Advanced SDPs	**Lecture 3 : February 23, 2017**
Lecturers: Nikhil Bansal and Daniel Dadush	*Scribe: Nikhil Bansal*

3.1 Laplacian

The Laplacian of a graph is defined as $D - A$ where D is a diagonal matrix with entries as degrees. $L_G = D - A = \sum_{(i,j)\in E}(e_i - e_j)(e_i - e_j)^T$ which is $dI - A$ if G is regular.

Note that L_G is PSD, and that $x^T L_G x = \sum_{(u,v)\in E}(x_u - x_v)^2$. So we could write max-cut as $\max_{x\in\{0,1\}^n} x^T L_G x$.

For our purposes, it will be convenient to work with the normalized Laplacian. If G is d-regular we will define the normalized Laplacian as $I - A/d$. More generally, one defines this at $I - D^{-1/2}AD^{-1/2}$, where D is the diagonal matrix with entry D_{ii} = degree of vertex i.

We will consider the d-regular case only. Everything we say works for general degrees using the modified definition, but things get slightly more tedious. For this reason, in most lecture notes on spectral graph theory you will see the d-regular case, but research papers will show the same result for general case.

As we will only consider the normalized Laplacian and regular graphs henceforth, we denote $L_G = I - A/d$.

Some facts that you will prove in the exercises:

Theorem 1 *(Variational Characterization of Eigenvalues) Let M be a symmetric real matrix with eigenvaues $\lambda_1 \leq \ldots \leq \lambda_n$. Then*

$$\lambda_1 = \min_{x\in\mathbf{R}^n-\{\mathbf{0}\}} \frac{x^T M x}{x^T x}$$

$$\lambda_k = \max_{S:dim(S)=n-k+1} \min_{x\in S-\{\mathbf{0}\}} \frac{x^T M x}{x^T x}$$

$$\lambda_k = \min_{S:dim(S)=k} \max_{x\in S-\{\mathbf{0}\}} \frac{x^T M x}{x^T x}$$

Hint: Use the spectral decomposition for symmetric matrices that $M = \sum_i \lambda_i u_i u_i^T$ where λ_i are real, and u_i are orthonormal.

Theorem 2 *Let $\lambda_1 \leq \ldots \leq \lambda_n$ denote the eigenvalues of L_G. Then (i) $0 = \lambda_1$ and $\lambda_n \leq 2$. The all 1 vector is an eigenvector for λ_1. (ii) For any integer k, $\lambda_k = 0$ iff G has at least k components. (iii) $\lambda_n = 2$ iff G has a component that is bipartite.*

3.2 Conductance and Edge-Expansion

Given a graph G we define conductance of a set S as

$$\phi(S) = \frac{E(S, \bar{S})}{\sum_{v \in S} d_v} = \frac{E(S, \bar{S})}{vol(S)}$$

The fraction of edges across cut divided by total possible edges from S.

The conductance of G is defined as

$$\phi(G) = \min_{S: Vol(S) \leq Vol(G)/2} \phi(S)$$

When G is regular this is

$$\phi(G) = \min_{S: |S| \leq n/2} \frac{E(S, \bar{S})}{d|S|}$$

Examples: Conductance of cycle = 2/n. Hypercube = $1/\log n$. Another useful fact is that a random degree d graph for $d \geq 3$ has $\Omega(1)$ with high probability.

A graph with conductance $\Omega(1)$ is called an expander, and these are very useful objects with various applications. Understanding the conductance of a graph plays a major role is various areas, graph partitioning, markov chain mixing and so on.

So, how do we get a handle on the conductance of a graph?

As discussed in the previous lecture, we will discuss three approaches: Cheeger's inequality, which related expansion to the second smallest eigenvalue of L_G. An LP and metric-embedding based $O(\log n)$, and an SDP based $O(\sqrt{\log n})$-approximation which is a landmark result in approximation algorithms. Interestingly the latter result can be viewed as a combination of the previous two approaches. So we will develop a unified perspective on these three results.

3.3 Relaxations

Cheeger's inequality states the following.

Theorem 3
$$\lambda_2/2 \leq \phi(G) \leq \sqrt{2\lambda_2}$$

Before proving this, let us look at some relaxation of $\Phi(G)$.

The condition $|S| \leq n/2$ is the definition of $\Phi(G)$ is a bit pesky, so it is convenient to work with the following symmetric version, which is within a factor of 2 of $\Phi(G)$.

