In this lecture, we will see more connections between the combinatorial tools defined so far, specifically, we will see how to construction Extractors from error-correcting codes, extractors from Expanders, Samplers from Expanders and a-expanding graphs from spectral Expanders.

Useful results from the last lecture:

Lemma 0.1. Let $D$ be a distribution on $[m]$ with collision probability $\text{CP}(D) \leq \frac{1+4\varepsilon^2}{m}$. Then:

$$\left| D - U_m \right|_{TV} \leq \varepsilon$$

Lemma 0.2 (Expander Mixing Lemma). Let $G$ be an $(N,D,\alpha)$-spectral expander. Then for every $S,T \subseteq V(G)$:

$$\left| E(S,T) - \mu(S)\mu(T) \right| \leq \frac{\alpha}{D} \sqrt{\mu(S)\mu(T)}$$

where $\alpha \in [0,D]$, $\mu(S) = \frac{|S|}{N}$ and $E(S,T) = \{(u,v) \in E(G) \mid u \in S \land v \in T\}$.

1 Extractors

1.1 Extractors from Codes

General Level Idea: The extractor will be sampling indices from the output of a well-separated code.

Given: A code $C : \tilde{n}, n, \left(1 - \frac{1}{q} - \delta\right)\tilde{n}$ on alphabets in $\{0,1\}^q$ with the block length $n$, message length $\tilde{n}$ and the minimum distance $d = \left(1 - \frac{1}{q} - \delta\right)\tilde{n}$.

Construction: Given $C$, construct an extractor $\text{EXT} : \{0,1\}^{n \log(q)} \times \{0,1\}^{\log(\tilde{n})} \rightarrow \{0,1\}^{\log(q)}$ as follows:

$$\forall x \in \mathbb{F}_q^n, y \in [\tilde{n}], \quad \text{EXT}(x,y) = C(x)_{|y}$$

i.e. for input $(x,y)$, encode $x$ using $C$ and keep the $y$-th symbol.

Theorem 1.1. $\text{EXT}$ is a $\left(\log\left(\frac{1}{\delta}\right), \sqrt{\frac{\delta q}{2}}\right)$-strongly seeded extractor.
Proof. Let \( x \sim X \), with min-entropy \( H_\infty(X) \geq \log \left( \frac{1}{\delta} \right) \). Also, let \( y \) be uniformly sampled in \([\tilde{n}]\), i.e. \( y \sim U_{[\tilde{n}]} \). For the sake of notation, let us define \( K = 2^{H_\infty(X)} \), thus,

\[
\Pr[X = x] \leq \frac{1}{K} \leq \delta
\]  

(1)

We will be proving this using the lemma 0.1 by first bounding the collision probability as follows:

\[
\text{CP}(Y, \text{Ext}(X,Y)) = \Pr_{x,x' \sim X, y,y' \sim Y}[(y, \text{Ext}(x,y)) = (y', \text{Ext}(x',y'))]
\]

\[
= \frac{1}{n} \Pr_{x,x' \sim X, y \sim Y}[	ext{Ext}(x,y) = \text{Ext}(x',y)]
\]

\[
\leq \frac{1}{n} \left[ \Pr[x = x'] + \Pr[\text{Ext}(x,y) = \text{Ext}(x',y) | x \neq x'] \Pr[x \neq x'] \right]
\]

(using 1)

\[
\leq \frac{1}{n} \left[ \frac{1}{K} + \Pr[\text{Ext}(x,y) = \text{Ext}(x',y) | x \neq x'] \right]
\]

\[
\leq \frac{1}{n} \left[ \frac{1}{K} + \left( \frac{1}{q} + \delta \right) \right]
\]

\[
= \frac{1}{nq} \left[ 1 + \left( \delta + \frac{1}{K} \right) q \right]
\]

\[
\leq \frac{1 + 2\delta q}{nq} \quad \text{ (using 1)}
\]

where in the first step we conditioned on the event \( y = y' \) and later we used the inequality

\[
\Pr[\text{Ext}(x,y) = \text{Ext}(x',y)|x \neq x'] = \Pr[C(x)|y = C(x')|y | x \neq x'] \leq 1 - \frac{d}{n} = \left( \frac{1}{q} + \delta \right).
\]

