CS 6815: Lecture 13

Instructor: Eshan Chattopadhyay Scribe: Makis Arsenis and Ayush Sekhari

October 11, 2018

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 $CP(D) \leq \frac{1+4\varepsilon^2}{m}$. Then:*

$$\left|D - U_{[m]}\right|_{TV} \leq \varepsilon$$

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

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

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

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 $C : \left[\tilde{n},\ n,\ \left(1 - \frac{1}{q} - \delta\right)\tilde{n}\right]$ 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 $EXT : \{0,1\}^{n\log(q)} \times \{0,1\}^{\log(\tilde{n})} \mapsto \{0,1\}^{\log(q)}$ as follows:

$$\forall x \in \mathbb{F}_q^n,\ y \in [\tilde{n}], \qquad EXT(x,y) = C(x)\Big|_y$$

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

Theorem 1.1. *EXT is a $\left(\log(\frac{1}{\delta}), \sqrt{\frac{\delta q}{2}}\right)$-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,

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

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

$$\begin{aligned}
\mathrm{CP}(Y, EXT(X,Y)) &= \Pr_{\substack{x,x' \sim X \\ y,y' \sim Y}}\left[(y, EXT(x,y)) = \left(y', EXT(x',y')\right)\right] \\
&= \frac{1}{\tilde{n}} \Pr_{\substack{x,x' \sim X \\ y \sim Y}}\left[EXT(x,y) = EXT(x',y)\right] \\
&\leq \frac{1}{\tilde{n}}\left[\Pr\left[x = x'\right] + \Pr\left[EXT(x,y) = EXT(x',y) \mid x \neq x'\right]\Pr\left[x \neq x'\right]\right] \\
&\qquad \text{(using 1)} \\
&\leq \frac{1}{\tilde{n}}\left[\frac{1}{K} + \Pr\left[EXT(x,y) = EXT(x',y) \mid x \neq x'\right]\right] \\
&\leq \frac{1}{\tilde{n}}\left[\frac{1}{K} + \left(\frac{1}{q} + \delta\right)\right] \\
&= \frac{1}{\tilde{n}q}\left[1 + \left(\delta + \frac{1}{K}\right)q\right] \\
&\leq \frac{1 + 2\delta q}{\tilde{n}q} \qquad \text{(using 1)}
\end{aligned}$$

where in the first step we conditioned on the event $y = y'$ and later we used the inequality $\Pr[\mathrm{EXT}(x,y) = \mathrm{EXT}(x',y) | x \neq x'] = \Pr[C(x)|_y = C(x')|_y \mid x \neq x'] \leq 1 - \frac{d}{\tilde{n}} = (\frac{1}{q} + \delta)$.
Thus, using lemma 0.1, we get:

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

□

1.2 Extractors from Expanders

Given a graph G which is a (N, D, α)-spectral expander, we would like to construct an extractor $EXT : [N] \times [D] \mapsto [N]$. 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 contruction easier.

Bipartite representation of extractors Given an extractor $\mathrm{EXT} : [N] \times [D] \mapsto [M]$, consider the bipartite graph with vertex set $V = [N] \cup [M]$. Add edge (x, z) iff there exists a $y \in [D]$ such that $\mathrm{EXT}(x, y) = z$. If multiple $y \in [D]$ have this property then add one edge for each such y. This results in a D-regular bipartite multigraph. Conversly, given a D-regular bipartite graph one can recover a function $\mathrm{EXT} : [N] \times [D] \mapsto [M]$ 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 EXT is not necessarily an extractor for an arbitrary D-regular bipartite graph.

Construction of Extractors from Expanders Let G be an (N, D, α)-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 EXT_H the extractor function corresponding to H as described in the previous paragraph.

