Lectures on expanders

Marcin Kotowski, Michał Kotowski

April 28, 2012

Contents

1 Lecture 1: the basics

1.1 Graphs, eigenvalues and expansion

We assume that all graphs are unoriented simple graphs, i.e. have no self-loops or multiple edges (although it is easy to extend all the definitions to cover the more general situation). The size of a graph G, denoted $|G|$, is the number of vertices in G. We will be working mainly with *d-regular* graphs, in which each vertex has the same degree d. Usually d will be a fixed constant while the number of vertices in the graph goes to infinity.

Let $G = (V, E)$ be an unoriented simple graph. The *adjacency matrix* of G, denoted by A_G or simply A, has entries defined by: $A_{ij} = 1$ if $(i, j) \in E$ and $A_{ij} = 0$ otherwise, where $i, j \in V$.

Another closely related matrix is the *Laplacian* of G, denoted as Δ_G or Δ, defined as:

$$\Delta = I - D^{-1}A$$

where D is the degree matrix, i.e. diagonal matrix with $d_{ii} = \deg(i)$. In the case of d-regular graphs, this reduces to:

$$\Delta = I - \frac{1}{d}A$$

For irregular graphs the Laplacian is often more convenient to work with than the adjacency matrix.

We will think of vectors on which A acts as functions on the vertex set $f : V \to \mathbb{R}$. The space of such functions has dimension $|G|$ and is endowed with a natural inner product:

$$\langle f, g \rangle = \sum_{v \in V} f(v)g(v)$$

and the corresponding norm:

$$\|f\|^2 = \sum_{v \in V} |f(v)|^2$$

We will denote this space by $\ell^2(G)$. For a subset of vertices $S \subseteq V$, its characteristic function will be denoted by $\mathbb{1}_S$.

The adjacency matrix replaces the value of f in a vertex v with the sum of values over all neighbors of v:

$$(Af)(v) = \sum_{\substack{w \in V \\ w \sim v}} f(w)$$

An important feature of A is that it is symmetric with respect to the inner product on $\ell^2(G)$. For any two functions $f, g \in \ell^2(G)$ we have:

$$\langle Af, g \rangle = \sum_{v \in V} (Af)(v)g(v) = \sum_{v \in V} \sum_{\substack{w \in V \\ w \sim v}} f(w)g(v) = \sum_{(v,w) \in E} f(w)g(v) = \langle f, Ag \rangle$$

since edges in G are unoriented.

Since A is a symmetric matrix, it is diagonalizable, with (possibly repeated) real eigenvalues:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Our goal for the next two lectures will be building connections between the spectral properties of G, expressed by eigenvalues and eigenvectors of A, and its geometric and combinatorial properties.

First recall that we can always choose eigenvectors of A so that they are orthogonal in $\ell^2(G)$. The eigenvalues of symmetric matrices can be characterized by the following lemma:

Lemma 1.1. *For a symmetric matrix A we have:*

$$\lambda_1 = \max_{f \in \ell^2(G)} \frac{\langle Af, f \rangle}{\|f\|^2}$$

More generally, the k-th largest eigenvalue is characterized by:

$$\lambda_k = \max_{\substack{f \in \ell^2(G) \\ f \perp f_1, \ldots, f_k}} \frac{\langle Af, f \rangle}{\|f\|^2}$$

where the sum is over functions f orthogonal to the first k eigenfunctions $f_1 \ldots, f_k$.

We start with the following bounds on the eigenvalues of A:

Remark 1.1. For any d-regular graph G we have:

$$d \geq \lambda_1 \geq \ldots \geq \lambda_n \geq -d$$

Proof. For any $f \in \ell^2(G)$, we have:

$$|\langle Af, f \rangle| \leq \sum_{\substack{v,w \\ w \sim v}} |f(w)f(v)| \leq \frac{1}{2} \sum_{\substack{v,w \\ w \sim v}} \left(|f(w)|^2 + |f(v)|^2 \right) = d \sum_v |f(v)|^2 = d \|f\|^2$$

so all eigenvalues of A lie in the interval $[-d, d]$. □

It is straightforward to write down at least one eigenfunction of A - the constant function $\mathbb{1}_G$ has eigenvalue d. Equivalently, its Laplacian eigenvalue is 0 and this holds for all graphs, not necessarily regular (although for an irregular graph $\mathbb{1}_G$ usually will not be an eigenfunction of the adjacency matrix). Functions satisfying $\Delta f = 0$ are called *harmonic functions*.

A simple observation links λ_1 to connectednes of G:

Remark 1.2. For any d-regular graph G, $\lambda_1 = d$ and its multiplicity equals the number of connected components of G. In particular, G is connected if and only if λ_1 has multiplicity one.

Proof. As remarked above, $\lambda_1 = d$, as $A\mathbb{1}_G = d\mathbb{1}_G$. If G has k connected components $G_1, \ldots, G_k$, characteristic functions $\mathbb{1}_{G_i}$ are also orthogonal eigenfunctions with eigenvalue d, so multiplicity of d is at least k. Finally, we notice that on any connected component, the only functions with eigenvalue $\lambda_1 = d$ are constants. Each such function is harmonic and on a connected graph, every harmonic function is constant, which follows from the maximum principle for Δ (Δ cannot have strict local maxima). □

From now on, we assume that G is connected. Another easy observation links the smallest eigenvalue λ_n and bipartiteness of G:

Remark 1.3. For any d-regular graph G, $\lambda_n = -d$ if and only if G is bipartite.

Proof. If G is bipartite with bipartition $V = V_1 \cup V_2$, then the function $f = \mathbb{1}_{V_1} - \mathbb{1}_{V_2}$ (i.e. equal 1 on one bipartite component and -1 on the other) is an eigenfunction of A with eigenvalue $-d$. Conversely, if f is an eigenfunction with eigenvalue $-d$, $Af = -df$, then it is easy to see $|f|$ has no strict local maxima or minima, so by connectedness $|f| = \text{const}$. Defining $V_+ = \{v \in V : f(v) \geq 0\}$, $V_- = \{v \in V : f(v) < 0\}$ gives us the desired bipartition, since it's readily checked that $Af = -df$ precludes any edges between V_+ and V_-. □

Since G is connected, $d = \lambda_1 > \lambda_2$. Because of the variational characterization of eigenvalues (Lemma 1.1), the second largest eigenvalue λ_2 is the maximum of $\langle Af, f\rangle$ over all functions with norm 1 which are orthogonal to $\mathbb{1}_G$. The condition $\langle f, \mathbb{1}_G\rangle = 0$ implies that f has mean zero:

$$\langle f, \mathbb{1}_G\rangle = \sum_{v \in V} f(v) = 0$$

The quantity $d - \lambda_2$ is called the (1-sided) *spectral gap* of the graph. In some applications it is necessary to consider also the 2-sided spectral gap, defined as $d - \lambda$, where $\lambda = \max\{\lambda_2, |\lambda_n|\}$.

Intuitively, the spectral gap $\sigma = d - \lambda_2$ will quantify the connectivity of G - if σ is small, G can be disconnected by removing a small number of edges (indeed, as we have seen above, $\sigma = 0$ means that G is already disconnected). On the other hand, large σ will imply that G has high connectivity, so to disconnect it we must remove a large number of edges.

To make this idea precise, we introduce the notion of edge expansion of a graph. For a set $S \subset V$, denote by ∂S its *edge boundary*, defined as $\partial S = \{(u, v) \in E : u \in S, v \notin S\}$. This is the number of edges which have one endpoint in S and the other one outside of S.

Definition 1.2. *For a set $S \subseteq V$ of size $|S| \leq \frac{|V|}{2}$, its edge expansion is defined by:*

$$h(S) = \frac{|\partial S|}{|S|}$$

Edge expansion of a graph G is defined as:

$$h = \min_{\substack{A \subseteq V \\ |A| \leq \frac{|V|}{2}}} h(A)$$

If for any two sets S and T we denote the number of edges having one endpoint in S and the other one in T by $E(S, T)$, then we have the following useful expression for this quantity:

$$E(S, T) = \langle A\mathbb{1}_S, \mathbb{1}_T\rangle$$

In particular $|\partial S| = E(S, \overline{S}) = \langle A\mathbb{1}_S, \mathbb{1}_{\overline{S}}\rangle$.

A set S with high expansion $h(S)$ should be thought of as having large "perimeter" (expressed by its edge boundary ∂S) relative to its size, so to disconnect S from the rest of the graph one has to remove many edges. If a graph G has high edge expansion, it is difficult to cut it into two parts, since all sets (at least those smaller than $\frac{|V|}{2}$) have large boundaries.

Before we investigate connections between spectral and geometric properties of graphs more systematically, we study a few examples that confirm the intuition that large spectral gap should be qualitatively equivalent to large edge expansion.

Example 1.3. *For a complete graph K_n, we have:*

$$\lambda_1 = n-1,\ \lambda_2 = \lambda_3 = \ldots = \lambda_n = -1$$
$$h = 2\left(1 - \frac{1}{n}\right)$$

Thus, K_n enjoys a very large (2-sided) spectral gap $\sigma = n$ and is of course very difficult to disconnect. Note, however, that this comes at a price of having $\Omega(n^2)$ edges.

Exercise 1.1. *For a cycle C_n, we have:*

$$\lambda_k = 2\cos\left((k-1)\frac{2\pi}{n}\right), k = 1, \ldots, n$$
$$h = \frac{4}{n}$$

The fact that C_n can be disconnected by removing just 2 edges, so its edge expansion is $h = \Theta\left(\frac{1}{n}\right)$, is reflected by small spectral gap $\sigma = 2 - 2\cos\left(\frac{2\pi}{n}\right) = \Theta\left(\frac{1}{n^2}\right)$.

Exercise 1.2. *For a complete bipartite graph $K_{n,n}$, we have:*

$$\lambda_1 = n,\ \lambda_2 = \lambda_3 = \ldots = \lambda_{2n-1} = 0,\ \lambda_{2n} = -n$$
$$h = \frac{n}{2}$$

Exercise 1.3. *For a d-regular graph G consider its complement G^c, having the same vertex set and exactly those edges which are not in G. This is an $(n-d-1)$-regular graph. Show that:*

$$\lambda_i(G^c) = -1 - \lambda_{n+2-i}(G)$$

for $2 \leq i \leq n$.

