Lecture 9: Expanders Part 2, Extractors

Topics in Complexity Theory and Pseudorandomness (Spring 2013)
Rutgers University
Swastik Kopparty
Scribes: Jason Perry, John Kim

In this lecture, we will discuss further the pseudorandomness properties of expander graphs. Then we will begin on the topic of extractors.

1 More results on expanders

In the previous lecture we saw the definitions of eigenvalue expanders, edge expanders, and vertex expanders for d-regular graphs. To recap, in an eigenvalue expander, all except the first eigenvalue of the graph's adjacency matrix are bounded below d. In an edge expander, every subset of vertices of size below a certain threshold has a large number of edges "going out" of the subset. And in a vertex expander, every such subset of vertices has a large neighborhood.

We will show that λ eigenvalue expanders are edge and vertex expanders. The converse is also true: edge and vertex expanders are eigenvalue expanders. All these objects are strictly weaker than *absolute* eigenvalue expanders. For more theorems on the relations between expander definitions, see the lecture notes from the professor's graph theory course at `http://math.rutgers.edu/~sk1233/courses/graphtheory-F11/`.

1.1 Eigenvalue expanders are edge expanders

Let $e(S,T)$ denote the number of edges between subsets S and T of vertices in a graph, and let $S^C = V \backslash S$.

Theorem 1. *If $G(V,E)$ is a d-regular λ-absolute eigenvalue expander on n vertices, then for all $S \subseteq V$ where $|S| \leq n/2$, $e(S,S^C) \geq \frac{d-\lambda}{2} \cdot |S|$.*

Which is to say that G is a $(\frac{d-\lambda}{2} \cdot |S|, n/2)$ edge expander.

As an aside, we might try to get this theorem from the expander mixing lemma, which bounds the number of edges between *any* two subsets of vertices in G. We can consider $e(S,S)$, the number of edges internal to S, and subtract it from the number of outgoing edges, which has an upper bound of $d|S|$. So, by the expander mixing lemma:

$$\left| e(S,S) - \frac{d|S|^2}{n} \right| \leq \lambda |S|$$

$$e(S,S) \leq \frac{d|S|^2}{n} + \lambda |S|.$$

Then,

$$\begin{aligned} e(S, S^C) &\geq d|S| - \left(\frac{d|S|^2}{n} + \lambda|S|\right) \\ &\geq \left(d - \frac{d|S|}{n} - \lambda\right) \cdot |S| \\ &\geq \left(d\left(1 - \frac{|S|}{n}\right) - \lambda\right) \cdot |S| \\ &\geq \left(\frac{d}{2} - \lambda\right) \cdot |S| \end{aligned}$$

We did not reach our desired lower bound, but it's close if λ is much smaller than d (say $\sqrt{d}$.) We now proceed with the proof proper. We use a similar indicator function approach as in the proof of the expander mixing lemma.

Proof. Let $\mathbf{1}_S, \mathbf{1}_{S^C}$ be the 0/1 indicator vectors for vertex sets S and S^C, respectively. We can write

$$\begin{aligned} e(S, S) &= \mathbf{1}_S^T \mathbf{A} \mathbf{1}_S \\ e(S, S^C) &= \mathbf{1}_S^T \mathbf{A} \mathbf{1}_{S^C} = d|S| - \mathbf{1}_S^T \mathbf{A} \mathbf{1}_S. \end{aligned}$$

Now let's write $\mathbf{1}_S$ in the eigenbasis. Say $\mathbf{v}_i$ are the eigenvectors, and we know the following facts:

$$\begin{aligned} \mathbf{1}_S &= \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \\ \alpha_1 &= \langle \mathbf{1}_S, \mathbf{v}_1 \rangle = \frac{|S|}{n} \\ \sum_{i=1}^{n} \alpha_i^2 &= |S|. \end{aligned}$$

Now expand out the expression for $e(S, S)$:

$$\begin{aligned} \mathbf{1}_S^T \mathbf{A} \mathbf{1}_S &= \left(\sum_{i=1}^{n} \alpha_i \mathbf{v}_i\right) \mathbf{A} \left(\sum_{j=1}^{n} \alpha_j \mathbf{v}_j\right) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \ \mathbf{v}_i^T \mathbf{A} \mathbf{v}_j \\ &= \sum_{i=1}^{n} \alpha_i^2 \lambda_i, \end{aligned}$$

since only the terms where $i = j$ survive. We already know $\sum_i \alpha_i^2$. We isolate the first term,

recalling that $\lambda_1 = d$, and bound the rest:

$$\begin{aligned}\mathbf{1}_S^T \mathbf{A} \mathbf{1}_S &\leq \frac{|S|^2}{n} \cdot d + \lambda \left(\sum_{i=2}^{n} \alpha_i^2 \right) \\ &\leq \frac{|S|^2}{n} \cdot d + \lambda \left(|S| - \frac{|S|^2}{n} \right) \\ &\leq |S| \left(\frac{d|S|}{n} + \lambda \left(1 - \frac{|S|}{n} \right) \right).\end{aligned}$$