So let us define the (uniform) sparsity of G as

$$S(G) = \min_{S \neq \emptyset, V} \frac{n}{d} \frac{E(S, \bar{S})}{|S||V - S|}$$

Note that if $|S| \leq n/2$, then $(n/2)|S| \leq |S||V - S| \leq n|S|$, so

$$\phi(G) \leq S(G) \leq 2\Phi(G)$$

Now, an exact formulation for $S(G)$ on d-regular graphs would be the following.

$$\min_{S \subset V} \frac{n}{d} \frac{\sum_{(u,v) \in E \in} |1_S(u) - 1_S(v)|}{\sum_{(u,v)} |1_S(u) - 1_S(v)|}$$

There are two ways to relax this. We can interpret $|1_S(u) - 1_S(v)|$ as the distance $d(u, v)$ between vertices. If we only impose the constraint that $d(u, v)$ they be a metric, this will give the Leighton-Rao relaxation and the $O(\log n)$ approximation.

Or noting that $|1_S(u) - 1_S(v)| = (1_S(u) - 1_S(v))^2$, since these are 0-1, we can say

$$\frac{n}{d} \min_{x \in \{0,1\}^n} \frac{\sum_{(u,v) \in E} (x_u - x_v)^2}{\sum_{u,v} (x_u - x_v)^2}$$

Here x is supposed to be the indicator vector of set S. Relaxing the $\{0, 1\}$ condition and letting x_v be arbitrary reals gives Cheeger inequality. (One can also equivalently relax x_v to be arbitary vectors which can be written as an SDP).

Finally, we can combine the best of both, i.e. relax x_v to be vectors, and require that $d(u, v) = \|x_u - x_v\|^2$ satisfy the metric constraints. Then we get the ARV relaxation. In particular, this relaxation restricts the distance metric so that these distances can only be realized as squares of the euclidean distance between vectors.

3.4 Cheeger's inequality

Let us relax x to be arbitrary reals. Let

$$B = \frac{n}{d} \min_{x \in \mathbf{R}^n} \frac{\sum_{(u,v)\in E}(x_u - x_v)^2}{\sum_{(u,v)\in K}(x_u - x_v)^2}.$$

We first show that

Lemma 1 $\lambda_2 = B$

Proof: By the variational characterization of eigenvalues

$$\lambda_2 = \min_{x \in \mathbf{R}^n, x \perp \mathbf{1}} x^T L_G x / x^T x$$

Now consider the expression for B. The denominator can be written as $n \sum_u x_u^2 - (\sum_u x_u)^2$. Moreover, as the numerator does not change if x is shifted $c\mathbf{1}$ for any c, the denominator is maximized when $\sum_u x_u = 0$ and thus

$$\frac{n}{d} \min_{x \in \mathbf{R}^n} \frac{\sum_{(u,v)\in E}(x_u - x_v)^2}{\sum_{(u,v)\in K}(x_u - x_v)^2} = \min_{x \in \mathbf{R}^n, x \perp \mathbf{1}} \frac{x^T L_G x}{x^T x}$$

■

As B was a relaxation of $\Phi(G)$ (up to factor 2), this gives the easy side of Cheeger

$$\lambda_2/2 \le \Phi(G)$$

As $\frac{n}{d} \approx \frac{n(n-1)/2}{nd/2} = \frac{n(n-1)/2}{|E|}$, the lemma above also says that we can view λ_2 as follows: Put the vertices on a line, so that average squared distance between random pair of vertices is 1, but the average squared distance for an edge is as small as possible.

We now show the harder side. To do this, we need to find a cut with expansion at most $\sqrt{2\lambda_2}$.

3.4.1 Fiedler Cuts:

First we describe the algorithm, that is, and later we will see the proof.

Consider the eigenvector x corresponding to λ_2. Lay the entries $x_1, \ldots, x_n$ on a line, and look at the possible cuts $S_t = v : x_v \geq t$ for each $t \in R$. Clearly one needs to check at most n such cuts, and return the cut that minimizes

$$E[S_t, \bar{S}_t]/\min(|S_t|, |V - S_t|)$$

This heuristic is the basis of spectral partitioning and works very well in practice.

3.4.2 Proof:

For a vector x, let $R(x) = x^T L_G x / x^T x$ denote the Rayleigh quotient.

We need to find a cut with low sparsity and size at most $n/2$. The latter condition is important as otherwise, we might just return the entire vertex set. So, we first "balance" the vector x. Let c be some median of x. Consider $x' = x - c\mathbf{1}$. Then

Claim 1 $R(x') \leq R(x)$

Proof: The numerator can only increase as $x \perp \mathbf{1}$ and the numerator stays the same. ■

Let x^- with entries $x^i(i) = \max(-x(i), 0)$ be the non-positive part of x and $x^+(i) = \max(x(i), 0)$ be the positive part. Then $x = x + x'$ and $\|x\|^2 = \|x^+\|^2 + \|x^-\|^2$ (as the supports are disjoint). Moreover,

Claim 2 $x^T L_G x \geq x^+ L_g x^+ + x^- L_G x^-$.

