CS 6815: Lecture 11

Instructor: Eshan Chattopadhyay Scribes: Jason Gaitonde, Shawn Ong

October 2, 2018

1 Spectral Expansion implies Vertex Expansion

We recall from last class that if $p \in \mathbb{R}^n$ is a distribution over vertices of a graph $G = (V, E)$, where $V = [n]$, then the *support of* p is

$$\sup(p) = \{i \in [n] : p(i) > 0\}. \tag{1}$$

Moreover, we defined the Rényi entropy of a distribution p by

$$H_2(p) = \log\left(\frac{1}{\|p\|_2^2}\right) = \log\left(\frac{1}{\sum_{i=1}^n p(i)^2}\right). \tag{2}$$

From the homework, the quantity $\|p\|_2^2$ is the *collision probability* of p, which is the probability that two random samples from p are equal. From last time, we know that

$$\sup(p) \geq 2^{H_2(p)}. \tag{3}$$

We will use these notions to prove that any spectral expander is also a vertex expander.

Theorem 1. *Let $G = (V, E)$ be a (n, d, α)-graph. For any $\beta > 0$, $B \subseteq V$ with $|B| = \beta n$:*

$$|N(B)| \geq \frac{|B|}{\alpha^2(1-\beta) + \beta}$$

where α is the second-largest eigenvalue of $\hat{A}$.

Proof. Let p_B be the uniform distribution on $B \subseteq V$, so that $p_B(v) = \frac{1}{|B|}$ if $v \in B$ and 0 otherwise. Then note that $\|p_B\|_2^2 = \frac{1}{|B|}$ and therefore,

$$H_2(p_B) = \log(|B|). \tag{4}$$

Now, let $p = \hat{A}p_B$, where $\hat{A}$ is the normalized adjacency matrix of G. The normalized eigenvector of $\hat{A}$ is $\frac{1}{\sqrt{n}}\mathbf{1}$, and we can compute $\langle p_B, \frac{1}{\sqrt{n}}\mathbf{1}\rangle = \frac{1}{\sqrt{n}}\sum_{v\in B}\frac{1}{|B|} = \frac{1}{\sqrt{n}}$. Therefore, we can decompose p_B into orthogonal components

$$p_B = \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\mathbf{1}\right) + p_B^\perp = u + p_B^\perp, \tag{5}$$

where we define $u := \frac{1}{n}\mathbf{1}$ and note $\langle u, p_B^\perp\rangle = 0$. Then

$$\hat{A}p_B = \hat{A}(u + p_B^\perp) = \hat{A}u + \hat{A}p_B^\perp = u + \hat{A}p_B^\perp.$$

Moreover, we note that $\langle \hat{A}p_B^\perp, u\rangle = 0$ as well, because $\hat{A}$ is invariant on the orthogonal complement of u. In inner product notation, this follows because

$$\langle \hat{A}p_B^\perp, u\rangle = (p_B^\perp)^T \hat{A}^T u = (p_B^\perp)^T \hat{A} u = u^T \hat{A}^T (p_B^\perp) = \langle \hat{A}u, p_B^\perp\rangle = \langle u, p_B^\perp\rangle = 0, \tag{6}$$

where we use the fact $\hat{A}$ is symmetric and the orthogonality from before. Therefore, by Pythagoras' Theorem,

$$\|\hat{A}p_B\|_2^2 = \|u\|_2^2 + \|\hat{A}p_B^\perp\|_2^2 \tag{7}$$

We also have

$$\|p_B\|_2^2 = \|u\|_2^2 + \|p_B^\perp\|_2^2 \tag{8}$$

Define $\gamma := \|p_B^\perp\|/\|p_B\|$. Combining equations (7) and (8), we derive

$$\begin{aligned}
\|\hat{A}p_B\|_2^2 &= \|p_B\|_2^2 - \|p_B^\perp\|_2^2 + \|\hat{A}p_B^\perp\|_2^2 && (9)\\
&\le \|p_B\|_2^2 - \|p_B^\perp\|_2^2 + \alpha^2\|p_B^\perp\|_2^2 && (10)\\
&= \|p_B\|_2^2(1 - \gamma^2 + \alpha^2\gamma^2) && (11)\\
&= \|p_B\|_2^2(1 - \gamma^2(1 - \alpha^2)), && (12)
\end{aligned}$$

where we use the fact that α is the second largest eigenvalue in the first inequality and $p_B^\perp$ is a weighted combination of the other eigenvectors. To conclude, we observe that $\|p_B\|_2^2 = \frac{1}{|B|} = \frac{1}{\beta n}$, $\|p_B^\perp\|_2^2 = \frac{\gamma^2}{\beta n}$, $\|u\|_2^2 = \frac{1}{n}$, and therefore from equation (8),