Observe that this is a convex combination of d and λ with weight $|S|/n$. Let's relate this back to the expression for outgoing edges:

$$\begin{aligned}e(S, S^C) &\geq \left(d - \left(\frac{d|S|}{n} + \lambda \left(1 - \frac{|S|}{n} \right) \right) \right) \cdot |S| \\ &= \left(1 - \frac{|S|}{n} \right) \cdot (d - \lambda) \cdot |S|.\end{aligned}$$

Since $(1 - |S|/n)$ is at least $1/2$,

$$e(S, S^C) \geq \frac{d - \lambda}{2} \cdot |S|.$$

□

In fact, we could have gotten this result from the expander mixing lemma, by using our precise value $\sum_i \alpha_i^2 = |S| - |S|^2/n$, instead of upper-bounding it by $|S|$.

1.2 Aside: How good can an eigenvalue expander be?

The next result shows that a bounded degree graph cannot be too good an eigenvalue expander. In fact the same result holds for (non-absolute) eigenvalue expansion (but we will not show that here).

Lemma 2. *For all d-regular λ-absolute eigenvalue expanders, $\lambda = \Omega(\sqrt{d})$.*

Proof. To show that the adjacency matrix $\mathbf{A}$ of λ-absolute eigenvalue expander G has a large second eigenvalue, we look at its trace. However, the trace, which is equal to the sum of the eigenvalues, could be negative, so instead we take:

$$\begin{aligned}\mathrm{Tr}(\mathbf{A}^2) &= \sum_{i=1}^{n} \lambda_i^2 \\ &\leq d^2 + \lambda^2 (n-1).\end{aligned}$$

Note that the entries of $\mathbf{A}^2$, $(\mathbf{A}^2)_{ij} = \sum_k \mathbf{A}_{ik}\mathbf{A}_{kj}$, are the number of length-2 paths from i to j in G. The trace of $\mathbf{A}^2$ is the sum over all vertices i of length-2 paths from i to i. The only such paths are those that go out and come back along one edge, so the value of the trace in our d-regular graph is dn. So

$$\begin{aligned}dn &\leq d^2 + \lambda^2 (n-1) \\ \Rightarrow \lambda &\geq \Omega(\sqrt{d}).\end{aligned}$$

□

1.3 Eigenvalue expanders are vertex expanders

Define $\Gamma(S)$ as the neighborhood of S in G. Let us take G to be a d-regular, λ-absolute eigenvalue expander on n vertices, and see what lower bound we get on the size of $\Gamma(S)$.

First, we need to express $\Gamma(S)$ algebraically. Again taking the vector $\mathbf{1}_S$, note that the entries of $(\mathbf{A} \cdot \mathbf{1}_S)$ are the number of neighbors of each vertex in S. Therefore,

$$\text{Support}(\mathbf{A} \cdot \mathbf{1}_S) = \Gamma(S)$$

Our go-to trick for showing that support is large is Cauchy-Schwarz.

$$\text{Support}(\mathbf{A} \cdot \mathbf{1}_S) \geq \frac{\|\mathbf{A} \cdot \mathbf{1}_S\|_1^2}{\|\mathbf{A} \cdot \mathbf{1}_S\|_2^2}.$$

For the L^1 norm, observe that each vertex contributes 1 to each of its d neighbors:

$$\|\mathbf{A} \cdot \mathbf{1}_S\|_1 = d|S|$$

and

$$\begin{aligned}
\|\mathbf{A} \cdot \mathbf{1}_S\|_2^2 &= \left\| \sum_{i=1}^{n} \alpha_i \lambda_i \mathbf{v}_i \right\|_2^2 \\
&= \sum_i \alpha_i^2 \lambda_i^2 \\
&\leq \frac{|S|}{n} \cdot d^2 + \lambda \left(|S| - \frac{|S|^2}{n} \right).
\end{aligned}$$