Exercise 1.4. *The n-dimensional hypercube H_n has $\{0,1\}^n$ as its vertex set and two points $x, y \in \{0,1\}^n$ are connected by an edge if they differ in exactly one coordinate. Calculate the eigenvalues $\lambda_i(H_n)$ and the corresponding eigenvectors.*

1.2 Expanders and their properties

At least qualitatively, spectral gap is tightly linked to having high connectivity. We have seen that clique K_n enjoys very good spectral and isoperimetric properties - however, in many applications it is crucial to achieve such properties while having far fewer than $\Omega(n^2)$ edges. Ideally, we would like our graph to be sparse, having constant degree, which implies

having $O(n)$ edges. This seems to conflict with high connectivity, as we can expect that a graph with few edges would be very easy to cut into disjoint parts. Existence of sparse graphs with good spectral properties is highly nontrivial - such miraculous graphs motivate the most important definition in this course:

Definition 1.4. *A family of graphs $\{G_n\}_{n=1}^{\infty}$ with $|G_n| \to \infty$ is a family of 1-sided (resp. 2-sided) (d, ε)-expanders for some fixed $d, \varepsilon > 0$ if it satisfies the following conditions:*

- *all G_n are d-regular*
- *for all G_n we have $\lambda_2 \leq (1-\varepsilon)d$ (resp. $\max\{\lambda_2, |\lambda_n|\} \leq (1-\varepsilon)d$)*

We will often say that a single graph G is an expander, implicitly treating G as a member of some family of (d, ε)-expanders.

As mentioned before, requiring that the degree d be constant implies sparsity, which makes the task of constructing expanders considerably difficult.

How large can the spectral gap of a graph be? The following very simple lower bound shows that it cannot be arbitrarily large:

Exercise 1.5. *Let G be a d-regular graph. Prove that:*

(a) $\mathrm{Tr}A = \sum_{i=1}^{n} \lambda_i = 0$

(b) $\mathrm{Tr}A^2 = \sum_{i=1}^{n} \lambda_i^2 = nd$

(c) $\max\{|\lambda_2|, |\lambda_n|\} = \sqrt{d} - o(1)$

(d) give a combinatorial interpretation of $\mathrm{Tr}A^3$

This puts an asymptotic upper bound on the quality of expansion afforded by any d-regular graph. It turns out that the bound derived above is not optimal - we will return to the tight bound, called the Alon-Boppana bound, in the next lecture.

Remark 1.4. The technique of estimating the spectral gap by considering the trace of powers of A, illustrated in the simplest case in Exercise 1.5, is very useful and is known as the *trace method*. Often it is possible to evaluate or approximate $\mathrm{Tr}A^k$ by combinatorial techniques (for example, counting paths of given length) and thus extract information about the eigenvalues. It is also used extensively in random matrix theory.

The intuition that spectral gap is related to edge expansion is made precise by the following important inequality:

Theorem 1.5 (Cheeger's inequality)**.** *Let $\lambda = \lambda_2$ be the second largest eigenvalue of A_G. The spectral gap and edge expansion of G satisfy the following inequalities:*

$$\frac{d-\lambda}{2} \leq h \leq \sqrt{2d(d-\lambda)} \tag{1}$$

One direction of the inequality (large spectral gap implies good expansion) is easy to prove, but the other one (good expansion implies large spectral gap) is more difficult, as calculating λ_2 requires controlling $\langle Af, f\rangle$ for all functions f and having good expansion enables us to control this quantity essentially only for f which are indicators. We will prove this inequality in the next lecture.

It is interesting to note that this kind of inequality first appeared in the context of Riemannian manifolds, where one can also give appropriate definitions of the isoperimetric constant and the Laplacian ([HLW06, Section 4.4]). In this sense graphs can be sometimes thought of as discrete approximations of manifolds.

Apart from enjoying good isoperimetric properties expander graphs have other desirable features. One of them is that an expander looks roughly like a random graph with the same number of edges. If we pick the edges of a graph at random, taking each edge with probability $\frac{d}{n}$, then the average number of edges between any two sets S, T will be $\frac{d}{n}|S||T|$, while in the expander graph it is equal to $E(S,T)$. The Expander Mixing Lemma says that these two quantities are close to each other:

Exercise 1.6 (Expander Mixing Lemma)**.** *Let G be a d-regular graph with $|V| = n$ and $\lambda(G) = \lambda$. Then for any two subsets $S, T \subseteq V$ we have:*

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

Exercise 1.7. *Let G with n vertices be a two-sided ε-expander. Show that the size of any independent set in G is at most $(1-\varepsilon)n$ and that G has chromatic number at least $\frac{1}{1-\varepsilon}$.*

For any graph G its *diameter*, denoted $\operatorname{diam} G$, is the maximum of distance over all pairs of vertices, where the distance $d(u,v)$ between two vertices u, v is the length of the shortest path between them. Let $B(v,r) = \{u : d(u,v) \leq r\}$ denote the ball with center v and radius r. The following exercise shows that expanders have small diameter, so all vertices are close to each other:

Exercise 1.8. *Show that if G is an expander with n vertices, then balls in G grow exponentially - for every vertex v we have $|B(v,r)| \geq \min\{K^r, n\}$, where $K > 1$ is a constant independent of n. Conclude that* $\operatorname{diam} G = O(\log n)$.

Exercise 1.9. *Give an example of a family of graphs which have logarithmic diameter, but are not expanders.*

We now turn to discuss random walks on expanders. Random walks on graphs is a vast and fascinating topic, here we will only consider the property that walks on expanders have good *mixing time* - if we start from any vertex and perform a random walk, after not too many steps the probability distribution on the vertices will be close to uniform. This property is important from the point of view of applications.

Take a graph G and pick some starting vertex $v_0 \in V$. A *simple random walk* on G is a sequence of random variables $v_0, v_1, \ldots$ such that at each step the next vertex v_{k+1} is chosen

uniformly at random among neighbors of the currrent vertex v_k. For each $k \geq 0$ this defines a probability distribution on the vertex set:

$$\mu^{(k)}(v) = \mathbb{P}(v_k = v)$$

which we will call the distribution after k steps of the random walk.

If we treat each $\mu^{(k)}$ simply as a nonnegative function on G such that $\sum_{v \in V} \mu^{(k)}(v) = 1$, then it is easy to see that each step of the random walk is governed by the *transition matrix* $M = \frac{1}{d}A$:

$$\mu^{(k+1)} = M\mu^{(k)}$$

with $\mu^{(0)} = \delta_{v_0}$.

Because $\mathbb{1}_G$ is an eigenfunction of A with eigenvalue d, the uniform distribution $\pi = \frac{1}{|V|}\mathbb{1}_G$ on the vertex set satisfies $M\pi = \pi$, so, in the language of Markov chains, it is a *stationary distribution*. Unless G is bipartite, for any initial starting distribution (not necessarily a single vertex δ_{v_0}) the distribution in k steps will converge (in a suitable norm) to the uniform distribution as $k \to \infty$, and the speed of that convergence depends on the spectral gap of A.

There are various ways of quantifying how close to each other two probability distributions are - in many contexts a natural measure is the *total variation distance*:

$$\|\pi_1 - \pi_2\|_{TV} = \sup_{S \subseteq V} \left| \sum_{v \in S} \pi_1(v) - \sum_{v \in S} \pi_2(v) \right|$$

which can be expressed in terms of the ℓ^1 norm:

$$\|\pi_1 - \pi_2\|_{TV} = \frac{1}{2} \|\pi_1 - \pi_2\|_1$$

Proposition 1.6. *If A has two-sided spectral gap σ and $|V| = n$, then:*

$$\left\|\mu^{(k)} - \pi\right\|_1 \leq \sqrt{n} \left(\frac{\sigma}{d}\right)^k$$

Proof. Decompose δ_{v_0} as $\delta_{v_0} = \frac{1}{n}\mathbb{1}_G + f = \pi + f$, where $f = \delta_{v_0} - \frac{1}{n}\mathbb{1}_G$ is orthogonal to $\mathbb{1}_G$. Because $\mu^{(k)} = M^k \delta_{v_0}$, by Cauchy-Schwartz we have:

$$\left\|\mu^{(k)} - \pi\right\|_1 \leq \sqrt{n} \left\|M^k(\pi + f) - \pi\right\|_2 = \sqrt{n} \left\|M^k\pi + M^k f - \pi\right\|_2 =$$
$$= \sqrt{n} \left\|M^k f\right\|_2 \leq \sqrt{n} \left(\frac{\sigma}{d}\right)^k \|f\|_2 \leq \sqrt{n} \left(\frac{\sigma}{d}\right)^k$$

where w ehave used that $M^k\pi = \pi$ and A has spectral gap σ, so $\left\|A^k f\right\|_2 \leq \sigma^k \|f\|_2$. □

The proof works of course for any initial probability distribution.

This bound in particular implies that for a fixed $\varepsilon > 0$ after roughly $k \approx \log n$ we will have $\left\|\mu^{(k)} - \pi\right\|_1 \leq \varepsilon$. Compare this for example with the cycle C_n, for which $k \approx n^2$ steps

are needed to get the same error ε ([LPW09, Chapter 7]). For more about mixing times of random walks see [LPW09].

Note that in the bound above having only one-sided spectral gap is clearly insufficient, as on a bipartite graph the random walk will alternate between two sets of the bipartition, so it does not converge to the uniform distribution. However, a common method to get a two-sided spectral gap when we have only one-sided gap is to consider a *lazy random walk* instead of the ordinary walk. In the lazy random walk in each step with probability $1/2$ we stay at the current vertex and with probability $1/2$ we move as in the ordinary random walk. It is clear that the transition matrix for this walk is $M_L = \frac{1}{2}M + \frac{1}{2}I$ and it has a two-sided (although smaller) spectral gap if M had a one-sided gap.

1.3 Random graphs are expanders

So far we have proved several interesting properties of expanders, but it is not a priori clear that such families of graphs exist at all. Explicit constructions of expanders will be the topic of the next lectures - here we show by using the probabilistic method that families of expanders exist, although the proof is non-constructive.

We will show that a randomly chosen graph has expansion bounded away from 0 with nonzero probability (the bounds we use are very crude - actually the probability of getting an expander is very close to 1). We use the following model of a random d-regular graph - take a vertex set V of size n, with n even, and choose uniformly at random d perfect matchings on all $n/2$ pairs of vertices. The union of these matchings defines the edge set E of G. Note that the resulting graph may have multiple edges, but this is only a technicality and it is possible to fix the model to remove them, although for the sake of simplicity we will not do this here.

Theorem 1.7. *There exist d and $\eta > 0$ such that for any n a random d-regular graph G chosen from the model above has edge expansion greater than η with nonzero probability.*

Proof. Take any set $S \subseteq V$ of size $k \leq \frac{n}{2}$. We will show that with high probability we have $|N(S)| > \eta|S|$, where $N(S) = \{v : (u, v) \in E \,\text{for some}\, u \in S\}$ and $\eta > 1$. Because $|\partial S| \geq |N(S) \backslash S|$, this will imply the bound on expansion.

We want to bound from above the probability that there exists some $T \subseteq V$ with at most ηk vertices such that $N(S) \subseteq T$, since this would contradict high expansion. Pick any subset T with at most ηk vertices. Suppose we are choosing one of the matchings in G and let X_i denote the event that the i-th vertex from S is matched with some vertex in T.

Exercise 1.10. *Prove that:*

$$\mathbb{P}(X_1 \cap \ldots \cap X_{\lceil k/2 \rceil}) \leq \left(\frac{\eta k}{n}\right)^{k/2}$$

Exercise 1.11. *Show that for a given $1 \leq k \leq \frac{n}{2}$ the probability that sets S and T as above exist is bounded from above by 2^{-k} for sufficiently large (but constant) degree d. Conclude that the probability that such sets exist for any k is smaller than 1, so with nonzero probability we have high expansion.*

□

As we will see in the next lecture, a random regular graph is not only an expander with high probability, but an almost optimal expander, at least in terms of spectral gap. A lot more can be said about spectral properties of random graphs, but we will not pursue this direction here.

