Expander Graphs and their Applications

Lecture notes for a course by Nati Linial and Avi Wigderson
The Hebrew University, Israel.

January 1, 2003

Contents

List of Figures

Chapter 1

The Magical Mystery Tour

Notes taken by Ran Gilad-Bachrach

Summary:

Since the introduction of *Expander Graphs* during the 1970's they turn to be a significant tool both in theory and practice. They have been used in solving problems in communication and construction of error correcting codes as well as a tool for proving results in number theory and computational complexity. In this course we will explore *Expander Graphs,* both the properties and the use of such graphs will be studied.

The goal of this lecture is to sample the wide range of applications for expander graphs. This should serve as a motivation for the rest of the course.

1.1 Some Problems

To begin our tour we will look at three questions from three very different domains. Note that in these problems the connection to graph theory, and especially to expander graphs is not clear.

1.1.1 Hardness results for linear transformation

Maybe the most important open problem in mathematics these days is the famous $P = NP$ (or $P \neq NP$) problem. Although it has been studied for decades now, almost no significant progress has been made. One of the reasons lies in the fact that we have very few problems that are known to be hard. During the 1970'th, Leslie Valiant [Val76] addressed this problem. He defined the following simple problem:

Problem 1. Let $\mathcal{F}$ be a finite field. Let A be a linear transformation over $\mathcal{F}$, i.e. A is an $n \times n$ matrix. We would like to build a circuit which computes the transformation $x \mapsto Ax$. Each gate of this circuit computes addition or multiplication. How many gates do we need in this network?

Assume for instance that the transformation A represent the Fourier Transform. Cooley and Tukey [CW65] presented the Fast Fourier Transform (FFT) which computes the transformation using $O(n \log n)$ gates. However there is no matching lower bound so it might be possible to do the computation using only $O(n)$ gates. The implications of a Very Fast Fourier Transform, i.e an $O(n)$ algorithm for computing the transform are hard to over estimate.

By counting the number of circuits and comparing to the number of linear transformations it could be verified that the average size of such circuit is $O\left(n^2/\log n\right)$ gates, however we don't know of any transformation which needs more then $O(n)$ gates.

Valiant [Val76] tried to present transformations for which the number of gates needed is greater then $O(n)$. He suggested that super regular transformation have this property:

Definition 1.1 (Super Regular Matrix). A matrix A is *Super Regular* if any rectangular sub-matrix of A has full rank.

Figure 1.1: Leslie G. Valiant

The main observation of Valiant was that if we look at the graph layout of a circuit which computes a super regular matrix then this graph is a *Super Concentrator*:

Definition 1.2 (Super Concentrator). A graph G is a *Super Concentrator* if it has n input vertexes denoted by I and n output vertexes denoted by O such that for every k and every $S \subseteq I$ and $T \subseteq O$ of size k (i.e. $|S| = |T| = k$) there exists k paths in G from S to T which are vertexes disjoint.

Valiant conjectured that any *Super Concentrator* graph must have $\gg n$ edges and hence any circuit which computes a super regular matrix must have $\gg n$ gates. However, Valiant himself disproved the conjecture and presented super concentrators with $O(n)$ edges, and as you might have guessed this is where expanders come into the picture.

For the moment we will skip to a totally different problem.

1.1.2 Error Correcting Codes

One of the most fundamental problems in communication is noise. Assume that Alice has a message of k bits which she would like to deliver to Bob over some communication channel. The problem is that the channel might interfere in the way and thus the message that Bob receives might be different then the one that Alice sent.

During the 1940's Clude Elwood Shannon has developed the theory of communication which is called Information Theory. In his innovative paper "A Mathematical Theory of Communication" [Sha48] the problem of communication over noisy channel is one of the problems he addressed. Let us first define the problem[1]:

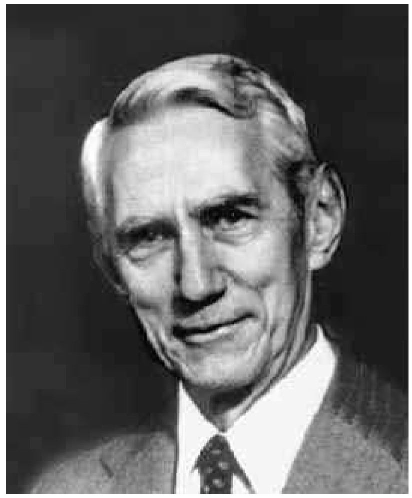
Problem 2 (communication over noisy channel). Alice and Bob can communicate over a noisy channel that might change a proportion p of the bits sent through it. How can Alice send Bob a message of k bits?

Shannon presented an answer to the above given question. He suggested building a dictionary (or code) $C \subseteq \{0,1\}^n$ such that $|C| = 2^k$. Every k-bits message is encoded by a code word in C and transmitted. Bob receives n bits and finds the closest code word in C 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). A good dictionary is one that is both big (i.e. $|C|$ is big) and at the same time the length of the words in C is small. This is seen from the next definition:

[1] The problem as described here is a simplification of the original problem presented by Shannon.

Figure 1.2: Clude Shannon.



Definition 1.3 (the rate and distance of a dictionary). Let $C \subseteq \{0,1\}^n$ be a dictionary. The **rate** of the dictionary is defined as

$$R = \frac{\log |C|}{n}$$

while the **distance** of the code is

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

where d_H is the hamming distance.

As we saw before the distance of a dictionary governs it's ability to overcome noisy channels while the rate counts the efficiency of the code. At this point we can refine the problem just state:

Problem 3 (refined communication problem). Is it possible to design a series of dictionaries $\{C_k\}_{k=1}^{\infty}$ 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 $R_0 > 0$.

We will see that a solution to this problem can be found using expander graphs. However, we will now present yet another problem.

1.1.3 De-randomizing Algorithms

Rabin [Rab80] presented in 1980 an algorithm for checking primality. 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 $f(x, r) = 1$, on the other hand if x is not primal $f(x, r) = 1$ with probability smaller then $1/4$. Applying this algorithm over and over again can reduce the error to be arbitrary small. However this process involves the use of more and more random bits.

The primality test is a special case of a Random Polynomial algorithm (RP). Let $L \subseteq \{0,1\}^k$ be a language. An algorithm which decides on $x \in \{0,1\}^k$ weather it is in L or not is Random Polynomial, if it runs in polynomial time and using poly(k) random bits and gives an answer which is always one if $x \in L$ and has probability smaller than $1/4$ to give 1 if $x \notin L$.

Problem 4 (Saving Random Bits). Assume that that $L \subseteq \{0,1\}^k$ has a random polynomial algorithm. How many random bits are needed in order to give an answer with probability of mistake smaller then $1/d$?

Figure 1.3: Michael Rabin

1.2 Magical Graphs

In the previous section we presented three problems which seems to be unrelated. We will now present a new object called Magical Graph which will enable us to find a solution for all these problems.

Definition 1.4 (Magical Graph). Let G be a two sided graph with n vertexes on each side. Let L be the vertexes on the left side and R the vertexes on the right. Assume that any vertex in L has d neighbors in R. We say that G is (d, n) **magical graph** if it has the following two properties

1. For any $S \subseteq L$ such that $|S| \leq \frac{n}{3d} \Longrightarrow |\Gamma(S)| \geq |S| \frac{d}{4}$
2. For any $S \subseteq L$ such that $\frac{n}{3d} < |S| \leq \frac{n}{2} \Longrightarrow |\Gamma(S)| \geq |S| + \frac{n}{3d}$

where $\Gamma(S)$ is the set of neighbors of S in G.

We will now turn to explore some properties of magical graphs.

Lemma 1.5. *For each $d \geq 8$ and sufficiently large n there exists a (d, n) magical graph.*

Proof. Construct a random graph as follows: for each vertex $v \in L$ choose randomly d vertexes in R and connect them with v. We claim that with high probability the graph generated by this process is a magical graph.

Let $S \subseteq L$ be such that $s = |S| \leq \frac{n}{3d}$. Let $T \subseteq R$ be such that $t = |T| < |S| \frac{d}{4}$.

Let $X_{S,T}$ be an indicator random variable for the event that all the edges from S go to T. It is clear that if $\sum X_{S,T} = 0$ then the first property in the definition of magical graphs hold.

The probability of the event $X_{S,T}$ is $\left(\frac{t}{n}\right)^{sd}$ and therefore using a union bound

$$
\begin{aligned}
\Pr\left[\sum_{S,T} X_{S,T} > 0\right] &\leq \sum_{S,T} \Pr\left[X_{S,T} = 1\right] \\
&= \sum_{S,T} \left(\frac{t}{n}\right)^{sd} \\
&\leq n^2 \binom{n}{\frac{n}{3d}} \binom{n}{\frac{n}{12}} \left(\frac{1}{12}\right)^{n/3} \\
&\cong n^2 2^{n\mathcal{H}\left(\frac{1}{3d}\right) + n\mathcal{H}\left(\frac{1}{12}\right) + \frac{n}{3}\log\frac{1}{12}} \\
&\leq n^2 2^{-n/8}
\end{aligned}
$$

When n is sufficiently large $n^2 2^{-n/8}$ is smaller then 0.25 and therefore the probability that requirement 1 in the definition of magical graph will holds is greater than 0.75.

We use the same technique to bound the probability that requirement 2 in the definition of magic graph hold. For every $S \subset L$ such that $\frac{n}{3d} < |S| \leq \frac{n}{2}$ and $T \subset R$ such that $|T| < |S| + \frac{n}{3d}$ let $Y_{S,T}$ be an indicator random variable for the event that all the edges from S go to T. As in the previous case, if $\sum Y_{S,T} = 0$ then the second property in the definition of magical graphs hold.

The probability of the event $Y_{S,T}$ is $\left(\frac{t}{n}\right)^{sd}$ and therefore using a union bound

$$\begin{aligned}
\Pr\left[\sum_{S,T} Y_{S,T} > 0\right] &\leq \sum_{S,T} \Pr\left[Y_{S,T} = 1\right] \\
&= \sum_{S,T} \left(\frac{t}{n}\right)^{sd} \\
&\leq n^2 \binom{n}{\frac{n}{2}} \binom{n}{\frac{n}{2}} \left(\frac{\frac{n}{2} + \frac{n}{3d}}{n}\right)^{\frac{n}{2}d} \\
&\cong n^2 2^{2n\mathcal{H}\left(\frac{1}{2}\right) + \frac{nd}{2}\log\left(\frac{1}{2} + \frac{1}{3d}\right)} \\
&\leq n^2 2^{n\left(2 - \frac{d}{3}\right)}
\end{aligned}$$

For n sufficiently large $n^2 2^{n\left(2-\frac{d}{3}\right)} \leq 0.25$ and hence the second property of magical graph holds with probability 0.75 at least.

Finally if we chose a sufficiently large n we get that the two requirements of magical graphs hold with probability greater than 0.5. Therefore not only that there exist an (n, d) magic graph but there are many of those.

□

Before introducing the solutions to the above mentioned problems, we will present a small variation on magical graphs. We will delete $\frac{n}{4d^2}$ vertexes from R, i.e. the right side of the graph such that the main properties of the graph will remain:

Lemma 1.6. *Let G be a magical 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. We will present an algorithm which constructs the set B. We begin by holding the two sets of vertexes $L_0 = L$ and $R_0 = R$ and we reset B to be the empty set.

At each iteration we choose $v \in R_i$ with degree at most $2d$. We add v to B and then construct L_{i+1} such that $L_{i+1} = L_i \setminus \Gamma(v)$ and $R_{i+1} = R_i \setminus \Gamma(\Gamma(v))$, i.e. we delete all the neighbors of v from L_i and delete all the second-degree neighbors of v from R_i. We keep the process running as long as there is a vertex $v \in R_i$ with degree at most $2d$.

From the way we constructed the set B it is clear that any $u \in L$ has at most one neighbor in B. We would like to count how many iterations can we do with the above algorithm, this will give us a lower bound on the size of B.

At each step we have that $|L_i| \geq n - 2di$ since the vertex v which we add to B has degree at most $2d$. Also we have that $|R_i| \geq n - d(n - |L_i|)$ since each vertex in L has degree d. Since the number of vertexes in the graph induced on $L_i \cup R_i$ is at most $d|L_i|$, the average degree of the vertexes in R_i is at most

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

and hence as long as $|L_i| \geq \left(1 - \frac{1}{2d-1}\right) n$ the average degree of the vertexes in R_i is at most $2d$ and therefore there exist a vertex with degree at most $2d$.

Since $|L_i| \geq n - 2di$ we have that as long as $i \leq \frac{n}{4d^2}$ that $|L_i| \geq \left(1 - \frac{1}{2d-1}\right) n$ as required. Therefore we can build a set B of size at $\frac{n}{4d^2}$.

□

We will call **modified magical graph** a magical graph that a set of size $\frac{n}{4d^2}$ of vertexes were deleted from the right side as described in lemma 1.6.

Since magical graphs exists as we saw in lemma 1.5 and can be modified as we saw in lemma 1.6 we now turn to use this construction to solve the problems presented in the first section of this lecture.

1.2.1 A Super Concentrator with $O(n)$ edges

As we saw in section 1.1.1 Valiant's conjecture was that a super concentrator must have many edges. This sounds reasonable from the definition of super concentrators (see definition 1.2). However we will see that using magical graphs it is possible to build such graphs with only $O(n)$ edges.

Let G be a modified magical graph such that there are n vertexes on the left side of G but only $n - \frac{n}{4d^2}$ vertexes on the right side as we saw in lemma 1.6.

By the construction of G we have that for each $S \subset L$ such that $|S| \leq \frac{n}{2}$ has $|\Gamma(S)| \geq |S|$. Hence by Hall's theorem [Die97, Theorem 2.1.2] for any $|S| \leq \frac{n}{2}$ in L there is a perfect matching from S to $\Gamma(S)$. We will use these facts to build a super concentrator.

The construction can be presented recursively. Let n_0 be the minimal size of modified magical graph. If we are required to build a super concentrator with less then n_0 vertexes we just return the full two sided graph. The full two sided graph is a concentrator with n_0^2 edges (we will use notation $S(n)$ for the number of edges).

Assume we would like to build a super concentrator with $n > n_0$ vertexes. Let G be a modified magical graph with n vertexes on the left side and $n - \frac{n}{4d^2}$ vertexes on the right. Let C be a super concentrator with $n - \frac{n}{4d^2}$ input and output edges. Such a concentrator exists according to our induction assumption.

Using G and C we will construct a new concentrator with n inputs and outputs. The inputs of the new concentrator will be the left side of G. We connect the right hand side of G to the inputs of C. We will place another copy of G on the outputs of C, this is illustrated in figure 1.4(a). Finally we add direct edges between the two copies of G, each vertex on the left side of G we placed in the input is connected to the matching vertex on the left side of G we placed in the output as illustrated in figure 1.4(b).

We would like to show that the graph constructed is indeed concentrator and count the number of edges in this graph. Let S be a set of vertexes from the input of the new graph and T be vertexes on the output such that $|S| = |T| = k$. If $k \leq n/2$ then due to the properties of the modified magical graph G we know that $|\Gamma(S)| \geq |S|$ and $|\Gamma(T)| \geq |T|$. Using Halls marriage theorem it is possible to construct a perfect matching between S and $\Gamma(S)$ and on the other side between T and $\Gamma(T)$. Since C is a super concentrator, the matches of S in $\Gamma(S)$ and of T in $\Gamma(T)$ can be connected by k disjoint paths and hence S and T can be connected by disjoint paths.

If the two sets S and T are big, i.e. $|S| = |T| = k > n/2$ then there must exists at least $k - n/2$ vertexes in S that are matched to vertexes in T by direct edges. These edges form paths and hence we can exclude these vertexes from S and their matches from T. After doing so we are left with groups of size $n/2$ at most which we already know how to treat.

After we proved that the graph we constructed is a super concentrator we turn to count the number of edges. Let $S(n)$ be the number of edges in the graph with n inputs. From the construction we know that $S(n) = |C| + 2|G|$. We also know that $|G| = nd$ and $|C| = S\left(n\left(1 - \frac{1}{4d^2}\right)\right)$. Hence we obtain a recursive formula for $S(n)$:

1. for $n > n_0$ we have that $S(n) \leq 2nd + S\left(n\left(1 - \frac{1}{4d^2}\right)\right)$.
2. for $n \leq n_0$ we have that $S(n) \leq n^2$.

Solving this recursive formula we get

$$S(n) \leq cn$$

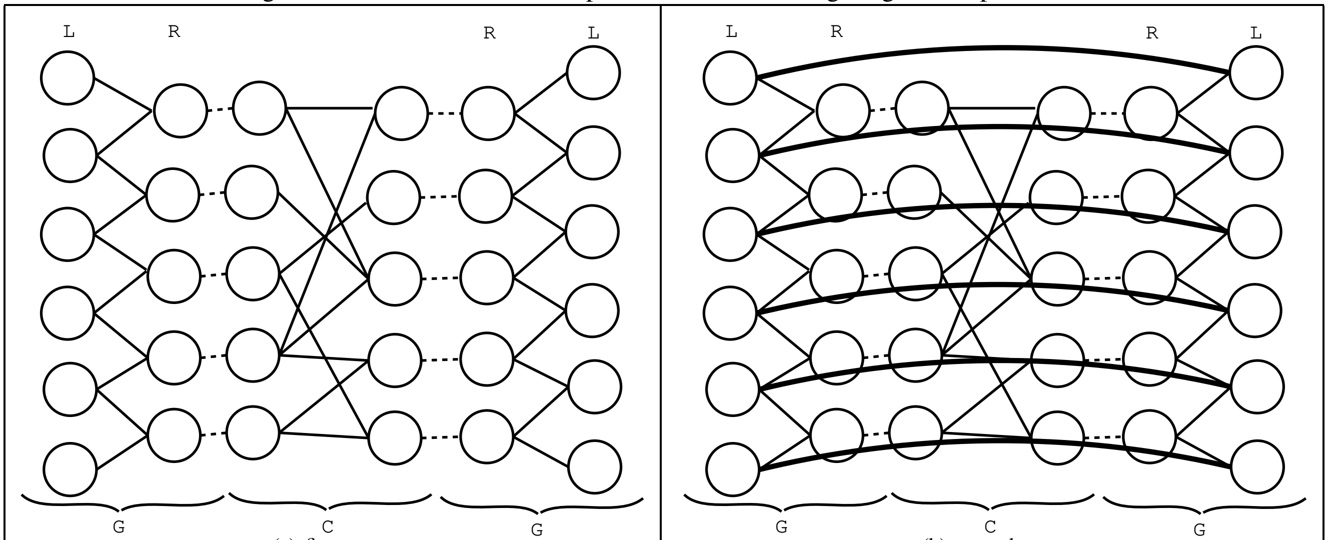
such that $c = n_0 + 8d^3$.

Therefore, using magical graphs it is possible to construct super concentrators with $O(n)$ edges.

1.2.2 Error Correcting Codes

We now turn to present a solution to Shannon's problem of correcting errors over communication channels. Again let G be a modified magical graph, i.e. a magical graph such that it's right side consist of less then $n(1 - \frac{1}{d^2})$ vertexes.

Figure 1.4: A construction of Super Concentrators using Magical Graphs



(a) first stage (b) second stage

Let $S \subset L$ be such that $|S| \leq \frac{n}{3d}$ then there exists $v \in S$ and $u \in R$ such that v is the only neighbor of u in the group S, i.e. $u \in \Gamma(S) \setminus \Gamma(S \setminus \{v\})$. This follows since $|\Gamma(S)| \geq (\frac{3d}{4} - 1)|S| > \frac{d}{2}|S|$. To prove the existence of u and v, consider the set E of edges between S and $\Gamma(S)$. Then $d|S| \geq |E| \geq 2|\Gamma(S)|$ and so $|\Gamma(S)| \leq \frac{d}{2}|S|$ which is a contradiction.

We use this construction to build a code $C \subset \{0,1\}^n$ of size 2^k such that the hamming distance between any two distinct code words is at least $\frac{n}{3d}$ so the code has distance $\frac{1}{3d}$ and the rate is $\frac{k}{n} = \frac{1}{3d}$.

Let G be a modified magical graph. We will view the graph G as a function from $\{0,1\}^n$ to $\{0,1\}^{n(1-\frac{1}{d^2})}$ by assigning a parity function to each vertex in the right side, i.e. $G(x)_u = \oplus_{v \in \Gamma(u)} x_v$.

$$C = \{x \in \{0,1\}^n \,|G(x) = 0\}$$

I.e. a word x is in the code C if the parity assigned to each vertex on the right side is zero. Figure 1.5 demonstrates the code.

C is a linear sub space of $\{0,1\}^n$ defined by $n(1 - \frac{1}{4d^2})$ linear equations and hence $|C| \geq 2^{n/4d^2}$. Since C is a linear code (i.e. a linear sub-space) the minimal distance between two code words in C is the minimal weight of a non-zero code word in C. Let $x \in \{0,1\}^n$. We can look at x as an indicating function of vertexes in L. Let S be the set of vertexes to which x assigns the value 1. If $|S| < \frac{n}{3d}$ then there exists a unique neighbor, i.e there exists $u \in R$ and $v \in S$ such that v is the only neighbor of u in S. Hence the parity function associated with u will assign the value 1 to x and hence $x \notin C$. Therefore the minimal distance between two code words in C is at least $n/3d$.

Therefore we presented a way to construct "good" dictionaries (codes) using magical graphs.

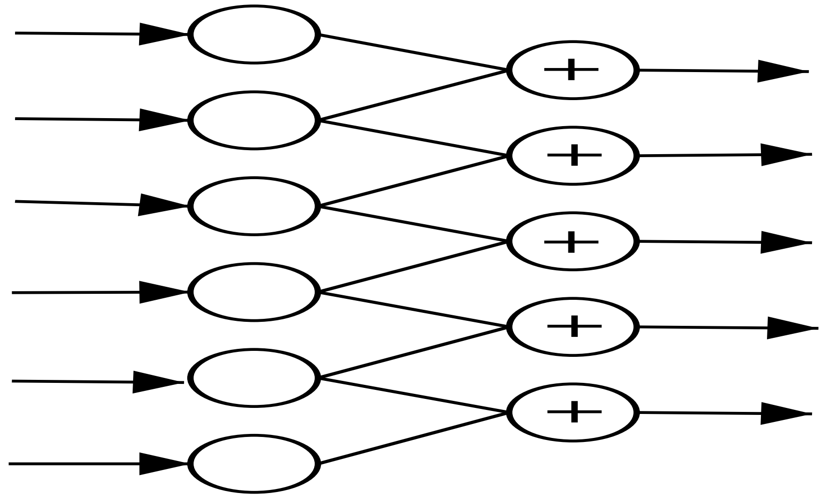
1.2.3 De-randomizing Random Algorithms

The last problem we presented was that of random algorithms. Let LANG be a language such that there exists an algorithm such that when it receives x of size k and a string r of k random bits, it calculates a function $f(x,r)$ such that if $x \in$ LANG then $f(x,r) = 1$ but if $x \notin$ LANG then $f(x,r) = 1$ with probability at most $1/12$ over the choice of r.

Let G be a magical graph over $n = 2^k$ vertexes. By the definition of such graphs we know that for every $S \subseteq L$ such that $|S| \geq \frac{n}{3d}$ we have that $|\Gamma(S)| \geq \frac{n}{12}$. Now assume that $x \notin$ LANG. Let B be the "bad" set i.e. $B = \left\{r \in \{0,1\}^k \mid f(x,r) = 1\right\}$. We know that $|B| \leq \frac{n}{12}$.

For each string of k random bits we assign a vertex in L. The graph G then direct us to d string of k bits which are associated to vertexes in R, we will call these strings $r_1, \ldots, r_d$. Our algorithm will apply the function $f(x, \cdot)$ with

Figure 1.5: A construction of an error correcting code



$r_1, \ldots, r_d$. If all the values we receive were 1 then we will predict that $x \in$ LANG. If $f(x, r_j) = 0$ for some j then we know for sure that $x \notin$ LANG.

Our algorithm will fail only if $x \notin$ LANG and $r_1, \ldots, r_d \in B$, i.e. $\Gamma(r) \subseteq B$. Let $S \subset L$ be the set of vertexes for which all their neighbors are in B, so $S = \{v \in L \mid \Gamma(v) \subseteq B\}$. Since $|B| \leq \frac{n}{12}$ we have that $|S| \leq \frac{n}{3d}$ and therefore our algorithm will fail with probability $\frac{1}{3d}$ at most, using only k random bits.

1.3 Conclusions

In the first section of this lecture we presented three problems of different nature. All these problem had no direct connection to graph theory. However we saw that by constructing magical graphs we could find a solution to these problems. During our discussion we explored some of the features of magical graphs.

What other magic can these graphs do? what properties do they have. Can we construct these graphs efficiently? All these questions are the topic of this course. We will explore both the theory and applications of magical graphs.

And one last word, the magical graphs we used in this lecture are a special case of the exciting family of ***Expanders***.

Chapter 2

Graph Expansion & Eigenvalues

Notes taken by Danny Harnik

Summary: After defining families of expander graphs and giving some examples of such families, we discuss some algebraic properties of graphs. Mainly, we discuss the connection between the expansion property of a graph to the eigenvalues of the graph's adjacency matrix. We also see an application of expander graphs for error amplification with a small amount of random bits.

2.1 Definitions

We begin with some notes and notations:

- Throughout this lecture (and course) we discuss d-**regular** graphs (graphs in which all vertices have the same degree d). denote a graph by $G = (V, E)$ and $|V| = n$. We allow self loops and multiple edges in the graph.
- Unlike the previous lecture, we discuss general graphs and not only bipartite graphs.
- For $S, T \subset V$ denote the set of all edges between S and T by $E(S, T) = \{(u, v) | u \in S, v \in T, (u, v) \in E\}$.

Definition 2.1.

1. The **Edge Boundary** of a set S, denoted ∂S, is $\partial S = E(S, \overline{S})$. This is actually the set of outgoing edges from S.
2. The **Expansion Parameter** of G, denoted $h(G)$, is defined as:

$$h(G) = \min_{\{S | |S| \leq \frac{n}{2}\}} \frac{|\partial S|}{|S|}$$

We note that there are other notions of expansion that can be studied. The most popular is counting the number of neighboring vertices of any small set S, rather than the number of outgoing edges.