So,

$$\begin{aligned}
|\Gamma(S)| = |\text{Support}(\mathbf{A} \cdot \mathbf{1}_S)| &\geq \frac{d^2 |S|^2}{\frac{|S|^2}{n} \cdot d + \lambda^2 \left(|S| - \frac{|S|^2}{n} \right)} \\
&= \left(\frac{d^2}{\frac{|S|}{n} \cdot d + \lambda^2 \left(1 - \frac{|S|}{n} \right)} \right) \cdot |S| \\
&= \left(\frac{1}{\frac{|S|}{n} + \frac{\lambda^2}{d^2} \left(1 - \frac{|S|}{n} \right)} \right) \cdot |S|,
\end{aligned}$$

which looks like an expansion factor of about d^2/λ^2. If $|S| \leq \alpha n$, the expansion factor is

$$\frac{d^2}{\alpha d^2 + (1-\alpha)\lambda^2},$$

which gives us $(1+\Omega(1))$-expansion when $\lambda < d$, $\alpha < 1$. If $\alpha = o(1)$ then this gives expansion nearly $\frac{d^2}{\lambda^2}$, which cannot be larger than $\frac{d}{4}$ (this is a sharper version of the bound on absolute eigenvalue expansion that we showed earlier).

Later we will see constructions of expanders such that all small sets have vertex expansion nearly equal to d.

1.4 Expander Tricks

Is there a way to make an eigenvalue an expander an absolute eigenvalue expander? Yes, by using the (somewhat dirty) trick of adding lots of self-loops at every vertex.

Trick 1. If G is a d-regular λ-eigenvalue expander with adjacency matrix $\mathbf{A}$, then the graph given by $(\mathbf{A} + d\mathbf{I})$ is a $2d$-regular, $(d + \lambda)$-eigenvalue expander.

To see this, note that the eigenvalues of $(\mathbf{A} + d\mathbf{I})$ are $\lambda_1 + d, \lambda_2 + d, ..., \lambda_n + d$. Since the smallest eigenvalue of $\mathbf{A}$ is at least $-d$, $(\mathbf{A} + d\mathbf{I})$ has no negative eigenvalues.

What we lose here is the ratio between λ and d. It's more like $d/2$ instead of the $\sqrt{d}$ we had originally. However, there is another trick that amplifies that gap:

Trick 2. The graph G^k, given by adjacency matrix $\mathbf{A}^k$, is a d^k-regular, λ^k-absolute eigenvalue expander.

This is the case since the eigenvalues of $\mathbf{A}^k$ are $\lambda_1^k, \lambda_2^k, ..., \lambda_n^k$.

1.5 Chernoff bounds for expander walks

We already saw in the last class that random walks on expanders give good sampling properties. In fact, expander walk sampling obeys Chernoff-like bounds.

Consider $G(V, E)$, a d-regular λ-absolute eigenvalue expander, and a set $S \subseteq V$. Pick $x_0 \in V$ uniformly, and let $x_0, x_1, ..., x_D$ be a random walk on G. Let X_i be the indicator of whether vertex i is in S. then $\mu = |S|/|V|$.

The result, due to Gillman, is that

$$\Pr_x \left[\left| \frac{(\#\text{ of } i \text{ with } x_i \in S)}{D} - \mu \right| > \epsilon \right] \leq \exp\left(-\epsilon^2 D \left(1 - \frac{\lambda}{\alpha} \right) /4 \right)$$

The proof of this will be developed in the homework.

2 Extractors

Extractors are also pseudorandom objects, and seem to bear some similarities to expanders we have seen, but their properties are different. They address the question, "Where do we obtain uniform random bits in the first place?" Physical phenomena seem to be good sources of randomness, but they do not provide randomness in the form of strings of independent, uniformly distributed bits. So we desire a way to take numbers from an unknown distribution—all that is known is that it has sufficiently high "entropy"—and produce sequences of uniform random bits. This is the job of extractors.

Using extractors, we have the result (due to Zuckerman) that deterministic polynomial time algorithms can simulate randomized algorithms using just a weak source of randomness. In complexity language, $P^{\text{teacup}} = BPP$.

As a toy example of the applications of extractors, consider a procedure for generating a random n-bit prime number. The obvious thing to do is to generate a random n-bit integer, test it for

primality, and repeat until success. By the prime number theorem, we expect to have to do this about n times, thus requiring about n^2 uniform random bits. Is there any way to decrease this? One way is that if, after a failure (finding that x is composite), we could somehow use the randomness "left over" in x. It has the entropy of a sample from the uniform distribution over all n-bit *composite* integers. This is what extractors will allow us to do (with some limitations.)

2.1 Min Entropy

Let X be a random variable over some domain W.

Definition 3. *The* min entropy *of X is defined by:*

$$H_\infty(X) = \min_{w \in W} \log_2 \left(\frac{1}{Pr(X = w)} \right).$$

In particular, note that:

1. $H_\infty(X) \geq k \Leftrightarrow$ All elements in W occur with probability at most 2^{-k}.
2. $H_\infty(U_m) = m$

where U_m is the uniform distribuition on m bits.

