Lecture 12: October 8

Lecturer: Eshan Chattopadhyay *Scribe: Benjamin Y Chan*

12.1 Overview

This lecture covers the following material:

- (continued from Lecture 11) Constructing seeded extractors from hash functions
- Constructing seeded extractors from expanders

Finally, recall the notation we used, where a capital $K = 2^k$ for some variable k. So $M = 2^m$, $D = 2^d$, etc.

12.2 Extractors from Universal Hash Families

Recall the following lemma which we proved last class, where if the collision probability is small, we can also bound the statistical distance.

Lemma 12.1 *If D is a distribution on $\{0,1\}_m$, with $\mathrm{cp}(D) \leq \frac{1+\delta}{M}$, then*

$$|D - U_{[M]}| \leq \sqrt{\delta}$$

Now, we try to construct an extractor from a universal hash family.

Lemma 12.2 (Leftover Hash Lemma) *Denote a universal hash family $\mathcal{H} : h : [N] \to [M]$ s.t. $|\mathcal{H}| = N$. For any (n,k)-source X, then*

$$H(X), H \approx_\epsilon U_m, H$$

where $\epsilon = 2^{\frac{m-k}{2}-1}$ and $H \xleftarrow{R} \mathcal{H}$ (and $\approx_\epsilon$ denotes statistical closeness).

Proof: We examine the collision probability

$$\begin{aligned}
\mathrm{cp}(H(X), H) &= \Pr_{\substack{h,h' \sim H \\ x,x' \sim X}}[(h(x), h) = (h'(x'), h')] \\
&= \mathrm{cp}(H) \cdot \Pr[h(x) = h(x')] \\
&\leq \mathrm{cp}(H) \left(\mathrm{cp}(X) + \max_{\substack{x \neq x' \\ x \in supp(X)}} \Pr_{h \sim H}[h(x) = h(x')] \right)
\end{aligned}$$

Observe that:

$$\begin{aligned}\mathrm{cp}(H) &= \frac{1}{N} \text{ (because there are } N \text{ hash functions)}\\ \mathrm{cp}(X) &= \sum_{x\in\{0,1\}^m} \left(\Pr_{x'\sim X}[x'=x]\right)^2\\ &\le \frac{1}{K}\end{aligned}$$

So then, by definition of universal hash function:

$$\begin{aligned}\mathrm{cp}(H(X),H) &\le \frac{1}{N}\left(\frac{1}{K}+\frac{1}{M}\right)\\ &= \frac{1}{MN}\left(1+\frac{M}{K}\right)\end{aligned}$$

Finally, by Lemma 12.1, then:

$$\begin{aligned}|(H(x),H)-(U_m,H)| &\le \sqrt{M/K}\\ &= 2^{\frac{m-k}{2}}\end{aligned}$$

Solving for m:

$$m = k - 2\log\left(\frac{1}{\epsilon}\right)$$

■

Lemma 12.3 *Denote* $\mathrm{Ext}:\{0,1\}^n\times\{0,1\}^d\to\{0,1\}^m$*, where* $\mathrm{Ext}(x,h)=h(x)$*, and* $d=n$*. Then* Ext *is a* (k,ϵ)*-strong seeded extractor for all* k,m *s.t.* $m=k-2\log(\frac{1}{\epsilon})$*.*

This extractor isn't the best we can do. The seed length is too long (since $d=n$). Recall that we showed an extractor with seed length $d=\log(n-k)+2\log(\frac{1}{\epsilon})+O(1)$ exists using the probabilistic method (in the previous lecture), and this universal hash extractor construction does not get very close.

12.3 Extractors from Expanders

Lemma 12.4 *For all* $n,k\in\mathbb{N}$*,* $\exists$ *explicit* (k,ϵ)*-seeded extractor* $\mathrm{Ext}:\{0,1\}^n\times\{0,1\}^d\to\{0,1\}^m$ *bits, where* $m=n$*, with (very large) seed length* $d=O(n-k+\log(\frac{1}{\epsilon}))$*.*

Proof: Let G be a (N,D,γ)-spectral expander. Turn G into a bipartite graph G' (in a natural way) s.t. both halves have N vertices, and edges originally in G now span the two halves. Clearly G' is D-left regular.