2 Lecture 2: Cheeger inequality and Alon-Boppana bound

2.1 Cheeger inequality

In this lecture, we will prove Cheeger's inequality, which gives quantitative relation between spectral gap and edge expansion.

Let $f : V \to \mathbb{R}$ be a function on vertices of G. Orient the edges of G arbitrarily and denote the set of oriented edges as E_+. Let $\nabla f : E_+ \to \mathbb{R}$ denote the gradient of f, given by:

$$\nabla f(e) = f(e_+) - f(e_-),\ e = (e_+, e_-)$$

For $f, g : E_+ \to \mathbb{R}$ their inner product is:

$$\langle f, g \rangle = \sum_{e \in E_+} f(e)g(e)$$

Lemma 2.1. *For f as above we have:*

$$\|\nabla f\|^2 = d \langle f, \Delta f \rangle$$

Proof.

$$\langle f, \Delta f \rangle = \sum_{v \in V} f(v)(\Delta f)(v) = \sum_{v \in V} f(v) \left(f(v) - \frac{1}{d} \sum_{\substack{u \in V \\ (v,u) \in E}} f(u) \right) =$$

$$\frac{1}{d} \sum_{(v,u) \in E} f(v) \left(f(v) - f(u) \right) = \frac{1}{d} \sum_{(v,u) \in E_+} \left(f(v) - f(u) \right)^2 = \frac{1}{d} \|\nabla f\|^2$$

□

As an aside, with the notation above having spectral grap σ can be expressed in terms of *Poincaré inequality* (which may be familiar from analysis) - for any function f orthogonal to $\mathbb{1}_G$ we have:

$$\|f\|^2 \leq \frac{1}{\lambda} \|\nabla f\|^2$$

Inequalities of this kind show up in other places, for example in study of infinite graphs or metric embeddings, although we won't go in any details here (see [Pet]).

We now move to the main course of this lecture.

Theorem 2.2 (Cheeger's inequality)**.** *Let $\lambda = \lambda_2$ be the second largest eigenvalue of A_G. The spectral gap and edge expansion of G satisfy the following inequalities:*

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

Proof. (a) $\frac{\lambda}{2} \leq h$:

This is the easy direction. For any set $S \subseteq V$, $|S| \leq \frac{|V|}{2}$ we have:

$$E(S,\bar{S}) = \langle \mathbb{1}_S, A\mathbb{1}_{\bar{S}} \rangle = \langle \mathbb{1}_S, A\mathbb{1}_G \rangle - \langle \mathbb{1}_S, A\mathbb{1}_S \rangle = d|S| - \langle \mathbb{1}_S, A\mathbb{1}_S \rangle$$

so:

$$\frac{E(S,\bar{S})}{|S|} = d - \frac{\langle \mathbb{1}_S, A\mathbb{1}_S \rangle}{\langle \mathbb{1}_S, \mathbb{1}_S \rangle} \tag{3}$$

On the other hand, $\mathbb{1}_S$ can be decomposed as a sum of its mean and the component orthogonal to constants:

$$\mathbb{1}_S = \frac{|S|}{|V|}\mathbb{1}_G + \left(\mathbb{1}_S - \frac{|S|}{|V|}\mathbb{1}_G\right)$$

which gives:

$$\langle \mathbb{1}_S, A\mathbb{1}_S \rangle = \left\langle \frac{|S|}{|V|}\mathbb{1}_G + \left(\mathbb{1}_S - \frac{|S|}{|V|}\mathbb{1}_G\right), A\left(\frac{|S|}{|V|}\mathbb{1}_G + (\mathbb{1}_S - \frac{|S|}{|V|}\mathbb{1}_G)\right)\right\rangle =$$
$$d\frac{|S|^2}{|V|} + \left\langle \mathbb{1}_S - \frac{|S|}{|V|}\mathbb{1}_G, A\left(\mathbb{1}_S - \frac{|S|}{|V|}\mathbb{1}_G\right)\right\rangle \leq d\frac{|S|^2}{|V|} + \lambda\left\langle \mathbb{1}_S - \frac{|S|}{|V|}\mathbb{1}_G, \mathbb{1}_S - \frac{|S|}{|V|}\mathbb{1}_G\right\rangle =$$
$$(d-\lambda)\frac{|S|^2}{|V|} + \lambda|S| \leq \frac{1}{2}(d+\lambda)|S|$$

which combined with 3 gives:

$$h = \min_{\substack{S\subseteq V \\ |S|\leq\frac{|V|}{2}}} \frac{E(S,\bar{S})}{|S|} \geq \frac{d-\lambda}{2}$$

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

This is the harder direction. To bound expansion from above, we have to find a sparse cut, i.e. a set S for which $\frac{E(S,\bar{S})}{|S|}$ is small. The idea is to use the eigenvector f with eigenvalue λ_2 to approximate the optimal cut.

It will be easier to work with nonnegative functions, so decompose f into its positive and negative part, $f = f_+ - f_-$. First, note that since $f_+ \geq 0$, for all $v \in \text{supp} f_+$ we have:

$$(\Delta f_+)(v) \leq (\Delta f)(v)$$

and $\Delta f = \frac{1}{d}(d-\lambda)f$, so from Lemma 2.1:

$$\|\nabla f_+\|^2 = d\,\langle f_+, \Delta f_+\rangle \leq d\,\langle f_+, \Delta f\rangle = (d-\lambda)\,\langle f_+, f\rangle = (d-\lambda)\,\|f_+\|^2 \tag{4}$$

We will use f_+ to construct a sparse cut. Pretend for a while that f takes only two values - since f has mean zero, one must be positive and the other negative. Taking $S = \mathrm{supp} f_+$ gives the cut with $h(S) \leq d(d-\lambda)$ (in general the loss of square root is unavoidable, see Remark 2.1). In this model case f_+ is simply a characteristic function $\mathbb{1}_A$, so $\|\nabla f\|$ is proportional to $|\partial A|$ and we can bound h by $\frac{\|\nabla f_+\|^2}{\|f_+\|^2}$ immediately.

This is not true in general, but we can decompose f_+ into its *level sets*. To this end order vertices of $V = \{1, 2, \ldots, n\}$ so that $f_+(1) \geq f_+(2) \geq \ldots \geq f_+(n)$ and let S_i be the i-th level set of f_+, $S_i = \{1, \ldots, i\}$. Consider *threshold cuts* of the form $S = \{i \,:\, f_+(i) \geq t\}$ for $t \in [f_+(n), f_+(1)]$. We will show that if t is chosen at random from an appropriate probability distribution, then we have:

$$\frac{\mathbb{E}\left(E(S, \overline{S})\right)}{\mathbb{E}\min\{|S|, |\overline{S}|\}} \leq \sqrt{2d(d-\lambda)}$$

where $\mathbb{E}$ denotes the expectation with respect to a random choice of t. This can be rewritten as:

$$\mathbb{E}\left(E(S, \overline{S}) - \sqrt{2d(d-\lambda)}\min\{|S|, |\overline{S}|\}\right) \leq 0$$

from which it follows that there exists at least one t such that:

$$\frac{E(S, \overline{S})}{\min\{|S|, |\overline{S}|\}} \leq \sqrt{2d(d-\lambda)}$$

and this proves what we want.

The trick is to choose the right probability distribution for t. A natural choice would be to take t with uniform distribution in $[f_+(n), f_+(1)]$, but it turns out that this doesn't quite work. Instead we pick t with density proportional to $t\,dt$, so that $\mathbb{P}(b \geq t \geq a) \sim b^2 - a^2$. The normalization constant will cancel out in the end, so for simplicity we assume it is equal to 1.

Of course it is enough to consider values $t = f_+(i)$, $i = 1, \ldots, n$. We can without loss of generality assume that $\mathrm{supp} f_+ \leq \frac{n}{2}$ (otherwise take $-f$ and f_- instead of f and f_+), so that all the sums below are in fact from $i = 1$ to $\lfloor \frac{n}{2} \rfloor$.

First we have:

$$\mathbb{E}\left(E(S,\overline{S})\right) = \sum_{k=1}^{n-1} \mathbb{P}\left(f_+(k) \geq t > f_+(k+1)\right) E(S_k, \overline{S}_k) =$$

$$\sum_{k=1}^{n-1} \left(f_+^2(k) - f_+^2(k+1)\right) E(S_k, \overline{S}_k) = \sum_{k=1}^{n-1} \left(f_+^2(k) - f_+^2(k+1)\right) \sum_{(i,j)\in E} \delta_{\{i\leq k\leq j-1\}} =$$

$$\sum_{\substack{(i,j)\in E \\ i<j}} \sum_{k=i}^{j-1} \left(f_+^2(k) - f_+^2(k+1)\right) = \sum_{\substack{(i,j)\in E \\ i<j}} \left(f_+^2(i) - f_+^2(j)\right) = \sum_{(i,j)\in E_+} \left|f_+^2(i) - f_+^2(j)\right|$$

and:

$$\mathbb{E}\min\{|S|, |\overline{S}|\} = \sum_{k=1}^{n-1} \mathbb{P}\left(f_+(k) \geq t > f_+(k+1)\right) \min\{|S_k|, |\overline{S}_k|\} =$$

$$\sum_{i=1}^{n-1} \left(f_+^2(k) - f_+^2(k+1)\right) \min\{|S_k|, |\overline{S}_k|\} = \sum_{k=1}^{n-1} \left(f_+^2(k) - f_+^2(k+1)\right) k = \sum_{k=1}^{n} f_+^2(k) = \|f_+\|^2$$

where we have telescoped the sum.

We have the following:

Lemma 2.3. $\mathbb{E}\left(E(S,\overline{S})\right) \leq \sqrt{2d}\,\|\nabla f_+\| \cdot \|f_+\|$

Proof. We simply use Cauchy-Schwartz:

$$\mathbb{E}\left(E(S,\overline{S})\right) = \sum_{(i,j)\in E_+} \left|f_+^2(i) - f_+^2(j)\right| = \sum_{(i,j)\in E_+} |f_+(i) - f_+(j)| \cdot |f_+(i) + f_+(j)| \leq$$

$$\left(\sum_{(i,j)\in E_+} |f_+(i) - f_+(j)|^2\right)^{1/2} \left(\sum_{(i,j)\in E_+} |f_+(i) + f_+(j)|^2\right)^{1/2} =$$

$$\|\nabla f_+\| \cdot \left(\frac{1}{2}\sum_{(i,j)} A_{ij} \left|f_+(i) + f_+(j)\right|^2\right)^{1/2} \leq \|\nabla f_+\| \cdot \left(\sum_{(i,j)} A_{ij} \left(f_+(i)^2 + f_+(j)^2\right)\right)^{1/2} =$$

$$\|\nabla f_+\| \left(2d \sum_i f_+(i)^2\right)^{1/2} = \sqrt{2d}\,\|\nabla f_+\| \cdot \|f_+\|$$

□

Putting everything together, we arrive at:

$$\frac{\mathbb{E}\left(E(S,\overline{S})\right)}{\mathbb{E}\min\{|S|, |\overline{S}|\}} \leq \frac{\sqrt{2d}\,\|\nabla f_+\| \cdot \|f_+\|}{\|f_+\|^2} = \sqrt{2d}\frac{\|\nabla f_+\|}{\|f_+\|} \leq \sqrt{2d} \cdot \sqrt{d-\lambda}$$

by Lemma 2.3 and inequality 4, so we are done.