Definition 2.2. Family Of Expander Graphs
A family of Expander graphs $\{G_i\}$ where $i \in \mathbf{N}$ is a collection of graphs with the following properties:

- The graph G_i is a d-regular graph of size n_i (d is the same constant for the whole family). $\{n_i\}$ is a monotone growing series that doesn't grow too fast (e.g. $n_{i+1} \leq n_i^2$).
- For all i, $h(G_i) \geq \epsilon > 0$.

When discussing a family of expander graphs one should also consider the time required to construct such a graph. There are two natural versions for the requirement on the constructibility of graphs:

Definition 2.3.

1. A family of expander graphs is called **Mildly Explicit** if there is a Polynomial-Time algorithm that given 1^i creates G_i.

2. A family of expander graphs is called **Very Explicit** if there is a Polynomial-Time algorithm that given (i, v, k) (where $i \in \mathbf{N}, v \in V$ and $k \in \{1, \cdots, d\}$) computes the kth neighbor of vertex v in the graph G_i.

The second definition is useful for very large graphs, where one cannot construct the whole graph, but rather works locally on a small part of the graph.

2.2 Examples of Expander Graphs

1. This family of graphs G_m lies on a grid: $V_m = \mathbf{Z}_m \times \mathbf{Z}_m$.
   The degree is $d = 4$ and the edges are described as follows:
   Vertex (x, y) has edges to $(x + y, y), (x - y, y), (x, y + x)$ and $(x, x - y)$ (all operations are done modulo m).
   Margulis (73) showed that this is an expander family.
   Gaber & Galil (80) showed that this is an ϵ-expander family (for a specific ϵ).

2. This family has graphs of size p (for all prime p). Here $V_p = \mathbf{Z}_p$ and $d = 3$. Each vertex x is connected to its neighbors and its inverse (i.e. $x + 1, x - 1$ and x^{-1}).
   This was shown to be an ϵ-expander family by Lubotsky, Philips and Sarnak (88).

2.3 The Spectrum of a Graph

The **Adjacency Matrix** of a graph G, denoted $A(G)$, is an $n \times n$ matrix that for each (u, v) contains the number of edges in G between vertex u and vertex v. Since the graph is d-regular, the sum of each row and column in $A(G)$ is d.

By definition the matrix $A(G)$ is symmetric and therefore has an orthonormal base $v_0, \cdots, v_{n-1}$, with eigenvalues $\mu_0, \mu_1, \cdots, \mu_{n-1}$ such that for all i we have $Av_i = \mu_i v_i$. Without loss of generality we assume the eigenvalues are sorted in descending order $\mu_0 \geq \mu_1 \geq \cdots \geq \mu_{n-1}$. The eigenvalues of $A(G)$ are called the **Spectrum** of the graph G.

The spectrum of a graph contains a lot of information regarding the graph. Here are some examples of observations that demonstrate this connection between the spectrum of a d-regular graph and its properties:

- $\mu_0 = d$
- The graph is connected iff $\mu_0 > \mu_1$
- The graph is bipartite iff $\mu_0 = -\mu_{n-1}$

In the rest of the lecture we will discuss the connection between the expansion of a graph and its spectrum. In particular, the graphs second eigenvalue is related to the expansion parameter of the graph.

Theorem 2.4.

$$\frac{d - \mu_1}{2} \leq h(G) \leq \sqrt{2d(d - \mu_1)}$$

This Theorem is due to Cheeger & Buser in the continuous case, and to Tamner, Alon & Milman in the discrete case.

The theorem actually proves that $d - \mu_1$, also known as the **Spectral Gap**, can give a good estimate on the expansion of a graph. Moreover, the graph is an expander ($h(G) > \epsilon$) if and only if the spectral gap is bounded ($d - \mu_1 > \epsilon'$). We do not prove this theorem at this stage (will be proved later in the course). Instead we show a Lemma that allows us to find connections between the expansion property and the second eigenvalue.

2.4 The Expander Mixing Lemma and Applications

Denote $\lambda = \max(|\mu_1|, |\mu_{n-1}|)$.
Since the eigenvalues are already sorted, this means that λ is larger than the absolute value of all eigenvalues (except $\mu_0 = d$).

Lemma 2.5. Expander Mixing Lemma *for all $S, T \subseteq V$:*

$$||E(S,T)| - \frac{d|S||T|}{n}| \leq \lambda\sqrt{|S||T|}$$

This lemma can be viewed as relating the second eigenvalue to the question of how "random" the graph is. The left hand side compares the expected number of edges between S and T in a random graph ($\frac{d|S||T|}{n}$) and the actual number of edges between the two sets ($|E(S,T)|$). This difference is small when λ is small. So a small λ (or large spectral graph) means a graph with allot of "randomness".

Proof. Denote by χ_S and χ_T the characteristic vectors of S and T (χ_S is a vector with ones for all $v \in S$ and zeros in all other places). Let $\chi_S = \Sigma_i \alpha_i v_i$ and $\chi_T = \Sigma_j \beta_j v_j$ be their representation as linear combinations of the orthonormal base $v_0, \cdots, v_{n-1}$, where $v_0 = \overline{1}/\sqrt{n}$. We have:

$$\begin{aligned} |E(S,T)| &= \chi_S A \chi_T \\ &= (\Sigma_i \alpha_i v_i) A (\Sigma_j \beta_j v_j) \end{aligned}$$

and since the v_i's are eigenvectors and orthonormal:

$$\begin{aligned} |E(S,T)| &= (\Sigma_i \alpha_i v_i)(\Sigma_j \beta_j A v_j) \\ &= (\Sigma_i \alpha_i v_i)(\Sigma_j \beta_j \mu_j v_j) \\ &= \Sigma_i \mu_i \alpha_i \beta_i \end{aligned}$$

Since $\alpha_0 = \langle \chi_S, \frac{\overline{1}}{\sqrt{n}} \rangle = \frac{|S|}{\sqrt{n}}$ and $\beta_0 = \frac{|T|}{\sqrt{n}}$:

$$\begin{aligned} |E(S,T)| &= \mu_0 \frac{|S||T|}{n} + \Sigma_{i=1}^{n-1} \mu_i \alpha_i \beta_i \\ &= d\frac{|S||T|}{n} + \Sigma_{i=1}^{n-1} \mu_i \alpha_i \beta_i \end{aligned}$$

Due to the triangle inequality and the definition of λ:

$$\begin{aligned} ||E(S,T)| - d\frac{|S||T|}{n}| &= |\Sigma_{i=1}^{n-1} \mu_i \alpha_i \beta_i| \\ &\leq \Sigma_{i=1}^{n-1} |\mu_i \alpha_i \beta_i| \\ &\leq \lambda \Sigma_{i=1}^{n-1} |\alpha_i \beta_i| \end{aligned}$$

And by the Cauchy-Schwartz inequality:

$$\begin{aligned} ||E(S,T)| - d\frac{|S||T|}{n}| &\leq \lambda ||\alpha||_2 ||\beta||_2 \\ &= \lambda ||\chi_S||_2 ||\chi_T||_2 \\ &= \lambda \sqrt{|S||T|} \end{aligned}$$

$\square$

Following is an example of an application of the lemma above:

2.4.1 Deterministic Error Amplification for BPP

In this example we are given a function in BPP. This means a function $f : \{0,1\}^n \to \{0,1\}$ and a probabilistic polynomial time algorithm $\mathcal{A}$ that approximates f in the sense that for random $r \in \{0,1\}^m$ we have:

$$Pr_r[\mathcal{A}(x,r) \neq f(x)] \leq \frac{1}{4}$$

Our goal is to reduce the probability of error. This can be achieved by simply repeating $\mathcal{A}$ with different random coins r, and taking a majority vote. Using t calls to $\mathcal{A}$ one can reduce the error to c^{-t} for some constant c. However this requires a large number of coin tosses ($t \cdot m$ in this case). The question is: can we make the error smaller with a small number of random coins?

We introduce an algorithm $\mathcal{B}$ that uses just m random coins: $\mathcal{B}$ uses a d-regular expander graph G_{2^m} (of size $N = 2^m$). The algorithm chooses a random vertex $v \in G_{2^m}$ and takes a majority vote on the output of $\mathcal{A}$ on each of the neighbors of v (denote by $\Gamma(v)$ the set of the neighbors of v in G_{2^m}).

$$\mathcal{B}(x,v) = Majority_{u\in\Gamma(v)}\mathcal{A}(x,u)$$

Claim 2.6.

$$Pr_v[\mathcal{B}(x,v) \neq f(x)] \leq 4(\frac{\lambda}{d})^2$$

Proof. Let S be the set of vertices that algorithm $\mathcal{B}$ makes errors on, and T be the set of vertices that algorithm $\mathcal{A}$ makes errors on ($S = \{v|\mathcal{B}(x,v) \neq f(x)\}$ and $T = \{u|\mathcal{A}(x,u) \neq f(x)\}$). By definition, every $v \in S$ has at least $\frac{d}{2}$ neighbors in T. So: $|E(S,T)| \geq |S|\frac{d}{2}$ On the other hand, due to the Expander Mixing Lemma and since $|T| \leq \frac{N}{4}$:

$$\begin{aligned} |E(S,T)| &\leq \frac{d|S||T|}{N} + \lambda\sqrt{|S||T|} \\ &\leq \frac{d|S|}{4} + \lambda\sqrt{|S|\frac{N}{4}} \end{aligned}$$

Combined together we get:

$$\frac{d|S|}{4} \leq \lambda\sqrt{|S|\frac{N}{4}}$$

And finally:

$$\begin{aligned} Pr[\mathcal{B}(x,v) \neq f(x)] &= \frac{|S|}{N} \\ &\leq 4(\frac{\lambda}{d})^2 \end{aligned}$$

□

The above claim shows that error amplification can be done, but gives a rather poor amplification rate. There are a few ways to improve this:

- One can take instead of G_{2^m}, with adjacency matrix A, the graph with matrix A^k. In this new graph (containing an edge for any path of length k in the original graph) we have degree d^k, but also second eigenvalue λ^k giving a reduced error of $(\frac{\lambda}{d})^{2k}$.
- It is worth noting that in a good expander we can achieve $\lambda \approx \frac{1}{\sqrt{d}}$ giving an error of about $\frac{1}{d}$. So to get an error smaller than a given ϵ one should use graphs of degree $\frac{1}{\epsilon}$.
- We will see in the next lecture constructions with better amplification using random walks on expander graphs.

2.5 How Big Can the Spectral Gap be?

We conclude with the question of how small can λ be.

As an example we can check the most connected graph - The Clique.
The n-clique K_n is the graph were every vertex is connected to all its neighbors and has degree $d = n - 1$. The spectrum of the Clique can be easily calculated by viewing the Clique's matrix as $J - I$ were J is the all ones matrix and I is the identity matrix. The spectrum of K_n is $[n-1, -1, -1, \cdots, -1]$. hence $\lambda = 1$.

But this is for $d \approx n$ and we are interested in the behavior of λ when d is much smaller (usually a constant). This case is discussed in the following Theorem due to Alon-Boppana:

Theorem 2.7. *for every d-regular graph:*

$$\lambda \geq 2\sqrt{d-1} - o_n(1)$$

We will not prove this theorem here, but instead show a weaker statement:

Claim 2.8. *for every d-regular graph:*

$$\lambda \geq \sqrt{d}(1 - o_n(1))$$

Proof.
note: In this discussion we don't allow multiple edges (the Adjacency matrix contains only zeros and ones).

Given a d-regular graph G with adjacency matrix A, we look at the trace of A^2 (trace is the sum of the values in the diagonal) . On one hand, since the matrix is symmetric, and the sum of each row/column is d we have $(A^2)_{ii} = d$ for all i:

$$Trace(A^2) = n \cdot d.$$

On the other hand:

$$\begin{aligned} Trace(A^2) &= \sum_i \mu_i^2 \\ &\leq d^2 + (n-1)\lambda^2 \end{aligned}$$

together we get

$$\lambda^2 \geq d\frac{n-d}{n-1}$$

and:

$$\lambda \geq \sqrt{d}(1 - o_n(1))$$

$\square$

Chapter 3

Random Walks on Expander Graphs

Notes taken by Boaz Barak and Udi Wieder

Summary: In this lecture we consider random walks on expander graphs. We will see that the t vertices on a length t random walk on a expander graph "look like" (in some respects) t random independently chosen vertices. This occurs even though sampling a length t walk on a (constant-degree) expander requires a significantly smaller number of random bits than sampling t random vertices. We will use these properties for two applications. The first application is a randomness-efficient error reduction procedure for randomized algorithms. The second application is proving a strong hardness-of-approximation result for the maximum clique problem.

3.1 Preliminaries

(n, d, α) graphs. For a graph G on n vertices we denote by $\lambda_0(G), \ldots, \lambda_{n-1}(G)$ the eigenvalues of the adjacency matrix of G, where $\lambda_0(G) \geq \lambda_1(G) \geq \ldots \geq \lambda_{n-1}(G)$ (recall that all the eigenvalues are real numbers since G is undirected and so the adjacency matrix is symmetric). We say that a graph G on n vertices is an (n, d)-*graph* if it is d-regular. In this case $\lambda_0(G) = d$. For a number $\alpha < 1$, we say that G is an (n, d, α)-*graph* if G is an (n, d)-graph and $\max(|\lambda_1(G)|, |\lambda_{n-1}(G)|) \leq \alpha d$.

Vectors and norms. For two vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$, we define the *dot product* of $\vec{u}$ and $\vec{v}$, denoted $\langle \vec{u}, \vec{v} \rangle$ to be $\sum_{i=1}^n \vec{u}_i \vec{v}_i$. For a vector $\vec{u} \in \mathbb{R}^n$ we define the l_1, l_2 and l_∞ norms of $\vec{u}$ as:

$$\|\vec{u}\|_1 \stackrel{def}{=} \sum_{i=1}^n |\vec{u}_i|$$

$$\|\vec{u}\|_2 \stackrel{def}{=} \sqrt{\langle \vec{u}, \vec{u} \rangle} = \left(\sum_{i=1}^n \vec{u}_i^2 \right)^{1/2}$$

$$\|\vec{u}\|_\infty \stackrel{def}{=} \max_{1 \leq i \leq n} |\vec{u}_i|$$

Probability vectors. We say that a vector $\vec{p} \in \mathbb{R}^n$ is a *probability vector* if for every $1 \leq i \leq n$, $\vec{p}_i \geq 0$ and $\sum_{i=1}^n \vec{p}_i = 1$. We denote by $\vec{u}$ the probability vector that corresponds to the uniform distribution. That is, $\vec{u} = \frac{1}{n}(1, \ldots, 1)$.

3.2 A random walk on an expander is rapidly mixing.

In this section we show that a random walk on the vertices of an expander mixes rapidly towards the stationary distribution. Let G be an (n, d, α) expander, and let A be it's adjacency matrix.

Definition 3.1. A random walk over the vertices of G is a stochastic process defining a series of vertices $(X_0, X_1, \ldots)$ in which X_0 is a vertex of G chosen by some initial distribution and X_{i+1} is chosen uniformly at random from the neighbors of X_i.

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.

Definition 3.2. The *normalized adjacency matrix* of G is defined to be $\frac{1}{d}A$ and is denoted by $\hat{A}$.

The following facts are easy to verify:

1. $\hat{A}$ is double stochastic; i.e. every column and every row sums up to 1.
2. Denote by $\hat{\lambda}_0, \ldots, \hat{\lambda}_n$ the eigenvalues of $\hat{A}$, then $\hat{\lambda}_0 = 1$ and $\max\{|\hat{\lambda}_1|, |\hat{\lambda}_n|\} = \alpha$.

$\hat{A}$ can be viewed as the transition matrix of the Markov chain defined by the random walk over the vertices of G. In other words let X be a random vertex in G with probability vector $\vec{p}$. Let Y be a uniformly chosen neighbor of X. We claim that the probability vector of Y is given by $\hat{A}\vec{p}$. To see this write the Bayesian equation:

$$\begin{aligned} Pr[Y = i] &= \sum_j Pr[Y = i | X = j] \cdot Pr[X = j] \\ &= \sum_j \hat{A}_{ij} p_j \\ &= (\hat{A}\vec{p})_i \end{aligned}$$

A similar argument shows that $\hat{A}^t$ is the transition matrix of the Markov chain defined by random walks of length t; i.e. $(\hat{A}^t)_{ij}$ is the probability a random walk starting at i reached j in exactly t steps.
Clearly $\hat{A}\vec{u} = \vec{u}$ therefore the following theorem holds:

Theorem 3.3. *The stationary distribution of the random walk on G is the uniform distribution.*

The main result of this section is the following:

Theorem 3.4. $\|\hat{A}^t\vec{p} - \vec{u}\|_1 \leq \sqrt{n} \cdot \alpha^t$ *for any distribution vector $\vec{p}$.*

In other words theorem 3.4 states that it doesn't matter what the initial distribution of the random walk is (it might be concentrated in one vertex), 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 $(\vec{v_0}, \vec{v_1}, \ldots, \vec{v_n})$.

Decompose $\vec{p}$ into the sum of the uniform distribution $\vec{u}$ and an error vector $\vec{\epsilon} = \vec{p} - \vec{u}$. The sum of the coordinates of $\vec{p}$ is 1, and the same is true for $\vec{u}$ therefore we have that the sum of the coordinates of $\vec{\epsilon}$ is 0. This means that $\langle \vec{u}, \vec{\epsilon} \rangle = 0$. In other words $\vec{\epsilon}$ is spanned by $(\vec{v_1}, \ldots, \vec{v_n})$. So we have:

$$\begin{aligned} \hat{A}\vec{p} - \vec{u} &= \hat{A}(\vec{u} + \vec{\epsilon}) - \vec{u} \\ &= \hat{A}\vec{u} + \hat{A}\vec{\epsilon} - \vec{u} \\ &= \vec{u} + \hat{A}\vec{\epsilon} - \vec{u} \\ &= \hat{A}\vec{\epsilon} \end{aligned}$$

Therefore we have:

$$\begin{aligned} \|\hat{A}\vec{p} - \vec{u}\|_2 &= \|\hat{A}\vec{\epsilon}\|_2 \\ &\leq \alpha \|\vec{\epsilon}\|_2 \\ &\leq \alpha \|\vec{p}\|_2 \end{aligned}$$

where in the first inequality we used the fact that $\vec{\epsilon}$ is spanned by an orthonormal set of vectors for which the largest eigenvalue is α. We deduce that

$$\|\hat{A}^t\vec{p} - \vec{u}\|_2 \leq \alpha^t$$

and therefore that

$$\|\hat{A}^t\vec{p} - \vec{u}\|_1 \leq \sqrt{n} \cdot \alpha^t$$

$\square$

3.3 A random walk on an expander yields a sequence of "good samples"

Consider the following problem: In an (n, d, α)-graph G there is a large set of good vertices (satisfying some condition) and we wish to find one of them. Let $B \subseteq V$ be the set of bad vertices, and assume that $\frac{|B|}{n} = \beta$. Let $x_1, \ldots, x_l$ be l vertices chosen uniformly at random from V, then $Pr[\forall i \ x_i \in B] \leq \beta^l$. This approach uses $l \log n$ random bits. We will show that by choosing one vertex randomly, and then performing a random walk in G of length l, the probability that the random walk is confined to B is exponentially small in l.
First some intuition. Recall the expander mixing lemma, proven in the previous lecture:

Theorem 3.5 (Expander Mixing Lemma). *Let G be an (n, d, α)-graph, then for every $S, T \subseteq V(G)$ it holds that*

$$\left| \frac{d|S||T|}{n} - E(S,T) \right| \leq \alpha d \sqrt{|S||T|} \leq \alpha d n$$

Previously the expander mixing lemma was interpreted as saying that the number of edges between any two sets of vertices is not far from the expected for sets of those sizes. Now divide the inequality by dn to receive the following inequality:

$$\left| \frac{|S||T|}{n^2} - \frac{E(S,T)}{dn} \right| \leq \alpha \tag{3.1}$$

Consider the following test: select uniformly at random (i, j) two vertices of G. Check whether $i \in S$ and $j \in T$. The term $\frac{|S||T|}{n^2}$ can be interpreted as the probability that this test succeeds.
Now consider a different test: select uniformly at random (i, j) an **edge** in G. Check whether $i \in S$ and $j \in T$. The term $\frac{E(S,T)}{dn}$ can be interpreted as the probability that this test succeeds. Note that the size of the probability domain of the first test is n^2 while the size of the probability domain of the second test is only nd, yet the difference between success probabilities is only a small constant α. In other words a random walk of length 1 can be viewed as discrepancy sets over sets of two vertices. Next we will show that random walks of length t are in fact discrepancy sets for sets of t vertices.
Let G be an (n, d, α)-graph, and $B \subset V$ with density $\beta = \frac{|B|}{|V|}$. Choose $X_0 \in_R V$ uniformly at random and let $X_0, \ldots, X_t$ be a random walk on G starting at X_0. Denote by (B, t) the event that the random walk is confined to B; i.e. that $\forall i \ X_i \in B$.

Theorem 3.6. $Pr[(B,t)] \leq (\beta + \alpha)^t$

Let $P = P_B$ be a projection on the space of vectors supported in B, i.e.

$$P_{ij} = \begin{cases} 1 & \text{if } i = j \in B \\ 0 & \text{otherwise} \end{cases}$$

If v is a distribution vector, then Pv is the residual distribution vector of the distribution v conditioned on being in the set B. We need two lemmas:

Lemma 3.7. $Pr[(B,t)] = \|(P\hat{A})^t P\vec{u}\|_1$

Proof. The action of P over a probability vector $\vec{v}$ is to nullify all the coordinated outside the set B, this transforms a probability vector into the residual probability vector of the same distribution, but conditioned on being in B. Thus $P\vec{u}$ is the residual probability of the uniform distribution conditioned to be in B. $\hat{A}P\vec{u}$ is the residual distribution after a random step has been taken. $P\hat{A}P\vec{u}$ is the residual probability conditioned on the random step remaining in B. Repeating this we see that $(P\hat{A})^t P\vec{u}$ is the residual probability vector of the random initial point and all t steps being in B. Since we don't care where in B we end up, we need to sum the coordinates of this vector, hence $Pr[(B,T)]$ is indeed given by $\|(P\hat{A})^t P\vec{u}\|_1$. ∎

Lemma 3.8. *For any non negative vector v:*

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

Proof. The idea of the proof is that $\hat{A}$ shrinks all components of a vector except the uniform distribution component, whereas P shrinks the uniform component without increasing anything else. Together they reduce all parts of the vector.
Decompose $P\vec{v}$ into $P\vec{v} = (P\vec{v})_{\|} + (P\vec{v})_{\perp}$ where $(P\vec{v})_{\|}$ is the projection of $P\vec{v}$ on $\vec{u}$ and $\langle (P\vec{v})_{\|}, (P\vec{v})_{\perp} \rangle = 0$. By the triangle inequality we know that

$$\|P\hat{A}P\vec{v}\|_2 \le \|P\hat{A}(P\vec{v})_{\|}\|_2 + \|P\hat{A}(P\vec{v})_{\perp}\|_2$$

First we look how P affects $(P\vec{v})_{\|}$. Since $\vec{u}$ is an eigenvector of $\hat{A}$ with eigenvalue 1 we know:

$$\|P\hat{A}(P\vec{v})_{\|}\|_2 = \|P(P\vec{v})_{\|}\|_2 = \sqrt{\beta} \cdot \|(P\vec{v})_{\|}\|_2$$

where the second equality is true since $(P\vec{v})_{\|}$ is of the form $(a, a, \ldots, a)$ where a is some scalar.
If we fix $\|v\|_2$ then $\|(P\vec{v})_{\|}\|_2$ is maximized when $\langle v, \vec{u} \rangle$ is maximized. In other words $\|(P\vec{v})_{\|}\|_2$ is maximized when $\vec{v}$ is some scalar multiple of $\vec{u}$. Therefore

$$\|P\hat{A}(P\vec{v})_{\|}\|_2 \le \beta\|\vec{v}\|_2.$$

Next we look at the effect on $(P\vec{v})_{\perp}$. P's effect is to multiply some coordinates by 0 without changing the others, so P can only shrink a vector. Since $(P\vec{v})_{\perp}$ is perpendicular to $\vec{u}$, it is spanned by the remaining eigenvectors of $\hat{A}$, all with eigenvalues at most α. This implies that

$$\|P\hat{A}(P\vec{v})_{\perp}\|_2 \le \|\hat{A}(P\vec{v})_{\perp}\|_2 \le \alpha\|(P\vec{v})_{\perp}\|_2$$

We note that $\|(P\vec{v})_{\perp}\|_2 \le \|\vec{v}\|_2$ and conclude that $\|P\hat{A}(P\vec{v})_{\perp}\|_2 \le \alpha\|\vec{v}\|_2$. Adding the two parts together we conclude that:

$$\|P\hat{A}P\vec{v}\|_2 \le (\beta + \alpha) \cdot \|\vec{v}\|_2$$

□

Now we use the lemma to prove theorem 3.6:

Proof. (theorem 3.6)

$$\begin{aligned}
\|(P\hat{A})^t P\vec{u}\|_1 &\le \sqrt{n} \cdot \|(P\hat{A})^t P\vec{u}\|_2 \\
&= \sqrt{n} \cdot \|(P\hat{A}P)^t \vec{u}\|_2 \\
&\le \sqrt{n} \cdot (\beta + \alpha)^t \|\vec{u}\|_2 \\
&= (\beta + \alpha)^t
\end{aligned}$$

□

3.3.1 Application: amplifying the success probability of random algorithms

Let L be some language in RP and assume that A is a randomized algorithm that decides whether $x \in L$ with a one sided error. Assume that A tosses m coins and has an error probability of β. Build an (n, d, α)-graph such that $V = \{0, 1\}^m$; i.e. the vertex set of the graph is the probability domain of A's coin tosses. Fix some input x and let B be all the coin tosses for which $A(x)$ is wrong. Now let A' be the following algorithm:

1. pick a vertex $v_0 \in V$ uniformly and at random.
2. perform a random walk of length t resulting with the set of vertices $(v_0, v_1, \ldots, v_t)$.
3. return $\bigcup_{i=0}^{t} A(x, v_i)$

A direct implication of theorem 3.6 yields that

$$Pr[A' \text{ fails }] = Pr[\forall i \; v_i \in B] \le (\beta + \alpha)^t$$

The error probability is reduced exponentially while the number of random bits used is only $m + t\log d = m + O(t)$.
Next we will show that the same trick can amplify the success probability of a two-sided error algorithm. In order to show this we need to restate theorem 3.6 in a stronger version.

