CS 798 - Algorithmic Spectral Graph Theory, Fall 2015, Waterloo

Lecture 11: Small-set expanders

We will show that the noisy hypercube is a small-set expander, using basic results about Cayley graphs and hypercontrative inequality. The noisy hypercube shows that higher-order Cheeger's inequality is tight.

We also discuss other hard examples for random walks and eigenspace enumeration.

Cayley graphs

Let Γ be a group with a set of elements V and an operation $\cdot$ that combines two elements to form another element that satisfies some axioms (closure, associativity, existence of identity and inverse).

For example, the group $\mathbb{Z}/n\mathbb{Z}$ is defined on the set $\{0, \ldots, n-1\}$ with the operation being addition modulo n, when n is prime, the group $(\mathbb{Z}/n\mathbb{Z})^\times$ is defined on the set $\{0, \ldots, n-1\}$ with the operation being multiplication modulo n (since multiplicative inverse exists when n is prime).

A set $S \subseteq V$ is symmetric if $s \in S$ implies that $s^{-1} \in S$.

<u>Definition</u> Given a group $\Gamma = (V, \cdot)$ and a symmetric subset $S \subseteq V$, the Cayley graph $Cay(\Gamma, S)$ has vertex set V, and there is an edge between two elements a and b if and only if $a \cdot s = b$ for some $s \in S$.

So, the Cayley graph $Cay(\Gamma, S)$ is $|S|$-regular, and is undirected because S is symmetric.

We can also define a weighted Cayley graph $Cay(\Gamma, w)$, where $w: V \to \mathbb{R}$ is a symmetric weight function where $w(a) = w(a^{-1})$ for all $a \in V$, and an edge ab has weight $w(a^{-1}b)$.

<u>Definition</u> A function $\chi: V \to \mathbb{C}$ is called a character of the group if

$$\chi(a \cdot b) = \chi(a)\chi(b) \quad \text{for all } a, b \in V$$

The following lemma shows that a character defines an eigenvector of the Cayley graph $Cay(G, S)$ for any S.

<u>Lemma</u> Let $\chi: V \to \mathbb{C}$ be a character of the group Γ. Think of χ as a $|V|$-dimensional vector. Then χ is an eigenvector of $Cay(\Gamma, S)$ with eigenvalue $\frac{1}{|S|}\sum_{s \in S} \chi(s)$ for the normalized adjacency matrix of $Cay(\Gamma, S)$.

proof Let $\mathcal{A}$ be the normalized adjacency matrix of $Cay(\Gamma, S)$, with entry $\frac{1}{|S|}$ for an edge.

Then $$(\mathcal{A}\chi)_i = \sum_{s\in S} \mathcal{A}_{i,i\cdot s}\, \chi(i\cdot s) = \sum_{s\in S} \frac{1}{|S|} \chi(i)\cdot\chi(s)$$ by definition of character
$$= \Big(\sum_{s\in S} \frac{1}{|S|}\chi(s)\Big)\chi_i.$$

So, χ is an eigenvector with eigenvalue $\sum_{s\in S}\frac{1}{|S|}\chi(s)$. □

The important point is that χ is an eigenvector of $Cay(\Gamma, S)$ regardless of S, but the eigenvalue of χ depends on the choice of S.

It can be proved that two different characters are orthogonal (will see special cases).

Furthermore, for Abelian groups (satisfy the commutative property that $a\cdot b = b\cdot a$ $\forall\, a,b\in V$), there are exactly $|V|$ different characters.

Hence, for Abelian groups, we have a complete description of the spectrum of its Cayley graphs, using the characters of the group.

Hypercube

The n-dimensional hypercube is the graph with 2^n vertices, where each vertex corresponds to an n-bit string, and two strings have an edge if and only if they differ in exactly one bit.

Note that the n-dimensional hypercube is a Cayley graph of the group $(\mathbb{Z}/2\mathbb{Z})^n$, whose elements are the n-bit strings and the group operation is bit-wise addition modulo 2 (i.e. for two elements/strings $\vec{a} = (a_1,\dots,a_n)$ and $\vec{b} = (b_1,\dots,b_n)$, $\vec{a}\cdot\vec{b} = (a_1\oplus b_1,\dots,a_n\oplus b_n)$, where $\oplus$ denotes addition modulo 2).