□

Remark 2.1. In general, Cheeger's inequality is optimal. The easy direction is tight for the hypercube H_n, which has $h = 1$ and $d - \lambda = 2$. The hard direction is tight, up to constants, for the cycle C_n, which has $h = \Theta\left(\frac{1}{n}\right)$ and $d - \lambda = \Theta\left(\frac{1}{n^2}\right)$.

The proof of Cheeger's inequality actually gives a fast algorithm for finding the approximately minimal cut - compute the second largest eigenvector f, sort vertices with respect to $f(v)$ and return the level set S achieving minimal isoperimetric ratio $h(S) = \frac{|\partial S|}{|S|}$. Cheeger's inequality guarantees that such a solution is at most quadratically worse than the optimum, since:

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

Computing the optimal value of minimal cut is NP-hard - compare this with computing the spectral gap, which can be done in polynomial time to arbitrary precision simply by diagonalizing the adjacency matrix A. A closely related problem is computing the *sparsest cut*, defined as:

$$\Phi = \min_{S \subseteq V} \frac{n|\partial S|}{|S||\overline{S}|}$$

Since $\min\{|S|, |\overline{S}|\} \leq \frac{n}{2}$, we have $\Phi \leq 2h$. The spectral partitioning algorithm can be viewed as based on *relaxation* of sparsest cut problem to spectral gap problem, in the sense that:

$$\Phi = \min_{S \subseteq V} \frac{n|\partial S|}{|S||\overline{S}|} = \min_{S \subseteq V} \frac{\sum\limits_{u \sim v} |\mathbb{1}_S(u) - \mathbb{1}_S(v)|^2}{\frac{1}{n} \sum\limits_{u,v \in V} |\mathbb{1}_S(u) - \mathbb{1}_S(v)|^2}$$

while:

$$\sigma = \min_{\substack{f \in \ell^2(G) \\ f \perp \mathbb{1}_G}} \frac{\|\nabla f\|^2}{\|f\|^2} = \min_{f \in \ell^2(G)} \frac{\|\nabla f\|^2}{\left\|f - \bar{f}\mathbb{1}\right\|^2} = \min_{f \in \ell^2(G)} \frac{\sum\limits_{u \sim v} |f(u) - f(v)|^2}{\frac{1}{n} \sum\limits_{u,v \in V} |f(u) - f(v)|^2}$$

where replacing f with $f - \bar{f}\mathbb{1}_G$, for $\bar{f} = \frac{1}{n}\sum_v f(v)$, automatically guarantess the condition $f \perp \mathbb{1}_G$. Thus, σ solves the same optimization problem as Φ, but with relaxed constraints - instead of optimizing only over characteristic functions $\mathbb{1}_S$, we optimize over all functions $f \in \ell^2(G)$. Other partitioning algorithms exist, based on different notions of relaxation (see e.g. [Tre]).

A good question is whether higher eigenvalues λ_k can be given geometric interpretation similar to connection between h and λ_2. Define the *k-way expansion* of a graph:

$$h_k(G) = \min_{S_1, \ldots, S_k \subseteq V} \max\{h(S_i) : i = 1, \ldots, k\}$$

where the minimum is taken over all partitions of V into disjoint subsets $S_1, \ldots, S_k$. In other words, we are interested in partitioning V into k subsets such that each subset defines a sparse cut. Very recently, the following k-way Cheeger inequality has been proven [LGT11]:

$$\frac{d-\lambda_k}{2} \leq h_k(G) \leq O(k^2)\sqrt{d(d-\lambda_k)}$$

Although we have not used it in these lectures, there is another notion of expansion known as *vertex expansion*. It is in general more difficult to handle, but it is instructive to see how it is related to the spectral gap. For a set $S \subseteq V$ define its neighborhood $N(S) = \{v \in V : \exists s \in S\,(s,v) \in E\}$.

Definition 2.4. *For a graph G its vertex expansion $h_V(G)$ is defined as:*

$$h_V = \min_{\substack{S \subseteq V \\ |S| \leq \frac{|V|}{2}}} \frac{|N(S) \backslash S|}{|S|}$$

The following exercise gives the analog of Cheeger inequality for vertex expansion:

Exercise 2.1. *For a graph G let $\lambda = \max\{|\lambda_2(G)|, |\lambda_n(G)|\}$ and let $h_V = h_V(G)$. Show that:*

$$\frac{d^2-\lambda^2}{d^2+\lambda^2} \leq h_V \leq \sqrt{2}d\sqrt{d^2-\lambda^2}$$

2.2 Alon-Boppana bound

In the exercise 1.5 we derived a crude lower bound on the second eigenvalue:

$$\lambda_2(G) \geq \sqrt{d} - o(1) \tag{5}$$

As mentioned there, this inequality is not optimal. We will now proceed to derive the optimal bound, called the Alon-Boppana bound - in the process, we will touch upon fascinating topics such as graph covers, spectral properties of infinite graphs and Ramanajun graphs, although they are unfortunately beyond the scope of these lectures.

Theorem 2.5 (Alon-Boppana bound)**.** *Let G be a d-regular graph with diameter δ. Then:*

$$\lambda(G) \geq 2\sqrt{d-1}\left(1 - O\left(\frac{\log \delta}{\delta}\right)\right) = 2\sqrt{d-1}\,(1-o(1))$$

where $\lambda = \max\{\lambda_2, |\lambda_n|\}$.

Proof. The proof of bound 5 proceeded (essentially) by counting the number of closed paths of length 2, expressed by $\mathrm{Tr}A^2$, and using it to bound λ_2, thus discarding all other eigenvalues. Taking higher powers of A, or closed path of larger length, might help to improve the bound.

Let $s, t \in V$ be two vertices with distance δ from each other. Consider the function $f = \mathbb{1}_s - \mathbb{1}_t$ (i.e. $f(s) = 1, f(t) = -1$ and 0 on all other vertices)- it has mean zero, so we can use it as a test function to lower bound λ:

$$\lambda^{2k}(A) = \lambda(A^{2k}) \geq \frac{\langle f, A^{2k} f \rangle}{\|f\|^2} = \frac{A_{ss}^{2k} + A_{tt}^{2k} - 2A_{st}^{2k}}{2} \tag{6}$$

We can kill the term A_{st}^{2k} by taking $k = \lfloor \frac{\delta - 1}{2} \rfloor$, so there are no paths of length $2k$ from s to t. We are left with estimating A_{ss}^{2k}, A_{tt}^{2k}, which are numbers of closed paths of length $2k$.

Consider the infinite d-regular tree T_d and let t_{2k} denote the number of closed paths in T_d starting at the root. Combinatorial properties of t_{2k} are completely known - including asymptotics, generating functions and recursion equations - but we will be satisfied here with a simple bound. Note that every closed path in T_d determines a *sign pattern* $\pi = (s_1, \ldots, s_{2k})$, where $s_i = \pm 1$ depending on whether the i-th step is toward or away from the root. Each sign pattern has the property that the sum of its every prefix is nonnegative - thus, the number of valid sign patterns is the Catalan number C_k:

$$C_k = \frac{\binom{2k}{k}}{k+1}$$

Every sign pattern determines at least $(d-1)^k$ closed paths, since there are k of $+1$ steps and at each such step there are at least $d-1$ possible choices of moving away from the root.

Now, clearly $A_{ss}^{2k} \geq t_{2k}$, since presence of cycles in G can only increase the number of closed paths. Combining this with 6 and using well known asymptotics of C_k:

$$C_k \sim \Theta\left(\frac{4^k}{k^{3/2}}\right)$$

we get:

$$\lambda^{2k}(G) \geq t_{2k} \geq C_k (d-1)^k = \Theta\left((2\sqrt{d-1})^{2k} \cdot k^{-\frac{3}{2}}\right)$$

which, after taking the $2k$-th root and plugging in $k = \lfloor \frac{\delta - 1}{2} \rfloor$ leads to:

$$\lambda(G) \geq 2\sqrt{d-1} \cdot \left(1 - O\left(\frac{\log \delta}{\delta}\right)\right)$$

□

The gist of the above proof lies in comparing behavior of the random walk on G to the random walk on its *universal cover* T_d, as we are essentially calculating the probability of a simple random walk returning to its starting point after $2k$ steps. This idea can be developed further, but we will need the definition of a graph covering:

Definition 2.6. *Let G be a finite graph and H a possibly infinite graph. A map $\pi : H \to G$ is called a covering map if for every vertex $v \in H$, π maps the set of edges incident to v bijectively onto the set of edges incident to $\pi(v)$.*

For example, a single vertex with d loops to itself is covered by any d-regular graph and every d-regular graph is covered by the infinite d-regular tree T_d (in other words, T_d is the universal cover of d-regular graphs, since it is simply connected).

Exercise 2.2. *Let $\pi : H \to G$ be a covering, with G and H finite. Prove that $\lambda_2(H) \geq \lambda_2(G)$.*

For an infinite graph H, the analogue of λ_2 comes in the form of *spectral radius* $\rho(H)$. Spectral radius can be defined in various ways, the most natural being the spectral norm of A_H treated as a self-adjoint operator on an (infinite-dimensional) space $\ell^2(H)$, but this direction would take us too far into functional analysis. Instead, we give the following definition:

Definition 2.7. *Consider a simple random walk on H. For $x \in H$, let $p_{2n}(x, x)$ be the probability of returning to x after $2n$ steps. Then:*

$$\rho(H) = \lim_{n\to\infty} (p_{2n}(x, x))^{\frac{1}{2n}}$$

Exercise 2.3. *Let $\pi : H \to G$ be a covering, with G and H infinite. Prove that $\rho(G) \geq d\rho(H)$.*

Combined with the fact that every d-regular graph is covered by T_d and that $\rho(T_d) = \frac{2\sqrt{d-1}}{d}$ (essentially contained in estimates for t_{2k}), this gives another way of viewing the Alon-Boppana bound and explains the origin of the factor $2\sqrt{d-1}$.