Theorem 3.9. *Let $B_0, B_1, \ldots, B_t$ be subsets of V such that $\frac{|B_i|}{n} = \beta_i$. Define (B,t) to be the event that a random walk $(X_0, X_1, \ldots, X_t)$ has the property that $\forall i \;\; X_i \in B_i$. It holds that*

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

Note that one should be able to strengthen that by a factor of $\sim \sqrt{\beta_0 \beta_t}$, but our simple argument used in the proof of Lemma 3.8 seems to lose that.

The proof of theorem 3.9 is indeed similar to the proof of theorem 3.6. Let P_i be the projection matrix corresponding to the set B_i. Lemma 3.7 should be restated such that

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

The analouge of Lemma 3.8 is

$$\| P_{i+1} \hat{A} P_i \vec{v} \|_{\leq} (\sqrt{\beta_i \beta_{i+1}} + \alpha) \|v\|_,$$

and therefore theorem 3.9 follows.

Now let L be a language in BPP and assume that A is a randomized algorithms that decides whether $x \in L$ with a two sided error probability of $\beta \leq \frac{1}{10}$. As before assume that A tosses m coins and build an (n,d,α)-graph such that $V = \{0,1\}^m$; i.e. the vertex set of the graph is the probability domain of A's coin tosses. Fix some input x and let B be all the coin tosses for which $A(x)$ is wrong. Now let A' be the following algorithm:

1. pick a vertex $v_0 \in V$ uniformly and at random.
2. perform a random walk of length t resulting with the set of vertices $(v_0, v_1, \ldots, v_t)$.
3. return $majority\{A(x, v_i)\}$

A' fails iff a majority of the v_i's are in B. Fix a set of indices $K \subset [t]$ such that $|K| \geq \frac{t}{2}$. For each $i \in K$ let $B_i = B$. We deduce from Theorem 3.6 that

$$Pr[\forall i \in K \; v_i \in B] \leq (\beta + \alpha)^{|K|} \leq (\beta + \alpha)^{\frac{t}{2}}$$

(note that sometimes the walk makes more that one transition before testing for membership in B). By assuming that α is small enough such that $\alpha + \beta \leq \frac{1}{8}$ and applying the union bound we deduce that:

$$Pr[A' \text{ fails }] \leq 2^t \cdot (\beta + \alpha)^{\frac{t}{2}} \leq 2^t \cdot \left(\frac{1}{8}\right)^{\frac{t}{2}} = \left(\frac{1}{2}\right)^{\frac{t}{2}}$$

We achieve an exponential reduction in the error probability using only $m + O(t)$ random bits. The following table sums up the parameters of the techniques presented for error reduction:

Method	**Error Probability**	**No. of random bits**
random algorithm A	$\frac{1}{10}$	m
t independent repetitions of A	2^{-t}	$t \cdot m$
Sampling a point and it's neighbors in an $(n, t, \frac{1}{\sqrt{t}})$-graph.	$\frac{1}{t}$	m
A random walk of length t on an $(n, d, \frac{1}{40})$-graph	$2^{-\frac{t}{2}}$	$m + O(t)$

3.4 Using expanders for hardness of approximation.

In this section we show another application for random walks on expanders. We will show that we can use such walks in order to establish a hardness of approximation result for an $\mathbf{NP}$ optimization problem - the maximum clique problem.

For a graph G, we define $\omega(G)$ to be the *clique number* of G. That is, $\omega(G)$ is the size of the maximum set $S \subseteq V(G)$ such that all vertices in S are neighbors of each other. Computing exactly the clique number of a graph is $\mathbf{NP}$-hard. Using the $\mathbf{PCP}$ Theorem, it is possible to prove that it is even hard to *approximate* the clique number to within a constant factor. That is, we have the following theorem:

Theorem 3.10. *There exists a number $0 < a < 1$, such that it is NP-hard to distinguish between the following cases:*

1. $\omega(G) \leq an$

   and

2. $\omega(G) \geq 1.1an$

In this section we will show that even obtaining a *very rough* approximation for $\omega(G)$ is $\mathbf{NP}$-hard. That is, we will show that it is $\mathbf{NP}$-hard to approximate $\omega(G)$ even within a factor of n^ϵ for some $\epsilon > 0$. That is, we will prove the following theorem:

Theorem 3.11. *There exists a number $\epsilon > 0$, such that if there is a polynomial-time algorithm A such that for every graph G with n vertices*

$$n^{-\epsilon} \leq \frac{A(G)}{\omega(G)} \leq n^\epsilon$$

then $\mathbf{NP} = \mathbf{P}$.

Note: The results of this section have been superseded by a result of Håstad that it is $\mathbf{NP}$-hard to approximate $\omega(G)$ even within a factor $n^{1-\alpha}$ for *every* $\alpha > 0$. However, our approach will involve simpler analysis (and of course, expander graphs).

3.4.1 Proof of Theorem 3.11.

To illustrate the main ideas behind the proof of Theorem 3.11, we will prove a weaker version of this theorem. In the weak version we will prove under the same assumption the weaker conclusion that $\mathbf{NP} \subseteq \mathbf{RP}$ (instead of $\mathbf{NP} = \mathbf{P}$).

Lemma 3.12 (Theorem 3.11, weak version). *There exists a number $\epsilon > 0$, such that if there is a polynomial-time algorithm A such that for every graph G with n vertices*

$$n^{-\epsilon} \leq \frac{A(G)}{\omega(G)} \leq n^\epsilon$$

then $\mathbf{NP} \subseteq \mathbf{RP}$.

After we prove Lemma3.12, we will use the ideas of the proof, along with random walks on expanders to obtain Theorem 3.11, which can be looked at as a derandomized version of Lemma 3.12.

Proof of Lemma 3.12

(Since this proof will be superseded by the proof of Theorem 3.11, we allow ourselves some slackness.)

We will let ϵ be some constant, whose value will be determined later. Suppose that there exists a polynomial-time algorithm A that distinguishes between the two cases of Theorem 3.11. We will show that there exists a probabilistic polynomial-time algorithm B to distinguish between the two cases of Theorem 3.10, thus showing that $\mathbf{NP} \subseteq \mathbf{RP}$.[1]

Our algorithm B will work as follows:

[1] It may seem as if we only show that $\mathbf{NP} \subseteq \mathbf{BPP}$ but it is not hard to show (using the self-reducibility of $\mathbf{NP}$-complete problems) that if $\mathbf{NP} \subseteq \mathbf{BPP}$ then $\mathbf{NP} \subseteq \mathbf{RP}$.

Algorithm B

- Input: A graph G on n vertices.

1. Construct the graph H, where H is the following t-th power of G, for $t = \Theta(\log n)$: The vertex set $V(H)$ is the set V^t of all t-tuples in V. The edge set $E(H)$ is defined as follows: $\langle (v_1, \ldots, v_t), (u_1, \ldots, u_t) \rangle$ is an edge in $E(H)$ iff the set $\{v_1, \ldots, v_t\} \cup \{u_1, \ldots, u_t\}$ is a clique in G.

2. Let H' be the graph obtained from H by sampling $m = \Theta(a^{-t})$ $(= n^{\Theta(1)})$ vertices from H at random and taking their induced graph.

3. Return 1 if $A(H') > n^\epsilon$ and 0 otherwise. (ϵ will be determined later.)

It may seem like Algorithm B runs in time $n^{O(\log n)}$ instead of polynomial-time because the size of the graph H will be $n^{O(\log n)}$. However the construction of H in Step 1 can be done *implicitly* (that is, we don't to write out the full graph H) and so Algorithm B can be implemented in probabilistic polynomial-time.

We have the following claim:

Claim 3.12.1. $\omega(H) = \omega(G)^t$

Proof. Clearly if S is a clique in G then the set S^t is a clique. Therefore $\omega(H) \geq \omega(G)^t$.

On the other hand we claim that $\omega(H) \leq \omega(G)^t$. Indeed, if S' is a clique in H then the union of all tuples in S' is a clique in G. If $|S'| > k^t$ then it must be that this union contains more than k elements. $\square$

For every clique $S \subseteq V(H)$, the expected fraction of vertices in H' that are in S is $\frac{|S|}{|V|^t}$. With high probability we will have that for *every* clique $S \subseteq V(H)$, the fraction of vertices in S chosen to be in H' is $\Theta(\frac{|S|}{|V|^t})$. Therefore we have that with high probability we will have that $\omega(H') = \Theta(\frac{\omega(H)}{|V|^t} \cdot m) = \Theta(\frac{\omega(G)^t}{n^t} \cdot m)$. We see that:

1. If $\omega(G) \leq an$ then $\omega(H') \leq a^t m = \Theta(1)$.
2. If $\omega(G) \geq 1.1an$ then $\omega(H') \geq 1.1^t \Theta(1) \sim m^{2\epsilon}$ (for $\epsilon \sim \frac{\log 1.1}{-2 \log a}$).

We see that with high probability Algorithm B will return 1 if $\omega(G) > 1.1an$ and 0 if $\omega(G) < an$ which is what we wanted to prove.

The Actual Proof

Now that we have proved Lemma 3.12, we will now use the ideas of this proof, along with random walk on expander graphs, to prove Theorem 3.11. We will again let ϵ be some constant, whose value will be determined later, and assume that there exists a polynomial-time algorithm A that distinguishes between the two cases of Theorem 3.11. We will use A this time to show that there exists a *deterministic* polynomial-time algorithm B' to distinguish between the two cases of Theorem 3.10, thus showing that $\mathbf{NP} = \mathbf{P}$.

The only difference between Algorithm B' and Algorithm B, described in Section 3.4.1, is that in Step 2, Algorithm B' will use a *derandomized sampling* to construct the graph H'. The sampling will work in the following way. We will construct a (n, d, α)-expander $\mathcal{G}$ such that $V(\mathcal{G}) = V(G)$. We will then choose the set of t-tuples to be sampled in H' as the set of all t-tuples that represent a *length t walk* in the graph $\mathcal{G}$. We again let m denote the number of vertices in H', note that $m = nd^{t-1}$ (where d is the degree of $\mathcal{G}$). If $t = \Theta(\log n)$ and d is constant then this value is polynomial in n. What we have already seen is that a random length t walk in $\mathcal{G}$ does sometimes behave similarly to a random t-tuple. We need to show that this holds also in this context.

We start with the following claim:

Claim 3.13. *Suppose that $\omega(G) \leq an$. For every clique $S \subseteq V(G)$ in G, the probability that a length t random walk in $\mathcal{G}$ doesn't leave S (i.e., is contained in S^t) is at most $(a + \alpha)^t$.*

Proof. This is a direct application of Theorem 3.6. $\square$

As a corollary we obtain the following:

Corollary 3.14. *If $\omega(G) \leq an$ then $\omega(H') \leq (a+\alpha)^t m$*

Proof. A set $U \subseteq V(H')$ is a clique in H' if and only if all tuples in U are part of the same clique in G. Since we assume that $\omega(G) \leq an$, this means that $\frac{|U|}{m}$ can be at most $(a+\alpha)^t$. $\square$

For the other direction, we need to prove the following claim:

Claim 3.15. *Suppose that $\omega(G) \geq 1.1an$. Let $S \subseteq V(G)$ be a maximum sized clique in G (i.e., $|S| \geq 1.1an$). The probability that a length t random walk in $\mathcal{G}$ doesn't leave S (i.e., is contained in S^t) is at* least $(1.1a - 2\alpha)^t$.

Proof. This is an application of the following theorem that is analogous to Theorem 3.6:

Theorem 3.16. *In the notation of Theorem 3.6, suppose that $\beta > 6\alpha$. Then,*

$$Pr[(B,t)] \geq (\beta - 2\alpha)^t$$

Theorem 3.16 provides a lower bound on the probability that a random walk does not leave a specified set of fraction β. It shows that this probability is not much smaller than than β^t (which is what happens if we choose t independent random vertices). The proofs of Theorem 3.6, 3.16 can be found in the paper "Derandomized Graph Products" by Alon, Feige, Wigderson and Zuckerman.[2] We remark that an analogous theorem to Theorem 3.9 also holds (i.e., a lower bound on the probability to stay in *changing* sets).

$\square$

We now have the following corollary:

Corollary 3.17. *If $\omega(G) \geq 1.1an$ then $\omega(H') \geq (1.1a - \alpha)^t m$.*

Using both corollaries we see that if we choose α small enough such that $\alpha < 1 - a$ and $\alpha < a/30$ (we can take the graph $\mathcal{G}^c$ for some constant c to ensure this) then we get that

1. If $\omega(G) \leq an$ then $\omega(H') \leq \beta^t m$ for some constant $\beta < 1$.
2. If $\omega(G) \geq 1.1an$ then $\omega(H') \geq \gamma^t m$ for some constant $\gamma > \beta$.

Since $(\gamma/\beta)^t = n^{\epsilon'} = m^{2\epsilon}$ for some constants ϵ', ϵ we see that we can use A to distinguish between the two cases.

[2] Available from Avi Wigderson's homepage on `http://www.math.ias.edu/~avi/PUBLICATIONS/`.

Chapter 4

A Geometric View of Expander Graphs

Notes taken by Eran Ofek and Erez Waisbard

Summary: In the previous lectures we dealt with expander graphs in the combinatorial aspect (algorithmic and complexity applications), the algebraic aspect (spectral gap) and probabilistic aspect (rapidly mixing random walks). In this lecture we start dealing with the geometric/differential aspect of expander graphs. We introduce the construction of Margulis for expander graphs which is in fact a continuous graph with an expansion property. We show an analogy between expansion in graphs and the Cheeger constant which is defined for Riemannian surfaces. We also show the connection between the expansion constant and the spectral gap.

4.1 The Classical Isoperimetric Problem

A very natural (and ancient) question in geometry is the following:
Of all simple closed curves in the plane of a given length, which curve encloses the greatest area?

The solution to this question is obviously a circle. Although this fact was already known to the Greeks, they could not prove it. The first proof for this fact (which can be considered rigorous) is the proof of Jacob Steiner (1800's). This proof uses a method called Steiner Symmetrization. We will briefly sketch the idea of this method. Let K be a closed plane curve, let l be a line in R^2. The *symmetrization* of K with respect to l which we denote by K^* is a region in R^2 which is symmetric about l such that any line perpendicular to l intersects K iff it intersects K^*, and the intersections have the same length; furthermore the intersections of lines perpendicular to l with K^* are connected. It can shown (via calculus) that K^* has the same area as K and it's boundary length does not increase with respect to K. In fact, if l is not parallel to a line of symmetry of K then symmetrization decreases boundary length.
To gain some intuition for the correctness of this statement, we will show it for the special case in which K is a polygon which is composed of parallel trapezoids as demonstrated in figure 4.1.

In this case K^* is accepted from K if we transform each trapezoid into a symmetric trapezoid (a trapezoid of equal sides length) with the same bases and height as illustrated in figure 4.1. The area of the new trapezoid remains the same. Furthermore, the sum of the side lengths can only decrease (this fact can be easily verified). It follows that the boundary length of K^* is less or equal to that of K.

We remark that the area and boundary length of any closed curve is accepted by considering the limit of the area and boundary length of shapes of this special form (see figure 4.2). This gives intuition for the correctness of the claim in general case.

4.2 Graph Isoperimetric problems

In the spirit of the last section, one can define an analogous problem in graphs (rather than in the Euclidean space). The graph is analogous to the plane, closed curves are analogous to subsets of vertices, the "area" of a subset of vertices is

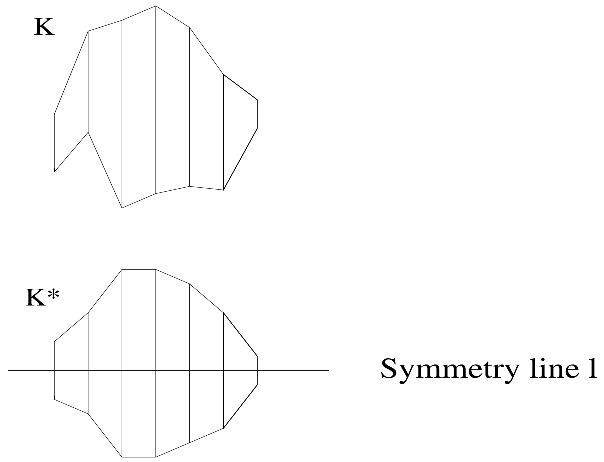


Figure 4.1: Steiner Symmetrization

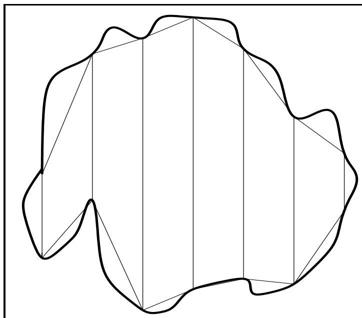


Figure 4.2:

it's cardinality and the "boundary length" of a subset is the number of edges which go out from it (or the vertices of these edges which are outside the set). More specifically we define the following isoperimetric problems:

Definition 4.1. *The edge isoperimetric problem*: given a graph G and a number k find

$$\Phi_E(G,k) = \min_{S\subset V}\{|E(S,\overline{S})| \ : \ |S| = k\}$$

Definition 4.2. *The vertex isoperimetric problem*: given a graph G and a number k find

$$\Phi_V(G,k) = \min_{S\subset V}\{|\Gamma(S)\setminus S| \ : \ |S| = k\}$$

4.2.1 The discrete cube

Let us consider first a well known graph for which the isoperimetric problem is partially solved. The discrete cube graph G_d is formally defined as:

$$V(G_d) = \{0,1\}^d$$

$$E(G_d) = \{(v_1,v_2) : \ v_1,v_2 \in \{0,1\}^d, \text{the Hamming distance between } v_1,v_2 \text{ is } 1\}$$

An equivalent definition for the d-dimensional cube graph is by recursion:

- G_1 equals K_2 (i.e. two vertices connected by an edge).
- G_{d+1} is accepted by taking two copies of G_d and connecting vertex i in the first copy to vertex i in the second copy (for all i).

It is known that if $k = 2^l$ then $\Phi_E(G_d, k)$ is achieved by a set of 2^l vertices which induces an l-dimensional cube. In this case $\Phi_E(G_d, 2^l) = 2^l(n-l)$.

For $k = \binom{d}{0} + \binom{d}{1} + ... + \binom{d}{r}$ the vertex expansion $\Phi_V(G_d, k)$ is achieved by any set S which is a ball of radius r around some vertex v_0:

$$S = \{v : \; v \in \{0,1\}^d, \text{The Hamming distance between } v, v_0 \leq r\}$$

4.3 The construction of Margulis, Gabber-Galil

In this section we describe one of the first explicit construction of an expander graph. In contrast to the expanders we encountered so far in this course, the construction they give is over a continuous set.

We denote by I the interval $(0,1)$. The set of vertices is all the points in the continuous cube $I \times I$. Two linear transformations define the edges:

$$T(x,y) \to (x+y, y) \text{ mod } 1$$

$$S(x,y) \to (x, x+y) \text{ mod } 1$$

The neighbors of a point (x,y) are the points: $T(x,y), S(x,y), T^{-1}(x,y), S^{-1}(x,y)$. Thus the graph is 4-regular. The expansion property of this graph is described by the following theorem:

Theorem 4.3. ***(Margulis,Gabber-Galil)***
There exists some $\epsilon > 0$ such that for any measurable set $A \subset I \times I$ with $\mu(A) \leq \frac{1}{2}$ (μ denotes the Lebegue measure) the following holds:

$$\mu(\Gamma(A) \cup A) \geq (1+\epsilon)\mu(A),$$

where $\Gamma(A) = S(A) \cup T(A) \cup S^{-1}(A) \cup T^{-1}(A)$ is the set of all points which are neighbors of points in A.

It s worthwile to mention here the following conjecture:

Conjecture 4.4. *(Linial) For any measurable subset A, such that $\mu(A) \leq \frac{1}{2}$,*

$$\mu(A \cup S(A) \cup T(A)) \geq \frac{4}{3}\mu(A),$$

with equality achieved by the hexagon whose vertices are $(0,a)$, $(a,0)$, $(a,-a)$, $(0,-a)$, $(-a,0)$, $(-a,a)$ for some small $a > 0$.

4.4 The Cheeger constant, Cheeger inequality

In this section we introduce the Cheeger constant. Loosely speaking, this constant represents the "expansion" of a curve.

Definition 4.5. Let $M^{(n)}$ be an n-dimensional Riemann surface. The **Cheeger constant** of M is defined to be:

$$h(M) = \min_{\substack{C \text{ is an} \\ (n-1)\text{-dimensional} \\ \text{surface which divides} \\ M \text{ into } M_1, ...M_t}} \frac{\mu_{n-1}(C)}{\min_i \; \mu_n(M_i)}$$

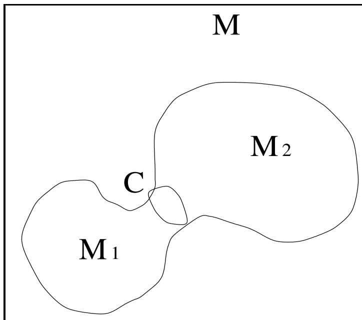


Figure 4.3: C divides M into M_1, M_2

where $\mu_{n-1}(C)$ is the area of C and $\mu_n(M_i)$ is the volume of M_i.
An intuitive demonstration of the definition is illustrated in figure 4.3
The analogy between the Cheeger constant and expansion is as follows:
$M \sim G$,
C is analogous to a cut in G,
M_1, M_2 are analogous to $S, \overline{S}$,
$\mu_{n-1}(C) \sim |e(S, \overline{S})|$,
$\min_i \mu_n(M_i) \sim \min\{|S|, |\overline{S}|\}$.

Given an n-dimensional Riemann surface $M(n)$, and a function $f : M(n) \to \mathbb{R}$, then its Laplacian is $\Delta(f) = \text{div}(\text{grad}(f))$. The Laplacian is a linear operator, and its eigenvalues are all the numbers λ, for which there is a function $f : M(n) \to \mathbb{R}$ satisfying $\Delta f = \lambda f$. All its eigenvalues are non-negative, and its lowest eigenvalue is zero, corresponding to the constant function.

Theorem 4.6. *Let M be a Riemann surface as described before, denote by λ the lowest positive eigenvalue of the Laplacian of M, then $\lambda \geq \frac{h^2}{4}$.*

We will now explain the discrete analogs for the gradient and divergence operators in graphs. Let $G = (V, E)$ be an undirected graph. Select an orientation for the edges of G. Let M be the $V \times E$ adjacency matrix of G where the entry $M_{v,e}$ equals 1 (-1) if the edge e enters (leaves) v and 0 otherwise.

The gradient: Let $f : V \to R$ be a function on the vertices of G. f can be thought of as a row vector with V entries. The gradient operator is $f \mapsto fM$. The gradient of f is a vector with E entries which tells us how does f change along the edges of the graphs. I.e., if e is the edge from u to v, then $(fM)_e = f_v - f_u$.

The divergence: Let $g : E \to R$ be a function on the edges of G. g is a column vector with E entries. The divergence operator is $g \mapsto Mg$. The divergence of g is a V dimensional vector with $Mg_v = \sum_{\text{e enters v}} g(e) - \sum_{\text{e leaves v}} g(e)$.

The Laplacian: If we go through with the analogy between real functions in R^n and functions on the vertices of a graph, then the discrete analog of the Laplacian will be: $f \mapsto MM^t f$ (for $f : V \to R$). The matrix $L = MM^t$ is called the Laplacian of G. A simple calculation shows that L is the following symmetric matrix:

$$L_{i,j} = \begin{cases} -1, & (i,j) \in E \\ \text{deg}(i), & i = j \end{cases}$$

In the case that G is a d-regular graph (with A_G it's corresponding adjacency matrix):

- $L = d \cdot I - A_G$.
- The spectrum of L is in $[0, +2d]$ (since the spectrum of A_G is in $[-d, +d]$).

- $\lambda_1(A_G) = d$ corresponds to $\lambda_n(L) = 0$ and in general $\lambda_i(A_G) = d - \lambda_{n-i+1}(L)$.
- The spectral gap of G ($\lambda_1(A_G) - \lambda_2(A_G)$) equals to the lowest positive eigenvalue of L.

Notice the similarity between Theorem 4.6 and the upper bound on h given by Theorem 4.9.

4.5 Expansion and the spectral gap

In this section we show the that a graph has high expansion (high Cheeger constant) iff it has a large spectral gap.

4.5.1 The Rayleigh quotient

For a real symmetric matrix A one can obtain the eigenvalues of A using a special quotient known as the Rayleigh quotient.

Theorem 4.7. *Let A be a real symmetric matrix and let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be its corresponding eigenvalues. Then:*

$$\lambda_1 = \max_{||x||=1} \frac{xAx^t}{||x||}, \quad \lambda_2 = \max_{||x||=1, x\perp x_1} \frac{xAx^t}{||x||} \quad \ldots \quad \lambda_n = \max_{||x||=1,\, x\perp x_1, \ldots, x\perp x_{n-1}} \frac{xAx^t}{||x||}$$

where x_i is an eigenvector corresponding to λ_i.

4.5.2 The main theorem

Before stating the theorem, we formally define the expansion constant:

Definition 4.8. The expansion constant of a graph $G = (V, E)$, is

$$h(G) = min_{S\subseteq V, |S| \leq \frac{|V|}{2}} \frac{|E(S, \overline{S})|}{|S|}$$

Theorem 4.9. *Let $G = (V, E)$ be a finite, connected, k-regular graph without loops. Let λ be the second eigenvalue of G. Then*

$$\frac{k-\lambda}{2} \leq h(G) \leq \sqrt{2k(k-\lambda)}$$

Proof. In this lecture we only prove the lower bound on h, showing that a large gap implies high expansion. In order to prove that $\lambda \geq\ k - 2h(G)$, we will give a vector $f \perp \vec{1}$ for which $\frac{fAf^t}{||f||} \geq k - 2h(G)$. For $S \subset V$ we define f to be the following weighted cut function:

$$f = |\overline{S}|1_S - |S|1_{\overline{S}}$$

using the Rayleigh quotient we get:

$$\lambda \geq \frac{fAf^t}{||f||^2}.$$

We will evaluate the rhs. Starting with the denominator we get that:

$$||f||^2 = |\overline{S}|^2|S| + |S|^2|\overline{S}| = |S||\overline{S}|(|S| + |\overline{S}|) = n|S||\overline{S}|$$

moving on to the numerator we get:

$$fAf^t = 2|E(S)||\overline{S}|^2 + 2|E(\overline{S})||S|^2 - 2|S||\overline{S}||E(S, \overline{S})| \tag{4.1}$$

Since G is a k-regular graph

$$k|S| = 2|E(S)| + |E(S, \overline{S})|$$

and

$$k|\overline{S}| = 2|E(\overline{S})| + |E(S, \overline{S})|$$

substituting $2|E(S)|$ and $2|E(\overline{S})|$ in (4.1) yields:

$$fAf^t = nk|S||\overline{S}| - n^2|E(S,\overline{S})|$$

We now plug it in and get

$$\lambda \geq \frac{fAf^t}{||f||^2} = \frac{nk|S||\overline{S}| - n^2|E(S,\overline{S})|}{n|S||\overline{S}|} = k - \frac{n|E(S,\overline{S})|}{|S||\overline{S}|}$$

Fix S to be a set for which

$$h(G) = \frac{|E(S,\overline{S})|}{|S|}$$

it follows that:

$$\lambda \geq k - \frac{nh(G)}{|\overline{S}|} \geq k - 2h(G)$$

(since $|\overline{S}| \geq \frac{n}{2}$). $\square$

Chapter 5

Expander graphs have a large spectral gap

Notes taken by Yael Vinner

Summary: In the previous lecture we started proving upper and lower bounds on $h(G)$, defined as $h(G) = \min_{|S| \leq \frac{n}{2}}(\frac{e(S,S^C)}{|S|})$. We proved a lower bound of $\frac{d-\lambda}{2} \leq h(G)$. In this lecture we will prove an upper bound of $h(G) \leq \sqrt{2d(d-\lambda)}$. We will also discuss the d-regular infinite tree, which is the optimal expander, and show that its spectrum is $[-2\sqrt{d-1}, 2\sqrt{d-1}]$.

