PCPs and Inapproximability: Expander Graphs and their Applications

My T. Thai

1 Introduction

Since the introduction of Expander Graphs during the 1970's, they turn to be a significant tool both in theory and practice. Last time, we saw how to use expander graphs to show the Hardness Approximation of MAX-3SAT(k). Indeed, they have been used in solving many problems in communication including topology design, error correcting, cryptographic hash function, worm propagation schemes. And of course, expander graphs are also used for hardness of approximation and gap amplification.

To begin, let us consider the following two problems and see how they connect to the expander graphs.

1.1 Some Problems

Error Correcting Codes.

Assume that Alice has a message of k bits which she would like to deliver to Bob over some communication channels. However, a proportion p of the bits may be changed due to the noise, thus the message that Bob receives might be different than the one that Alice sent. How can Alice send Bob a message of k bits so that Bob can correctly receive it? That is, what is the smallest number of bits that Alice can send so that Bob should be able to unambiguously recover the original k bits message.

During the 1940's, Clude Elwood Shannon has developed the theory of communication (which is called Information Theory). Shannon has presented an answer to this problem. He suggested building a dictionary (or code) $C \subseteq \{0,1\}^n$ of size $|C| = 2^k$ and using a bijective mapping ("an encoding") $\varphi : \{0,1\}^k \to C$. To send a message $x \in \{0,1\}^k$, Alice transmits the n-bit encoded message $\varphi(x) \in C$. Assume that Bob receives a string $y \in \{0,1\}^n$ that is a corrupted version of $\varphi(x)$. Bob will find the codeword $z \in C$ that is "closest" to y in terms of hamming distance and determines the k-bits associated with it. If the minimal distance between two words in C is greater then $2pn$, it is guaranteed that the k-bits that Bob will find is exactly the bits Alice encoded.

Therefore the problem of communicating over noisy channel is reduced to the problem of finding a good dictionary (code).

Problem 1. (Communication Problem.) Is it possible to design a series of dictionaries $\{C_k\}$ such that $|C_k| = 2^k$, the distance of each dictionary is greater than $\delta_0 > 0$ and the rate of each code is greater than $R_0 > 0$ where the rate of the dictionary is defined as $R = \frac{\log |C|}{n}$ and the distance of the code is defined as

$$\delta = \frac{\min_{c_1 \neq c_2 \in C} d_H(c_1, c_2)}{n}$$

where d_H is the hamming distance.

Now, let's consider another problem which seems to be unrelated:

De-randomizing Algorithms

Checking the primality (by Rabin in 1980): Given an integer x of k bits and a set of k random bits r, the algorithm computes a function $f(x, r)$ such that if x is primal, then $f(x, r) = 1$, else, $f(x, r) = 1$ with probability smaller than $1/4$. Applying this algorithm over and over again can reduce the error to arbitrary small. Clearly, this process involves the use of more and more random bits.

This primality test is a special case of Random Polynomial algorithm (RP). Let $L \subseteq \{0, 1\}^k$ be a language and consider an algorithm (called RP algorithm) which decides on $x \in \{0, 1\}^k$ whether it is in L or not as follows: RP runs in polynomial time and uses poly(k) random bits and gives an answer which is always 1 if $x \in L$ and if $x \notin L$, gives an answer 1 with probability smaller than $1/4$.

Is there anyway to save the random bits used?

Problem 2. (Saving Random Bits) Assume that $L \in \{0, 1\}^k$ has a RP algorithm. How many random bits are needed in order to give an answer with probability of mistake smaller than $1/d$ for some $d > 1$.

1.2 Solutions via a Special Graph

Let us consider the following special graph.

Definition 1 *(Special Graph). Let $G = (L, R; E)$ be bipartite graph where $|L| = n$, $|R| = m$, and each vertex in L has d neighbors in R. We say G is $(n.m; d)$-special graph if it has the following two properties:*

1. *For any $S \subseteq L$ such that $|S| \leq \frac{n}{10d}$, then $|N(S)| \geq |S|5d/8$*
2. *For any $S \subseteq L$ such that $n/(10d) < |S| \leq n/2$, $|N(S)| \geq |S|$ where $N(S)$ is the set of neighbors of S in G.*

Lemma 1 *There exists a constant n_0 such that for every $d \geq 32$ and $n \geq n_0$, $m \geq 3n/4$, a $(n, m; d)$-special graph exists.*

Proof. First, construct a random graph as follows. Let G be a random bipartite graph with n vertices on the left and m vertices on the right. For each $x \in L$, randomly choose d vertices in R and connect them with v. We claim that the generated graph is a $(n, m; d)$-special graph with high probability.

Let $S \subseteq L$ be such that $|S| = s \leq n/(10d)$. Let $T \subseteq R$ be such that $|T| = t < 5ds/8$. Let $X_{S,T}$ be an indicator random variable for the event that all the edges from S go to T. It is easy to see that if $\sum_{S,T} X_{S,T} = 0$ then property (1) holds. Else, using a union bound and the inequality $\binom{n}{k} \leq (ne/k)^k$, we have:

$$Pr\left[\sum_{S,T} X_{S,T} > 0\right] \leq \sum_{S,T} Pr[X_{S,T} = 1] = \sum_{S,T} (t/m)^{sd}$$

$$\leq \sum_{s=1}^{n/10d} \binom{n}{s}\binom{m}{5ds/8}\left(\frac{5ds}{8m}\right)^{sd}$$

$$\leq \sum_{s=1}^{n/10d} \left(\frac{ne}{s}\right)^{s}\left(\frac{8me}{5ds}\right)^{5ds/8}\left(\frac{5ds}{8m}\right)^{sd} \leq 1/10$$

The last inequality follows since the s^{th} term is bounded by 20^{-s}.

To show that property (2) holds with high probability, we use the same arguments as above.