Graphs which achieve the asymptotic bound for the second eigenvalue, i.e. have $\lambda_2(G) \leq 2\sqrt{d-1}$, are called *Ramanujan graphs*. Constructing explicitly such graphs is a nontrivial task, but can be done using deep connections with number theory and group theory. The best known construction by Lubotzky, Phillips and Sarnak (see [DSV03]), based on Cayley graphs of $PGL_2(\mathbb{F}_q)$ (where $\mathbb{F}_q$ is a finite field), gives a family of Ramanujan graphs for all d such that $d-1$ is a prime number (this can be extended to $d-1$ being a prime power). These graphs also enjoy the property of having high girth and high chromatic number, which solves a long-standing problem in combinatorics (previously the only known examples of such graphs were random graphs).

Although we will not cover spectral theory of random graphs here, it is interesting that in a model of d-regular random graphs similar to the one defined in Section 1.3, almost all graphs are almost Ramanujan, in the sense that for any $\varepsilon > 0$ we have:

$$\mathbb{P}\left(\lambda_2(G) \leq 2\sqrt{d-1} + \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0$$

where G is a random d-regular graphs with size going to infinity. This has been proved only fairly recently by Friedman and it is a difficult result.

For very good treatment of spectral properties of infinite graphs and connections to probability and geometric group theory, see [Pet].

Exercise 2.4. *Prove a bound similar to the Alon-Boppana bound for bipartite (k, l)-regular graphs, i.e. graphs where each left vertex has degree k and each right vertex has degree l.*

3 Lecture 3: algebraic constructions

3.1 Cayley graphs

Having discussed basic properties of expanders in previous two lectures, we turn to the question of giving explicit constructions of expander families. It turns out that a rich source of such constructions comes from group theory, specifically the study of Cayley graphs of finite groups. The highly symmetric nature of such graphs makes them amenable to precise analysis, using tools from harmonic analysis and representation theory. In fact, the very first constructions of expanders were based on group theory. In this lecture, we present a fairly elementary construction of expander family due to Margulis. There is a vast array of other group-theoretic techniques for producing expanders (even skimming through them would take at least one semester course), including:

- representation theory and Kazhdan property (T) (see [Tao], Notes 2, for good exposition)
- expanders from explicit generating sets for finite simple groups, e.g. symmetric and alternating groups S_n, A_n ([Kas07])
- expansion in Lie type groups, e.g. $SL_n(\mathbb{F}_p)$ ([Tao])

Definition 3.1. *Let G be a group generated by a finite set S. The Cayley graph $Cay(G,S) = (V,E)$ is a (directed) graph whose vertex set V is G and edge $(g,g') \in E$ if and only if $g = g's$ for some $s \in S$.*

Note that $Cay(G,S)$ is always $|S|$-regular. We will always assume that $e \notin S$ (so there are no loops) and that S is symmetric, $s \in S \Leftrightarrow s^{-1} \in S$. With this assumption, $Cay(G,S)$ can be viewed as an undirected graph since $(g,g') \in E \Leftrightarrow (g',g) \in E$. The graph depends on the generating S and constructing an expander out of G will often require a judicious choice of S.

A related more general notion is that of a Schreier graph:

Definition 3.2. *Let G be a group which is generated by a finite set S and acts on a set X. Assume that S acts freely on X, so for all $s,s' \in S, s \neq s'$ and $x \in X$ we have $s \cdot x \neq s \cdot s'$ and $s \cdot x \neq x$. The Schreier graph $Sch(X,S) = (V,E)$ is a graph whose vertex set V is X and edge $(x,y) \in E$ if and only if $x = s \cdot y$ for some $s \in S$.*

Example 3.3. *The cycle C_n is the Cayley graph $Cay(\mathbb{Z}_n, S)$ for the standard generating set $S = \{-1,1\}$. As we have seen before, this graph is not an expander.*

In general, obtaining a family of constant-degree expanders out of Cayley graphs requires turning to non-Abelian groups:

Exercise 3.1. *Show that if G_n is a sequence of Abelian groups, then $Cay(G_n, S_n)$ does not form a family of expanders.*

3.2 Fourier analysis on Abelian groups

The expander construction which we will present in this lecture relies heavily on Fourier analysis, so we start with introducing the key notions of harmonic analysis on finite Abelian groups (the whole theory can be developed in a much more general setting of locally compact Abelian groups).

Let G be an Abelian group. For functions $f, g : G \to \mathbb{C}$, we introduce the inner product:

$$\langle f, g \rangle = \sum_{f \in G} f(h)\overline{g(h)}$$

and the associated norm:

$$\|f\|_2^2 = \langle f, f \rangle$$

Definition 3.4. *Let G be an Abelian group. A character χ is a homomorphism $\chi : G \to \mathbb{C}^*$, i.e. a function satisfying $\chi(x+y) = \chi(x)\chi(y)$ for every $x, y \in G$.*

In the simplest case $G = \mathbb{Z}_n$ it is straightforward to check all characters are given up to normalization by:

$$\chi_a(k) = \omega^{ak}$$

where $a \in \mathbb{Z}_n$ and $\omega = e^{\frac{2\pi i}{n}}$ is the n-th root of unity.

The important property of characters is that they are orthogonal as functions in $\ell^2(G)$:

Exercise 3.2. *Prove that if G is a finite Abelian group then:*

(a) if χ is a character which is not identically 1, then $\sum_a \chi(a) = 0$

(b) if χ_1, χ_2 are two different characters, then $\langle \chi_1, \chi_2 \rangle = 0$

From the point of view of spectral graph theory, knowing characters of G enables us to find eigenvectors and eigencalues of any Cayley graph of G:

Exercise 3.3. *Let $Cay(G, S)$ be a Cayley graph of a finite Abelian group G. Show that for any character χ the function $f : G \to \mathbb{C}$ defined by $f(a) = \chi(a)$ is an eigenvector of the adjacency matrix A_G with eigenvalue $\sum_{s \in S} \chi(s)$.*

From the fact that characters are orthogonal it follows that there are at most $|G|$ of them, since this is the dimension of $\ell^2(G)$. It is easy, using the fact that every finite Abelian group is of the form $\mathbb{Z}_{n_1}^{k_1} \times \ldots \times \mathbb{Z}_{n_m}^{k_m}$, to write them down explicitly and check that there are in fact exactly $|G|$ of them. Therefore characters form (after proper normalization) an orthonormal basis of $\ell^2(G)$.

Example 3.5. *Since characters of $\mathbb{Z}_n$ are of the form $\chi(k) = \omega^{ak}$, $a \in \mathbb{Z}_n$, we immediately see that the eigenvalues of the cycle C_n are given by $\omega^a + \omega^{-a} = 2\cos(\frac{2\pi a}{n})$, as we found in Exercise 1.1.*

The hypercube H_n is the Cayley graph of $\mathbb{Z}_2^n$ with the standard generating set. Characters of $\mathbb{Z}_2^n$ are given by:

$$\chi_a)(x) = (-1)^{\langle a,x\rangle}$$

where $a = (a_1, \ldots, a_n) \in \mathbb{Z}_2^n$, $x = (x_1, \ldots, x_n) \in \mathbb{Z}_2^n$) and $\langle a, x\rangle = a_1x_1 + \ldots + a_nx_n$. From this we find that the $n - 2|a|$ (where $|a|$ is the number of ones in a) is an eigenvalue of the hypercube with multiplicity $\binom{n}{|a|}$, as in Exercise 1.4.

We now turn to the case of $G = \mathbb{Z}_n^2$, which we will use later, although all the properties discussed below hold in the general setting.

Normalized characters of $\mathbb{Z}_n^2$ are given by:

$$\chi_{(a,b)}(x, y) = \frac{1}{n}\omega^{ax+by}$$

where $(a, b) \in \mathbb{Z}_n^2$.

Because $\chi_{(a,b)}$ form an orthonormal set, every function $f \in \ell^2(\mathbb{Z}_n^2)$ can be decomposed as a sum of characters:

$$f = \sum_{(a,b)\in\mathbb{Z}_n^2} \widehat{f}(a, b)\chi_{(a,b)}$$

where the coefficients $\widehat{f}(a, b)$ are called the *Fourier coefficients* of f. They define $\widehat{f} : \mathbb{Z}_n^2 \to \mathbb{C}$, called the *Fourier transform* of f. Since $\chi_{(a,b)}$ form an orthonormal basis, $\widehat{f}$ can be computed explicitly:

$$\widehat{f}(a, b) = \left\langle f, \chi_{(a,b)}\right\rangle = \sum_{(x,y)\in\mathbb{Z}_n^2} f(x, y)\overline{\chi_{(a,b)}(x, y)} = \frac{1}{n}\sum_{(x,y)\in\mathbb{Z}_n^2} f(x, y)\omega^{-ax-by}$$

The Fourier transform can be viewed as an isometry $\widehat{\ }: \ell^2(\mathbb{Z}_n^2) \to \ell^2(\mathbb{Z}_n^2)$ as it satisfies the *Plancherel identity*:

$$\|f\|_2 = \|\widehat{f}\|_2$$

A crucial property that makes the Fourier transform useful is that it changes translation to multiplication by a phase factor. More precisely, for $h = (h_1, h_2) \in \mathbb{Z}_n^2$ denote by f_h the translation of f, $f_h(x) = f(x + h)$. We then have:

$$\widehat{f_h}(a, b) = \omega^{-ah_1-bh_2}\widehat{f}(a, b)$$

Of particular interest is the zero Fourier coefficient:

$$\widehat{f}(0, 0) = \frac{1}{n}\sum_{(x,y)\in\mathbb{Z}_n^2} f(x, y)$$

which is proportional to the mean of f. If f has mean zero, we have $\widehat{f}(0, 0) = 0$.

3.3 Margulis expander

We will now provide a relatively simple and explicit construction of expanders due to Margulis. The construction will be based on the natural action of the affine group $SL_2(\mathbb{Z}) \ltimes \mathbb{Z}^2$ on the integer lattice $\mathbb{Z}^2$, quotiented to $\mathbb{Z}_n^2$.