5.1 Comments about the previous lecture

Given a graph G, we choose an arbitrary orientation for its edges, and define the matrix $M_{n\times m}$ where $n = |V(G)|$ and $m = |E(G)|$. The entries M_{ue} for $u = 1 \dots n$ and $e = 1 \dots m$, are defined as follows:

$$M_{ue} = \begin{cases} 1 & e = (u \to v) \\ -1 & e = (v \to u) \\ 0 & otherwise \end{cases}$$

The Laplacian of G is defined as the matrix $L = M \cdot M^T$. Then for any $f : V \to \mathbb{R}$, we have

$$fLf^T = fMM^Tf^T = < fM, fM > = \| fM \|^2$$

where $\| \cdot \|$ is the l_2 norm. Furthermore, $(fM)_{e=(x\to y)} = f(x) - f(y)$, and therefore we can write

$$\| fM \|^2 = \sum_{(x,y)\in E} (f(x) - f(y))^2$$

Another comment, is about the variational description of the eigenvalues of a matrix, which is

$$\lambda_k = \max_{x\perp x_1,\dots,x_{k-1}} \left(\frac{xAx^T}{\| x \|^2}\right)$$

This can be re-written in the following way:

$$\lambda_k = \min_{F, dim(F)=n-k+1} \max_{x\in F} \left(\frac{xAx^T}{\| x \|^2}\right)$$

5.2 An upper bound on $h(G)$

Theorem 5.1. *For any connected graph G, define $h(G) = \min_{|S|\leq\frac{n}{2}}(\frac{e(S,S^C)}{|S|})$. Then*

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

Proof.

Definition 5.2. Given a function $f : V \to \mathbb{R}$, define

$$B_f = \sum_{(x,y)\in E} |f^2(x) - f^2(y)|$$

Definition 5.3. Let $\beta_0 < \beta_1 < \cdots < \beta_r$ be the different values achieved by f over V. Then we define L_i for $i = 1, \dots, r$ as follows:

$$L_i = \{x \in V | f(x) \geq \beta_i\}$$

This definition leads to $L_0 \supseteq L_1 \supseteq \cdots \supseteq L_r$.

We will use the three following claims.

Claim 5.4. $B_f = \sum_{i=1}^{r} e(L_i, L_i^C)(\beta_i^2 - \beta_{i-1}^2)$.

Proof. For $(x,y) \in E$, if $f(x) = \beta_p > \beta_q = f(y)$, (x,y)'s contribution to B_f is

$$(\beta_p^2 - \beta_q^2) = (\beta_p^2 - \beta_{p-1}^2) + (\beta_{p-1}^2 - \beta_{p-2}^2) + \cdots + (\beta_{q+1}^2 - \beta_q^2)$$

When we sum over E, $(\beta_i^2 - \beta_{i-1}^2)$ appears once for every edge (x,y) such that $f(x) = \beta_p > \beta_q = f(y)$ and $p \geq i > q$. In other words, $(\beta_i^2 - \beta_{i-1}^2)$ appears exactly once for every edge (x,y) such that $x \in L_i$ and $y \notin L_i$, and in total $e(L_i, L_i^C)$ times, which gives us the desired result. □

Claim 5.5. $B_f \leq \sqrt{2d} \cdot \| fM \| \cdot \| f \|$.

Proof. Using the Cauchy-Schwartz inequality, we have

$$B_f = \sum_E |f^2(x) - f^2(y)| = \sum_E |f(x) + f(y)| \cdot |f(x) - f(y)| \leq$$

$$\leq \sqrt{\sum_E (f(x) + f(y))^2} \sqrt{\sum_E (f(x) - f(y))^2}$$

We know that

$$\sqrt{\sum_E (f(x) - f(y))^2} = \| fM \|$$

and evaluating the second term in the product gives us

$$\sqrt{\sum_E (f(x) + f(y))^2} \leq \sqrt{2\sum_E (f^2(x) + f^2(y))} = \sqrt{2d \sum_V f^2(x)} = \sqrt{2d} \cdot \| f \|$$

from which we can conclude the inequality in the claim. □

Claim 5.6. *If $f \geq 0$ and $|Supp(f)| \leq \frac{n}{2}$, where $Supp(f)$ is the subset of V where $f(x) \neq 0$, then*

$$B_f \geq h(G) \| f \|^2$$

Proof. Since f equals zero on more than half of the coordinates, and is positive on the rest, $\beta_0 = 0$, and for every $i \geq 1$ $|L_i| \leq \frac{n}{2}$. Therefore $\frac{e(L_i, L_i^C)}{|L_i|} \geq h(G)$. Plugging this inequality into claim 5.4 yields

$$B_f = \sum_{i=1}^{r} e(L_i, L_i^C)(\beta_i^2 - \beta_{i-1}^2) \geq h(G) \sum_{i=1}^{r} |L_i|(\beta_i^2 - \beta_{i-1}^2) =$$

$$= h(G) \sum_{i=1}^{r} \beta_i^2(|L_i| - |L_{i+1}|) = h(G) \sum_{i=1}^{r} \beta_i^2 \cdot |\{x | f(x) = \beta_i\}| = h(G) \| f \|^2$$

□

For each eigenvalue λ_i of A, $d - \lambda_i$ is an eigenvalue of the Laplacian, $L = dI - A$. Let g be an eigenvector of L (and A) with eigenvalue $d - \lambda$, where λ is the second largest eigenvalue of A. Define $f = g^+$, i.e. equal to g where g is positive, and zero elsewhere. Without loss of generality, $|Supp(f)| \leq \frac{n}{2}$, since otherwise we would look at $-g$ which is also an eigenvector with the same eigenvalue. Define $V^+ = Supp(f)$. Then for every $x \in V^+$ we can write

$$(Lf)(x) = df(x) - \sum_{y \in V} a_{xy} f(y) = dg(x) - \sum_{y \in V^+} a_{xy} g(y) \leq$$

$$\leq dg(x) - \sum_{y \in V} a_{xy} g(y) = (Lg)(x) = (d - \lambda) g(x)$$

Since $f(x) = 0$ for any $x \notin V^+$, we can write

$$\| fM \|^2 = fLf^T = \sum_{x \in V} f(x) \cdot (Lf)(x) \leq (d - \lambda) \sum_{x \in V^+} g^2(x) = \tag{5.1}$$

$$= (d - \lambda) \sum_{x \in V} f^2(x) = (d - \lambda) \| f \|^2$$

From claim 5.5 we have

$$B_f \leq \sqrt{2d} \cdot \| fM \| \cdot \| f \|$$

and from claim 5.6 we have

$$B_f \geq h(G) \| f \|^2$$

If we combine these results we get

$$h(G) \| f \|^2 \leq \sqrt{2d} \cdot \| fM \| \cdot \| f \|$$

If we square this equation and combine with (5.1) we get

$$h^2(G) \| f \|^2 \leq 2d \cdot \| fM \|^2 \leq 2d(d - \lambda) \| f \|^2$$

and therefore

$$h^2(G) \leq 2d(d - \lambda)$$

□

5.3 The Infinite d-Regular Tree

Let us look at the infinite adjacency matrix A_T of the infinite d-regular tree T. The infinite vectors we work with are those in $l_2(V(T))$:

$$l_2(V(T)) = \{x : V(T) \to \mathbb{R} | \sum_{v \in V(T)} x_v^2 < \infty\}$$

Define $\sigma(A_T)$ to be the set of all λ's such that $(A_T - \lambda I)$ is non invertible. This is the set of all λ's such that $(A_T - \lambda I)$ is not one to one, i.e. there is a vector $u \in l_2$ such that $(A_T - \lambda I)u = 0$, or $(A_T - \lambda I)$ is not onto l_2.

Theorem 5.7.

$$\sigma(A_T) = [-2\sqrt{d-1}, 2\sqrt{d-1}]$$

Given $v \in V(T)$ (the root of the tree),

$$\lambda \in \sigma(A) \iff \delta_v \notin Range(\lambda I - A)$$

where δ_v is defined as follows:

$$\delta_v(u) = \begin{cases} 1 & u = v \\ 0 & u \neq v \end{cases}$$

The direction $\Leftarrow$ is easy. The other direction requires a proof which will not be given here.

We wish to find out for which values of λ, $\delta_v \in Range(\lambda I - A_T)$.
We are trying to find a function $f \in l_2$ such that

$$\delta_v = (\lambda I - A)f \tag{5.2}$$

Without loss of generality, f is spherical, meaning if u, w are the same distance from v then $f(u) = f(w)$. This is true since if g is a solution to (5.2) then so is f which is the spherical symmetrization of g (for all vertices with a given distance from v, f will be the average of g on these vertices). Therefore, $f(u)$ depends only on the distance $d_T(u, v)$. We need to define a sequence of numbers $x_0, x_1, \ldots$ such that all vertices u with $d_T(u, v) = r$ will have $f(u) = x_r$. Substituting the sequence $\{x_i\}_{i=0}^{\infty}$ for f in (5.2), we get the following recursion:

$$\lambda x_0 = dx_1 + 1$$

$$\lambda x_i = x_{i-1} + (d-1)x_{i+1}$$

We will try to find two numbers ρ_1, ρ_2, such that for every i, $x_i = A\rho_1^i + B\rho_2^i$. To find these numbers, we need to solve the equation

$$\lambda\rho = 1 + (d-1)\rho^2$$

The solutions to this equation are

$$\rho_{1,2} = \frac{1}{2(d-1)}(\lambda \pm \sqrt{\lambda^2 - 4(d-1)})$$

If $\lambda^2 < 4(d-1)$ ($\lambda \in \sigma(A)$) then $\rho_{1,2}$ are complex and $|\rho_1| = |\rho_2| = \frac{1}{\sqrt{d-1}}$. In this case, f is not in l_2: $|x_i| = \Theta((d-1)^{-\frac{i}{2}})$, and for every i x_i appears $\Theta((d-1)^i)$ times in $\| f \|$, each time contributing $\Theta(((d-1)^{-\frac{i}{2}})^2)$ to the sum. This means that each level i in the tree contributes $\Theta(1)$ to the sum, which results in the sum being infinite.

On the other hand, if $\lambda^2 > 4(d-1)$, then one of the roots ρ_1, ρ_2, say ρ_1, is less than $\frac{1}{\sqrt{d-1}}$ in absolute value, in which case we can choose $B = 0$, so the contribution of the i'th level of the tree to f's norm will be exponentially small in i, and therefore $f \in l_2$. Also, there is a solution to $\lambda A = dA\rho + 1$, since $\lambda \neq d\rho$:

$$\lambda \stackrel{?}{=} \frac{d}{2(d-1)}(\lambda - \sqrt{\lambda^2 - 4(d-1)})$$

$$\sqrt{1 - \frac{4(d-1)}{\lambda^2}} \stackrel{?}{=} 1 - \frac{2(d-1)}{d}$$

This equality cannot hold for $d > 2$, since the r.h.s. is negative.

Chapter 6

Upper Bound on Spectral Gap

Notes taken by Yishai Beeri

Summary: This lecture presents a lower bound for λ_1, the second-largest eigenvalue of a d-regular graph: $\lambda_1 \geq 2\sqrt{(d-1)}(1-\frac{c}{\Delta^2})$, which is related to the graph's diameter Δ. As the diameter grows, so does the lower bound for λ_1. This can also be viewed as an upper bound on the graph's spectral gap.

6.1 Reminder to previous lecture

In the previous lecture we discussed the infinite d-regular tree and the spectrum of its (infinite) adjacency matrix A. We defined the spectrum $\sigma(A)$ as follows:

$$\sigma(A) := \{\lambda | (\lambda I - A)\ is\ not\ invertible\ \}$$

And we showed that for the d-regular infinite tree:

$$\sigma(A) = [\ -2\sqrt{d-1}\ ,\ 2\sqrt{d-1}\ ]$$

In this lecture we will use the finite (k-tall) d-regular tree to prove a lower bound on λ_1 for general graphs.

6.2 Lower bound on λ_1

We show a lower bound on λ_1 for general d-regular graphs: [1]

Theorem 6.1. [2] *There exists a constant c s.t. for any d-regular graph G of size n and diameter Δ:*

$$\lambda_1 \geq 2\sqrt{(d-1)}(1-\frac{c}{\Delta^2})$$

Where λ_1 is the second largest eigenvalue of G.

Notes:

- It is easy to show that in the above graph $Diam(G) > \Omega(\log_{d-1} n)$, and hence it follows that for any fixed $d > 2$:

$$\lambda_1 \geq 2\sqrt{(d-1)}(1-O(\frac{1}{\log^2 n})).$$

[1] The original statement of the lower bound is due to N. Alon and R. Boppana, and appears in A. Nilli *On the second eigenvalue of a graph* Discrete Math., 91(2):207-210, 1991

[2] This stronger statement and proof is taken from J. Friedman *Some geometric aspects of graphs and their eigenfunctions* Duke Math. J. 69(3):487-525, 1993.

- In the case that $n = d + 1$ (G is the n-clique) the eigenvalues of G are $\{ n - 1, -1, \ldots, -1 \}$, since if A is the adjacency matrix for G, then $A + I = J$ and J's eigenvalues are, naturally, $\{ n, 0, \ldots, 0 \}$. While at first this may seem to be a counter-example to our theorem, note that the theorem deals with the case where d is fixed and n, Δ are going to infinity.
- Since $\lambda_1 = \max_{x \perp \vec{1}} \{ \frac{xAx^T}{||x||^2} \}$ we expect a proof for the above bound might use a specific "test function" (=eigenvector), e.g. find a vector f s.t. $\sum_{x \in V(G)} f(x) = 0$ and $\frac{fAf^T}{||f||^2} \geq 2\sqrt{d-1}(1 - \frac{c}{\Delta^2})$.

Proof Sketch: Taking two nodes s, t with $d(s,t) = \Delta$, we build a spherical function f that will be positive for the nodes within a distance of $k = \lfloor \frac{\Delta}{2} \rfloor - 1$ from s, and negative on the nodes that are within a distance k from t. The values of f will be derived from the spherical function g with maximal eigenvalue μ for the d-regular tree of height k, treating s and t as roots of (separate) k-tall trees. We will show that for the positive part of f we have $Af \geq \mu f$, and likewise for the negative part of f we have $Af \leq -\mu f$, so that $fAf^t \geq \mu f f^t = \mu ||f||^2$, giving $\frac{fAf^T}{||f||^2} \geq \mu$. Finally, the positive and negative parts of f will be normalized to ensure $\sum f(x) = 0$, letting μ apply as a lower bound for λ_1. $\square$

Proof. Set $k = \lfloor \frac{\Delta}{2} \rfloor - 1$. Select two nodes $s, t \in G$ with distance $d(s,t) = \Delta$. For all $0 \leq i \leq k$ define:

$$S_i := \{ v | d(s,v) = i \}$$

and

$$T_i := \{ v | d(t,v) = i \}$$

and in addition define

$$Q := V(G) \setminus (\bigcup_{0 \leq i \leq k} S_i \cup T_i)$$

> **Note:**This is simply a breadth-first-search dividing the graph into layers according to the distance from s and t. Q represents the "middle ground" with at least 1 layer of nodes. There are, of course, no edges between any S_i and any T_i.

Denote T to be the finite tree of height k, and mark A_T to be its adjacency matrix.

Claim 6.2. *There is a single spherical function $g : [0, \ldots, k+1] \to \mathbb{R}$ on T that satisfies:*

$$g(k+1) = 0$$

(we extend g's domain in include $k+1$ even though it is not part of A_T), and

$$A_T g = \mu g$$

(g is an eigenfunction of A_T with eigenvalue μ), where μ is the maximal eigenvalue of A_T.

This claim can be proved using the same technique as shown in the previous lecture dealing with the infinite d-tree. We will not prove this claim here, but will show a specific function g that displays these properties.

Definition 6.3. Define $f : V(G) \to \mathbb{R}$ as follows:

$$f(v) = \begin{cases} c_1 g(i) & \exists i, v \in S_i \\ -c_2 g(i) & \exists i, v \in T_i \\ 0 & otherwise \end{cases}$$

where $c_1, c_2 \geq 0$.

Since f is spherical, it will be convenient to define $F : [0, ..., k] \to \mathbb{R}$ with $f(v) = F(i) \iff v \in S_i$. We will next show that f gives us the desired properties:

Lemma 6.4. *If g above is non-negative and monotonically non-increasing, then:*

$$v \in S_i => (Af)(v) \geq \mu f(v)$$

and

$$v \in T_i => (Af)(v) \leq \mu f(v)$$

Proof. Let $v \in S_i$ for some i. G is d-regular, so v has $1 \leq p \leq d$ neighbors in level $i-1$, another $0 \leq q \leq d-p-1$ neighbors in the same level i, and the remaining $d-p-q$ neighbors in level $i+1$, giving:

$$\begin{aligned}(Af)(v) &= pF(i-1) + qF(i) + (d-p-q)F(i+1) \\ &= pc_1 g(i-1) + qc_1 g(i) + (d-p-q)c_1 g(i+1)\end{aligned}$$

but for g and the matrix A_T (of the k-tall d-tree), we know that for a node u of level i:

$$\mu g(i) = (A_T g)(i) = g(i-1) + (d-1)g(i+1)$$

since each node has exactly one neighbor in the previous level, and $d-1$ neighbors in the next level. As g is non-negative and non-increasing we get:

$$\begin{aligned}(Af)(v) &= c_1[pg(i-1) + qg(i) + (d-p-q)g(i+1)] \\ &\geq c_1[g(i-1) + (d-1)g(i+1)] \\ &= c_1(A_T g)(i) = c_1 \mu g(i) \\ &= \mu f(v)\end{aligned}$$

The same argument is used for $v \in T_i$, with the approriate redefinition of the function F, and using c_2.

□

Corollary 6.5. $\frac{fAf^T}{||f||^2} \geq \mu$ *and* $f \perp \vec{1}$.

Proof. The previous lemma gives $|(Af)(v)| \geq |\mu f(v)|$ for $v \in V(G) \setminus Q$. For $v \in Q$, note that $f(v) = 0$, in which case $|(Af)(v)| \geq |\mu f(v)|$ is trivial. From this follows:

$$\begin{aligned}fAf^T &= <f, Af> = \sum_{v \in V(G)} f(v)(Af)(v) \\ &= \sum_{\exists i, v \in S_i} f(v)(Af)(v) + \sum_{\exists i, v \in T_i} f(v)(Af)(v) + \sum_{v \in Q} f(v)(Af)(v) \\ &\geq \sum_{\exists i, v \in S_i} f(v)\mu f(v) + \sum_{\exists i, v \in T_i} f(v)\mu f(v) \\ &= \mu f f^T\end{aligned}$$

since for $v \in T_i$ $f(v) \geq 0$ and for $v \in Q$ $f(v) = 0$. From this we get:

$$fAf^T \geq \mu f f^T => \frac{fAf^T}{||f||^2} \geq \mu$$

Next, we show that $f \perp \vec{1}$, (namely: $\sum f(v) = 0$). This is easily achieved by selecting c_1 and c_2 such that

$$\sum_{v \in S_i} f(v) = -\sum_{v \in T_i} f(v)$$

□

Definition 6.6. Finally, we will find the spherical function g that is non negative, non-increasing, that satisfies $g(k+1)=0$, and that has the maximal eigenvalue μ. We'll also show that this μ yields the desired bound:

$$g(i) := (d-1)^{-i/2}\sin(\theta(k+1-i))$$

where $\theta = \frac{\pi}{2k+2}$.

It is easy to check that g is non-negative, non increasing, and that $g(k+1)=0$. We now show that $A_T g = \mu g$ with $\mu = 2\sqrt{d-1}\cos(\theta)$. For all j:

$$\begin{aligned}
(A_T g)(j) &= 1g(j-1) + (d-1)g(j+1) \\
&= (d-1)^{-(j-1)/2}\sin(\theta(k+1-(j-1))) + (d-1)(d-1)^{-(j+1)/2}\sin(\theta(k+1-(j+1))) \\
&= (d-1)^{(-j+1)/2}\sin(\theta(k+2-j)) + (d-1)^{(-j+1)/2}\sin(\theta(k-j) \\
&= \sqrt{d-1}(d-1)^{-j/2}(\sin(\theta(k+2-j)) + \sin(\theta(k-j)))
\end{aligned}$$

And since $\sin(\alpha)+\sin(\beta) = 2\sin(\frac{\alpha+\beta}{2})\cos(\frac{\alpha-\beta}{2})$, we get for all j:

$$\begin{aligned}
(A_T g)(j) &= 2\sqrt{d-1}\cos(\theta)(d-1)^{-j/2}\sin(\theta(k+1-j)) \\
&= 2\sqrt{d-1}\cos(\theta)g(j)
\end{aligned}$$

To arrive at our bound, we need to show that $\cos(\theta) \geq (1-\frac{c}{\Delta^2})$. Recall:

$$\theta = \frac{\pi}{2k+2} \simeq \frac{\pi}{\Delta}$$

since $k = \lfloor\frac{\Delta}{2}\rfloor - 1$, and use the Taylor expansion for the fucntion $\cos(x)$:

$$\cos(x) = 1 - \frac{x^2}{2} + o(x^2)$$

to get:

$$\cos(\theta) \geq 1 - \frac{\pi^2}{2\Delta^2} = (1 - \frac{c}{\Delta^2})$$

with $c \approx \frac{\pi^2}{2}$. □

Chapter 7

The Margulis construction

Notes taken by Statter Dudu

Summary: We define an explicit family of 8-regular graphs on the torus $\mathbb{Z}_n \times \mathbb{Z}_n$, and prove that this is a family of expander graphs.

Construction 7.1. Let

$$T_1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix},\ T_2 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix},\ e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

and define the following 8-regular graph G=(V,E) on the vertex set $V = \mathbb{Z}_n \times \mathbb{Z}_n$. Each vertex $v = (x, y)$ is adjacent to the four vertices

$$T_1 v,\ T_2 v,\ T_1 v + e_1,\ T_2 v + e_2,$$

where all the calculations are performed mod n. The other four neighbours of v are obtained by the four inverse transformations. (Note that this is an 8-regular undirected graph, that may have multiple edges and self loops.)

Theorem 7.2. $\lambda_2(G) \leq 5\sqrt{2} < 8$

We have already seen that if G is a d-regular graph of size n and eigenvalues (of its adjacency matrix) $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_n$, then $\lambda_1 = d$, then its Cheeger constant satisfies $h(G) \geq (d - \lambda_2)/2$. Therefore, a lower bound on the gap between λ_2 and d that is independent of n implies that this is a family of expanding graphs. Margulis proved a similar result in 1973 but couldn't give an explicit lower bound on the gap. Galil and Gabber (1981) used continuous harmonic analysis to derive a lowerbound on the gap. Boppana simplified the proof, and Jimbo, Marouka (1985) improved it furthermore using discrete Fourier transform.

As we have seen before, by the Rayleigh quotient Theorem,

$$\begin{aligned}
\max\{|\lambda_2|, |\lambda_n|\} &= \max\{| < Af, f > | :< f, v_1 >= 0, ||f|| = 1\} \\
&= 2\max\{| \sum_{(i,j)\in E} \overline{f(i)} f(j)| :< f, v_1 >= 0, ||f|| = 1\},
\end{aligned}$$

where $v_1 = \overline{1}/\sqrt{n}$ is the eigenvector corresponding to λ_1, and the maxima are taken over all $f : \mathbb{Z}_n^2 \to \mathbb{C}$. It follows, that it is sufficient to prove that for every complex function $f : \mathbb{Z}_n^2 \to \mathbb{C}$ satisfying $\sum_x f(x) = 0$,

$$| \sum_{(x,y)\in E} \overline{f(x)} f(y)| \leq \frac{5\sqrt{2}}{2} \sum |f(x)|^2.$$

By the definition of our graph, this is equivalent to:

Theorem 7.3. *For any $f : \mathbb{Z}_n^2 \to \mathbb{C}$ satisfying $\sum_\vartheta f(\vartheta) = 0$, the following inequality holds:*

$$| \sum_{\vartheta \in \mathbb{Z}_n^2} \overline{f(\vartheta)} \big(f(T_1\vartheta) + f(T_1\vartheta + e_1) + f(T_2\vartheta) + f(T_2\vartheta + e_2) \big) | \leq \frac{5\sqrt{2}}{2} \sum |f(x)|^2 \tag{7.1}$$

Discrete Fourier transform It is hard to use the condition $\sum_x f(x) = 0$ so we will move to a new space using the discrete Fourier transform.

A character of an abelian group G is a homomorphism $\chi : G \to \mathbb{C}$, mapping addition in G to multiplication in $\mathbb{C}$. It can be seens, that the characters of $\mathbb{Z}_n^k$ are: $\chi_b : \mathbb{Z}_n^k \to \mathbb{C}$ for $b \in \mathbb{Z}_n^k$, where $\chi_b(a) = \omega^{<a,b>}$. Here, ω is the nth root of the unity ($\omega = e^{2\pi i/n}$), and for any $a, b \in \mathbb{Z}_n^k$ their inner product is $< a, b >= \sum_{j=1}^k \overline{a_i} b_i$.