For every $S \subset L$ such that $n/(3d) < |S| \leq n/2$ and $T \subset R$ such that $|T| < |S|$, let $Y_{S,T}$ be an indicator random variable for the event that all the edges from S go to T. Then if $\sum Y_{S,T} = 0$ then the second property in property (2) holds. Else, we have:

$$Pr\left[\sum_{S,T} Y_{S,T} > 0\right] \leq \sum_{S,T} Pr[Y_{S,T} = 1] = \sum_{S,T} (t/m)^{sd} \leq ... \leq 1/10$$

□

Question 1: Using the same construction and similar argument, can we find a special graph $G = (L, R; E)$ such that $|L| = |R| = n$, each node in L has a degree d and satisfying the following two properties:

1. For any $S \subseteq L$ such that $|S| \leq \frac{n}{3d}$, then $|N(S)| \geq |S|d/4$
2. For any $S \subseteq L$ such that $n/(3d) < |S| \leq n/2$, $|N(S)| \geq |S| + n/(3d)$ where $N(S)$ is the set of neighbors of S in G.

Question 2. Consider the graph G in question 1. Does G has the following property:

Lemma 2 *Let G be a special graph then there exists $B \subset R$ such that $|B| \geq \frac{n}{4d^2}$ and for each vertex $v \in L$, there is at most one neighbor in B.*

Proof. Consider the following algorithm $\mathcal{A}$: Initialize: $L_0 = L, R_0 = R, B = \emptyset$. While there exists $v \in R_i$ with degree at most $2d$, let $B = B \cup \{v\}$, $L_{i+1} = L_i \setminus N(v)$ and $R_{i+1} = R_i \setminus N(N(v))$. Clearly, when $\mathcal{A}$ terminates, we have a set B such that any $u \in L$ has at most one neighbor in B.

Now, we show that $|B| \geq \frac{n}{4d^2}$, that is, we need to show that $\mathcal{A}$ runs $\frac{n}{4d^2}$ iterations. At each iteration i, we have $|L_i| \geq n - 2di$ since v has a degree at most $2d$. We also have $|R_i| \geq n - d(n - |L_i|)$ since each vertex in L has degree d. Note that the number of edges in $G[L_i \cup R_i]$ is at most $d|L_i|$, the average degree of vertices in R_i is upper bounded (at most):

$$\frac{d|L_i|}{|R_i|} \leq \frac{d|L_i|}{n - d(n - |L_i|)}$$

In order to have the average vertex degree in R_i is at most $2d$ (then there exists a vertex with a degree at most $2d$), $|L_i| \geq (1 - 1/(2d-1))n$. Note that $|L_i| \geq n - 2di$, then we have $i \leq \frac{n}{4d^2}$.

Thus we can build such a set B. □

Now consider the special graph G. We will show how to use it to solve the two mentioned problems.

Solution to Problem 1.

Let $G = (L, R; E)$ be such a graph with n left vertices and $3n/4$ right vertices. We first show one useful property of G, that is, for every non-empty $S \subset L$ with $s = |S| \leq n/(10d)$, there exists a vertex $u \in R$ with exactly one neighbor in S, that is, $|N(S) \cap S| = 1$. Note that the number of edges between S and $N(S)$ is equal to ds. Since $N(S) \geq 5ds/8$, the average number of neighbors in S that a vertex in $N(S)$ has is at most $8/5 < 2$. But every vertex in $N(S)$ has at least one neighbor in S, so there must be some vertices in $N(S)$ with exactly one neighbor in S.

Now, use G to construct a code $C \subset \{0,1\}^n$ with rate at least $1/4$ and distance at least $1/(10d)$ as follows. Represent G by a matrix A with R rows and L columns. Let $a_{ij} = 1$ if the i^{th} vertex in R adjacent to the j^{th} vertex in L, else, $a_{ij} = 0$. The code is defined as the nullspace (right kernel) of A: $C = \{x \in \{0,1\}^n \mid Ax = 0\}$.

(1) **Show rate** $\geq 1/4$: Note that C is a linear sub space of $\{0,1\}^n$ defined by $3n/4$ linear equations (linear sub space of dimensions $\geq n/4$) and hence $|C| \geq 2^{n/4}$, thus the rate $\geq 1/4$.

(2) **Show distance** $\geq 1/10d$: Since C is a linear code, the smallest distance between two of its codewords equals to the smallest weight of a non-zero codeword. Let $x \neq 0$ be an n-bit vector with support $S = \{j \in L \ : \ x_j = 1\}$. If $|S| < n/10d$, there is some $i \in R$ with $|N(i) \cap S| = 1$. It means that the i^{th} coordinate in Ax is 1, and so x cannot be a codeword. Therefore, the smallest weight of any codeword in C must be $\geq n/10d$. Thus the distance $\geq 1/10d$.

□

Solution to Problem 2. This will be a homework problem for the assignment 2.

2 Graph Expansion and a Combinatorial Definition of Expanders

Given a graph $G = (V, E)$ which may have multiple edges. For any $S \subset V$, let δS be the set of edges emanating from S to its complement and $\Gamma(S)$ be

the set of adjacent vertices of S. Note that $\Gamma(S) \cap S$ may not be empty.

Definition 2 *(Edge Expansion)* $h(G) = \min_{|S| \leq n/2} \frac{|\delta S|}{|S|}$.

Definition 3 *(Vertex Expansion)* $g(G) = \min_{|S| \leq n/2} \frac{|\Gamma(S) \backslash S|}{|S|}$.

An (n, d, c)-edge expander is a d-regular graph G with n vertices and $h(G) \geq c$. Likewise, an (n, d, c)-vertex expander is a d-regular graph G with n vertices and $g(G) \geq c$.

c is also called an expansion rate of G. Note that usually n is a sufficient large number where c and d are some constants.