Thus, using lemma 0.1, we get:

\[
\left| (\text{Ext}(x,y),y) - (U_q,U_{[\tilde{n}]}) \right|_{TV} \leq \sqrt{\frac{\delta q}{2}}
\]

\[\square\]

1.2 Extractors from Expanders

Given a graph \( G \) which is a \((N,D,\alpha)\)-spectral expander, we would like to construct an extractor \( \text{Ext} : [N] \times [D] \mapsto [M] \). In order to do that, let’s first examine a way of representing Extractors as bipartite graphs. This representation will make the description and analysis of the construction easier.

**Bipartite representation of extractors**

Given an extractor \( \text{Ext} : [N] \times [D] \mapsto [M] \), consider the bipartite graph with vertex set \( V = [N] \cup [M] \). Add edge \((x,z)\) iff there exists a \( y \in [D] \) such that \( \text{Ext}(x,y) = z \). If multiple \( y \in [D] \) have this property then add one edge for each such \( y \). This results in a \( D \)-regular bipartite multigraph. Conversely, given a \( D \)-regular bipartite graph one can recover a function \( \text{Ext} : [N] \times [D] \mapsto [M] \) in the obvious way by labeling the edges incident to every node with numbers from the set \([D]\) in an arbitrary way\(^1\).

\(^1\)Notice that the function \( \text{Ext} \) is not necessarily an extractor for an arbitrary \( D \)-regular bipartite graph.
Construction of Extractors from Expanders  Let $G$ be an $(N,D,\alpha)$-spectral expander. Construct two copies $G^1, G^2$ of $G$ and remove all edges from both copies. If $(i,j) \in E(G)$ then add edge $(i^1, j^2)$ between $i^1 \in V(G^1)$ and $j^2 \in V(G^2)$. Call the resulting bipartite graph $H$ and denote by $\text{EXT}_H$ the extractor function corresponding to $H$ as described in the previous paragraph.

Lemma 1.2. Let $G$ be an $(N,D,\alpha)$-spectral expander and let $\text{EXT}_H : [N] \times [D] \mapsto [N]$ be the function constructed as described above. Then $\text{EXT}_H$ is a $(k,\epsilon)$-extractor for every $k, \epsilon > 0$ such that: $\alpha = D\epsilon \sqrt{\frac{2\epsilon}{N}}$.

Proof. We need to prove that for every source $X$ on $[N]$ with $H_\infty(H) \geq k$ and $Y \sim U[D]$:

$$|\text{EXT}(X,Y) - U[N]|_{TV} \leq \epsilon$$

By definition of the total variation distance, this is equivalent to the following condition holding for every $T \subseteq [N]$:

$$|\Pr[\text{EXT}(X,Y) \in T] - \mu_T| \leq \epsilon$$

(2) where $\mu_T = \frac{|T|}{N}$.

As discussed in previous lectures, we can assume without loss of generality that $X$ is a flat distribution on a set $S \subseteq [N]$ of size $|S| \geq 2^k$ since $H_\infty(X) \geq k$. So, by construction, proving (2) reduces to proving that for every $S \subseteq [N]$ such that $\mu_S = \frac{|S|}{N} \geq \frac{2^k}{N}$, the following holds:

$$\left| \frac{E(S,T)}{|S|D} - \mu_T \right| \leq \epsilon \iff \left| \frac{E(S,T)}{ND} - \mu_S \mu_T \right| \leq \epsilon \mu_S$$

(3)

To prove this, we use the Mixing Lemma (0.2) for $G$ which gives us the following inequality:

$$\left| \frac{E(S,T)}{ND} - \mu_T \mu_S \right| \leq \frac{\alpha}{D} \sqrt{\mu_S \mu_T}$$