Since the characters χ_b are an orthonormal basis, we can express any $f : \mathbb{Z}_n^k \to \mathbb{C}$ as $f = \sum_b \widehat{f}(b)\chi_b$, where

$$\widehat{f}(a) =< f, \chi_b >= \frac{1}{n^{k/2}} \sum_b \overline{f(b)} \cdot \chi_b(a) = \frac{1}{n^{k/2}} \sum_b \overline{f(b)} \cdot \omega^{<a,b>}.$$

$\widehat{f} : \mathbb{Z}_n^k \to \mathbb{C}$ is called the Discrete Fourier Transform - (DFT) of $f : \mathbb{Z}_n^k \to \mathbb{C}$.

To prove Theorem 7.3, we express the condition of equation (7.1) using the Fourier coefficients of f. The condition $\sum_\vartheta f(\vartheta) = 0$ is equivalent to $\widehat{f}(0,0) = 0$. Using identities 2,4 in appendix 1, one can easily prove that Theorem 7.3 reduces to:

Theorem 7.4. *Any function $F : \mathbb{Z}_n^2 \setminus (0,0) \to \mathbb{C}$ must satisfy*

$$|\sum_{\vartheta=(\vartheta_1,\vartheta_2)\in\mathbb{Z}_n^2\setminus(0,0)} \overline{F(\vartheta)} \cdot [F(T_2^{-1}\vartheta)(1+\omega^{-\vartheta_1}) + F(T_1^{-1}\vartheta)(1+\omega^{-\vartheta_2})]| \le \frac{5\sqrt{2}}{2} \sum_{\vartheta\in\mathbb{Z}_n^2\setminus(0,0)} |F(\vartheta)|^2.$$

Let $G = |F|$. Then G is a real function $G : \mathbb{Z}_n^2 - (0,0) \to \mathbb{R}$. Moving the absolute value inside the summation, and using the equality

$$|1+\omega^{-a}| = 2|\cos(\frac{\pi a}{n})|,$$

it follows that it is sufficient to prove:

Theorem 7.5. *Any function $G : \mathbb{Z}_n^2 - (0,0) \to \mathbb{R}$, must satisfy*

$$\sum_\vartheta G(\vartheta) \cdot [G(T_2^{-1}\vartheta)|\cos\frac{\pi\vartheta_1}{n}| + G(T_1^{-1}\vartheta)|\cos\frac{\pi\vartheta_2}{n}|] \le \frac{5\sqrt{2}}{4} \sum G^2(\vartheta).$$

We would like to convert the LHS into a sum of squares that will match the RHS. To do this we use the following inequality, valid for any real a, b, γ

$$2ab \le \gamma a^2 + \frac{1}{\gamma} b^2.$$

Let $\gamma : (\mathbb{Z}_n^2)^2 \to \mathbb{R}$ be a function satisfying for all $x, y \in \mathbb{Z}_n^2$

$$\gamma(x,y) \cdot \gamma(y,x) = 1, \tag{7.2}$$

and in addition, that for all $\vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{Z}_n^2 \setminus (0,0)$

$$|\cos\frac{\pi\vartheta_1}{n}| \cdot [\gamma(\vartheta, T_2\vartheta) + \gamma(\vartheta, T_2^{-1}\vartheta)] + |\cos\frac{\pi\vartheta_2}{n}| \cdot [\gamma(\vartheta, T_1\vartheta) + \gamma(\vartheta, T_1^{-1}\vartheta)] \le \frac{5\sqrt{2}}{2}. \tag{7.3}$$

Since for any $p, q \in \mathbb{Z}_n^2 \setminus (0,0)$

$$2G(p)G(q) \le \gamma(p,q)G^2(p) + \gamma(q,p)G^2(q),$$

it follows that

$$\begin{aligned} 2 \cdot LHS \le \sum_\vartheta & |\cos\frac{\pi\vartheta_1}{n}| \cdot [\gamma(\vartheta, T_2^{-1}\vartheta)G^2(\vartheta) + \gamma(T_2^{-1}\vartheta, \vartheta)G^2(T_2^{-1}\vartheta)] \\ + & |\cos\frac{\pi\vartheta_2}{n}| \cdot [\gamma(\vartheta, T_1^{-1}\vartheta)G^2(\vartheta) + \gamma(T_1^{-1}\vartheta, \vartheta)G^2(T_1^{-1}\vartheta)]. \end{aligned}$$

Since T_1 doesn't change ϑ_2, and T_2 doesn't change ϑ_1, it follows that:

$$\begin{aligned} 2 \cdot LHS &\leq \sum_{\vartheta} G^2(\vartheta) \cdot |\cos \frac{\pi\vartheta_1}{n}| \cdot [\gamma(\vartheta, T_2\vartheta) + \gamma(\vartheta, T_2^{-1}\vartheta)] + G^2(\vartheta) \cdot |\cos \frac{\pi\vartheta_2}{n}| \cdot [\gamma(\vartheta, T_1\vartheta) + \gamma(\vartheta, T_1^{-1}\vartheta)] \\ &\leq \frac{5\sqrt{2}}{2} \sum_{\vartheta} G^2(\vartheta). \end{aligned}$$

Therefore, if we just find a function γ satisfying the requirements (7.2) and (7.3), Theorem 7.2 would follow, and we are done. To define γ, we first define a partial order on $\mathbb{Z}_n^2$. So, let

$$a(x) = \begin{cases} x & \text{if } \frac{n}{2} \geq x \geq 0 \,, \\ n - x & \text{if } n \geq x \geq \frac{n}{2}. \end{cases}$$

(Notice that $a(x)$ is invariant under mod n, i.e. $a(x \mod n) = a(x)$.) Then we say that $(\vartheta_1, \vartheta_2) > (\vartheta_1', \vartheta_2')$ if $a(\vartheta_1) \geq a(\vartheta_1')$ and $a(\vartheta_2) \geq a(\vartheta_2')$ and at least one of the inqualities is strong. That is the distance to the X axis and/or to the Y axis is bigger.

The definition of γ is:

$$\gamma((\vartheta_1, \vartheta_2), (\vartheta_1', \vartheta_2')) = \begin{cases} \alpha & \text{if }, (\vartheta_1, \vartheta_2) > (\vartheta_1', \vartheta_2') \\ \frac{1}{\alpha} & \text{if }, (\vartheta_1, \vartheta_2) < (\vartheta_1', \vartheta_2') \\ 1 & \text{otherwise.} \end{cases}$$

This definition of γ obviously satisfies (7.2). We will show that for $\alpha = \frac{5}{4}$, also (7.3) is satisfied for any $\vartheta \in \mathbb{Z}_n^2 \setminus (0,0)$. We define the diamond to be the set of all $\vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{Z}_n^2 \setminus (0,0)$, satisfying

$$a(\vartheta_1) + a(\vartheta_2) \leq \frac{n}{2}.$$

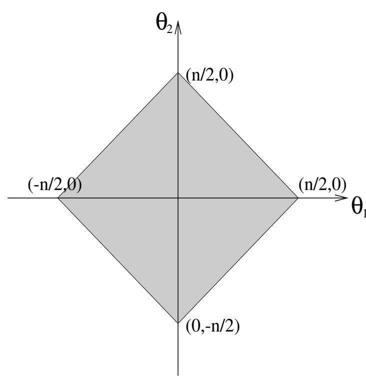


Figure 7.1: The diamond

We analyze the case where ϑ is outside or inside the diamond separately.

- **Outside the diamond**

  (7.3) will follow from the fact that the cosines are small. So, it is sufficient to prove that

$$2\alpha \cdot (|\cos \frac{\pi\vartheta_1}{n}| + |\cos \frac{\pi\vartheta_2}{n}|) \leq \frac{5\sqrt{2}}{2},$$

  or that

$$|\cos \frac{\pi\vartheta_1}{n}| + |\cos \frac{\pi\vartheta_2}{n}| \leq \sqrt{2}. \tag{7.4}$$

We can assume w.l.o.g. that we are in the first quadrant. Since $\cos\frac{\pi\vartheta_2}{n}$ is decreasing, for any given value of ϑ_1, the LHS of (7.4) is maximized, when $\vartheta_2 = n/2 - \vartheta_1$. Therefore, using the convexity of $\cos$ we get that

$$\cos\frac{\pi\vartheta_1}{n} + \cos\frac{\pi\vartheta_2}{n} \leq \cos\frac{\pi\vartheta_1}{n} + \cos\frac{\pi(\frac{n}{2} - \vartheta_1)}{n} \leq 2\cos\frac{\pi}{4} = \sqrt{2}.$$

.

- **Inside the diamond**

  In this case, we just bound the cosines by 1, and would like to prove that

  $$\gamma(\vartheta, T_1\vartheta) + \gamma(\vartheta, T_1^{-1}\vartheta) + \gamma(\vartheta, T_2\vartheta) + \gamma(\vartheta, T_2^{-1}\vartheta) \leq \frac{5\sqrt{2}}{2}. \tag{7.5}$$

  It is not difficult to verify that for every ϑ satisfying $a(\vartheta_1) + a(\vartheta_2) < \frac{n}{2}$, one of the following two cases must hold:

  1. Three of the four points $T_1\vartheta$, $T_2\vartheta$, $T_1^{-1}\vartheta$ and $T_2^{-1}\vartheta$ are $> \vartheta$ and one is $< \vartheta$.
  2. Two of the four points $T_1\vartheta$, $T_2\vartheta$, $T_1^{-1}\vartheta$ and $T_2^{-1}\vartheta$ are $> \vartheta$ and two are incomparable with ϑ.

  In the first case, the LHS of (7.5) is $\leq \frac{3}{\alpha} + \alpha$, while in the second case it is $\leq \frac{2}{\alpha} + 2$. Substituting $\alpha = 5/4$, this gives a upper bound of 3.65 or the LHS of (7.3). This is not as good as $5\sqrt{2}/2 = 3.53\ldots$, but it does prove that the graphs are a family of good expanders.

Appendix

Using the same definitions as before the following properties of the discrete Fourier transform can by easily concluded from the definitions:

1. $\sum_a f(a) = 0 \Leftrightarrow \widehat{f}(0) = 0$
2. The dot product of two functions remains invariant under the transform i.e for any f, g

   $$\sum_a \overline{f(a)}g(a) = \sum_a \overline{\widehat{f}(a)}\widehat{g}(a).$$

3. Parseval's identity: A special case of 2, where $f = g$

   $$\sum_a |f(a)|^2 = \sum_a |\widehat{f}(a)|^2.$$

4. The shift property : If A is a non-singular $k \times k$ matrix with entries over $\mathbb{Z}_n$, $b \in \mathbb{Z}_n^k$ and $g(x) = f(Ax + b)$ then

   $$\widehat{g}(y) = \omega^{-<A^{-1}b,y>}\widehat{f}((A^{-1})^t y).$$

5. The inverse formula :

   $$f(a) = \frac{1}{n^{k/2}}\sum_b \widehat{f}(a)\omega^{-<a,b>}.$$

Chapter 8

The Zig Zag Product

Notes taken by Eyal Bigman

Summary: In this lecture we shall introduce a new kind of graph product called the *Zig Zag Product*. We shall analyze the expansion properties of the zig zag product of expanding graphs and use them to create an explicit recursive construction of a family of good expanders. Furthermore we will connect the entropy of a random walk on the graph and its expansion.

8.1 Introduction

We begin with some standard definitions: Let $G =< V, E >$ be a d-regular graph ($|V| = n$). The adjacency matrix $A(G) = (a_{ij})_{i,j=1}^{n}$ of the graph G is a symmetric $n \times n$ non-negative integer matrix such that $a_{ij} = k$ iff there are k edges between vertices i and j. It follows from regularity that the sum of every row is d. The matrix $\hat{G} = \frac{1}{d}A$ is the *normalized adjacency matrix*, it is the transition matrix of a random walk on G. Thus if $p \in \Re^n$ is a probability distribution on the vertices at time t then $\hat{G}p$ is the distribution at time $t + 1$.

G is a (n, d, α)-expander if α is an upper bound on the second eigenvalue of the normalized adjacency matrix $\lambda_2 \leq \alpha$ ($\lambda_1 = 1$). It follows from the spectral gap theorem that $(1 - \alpha)d/2 \leq h(G)$ where $h(G)$ is the expansion of G. Thus G is a better expander when α is close to 0.

The *square* $G^2 = (V, E')$ is a graph on the same vertices and $(u, w) \in E'$ iff there is a path of length 2 in G from u to w. If A is the adjacency matrix of G then A^2 is the adjacency matrix of G^2. It is easy to see that G^2 is a (n, d^2, α^2)-expander.

The zig zag product will be an unsymmetric binary function, that given an m-regular graph of size n and a d-regular graph of size m, it yields a d^2-regular graph of size mn. After we define the zig zag product we will prove:

Theorem 8.1 (The Zig Zag Theorem). *Let G be a (n, m, α)-expander and H be a (m, d, β)-expander then $G ⓩ H$ is a $(nm, d^2, f(\alpha, \beta))$-expander where $f(\alpha, \beta) = \alpha + \beta + \beta^2$.*

We will present a proof that uses only elementary linear algebra. With some more algebra this bound is improved in [RVW02] to $f(\alpha + \beta) = \alpha + \beta$, and if $\alpha, \beta < 1$, it can be shown $f(\alpha + \beta) < 1$.

Before we proceed with the definition let us show how it can be used for an explicit construction of a family of expanders with constant degree. For d constant let H be a $(d^4, d, \frac{1}{4})$-expander, there is a probabilistic proof that such an expander exists and since d is constant we can find such a graph by an exhaustive search in constant time, there are also efficient constructions of such graphs - see [RVW02]. Define recursively:

$$G_1 = H^2$$
$$G_{n+1} = (G_n)^2 ⓩ H$$

Proposition 8.2. *G_n is a $(d^{4n}, d^2, \frac{1}{2})$-expander for all n*

Proof. By induction. The case $n = 1$ follows from the definition. If we assume the proposition for n then G_n^2 is a $(d^{4n}, d^4, \frac{1}{4})$-expander and from theorem 8.1 it follows that G_{n+1} is a $(d^{4(n+1)}, d^2, \frac{1}{2})$ graph ∎

8.2 The Construction

Let G be a (n, m, α)-expander and H be a (m, d, β)-expander. For every vertex $v \in V(G)$ let $e_v^1, \ldots e_v^m$ be the edges connected to the vertex. Also, we regard the vertices of H as the integers $1, \ldots, m$, denoted by $[m]$. To obtain $G ⓩ H$, we replace every vertex v with a cloud of m vertices $(v, 1), \ldots (v, m)$ one for every edge connected. The vertices within a cloud are connected by "miniedges" so that every cloud forms a mini copy of H. The edges of G are augmented on both sides by these miniedges to form the edges of $G ⓩ H$. More formally:

Definition 8.3. $G ⓩ H =< V(G) \times [m], E' >$, where $((v, i), (u, j)) \in E'$ iff there are some $k, l \in [m]$ such that $(i, k), (l, j) \in V(E)$ and $e_v^k = e_u^l$.

It is essential of course for the sake of well definedness that the degree of G equals the size of H.

Proof of the Zig Zag theorem. It is easy to see that $GⓩH$ is a graph of size mn and degree d^2. The expansion constant of $G ⓩ H$ is a bound on the second eigenvalue of the transition matrix of the random walk on the graph. Each step of the random walk on an edge of $G ⓩ H$ can be regarded as a random step on a miniedge within a cloud, a deterministic step on an edge connecting two clouds and another random step within a cloud. Thus, the transition from (v, i) to (u, j) consists of a random step from (v, i) to (v, k) for some $k \in [m]$, a deterministic step from (v, k) to (u, l) and another random step from (u, l) to (u, j).

The transition matrix of the random walk on $GⓩH$ will therefore be the product of the transition matrices of these three steps, i.e. $\widehat{G ⓩ H} = \tilde{H}\tilde{G}\tilde{H}$, where $\tilde{H} = \hat{H} \otimes I_n$ is the transition matrix of a random step in each cloud, and

$$\tilde{G}_{(v,k),(u,l)} = \begin{cases} 1 & \text{if } e_v^k = e_u^l \\ 0 & \text{otherwise} \end{cases}$$

is a matrix of transpositions. It is easy to see that $\| \tilde{H} \|, \| \tilde{G} \| \leq 1$.

$G ⓩ H$ is a regular graph thus the constant vector 1_{mn} is an eigenvector, it follows from Rayleigh's theorem that $\lambda_2 = max_{f \perp 1_{mn}} \frac{|f \widehat{GⓩH} f|}{\|f\|^2}$. Therefore it suffices to show $\frac{|f \widehat{GⓩH} f|}{\|f\|^2} \leq \alpha + \beta + \beta^2$ for every $f \perp 1_{mn}$.

We can write any vector f as $f = f^{\|} + \tilde{H} f^{\perp}$, where $f^{\|}$ is a vector that is constant within each cloud, and $\tilde{H} f^{\perp}$ sums up to zero within each cloud. Since 1_m is an eigenvector of H with eigenvalue one, we get that $\tilde{H} f^{\|} = f^{\|}$.

Therefore, for every $f \perp 1_{mn}$:

$$\begin{aligned} |f \widehat{G ⓩ H} f| &= |f \tilde{H}\tilde{G}\tilde{H} f| \\ &= |f^{\|} \tilde{H}\tilde{G}\tilde{H} f^{\|}| + 2|f^{\|} \tilde{H}\tilde{G}\tilde{H} f^{\perp}| + |f^{\perp} \tilde{H}\tilde{G}\tilde{H} f^{\perp}| \\ &= |f^{\|} \tilde{G} f^{\|}| + 2|f^{\|} \tilde{G}\tilde{H} f^{\perp}| + |f^{\perp} \tilde{H}\tilde{G}\tilde{H} f^{\perp}| \end{aligned}$$

Now, since $f^{\|} \perp 1_{mn}$, we know that $\| \tilde{G} f^{\|} \| \leq \alpha \| f^{\|} \|$. Also, since $f^{\perp}$ sums up to zero in each cloud, we have $\| \tilde{H} f^{\perp} \| \leq \beta \| f^{\perp} \|$. Therefore:

$$|f^{\|} \tilde{G} f^{\|}| \leq \alpha \| f^{\|} \|^2$$

$$|f^{\|} \tilde{G}\tilde{H} f^{\perp}| \leq \| \tilde{G} \| \cdot \| f^{\|} \| \cdot \| \tilde{H} f^{\perp} \| \leq \beta \| f^{\|} \| \cdot \| f^{\perp} \|$$

$$|f^{\perp} \tilde{H}\tilde{G}\tilde{H} f^{\perp}| \leq \| \tilde{G} \| \cdot \| \tilde{H} f^{\perp} \|^2 \leq \beta^2 \| f^{\perp} \|^2$$

The inequalities follow from Cauchy Schwartz and the definition of the operator norm.

$$\begin{aligned} |f \widehat{G ⓩ H} f >| &\leq \alpha \| f^{\|} \|^2 + 2\beta \| f^{\|} \| \cdot \| f^{\perp} \| + \beta^2 \| f^{\perp} \|^2 \\ &\leq \alpha \| f \|^2 + 2\beta \| f^{\|} \| \cdot \| f^{\perp} \| + \beta^2 \| f \|^2 \\ &= \alpha \| f \|^2 + \beta \cdot (\| f^{\|} \|^2 + \| f^{\perp} \|^2 - (\| f^{\|} \| - \| f^{\perp} \|)^2) + \beta^2 \| f \|^2 \\ &= (\alpha + \beta + \beta^2) \cdot \| f \|^2 - \beta \cdot (\| f^{\|} \| - \| f^{\perp} \|)^2 \\ &\leq (\alpha + \beta + \beta^2) \cdot \| f \|^2 \end{aligned}$$

□

8.3 Entropy Analysis

There are several different definitions of entropy for distribution p, the classical definitions $H(p) = -\sum_{i=1}^{n} p_i \log(p_i)$, the H_2 entropy $H_2(p) = -\log(\| p \|_2)$ and the min-entropy $H_\infty(p) = -\log(\| p \|_\infty)$. For any transition matrix A with expansion constant α and distribution $p = \frac{1}{n}1_n + f$

$$\| Ap \|^2 = \| \frac{1}{n}1_n + Af \|^2 \leq \| \frac{1}{n}1_n \|^2 + \| Af \|^2 \leq ((1-\lambda^2) + \alpha^2\lambda^2) \| p \|^2$$

where $\lambda = \frac{\|f\|}{\|p\|} \leq 1$. Hence

$$H_2(Ap) = H_2(p) + \log((1-\lambda^2) + \alpha^2\lambda^2) = H_2(p) + \log(1-(1-\alpha^2)\lambda^2) = H_2(p) - \Delta_E$$

$0 < 1-\alpha^2 \leq 1$ implies $0 \leq 1-(1-\alpha^2)\lambda^2 < 1$ hence $\Delta_E = -\log((1-\lambda^2)+\alpha^2\lambda^2) > 0$ as long as $\lambda > 0$. This shows that the entropy increases by at least Δ_E as long as there is a positive nonuniform component. It follows that for better expanders (α smaller) the H_2 entropy grows faster. It can be shown that the H_2 and the H_∞ are correlated, therefore the same increase is true for the min-entropy.

We will show next that the classical entropy also grows when $\lambda > 0$ but we shall make very different considerations. It is currently unknown how fast the entropy grows and how the growth rate relates to the expansion constant. We note that for all the definitions of entropy, the increase of entropy for nonuniform distributions is essentially the second law of thermodynamics.

For random variables X and Y with some joint distribution we have the standard entropy equation $H(X,Y) = H(X) + H(Y|X)$ thus the entropy of the joint distribution of X and Y is the sum of the entropy of X and the conditioned entropy of $Y|X$. In the case of the zig zag graph the set of vertices is $V \times [m]$, the random variables we will analyze will be the projections X and Y to the first and second coordinates. We will see that a random step on the zig zag graph will increase both the entropy of X and the conditioned entropy $H(Y|X)$ and thus increase the entropy of the joint distribution.

As we mentioned before a random step consists of a random step within a cloud (zig) a deterministic step between clouds and another random step within a cloud (zag). Since the zig and the zag steps are within a cloud, they only affect the second coordinate. They are random steps therefore they increase the conditioned entropy of $Y|X$ as long as that is less than maximal. For a distribution which is uniform on every cloud the zig and zag steps have no effect.

We can think of X and Y as basins of entropy, taking either a zig or a zag step is like pouring entropy from an external source into the basin. How much entropy can be poured in? Well that depends on the capacity of the random variable (maximum entropy) and the amount of entropy present in the basin, i.e. how far is the distribution on every cloud from uniform.

What about the deterministic step between the clouds? Well this step is deterministic and therefore it does not change the total amount of entropy in the system, but that does not mean it cannot change the division of the entropy between X and Y.

Let us think of the extreme case $p = e_v \otimes \frac{1}{m}1_m$ where the entropy of X is zero and the entropy of Y is maximal, i.e. the distribution is concentrated uniformly on the vertices of the cloud v (obviously p remains unchanged by the zig step). In this case $H(Y|X) = \log(m)$ and $H(X,Y) = H(X) + H(Y|X) = 0 + \log(m)$. After the deterministic step there is equal probability to be in any one of v's neighboring clouds, but within these clouds the distribution is concentrated on the vertex that is actually connected to a vertex in v. The entropy in these clouds is 0. The entropy of clouds not connected to v remains zero, and the entropy in v is zero. Thus we see that the entropy of X goes from zero to $\log(m)$ and the entropy of Y plummets to zero. We still have the same entropy in the system $H(X,Y) = H(X)+H(Y|X) = \log(m)+0$, but the division between X and Y changes, all the entropy of Y is passed to X. In the general case the same thing happens, entropy is exchanged between X and Y. Since deterministic steps are reversible, not necessarily is the entropy transfered from Y to X, or from a variable with high entropy to a variable with low entropy. Nevertheless, it can be shown that if the entropy of $Y|X$ is maximal and the joint distribution is not uniform then the deterministic step reduces the entropy of $Y|X$. Thus we see that the deterministic step does not change the entropy in the systems but pours entropy from one vessel to another, and in the case that Y is full and X is not, some entropy is poured from Y to X.

We can see now the effect of a random step, first entropy is poured into Y, if it is not full then some entropy is added to the joint distribution. Next entropy is exchanged between X and Y, necessarily making room for more entropy in Y if it is full and X is not. Finally yet more entropy is poured into Y and again if it is not full already then the entropy of the joint distribution is increased.

Thus we see that the total amount of entropy in the system necessarily increases unless the two basins are full already. Why do we have to pour in entropy from an external source twice? Because in cases like p above, whatever entropy was poured while Y was full will be wasted and the entropy of the whole system will remain the same. Only the second step will make room in Y for more entropy, thus if we avoid the zag step the entropy in the system will not increase. On the other hand, since the deterministic step is reversible it could also be the case that Y will be filled from X and thus the entropy poured in the zag step will be wasted. It can be shown (or follows immediatly form reversability) that in such a case Y is not full from the start. Thus if we don't pour entropy on Y at the beginig it may not be possible later. Therefore both the zig and the zag steps are essential in order to assure that the entropy will actually increase.

Chapter 9

Metric Embedding

Notes taken by Tamir Hazan

Summary: We can embed any metric space into $\mathbb{R}^n$, but with some distortion of the distances. We show that the graph metric of expander graphs is the hardest metrics to embed, in the sense that of all finite metric spaces on a given number of points, expanders require the largest distortion.

9.1 Basic Definitions

(X, d) is a metric space if

- $d : X \times X \to \mathbb{R}^+$.
- $d(x, y) = 0$ if and only if $x = y$.
- $d(x, y) = d(y, x)$.
- $d(x, y) \leq d(x, z) + d(z, y)$.

In this lecture we will examine how to approximate a finite metric (X, d) by the metric space l_2. l_2 is the metric space $(\mathbb{R}^n, ||\cdot||)$ such that for every $y, z \in \mathbb{R}^n$, $||y - z||^2 = \sum_{i=1}^n (y_i - z_i)^2$.