$$\frac{1}{\beta n} = \frac{1}{n} + \frac{\gamma^2}{\beta n}, \tag{13}$$

and therefore, $\gamma^2 = 1 - \beta$. Then we have

$$H_2(\hat{A}p_B) \ge H_2(p_B) - \log(1 - (1-\beta)(1-\alpha^2)). \tag{14}$$

As a result,

$$\begin{aligned}
|\mathrm{sup}(\hat{A}p_B)| &\ge 2^{H_2(\hat{A}p_B)}\\
&\ge 2^{H_2(p_B) - \log(1-(1-\beta)\alpha^2)}\\
&= \frac{|B|}{1-(1-\beta)(1-\alpha^2)}.
\end{aligned}$$

Rearranging $1 - (1-\beta)(1-\alpha^2) = 1 - (1 - \beta - \alpha^2 + \beta\alpha^2) = \alpha^2(1-\beta) + \beta$ yields the theorem. □

2 Random Walks on Expanders

We now show that expanders are highly connected in the sense that the probability that we stay inside a subset while doing a random walk goes down exponentially in the number of steps.

Theorem 2. *Let $G = (V, E)$ be a (n, d, α)-graph. For any $\beta > 0$, $B \subseteq V$ with $|B| = \beta n$:*

$$\Pr[E_{B,t}] \le (\beta + \alpha)^t$$

where $E_{B,t}$ is defined to be the event that a random walk from a vertex randomly selected from V is in B for all t steps of the random walk.

Proof. Observe first that $(\hat{A})^t_{ij}$ gives the probability that a random walk starting at i arrives at j after t steps. Let M_B be the $n \times n$ matrix with:

$$(M_B)_{ij} = \begin{cases} 1 & \text{if } i = j \in B \\ 0 & \text{otherwise} \end{cases}$$

In other words, $M_B v$ is the projection of v onto the elements of B. Observe that then $v' = (M_B \hat{A} M_B)v$, if $\|v\|_1 = 1$, is a vector representing the probabilities of selecting a vertex from B according to the distribution corresponding to v, and taking a random edge from there to another vertex in B. In particular, the value of v'_j corresponds to the probability of such a walk ending at j. It follows that $(M_B \hat{A} M_B)^t v$ is the equivalent probability for a random walk of length t whose vertices are all contained in B. Then $\|(M_B \hat{A} M_B)u\|_1$ gives the sum over elements of B of the probabilities that a random walk starting from a vertex chosen uniformly at random never leaves B. More formally,

$$\Pr[E_{B,t}] = \|(M_B \hat{A} M_B)^t u\|_1 \tag{15}$$

We next demonstrate that, for any $v \in \mathbb{R}^n$ with nonnegative coordinates,

$$\|(M_B \hat{A} M_B)u\|_2 \leq (\alpha + \beta)\|v\|_2 \tag{16}$$

To this end, we can write decompose v into orthogonal components $v = \langle v, \hat{u}\rangle \hat{u} + v^\perp$, where $\hat{u} = \frac{1}{\sqrt{n}}\mathbf{1}$ is the ℓ_2-normalized eigenvector (for now, we write $u' = \langle v, \hat{u}\rangle \hat{u}$). Without loss of generality, we may assume that v is supported on B, that is, $M_B v = v$. If this is not the case, then $\|M_B v\|_i \leq \|v\|_i$ for $i = 1, 2$, and the bound we show will still be valid. If $M_B v = v$, we have:

$$M_b \hat{A} v = M_B \hat{A} u' + M_B \hat{A} v^\perp$$

Since u' is a scalar multiple of $\mathbf{1}$ and $\hat{A}\mathbf{1} = \mathbf{1}$:

$$\begin{aligned} M_b \hat{A} v &= M_B u' + M_B \hat{A} v^\perp \\ \|M_b \hat{A} v\|_2 &\leq \|M_B u'\|_2 + \|M_B \hat{A} v^\perp\|_2 \end{aligned} \tag{17}$$

Now consider each component separately:

$$\begin{aligned} \|M_B u'\|_2 &= \sqrt{\sum_{i \in B} (u'_i)^2} \\ &= \sqrt{\sum_{i \in B} \left(\frac{\langle v, \hat{u}\rangle}{\sqrt{n}}\right)^2} \end{aligned}$$