Question: Is there any relationship between edge expansion and vertex expansion. Indeed, they are equivalent. Prove it by showing the following:

1. Show that if G is a (n, d, c)-edge expander, then G is also an (n, d, c')-vertex expander for some constant c'

2. Show that if G is an (n, d, c)-vertex expander, then G is also an (n, d, c')-edge expander.

Due to this equivalent, for the rest of this note, we only consider edge expander graphs. In the previous section, we see one example of constructing an expander graph. Now, let us consider another construction, used to prove the following theorem:

Theorem 1 *For all $c > 0$, there exists a constant $d_0 > 0$ and $n_0 > 0$ such that an (n, d, c)-edge expander graph exits for all $d \geq d_0$ and $n \geq n_0$.*

Proof. Similar to the one you have seen before, we will somehow construct a graph randomly and prove that this graph is an expander with high probability when n and d are big enough.

Let us consider this construction. WOLG, assume that nd is even. Consider V with n vertices. For each $v \in V$, replace v by d dummy vertices $v_1, ..., v_d$. Let $U = \{v_1, ..., v_d \mid \forall v \in V\}$. Clearly, $|U| = dn$. Now, find a perfect matching on U. Then compress all these dummy vertices back to each v to obtain the graph G. (That is, for each pair $u, v \in V$, if there is a perfect matching between U_u to U_v, then create an edge between u and v.) Note that G can have loop and multi-edges.

Now, we will show that the probability for G not being an (n, d, c)-expander is very small! Can we use a similar method of union bound and the inequality $\binom{n}{k} \leq (ne/k)^k$ for all $n \geq k$ as before? **Homework Problem!**

□

3 Graph Spectrum and an Algebraic Definition of Expansion

Let A be an adjacency matrix for G where a_{ij} = number of edges between i and j. We will sometimes denote $A(G)$ as the adjacency matrix for G. Note that A is symmetric and has n real eigenvalues, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Note that we can also associate with it an orthonormal system of eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$. Often, the eigenvalues of $A(G)$ are referred as the **Spectrum** of G

Example. Let A be the adjacency matrix of a d-regular graph. Then $A(1, 1, \cdots, 1)^T = (d, d, \cdots, d)^T = d(1, 1, \cdots, 1)^T$. Thus vector $(1, 1, \cdots, 1)^T$ is an eigenvector of A with corresponding eigenvalue d.

Here are some simple illustrations how certain properties of a d-regular graph are reflected in its spectrum:

- $\lambda_1 = d$ and the corresponding eigenvector is $\mathbf{x}_1 = \mathbf{1}/\sqrt{n} = (1/\sqrt{n}, \ldots, 1/\sqrt{n})$
- The graph is connected iff $\lambda_1 > \lambda_2$
- The graph is bipartite iff $\lambda_1 = -\lambda_n$

As shown in the following theorem, the graph's second eigenvalue is closely related to its expansion parameters.

Theorem 2 *(***Cheeger-Alon-Milman***) Let G be a d-regular graph with spectrum $\lambda_1 \geq \cdots \geq \lambda_n$, then*

$$\frac{d - \lambda_2}{2} \leq h(G) \leq \sqrt{2d(d - \lambda_2)}$$

Note that $d - \lambda_2$ is known as the **Spectral Gap**. The Theorem 2 implies that the spectral gap can give a good estimate on the expansion of a graph. Before proving Theorem 2, we first introduce how to compute eigenvalues using Rayleigh quotient.

Consider a matrix $A \in \mathbb{R}^{n\times n}$ and be symmetric with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. The Rayleigh quotient is defined as follows:

$$R_A(\mathbf{x}) = \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \quad \mathbf{x} \in \mathbb{R}^n$$

And we have:

$$\lambda_1 = \max_{\mathbf{x}\neq\mathbf{0}} R_A(\mathbf{x}) \text{ and } \quad \lambda_n = \min_{\mathbf{x}\neq\mathbf{0}} R_A(\mathbf{x})$$

More general, we have:

$$\lambda_k = \max_{\mathbf{x}\neq\mathbf{0},\mathbf{x}\perp\mathbf{x}_1,\cdots,\mathbf{x}_{k-1}} R_A(\mathbf{x}) = \min_{\mathbf{x}\neq\mathbf{0},\mathbf{x}\perp\mathbf{x}_{k+1},\cdots,\mathbf{x}_n} R_A(\mathbf{x})$$

Here are some proofs for you to read.

Lemma 3 *Let $A \in \mathbb{R}^{n\times n}$ and be symmetric. Then $\lambda_1 = \max_{\mathbf{x}\in\mathbb{R}^n,|\mathbf{x}|=\mathbf{1}}\{\mathbf{x}^T A\mathbf{x}\}$*

Proof. Remember that we still assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. We have $\mathbf{x}_1^T A\mathbf{x}_1 = \lambda_1\mathbf{x}_1^T\mathbf{x}_1 = \lambda_1$, thus $\max_{\mathbf{x}\in\mathbb{R}^n,|\mathbf{x}|=\mathbf{1}}\{\mathbf{x}^T A\mathbf{x}\} \geq \lambda_1$.

Conversely, let $\mathbf{x} \in \mathbb{R}^n$ and $|\mathbf{x}| = \mathbf{1}$ ($\mathbf{x}$ be any vector of length one. Note that a multiple of an eigenvector is also an eigenvector and therefore we can assume wlog that all the x_i have length one.) Let $\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n$, we have:

$$\mathbf{x}^T A\mathbf{x} = (\sum_i a_i\mathbf{x}_i)^T A(\sum_i a_i\mathbf{x}_i)$$

$$= (\sum_j a_i\mathbf{x}_j)^T(\sum_i \lambda_i a_i\mathbf{x}_i) = \sum_i \lambda_i a_i^2 \leq \max_i \lambda_i \sum_i a_i^2 = \lambda_1$$

It implies that $\max_{\mathbf{x}\in\mathbb{R}^n,|\mathbf{x}|=\mathbf{1}}\{\mathbf{x}^T A\mathbf{x}\} \leq \lambda_1$. □

Indeed, we can also prove that $\lambda_2 = \max_{\mathbf{x}\in\mathbb{R}^n,|\mathbf{x}|=\mathbf{1},\mathbf{x}\perp\mathbf{x_1}}\{\mathbf{x}^T A\mathbf{x}\}$. Can you? Shall we?