Let S be the standard basis, i.e. $S := \{(1,0,\dots,0), (0,1,0,\dots,0), \dots, (0,0,\dots,0,1)\}$.

Then, observe that hypercube is $Cay((\mathbb{Z}/2\mathbb{Z})^n, S)$, as each bit string is connected to the n bit strings that differ on exactly one bit.

It is known that the characters of the group $(\mathbb{Z}/2\mathbb{Z})^n$ are: for each n-bit string b, $\chi_b(y) = (-1)^{\langle b,y\rangle}$, where y is an n-bit string and $\langle b,y\rangle$ denotes the standard inner product of two vectors. We can think of χ_b as a 2^n-dimensional vector.

Note that there are 2^n different characters, one for each n-bit string b.

To check that χ_b is a character, just see that $\chi_b(y\cdot z) = (-1)^{\langle b,y\cdot z\rangle} = (-1)^{\langle b,y\rangle}(-1)^{\langle b,z\rangle}$
$$= \chi_b(y)\,\chi_b(z).$$

We can also check that χ_a, χ_b are orthogonal (although it follows from general theory):

$$\langle \chi_a, \chi_b \rangle = \sum_y \chi_a(y)\chi_b(y) = \sum_y (-1)^{\langle a,y\rangle}(-1)^{\langle b,y\rangle} = \sum_y (-1)^{\langle a\cdot b, y\rangle} = \langle \chi_{a\cdot b}, \vec{1} \rangle = 0.$$

Therefore, we have an orthonormal basis of eigenvectors $\{\chi_b \mid b \text{ is an } n\text{-bit string}\}$ for the hypercube, and we can use these to easily compute the spectrum.

Suppose b is an n-bit string with k ones, say $b_1 = b_2 = \dots = b_k = 1$ and $b_{k+1} = \dots = b_n = 0$.

By the lemma, the eigenvalue for χ_b is equal to $\frac{1}{|S|}\sum_{i=1}^{n} \chi_b(\vec{e_i})$ where $\vec{e_i}$ is the i-th unit vector

$$= \frac{1}{n}\sum_{i=1}^{n} (-1)^{\langle b, \vec{e_i}\rangle} = \frac{1}{n}\sum_{i=1}^{n} (-1)^{b_i} = \frac{1}{n}\left(k(-1) + (n-k)(1)\right) = 1 - \frac{2k}{n}.$$

So, the spectrum of the hypercube has $1 - \frac{2k}{n}$ as eigenvalue with multiplicity $\binom{n}{k}$, for $0 \le k \le n$.

Fourier analysis

Given a boolean function $f: \{0,1\}^n \to \{0,1\}$, we can think of f as a 2^n-dimensional vector.

The characters χ_b for the hypercube are orthogonal vectors, and thus any f can be written as $f = \sum_b \langle f, \chi_b\rangle \chi_b = \sum_{S \subseteq V} \langle f, \chi_S \rangle \chi_S$ where for each string b, S is the set of ones in b.

The basis χ_S is called the Fourier basis, and $\hat{f}_S := \langle f, \chi_S\rangle$ are called the Fourier coefficients.

This is a very useful way to represent a boolean function with many applications; see [OD].

Noisy hypercube (optional)

We will see that the higher-order Cheeger's inequality $\phi_{\frac{k}{2}}(G) = O(\text{polylog}(k)\sqrt{\lambda_k})$ is tight, by showing that in the noisy hypercube all sets of size $\frac{cn}{k}$ for some $c \ge 1$ have conductance $\Omega(\sqrt{\lambda_k \log k})$.

<u>Definition</u> The noisy hypercube is a weighted graph in which every vertex corresponds to an n-bit string. Let $|x \oplus y|$ be the number of bits differed in x and y (the Hamming distance). The weight between two vertices x and y is $\varepsilon^{|x\oplus y|}$.

We can think of the weight of an edge is the probability that we get y from x, if we flip each bit of x with probability ε, and hence the name noisy hypercube as we add some noise on each coordinate (except that the probabilities don't sum to one).