Note that $\langle v, \hat{u}\rangle = \sum_{i=1}^{n} \frac{v_i}{\sqrt{n}} = \frac{\|v\|_1}{\sqrt{n}}$. Additionally, $|B| = n\beta$:

$$\begin{aligned} &= \sqrt{n\beta \frac{\|v\|_1^2}{n^2}} \\ &= \sqrt{\frac{\beta}{n}}\|v\|_1 \end{aligned} \tag{18}$$

We can use Cauchy-Schwarz and the fact that v is supported on B to obtain:

$$\begin{aligned}
\|v\|_1 &= \sum_{i=1}^{n} |v_i| \\
&= \sum_{i\in B} |v_i| \cdot 1 + \sum_{i\notin B} 0 && \text{(since } v_i = 0 \text{ for } i \notin B) \\
&= \langle v, M_B\mathbf{1}\rangle \\
&\leq \|v\|_2 \cdot \|M_B\mathbf{1}\|_2 && \text{(Cauchy-Schwarz)} \\
&= \|v\|_2 \cdot \left(\sum_{i\in B} 1^2\right)^{1/2} \\
\|v\|_1 &\leq \|v\|_2\sqrt{\beta n} && (19)
\end{aligned}$$

From (18) and (19), we have:

$$\|M_B u'\|_2 = \sqrt{\frac{\beta}{n}}\|v\|_1 \leq \sqrt{\frac{\beta}{n}}\sqrt{\beta n}\|v\|_2 = \beta\|v\|_2 \tag{20}$$

The $\|M_B\hat{A}v^{\perp}\|_2$ term is simpler. Since $M_B v$ is a projection of v onto a subset of its components, $\|M_B v\|_2 \leq \|v\|$ for any v. In particular:

$$\|M_B\hat{A}v^{\perp}\|_2 \leq \|\hat{A}v^{\perp}\|_2 \leq \alpha\|v^{\perp}\|_2 \leq \alpha\|v\|_2 \tag{21}$$

The second inequality is justified in the same way as (10), as α is the second-largest eigenvalue of $\hat{A}$; the last inequality follows from the fact that $v^{\perp}$ is a projection of v. Now combining (17), (20), and (21), we obtain:

$$\|M_b\hat{A}v\|_2 \leq \|M_B u'\|_2 + \|M_B\hat{A}v^{\perp}\|_2 \leq (\alpha+\beta)\|v\|_2 \tag{22}$$

Note that we may inductively apply (22) and the fact that $\|M_B v\|_2 \leq \|v\|_2$, as M_B is a projection, to obtain the general result:

$$\|(M_b\hat{A}M_b)^t v\|_2 \leq (\alpha+\beta)^t\|v\|_2 \tag{23}$$

In particular, we have, using (15):

$$\begin{aligned}
\Pr[E_{B,t}] &= \|(M_B\hat{A}M_B)^t u\|_1 \\
&\leq \sqrt{n}\|(M_B\hat{A}M_B)^t u\|_2 \\
&= \sqrt{n}(\alpha+\beta)^t\|u\|_2 && \text{(From (23))} \\
&= (\alpha+\beta)^t
\end{aligned}$$

□

Observation 2.1. *This theorem has applications for one-sided error reduction. For instance, suppose $L \in$ **RP** via a randomized algorithm $A(x,r)$, where $x \in \{0,1\}^n$ is the input and $r \in \{0,1\}^k$ is the random bit string. If $x \in L$, then $A(x,r) = 1$ for all r, and if $x \notin L$, then the probability that $A(x,r) = 1$ is at most $\beta < 1$; that is, at most β fraction of random strings will give the wrong answer. To drive down error, take a spectral expander $G = (\{0,1\}^k, E)$ be a $(2^k, d, \alpha)$-spectral expander. By letting B be the set of "bad" witnesses of an $x \notin L$, take a random walk of length t as specified in the theorem and get strings $r_1, \ldots, r_t$. We then compute $A(x, r_i)$ for each i and output 1 if each computation was 1. By the previous theorem, the probability this procedure mislabels $x \notin L$ is less than $(\beta+\alpha)^t$, which goes down exponentially fast for small enough β and α.*

Moreover, this construction requires $k + t\log d$ random bits to sample the first vertex uniformly and $\log d$ to take a step in the walk.

3 Randomness Extractors

In general, natural sources of randomness are often defective. When we use randomness for various applications in randomized algorithms, cryptography, distributed computing, or other settings, we may require purely random bits. Our next problem will be to determine how to model a defective source and a way to extract pure randomness from it.

One early such model, due to von Neumann, is as follows:

1. Model the defective source as a stream of independent p-biased bits $x_1, \dots, x_i, \dots$ such that the x_i's are i.i.d. and $\Pr[x_i = 1] = p, \Pr[x_i = 0] = 1 - p$.