Proof: For any edge (u, v), if u, v are both in the same part then the contribution is same on both sides, else if $x_u < 0$ and $x_v > 0$, then left contribution is $(x_v - x_u)^2$ while the right contribution is $x_v^2 + x_u^2$. ■ We now state a useful inequality that we

will use often. If $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are non-negative numbers then,

$$\min_i \frac{a_i}{b_i} \leq \frac{\sum_i a_i}{\sum_i b_i} \leq \max_i \frac{a_i}{b_i} \tag{3.1}$$

Note that this may not be true if the numbers could have different signs. E.g. $a_1 = b_1 = 2$ and $a_2 = -2$ and $b_2 = -1$.

Applying this with the two claims above, we get $\min(R(x^+), \mathbf{R}(x^-)) \leq R(x)$. Wlog suppose the support of x^+ is at most $n/2$. So we will try to find the cut only in this half of the vertices (which will sure $|S| \leq n/2$).

Let us make another observation which will be very useful. If $0 \leq y_1 \leq y_2 \ldots \leq y_n$ are non-negative numbers. Let E be some collection of edges on vertex set [n]. Then,

Claim 3 *Let $S_{\geq i} = \{i, \ldots, n\}$, and let $\delta(S)$ denote the set of edges between S and $V - S$. Then, it holds that*

$$\sum_{(i,j)\in E, i<j} y_j - y_i = \sum_{i=1}^{n} y_i - y_{i-1}\delta(S_{\geq i})$$

In fact, this does not require y_i to be non-negative. Second,

$$\sum_i y_i = \sum_{i=1}^{n} (y_i - y_{i-1})|S_{\geq i}|$$

where $y_0 = 0$.

Let us scale x^+ so that the max-entry is 1. This does not affect $R(x^+)$. Define the level cut $S_t = v : x_v^+ > t$. We will let t range from $(0, 1)$.

Wlog let us relabel the vertices s.t. $x_1^+ \leq \ldots \leq x_n^+$ Suppose our objective was

$$\frac{\sum_{(u,v)\in E, u<v} |y_v - y_u|}{d \sum_v y_v} \tag{3.2}$$

Then by Claim 3 the numerator can be written as $\int_{0,1} |\delta(S_{\geq t}|)dt$ and the denominator is $\int_0^1 d|S_{\geq t}|dt$. Applying the inequality 3.1 we could then recover a cut S with objective value at most in (3.2).

So, the idea will be to set $y_v = x_v^+$ (let us just denote x_v^+ as x_v to keep the superscripts under control. The objective (3.2) becomes $\frac{\sum_{(u,v)\in E, u\leq v} (x_v^2 - x_u^2)}{d \sum_v x_v^2}$. But by Cauchy Schwarz, the numerator is at most

$$\sum_{(u,v)\in E, u\leq v} (x_v^2 - x_u^2) \leq (\sum_{(u,v)\in E, u\leq v} (x_u - x_v)^2)^{1/2} (\sum_{(u,v)\in E} (x_u + x_v)^2)^{1/2}$$

But

$$(\sum_{(u,v)\in E} (x_u + x_v)^2)^{1/2} \leq \sum_{(u,v)\in E} (2x_u^2 + 2x_v^2)^{1/2} \leq (2d \sum_v x_v^2)^{1/2}$$

Thus we have that

$$\frac{\sum_{(u,v)\in E, u\leq v}(x_v^2 - x_u^2)}{d\sum_v x_v^2} \leq \frac{(2\sum_{(u,v)\in E, u\leq v}(x_u - x_v)^2)^{1/2}}{(d\sum_v x_v^2)^{1/2}} = \sqrt{2R(x^+)}$$

Thus we are done.

Remark: After reading further on the equivalence of ℓ_1-embeddings and cuts, the proof above can be viewed as mapping x^+ on the positive part to $y_v = x_v^2$ and viewing it as ℓ_1 distances on a line. The Cauchy-Schwarz argument bound the loss in the objective with this embedding.

3.5 Intuition and Tight Example

Cheeger's inequality should be viewed as a robust version of the fact that $\lambda_2 = 0$ for a disconnected graph.

Both sides of Cheeger's inequality are tight. For a cycle, $\lambda_2 = O(1/n^2)$ (show this by giving an embedding of vertices on a line and using the interpretation of λ_2 discussed above. It might be easier to first show this for the graph that is a line). However, Sparsity $= 4/n$.

For d-dimension Hypercube, it turns out that sparsity $= 1/d$ and $\lambda_2 = 2/d$. This shows tightness of the lower bound for Cheeger.

Moreover, for the hypercube there is one eigenvector corresponding to λ_2, such that if you look at the best Fiedler cut, you get $\Omega(1/\sqrt{d})$ sparsity. This is the majority cut. This also shows tightness of the analysis of Fiedler's cut (just based on using some eigenvector for λ_2).