The first ingredient in our construction is examining the action of $SL_2(\mathbb{Z})$ on $\mathbb{Z}^2$. Recall that $SL_2(\mathbb{Z})$ is the group of all 2×2 matrices with integer coefficients and determinant 1. Consider the following elements $a, b \in SL_2(\mathbb{Z})$:

$$a = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \; b = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

and sets $A, B \subseteq \mathbb{Z}^2$ defined as $A = \{(x, y) : |x| < |y|\}, B = \{(x, y) : |x| > |y|\}$. We have the following easily proved proposition (which can be used, with help of the *ping-pong lemma*, to show that a and b generate a free subgroup of $SL_2(\mathbb{Z})$):

Proposition 3.6. *For any $n > 0$ we have $a^n A \subseteq B, \; b^n B \subseteq A$.*

The existence of this ,,ping-pong decomposition" is crucial in the following lemma, which says that every ,,almost invariant" probability measure must have a large point mass in $\{0\}$:

Lemma 3.7. *Choose a finite generating set S for $SL_2(\mathbb{Z})$. Suppose that μ is a probability measure on $\mathbb{Z}^2$ such that for every $s \in S$:*

$$\|s_* \mu - \mu\|_{TV} \leq \varepsilon$$

where $s_* \mu$ denotes the pushforward measure $(s_* \mu)(A) = \mu(s^{-1} A)$. Then:

$$\mu(\{0\}) = 1 - O(\varepsilon)$$

Proof. Take a, b, A, B as in Proposition 3.6. Since $aA \subseteq B$, we have:

$$\mu(B) \geq \mu(aA) = \mu(A) + O(\varepsilon)$$

and likewise:

$$\mu(A) \geq \mu(bB) = \mu(B) + O(\varepsilon)$$

so:

$$\mu(B) = \mu(aA) + O(\varepsilon)$$

In the same vein, we can obtain:

$$\mu(B) = \mu(a^2 A) + O(\varepsilon)$$

Now, $a^2 A \subseteq B$, so $\mu(B \backslash a^2 A) \leq O(\varepsilon)$ and $C \subseteq B \backslash a^2 A$, where $C = \{(x, y) : |y| < |x| < |3y|\}$. It follows that $\mu(C) \leq O(\varepsilon)$. We can translate C by a fixed number of elements in $SL_2(\mathbb{Z})$ to cover all $\mathbb{Z}^2$ but 0, from which it follows that $\mu(\mathbb{Z}^2 \backslash \{0\}) = O(\varepsilon)$. $\square$

We are now ready to prove the main theorem of this lecture. Recall that $SL_2(\mathbb{Z}) \ltimes \mathbb{Z}^2$ is the affine group, i.e. set of all affine transformations $f : \mathbb{Z}^2 \to \mathbb{Z}^2$ given by:

$$f(x) = Ax + b$$

where $A \in SL_2(\mathbb{Z}),\ b \in \mathbb{Z}^2$. In analogous fashion, we can reduce the coefficients modulo n and consider $SL_2(\mathbb{Z}_n) \ltimes \mathbb{Z}_n^2$ acting on $\mathbb{Z}_n^2$ in a natural manner. Since $SL_2(\mathbb{Z}_n) \ltimes \mathbb{Z}_n^2$ acts on $\mathbb{Z}_n^2$, we also have a natural action of $SL_2(\mathbb{Z}_n) \ltimes \mathbb{Z}_n^2$ on $\ell^2(\mathbb{Z}_n^2)$, which we will denote by $\rho_n(g)$, given by:

$$(\rho_n(g)f)(x) = f(g^{-1} \cdot x)$$

Theorem 3.8. *Choose any symmetric finite generating set S for $SL_2(\mathbb{Z}) \ltimes \mathbb{Z}^2$. Let $\pi_n : SL_2(\mathbb{Z}) \ltimes \mathbb{Z}^2 \to SL_2(\mathbb{Z}_n) \ltimes \mathbb{Z}_n^2$ be the natural projection map. The family of Schreier graphs $G_n = Sch(SL_2(\mathbb{Z}_n) \ltimes \mathbb{Z}_n^2, \pi_n(S))$ forms a family of one-sided expanders.*

Proof. The proof will proceed as follows. By contradiction, we assume that G_n are not expanders, so we can find a sequence of "almost invariant" functions f_n of mean zero. Using Fourier transform, we turn each f_n into an "almost invariant" measure μ, which, by Lemma 3.7, will have most of its mass concentrated on $\{0\}$. On the other hand, f_n has mean zero and properties of Fourier transform will imply that $\mu(\{0\}) = 0$, a contradiction.

Suppose by contradiction that $G_n = Sch(SL_2(\mathbb{Z}_n) \ltimes \mathbb{Z}_n^2, \pi_n(S))$ is not a family of expanders. By passing to a subsequence, we can assume that $\lambda(G_n) = o(1)$. It then follows that if f_n is the second largest eigenvector associated to G_n, we have for all $s \in S$:

$$\|\rho_n(s)f_n - f_n\|_{\ell^2(\mathbb{Z}_n^2)} = o(1) \tag{7}$$

In particular, if we take $e_i, i = 1, 2$ to be translations along each coordinate, we have:

$$\|\rho_n(e_i)f_n - f_n\|_{\ell^2(\mathbb{Z}_n^2)} = o(1)$$

Now, take the Fourier transform of the left side and recall that Fourier transforming changes translation to multiplication by phase, by Plancherel identity we get:

$$\left\|(e^{-\frac{2\pi a i}{n}} - 1)\widehat{f_n}\right\|_{\ell^2(\mathbb{Z}_n^2)} = o(1) \tag{8}$$

This implies that $\widehat{f_n}$ must have almost all of its mass concentrated on a ball B_n of radius $o(n)$. More precisely:

$$\left\|\widehat{f_n}\right\|_{B_n} = 1 - o(1)$$

Indeed, note that:

$$\left\|(e^{-\frac{2\pi a i}{n}} - 1)\widehat{f_n}\right\|_{\ell^2(\mathbb{Z}_n^2)} = \left\|2\sin\frac{\pi a}{n}\widehat{f_n}\right\|_{\ell^2(\mathbb{Z}_n^2)}$$

so if $\widehat{f_n}$ had a constant fraction of mass on some set $A = \{(a, b)\} \subseteq \mathbb{Z}_n^2$ where, $|a| > cn$ or $|b| > cn$ for some constant $c > 0$, we would have $\left|\sin\frac{\pi a}{n}\right| \geq C$ on A, so A gives a constant contribution to $\left\|2\sin\frac{\pi a}{n}\widehat{f_n}\right\|_{\ell^2(\mathbb{Z}_n^2)}$, which contradicts 8.

We can embed $\mathbb{Z}_n^2$ into $\mathbb{Z}^2$, so every f_n is also a function on $\mathbb{Z}^2$. We would like 7 to hold also if f_n is treated as a function on $\mathbb{Z}^2$, i.e.:

$$\|\rho_n(s)f_n - f_n\|_{\ell^2(\mathbb{Z}^2)} = o(1) \tag{9}$$

In principle, passing from $\mathbb{Z}_n^2$ to $\mathbb{Z}^2$ could increase this norm (think of $f = \mathbb{1}_A$, which translated on $\mathbb{Z}^2$ and rounded modulo n has larger overlap with itself than without the rounding). However, f_n has almost all of its mass concentrated on a ball B_n of radius $o(n)$, so, for sufficiently large n, this cannot happen.

Let g_n equal $\widehat{f_n}$ restricted to B_n, now treated as a function on $\mathbb{Z}^2$. From 9, it follows that after Fourier transform we get:

$$\|g_n \circ s^* - g_n\|_{\ell^2(\mathbb{Z}^2)} = o(1) \tag{10}$$

Consider a measure on $\mathbb{Z}^2$ with "density" $|g_n|^2$:

$$\mu(A) := \int_A |g_n|^2$$

Now, 10 implies that $\|s_*\mu - \mu\|_{TV} = o(1)$, so using Cauchy-Schwartz:

$$\|s_*\mu - \mu\|_{TV} = \sup_{A\subseteq\mathbb{Z}^2} |s_*\mu(A) - \mu(A)| = \sup_{A\subseteq\mathbb{Z}^2} \left| \int_{s^{-1}A} |g_n|^2 - \int_A |g_n|^2 \right| =$$

$$\sup_{A\subseteq\mathbb{Z}^2} \left| \int_A \left(|g_n \circ s^*|^2 - |g_n|^2 \right) \right| \leq \sup_{A\subseteq\mathbb{Z}^2} \left(\int_A \left(|g_n \circ s^*| - |g_n| \right)^2 \right)^{\frac{1}{2}} \left(\int_A \left(|g_n \circ s^*| + |g_n| \right)^2 \right)^{\frac{1}{2}} \leq$$

$$\left(\int_{\mathbb{Z}^2} (g_n \circ s^* - g_n)^2 \right)^{\frac{1}{2}} \left(2 \int_{\mathbb{Z}^2} \left(|g_n \circ s^*|^2 + |g_n|^2 \right) \right)^{\frac{1}{2}} \leq$$

$$\sqrt{2}\, \|g_n \circ s^* - g_n\|_{\ell^2(\mathbb{Z}^2)} = o(1)$$

By Lemma 3.7, this implies that $\mu(\{0\}) = 1 - o(1)$, so $g_n(0) = 1 - o(1)$. But recall that f_n has mean zero, so $\widehat{f_n}(0) = 0$ - from this it follows that $g_n(0) = 0$, which gives a contradiction $\square$

Although the proof we have presented does not give an estimate of the spectral gap, it can be obtained using similar Fourier-analytic techniques [HLW06, Section 8].

4 Lecture 4: zig-zag product

In the previous lecture we saw that expanders can be explicitly constructed by algebraic means. The drawback of this approach is that spectral properties of Cayley or Schreier graphs used there depend on rather intricate details of their algebraic structure and quality of the expander obtained is not easy to analyze. A more direct combinatorial construction would be desirable. Finding such a construction has been a major breakthrough in the field

([RVW02]) and can be achieved by a simple iterative procedure known as the zig-zag product of graphs.

The construction relies on a few combinatorial operations on graphs, used to build larger graphs from smaller graphs while preserving good expansion. The first operation is the *graph product*.

Definition 4.1. *Let $G = (V, E)$ be a graph with n vertices and degree d. The k-th power of G, denoted G^k, is a graph whose vertex set is V and two vertices u, v are connected by an edge if they are connected by a path of length k in G. G^k has degree d^k and if A is the adjacency matrix of G, then A^k is the adjacency matrix of G^k.*

The new graph G^k may have multiple edges or loops, but this will not be a problem in the construction that follows.

Taking the product of a graph with itself improves the expansion, as the eigenvalues are raised to some power, but at the cost of blowing up the degree. We will need an operation which enables us to keep good expansion and small degree while enlarging the graph.

This is the *replacement product*. It takes two graphs G and H, with G having n vertices and degree m and H having m vertices and degree d and, and returns a new graph $G \textcircled{r} H$ with nD vertices and degree $d+1$. We will think of G as being a "large" graph and of H as being "small". The replacement product takes G and replaces its every vertex by a copy of H. For a vertex $v \in G$ we will call the copy of H corresponding to v the *cloud* of v. Because G has the same degree as the number of vertices in H, we will connect the vertices in the clouds so that the new graph will have degree d.

This is done in the following way. For any vertex $v \in G$ order its neighbors in an arbitrary way - we will denote them by $e_v^1, \ldots, e_v^m$ (here we treat each unoriented edge as a pair of oriented edges, so that u as a neighbor of v may have a different number than v as a neighbor of u). Let $\{(v, i) : i = 1, \ldots, m\}$ denote the cloud of v. Two vertices (u, i) and (v, j) are connected in the product if v is the i-th neighbor of u and u is the j-th neighbor of v or, in other words, $u = e_v^j$ and $v = e_u^i$. Also, vertices (v, i) and (v, j) are connected if i and j are connected in H.