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

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

$$|\mathrm{EXT}(X, Y) - U_{[N]}|_{\mathrm{TV}} \leq \varepsilon$$

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

$$|\Pr[\mathrm{EXT}(X, Y) \in T] - \mu_T| \leq \varepsilon \tag{2}$$

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 $H_\infty(X) \geq k$.[2]. 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:

$$\left|\frac{E(S, T)}{|S|D} - \mu_T\right| \leq \varepsilon \;\Leftrightarrow\; \left|\frac{E(S, T)}{ND} - \mu_S\mu_T\right| \leq \varepsilon\mu_S \tag{3}$$

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

1.3 Samplers from Extractors

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

Definition 1.3 ((ε, δ)-sampler)**.** *A function $SAMP : \{0, 1\}^n \mapsto [M]^D$ is an (ε, δ)-sampler if for all functions $f : [M] \mapsto [0, 1]$,*

$$Pr_{z_1,\dots z_D \leftarrow SAMP(U_n)}\left[\left|\frac{1}{D}\sum_{i=1}^{D} f(z_i) - \mu_f\right| > \varepsilon\right] \leq \delta$$

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

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

Lemma 1.4. *Consider a (k, ε')-extractor EXT. Also, define the function $SAMP : [N] \mapsto [M]^D$ as:*

$$SAMP(x) = (EXT(x, 1), EXT(x, 2) \dots, EXT(x, D))$$

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

[2]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:

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

First note that:

$$\begin{aligned}
\Pr_{z_1,\ldots z_D \leftarrow \text{SAMP}(U_{[N]})} \left[\left| \frac{1}{D} \sum_{i=1}^{D} f(z_i) - \mu_f \right| > \varepsilon \right] &= \Pr_{\substack{z_1,\ldots z_D \leftarrow \text{SAMP}(x) \\ x \sim U_n}} \left[\left| \frac{1}{D} \sum_{i=1}^{D} f(z_i) - \mu_f \right| > \varepsilon \right] \\
&= \Pr_{x \sim U_n}[x \in \text{BAD}] \\
&\leq \frac{|\text{BAD}|}{N} \qquad (5)
\end{aligned}$$

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 $U_X :=$ uniform distribution on the set X.

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

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

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.

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

□

Thus (6) is implies:

$$\begin{aligned}
&\left| \mathbb{E}_{(x,y) \sim (U_X, U_d)}[f(\text{EXT}(x,y))] - \mu_f \right| \leq 2\varepsilon' = \varepsilon \\
\text{or,} \quad &\left| \mathbb{E}_{x \in D_X} \left[\frac{1}{D} \sum_{i=1}^{D} f(\text{SAMP}(x,i)) \right] - \mu_f \right| < \varepsilon
\end{aligned}$$

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 $\left(\varepsilon, \frac{K}{N}\right)$-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*

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

i.e., all vertex subsets S, T of size greater than or equal to a have an edge between them [3].

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*[4] *G is also a-expanding for $D \geq \frac{4N^2}{a^2}$.*

Proof. G is a (N, D, α)-spectral expander, thus, by the *Expander Mixing Lemma*,

$$\forall S, T \subseteq [N], \quad \left| \frac{E(S,T)}{ND} - \mu(S)\mu(T) \right| \leq \frac{\alpha}{D}\sqrt{\mu(S)\mu(T)} \tag{8}$$

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

$$\begin{aligned}
\frac{E(S,T)}{ND} &\geq \mu(S)\mu(T) - \frac{\alpha}{D}\sqrt{\mu(S)\mu(T)} \\
\implies E(S,T) &\geq ND\left(\frac{a^2}{N^2} - \frac{\alpha}{D}\frac{a}{N}\right) && (\text{as } |S| = |T| = a) \\
&\geq ND\left(\frac{a^2}{N^2} - \frac{2}{D}\sqrt{D-1}\frac{a}{N}\right) \\
&\geq ND\left(\frac{a^2}{N^2} - \frac{2}{\sqrt{D}}\frac{a}{N}\right)
\end{aligned}$$

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

$$\geq 0$$

[3] An edge $e = (v_i, v_j) \in E(S,T)$ if $v_i \in S$ and $v_j \in T$ and $e \in E$

[4] 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.

$\square$

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

continued in the next lecture . . .