2.2 Data Processing Inequalities for Min Entropy

Let Y be a random variable over a domain W'. Let f be any function on W and let g be any function on $W \times W'$. Then:

1. $H_\infty(f(X)) \leq H_\infty(X)$
2. $H_\infty(g(X,Y)) \leq H_\infty(X) + H_\infty(Y)$

2.3 A Wishful Extractor

An ideal extractor would take any random variable with sufficiently large min entropy to a uniform distribution. More precisely,

Definition 4. *A* k-extractor *is a function $f : 0,1^n \to 0,1^m$ such that for every X distributed on $0,1^n$ with $H_\infty(X) \geq k$, $f(X)$ is uniform on $0,1^m$.*

Unfortunately, k-extractors do not exist. This is for trivial reasons: asking for *exact* uniformity turns out to be too much.

For practical purposes, we don't need extractors to map exactly to the uniform distribution. Perhaps there are extractors that take random variables with large min entropy to distributions that are close to uniform.

Definition 5. *A* (k, ϵ)-extractor *is a function* $f : 0,1^n \to 0,1^m$ *such that for every* X *distributed on* $0,1^n$ *with* $H_\infty(X) \geq k$, $f(X)$ *is* ϵ*-close to uniform on* $0,1^m$.

such (k, ϵ)-extractors also do not exist! To see this, take $m = 1$. Then either $f^{-1}(0)$ or $f^{-1}(1)$ has size at least 2^{n-1}. WLOG, suppose it is $f^{-1}(0)$. Let X be the uniform distribuition on $f^{-1}(0)$. Then $H_\infty(X) \geq n - 1$.

However, f(X)=0, so the statistical distance from $f(X)$ to U_1 is 1.

2.4 Extractors and Their Existence

To get a useful extractor that actually exists, we need to add a small number of "seed" random bits as an investment to our extractor.

Definition 6 (the real thing)**.** *A* (k, ϵ)-extractor *is a function* $f : 0,1^n \times 0,1^d \to 0,1^m$ *such that for every* X *distributed on* $0,1^n$ *with* $H_\infty(X) \geq k$, $f(X, Y)$ *is* ϵ*-close in* L_1 *to uniform on* $0,1^m$, *where* $Y = U_d$ *is a random variable independent of* X.

The main theorem is that (k, ϵ)-extractors exist.

Theorem 7. *Random functions are good* (k, ϵ)*-extractors.*

In fact, they are amazing extractors. One can take $d = O(\log n + \log \frac{1}{\epsilon}$ (so there are only about $\log n$ bits of randomness used as seed investment), and, more surpisingly, we can have $m = d + k - 2\log \frac{1}{\epsilon} - O(1)$ (so we get nearly as many output bits as imaginable: all the d bits of seed, and nearly all the k bits of weak randomness in the weak random source!).

Directly applying the probabilistic method the usual way is tricky: the number of potential weak random sources X the extractor should work with is infinite (and so it won't support a union bound across all weak random sources). Thus, before starting with our use of the probabilistic method, we need a few simple observations.

Observation 8. *Every random variable* X *with* $H_\infty(X) \geq k$ *is a convex combination of uniform distributions over sets of size* 2^k *(requires* $2^k \in \mathbb{Z}$.

Observation 9. *If* $f(X_i, Y)$ *is* ϵ*-close to uniform* $\forall i$ *and* X *is a convex combination of the* X_i, *then* $f(X, Y)$ *is* ϵ*-close to uniform.*

Proof. Write $X = \sum_i \alpha_i X_i$ where $\sum_i \alpha_i = 1$. Then

$$f(X, Y) = \sum_i \alpha_i f(X_i, Y).$$

So we have:

$$\begin{aligned}
\|f(X,Y) - U_m\| &= \left\| \sum_i \alpha_i \left(f(X_i, Y) - U_m \right) \right\| \\
&\leq \sum_i \alpha_i \left\| f(X_i, Y) - U_m \right\| \\
&\leq \sum_i \alpha_i \epsilon \\
&= \epsilon.
\end{aligned}$$

□

Proof. of Theorem 7

For simplicity of notation, we write $N = 2^n, K = 2^k, D = 2^d, M = 2^m$. Let X be any uniform distribution on some set of size K. It follows from the two observations that it suffices to prove that for a random function f, $f(X,Y)$ is ϵ-close to U_m in L_1 with high probability.

