( x_i \) will be at least \( |S_i| > t_i \).

**References**