Given the metric spaces (X, d) and $(\mathbb{R}^n, l_2)$ and a transformation $f : X \to \mathbb{R}^n$ we define:

- $expansion(f) = max_{x_1, x_2 \in X} \frac{\|f(x_1) - f(x_2)\|}{d(x_1, x_2)}$
- $contraction(f) = max_{x_1, x_2 \in X} \frac{d(x_1, x_2)}{\|f(x_1) - f(x_2)\|}$
- $distortion(f) = expansion(f) \cdot contraction(f)$

It is clear that there are metric spaces that need to be embedded with distortion. E.g the metric $(\{1, 2, 3, 4\}, d)$ with $d(1, 4) = d(2, 4) = d(3, 4) = 1$, and $d(i, j) = 2$ otherwise, since $\{1, 2, 4\}$, $\{1, 3, 4\}$, $\{2, 3, 4\}$ must be on the same line in $\mathbb{R}^n$.

9.2 Finding the Minimal Distortion

In this section we will present some of the properties of embeddings in l^2. Given a metric space (X, d) we denote its minimal distortion by $C_2(X, d)$.

Theorem 9.1. *Bourgain (1985) Any n-point metric space (X, d) can be embedded into $O(\log n)$ dimensional Euclidean space with $O(\log n)$ distortion.*

Theorem 9.2. *Linial, London, and Rabinovich (1995) There is a polynomial time algorithm which computes $C_2(X, d)$.*

Proof. The proof is based on semi definite programming. Let (X, d) be a metric space with $|X| = n$. Let $f : X \to l_2$. Since we can assume that without loss of generality that $contraction(f) = 1$, then $distortion(f) \leq \gamma$ if and only if for every $1 \leq i < j \leq n$:

$$(*) \quad d(x_i, x_j)^2 \leq \|f(x_i) - f(x_j)\|^2 \leq \gamma^2 d(x_i, x_j)^2$$

We say that a matrix Z is positive semi definite (denoted PSD) if Z is symmetric and $x^T Z x \geq 0$ for every $x \in \mathbb{R}^n$.

Let $u_i = f(x_i)$ be the rows of the matrix U. Let $Z = U \cdot U^T$. It is clear that $Z \in PSD$ since it is symmetric, and for every $x \in \mathbb{R}^n$, $x^T Z x = x^T U \cdot U^T x = (U^T x)^T \cdot (U^T x) = \|U^T x\|^2 \geq 0$.

But also the converse is true, if $Z \in PSD$ then $Z = U \cdot U^T$ for some matrix U. To see that, note that Z is symmetric, and therefore diagonizable. Thus $Z = A \Lambda A^T$ for some matrix A and a diagonal matrix Λ. Let $\sqrt{\Lambda}$ be the diagonal matrix which $(\sqrt{\Lambda})_{ii} = \sqrt{(\Lambda)_{ii}}$. Then $Z = A\sqrt{\Lambda}\sqrt{\Lambda}^T A^T = (A\sqrt{\Lambda}) \cdot (A\sqrt{\Lambda})^T$.

Therefore instead of finding $u_i = f(x_i)$ which satisfies $(*)$ we can find a $Z \in PSD$ that satisfies

$$(**) \quad d(x_i, x_j)^2 \leq z_{ii} + z_{jj} - 2z_{ij} \leq \gamma^2 d(x_i, x_j)^2,$$

since $\|u_i - u_j\|^2 = z_{ii} + z_{jj} - 2z_{ij}$ for $Z = U \cdot U^T$.

Thus we conclude that $C_2(X, d) \leq \gamma$ if and only if there is a positive semi definite matrix Z such that for all i, j (**) holds. This is a linear programming problem (more precisely a convex programming problem) which can be solved in polynomial time by the ellipsoid algorithm. □

The algorithm above constructs a primal problem and solves it by the ellipsoid algorithm. Looking at the dual problem gives us an interesting characterization of $C_2(X, d)$. When we transform a primal problem to its dual we take a non negative combination of its constraints. But how do we look at the constraint $Z \in PSD$?

Lemma 9.3. *$Z \in PSD$ if and only if for all $Q \in PSD$, $\sum_{i,j} q_{ij} z_{ij} \geq 0$.*

Proof. Let Q be a matrix such that $Q_{ij} = q_i \cdot q_j$. Previously we showed that such a matrix is PSD. Therefore $q^T Z q = \sum_{i,j} q_{ij} z_{ij} \geq 0$ implying that $Z \in PSD$.

It can be easily seen that any $Q \in PSD$ of rank 1 must have the form $Q_{ij} = q_i \cdot q_j$ for some values $q_1, ..., q_n$. Thus, if Z is PSD then for every symmetric matrix $Q \in PSD$ of rank 1, $\sum_{i,j} q_{ij} z_{ij} \geq 0$. The lemma follows from the fact that any PSD matrix is a non negative linear combination of rank 1 PSD matrices. To see this, note that any $P \in PSD$ can be written as $A \Lambda A^T$, where Λ is a diagonal matrix of the same rank as P. Therefore $P = \sum_{i=1}^{rank(P)} A \Lambda_{ii} A^T$.

□

Our primal problem is:

- $\sum_{ij} q_{ij} z_{ij} \geq 0$ for all $Q \in PSD$.
- $d(x_i, x_j)^2 \leq z_{ii} + z_{jj} - 2z_{ij}$ for all i, j.
- $z_{ii} + z_{jj} - 2z_{ij} \leq \gamma^2 d(x_i, x_j)^2$ for all i, j.

We proceed by deriving an explicit formula for $C_2(X, d)$ from the dual problem:

Theorem 9.4.

$$C_2(X, d) = \max_{P \in PSD,\, P \cdot \vec{1} = \vec{0}} \sqrt{\frac{\sum_{p_{ij} > 0} p_{ij} d(x_i, x_j)^2}{-\sum_{p_{ij} < 0} p_{ij} d(x_i, x_j)^2}}$$

Proof. In the following proof we shall assume that $\gamma < C_2(X, d)$. Therefore when we inspect a non negative combination of the constraints of the primal problem (the dual problem) we must get a contradiction.

Let us look at the constraints of the first type (for all $Q \in PSD$, $< q, z >= \sum_{ij} q_{ij} z_{ij} \geq 0$). Since the collection of PSD matrices is convex, a non negative combination $\sum_k a_k < q, z >$ is equal to $< p, z >$ for some matrix $p \in PSD$. Thus the combination of the first type constraints gives us $\sum_{ij} p_{ij} z_{ij} \geq 0$ for some $P \in PSD$.

A contradiction can be reached if a combination of all the constraints result in $0 > 0$. Unfortunately so far we have $\sum_{ij} p_{ij} z_{ij} \geq 0$ for some $P \in PSD$. So to eliminate the z_{ij} for $i \neq j$, we take the following linear combination of the rest of the constraints:

- If $p_{kl} = 0$ then we multiply the constraints involving z_{kl} by zero.
- If $p_{kl} > 0$ then we multiply by $p_{kl}/2$ the constraint $d(x_k, x_l)^2 \leq z_{kk} + z_{ll} - 2z_{kl}$
- If $p_{kl} < 0$ then we multiply by $p_{kl}/2$ the constraint $z_{kk} + z_{ll} - 2z_{kl} \leq \gamma^2 d(x_k, x_l)^2$

To eliminate z_{ii}, we have to choose P such that

$$p_{ii} + \sum_{j \neq i} p_{ij} = 0.$$

Therefore the combination of all the constraints gives

$$(*) \quad \sum_{p_{ij}>0} p_{ij} d(x_i, x_j)^2 + \gamma^2 \sum_{p_{ij}<0} p_{ij} d(x_i, x_j)^2 \leq 0.$$

We get our contradiction if $(*)$ is violated. Thus we conclude the theorem.

□

9.3 Embeddings in l_2

9.3.1 Embedding the cube

Given a cube $\{0,1\}^r$ we can easily find an embedding in l_2 with distortion $\sqrt{r}$. Given the embedding $id : \{0,1\}^r \to \mathbb{R}^r$ such that $id(x) = x$, we get that $contraction(id) = \sqrt{r}$ and $expansion(id) = 1$. Using our main theorem we can show that this is the best embedding of the cube. Let define the $2^r \times 2^r$ matrix P:

- $P(i,j) = -1$ if $d(i,j) = 1$.
- $P(i,j) = r - 1$ if $i = j$.
- $P(i,j) = 1$ if $d(i,j) = r$.
- $P(i,j) = 0$ otherwise.

It is easy to check that $P\vec{1} = 0$, and that $P \in PSD$ (the later holds, since P has the same eigenvectors as the cube). Since $\sum_{p_{ij}>0} p_{ij} d(x_i, x_j)^2 = 2^r \cdot r^2$, and $-\sum_{p_{ij}<0} p_{ij} d(x_i, x_j)^2 = 2^r \cdot r$, we get that $C_2(X, d) \geq \sqrt{r}$.

9.3.2 Embedding expander graphs

Let $G = (V, E)$ be a k-regular graph, $|V| = n$, with $\lambda_2 \geq k - \epsilon$. As before it is simple to see that an expander can be embedded with distortion $O(\log n)$ in l_2. Indeed, take the expander and put it as a simplex in $\mathbb{R}^n$. Since every two nodes of the simplex have distance 1 we get that $expansion = 1$, and $contraction = diam(G)$. Since G is an expander then $diam(G) = O(\log n)$. As before this result is tight.

Lemma 9.5. *Let $H = (V, E')$ be a graph with the same vertex set as G. 2 vertices are adjacent in H if their distance in G is at least $\log_k n - 2$. Then H has a matching B of $n/2$ edges.*

Proof. G is k-regular graph thus every vertex has at most k^r vertices at distance $\leq r$ from it. If $r = \log_k n - 2$ then there are at most $n/2$ nodes at this distance. Since all the vertices of H have degree $\geq n/2$ then it has a matching of the desired size. This follows from Dirac's sufficient condition for a Hamiltonian circuit. (Modern Graph Theory B. Bollobas p. 106-107). □

Theorem 9.6. *Let $G = (V, E)$ be a k-regular graph, $|V| = n$, with $\lambda_2 \geq k - \epsilon$. Then $C_2(G) = \Omega(\log n)$.*

Proof. Let B be the adjacency matrix of the matching we found in H. For simplicity we assume that B is a complete matching in H. Let $P = kI - A_G + \frac{\epsilon}{2}(B - I)$ in H. It is easy to see that $P\vec{1} = 0$. $P \in PSD$ since for every $x \perp \vec{1}$,

$$x^T P x = x^T(kI - A_G)x + x^T(\epsilon/2)(B - I)x$$

$x^T(kI - A_G)x \geq \epsilon ||x||^2$ since G is expander.

$$x^T(B - I)x = \sum_{(i,j)\ edge\ in\ B} (2x_ix_j - x_i^2 - x_j^2) \geq -2\sum(x_i^2 + x_j^2) = 2||x||^2$$

To get the lower bound on $C_2(G)$ we note that

$$\sum_{p_{ij}>0} d(i,j)^2 p_{ij} \geq \frac{\epsilon}{2} \cdot n(\log_k n - 2)^2$$

since the distances of the entries in B are at least $\log_k n - 2$.

$$-\sum_{p_{ij}<0} d(i,j)^2 p_{ij} = kn$$

Thus $C_2(G) = \Omega(\log n)$.

□

Chapter 10

Error Correcting Codes

Notes taken by Elon Portugaly

Summary: An error correcting code is set of words in $\{0,1\}^n$. Its distance is the minimal Hamming distance between two codewords. Therefore, if we transmit a codeword through a noisy channel that flips some of the bits, then we can correct the errors, as long as the number of bits flipped is bounded by half the distance. There is a trade-off between the size of the code and the number of errors it can correct.

There are lower bounds (Gilbert Varshamov) and upper bounds (The Balls Bound, MRRW) on the size and correction capabilities of codes.

A linear code is an error correcting code that is also linear subspace of $\{0,1\}^n$.

Expanders can be used to build error correcting codes, that have large size and distance. These codes are also efficiently decodable.

10.1 Definition of Error Correcting Codes

Definition 10.1. A Code is a set $C \subseteq \{0,1\}^n$.

Definition 10.2. The distance of C is $d \equiv dist(C) \equiv \min_{\substack{x \neq y \\ x,y \in C}} d_H(x,y)$, where $d_H(x,y)$ is the *hamming* distance between x and y (the number of coordinates x and y differ on).

Definition 10.3. The rate of C is $r \equiv rate(C) \equiv \frac{\log |C|}{n}$.

When defining a code, we desire both $|C|$ and d to be as large as possible.

10.1.1 Motivation

The setting we look at, is as follows:

We would like to send information through a noisy channel, that may flip some of the bits we send. We code our information using a set of words $C \subset \{0,1\}^n$ that we transmit through the channel, and assume that the number of bits the channel may flip in the transmitted word is bounded from above.

In this setting, it is clear that we would like the number of codewords available to be as large as possible, therefore we want $|C|$ to be large. We also would like to be able to reconstruct the codeword that was sent from the corrupted codeword received. Therefore, we would like that no two codewords could appear the same after they have been corrupted by the noise. If the number of bits the channel can flip is limited by k, and $d > 2k$, the last requirement is fulfilled. When d is larger, we can deal with noisier channels.

$|C|$ strings can be encoded using C. Therefore, we can transmit $\log |C|$ information bits, using n channel bits. This achieves channel utilization of $\frac{\log |C|}{n}$, which is the rate of the code.

Note: A Code refers to the set of codewords and not to the process of encoding/decoding. However, in any practical aplication, we would like the code to be efficiently encodable and decodable.

10.2 Asymptotic bounds

10.2.1 A lower bound: Gilbert Varshamov

This bound shows that good codes can be built.

Theorem 10.4. *One can build a length n code with distance d and size* $\geq \frac{2^n}{\textit{volume of a radius d hamming ball}}$.

Proof. Following is an exponential time greedy algorithm that builds such a code:

- Initiate $S = \{0,1\}^n, C = \emptyset$.
- Repeat until S is empty:
  - Pick any point $x \in S$, and add it in the code.
  - Remove all the points in S that are within distance d from x.

Analysis: The volume of a hamming ball of radius d in $\{0,1\}^n$ is $\sum_{i=0}^{d} \binom{n}{i}$. Therefore, at most $\sum_{i=0}^{d} \binom{n}{i}$ points are removed from S in each iteration, and since the number of points in S at the beginning is 2^n, the number of iterations and thus the size of C at the end of the process must be at least $\frac{2^n}{\sum_{i=0}^{d}\binom{n}{i}} \equiv \frac{2^n}{\text{volume of a radius } d \text{ ball}}$. □

Define $\delta \equiv \frac{d}{n}$. For $d \leq \frac{n}{2}$ we have $\sum_{i=0}^{d} \binom{n}{i} \approx \binom{n}{\delta n} \approx 2^{nH(\delta)}$, where H is the binary entropy function

$$H(x) = -x \log x - (1-x)\log(1-x).$$

Therefore

$$2^{rn} \geq \frac{2^n}{\binom{n}{\delta n}} \approx 2^{n(1-H(\delta))},$$

implying that for large n, the rate of the above code satisfies $r \geq 1 - H(\delta)$.

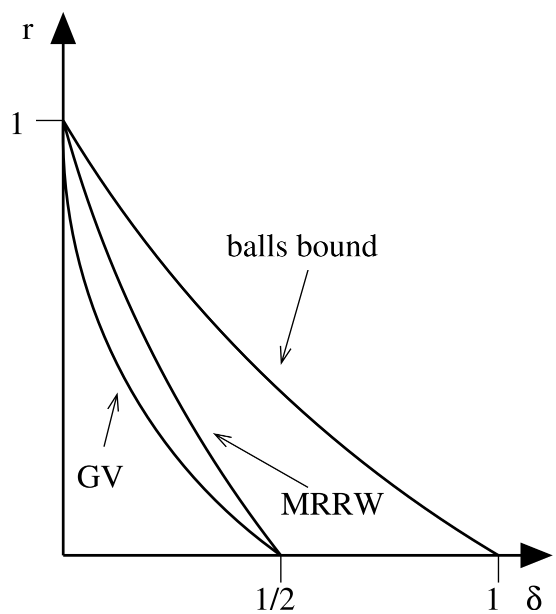


Figure 10.1: Illustrating Upper and Lower Bounds on rate vs. the relative distance

10.2.2 An Upper Bound - The Balls Bound

Theorem 10.5. *For any length n, distance d code C we have $|C| \leq \frac{2^n}{\text{volume of a ball of radius } \frac{d}{2}}$.*

This theorem implies the asymptotic bound: $r \leq 1 - H(\delta/2)$.

Proof. Given a distance d code C, a draw ball of radius $d/2$ around each of the points of C. If any two balls intersect, then the distance between them is smaller than d, which contradicts the definition of d. Therefore, the balls are disjoint, and their number is limited by the overall volume divided by the volume of each ball. □

A much stronger upper bound was shown by McEliece, Rodemich, Rumsey and Welsh 77 (MRRW). The relations between the three bounds is illustrated in Figure 10.1.

10.3 Linear Codes

A linear code is a code that is a linear subspace of the n-dimensional space $\{0,1\}^n$. (In other words, it is closed under coordinate-wise addition mod 2.) Such codes have a polynomial size (in n) description. (For instance by specifying a basis.)

The Gilbert Varshamov algorithm can be modified to generate linear codes, and thus the Gilbert Varshamov bound applies to linear codes. Obviously, any general upper bound applies to linear codes as well.

Observation 10.6. *Since any linear code C must include the 0 codeword, then for linear codes,*

$$dist(C) = \min_{0 \neq x \in C} weight(x),$$

where $weight(x)$ is the number of non-zero coordinates of x.

> **Note:** Although, for a linear code, the encoding can be done in polynomial time, the decoding is in general NP-hard. I.e., given a linear code C (using some reasonable representation), and a vector x, the problem of finding the element of C that is closest to x is an NP-hard problem.

10.4 Using Expanders to generate Error Correcting Codes

10.4.1 Defining Codes using Bipartite Graphs

Consider a bipartite graph with n vertices on the left and m vertices on the right side. We call the vertices on left *variables*, and the vertices on the right *constraints*. Each variable may assume the value of 0 or 1, and we say that a constraint is satisfied if the sum of all the variables adjacent to it is zero mod 2.

Example 10.7. The graph in Figure 10.2 illustrates such a variables and constraints graph. For this graph, constraint $y4$ is satisfied iff $x3 + x6 + x9 + x14 = 0$ while constraint $y10$ is satisfied iff $x12 + x14 = 0$. Note that the same variable may appear in more than one constraint.

To define the code we refer to the variables as the coordinates of a vector $x \in \{0,1\}^n$. A vector x is in the code iff it satisfies all the constraints defined by the graph. We denote by $C(G)$, the code defined by G. The code then, is the set of all solutions of m linear equations on n variables. Therefore, $|C| \geq 2^{n-m}$ or $r \geq \frac{n-m}{n}$.

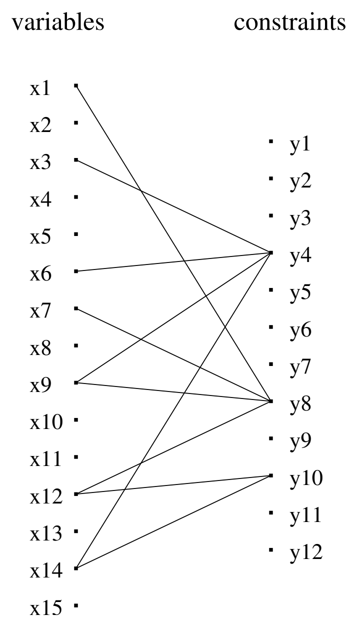


Figure 10.2: A Variables and Constraints Graph

10.4.2 Codes Using Left Side Expanders

Theorem 10.8. Sipser and Spielman (95): *Let $G = (V_L; V_R, E)$ be a bipartite graph of size $|V_L| = n$, $|V_R| = m$, that is k-regular on the left. Assume furthermore, that for any $S \subseteq V_L$ of size at most αn, $|\Gamma(S)| > \frac{k}{2}|S|$. Then $dist(C(G)) > \alpha n$.*

Proof.

Lemma 10.9. *Every set $S \subseteq V_L$ of size at most αn satisfies the* Unique Neighbor Property. *I.e, there exists $y \in V_R$ such that $|\Gamma(y) \cap S| = 1$*

> **Proof:** Consider a set $S \subseteq V_L$ of size $\leq \alpha n$. If the *Unique Neighbor Property* does not hold for S, then each $y \in \Gamma(S)$ has at least two neighbors in S. Therefore the number of edges leaving S is at least $2\Gamma(S) > 2\frac{k}{2}|S| = k|S|$, which contradicts the left regularity of G. □

Assume for the purpose of contradiction that $dist(C) \leq \alpha n$. Then there exists a nonzero codeword whose weight is at most αn (Observation 10.6). Let w be such a codeword, and X be its support (the set of coordinates v where $x_v = 1$), and let $y \in \Gamma(X)$ be a vertex guaranteed by the Unique Neighbor Property for X. The constraint defined by y is that the sum of all the variables that are neighbors of y is even, but in the assignment defined by w, only one of those variables is assigned the value 1. Therefore, the constraint cannot be satisfied, and w cannot be a codeword. A contradiction. □

Theorem 10.10. Efficient Decoding, Sipser and Speilman (95): *In the above conditions, if the expansion of sets of size at most αn is $> \frac{3}{4}k$, and if the distance of the input word ω' from a codeword ω is at most $\frac{\alpha}{2}n$, then the following decoding algorithm will return ω in a linear number of iterations:*

While there exists a variable such that most of its neighboring constraints are not satisfied, flip it.

Proof. Let A be the set of errors in ω', i.e. $A = \{v : \omega'_v \neq \omega_v\}$. If A is empty, we are done. Otherwise, assume that $|A| \leq \alpha n$. (We need the assumption to guaranty the expansion, and we will prove later that this assumption holds throughout the running of the algorithm.)

Partition $\Gamma(A)$ to satisfied neighbors S and unsatisfied neighbors U. Then:

$$|U| + |S| = |\Gamma(A)| > \frac{3}{4}k|A|. \tag{10.1}$$

Now, count the edges between A and $\Gamma(A) = U \cup S$. There are at least $|U|$ edges leaving U, and at least $2|S|$ edges leaving S (every vertex in S must have at least two neighbors in A). Therefore,

$$|U| + 2|S| \leq k|A|.$$

Combining this with (10.1) we get that

$$k|A| - |U| \geq 2|S| > 2(\frac{3}{4}k|A| - |U|),$$

and therefore,

$$|U| > \frac{1}{2}k|A|. \tag{10.2}$$

So more than $\frac{1}{2}k|A|$ neighbors of the $|A|$ members of A are unsatisfied. Therefore there is a variable in A that has $> \frac{1}{2}k$ unsatisfied neighbors. That means that as long as there are errors, there is a variable that most of its neighbors are unsatisfied. Since by definition, $|U|$ decreases with every step, we deduce that:

Corollary 10.11. *If the distance from ω does not exceed αn throughout the run of the algorithm, then U will reach the empty set, and the algorithm will halt with the codeword ω.*

To show that A is always $\leq \alpha n$, note that in the beginning, $|A_0| \leq \frac{\alpha}{2}n$, and therefore $|U_0| \leq |\Gamma(A_0)| \leq k\frac{\alpha}{2}n$. Therefore, since $|U|$ is decreasing, throughout the running of the algorithm

$$|U| \leq k\frac{\alpha}{2}n. \tag{10.3}$$

We know that at any step $|A_i|$ changes by ± 1. Therefore, if at any time $|A|$ exceeds αn, there must be a time i when $|A_i| = \alpha n$ (we can assume that αn is an integer). Then by (10.2), $|U| > k\frac{\alpha}{2}n$, which is a contradiction to (10.3). $\square$

Chapter 11

Lossless Conductors and Expanders

Notes taken by Ariel Elbaz, Yuval Filmus and Michal Igell

Summary: In this lecture we define conductors, a generalization of expanders. Extractors, dispersers and condensers are all types of conductors. We use the new structures to explicitly construct lossless expanders.

11.1 Min-entropy

Min-entropy measures the rarity of the most probable event. If all events occur at probability at most 2^{-k}, then the min-entropy is at most k, and vice versa. For a random variable X (or a distribution) over some finite set S, let

$$\mathrm{Supp}(X) = \{x \in S : \Pr[X = x] > 0\}.$$

Definition 11.1. The min-entropy of a distribution X

$$H_\infty(X) = \min_x \left\{ \log \frac{1}{\Pr[X = x]} \right\} = -\max_x \left\{ \log \big(\Pr[X = x] \big) \right\}$$

where minimum and maximum are taken over $x \in \mathrm{Supp}(X)$.

Note: throughout this work, $\log$ is always taken to base 2.

Definition 11.2. The Rényi entropy of a distribution X is

$$H_2(X) = -\log \sum_x (\Pr[X = x])^2 = -\log \mathbb{E}\Big[\Pr[X = x] \Big].$$

Lemma 11.3. *The min-entropy and Rényi entropy of a distribution X obey the following inequality:*

$$H_\infty(X) \le H_2(X) \le 2H_\infty(X)$$

The left inequality is tight iff X is uniformly distributed on some set S.

Proof. Since $\log$ is a monotone increasing function, we have

$$\begin{aligned} H_2(X) &= -\log \mathbb{E}[\, \Pr[X = x]] \\ &\ge -\log \max_{x \in \mathrm{Supp}(X)} \Pr[X = x] = H_\infty(X) \end{aligned}$$

and equality holds iff X is uniformly distributed over $\mathrm{Supp}(X)$.
On the other hand, let $x_M = \mathrm{argmax}_x \Pr[X = x]$. We have

$$\begin{aligned} H_2(X) &= -\log\Big(\sum_x (\Pr[X=x])^2\Big) \\ &= -\log\Big(\sum_{x \neq x_M} (\Pr[X=x])^2 + (\Pr[X = x_M])^2\Big) \\ &\leq -\log\Big((\Pr[X=x_M])^2\Big) \\ &= -2\log\Big(\Pr[X = x_M]\Big) = 2H_\infty(X) \end{aligned}$$

□

A distribution which is uniformly distributed on some set S is called a flat distribution, and both its min-entropy and its Rényi entropy are $\log|S|$. These are maximal among distributions with support S.