If $f(X,Y)$ is not ϵ-close to U_m then $\exists S \subseteq \{0,1\}^m$ such that

$$\Pr_{X,Y}[f(X,Y) \in S] - \Pr[U_m \in S] \geq \epsilon.$$

So for a random function f, we define bad events $B_S = \{f : \Pr_{X,Y}[f(X,Y) \in S] - \Pr[U_m \in S] \geq \epsilon\}$.

For fixed $S \subseteq \{0,1\}^m$, the Chernoff Bound gives:

$$\Pr_f B_S \leq e^{-\epsilon^2 KD}.$$

Taking the union bound over all $\binom{N}{K}$ choices of X and 2^M choices of S, we have:

$$\begin{aligned}
& \Pr[f \text{ is not a } (k,\epsilon)\text{-extractor}] \\
\leq\ & \Pr_f[\exists X, S \text{ such that } B_S \text{ occurs}] \\
\leq\ & \binom{N}{K} 2^M e^{-\epsilon^2 KD} \\
\leq\ & \left(\frac{eN}{K} \right)^K 2^M e^{-\epsilon^2 KD} \\
=\ & 2^{c + K \log \frac{N}{K} + M - c\epsilon^2 KD}
\end{aligned}$$

where $c = \log_2 e$.

We need this probability to be less than 1 for a random f to be a (k,ϵ)-extractor. That is, we require:

$$c + K \log \frac{N}{K} + M - c\epsilon^2 KD < 0.$$

This may be written as:

$$D > \frac{1}{c\epsilon^2} \log \frac{N}{K} + \frac{M+c}{c\epsilon^2 K},$$

which is satisfied if each of the two parts of the sum is less than $D/2$. Thus, as long as both

$$\begin{aligned} d &> \log(n-k) + 2\log\frac{1}{\epsilon} + O(1) \\ d &> m - k + 2\log\frac{1}{\epsilon} + O(1) \end{aligned}$$

are satisfied, then f is a good extractor with positive probability. Choosing d and m accordingly completes the proof.

□

Finding explicit extractors matching the above parameters has been a major research direction for the last two decades.

2.5 Extractors from Pairwise Independence

We now give an example of an explicit extractor. This extractor has very poor seed length, but the amount of entropy extracted is as high as possible.

Let H be a pairwise independent family of functions $\{h : \{0,1\}^n \to \{0,1\}^m$, and let $d = \log \|H\|$.

Define $f : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{m+d}$ by:

$$f(x,h) = (h, h(x)).$$

Theorem 10. *f is a $\left(k, \sqrt{\frac{M}{DK}}\right)$-extractor. In particular, we have optimal output length $m = d + k - 2\log\frac{1}{\epsilon}$.*

Proof. We bound the L_2 distance from f to U_{m+d}:

$$
\begin{aligned}
& \sum_{\alpha\in\{0,1\}^{m+d}} \left(\Pr_{x\in X, h\in H}[f(x,h)=\alpha] - \frac{1}{2^{m+d}} \right)^2 \\
= & \sum_{\alpha\in\{0,1\}^{m+d}} \left(\Pr_{x\in X, h\in H}[f(x,h)=\alpha] \right)^2 - \frac{1}{2^{m+d}} \\
= & \Pr_{x_1,h_1,x_2,h_2}[f(x_1,h_1)=f(x_2,h_2)] - \frac{1}{2^{m+d}} \\
= & \Pr_{x_1,h_1,x_2,h_2}[(h_1,h_1(x_1))=(h_2,h_2(x_2))] - \frac{1}{2^{m+d}} \\
= & \Pr_{h_1,h_2}[h_1=h_2] \Pr_{x_1,h_1,x_2,h_2}[h_1(x_1)=h_2(x_2)] - \frac{1}{2^{m+d}} \\
= & \frac{1}{2^d} \Pr_{h,x_1,x_2}[h(x_1)=h(x_2)] - \frac{1}{2^{m+d}} \\
\leq & \frac{1}{2^d}\left(\frac{1}{2^k}+\frac{1}{2^m}\right) - \frac{1}{2^{m+d}} \\
= & \frac{1}{2^{d+k}}.
\end{aligned}
$$

We can now use Cauchy-Schwarz and the L_2 distance to bound the L_1 distance:

$$
\begin{aligned}
& \Rightarrow \quad \|f(X,H)-U_m\|_2 \leq \frac{1}{\sqrt{DK}} \\
& \Rightarrow \quad \|f(X,H)-U_m\|_1 \leq \sqrt{\frac{M}{DK}} \\
& \Rightarrow \quad m = d+k-2\log\frac{1}{\epsilon}.
\end{aligned}
$$

□