The weighted degree of a vertex x is $\sum_y \varepsilon^{|x \oplus y|} = \sum_{i=0}^{n} \binom{n}{i} \varepsilon^i = (1+\varepsilon)^n$.

Clearly, the graph is undirected and regular.

The spectrum

First, observe that the noisy hypercube is a weighted Cayley graph.

The group is still $(\mathbb{Z}/2\mathbb{Z})^n$ with the group operation being bit-wise modulo 2 (bit-wise xor).

Each bit string b with k ones has weight $w(b) = \varepsilon^k$.

Let the adjacency matrix of the noisy hypercube be A.

Let χ_b be the vector defined by the character as before.

The eigenvalue of χ_b is $\sum_{y \in \{0,1\}^n} w(y) \chi_b(y) = \sum_{y \in \{0,1\}^n} \varepsilon^{|y|} (-1)^{\langle b, y \rangle}$ where $|y|$ denotes number of ones in y.

We divide the strings into disjoint subsets, where each subset S_ℓ consists of strings that have ℓ common ones with b, where $\ell \le k$ where k is the number of ones in b.

Then, $\sum_{y \in S_\ell} \varepsilon^{|y|} (-1)^{\langle b, y \rangle} = \binom{k}{\ell} \sum_{i=0}^{n-k} \binom{n-k}{i} \varepsilon^{\ell+i} (-1)^\ell$ ($\binom{k}{\ell}$ choices for common ones, $\binom{n-k}{i}$ choices of i ones outside)

$$= (-1)^\ell \varepsilon^\ell \binom{k}{\ell} \sum_{i=0}^{n-k} \binom{n-k}{i} \varepsilon^i = (-\varepsilon)^\ell \binom{k}{\ell} (1+\varepsilon)^{n-k}$$

So, $\sum_y \varepsilon^{|y|} (-1)^{\langle b, y \rangle} = \sum_{\ell=0}^{k} \sum_{y \in S_\ell} \varepsilon^{|y|} (-1)^{\langle b, y \rangle} = \sum_{\ell=0}^{k} (-\varepsilon)^\ell \binom{k}{\ell} (1+\varepsilon)^{n-k} = (1+\varepsilon)^{n-k} (1-\varepsilon)^k$.

Hence, the eigenvalue for χ_b in the normalized adjacency matrix is $(1+\varepsilon)^{n-k} (1-\varepsilon)^k / (1+\varepsilon)^n = \left(\frac{1-\varepsilon}{1+\varepsilon}\right)^k$, and so the eigenvalue $\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^k$ is of multiplicity $\binom{n}{k}$.

In particular, the eigenvalue $\frac{1-\varepsilon}{1+\varepsilon}$ is of multiplicity n, and so $\lambda_n(H) \le \varepsilon$ for the normalized Laplacian matrix. To summarize, there are n small eigenvalues.

Small-set expansion

Let $S \subseteq V$ be a subset of vertices in the noisy hypercube, where $|V| = 2^n$.

To argue that it has large expansion, it is equivalent to argue that the weight of induced edges $w(E(S,S))$ is small.

Let $\vec{1}_S$ be the characteristic vector of S, i.e. $\vec{1}_S(i) = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases}$ - not to be confused with character.

We write $\vec{1}_S$ using the Fourier basis as $\vec{1}_S = \sum_{T \subseteq V} \langle \vec{1}_S, \chi_T \rangle \chi_T$.

We will use the hypercontracity inequality to bound $\vec{1}_S^T A 1_S = 2w(E(S,S)) / (1+\varepsilon)^n$.

Hypercontractivity and noise operator

For a random variable X, the power mean inequality says that $\left(E[X^p]\right)^{\frac{1}{p}} \geq \left(E[X^q]\right)^{\frac{1}{q}}$ for $p \geq q$.

In particular, we have $E[X^4] \geq E[X^2]^2$.

Informally, we say that a random variable is (p, q, C)-hypercontractive if $\left(E[X^p]\right)^{\frac{1}{p}} \leq C\left(E[X^q]\right)^{\frac{1}{q}}$.