Flat distributions combine to make the most general distributions:

Lemma 11.4. *Every distribution over a finite set is a convex combination of a finite number of flat distributions. In other words, if X is a distribution then $X = \sum p_i U(S_i)$, where $S_i \subset \mathrm{Supp}(X)$, $p_i \geq 0$ and $\sum p_i = 1$.*

Proof. The proof goes by induction on $|\mathrm{Supp}(X)|$. If $\mathrm{Supp}(X) = \{x\}$ then $X = U(x)$. Otherwise, let $x_m = \mathrm{argmin}\Pr[X = x]$, and let $p_m = \Pr[X = x_m]$. We have $X = p_m U(\mathrm{Supp}(X)) + (1 - p_m)Y$, where $\mathrm{Supp}(Y) \subsetneq \mathrm{Supp}(X)$ since $x_m \notin \mathrm{Supp}(Y)$. □

If we combine a distribution Y with a flat distribution X, the joint distribution has min-entropy equal to $H_\infty(Y) + H_\infty(X)$:

Lemma 11.5. *Suppose X is a flat distribution, with $H_\infty(X) = c$, and let Y be another distribution. Then $H_\infty(X, Y) \leq k$ iff for every $x \in \mathrm{Supp}(X)$, we have $H_\infty(Y|X = x) \leq k - c$.*

Proof. Suppose first that $H_\infty(X, Y) \leq k$. If $H_\infty(Y|X = x) \geq k - c$ for some $x \in \mathrm{Supp}(X)$, then $\Pr[Y = y|X = x] \leq 2^{c-k}$ for some $y \in \mathrm{Supp}(Y)$. Since X is flat, $\Pr[X = x, Y = y] \leq 2^{c-k}2^{-c} = 2^{-k}$, contradicting the promise $H_\infty(X, Y) \leq k$.

Next, suppose that $H_\infty(Y|X = x) \leq k - c$ for all $x \in \mathrm{Supp}(X)$. Then for all $y \in \mathrm{Supp}(Y)$, we have $\Pr[Y = y|X = x] \geq 2^{c-k}$. Since X is flat, $\Pr[Y = y] \geq 2^{c-k}2^{-c} = 2^{-k}$. Therefore, $H_\infty(Y) \leq k$.

□

We complete this section with a technical lemma, showing how to divide a joint distribution according to conditional min-entropy.

Definition 11.6. Two distributions X and Y over the same set S are said to be ϵ-close if $|\Pr[X \in P] - \Pr[Y \in P]| \leq \epsilon$ for every subset $P \subset S$. Alternatively, X and Y are ϵ-close if $\sum_s |\Pr[X = s] - \Pr[Y = s]| \leq 2\epsilon$. We leave the reader to show that the two definitions are equivalent.

Lemma 11.7. *Let X_1 and X_2 be two distributions. Given $\epsilon > 0$ and a, there exist distributions Y_1 and Y_2 such that*

- *The joint distributions (X_1, X_2) and (Y_1, Y_2) are ϵ-close;*
- *The joint distribution (Y_1, Y_2) is a convex combination of two other joint distributions $(\hat{Y}_1, \hat{Y}_2)$ and $(\check{Y}_1, \check{Y}_2)$, both having min-entropy at least $H_\infty(X_1, X_2) - \log\frac{1}{\epsilon}$;*
- *For all $x \in \mathrm{Supp}(\hat{Y}_1)$ we have $H_\infty(\hat{Y}_2|\hat{Y}_1 = x) \geq a$;*
- *For all $x \in \mathrm{Supp}(\check{Y}_1)$ we have $H_\infty(\check{Y}_2|\check{Y}_1 = x) \leq a$;*

Proof. We begin by constructing $(\hat{Y}_1, \hat{Y}_2)$ and $(\check{Y}_1, \check{Y}_2)$. We split $\mathrm{Supp}(X_1)$ according to $H_\infty(X_2|X_1 = x)$:

$$\hat{X} = \{x : H_\infty(X_2|X_1 = x) \geq a\},$$
$$\check{X} = \{x : H_\infty(X_2|X_1 = x) < a\}.$$

Now we can define

$$\Pr[\hat{Y}_1 = y_1, \hat{Y}_2 = y_2] = \Pr[X_1 = x_1, X_2 = x_2 | x_1 \in \hat{X}],$$
$$\Pr[\check{Y}_1 = y_1, \check{Y}_2 = y_2] = \Pr[X_1 = x_1, X_2 = x_2 | x_1 \in \check{X}].$$

In other words, $\hat{Y}_1$ gets only values in $\hat{X}$, and $\check{Y}_1$ is restricted to $\check{X}$.

If $p = \Pr[X_1 \in \hat{X}]$, then the probability of each event in $(\hat{Y}_1, \hat{Y}_2)$ is multiplied by $1/p$, and the probability of each event in $(\check{Y}_1, \check{Y}_2)$ is multiplied by $1/(1-p)$. Therefore, if $\epsilon \leq p \leq 1-\epsilon$ then the min-entropy of $(\hat{Y}_1, \hat{Y}_2)$ and $(\check{Y}_1, \check{Y}_2)$ is reduced by at most $\log 1/\epsilon$. Since $(X_1, X_2) = p(\hat{Y}_1, \hat{Y}_2) + (1-p)(\check{Y}_1, \check{Y}_2)$ we can take $(Y_1, Y_2) = (X_1, X_2)$.

If, for example $p < \epsilon$, $(\check{Y}_1, \check{Y}_2)$ still has high enough min-entropy, and so we take $(Y_1, Y_2) = (\check{Y}_1, \check{Y}_2)$. This distribution is ϵ-close to (X_1, X_2):

$$\sum_{x_1 \in \hat{X}, x_2} \left| \Pr[X_1 = x_1, X_2 = x_2] - \Pr[\check{Y}_1 = x_1, \check{Y}_2 = x_2] \right| +$$
$$\sum_{x_1 \in \check{X}, x_2} \left| \Pr[X_1 = x_1, X_2 = x_2] - \Pr[\check{Y}_1 = x_1, \check{Y}_2 = x_2] \right| =$$
$$p + \left(\frac{1}{1-p} - 1 \right)(1-p) = 2p < 2\epsilon.$$

$\square$

11.2 Conductors and lossless expanders

11.2.1 Conductors

Loosely speaking, a conductor is a function that transfers entropy from its inputs to its output. In other words, if the input distributions have high min-entropy, then the output distribution will have high min-entropy.

Definition 11.8 (conductors). A function $E\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $(k_{\max}, a, \epsilon)$-conductor if for any distribution X on $\{0,1\}^n$ satisfying $H_\infty(X) = k \leq k_{\max}$, the distribution $E(X, U_d)$ is ϵ-close to a distribution Y whose min-entropy is at least $k + a$.

Remark: In this text we identify a distribution with a random variable sampled from it.

In other words, E gets two inputs: a distribution X with min-entropy $k \leq k_{\max}$, and the uniform distribution U_d (with min-entropy d). The output is ϵ-close to a distribution with min-entropy at least $k + a$.

The following definitions (definitions 3.4 to 3.7 from [CRSW02]) are several special cases of conductors, which we will later use.

Definition 11.9 (extracting conductors). A function $E\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is an (ϵ, a) extracting conductor if for any $0 \leq k \leq m - a$, and any k-source X over $\{0,1\}^n$, the distribution $E(X, U_d)$ is a $(k + a, \epsilon)$-source.

Note that if $E\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is an (ϵ, a) extracting conductor, it is also an $(m - a, \epsilon)$ extractor.

Definition 11.10 (lossless conductors). A function $E\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is an $(k_{\max}, \epsilon)$ lossless conductor if for any $0 \leq k \leq k_{\max}$, and any k-source X over $\{0,1\}^n$, the distribution $E(X, U_d)$ is a $(k + d, \epsilon)$-source.

The next two definitions require that the prefix of the output is an extracting conductor:

Definition 11.11 (buffer conductors). A pair of functions $\langle E, C\rangle\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m \times \{0,1\}^b$ is an $(k_{\max}, a, \epsilon)$ buffer conductor if E is an (ϵ, a) extracting conductor, and $\langle E, C\rangle$ is an $(k_{\max}, \epsilon)$-lossless conductor.

Definition 11.12 (permutation conductors). A pair of functions $\langle E, C\rangle\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m \times \{0,1\}^b$, where $n + d = m + b$ is an (ϵ, a) permutation conductor if E is an (ϵ, a) extracting conductor, and $\langle E, C\rangle$ is a permutation over $\{0,1\}^{n+d}$.

11.2.2 Lossless expanders

Having defined conductors, we define lossless expanders and reveal the connection.

Definition 11.13. A d-regular bipartite graph is a $(k_{\max}, \epsilon)$-lossless expander if every set of $k \leq k_{\max}$ vertices on the left side has at least $(1-\epsilon)d \cdot k$ neighbors.

That is, a lossless expander has almost the maximal expansion possible for a d-regular graph, for small enough subsets. An alternative view is that for every subset on the left side, most neighbors will be unique neighbors, i.e. neighboring a single vertex of the set. Naturally, $k_{\max}$ should be somewhat smaller than m/d for this to be possible, where m is the number of vertices on the right side.

Expanders, condensers and dispersers (which we don't define here) can be viewed as special cases of conductors. For example, a lossless conductor E can be viewed as a 2^d-regular bipartite graph with 2^n vertices on the left side and 2^m vertices on the right side, where each set of $2^k \leq 2^{k_{\max}}$ vertices has at least $(1-\epsilon)2^{k+d}$ neighbors on the right side. In other words, we get a $(2^{k_{\max}}, \epsilon)$-lossless expander.

We can explicitly construct constant-degree lossless expanders which losslessly expand sets of size $O(M/D)$, where $M = 2^m$ is the number of right vertices, and $D = 2^d$ is the left degree.

Theorem 11.14. *For any $\epsilon > 0$, there is an explicit family of $D = (N/\epsilon M)^c$-regular bipartite graphs which are $(O(\epsilon M/D), \epsilon)$-lossless expanders, where N is the number of vertices on the left side, $M < N$ is the number of vertices on the right side, $N = O(M)$ and c is a constant. Note that since $N = O(M)$, the degree itself is bounded by a constant depending only on ϵ.*

We will prove the theorem by constructing an explicit family of $(\log \frac{\epsilon M}{D}, \log D, \epsilon)$-conductors.

11.3 The Construction

The required lossless conductors will be constructed using a zigzag product. Let us recall the zigzag product for expanders:

Definition 11.15 (The zigzag product of two regular bipartite graphs). Let H be a d-regular bipartite graph with s vertices on each side, and let G be an s-regular bipartite graph with n vertices on each side.

The zigzag product $G \textcircled{z} H$ is a d^2-regular bipartite graph with sn vertices on each side, which we may conceive as n copies of H, one per each vertex of G. Pick a left vertex $(x, y) \in [n] \times [s]$. The edges emanating from (x, y) are labeled using labels from $[d] \times [d]$. The edge labeled (a, b) is determined as follows:

1. Take a left to right step in the local copy of H (use a to choose an edge).
2. Take a left to right step on G, that is between copies of H.
3. Take a left to right step in the new local copy of H (use b to choose an edge).

We expand on the second step. Suppose after the first step we are at (x, y'). Let $x' \in G$ be the y'-th neighbor of x, and let x be the z-th neighbor of x'. Then the second step takes us from (x, y') to (x', z).

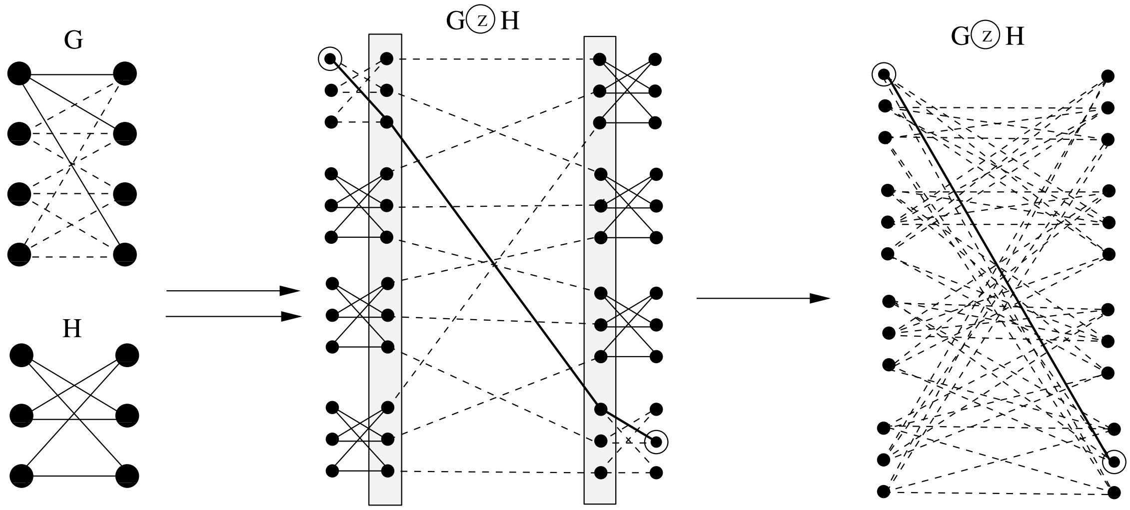


Figure 11.1: Zigzag product of bipartite graphs

11.3.1 The zigzag product for conductors

We now define the zigzag product for conductors, and show that composing conductors with carefully selected parameters, we get constant degree lossless expanders, as in theorem 11.14.

Recall that the zigzag theorem, proved in a previous lecture, shows that $G \textcircled{z} H$ is an expander if both G and H are such. Moreover, the degree of $G \textcircled{z} H$ is related to H, while its size and expansion are related to both G and H. Unfortunately, while $\deg(G \textcircled{z} H) = \deg^2(H)$, the expansion of $G \textcircled{z} H$ is the minimum between the expansion of H and the expansion of G. That is, the expansion of $G \textcircled{z} H$ is at most the square root of $deg(G \textcircled{z} H)$ (this can be seen by considering a set consisting of a single copy of H). This expansion is too low for lossless expanders, which require expansion almost as big as the degree.

Recall also that in the proof of the zigzag theorem, of the two random steps (steps 1 and 3 in definition 11.15), only one is "used" and contributes to the output entropy. That is, out of d^2 choices, only d are surely increasing the entropy.

Here we try to avoid this loss of entropy by buffering the random choices made at each step, and then using a lossless conductor, together with some fresh truly random bits, to condense the leftovers of entropy.

The name "conductor" suggests an analogy to electricity or water conductors. Another analogy to water is to think of the lossless conductor construction as putting a bucket beneath each object, so that when we pour randomness (water) into it, the leftovers (unused randomness, beyond the $k_{\max}$ bound), are stored for later use.

In the zigzag product for conductors we make use of several objects. We remark that they can be explicitly constructed using lemmas from [CRSW02].

Let us then assume that we have in our hands the following objects:

1. $\langle E_1, C_1 \rangle \colon \{0,1\}^{n_1} \times \{0,1\}^{d_1} \to \{0,1\}^{m_1} \times \{0,1\}^{b_1}$, a permutation conductor that can be taken from lemma 4.4 in [CRSW02];

2. $\langle E_2, C_2 \rangle \colon \{0,1\}^{n_2} \times \{0,1\}^{d_2} \to \{0,1\}^{d_1} \times \{0,1\}^{b_2}$, a buffer conductor;

3. $E_3 \colon \{0,1\}^{b_1+b_2} \times \{0,1\}^{d_3} \to \{0,1\}^{m_3}$, a lossless conductor.

Both E_2 and E_3 can be taken from lemma 4.13 in [CRSW02].

We describe the zigzag product for conductors.

Set $n = n_1 + n_2$, $d = d_2 + d_3$ and $m = m_1 + m_3$. For $x_1 \in \{0,1\}^{n_1}$, $x_2 \in \{0,1\}^{n_2}$, $r_2 \in \{0,1\}^{d_2}$ and $r_3 \in \{0,1\}^{d_3}$, define

$$E \colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$$

by $E(x_1x_2, r_2r_3) \stackrel{def}{=} y_1y_3$, where

- $(y_2, z_2) = \langle E_2, C_2\rangle(x_2, r_2)$
- $(y_1, z_1) = \langle E_1, C_1\rangle(x_1, y_2)$
- $y_3 = E_3(z_1z_2, r_3)$

z_1 and z_2 are buffers of $\langle E_1, C_1\rangle$ and $\langle E_2, C_2\rangle$.

Figure 2 shows an example of this construction.

Our notation in figure 2 is that (y_i, z_i) are the output of $\langle E_i, C_i\rangle$ on the inputs (x_i, r_i) (except for E_3 which has only one output). As $\langle E_1, C_1\rangle$ gets it's seed from $\langle E_2, C_2\rangle$, we get that $r_1 = y_2$, and since E_3 gets it's input from $\langle E_1, C_1\rangle$ and $\langle E_2, C_2\rangle$, we get that $x_3 = z_1z_2$.

Recall that the zigzag product for bipartite graphs, $G \circledz H$, uses H twice. In the new construction, the first use is replaced with E_2. This ensures that when x_2 has high min-entropy, y_2 is close to uniform, and is a good seed for E_1. The second use of H is replaced with E_3, which is a lossless conductor. The role of E_3 is to transfer entropy lost in E_1 and E_2 to the output.

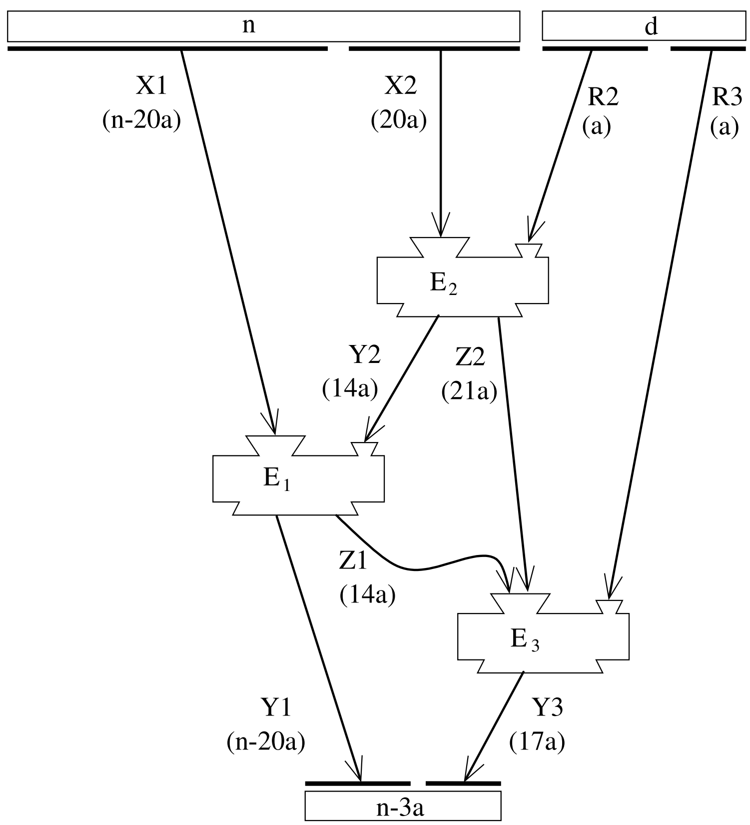


Figure 11.2: Entropy flow in a lossless conductor

11.3.2 A specific example of a lossless conductor

It will be constructive for the sake of explanation to look at a concrete example of the construction.

Set $a = 1000\log(\frac{1}{\epsilon})$ and $d = 2a$. Then we have

- $\langle E_1, C_1\rangle \colon \{0,1\}^{n-20a} \times \{0,1\}^{14a} \to \{0,1\}^{n-20a} \times \{0,1\}^{14a}$, an $(n-30a, 6a, \epsilon)$ permutation conductor;

- $\langle E_2, C_2\rangle\colon \{0,1\}^{20a} \times \{0,1\}^a \to \{0,1\}^{14a} \times \{0,1\}^{21a}$, a $(14a, 0, \epsilon)$-buffer conductor;
- $E_3\colon \{0,1\}^{35a} \times \{0,1\}^a \to \{0,1\}^{17a}$, a $(15a, a, \epsilon)$ lossless conductor.

The result is

$$E\colon \{0,1\}^n \times \{0,1\}^{2a} \to \{0,1\}^{n-3a}$$

which is a $(n-30a, 2a, 4\epsilon)$-lossless conductor.

Let us try and follow the entropy flow from the input (x_1x_2, r_2r_3) to the output y_1y_3.

Let $k = H_\infty(X_1, X_2)$, then we show that if we start with min-entropy k, we end up with min-entropy of $k + 2a$. Two main ideas for this example are emphasized:

1. The entropy is conserved by $\langle E_1, C_1\rangle, \langle E_2, C_2\rangle$, because $\langle E_1, C_1\rangle$ is a permutation conductor, and $\langle E_2, C_2\rangle$ is a buffer conductor. Therefore, we get

$$k + a = H_\infty(X_1, X_2, R_2) = H_\infty(X1, Y_2, Z_2) = H_\infty(Y_1, Z_1, Z_2)$$

2. We want to verify that enough entropy is transferred, using E_1 and E_2, to Y_1. We know that $H_\infty(Y_1, Z_1, Z_2) = k + a$, and if we prove that $H_\infty(Y_1) \geq k - 14a$, then $H_\infty(Z_1, Z_2|Y_1) \leq 15a$. In that case E_3, which is a $(15a, a, \epsilon)$-conductor will conduct a bits of entropy from R_3 to Y_3. That is, all the entropy of Z_1, Z_2 will be transferred to the output Y_3, without any entropy loss, as we want.

To prove that $H_\infty(Y_1) \geq k - 14a$, we look at the two cases, which lemma 11.7 essentially shows that are sufficient to prove the general case.

Case 1 For all $x_1 \in \mathrm{Supp}(X_1)$, we have $H_\infty(X_2|X_1 = x_1) \geq 14a$.

In this case, $H_\infty(Y_2|X_1 = x_1) = 14a$, for any $x_1 \in \mathrm{Supp}(X_1)$. Therefore Y_2 can be used as a seed for $\langle E_1, C_1\rangle$ for any $x_1 \in \mathrm{Supp}(X_1)$. We know that $H_\infty(X_1) \geq k - 20a$, and therefore E_1 conducts $6a$ bits of entropy from the seed into Y_1, and we get that $H_\infty(Y_1) \geq k - 14a$.

Case 2 For all $x_1 \in \mathrm{Supp}(X_1)$, we have $H_\infty(X_2|X_1 = x_1) \leq 14a$.

Since $H_\infty(X_1, X2) = k$, it follows that $H_\infty(X_1) \geq k - 14a$. Therefore, since E_2 is a lossless extractor, $H_\infty(Y_2|X_1 = x_1) \geq H_\infty(X_2|X_1 = x_1)$ for any $x_1 \in \mathrm{Supp}(X_1)$. It follows that $H_\infty(X_1, Y_2) \geq H_\infty(X_1, X_2) = k$. Since $\langle E_1, C_1\rangle$ is a permutation, also $H_\infty(Y_1, Z_1) \geq k$, and again we get that $H_\infty(Y_1) \geq k - 14a$.

To complete the example, for any $y_1 \in \mathrm{Supp}(Y_1)$,

$$H_\infty(Z_1, Z_2|Y_1 = y_1) \leq H_\infty(Y_1, Z_1, Z_2) - H_\infty(Y_1) \leq (k + a) - (k - 14a) = 15a.$$

Therefore, the lossless extractor E_3 transfers a bits of entropy from R_3 to Y_3, and we get that $H_\infty(Y_3|Y_1 = y_1) \geq H_\infty(Z_1, Z_2|Y_1 = y_1) + a$. This implies that, $H_\infty(Y_1, Y_3) \geq k + 2a$, as needed.

Chapter 12

Cayley graph expanders

Notes taken by Eyal Rozenman

Summary: We describe ideas leading to an elementary construction of Cayley graphs which are expanders with relatively small degree.

A set of elements S in a group H is a *generating set* if every element of $h \in H$ can be written as $h = s_1 \cdot s_2 \ldots \cdot s_k$ with $s_i \in S$.

Definition 12.1. The *Cayley graph* $C(H, S)$ of a group H and a generating set S is a graph whose vertices are the elements of H, and where (g, h) is an edge if $g \cdot s = h$ for some $s \in S$.

This generally defines a directed graph. If the set of generators S is *symmetric* - i.e. $s \in S$ iff $s^{-1} \in S$, then (g, h) is an edge iff (h, g) is, and we have an undirected graph. It is a regular graph of degree $|S|$.

Example 12.2.

- The additive cyclic group $C_d = \mathbb{Z}/d\mathbb{Z}$ with generators $S = \{+1, -1\}$ is the cycle on d vertices.
- The additive group of the vector space $(F_2)^d$ over the field with two elements is generated by the standard basis vectors $e_1 = (1, 0, \ldots, 0), e_2, \ldots, e_d$. The Cayley graph is the discrete cube - a d-regular graph.

Consider the following construction, resulting in the graph depicted in figure 12.1. The degree of the discrete cube is equal to the number of vertices in the d-cycle, so we can form a zigzag product of the two. Let's look at the simpler replacement product. In this product we replace every vertex of $(F_2)^d$ by a cloud of d vertices representing C_d. On each cloud we preserve the edges of the original C_d. We also connect each vertex in the cloud to one of the d neighbors of the cloud in F_2^d. For example, let's connect vertex (v, h) to $(v + e_h, h)$. Like the zigzag construction, this product is an expander if the original two graphs are expanders.

We started with two Cayley graphs and created a third graph by a graph-theoretic construction. Is the resulting graph also a Cayley graph of some group? The answer is yes. It's the *semidirect product* of C_d and F_2^d. To define this product we shall need (alas) some more definitions:

Definition 12.3. An *action* of a group B on a group A is a group homomorphism $\phi : B \to Aut(A)$. In other words, each element $b \in B$ corresponds to an automorphism ϕ_b of A, and we demand that $\phi_{b_1 \cdot b_2} = \phi_{b_1} \phi_{b_2}$.

Definition 12.4. Suppose a group B acts on a group A. The *semidirect product* $A \rtimes B$ is a group whose elements are pairs (a, b) where $a \in A$ and $b \in B$. We define

$$(a_1, b_1) \cdot (a_2, b_2) = (a_1 \cdot \phi_{b_1}(a_2), b_1 \cdot b_2).$$

Example 12.5.

- The direct product of two groups $A \times B$ is a special case of a semidirect product where ϕ_b is the identity automorphism of A for all $b \in B$. In this case $(a_1, b_1) \cdot (a_2, b_2) = (a_1 \cdot a_2, b_1 \cdot b_2)$.