2. Extract randomness by maintaining a bit which stores the value of $\bigoplus_{k=1}^{i} x_k$.

To see how this works, we note that we can model the parity after i bits using some ideas from Markov chains. Define

$$A = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}. \tag{24}$$

The states are even and odd, respectively, and the transitions are just that the parity remains the same with probability $1 - p$ and changes with probability p. It is easy enough to show that the probabilities that $\bigoplus_{k=1}^{i} x_k = 0$ and 1 respectively are given by the vector $A^i e_1$, where e_1 is the vector $(1, 0)^T$.

By the Perron-Frobenius Theorem, any ergodic Markov chain has a unique stationary distribution π with eigenvalue 1, and moreover, for any distribution q on states, $\lim_{i \to \infty} A^i q = \pi$. Here, ergodicity just corresponds to $p \neq 0$ or 1. It is easy to compute that the eigenvalues are 1 and $1 - 2p$, and the largest eigenvalue corresponds to the uniform distribution u and the latter is $v = (1/2, -1/2)$ using orthogonality by the Spectral Theorem, as A is conveniently also symmetric. Note that we can write $e_1 = u + v$ and then $A^i e_1 = u + (1 - 2p)^i v$. As a result, we see that the coefficient of the second term can be made less than ϵ in absolute value in $\frac{\log(1/\epsilon)}{\log(1/|1-2p|)}$ steps. For fixed p, this is just $O(\log(1/\epsilon))$ steps, and the constant gets better as $p \to 1/2$.

Another way to extract randomness is to take two bits at a time and output 0 if we see the pair 01 first and 1 if we see 10 first (in the case of 11 or 00, we just try again). These events are equally probable with probability $p(1 - p)$ each, and it can be shown that the expected number of pairs needed before seeing one or the other is $\frac{1}{2p(1-p)}$. It's not too hard to get high probability statements using Chernoff bounds.

Definition 3.1. *Let D be a distribution on $\{0, 1\}^n$. Define the **min-entropy** of D to be:*

$$H_\infty(D) = \min_{x \in sup(D)} (-\log(\Pr[D = x])) = \min_{x \in sup(D)} \left(\log \frac{1}{\Pr[D = x]} \right)$$

Intuitively, the larger $H_\infty(D)$ is, the less probably any particular outcome can be. We can formalize this as follows:

Example 3.2. *For any distribution D, $H_\infty(D) \geq k \implies \forall x, \Pr[D = x] \leq 2^{-k}$.*

Proof. Let $x \in \{0, 1\}^n$. By definition, $-\log(Pr[D = x]) \geq k$. This gives:

$$\begin{aligned} \log(Pr[D = x]) &\leq -k \\ 2^{\log(Pr[D=x])} &\leq 2^{-k} \\ Pr[D = x] &\leq 2^{-k} \end{aligned}$$

□

Definition 3.3. *An* (n, k)*-**source** is a distribution on* n *bits with min-entropy at least* k*.*

Our goal will be to create extractors which produce random bits using (n, k)-sources. We may attempt to define such an extractor as follows:

(Tentative) Definition 3.4. *An **extractor** is a function* $Ext : \{0, 1\}^n \to \{0, 1\}$ *such that, for any* (n, k)*-source* X*,* $0.49 \leq \Pr[Ext(X) = 1] \leq 0.51$*.*

However, it turns out that this definition will not be sufficient for our purposes. In fact, we will show that no such extractor can exist for $k = n - 1$; note that if $k = n$, then our source is already random (by Example 3.2, the probability of any outcome will be exactly 2^{-n}).

Claim 3.5. *There is no extractor satisfying (Tentative) Definition 3.4 with* $k = n - 1$*.*

Proof. By contradiction. Suppose that such an extractor $Ext : \{0, 1\}^n \to \{0, 1\}$ existed. But $|Ext^{-1}(0)| + |Ext^{-1}(1)| = 2^n$, so one of these two sets must contain at least half of the elements. Assume without loss of generality that it is $Ext^{-1}(0)$; we then have:

$$|Ext^{-1}(0)| \geq 2^{n-1}$$

Consider the distribution X which is uniform on $Ext^{-1}(0)$ and 0 elsewhere. Note that since the maximum probability of any outcome is $2^{-|Ext^{-1}(0)|} \leq 2^{-(n-1)}$, we have $H_\infty(X) \geq n - 1$. But by definition, we have $Ext(X) = 0$, so $\Pr[Ext(X) = 1] < 0.49$, contradicting the assumption that Ext was an (n, k)-extractor.

□