To do that, for the first step, we simply show that $\mathbf{x}_2^T A\mathbf{x}_2 = \lambda_2\mathbf{x}_2^T\mathbf{x}_2 = \lambda_2$, thus $\max_{\mathbf{x}\in\mathbb{R}^n,|\mathbf{x}|=\mathbf{1},\mathbf{x}\perp\mathbf{x_i}}\{\mathbf{x}^T A\mathbf{x}\} \geq \lambda_2$.

For the second step, we have $\mathbf{x}^T A\mathbf{x} = \sum_{i=2}^n \lambda_i a_i^2 = \lambda_2$.

Now, let us show that $\frac{2|E(G)|}{n} \leq \lambda_1 \leq \Delta(G)$ where $\Delta(G)$ is the max-degree of G.

Proof. Note that when $\mathbf{x}$ is an eigenvector of A with eigenvalue λ (that is, $A\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{x} \neq \mathbf{0}$, sometimes $\mathbf{x}$ is called λ-eigenvector of A), then we have:

$$\lambda x_u = \sum_{v \in \Gamma(u)} a_{uv} x_v, \forall u \in V(G)$$

Let u be a vertex in V with maximum value $|x_u|$, we have:

$$|\lambda||x_u| \leq \sum_{v \in \Gamma(u)} a_{uv}|x_v| \leq \deg(u)|x_u| \leq \Delta(G)|x_u|$$

Thus $|\lambda_1| \leq \Delta(G)$. We have proven the upper bound.

Now we show the lower bound, that is $\frac{2|E(G)|}{n} \leq \lambda_1$, by using the Rayleigh quotient as mentioned earlier. Let $\mathbf{1} = (1, \cdots, 1)$, we have:

$$\lambda_1 \geq \frac{\mathbf{1}^T A \mathbf{1}}{\mathbf{1}^T \mathbf{1}} = \frac{2|E(G)|}{n}$$

□

We are now ready to prove Theorem 2.

Proof. Let us prove the first part: $\frac{d-\lambda_2}{2} \leq h(G)$. Based on the definition of $h(G)$, we have:

$$h(G) = \min_{|S| \leq n/2} \frac{|E(S, \bar{S})|}{|S|} \geq \min_{|S| \leq n/2} \frac{|E(S, \bar{S})|}{|S|\frac{2|\bar{S}|}{n}} = \frac{1}{2} \min_{S \subset V} \frac{|E(S, \bar{S})|}{\frac{1}{n}|S||\bar{S}|} = \frac{1}{2}\phi(G)$$

where $\phi(G) = \min_{S \subset V} \frac{|E(S, \bar{S})|}{\frac{1}{n}|S||\bar{S}|}$

Consider a vector $\mathbf{x} \in \{0, 1\}^n$ represent a set of vertices in S and $\bar{S}$. Imagine that each vertex in S has a value 1 and each vertex in $\bar{S}$ has a value 0. Then we can re-write $\phi(G)$ as follows:

$$\phi(G) = \min_{\mathbf{x} \in \{0,1\}^n, \mathbf{x} \neq \mathbf{0}, \mathbf{x} \nparallel \mathbf{1}} \frac{\sum_{uv \in E}(x_u - x_v)^2}{\frac{1}{2n}\sum_u \sum_v (x_u - x_v)^2}$$

Since $\{0, 1\}^n \in \mathbb{R}^n$, we have:

$$\phi(G) \geq \min_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq\mathbf{0}, \mathbf{x}\not\parallel\mathbf{1}} \frac{\sum_{uv\in E}(x_u - x_v)^2}{\frac{1}{2n}\sum_u\sum_v(x_u - x_v)^2}$$

Replace $\mathbf{x} \in \mathbb{R}^n$ by $\mathbf{y} \in \mathbb{R}^n$ such that $y_u = x_u - \frac{1}{n}\sum_v x_v$. Note that $\sum_u y_u = 0$ (that is, $\mathbf{y}\perp\mathbf{1}$). Thus we have:

$$\phi(G) \geq \min_{\mathbf{y}\in\mathbb{R}^n, \mathbf{y}\neq\mathbf{0}, \mathbf{y}\perp\mathbf{1}} \frac{\sum_{uv\in E}(y_u - y_v)^2}{\frac{1}{2n}\sum_u\sum_v(y_u - y_v)^2}$$

Since $\mathbf{y}\perp\mathbf{1}$, we have $\frac{1}{2n}\sum_u\sum_v(y_u - y_v)^2 = \mathbf{y}^T\mathbf{y}$.

Also G is a d-regular graph, we have: $\sum_{uv\in E}(y_u - y_v)^2 = d\mathbf{y}^T\mathbf{y} - \mathbf{y}^T A\mathbf{y}$. Therefore, we have:

$$\phi(G) \geq \min_{\mathbf{y}\in\mathbb{R}^n, \mathbf{y}\neq\mathbf{0}, \mathbf{y}\perp\mathbf{1}} \left(d - \frac{\mathbf{y}^T A\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\right) = d - \max_{\mathbf{y}\in\mathbb{R}^n, \mathbf{y}\neq\mathbf{0}, \mathbf{y}\perp\mathbf{1}} \frac{\mathbf{y}^T A\mathbf{y}}{\mathbf{y}^T\mathbf{y}} = d - \lambda_2$$

Thus we obtain the lower bound.