Definition 4.2. *For an m-regular graph $G = (V_G, E_G)$ with $|V_G| = n$ and a d-regular graph $H = (V_H, E_H)$ with $|V_H| = m$, the replacement product $G \textcircled{r} H$ is a $d+1$-regular graph with a vertex set $V_G \times V_H$ and the set of edges E defined in the following way (given an ordering of neighbors of each vertex as above):*

- *for all $v \in G$ we have $(v, i) \sim (v, j)$ in $G \textcircled{r} H$ whenever $(i, j) \in E_H$*
- *if $u = e_v^j$ and $v = e_u^i$, then $(u, i) \sim (v, j)$ in $G \textcircled{r} H$*

Figure 1, showing the replacement product of a clique $G = K_5$ with a cycle $H = C_4$, perhaps makes the definition clearer.

The last operation is the *zig-zag product*, which will be the most important ingredient in the construction. The simplest way to describe is to by using the replacement product $G \textcircled{r} H$ - the zig-zag product $G \textcircled{z} H$ has the same vertex set and two vertices (u, i), (v, j)

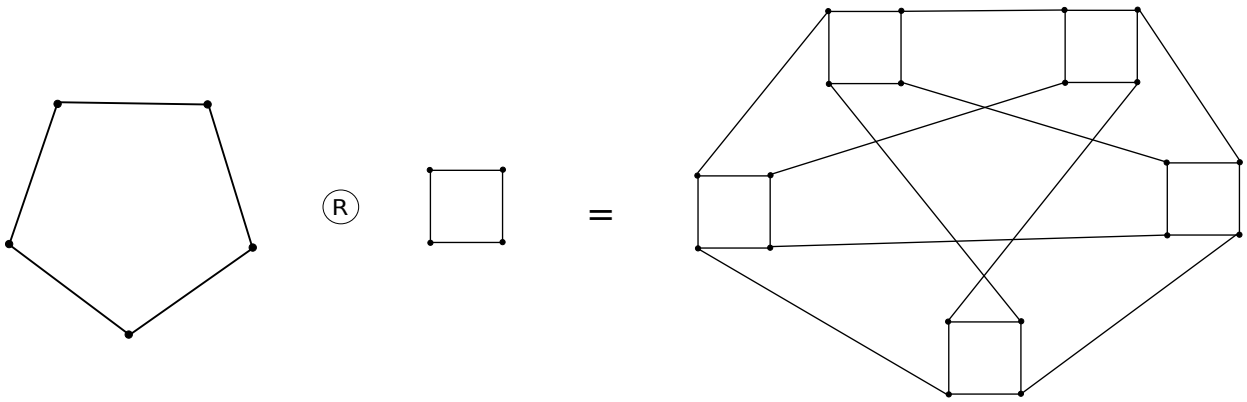

Figure 1: Replacement product of a clique K_5 with a 4-cycle C_4

are connected by an edge if there is a path of length 3 in $G \textcircled{r} H$ such that the first edge connects (u, i) to a vertex in the same cloud, the second edge leads from the cloud of u to the cloud of v and the third edge leads to (v, j). This justifies the name of this operation - the first and the third step are "cloud-type" edges (defined only by edges of H) and the second step is an "inter-cloud" step (defined by the edges of G). Note that the middle step is uniquely determined, as for given (v, l) there is exactly one (u, k) such that $e_u^k = e_v^l$.

Definition 4.3. *For an m-regular graph $G = (V_G, E_G)$ with $|V_G| = n$ and a d-regular graph $H = (V_H, E_H)$ with $|V_H| = m$, the zig-zag product $G \textcircled{z} H$ is a d^2-regular graph with a vertex set $V_G \times V_H$ and the following set of edges E: there is an edge $(u, i) \sim (v, j)$ if there exist some $k, l \in \{1, \ldots, m\}$ such that $(j, k), (i, l) \in E_H$ and $e_u^k = e_v^l$.*

The crucial property of the zig-zag product is that it preserves the degree of the graph and enlarges it by a constant factor while making the expansion not much worse than in the original graph.

For the sake of brevity we introduce a bit of notation - if G has n vertices, degree d and its second largest eigenvalue in absolute value is at most αd, then we will call G an (n, d, α)-graph.

The following lemma is the most important part of the construction:

Lemma 4.4. *If G is an (n, m, α)-graph and H is an (m, d, β)-graph, then $G \textcircled{z} H$ is an $(nm, d^2, \alpha + \beta + \beta^2)$-graph.*

With this lemma constructing a family of expanders is easy. We start with a fixed-size base graph H with reasonably good expansion and then iteratively at each step take the zig-zag product of a power of the previous graph with H. The degree is kept constant at each step, while the number of vertices grows exponentially and the spectral gap is preserved.

Theorem 4.5. *Let H be a $(d^4, d, 1/5)$-graph for some constant d. If we define a family of graphs G_n by:*

$$G_1 = H^2, \quad G_{n+1} = G_n^2 \textcircled{z} H$$

then G_n is a $(d^{4n}, d^2, 1/2)$-graph. In particular, G_n form a family of expanders.

Proof. We proceed by induction. The claim is true for $n = 1$ by definition of H. Now suppose that G_n is a $(d^{4n}, d^2, 1/2)$-graph. Then G_n^2 is a $(d^{4n}, d^4, 1/4)$-graph and it has the same degree as the number of vertices of H, so we can take the zig-zag product. By Lemma 4.4 $G_n^2 ⓩ H$ is a $(d^{4n+1}, d^2, 1/2)$-graph, since the new expansion parameter is at most $1/4 + 1/5 + (1/5)^2 \leq 1/2$. □

Since the base graph needed in the construction has constant size, one can find such a graph by exhaustive search, since we know that a randomly chosen graph with high probability will have good expansion. Another option is to use a simple deterministic algebraic construction which has good enough parameters:

Exercise 4.1. *Let $\mathbb{F}_p$ be a finite field with p elements, where p is a prime, and let $t < p$. Consider the set $\mathbb{F}_p^{t+1}$ and for each $x \in \mathbb{F}_p$, $a, b \in \mathbb{F}_p$ introduce edges $(x, x + b)$, $(x, x + ab)$, ..., $(x, x + a^t b)$. Show that the resulting graph is a $(p^{t+1}, p^2, \frac{t}{p})$-graph.*

Thus to satisfy the assumptions on the base graph in Theorem 4.5 it is enough to take, for example, $p = 37$, $t = 7$.

There are several improvements to this construction which give graphs with smaller degree, faster growing size and more succinct representation in terms of memory needed to store the graph (some of these modifications are particularly important from the point of view of applications in complexity theory, see [HLW06, 9.5]).

It remains to prove Lemma 4.4.

Proof of Lemma 4.4. Throughout the proof it will be convenient to work with transition matrices, i.e. adjacency matrices divided by the degree of the graph. Denote by A the transition matrix of A and by B the transition matrix of H.

By the way the zig-zag product is defined each step of a random walk in $G ⓩ H$ can be decomposed into three steps: a random step in the cloud of the initial vertex, a deterministic step to another cloud and again a random step in the new cloud. Thus if we denote the transition matrix of $G ⓩ H$ by M, then we have $M = \widetilde{B} P \widetilde{B}$, where $\widetilde{B}$ is a block matrix consisting of $|G|$ blocks equal to B (this corresponds to the walk on each cloud) and P is a permutation matrix defined by:

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

Let $\mathbb{1}_M$ denote the constant vector on $G ⓩ H$. We want to show that the second largest eigenvalue of M is at most $\alpha + \beta + \beta^2$ or in other words:

$$\frac{|\langle f, Mf \rangle|}{\|f\|^2} \leq \alpha + \beta + \beta^2$$

for every $f \perp \mathbb{1}_M$.

For any f we define $\overline{f}$ as its average over each cloud:

$$\overline{f}(v,i) = \frac{1}{m}\sum_{j=1}^{m} f(v,j)$$

so that $\overline{f}$ is constant on each cloud. Let $f' = f - \overline{f}$. By definition f' sums up to zero on each cloud. We have:

$$|\langle f, Mf\rangle| = |\langle f, \widetilde{B}P\widetilde{B}f\rangle| \leq |\langle \overline{f}, \widetilde{B}P\widetilde{B}\overline{f}\rangle| + 2|\langle \overline{f}, \widetilde{B}P\widetilde{B}f'\rangle| + |\langle f', \widetilde{B}P\widetilde{B}f'\rangle|$$

Because $\overline{f}$ is constant on each cloud, we have $\widetilde{B}\overline{f} = \overline{f}$. Similarly, because f' is orthogonal to constant on each cloud, we have $\|\widetilde{B}f'\| \leq \beta\|f'\|$ by definition of the spectral gap for H. Thus:

$$|\langle f, Mf\rangle| \leq |\langle \overline{f}, P\overline{f}\rangle| + 2|\langle \overline{f}, P\widetilde{B}f'\rangle| + |\langle \widetilde{B}f', P\widetilde{B}f'\rangle| \leq |\langle \overline{f}, P\overline{f}\rangle| + 2\beta\|\overline{f}\| \cdot \|f'\| + \beta^2\|f'\|^2$$

It remains to estimate the first term. Let $g(v) = \sqrt{m} \cdot \overline{f}(v,i)$ be a function on G. We have $\|g\| = \|\overline{f}\|$. Now note that $\langle \overline{f}, P\overline{f}\rangle = \langle g, Ag\rangle$, because P essentially encodes the information about edges in G. We assumed that $f \perp \mathbb{1}_M$, so $\overline{f} \perp \mathbb{1}_M$ and $g \perp \mathbb{1}_G$. Therefore $|\langle g, Ag\rangle| \leq \alpha\|g\|^2$ by definition of the spectral gap for G. This implies $|\langle \overline{f}, P\overline{f}\rangle| \leq \alpha\|\overline{f}\|^2$, so:

$$|\langle f, Mf\rangle| \leq \alpha\|\overline{f}\|^2 + 2\beta\|\overline{f}\| \cdot \|f'\| + \beta^2\|f'\|^2$$

Since $\overline{f}$ and f' are orthogonal, so $\|f\|^2 = \|\overline{f}\|^2 + \|f'\|^2$, it is straightforward to see that the right hand side of the expression above is at most $\alpha + \beta + \beta^2$ if $\|f\| = 1$, so we are done. □

5 Lecture 5: Selected applications

5.1 Error reduction in randomized algorithms

Suppose that we want to solve an algorithmic problem, e.g. deciding if an input graph satisfies some property P, efficiently (in polynomial time). Quite often, we can come up with an efficient randomized algorithm A that, on input x, samples a random bit string $r \in \{0,1\}^k$ and deterministically computes the answer $A(x,r)$. Usually, there will be some nonzero probability that the algorithm gives the wrong answer. For a moment let us deal with one-sided error, i.e. if the input x has the property we are considering, A always outputs "yes", regardless of the random string r, but if x does not have the property, A outputs the wrong answer with some probability $\beta < 1$. We are interested in reducing the probability of error to arbitrarily small number using as few random bits as possible.

One obvious approach is to run the algorithm t times, each time with a new random string, and output "no" if at least one of the runs answered "no". The total probability of failure will be β^t, which can be made arbitrarily small by choosing t large enough. However, this approach requires $O(tk)$ random bits, k for each of the runs. In computational settings

when random bits are expensive, we would like to achieve similar exponential error reduction using much fewer random bits.