From G' we construct an extractor $\mathrm{Ext}(x,i)=i$th neighbor of vertex x in G'. (Intuitively, the input chooses a vertex and the seed chooses a random neighbor.) Then, we want to prove that $\mathrm{Ext}(X,U_d)$ and U_n's statistical distance is small (interpreting T intuitively as any subset of the right vertices):

$$|\mathrm{Ext}(X,U_d)-U_n|\le\epsilon$$

$$\forall T\subseteq[N],\ \left|\Pr[\mathrm{Ext}(X,U_d)\in T]-\frac{|T|}{N}\right|\le\epsilon$$

Then, note that

$$\Pr[\mathrm{Ext}(X, U_d) \in T] = \frac{e(S,T)}{K \cdot D}$$

Plugging it in, and rearranging, where $|S| = K$::

$$\left| \frac{e(S,T)}{K \cdot D} - \frac{|T|}{N} \right| \leq \epsilon$$
$$\left| \frac{e(S,T)}{DN} - \frac{K|T|}{N^2} \right| \leq \frac{\epsilon K}{N}$$

Recall the statement of the expander mixing lemma,

$$\left| \frac{e(S,T)}{N \cdot D} - \frac{|T|}{N} \cdot \frac{|S|}{N} \right| \leq \lambda \frac{\sqrt{|S||T|}}{N}$$

To get the proper ϵ, then we can we solve:

$$\lambda \frac{\sqrt{KN}}{N} \leq \epsilon \frac{K}{N}$$
$$\lambda \leq \epsilon \sqrt{\frac{K}{N}}$$

So we want λ of the form $\epsilon \cdot 2^{\frac{k-n}{2}}$. Recall that $\lambda \geq c \cdot \frac{1}{\sqrt{D}}$, if we start with a good enough spectral expander. Then $D = c \cdot \frac{1}{\lambda_2} = O(\frac{1}{\epsilon^2} 2^{n-k})$ so we have seed length:

$$d = O(n - k + \log(1/\epsilon))$$

■

A brief note on the efficiency of this extractor. We want it to run in $poly(n)$ time. So we need spectral expanders that are strongly explicit, so that given a vertex we can find its neighbors in efficiently.

12.4 One More Construction From Expanders

Lemma 12.5 *For all $\alpha > 0$, there exists $\beta > 0$ s.t. $\forall n, k \in \mathbb{N}$, there exists a (k, ϵ)-seeded extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$, where $k \geq (1-\beta)n$, $\epsilon = O(1)$, $D = 2^d = \alpha n$ (so $d = c\log(n)$ bits), for $m = (1-\alpha)n$.*

Proof: Let G be a 2^c -regular on M nodes, where c is. constant and $\lambda = \lambda(G) < 1$.

Let X an (n, k)-source. Use the first m bits of the source to select a vertex x on an expander. Divide the last $n - m$ bits into multiple sections, each of length c bits, s.t. $n = m + c \cdot D$.

The construction is then to start at vertex x, and then do a D-length random walk (where each step in the walk selected by a group of c bits); each neighbor we traverse during the walk we add to the right set of a bipartite graph G', and then we can use the seed to choose one of these D neighbors and extract one as the output.

Choose any $T \subseteq [M]$. Let $\widetilde{\mu}$ denote the fraction of neighbors of vertex x that are also in T when we use bits drawn from U_n to do the random walk (instead of X). Then, assuming n is large enough,

$$\begin{aligned}\widetilde{\mu} &= \Pr[|\widetilde{\mu} - \mu_T| > \epsilon] \\ &\leq 2 \cdot exp\left(\frac{-(1-\lambda)\epsilon^2 D}{4}\right) \text{ (by Expander Chernoff Bound)} \\ &= 2^{-c'n}\end{aligned}$$

(If n is too small, we can brute force an expander.)

Define $Bad_T \subseteq \{0,1\}^m$, $Bad_T = \left\{x \in \{0,1\}^n \mid \left|\frac{|N(x) \cup T|}{D} - \frac{|T|}{M}\right| > \epsilon\right\}$. There are very few bad Ts:

$$\frac{|Bad_T|}{2^n} < 2^{-c'n}$$
$$\Pr[x \in Bad_T] \leq \frac{|Bad_T|}{2^n} \leq 2^{-cn}/2^{-\beta m}$$

So we choose $\beta < c$.

■