Now, we prove the upper bound $h(G) \leq \sqrt{2d(d-\lambda_2)}$. We will prove the following three claims:

Claim 1. For all $\mathbf{y} \in \mathbb{R}^\mathbf{n}$, we have:

$$\sum_{u,v} a_{uv}|y_u^2 - y_v^2| \leq \sqrt{2d\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T A\mathbf{y}}\sqrt{4d\mathbf{y}^T\mathbf{y}}$$

To prove Claim 1, we will use the Cauchy-Schwarz inequality.

$$\begin{aligned}
\sum_{u,v} a_{uv}|y_u^2 - y_v^2| &= \sum_{u,v} a_{uv}|y_u - y_v||y_u + y_v| \\
&\leq \sqrt{\sum_{u,v} a_{uv}(y_u - y_v)^2}\sqrt{\sum_{u,v} a_{uv}(y_u + y_v)^2} \\
&\leq \sqrt{\sum_{u,v} a_{uv}(y_u - y_v)^2}\sqrt{\sum_{u,v} 2a_{uv}(y_u^2 + y_v^2)} \\
&= \sqrt{2d\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T A\mathbf{y}}\sqrt{4d\mathbf{y}^T\mathbf{y}}
\end{aligned}$$

Claim 2. Suppose $\mathbf{x}$ is an eigenvector corresponding to eigenvalue λ_2 ($A\mathbf{x} = \lambda_2\mathbf{x}$) such that $|\{v : x_v > 0\}| \leq n/2$. (For the second condition, we can replace $\mathbf{x} = -\mathbf{x}$ to make it satisfy.) Define a vector $\mathbf{y}$ by $y_v = \max\{x_v, 0\}$. Then $A\mathbf{y} \geq \lambda_2\mathbf{y}$.

This claim is easy to prove. If $x_v \geq 0$, then $(A\mathbf{y})_v \geq (A\mathbf{x})_v = \lambda_2 x_v = \lambda_2 y_v$. If $x_v < 0$, then $y_v = 0$ and $(A\mathbf{y})_v \geq 0$.

Claim 3. For the vector $\mathbf{y}$ defined in Claim 2, we have:

$$\sum_{u,v} a_{uv}|y_u^2 - y_v^2| \geq 2h(G)\mathbf{y}^T\mathbf{y}$$

Proof of Claim 3: Let us arrange the component of y in non-increasing order: $y_{v_1} \geq y_{v_2} \geq \cdots \geq y_{v_n}$.

Suppose t of these components are strictly positive, that is, $y_{v_t} > y_{v_{t+1}} = \cdots = y_{v_n} = 0$. Let K be the set where the jumps occur, that is $K = \{k \mid y_{v_k} > y_{v_{k+1}}\}$. We will rewrite the sum $\sum_{u,v} a_{uv}|y_u^2 - y_v^2|$ as follows:

$$2\sum_{i=1}^{t}\sum_{j=i+1}^{n} a_{v_iv_j}(y_{v_i}^2 - y_{v_j}^2) = 2\sum_{k\in K}\sum_{i\leq k}\sum_{j>\ k} a_{v_iv_j}(y_{v_k}^2 - y_{v_{k+1}}^2)$$

Note that the last equality follows the fact that in case there are several elements of K between two indices i and j, the sum $\sum_{k\in K, i\leq k<j}(y_{v_k}^2 - y_{v_{k+1}}^2)$ goes to $y_{v_i}^2 - y_{v_j}^2$

Now for each $k \in [1, n]$, let $L_k = \{v_i \mid i \leq k\}$. For $k = 0$, let $L_0 = \emptyset$. We have $\sum_{i\leq k}\sum_{j>\ k} a_{v_iv_j} \geq h(G)|L_k|$. Therefore,

$$2\sum_{k\in K}\sum_{i\leq k}\sum_{j>\ k} a_{v_iv_j}(y_{v_k}^2 - y_{v_{k+1}}^2) \geq 2\sum_{k\in K} h(G)|L_k|(y_{v_k}^2 - y_{v_{k+1}}^2)$$

$$= 2h(G)\sum_{k\in K}(|L_k| - |L_{k'}|)y_{v_k}^2 = 2h(G)\sum_{k\in K}|\{v \mid y_v = y_{v_k}\}|y_{v_k}^2$$

$$= 2h(G)\sum_{v} y_v^2 = 2h(G)\mathbf{y}^T\mathbf{y}$$

where k' be the element of K preceding k.

Now using the above three claims, we can finish our proof as follows:

$$h(G) \leq \frac{\sum_{u,v} a_{uv}|y_u^2 - y_v^2|}{2\mathbf{y}^T\mathbf{y}} \leq \frac{\sqrt{2d\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T A\mathbf{y}}\sqrt{4d\mathbf{y}^T\mathbf{y}}}{2\mathbf{y}^T\mathbf{y}}$$

$$\leq \frac{\sqrt{2d\mathbf{y}^T\mathbf{y} - 2\lambda_2\mathbf{y}^T\mathbf{y}}\sqrt{4d\mathbf{y}^T\mathbf{y}}}{2\mathbf{y}^T\mathbf{y}} = \sqrt{2d(d-\lambda_2)}$$

□

4 Random Walks on Expander Graphs

Let us first introduction some definitions.

(n, d, α)-**graph**. Graph G is said (n, d, α)-graph if G is d-regular with n vertices and $\lambda = \max(|\lambda_2|, |\lambda_n|) \leq \alpha d$ where $\alpha < 1$.