Intuitively, a random variable is not hypercontractive only if X is concentrated on a few values, and is hypercontractive if the distribution is reasonable (like Gaussian distribution).

Let X be a sum of independent random variables $\frac{1}{\sqrt{n}}(X_1 + \dots + X_n)$, then it is true that X is hypercontractive. More generally, it is true if X is a low-degree polynomial of independent random variable.

Another source of hypercontractivity comes from applying noise to the output

Let $f: \{0,1\}^n \to \mathbb{R}$ be a function and let $T_\varepsilon f(x) = \underset{y \in N_\varepsilon(x)}{E}[f(y)]$ where $N_\varepsilon(x)$ is the random string obtained from x by flipping each bit independently with $1-\varepsilon$.

Informally, adding noise to a function makes it more like a low-degree function, as the high degree part will have smaller contribution after adding noise.

The version of the hypercontractivity inequality that we use is the following (see [LOT, OD]).

Theorem Let $f: \{0,1\}^n \to \mathbb{R}$ and $1 \leq p \leq q < \infty$. Then $\overline{\|T_\eta f\|_q} \leq \overline{\|f\|_p}$ for $0 \leq \eta \leq \sqrt{\frac{p-1}{q-1}}$.

Here, for $x \in \mathbb{R}^n$, we write $\overline{\|x\|_p}$ to denote the expectation norm, i.e. $\overline{\|x\|_p} = \left(\frac{1}{n}\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}$

(By the power mean inequality, we have $\overline{\|f\|_q} \geq \overline{\|f\|_p}$ for $q \geq p$, but the theorem says that if we add some noise, then the reverse inequality is true for some p, q.)

In particular, we will set $q=2$ and use $\overline{\|T_\eta f\|_2} \leq \overline{\|f\|_{1+\eta^2}}$.

Small-set expansion (continued)

Recall that we want to bound $\vec{1}_S^T A \vec{1}_S = 2w(E(S,S))/(1+\varepsilon)^n$.

By writing $\vec{1}_S$ using the eigenbasis (the Fourier basis) as $\vec{1}_S = \sum_{T \subseteq V} \langle \vec{1}_S, \chi_T \rangle \chi_T$, we have

$$\frac{1}{2^n}\vec{1}_S^T A \vec{1}_S = \sum_{T \subseteq V} \langle \vec{1}_S, \chi_T \rangle^2 \left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{|T|}$$ as $\|\chi_T\|^2 = 2^n$ and $\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{|T|}$ is the eigenvalue of χ_T

$$= \overline{\left\| T_{\sqrt{\frac{1-\varepsilon}{1+\varepsilon}}} \vec{1}_S \right\|_2}^2$$ write $\vec{1}_S$ using the Fourier basis and apply noise on each monomial

$$\leq \overline{\|\vec{1}_S\|_{\frac{2}{1+\varepsilon}}}^2$$ using hypercontractivity theorem with $q=2$

$$= \left(\frac{|S|}{2^n}\right)^{1+\varepsilon}$$

Now, $$\phi(S) = \frac{w(E(S,\bar{S}))}{\mathrm{vol}(S)} = 1 - \frac{2w(E(S,S))}{\mathrm{vol}(S)} = 1 - \frac{(1+\varepsilon)^n \vec{1}_S^T \mathcal{A} \vec{1}_S}{(1+\varepsilon)^n |S|} \geq 1 - \frac{2^n}{|S|}\left(\frac{|S|}{2^n}\right)^{1+\varepsilon} = 1 - \left(\frac{|S|}{2^n}\right)^{\varepsilon}.$$

For $|S| \leq c2^n/k$ ($\frac{c}{k}$-fraction of the noisy hypercube), we have $\phi(S) \geq 1 - \left(\frac{k}{c}\right)^{-\varepsilon}$.

Setting $\varepsilon = \ln(2)/\ln(k/c)$, we have $\phi(S) \geq \frac{1}{2}$ for all sets S with $|S| \leq c2^n/k$.

Therefore, using this choice of ε, we have $\lambda_k = O\left(\frac{1}{\ln k}\right)$ while all sets with $\frac{c}{k}$-fraction of vertices have conductance $\geq \frac{1}{2}$.