Setting $\alpha = D\epsilon \sqrt{\frac{2\epsilon}{N}}$ and noticing that $\mu_S \geq \frac{2^k}{N}$ and $\mu_T \leq 1$ proves (3) and concludes the proof. \qed

1.3 Samplers from Extractors

In this section, we will show how to construct Samplers from Extractors. As a recap,

Definition 1.3 ($(\epsilon, \delta)$-sampler). A function $\text{SAM} : \{0,1\}^n \mapsto [M]^D$ is an $(\epsilon,\delta)$-sampler if for all functions $f : [M] \mapsto [0,1]$,

$$\Pr_{z_1,\ldots,z_D \sim \text{SAM}(U_n)} \left[ \left| \frac{1}{D} \sum_{i=1}^{D} f(z_i) - \mu_f \right| > \epsilon \right] \leq \delta$$

where $\mu_f := \mathbb{E}_{x \sim U[M]}[f(x)]$.

We will now see how to construct a sampler from an extractor $\text{EXT} : [N] \times [D] \mapsto [M]$.

Lemma 1.4. Consider a $(k,\epsilon')$-extractor $\text{EXT}$. Also, define the function $\text{SAM} : [N] \mapsto [M]^D$ as:

$$\text{SAM}(x) = (\text{EXT}(x,1), \text{EXT}(x,2), \ldots, \text{EXT}(x,D))$$

for all $x \in [N]$. Then, $\text{SAM}$ is an $(\epsilon = 2\epsilon', \delta = \frac{K}{N})$-sampler, where $K = 2^k$.

\footnote{Every distribution with $H_\infty(X) \geq k$ is a convex combination of flat sources on sets of size $K = 2^k$.}
Proof. We will prove this by restricting the size of \( x \in [N] \) for which SAMP behaves in an unexpected way. Let us define the set BAD as follows:

\[
\text{BAD} = \left\{ x \in [N] \mid \left| \frac{1}{D} \sum_{i=1}^{D} f(z_i) - \mu_f \right| > \varepsilon \text{ for } (z_1, \ldots, z_d) \leftarrow \text{SAMP}(x) \right\} \tag{4}
\]

First note that:

\[
\Pr_{z_1, \ldots, z_D \leftarrow \text{SAMP}(U_{[N]})} \left[ \left| \frac{1}{D} \sum_{i=1}^{D} f(z_i) - \mu_f \right| > \varepsilon \right] = \Pr_{x \sim U_{\mathbb{N}}} \left[ x \in \text{BAD} \right] \leq \frac{|\text{BAD}|}{N} \tag{5}
\]

We will now complete the proof by upper bounding the size of BAD by \( K \). For assume that \( |\text{BAD}| \geq K \).

Let us define a set \( X \subseteq \text{BAD} \) such that \( |X| = K = 2^k \). Also, define the distribution \( U_X := \text{uniform distribution on the set } X \).

Thus, \( H_{\infty}(U_X) = k \) and correspondingly by the definition of EXT as an extractor, we have:

\[
|\text{EXT}(U_X, U_d) - U_{[M]}|_{TV} \leq \varepsilon' \tag{6}
\]

**Lemma 1.5.** Suppose \( D_1, D_2 \) are distributions on \([M]\) with \( |D_1 - D_2|_{TV} \leq \varepsilon \). Then:

\[
|\mathbb{E}[f(D_1)] - \mathbb{E}[f(D_2)]| \leq 2\varepsilon
\]

**Proof.**

\[
\left| \sum_x f(x)(\Pr[D_1 = x] - \Pr[D_2 = x]) \right| \leq \sum_x f(x)|\Pr[D_1 = x] - \Pr[D_2 = x]| \leq |D_1 - D_2|_1 = 2|D_1 - D_2|_{TV} = 2\varepsilon
\]