Vectors and Norms. For two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the dot product of $\mathbf{u}, \mathbf{v}$ is defined as: $\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^n u_i v_i$. For a vector $\mathbf{u} \in \mathbb{R}^n$, the l_1, l_2, and l_∞ are defined as followed:

$$\|\mathbf{u}\|_1 = \sum_{i=1}^n |u_i|$$

$$\|\mathbf{u}\|_2 = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \left(\sum_{i=1}^n u_i^2\right)^{1/2}$$

$$\|\mathbf{u}\|_\infty = \max_{1 \leq i \leq n} |u_i|$$

Probability (or Distribution) Vectors. A vector $\mathbf{p} \in \mathbb{R}^n$ is a probability vector if for every $1 \leq i \leq n$, $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$. Here $\mathbf{p}$ represents a probability distributed over V. When $\mathbf{p} = \frac{1}{n}(1, \cdots, 1)$, $\mathbf{p}$ is the probability vector that corresponds to the uniform distribution. Let us denote a uniform distribution vector as $\mathbf{u}$.

Random Walk. A random walk over the vertices of G is a stochastic process defining a series of vertices $(v_0, v_1, \cdots, v_n)$ in which v_0 is a vertex of

G chosen by some initial distribution and v_{i+1} is chosen uniformly at random from the neighbors of x_i. Informally, a random walk has a starting point randomly chosen over V (maybe uniformly) and at each step, it chooses with equal probability one of the d edges incident to the current vertex.

(A random walk is in fact a Markov chain where the set of states of the chain is the set of vertices of the graph.)

Normalized Adjacency Matrix. Let A be an adjacency matrix of G, then its normalized adjacency matrix $\hat{A} = \frac{1}{d}A$.

That is, each entry $\hat{A}_{uv}$ equals to the number of edges between u and v divided by the degree, i.e., the probability that a random walk currently at u proceeds to v in the next step. Similarly, we have: $\hat{A}^2_{uv} = \sum_{w \in V} \hat{A}_{uw}\hat{A}_{wv}$ equals to the probability that a random walk at u moves to v in two steps. Generally, we have $\hat{A}^k_{uv}$ equals to the probability of a transition from u to v in k steps. (Thus sometimes, $\hat{A}$ is called a transition matrix.) Also note that if A has eigenvalues $\lambda_1 = d, \lambda_2, \cdots, \lambda_n$, then $\hat{A}$ has eigenvalues $1, \frac{\lambda_2}{d}, \cdots, \frac{\lambda_n}{d}$.

Also note that $\hat{A}\mathbf{p} = \sum_{v \in\ V} \hat{A}_{uv}p_u$ equals to the probability vector derived from sampling a starting point from distribution $\mathbf{p}$ and taking one step of the random walk defined by $\hat{A}$. Similarly, $\hat{A}^k\mathbf{p}$ equals to the distribution vector of a walk starting at a vertex sampled from $\mathbf{p}$ and taking k steps according to $\hat{A}$.

The following facts are easy to see:

- $\hat{A}$ is double stochastic, i.e., every column and every row sums up to 1
- $\max\{|\hat{\lambda}_2|, |\hat{\lambda}_n|\} = \alpha$ where $\hat{\lambda}_i$ is the i^{th} eigenvalue of $\hat{A}$.

4.1 A Random Walk on an Expander is Rapidly Mixing.

We will show that a random walk on the vertices of an expander mixes rapidly towards the stationary distribution.

Theorem 3 *The stationary distribution of a random walk on G is the uniform distribution*

Proof. It is easy to see since $\hat{A}\mathbf{u} = \mathbf{u}$.

□

Theorem 4 $\|\hat{A}^k\mathbf{p} - \mathbf{u}\|_1 \leq \sqrt{n} \cdot \alpha^k$ *for any distribution vector* $\mathbf{p}$.

The above Theorem states that it does not matter what the initial distribution of the random walk is, if $\alpha < 1$, we need to take only a logarithmic number of steps to get a distribution which is close to the uniform up to a polynomial factor.

Proof. Since $\hat{A}$ is symmetric, it has an orthonormal base $(\hat{x}_1, \cdots, \hat{x}_n)$. Let $\mathbf{p} = \mathbf{u} + \mathbf{v}$ where vector $\mathbf{v} = \mathbf{p} - \mathbf{u}$. Note that $\mathbf{u}$ is an uniform probability vector and $\sum_i p_i = \sum_i u_i = 1$, thus $\sum_i v_i = 0$ and $\mathbf{u} \cdot \mathbf{v} = 0$. So we have:

$$\hat{A}\mathbf{p} - \mathbf{u} = \hat{A}(\mathbf{u} + \mathbf{v}) - \mathbf{u} = \hat{A}\mathbf{u} + \hat{A}\mathbf{v} - \mathbf{u} = \mathbf{u} + \hat{A}\mathbf{v} - \mathbf{u} = \hat{A}\mathbf{v}$$

Therefore,

$$\|\hat{A}\mathbf{p} - \mathbf{u}\|_2 = \|\hat{A}\mathbf{v}\|_2 \leq \alpha\|\mathbf{v}\|_2 \leq \alpha\|\mathbf{p}\|_2$$

(Why the first inequality? Anything to do with the fact that $\mathbf{u} \cdot \mathbf{v} = 0$ and $\hat{\lambda} = \alpha$)

Thus $\|\hat{A}^k\mathbf{p} - \mathbf{u}\|_2 \leq \alpha^k$. And hence, $\|\hat{A}^k\mathbf{p} - \mathbf{u}\|_1 \leq \sqrt{n} \cdot \alpha^k$.

□

4.2 A Random Walk Yields a Sequence of Good Samples

Theorem 5 *Let* $G = (V, E)$ *be an* (n, d, α)*-graph and* $B \subset V$ *with density* $\beta = \frac{|B|}{|V|}$*, we have* $Pr[(B, k)] \leq (\beta + \alpha)^k$ *where* (B, k) *denotes an event that the random walk is completely contained in* B*.*

What does this theorem mean? Consider the following problem: There is a large set of good vertices in G (note that G is (n, d, α)-graph) and we wish to find one of them (for good samples). Let $B \subset V$ be the set of bad vertices. The theorem says that by choosing one vertex randomly, and then performing a random walk in G with length k, the probability that the random walk is completely contained in B is exponentially small in l. In other ways, using random walks can yield a good sample.