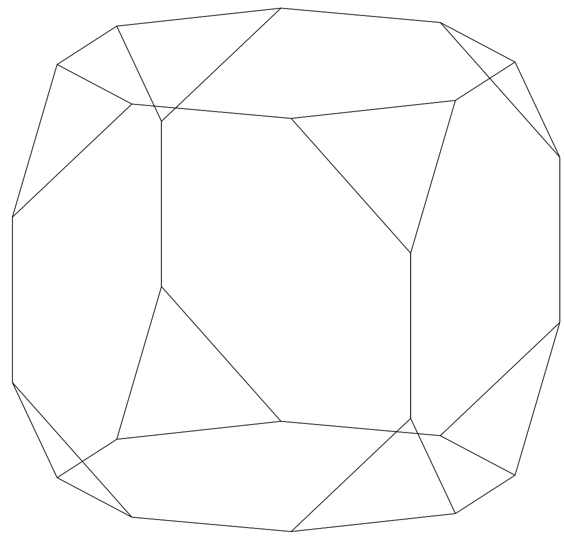


Figure 12.1: *A Cayley graph of $F_2^3 \rtimes C_3$*

- We have an action of C_d on F_2^d: the element $h \in C_d$ cyclically permutes the coordinates of F_2^d by h places. We can thus define $F_2^d \rtimes C_d$.

It turns out that under certain conditions there is a close relation between the semidirect product of groups and the replacement product of their Cayley graphs. Suppose a group B acts on a group A. The *orbit* of an element $a \in A$ under the action of B is the set $\{\phi_b(a)|b \in B\}$. For example, the orbit of $v \in F_2^d$ under the action of C_d is the set of all cyclic shifts of v

Claim 12.6. *Suppose we have two groups A, B with sets of generators S_A, S_B, such that $|B| = |S_A|$. Further, suppose that B acts on A in such a way that S_A is the orbit of one of the elements $x \in S_A$. Then $S := (1, S_B) \cup \{(x, 1)\}$ generates $A \rtimes B$, and $C(A \rtimes B, S)$ is a replacement product of $C(A, S_A)$ and $C(B, S_B)$.*

Proof. To see that S indeed generates $A \rtimes B$ notice that the elements $(S_A, 1) \cup (1, S_B)$ generate it. Now observe that $(1, b) \cdot (x, 1) \cdot (1, b^{-1}) = (\phi_b(x), 1)$, so we can indeed generate all of $(S_A, 1)$ starting with $(x, 1)$, so $(1, S_B) \cup \{(x, 1)\}$ is generating. Look at $C(A \rtimes B, S)$. It consists of clouds of the elements of B, with the graph of $C(B, S_B)$ on them, since $(a, b) \cdot (1, s_b) = (a, b \cdot s_b)$. Between clouds we have edges like $(a, b) \cdot (x, 1) = (a \cdot \phi_b(x), b)$. So the cloud of the element a is indeed connected by one edge to each of the clouds of the neighbors of a in the graph $C(A, S_A)$, and this is a replacement product. $\square$

Exercise 12.7. Under the assumptions of the claim, we can also describe the zigzag product of two Cayley graphs as a Cayley graph on $A \rtimes B$. Which generating set do we need?

Example 12.8. Look at F_2^3 and C_3 with the generators used above. In the Cayley graph of $F_2^3 \rtimes C_3$ with generators as in the claim, the neighbors of the cloud of $v = (1, 0, 0) \in F_2^3$ are:

$$
\begin{aligned}
(v, 0) \cdot (e_0, 0) &= (v + e_0, 0) \\
(v, 1) \cdot (e_0, 0) &= (v + e_1, 1) \\
(v, 2) \cdot (e_0, 0) &= (v + e_2, 2)
\end{aligned}
$$

And this is indeed a replacement product.

Recall that a replacement product of two expander graphs is again an expander. So we can try to make an expander graph which is also a Cayley graph using the semidirect product. As a starter, we look for generators for F_2^d that make it an expander. Since this is an abelian group, then for every set of generators the eigenvectors of the graph are the Fourier basis - which are vectors of the type $f_y(x) = (-1)^{<x,y>}$. For the standard basis $e_1, \dots e_d$ the eigenvalues are

$$\sum_i f_y(e_i) = \sum_i (-1)^{<e_i,y>} = d - 2|y|$$

Where $|y|$ is the Hamming weight of the binary vector y. So $\lambda_1 = d, \lambda_2 = d-2$. The second normalized eigenvalue is $(d-2)/d$ and obviously, this is not a good expander. How do we find another set of generators which does give an expander? Suppose we have a set $S = \{v_1, v_2, \dots, v_r\}$ of generators. Write an $|S| \times d$ matrix A whose rows are the elements of S. To be an expander, we need to satisfy, for every $y \neq 0$

$$\sum_i f_y(v_i) = \sum_i (-1)^{<v_i,y>} \leq c \cdot |S|$$

for some constant $c < 1$, independent of d. This means that the vectors Ax (with $x \neq 0$) don't have too much difference between the numbers of 0's and 1's. In particular, each word must have a sufficient number of 1's. In short - a good expander in this case gives an error correcting code. Using this intuition, one would try a random $2d \times d$ matrix A, which we know gives a good code, and indeed

Claim 12.9. *For a random $2d \times d$ binary matrix, almost surely $2d \cdot \delta < |Ax| < 2d \cdot (1-\delta)$ for all nonzero x (for some constant $\delta > 0$).*

So we have an expander. With two "small" flaws: (a) It's not explicit, and (b) The degree is too large. To get rid of flaw (b) we want not just an arbitrary set of generating vectors. We want the generators to be one orbit under the action of C_n, for example. Luckily, this also turns out to work

Claim 12.10. *Pick two random vectors $u, v \in F_2^d$. Consider the matrix A generated by the orbits of u and v under the action of C_d, that is, the $2d \times d$ matrix of the cyclic shifts of u and v. Then A (a.s.) satisfies $2d \cdot \delta < |Ax| < 2d \cdot (1-\delta)$ for all nonzero x (for some constant $\delta > 0$).*

So now, by using a semidirect product with C_d we get an expander. The only difference is that we used two orbits, instead of one as in claim 12.6.

Exercise 12.11. Claim 12.6 still holds when S_A is a union of k orbits under the action of B. The Cayley graph of the semidirect will have $|S_B| + k$ generators.

We can now use this idea to give a counterexample to

Conjecture 12.12. *If a group sequence G_n is an expander with one set of generators S_n of bounded size then it is also an expander with any other set of generators U_n of bounded size.*

Recall that we have two matrices that (with their inverses) make $SL_2(F_p)$ an expander. $SL_2(p)$ acts on the $p+1$ elements of the projective plane over F_p, so just as we did with C_d we can form the semidirect product $F_2^{p+1} \rtimes SL_2(p)$. It also turns out that there are two orbits of SL_2 that make F_2^{p+1} an expander (random orbits will do). So the semidirect product is an expander with 4 generators, for every p. On the other hand, we have another set of generators for which F_2^{p+1} is not an expander - the standard basis. This gives a set of 3 generators for $F_2^{p+1} \rtimes SL_2(p)$ which is not an expander (recall that the replacement product is never a better expander than its components).

Can we iterate the semidirect product construction to get a sequence of Cayley expander graphs the same way we did with the zigzag product? For a group G let $F_p[G]$ be the group ring over F_p. we would like to make the additive group an expander using a constant number of orbits under the action of G. If we could do that in general, we could define a sequence iteratively by $G_{i+1} = F_{p_i}[G_i] \rtimes G_i$. This is indeed possible with a proper choice of p_i and G_1:

Theorem 12.13. *$\lambda_2(C(G_n, S_n)) \leq 1/2$ and $S_n \leq log^{(n/2)}|G_n|$ where $log^{(n/2)}$ is the iterated logarithm.*

This is the (almost) best construction of this type we can hope to get using this construction, since the group G_n is a solvable group with solvability index at most n (as G_{n-1} is a normal subgroup with abelian quotient $F_{p_{n-1}}[G_{n-1}]$). In this case it is known that any generating set which gives $\lambda_2 \leq 1/2$ has cardinality at least $log^{(n)}|G_n|$

The property we need to make $F_p[G]$ an expander with a constant number of orbits is

Theorem 12.14. *Let $d_1, d_2, \ldots, d_t$ be the dimensions of the irreducible representations of G. if $|\{i : d_i < r\}| < c^r$ for all integer r and some constant c, then there is a constant number of orbits of $F_p[G]$ that make it an expander.*

Miraculously, $F_p[G]$ inherits this property if G is a **monomial group**, which means all its irreducible representations are induced from one-dimensional representations of subsets of G. Furthermore, $F_p[G]$ is also monomial! Even better, we can find the generating orbits explicitly. This gives a sequence of explicit Cayley graphs which are expanders with an "almost constant" number of generators.

Chapter 13

On eigenvalues of random graphs and the abundance of Ramanujan graphs

Notes taken by Danny Gutfreund

Summary: In this lecture we will be looking at eigenvalue distributions of random graphs and matrices. Our starting point is the question: Is it true that almost every graph is Ramanujan?

First, we consider generalizations of this question that look at the distribution of eigenvalues in general (and not only the second eigenvalue). We state "Wigner's semicircle law", and give a partial proof to a version of this law (By McKay) for regular graphs.

In the second part of the lecture, we define lifts of graphs and extend the definition of Ramanujan graphs to general (non-regular) graphs. We then state some conjectures and results regarding the abundance of Ramanujan graphs (under this new definition).

13.1 The eigenvalue distribution of random matrices and regular graphs

Open problem 13.1. Is it true that almost every d-regular graph is Ramanujan? More formally, is it true that,

$$\lim_{n\to\infty} Pr(\lambda_2(G_n) \leq 2\sqrt{d-1}) = 1$$

Where G_n is a random d-regular graph of size n.

Friedman gave a positive answer upto an ϵ additive factor,

Theorem 13.2. *For every $\epsilon > 0$,*

$$\lim_{n\to\infty} Pr(\lambda_2(G_n) \leq 2\sqrt{d-1} + \epsilon) = 1$$

Where G_n is a random d-regular graph of size n.

Extending this question to random symmetric matrices, and the distribution of all the eigenvalues, we have the following theorem by Wigner, known as "Wigner's semicircle law".

Theorem 13.3. *Let A_n be a $n \times n$ symmetric matrix over $\mathbb{R}$, where a_{ij} $(i \neq j)$ are sampled independently from a distribution F, and a_{ii} from a distribution G.*
Let $\lambda_1(A_n) \geq \cdots \geq \lambda_n(A_n)$ be the eigenvalues of A_n.
Define the empiric distribution,

$$W_n(x) = \frac{1}{n}|\{i : \lambda_i(A_n) \leq x\}|$$

Let,

$$W(x) = \lim_{n\to\infty} W_n(2x\sigma\sqrt{n})$$

where $\sigma^2 = var(F) = var(G)$.
If it holds that $\forall k, \int |x|^k dF, \int |x|^k dG < \infty$, *then* $W(x)$ *is continuous with density* $\frac{2}{\pi}\sqrt{1-x^2}$ *if* $|x| \leq 1$, *and 0 otherwise.*

Going back to eigenvalues of adjacency matrices of d-regular graphs, we have the following version of Wigner's semicircle law, proven by McKay.

Theorem 13.4. *Let* G_n *be an infinite sequence of d-regular graphs, such that, for all* $k \geq 3$, $C_k(G_n) = o(|G_n|)$, *where* $C_k(G_n)$ *is the number of length* k *cycles in* G_n.
Define,

$$F(G_n, x) = \frac{1}{|G_n|}|\{i : \lambda_i(G_n) \leq x\}|$$

Then for every x,

$$F(x) = \lim_{n\to\infty} F(G_n, x) = \int_{-2\sqrt{d-1}}^{x} \frac{d\sqrt{4(d-1)-z^2}}{2\pi(d^2-z^2)} dz$$

The idea of the proof is that under the assumption that there are very few short cycles, the neighborhood of almost every node looks like a tree. So we can count the number of cycle-free paths from a node to itself. We state this formally in the following lemma.

Lemma 13.5. *Let* G *be a d-regular graph, and let* v *be a node such that there are no cycles in its* r*-neighborhood. Then the number of paths of length* r *that start and end in* v *is 0 if* r *is odd, and if* $r = 2s$ *it is,*

$$\psi(r) = \sum_{j=1}^{s} \binom{2s-j}{s} \frac{j}{2s-j} d^j (d-1)^{s-j}$$

Proof. Clearly, if the r-neighborhood of v does not contain any cycles, then every length r path that starts and ends in v is cycle-free, and hence r must be even.

Let $r = 2s$. Every length r path that starts and ends in v, defines a sequence, $0 = \delta_0, \delta_1 \cdots \delta_r = 0$. Where δ_i is the distance from v after we did i steps on the path. Clearly, for every i, $\delta_i \geq 0$, and $|\delta_i - \delta_{i-1}| = 1$. We would like to count the number of such sequences in which exactly j out of $\delta_0 \cdots \delta_r$ are 0. This is a simple generalization of Catalan numbers and the answer is,

$$\binom{2s-j}{s} \frac{j}{2s-j}$$

We know that a path that defines such a sequence visits v exactly j times. Each time it leaves v it has exactly d possibilities for the next step. This gives the term d^j. In steps that go away from v, i.e. when $\delta_{i+1} - \delta_i = 1$, we have $d-1$ different nodes that we can continue too (we cannot backtrack to the previous node). and for steps that go toward v we have no choice because we have to go back on the same edge that we used before. There are $s-j$ steps of each type, altogether we have $(d-1)^{s-j}$ possibilities for such steps. □

If $C_k(G_n) = o(|G_n|)$ (for every $k \geq 3$), then for every constant r, almost every node has cycle-free r-neighborhood. Let $P_r(G_n)$ be the number of simple paths of length r from a node to itself in G_n. Then from Lemma 13.5 we conclude that,

$$\lim_{n\to\infty} \frac{P_r(G_n)}{G_n} = \psi(r)$$

Therefore, the function $F(x)$ must satisfy $\int x^r dF = \psi(r)$, for every r. In order to finish the proof of the theorem we need some inverse transformation, that calculates $F(x)$ out of its moments. This is achieved using the Chebyshev polynomials, but we do not include the details in this lecture note.

13.2 Random lifts and general Ramanujan graphs

We start by defining the notion of coverings and lifts.

Definition 13.6. Let G and H be two graphs. We say that a function $f : V(H) \to V(G)$ is a covering, if for every $v \in V(H)$, the restriction of f to the set of neighbors of v, $\Gamma_H(v)$, is one to one and onto $\Gamma_G(f(v))$.
If a covering function from H to G exists, we say that H lifts G.

Example 13.7. The zig-zag product of G and H (where H is the smaller of the two) lifts H^2.

Remark 13.8. If G is connected then every covering function on G has a covering number n, such that for every $v \in V(G)$, $|f^{-1}(v)| = n$, and for every $e \in E(G)$, $|f^{-1}(e)| = n$.

Definition 13.9. Let f be a covering function on G. For a node $v \in V(G)$, we say that $f^{-1}(v)$ is the fiber of v.

Let G be a connected graph. We denote by $L_n(G)$ the set of graphs that are lifts of G with covering number n. We would like to characterize the members of $L_n(G)$. If $H \in L_n(G)$ then $V(H) = V(G) \times [n]$. That is, for every $v \in V(G)$, the nodes $(v, 1), \cdots, (v, n)$ in $V(H)$ are the fiber of v. Next we define the edges of H, for every edge $(v, u) \in E(G)$ we take some permutation $\pi \in S_n$ and define the following edges in H, $((v, i), (u, \pi(i)))$ (for every $i \in [n]$). Thus every choice of permutations defines a member in $L_n(G)$. This also gives us a way to sample random elements from this set.

What can we say about the eigenvalues of lifts of G? We know that the eigenvalues of G are also the eigenvalues of its lifts. To see this, let $h : V(G) \to \mathbb{R}$ be an eigenfunction of G with eigenvalue λ, and let H cover G with the map $f : V(H) \to V(G)$, then $h \circ f$ is an eigenfunction of H with the same eigenvalue. Thus if H lifts G, we can talk about its *old* eigenvalues, i.e. those that were inherited from G, and its *new* eigenvalues, which are the rest of the eigenvalues.

Definition 13.10. The universal covering space of a graph G is a graph that lifts all the lifts of G.

Example 13.11. The infinite d-regular tree is the universal covering space of every d-regular graph.

Grienberg and Lubotzky gave the following definition which extends the definition of Ramanujan graphs to general (non-regular) graphs.

Definition 13.12. We say that a graph G is Ramanujan if the absolute value of every eigenvalue of G except λ_0 is at most the spectral radius [1] of the universal covering space of G.

Conjecture 13.13. *For every graphs G, if we lift it high enough then almost surely we will get a Ramanujan graph.*

Lubotzky and Nagnibeda falsified this conjecture by constructing an infinite tree T, that covers infinite number of finite graphs such that none of them is Ramanujan.

Friedman showed that the construction of Lubotzky and Nagnibeda, in fact, constructs a single graph G^* that is covered by all the graphs that T covers and has large second eigenvalue (i.e. larger than the spectral radius of T). Since all the other graphs inherit the eigenvalues of G^*, none of them can be Ramanujan. He then rephrased Conjecture 13.13 as follows,

Conjecture 13.14. *For every graph G, if we lift it high enough to a graph H, then the* new *eigenvalues of H will almost surely be at most the spectral radius of the universal covering of H.*

To support his conjecture, Friedman proved the following theorem.

Theorem 13.15. *Let G be a graph with a largest eigenvalue λ_0, and let ρ be the spectral radius of its universal covering space. Then in almost every (high enough) lift of G, every new eigenvalue μ satisfies,*

$$\mu \leq \sqrt{\lambda_0 \rho} + o(1)$$

This is a generalization of the following result by Broder and Shamir.

Theorem 13.16. *For almost every d-regular graph, the second eigenvalue is at most $O(d^{\frac{3}{4}})$.*

To see that Theorem 13.15 generalizes 13.16, note that for d-regular graphs, $\lambda_0 = d$ and $\rho = \sqrt{d}$.

[1] We refer the reader to lecture 5. There we defined the spectrum $\sigma(A_T)$ of the adjacency matrix of an infinite tree T. The spectral radius is the maximal (absolute) value in $\sigma(A_T)$.

Chapter 14

Some Eigenvalue Theorems

Notes taken by Yonatan Bilu

Summary: The last lecture in the course surveys several bounds on eigenvalues of symmetric matrices, and in particular those of graphs. Throughout this summary, the eigenvalues of a graph refer to the eigenvalues of its adjacency matrix, and are denoted by $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$.

The class ended in a picnic, where juicy watermelon slices were served on Harry Potter plates, along side Harry Potter napkins, to students wearing silly Harry Potter paper hats.

Theorem 14.1. *For every d and ϵ there is a $c = c(\epsilon, d)$, such that if G is a d-regular graph on n vertices, then the number of its eigenvalues with absolute value greater than $2\sqrt{d-1} - \epsilon$ is at least cn.*

Proof. See "Elementary Number Theory, Group Theory and Ramanujan Graphs", by G. Davidoff, P. Sarnak and A. Vallete. The proof makes use of the notorious Čebyshev Polynomials. □

Theorem 14.2. *(Füredi & Komloš '81) : Let P_1 be a random distribution on $\mathbb{R}$ with expectation μ and variance σ^2, and P_2 a random distribution on $\mathbb{R}$ with expectation ν and variance σ^2 as well. Assume further that both distributions are bounded. Let A be a real, $n \times n$ symmetric matrix, with off-diagonal entries chosen i.i.d. according to P_1, and diagonal entries chosen i.i.d. according to P_2. Then with probability tending to 1 as n tends to infinity the following holds:*

1. $\max_{i \geq 2} |\lambda_i| < 2\sigma n + O(n^{\frac{1}{3}} \log n)$

2. $\lambda_1 \sim N((n-1)\mu + \nu + \frac{\sigma^2}{n}, 2\sigma^2)$[1]

Proof. The proof is derived by looking at moments of increasing order. An alternative approach, by Kahn and Szemeredi, relies on the Rayleigh quotient to understand the behavior of random matrices. □

Theorem 14.3. *(Broder & Shamir '87) : For almost all d-regular graphs, $\lambda_2 = O(d^{\frac{3}{4}})$.*

Proof. Let G be a random $2d$-regular graph on n vertices, generated by choosing d random permutations, $\pi_1, ..., \pi_d$, and defining an edge $(v, \pi_i(v))$ for every $v \in [n]$ and $i \in [d]$. Let P be the transition matrix of the Markov chain defined by the graph, i.e. it's adjacency matrix divided by $2d$. Let $\mu_1 \geq \mu_2 \geq ... \geq \mu_n$ be its eigenvalues, and $\rho = \max\{|\mu_2|, |\mu_n|\}$. Since $\mu_1 = 1$, and for any k, $\{\mu_i^k\}_{i=1}^n$ are the eigenvalues of P^k, we have $\rho^{2k} \leq tr(P^{2k}) - 1$, and therefore:

$$\mathbb{E}(\rho) \leq (\mathbb{E}(\rho^{2k}))^{\frac{1}{2k}} \leq (\mathbb{E}(tr(P^{2k})) - 1)^{\frac{1}{2k}}, \tag{14.1}$$

where the inequality on the left follows from Jansen's inequality.

Let v be a vertex of G. A path in the graph that starts at v is uniquely defined be a word S over the alphabet $\{\pi_1, \pi_1^{-1}, ..., \pi_d, \pi_d^{-1}\}$ (these permutations label the edges of the path). Let us generate the graph by going along a

[1] $N(\mu, \sigma^2)$ refers to the Normal Distribution with expectation μ and variance σ^2.

random path, and choosing $\pi_i(u)$ uniformly among the currently allowed values. In other words, say we are currently at a vertex u. We choose uniformly at random $i \in [d]$ and $\epsilon \in \{-1, 1\}$. If $\pi_i^\epsilon(u)$ is already defined, we move to that vertex. We call this a "forced move". Otherwise, we choose $\pi_i^\epsilon(u)$ uniformly at random among the values not yet taken by π_i^ϵ. We call this a "free move".

Define $red(S)$ to be the reduction of S, that is, what remains of S if we repeatedly remove all two consecutive letters in it of the form π_i, π_i^{-1}.

Since we want to bound $tr(P^{2k})$, we would like to bound the probability that S induces a path that starts and ends in v. We divide this event into three:

1. $red(S)$ is the identity.
2. The path defined by S has exactly one (simple) loop.
3. The path defined by S has at least two loops.

We bound each of these in the following three lemmas:

Lemma 14.4. *Let S be a word of length $2k$ generated as above, then:*

$$\Pr[red(S) = \emptyset] \leq (\frac{2}{d})^k$$

Proof. The idea is to count closed paths in the infinite $2d$-regular tree, using Catalan numbers. □

Lemma 14.5. *Let S' be a word of length $2k$ generated as above, and let $S = red(S')$. Denote $s = |S|$, and assume $s > 1$, then:*

$$\Pr[S \textit{ start at 1, has exactly one loop, and ends in 1}] \leq \frac{1}{n} + O(\frac{s}{n^2})$$

Proof. Assume w.l.o.g. that S is a loop (otherwise, we argue for the loop, which is even shorter than s), and denote the vertices it visits $1 = v_0, ..., v_s = 1$. Since all these vertices are distinct, the choice made at v_{s-1} is free. Therefore, the probability that at this point vertex 1 is chosen is $\frac{1}{n-s} = \frac{1}{n} + O(\frac{s}{n^2})$ □

Lemma 14.6. *Let S be a word of length $2k$ generated as above,then:*

$$\Pr[S \textit{ has two loops}] \leq O(\frac{k^4}{n^2})$$

Proof. For S to have two loops there have to be two "free choices" where we choose a vertex that was already visited before. The probability of choosing a vertex that was already chosen before, is at most $\frac{2k}{n}$. The probability of this happening at two specific steps, is at most $\frac{4k^2}{n^2}$. By the union bound, the probability that it happens at some two steps is $O(\frac{k^4}{n^2})$. □

Note that Lemma 14.5 is not exactly what we need, since it deals with a reduced word. Broder and Shamir also bound the probability that when $|S| = 2k$, $|red(S)| = s$. With this, they show that the probability of the second case is bounded by $\frac{1}{n} + O(\frac{k2^k}{nd^k})$. This, together with the other two bounds yields:

$$\Pr[S \text{ starts at v and returns to v}] \leq (\frac{2}{d})^k + \frac{1}{n} + O(\frac{k2^k}{nd^k} + \frac{k^4}{n^2}) \tag{14.2}$$

Finally, we need to choose the k that minimizes the RHS, so we roughly need $(\frac{2}{d})^k = \frac{k^4}{n^2}$, or $k \sim (2 - o(1)) \log_{d/2} n$. Putting this back in 14.1, we get $\mathbb{E}(\rho) \leq (\frac{2}{d})^{\frac{1}{4}}(1 + o(1))$. The proof is finished by showing that ρ is concentrated around its mean by using martingales. □

Note: Currently, the best result is by Joel Friedman, who showed $\lambda_2 \leq \sqrt{2d-1} + \epsilon$, for all $\epsilon > 0$.

Note: The random model of Broder & Shamir, that of choosing $\frac{d}{2}$ random matchings, does not induce a uniform distribution on d-regular graphs. However, this model is *contiguous* to the uniform model, that is, any graph property that occurs in one of them w.h.p., occurs w.h.p. in the other as well. For details see chapter 9 in "Random Graphs" by S. Janson, T. Luczak and A. Ruciński.
Another source of reference is Nick Wormald's survey, "Models of random regular graphs", which appears in "Surveys in Combinatorics", 1999, J.D. Lamb and D.A. Preece, eds.

Bibliography

[CRSW02] M. Capalbo, O. Reingold, S. Vadhan and A. Wigderson, *Randomness Conductors and Constant-Degree Lossless Expanders*, Extended Abstract, STOC 2002. *Randomness Conductors and Constant-Degree Expansion Beyond the Degree/2 Barrier*.

[CW65] J. W. Cooley and Tukey J. W. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 90(19):297–301, 1965.

[Die97] R. Diestel. *Graph Theory*. Springer-Verlag, 1997.

[KMS94] D. Karger, R. Motwani and M. Sudan, *Approximate graph coloring by semidefinite programming* Proc. 35th IEEE Symposium on Foundations of Computer Science, pp. 2–13 (1994)

[Rab80] M. O. Rabin. Probabilistic algorithm for testing primality. *Journal of Number Theory*, 12(128-138), 1980.

[RVW02] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of Mathematics*, 155(1):157–187, 2002.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27, July and October 1948.

[Val76] L. G. Valiant. Graph-theoretic properties in computational complexity. *JCSS*, 13(3):278–285, 1976.

[Vla02] Vlad Ciubotariu, *The Perfect Graph Theorem*, University of Waterloo student project, (April 2002) from http://www.student.math.uwaterloo.ca/ cs762/Projects/index.php