To achieve such a reduction, we will use a suitable expander graph. Let G be a $(2^k, d, \alpha)$-expander with vertex set $V = \{0,1\}^k$ and α such that $\beta + \alpha < 1$ (note that this necessarily puts a lower bound on d). Now, consider the following algorithm on input x:

1. pick a uniformly random starting vertex $v_0 \in V$
2. starting from v_0, performs t steps of a random walk $(v_0, v_1, \ldots, v_t)$
3. output $\bigwedge_{i=0}^{t} A(x, v_i)$

If we denote by $B_x \subseteq \{0,1\}^k$ the subset of strings r such that $A(x,r)$ gives wrong answer, it is clear that the above algorithm is correct on input x if at least one vertex v_i visited by the random walk avoids B_x. To estimate the probability of this event, we use the following theorem (see [HLW06, 3.6] for the - not difficult - proof):

Theorem 5.1. *Let G be an (n, d, α)-graph and $B \subseteq V, |B| = \beta n$ for some $\beta > 0$. The probability that a t step random walk starting from a uniformly random vertex v is confined to B is bounded from above by:*

$$\mathbb{P}(\forall i \ v_i \in B) \leq (\beta + \alpha)^t$$

Since the original probability of error was β, we see that $|B_x| = \beta 2^k$ and applying the theorem gives us exponentially small error probability $\leq (\beta + \alpha)^t$. This approach uses only $k + t \log d = k + O(t)$ random bits (k for choosing the initial vertex and $\log d$ for sampling neighbors during each step of the walk). Note that in order to obtain an efficient algorithm, the expander used must be efficiently constructible, which can be achieved using an explicitly given expander arising e.g. from the zig-zag construction.

The case of two-sided error, when the algorithm can err also on „yes" instances, can be resolved in a similar fashion using majority voting - simply replace step 3 by taking the majority of $A(x, v_i)$. A simple union bound, which can be refined using Chernoff inequality, gives exponential error reduction (see [HLW06]).

5.2 Expander codes

Another application is closely related to the original motivation for considering expander graphs, namely error correcting codes. This is a huge topic in itself, so we only introduce basic definitions and show the construction of *expander codes*.

The setup is as follows - suppose we want to transmit some message $x \in \{0,1\}^k$ via a faulty communication channel. In the simplest model each bit of the input message is flipped independently with probability p, where p is the parameter of the channel. The receiver of the message sees an output string y which in general may be corrupted and we want to encode our message in a possibly larger number of bits so that the probability of incorect decoding is minimized.

An *error correcting code* is a subset $C \subseteq \{0,1\}^n$ (we think of each message as being associated to a unique *codeword* in C). The *minimum distance* of the code is:

$$\Delta = \min\{d(x,y) : x, y \in C\}$$

where $d(x,y)$ is the Hamming distance of two strings x, y. We will also sue the relative distance $\delta = \frac{\Delta}{n}$. Obviously greater minimum distance implies better error-correcting properties, as we may decode each output string to the closest (with respect to Hamming distance) codeword and then at least $\lceil \frac{\Delta}{2} \rceil$ bit flips must occur to cause a decoding error. At the same time we would like to use as short codewords as possible, maximizing the *rate* of the code, defined as $R = \frac{\log |C|}{n}$.

An important class of codes are *linear codes*, for which C is a linear subspace of $\{0,1\}^n$ (where we $\{0,1\}^n$ treat as a linear space over $\mathbb{Z}_2$). Note that for such codes the minimum distance Δ is equal to the lowest possible Hamming weight of a codeword. Each such a code can be represented by its *parity check matrix*, defined as an $m \times n$ matrix A such that $C = \ker A$, where we assume that $\dim C = m$ (this representation is of course not unique).

Two natural questions are:

(a) is it possible to have a family of codes $C_n \subseteq \{0,1\}^n$ with size $n \to \infty$ and $\delta_n \geq \delta_0$, $R_n \geq R_0$ for some constants δ_0, $R_0 > 0$?

(b) if yes, do these codes have efficient encoding and decoding algorithms?

Shannon showed that such asymptotically good codes as in (a) exist, and in fact a randomly chosen code will have good parameters. Unfortunately, this application of the probabilistic method gives no efficient decoding algorithm and it is known that the problem of decoding linear codes is NP-hard in general ([HLW06, Section 12]). Therefore to achieve asymptotically nonvanishing rate and distance and at the same time obtain an efficient decoding algorithm we need to construct the desired code in a more systematic way.

This can be achieved using expander graphs.

With each linear code we can associate a bipartite graph. If the code is specified by an $k \times n$ parity check matrix A, then $G = (V_L \cup V_R, E)$ has the left vertex set with $|V_L| = n$, the right vertex set with $|V_R| = k$ (we assume $k \leq n$) and there is an edge between $v_i \in V_L$ and $w_j \in V_R$ if $A_{ji} = 1$. We assume that each vertex on the left has degree d (having constant degree as $n \to \infty$ will be important for fast decoding).

If to each left vertex (row of A) we associate a variable x_i, then each right vertex (column of A) can be thought of as defining an equation involving x_i's, so that codewords all exactly those vectors of x_i's which satisfy all the equations.

The variant of expansion we need here is the vertex expansion for the left vertices, defined by:

$$h(\alpha) = \min_{\substack{S \subseteq V_L \\ |V_L| \leq \alpha}} \frac{|N(S)|}{|S|}$$

We have $h(\alpha) \leq d$. We will not prove it here, but using the zig-zag product machinery it can be shown that there exist graphs with any $k = \Omega(n)$ and $h(\alpha) > (1-\varepsilon)d$ for $\alpha = \Omega(n)$

([HLW06, Section 12]). Having $k = \Omega(n)$ is necessary to get nonvanishing rate R and now we show that the relative distance is also nonvanishing.

Lemma 5.2. *If $h(\alpha) > \frac{d}{2}$, then $\Delta \geq \alpha$. For $\alpha = \Omega(n)$ this in particular implies that the relative distance δ is asymptotically nonvanishing.*

Proof. First note that for each $S \subseteq V_L$ of size at most α there exists a neighbor $w \in V_R$ such that $|N(w) \cap S| = 1$. This is because there are $d|S|$ edges between S and $N(S)$ and by expansion $|N(S)| > \frac{d}{2}|S|$, so there must be at least one vertex in $N(S)$ with only one neighbor in S.

To show that $\Delta \geq \alpha$, we need to show that every nonzero codeword has Hamming weight at least α. Let x be a codeword and let S be the set of vertices corresponding to coordinates $x_i \neq 0$. If $|S| < \alpha$, then by the neighbor property above there exists some w which has only one neighbor in S. But this implies that the w-th coorodinate of Ax is equal to 1, so x cannot be a codeword. Therefore $|S| \geq \alpha$, which means that $\Delta \geq \alpha$. □

The main advantage of expander codes is that they can be decoded in polynomial time by means of the so-called *belief propagation* algorithm. Consider the following iterative decoding algorithm: given an input string y, if there is any variable which has more unsatisfied equations than satisfied ones, flip its value. Repeat this procedure until all the equations are satisfied and output the resulting string as the decoded codeword.

Lemma 5.3. *If $h(\Delta) > \frac{3}{4}d$ and y is a string whose distance from a codeword x is at most $\frac{\Delta}{2}$, then the algorithm above terminates on input y after a number of steps linear in n and returns x as output.*

Proof. Let $y^{(0)} = y$ denote the input string and let $y^{(j)}$ be the string after the j-th iteration of the algorithm. Let A_j denote the set of incorrect bits in $y^{(j)}$, i.e. $A_j = \{i : y_i^{(j)} \neq x_i\}$. We want to show that A_t is empty for some $t = O(n)$, so that $y^{(t)} = x$.

We first show that if at every step the distance of $y^{(j)}$ from x never exceeds Δ, then the algorithm terminates after a linear number of steps with x as the output. Take $A = A_j$ and divide $N(A)$ into sets S corresponding to satisfied equations and U corresponding to unsatisfied ones. We have $|A| \leq \Delta$, so by expansion:

$$|S| + |U| = |N(A)| > \frac{3}{4}d|A|$$

On the other hand there are at least $|U|$ edges going from U to A and at least $2|S|$ edges from S to A (note that each satisfied vertex must have an even number of neighbors in A). Since there are $d|A|$ edges going out of A, we have:

$$2|S| + |U| \leq d|A|$$

Combining these two inequalities gives:

$$|U| > \frac{d}{2}|A|$$

This implies that there exists some variable with more than $\frac{d}{2}$ unsatisfied neighbors, so it will be flipped (without the assumption $|A| \leq \Delta$ we wouldn't be able to employ expansion and the algorithm might get stuck with no variable to flip). This decreases $|U|$, so after at most $|V_R| = \Theta(n)$ steps we will have $|U| = 0$, which implies $|A| = 0$, so the output string is x.

To show that $|A| \leq \Delta$ for every iteration, note that at each step $|A_j|$ changes by ± 1, so if it exceeds Δ, there must be some t for which $|A_t| = \Delta$. By the inequality above we then have:

$$|U_t| > \frac{d\Delta}{2}$$

On the other hand we started with $|A_0| \leq \frac{\Delta}{2}$, so $|U_0| \leq |N(A_0)| \leq \frac{d\Delta}{2}$, since the left degree is d. We get $|U_t| > |U_0|$, but $|U_j|$ is always nonincreasing (regardless of the size of A_j), so this is a contradiction. Therefore $|A_j| \leq \Delta$ at each step of the algorithm. □

Of course this decoding procedure can be improved in terms of time performance and generally linear time decoding for expander-based codes is of great interest in theoretical computer science.

References

[DSV03] Giuliana Davidoff, Peter Sarnak, and Alain Valette, *Elementary number theory, group theory, and Ramanujan graphs*, London Mathematical Society Student Texts, vol. 55, Cambridge University Press, Cambridge, 2003.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson, *Expander graphs and their applications*, Bull. Amer. Math. Soc. (N.S.) **43** (2006), no. 4, 439–561 (electronic).

[Kas07] Martin Kassabov, *Symmetric groups and expander graphs*, Inventiones Mathematicae **170** (2007), 327–354, 10.1007/s00222-007-0065-y.

[LGT11] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan, *Multi-way spectral partitioning and higher-order cheeger inequalities*, CoRR **abs/1111.1055** (2011).

[LPW09] D.A. Levin, Y. Peres, and E.L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, 2009.

[Pet] Gábor Pete, *Probability and geometry on groups*, `http://www.math.bme.hu/~gabor/PGG.html`.

[RVW02] Omer Reingold, Salil Vadhan, and Avi Wigderson, *Entropy waves, the zig-zag graph product, and new constant-degree expanders*, Ann. of Math. (2) **155** (2002), no. 1, 157–187.

[Tao] Terence Tao, *Expansion in lie type groups*, `http://terrytao.wordpress.com/category/teaching/254b-expansion-in-groups/`.

[Tre] Luca Trevisan, *Cs359g: Graph partitioning and expanders*, `http://theory.stanford.edu/~trevisan/cs359g/`.