Let $v_0, \cdots, v_k$ be the vertices visited by the random walk under consider. Let P be the matrix (projection on the space of vectors supported in B) as

follows: $P_{uv} = 1$ if $u = v \in B$. Otherwise, $P_{uv} = 0$. In other words, $P\mathbf{x}$ sets all entries of $\mathbf{x}$ not in B to 0. We will prove Theorem 5 by proving the following two lemmas:

Lemma 4 $Pr[(B, k)] = \|(P\hat{A})^k P\mathbf{u}\|_1$

Proof. This proof is easy to see by the following arguments. Again, $P\mathbf{x}$ will set all the coordinated of vertices not in B to zeros. Thus the action of P over a probability vector $\mathbf{x}$ is to transform $\mathbf{x}$ into the residual probability vector of the same distribution, but conditioned on being in B. It implies that $P\mathbf{u}$ is the residual probability of the uniform distribution conditioned to be in B. $\hat{A}B\mathbf{u}$ is the residual distribution after a random step has been taken. $P\hat{A}P\mathbf{u}$ is the residual probability conditioned on the random step remaining in B. Then $(P\hat{A})^k P\mathbf{u}$ is the residual probability vector of the random initial point and all k steps being in B. Note that we do not consider the ending point of the random walk in B, thus $Pr[(B, k)] = \|(P\hat{A})^k P\mathbf{u}\|_1$

□

Lemma 5 *For any non-negative vector* $\mathbf{x} \in \mathbb{R}^n$*, we have:*

$$\|P\hat{A}P\mathbf{x}\|_2 \leq (\beta + \alpha) \cdot \|\mathbf{x}\|_2$$

Proof. Decompose $P\mathbf{x}$ into two parts: $P\mathbf{x} = (P\mathbf{x})_{\|} + (P\mathbf{x})_{\perp}$ where $(P\mathbf{x})_{\|}$ is the projection of $P\mathbf{x}$ on $\mathbf{u}$, that is $(P\mathbf{x})_{\|} = (\frac{1}{n}\sum_{v\in V}(P\mathbf{x})_v) \cdot \mathbf{1}$ and $(P\mathbf{x})_{\perp} = (P\mathbf{x}) - (P\mathbf{x})_{\|}$. As before, it is easy to see that $(P\mathbf{x})_{\|} \cdot (P\mathbf{x})_{\perp} = 0$.

By the triangle inequality, we have:

$$\|P\hat{A}P\mathbf{x}\|_2 \leq \|P\hat{A}(P\mathbf{x})_{\|}\|_2 + \|P\hat{A}(P\mathbf{x})_{\perp}\|_2$$

Now, we bound the first term $\|P\hat{A}(P\mathbf{x})_{\|}\|_2$. Note that $\hat{A}(P\mathbf{x})_{\|} = (P\mathbf{x})_{\|}$ due to the construction of $(P\mathbf{x})_{\|}$. Also, $(P\mathbf{x})_{\|}$ is of the form $(a, a, \cdots, a)$ where a is some scalar, thus $\|P(P\mathbf{x})_{\|}\|_2 = \sqrt{\beta} \cdot \|(P\mathbf{x})_{\|}\|_2$.

However, we have:

$$\|(P\mathbf{x})_{\|}\|_2 = \sqrt{n\left(\frac{1}{n}\sum_{v\in V}(P\mathbf{x})_v\right)^2} = \sqrt{\frac{1}{n}\left(\sum_{v\in V}(P\mathbf{x})_v\right)^2}$$

$$= \frac{1}{\sqrt{n}} \sum_{v \in V} (P\mathbf{x})_v$$

By Cauchy-Schwarz, we obtain:

$$\|(P\mathbf{x})_{\|}\|_2 \leq \frac{1}{\sqrt{n}} \sqrt{|B|} \sqrt{\sum_{v \in V} x_v^2} = \sqrt{\beta} \|\mathbf{x}\|_2$$

Therefore, we have:

$$\|P\hat{A}(P\mathbf{x})_{\|} \leq \sqrt{\beta} \|\mathbf{x}\|_2$$

Now we bound the second term $\|P\hat{A}(P\mathbf{x})_{\perp}\|_2$. Recall that $\hat{A}$ has the largest eigenvalue 1 and other eigenvalues $\hat{\lambda}_2/d, \cdots, \hat{\lambda}_n/d$, $|\hat{\lambda}_i/d| \leq \alpha$ for $1 \leq i \leq n$. Let $\mathbf{y_1}, \mathbf{y_2}, \cdots \mathbf{y_n}$ be an orthonormal basis of eigenvectors for $\hat{A}$. Let $\mathbf{z}$ be any vector perpendicular to $\mathbf{1}$ and spanned by $(\mathbf{y_1}, \cdots, \mathbf{y_n})$, then we have:

$$\|\hat{A}\mathbf{z}\|_2 = \|\frac{1}{d} \sum_{i=2}^{n} \hat{\lambda}_i z_i y_i\|_2 \leq \frac{1}{d} \sqrt{\sum_{i=2}^{n} z_i^2 \hat{\lambda}_i^2}$$

$$= \max_{i=2,\cdots,n} |\frac{\hat{\lambda}_i}{d}| \sqrt{\sum_{i=2}^{n} z_i^2} \leq \alpha \|\mathbf{z}\|_2$$

Apply the above inequality, we have:

$$\|P\hat{A}(P\mathbf{x})_{\perp}\|_2 \leq \|\hat{A}(P\mathbf{x})_{\perp}\|_2 \leq \alpha \|(P\mathbf{x})_{\perp}\|_2 \leq \alpha \|\mathbf{x}\|_2$$