Thus (6) is implies:

\[
|\mathbb{E}_{(x,y) \sim (U_X, U_d)}[f(\text{EXT}(x, y))] - \mu_f| \leq 2\varepsilon' = \varepsilon
\]

or,

\[
|\mathbb{E}_{x \in D_X} \left[ \frac{1}{D} \sum_{i=1}^{D} f(\text{SAMP}(x, i)) \right] - \mu_f| < \varepsilon
\]

which clearly contradicts the definition of BAD (see (4)), and thus,

\[
|\text{BAD}| < K \tag{7}
\]

Using 7 with 5, we get that SAMP is an \((\varepsilon, \frac{K}{N})\)-sampler.
Note: We proved the above for $x \sim U_n$ but the proof can go through even when $y$ is an $(n, k')$-source instead by relaxing the guarantee we get. More specifically, in that case we get:

$$\Pr[x \in \text{BAD}] \leq \frac{|\text{BAD}|}{2^{k'}} \leq \frac{2^k}{2^{k'}} = 2^{k-k'}$$

implying that extractors are $(\varepsilon, 2^{k-k'})$-weak samplers.

1.4 $a$-Expanding Graphs

Definition 1.6. Consider an undirected $D$-regular graph $G$ on $N$ vertices. $G$ is said to be $a$-expanding, if

$$\forall S, T \subseteq [N] \text{ with } |S| = |T| \geq a, \ E(S, T) > 0$$

i.e., all vertex subsets $S, T$ of size greater than or equal to $a$ have an edge between them.

A basic question we want to answer is how to construct $a$-expanding graphs, and more specifically, how large does $D$ need to be?

It is not hard to see that every $a$-expanding graph must have $D \geq \frac{N}{a}$, and consequently the probabilistic method suggests that random $\frac{N}{a} \log(N)$ regular graphs are $a$-expanding. In this lecture, we will show how to construct $a$-expanding graphs using spectral expanders, but under the assumption that $D \geq \frac{4N^2}{a^2}$.

Lemma 1.7. A $(N, D, 2\sqrt{D - 1})$ spectral expander is also $a$-expanding for $D \geq \frac{4N^2}{a^2}$.

Proof. $G$ is a $(N, D, \alpha)$-spectral expander, thus, by the Expander Mixing Lemma,

$$\forall S, T \subseteq [N], \ \left| \frac{E(S, T)}{ND} - \mu(S)\mu(T) \right| \leq \frac{\alpha}{D} \sqrt{\mu(S)\mu(T)}$$

implying that $\forall S, T \subseteq [N], \ |S| = |T| = a$,

$$\frac{E(S, T)}{ND} \geq \mu(S)\mu(T) - \frac{\alpha}{D} \sqrt{\mu(S)\mu(T)}$$

$$\implies E(S, T) \geq ND\left(\frac{a^2}{N^2} - \frac{\alpha a}{D N} \right)$$

$$\geq ND\left(\frac{a^2}{N^2} - \frac{2}{D} \sqrt{D - 1} \frac{a}{N} \right)$$

$$\geq ND\left(\frac{a^2}{N^2} - \frac{2}{\sqrt{D}} \frac{a}{N} \right)$$

when $D \geq \frac{4N^2}{a^2}$

$$\geq 0$$

\[\text{An edge } e = (v_i, v_j) \in E(S, T) \text{ if } v_i \in S \text{ and } v_j \in T \text{ and } e \in E\]

\[\text{Existence of such expanders was shown by the Alon-Boppana Lower Bound -} \]

https://lucatrevisan.wordpress.com/2014/09/01/the-alon-boppana-theorem-2/
The above expression thus implies that for $D \geq \frac{4N^2}{a^2}$, a $(N, D, 2\sqrt{D-1})$ spectral expander is $a$-expanding.

In the next lecture, we will see more explicit constructions of $a$-expanding graphs.

*continued in the next lecture...*