Hence, the $\frac{c}{k}$-small-set expansion is at least $\Omega\left(\sqrt{\lambda_k \log k}\right)$, showing that the analysis of higher-order Cheeger's inequality is tight.

<u>Hard example for random walks</u> (optional)

A large-alphabet version of the noisy hypercube gives a hard example for using random walks to solve the small-set expansion problem.

Consider the graph with k^n vertices, where each vertex corresponds to a n-symbol string where each symbol is from $\{0, \ldots, k-1\}$.

The random walk matrix of the graph is defined as follows: for each string, we flip each symbol independently with probability ε to a uniformly random symbol (so with probability $\frac{\varepsilon}{k}$ to a particular symbol (and do nothing to a symbol with probability $1-\varepsilon$).

For two strings x and y, the weight between x and y is the probability that we go from x to y in one step of random walk.

This is the weighted graph construction.

First, it is clear that there is a small sparse cut.

Consider the set of strings with the first symbol being zero.

This has only $\frac{1}{k}$-fraction of vertices and its conductance is at most ε.

On the other hand, if we start a random walk from a vertex, then it is not diffitult to see that all vertices with same Hamming distance to the starting vertex has the same probability all the time (since the graph is so symmetric)

This implies that when we apply sweep-cut on the random walk vectors, all the level sets will be Hamming balls, and one can prove that all Hamming balls of size up to $\frac{1}{k}$-fraction have conductance close to 1, using another version of hypercontractivity [CKL].

Therefore, even if the graph has a set S with $\frac{1}{k}$-fraction of vertices and conductance $\leq \varepsilon$, the random walk algorithm will output a set with conductance $\geq 1-\varepsilon$.

This shows that the random walk algorithm fails to disprove the small-set expansion conjecture.

Hard example for subexponential time algorithm (optional)

The key of the subexponential time algorithm for small-set expansion is the theorem that for $k \geq n^{\Omega(1)}$, there is a set of size $\approx \frac{n}{k}$ with conductance $O(\sqrt{\lambda_k})$.

If this bound can be improved to $n^{o(1)}$, then the small-set expansion conjecture is essentially disproved.

There is an example, called the "short code graph", which shows that k must be at least $2^{\log^{\Omega(1)} n}$, and thus the analysis of the subexponential time algorithm is almost tight.

The proof is very interesting, showing that one can construct a small-set expander from "locally testable codes", and then using result on testing Reed-Muller codes to give near optimal parameters. It shows a new connection between small-set expansion and locally testable codes [BGHMRS].

Discussions

All known constructions of (non-trivial) small-set expanders use some hypercontractivity inequality.

It would be very interesting to come up with new methods in proving small-set expansion.

The complexity of finding a small sparse cut is still wide open.

One promising approach is to develop algorithms for finding sparse vectors in a linear subspace.

To search for analytically sparse vectors, one can formulate an optimization problem called the 2-to-4 norm problem using semidefinite programming [BBHKSZ].

There are no counterexamples to this approach, and it is an important problem to resolve.

Graph expansion is a very interesting area, with many beautiful results and nice techniques developed through the study of it. Although we have spent half the course on it, there are still many results that we haven't discussed at all, such as linear programming, semidefinite programming, metric embedding, important applications of expander graphs such as the PCP theorem and the construction of expander codes, analysis of mixing time, etc.

But anyway, we will move on to the second half of the course, focusing on electrical networks, fast algorithms, graph sparsification and interlacing polynomials.

References

[OD] Analysis of boolean functions, by O'Donnell.

[LOT] Multi-way spectral partitioning and higher-order Cheeger inequalities, by Lee, Oveis Gharan, Trevisan, 2012.

[CKL] Random walks and evolving sets: faster convergences and limitations, by Chan, Kwok, Lau, 2015.

[BGHMRS] Making the long code shorter, with applications to the unique games conjecture, by Barak, Gopalan, Hastad, Meka, Raghavendra, Steurer, 2012.

[BBHKSZ] Hypercontractivity, sum-of-square proofs, and their applications, by Barak, Brandao, Harrow, Kelner, Steurer, Zhou, 2012.