The last inequality dues to the fact that $\|(P\mathbf{x})_{\perp}\|_2 \leq \|\mathbf{x}\|_2$.
Finally, we finish the proof:

$$\|P\hat{A}P\mathbf{x}\|_2 \leq \|P\hat{A}(P\mathbf{x})_{\|}\|_2 + \|P\hat{A}(P\mathbf{x})_{\perp}\|_2 \leq (\alpha + \beta)\|\mathbf{x}\|_2$$

□

Then Theorem 5 follows directly from the above two lemmas:

$$Pr[(B,k)] = \|(P\hat{A})^k P\mathbf{u}\|_1 \leq \sqrt{n} \cdot \|(P\hat{A})^k P\mathbf{u}\|_2 \quad \text{(by Cauchy-Schwarz)}$$

$$= \sqrt{n} \cdot (P\hat{A}P)^k \mathbf{u}\|_2 \leq \sqrt{n} \cdot (\alpha + \ \beta)^k \|\mathbf{u}\|_2 = (\alpha + \ \beta)^k$$

This completes the proof of Theorem 5.

□

4.3 Amplification of the Success Probability

4.3.1 One Sided Error

Let L be some language in co-RP and assume that randomized algorithm A running in poly time decides whether $x \in L$ with one sided error, that is:

- If $x \in \ L$, then $Pr[A \text{ accepts } x] = 1$
- If $x \notin L$, then $Pr[A \text{ accepts } x] \leq 1/2$

To reduce the error probability, we can run A t times independently and accept x iff A accepts for all t times. That is, if $x \notin L$, then $Pr[A \text{ accepts } x] \leq 1/2^t$. Thus the error probability is reduced exponentially. However, if the number of random bits that A needs for one time is r bits, then the total random bits for t times is tr. Can we reduce this number?

Using expanders, we can exponentially reduce the error probability while using a much smaller number of bits. Consider a $(2^r, d, \alpha)$-graph G, each node in G associates with r random bits. Perform a random walk with length t on G based on what we have discussed before. Then we can run A t times using the random bits on t vertices of this random walk and accept x iff A accepts for all t times. Then the total number of random bits is $r + t \log_2 d$ and the error probability is bounded by $(\alpha + \ \beta)^t$ based on Theorem 5 where $\beta = 1/2$.

4.3.2 Two Sided Error

What happens if A is a two-sided error algorithm? Can we directly apply the sequential repetition or repetition using expanders? How about PCP verifiers with completeness not equal to 1?

Now let L be a language in BPP and assume that A is a randomized algorithm running in poly time decides whether $x \in L$ with a two sides error probability as follows:

- If $x \in L$ then $Pr[A \text{ accepts } x] \geq 2/3$
- If $x \notin L$, then $Pr[A \text{ accepts } x] \leq 1/3$

We cannot directly apply the above simple sequence repetition. (**Why?**) Instead, we will modify it as follows: Let A runs t times and accept x if A accepts majority of all t times, else, do not accept x.

How do we analyze this method? Clearly, we will use tr bits. However, how much can we reduce the error probability? We will prove this using the Chernoff bound.

(Recall that Chernoff bound is defined as follows:

Let $X_1, \cdots, X_n$ be independent random variables with $Pr[X_i = 1] = p$ (and $Pr[X_i = 0] = 1 - p$). Note that $E[\sum_{i=1}^{t} X_i] = tp$. For all $\delta \in (0, 1)$, we have:

$$Pr[\sum_{i=1}^{t} X_i \geq (1+\delta)tp] \leq e^{-tp\delta^2/3}$$

$$Pr[\sum_{i=1}^{t} \leq (1-\delta)tp] \leq e^{-tp\delta^2/2}$$

)

Consider $x \in L$. Let $X_i = 1$ iff A does **not** accept x at the time $i, 1 \leq i \leq t$. Else, let $X_i = 0$. Let p be the probability that A does not accept x. Note that $p = Pr[X_i = 1] \leq 1/3$. Also A does not accept x iff $\sum_i X_i \geq t/2$. Therefore, the probability that $x \in L$ is not accepted is bounded by:

$$Pr[\sum_{i=1}^{t} X_i \geq t/2] \leq Pr[\sum_{i=1}^{t} X_i \geq (1+1/2)tp] \leq e^{-tp(1/2)^2/3}$$

Likewise for $x \notin L$ (but accepted - false positive).

How about expanders? Can we simply use Theorem 5 as above? If not, why?

Indeed, we also need to modify this Theorem in order to reduce the error probability in this case. Consider an (n, d, α)-graph G with a random walk of length t as mentioned above. Let A run t times according to t vertices in the random walk and accept x if A accepts majority of all t times. To

analyze the bound of error probability, we need to analyze the probability of a random walk with length t that is completely contained in B at least $t/2$. (Note that, this random walk can walk in and walk out from B several times, as long as the total number of steps that are in B at least $t/2$. We need to modify Theorem 5 as follows:

Theorem 6 *Let $B_0, B_1, \cdots, B_t$ be subsets of V such that $\beta_i = |B_i|/n$. Define (B, t) to be the event that a random walk $(v_0, v_1, \cdots, v_t)$ has the property that $\forall i, \quad v_i \in B_i$, we have:*

$$Pr[(B,t)] \leq \prod_{i=0}^{t-1} (\sqrt{\beta_i \beta_{i+1}} + \alpha)$$

Proof. The proof is similar to that of Theorem 5. Let P_i be the projection matrix corresponding to B_i. Then Lemma 4 can be modified as follows:

$$Pr[(B,t)] = \| \prod_{i=1}^{t} (P_i \hat{A}) P_0 \mathbf{u} \|_1$$

And the analogue of Lemma 5 is:

$$\|P_{i+1} \hat{A} P_i \mathbf{x}\|_2 \leq (\sqrt{\beta_i \beta_{i+1}} + \alpha) \|\mathbf{x}\|_2$$

□

Based on Theorem 6, we can achieve an exponential reduction in the error probability using only $r + O(t)$ random bits.