ORIE 6334 Bridging Continuous and Discrete Optimization Oct 7, 2019

Lecture 10

Lecturer: David P. Williamson *Scribe: Lijun Ding*

In this lecture, we continue the proof of Cheeger's inequality and explore similar bounds on the largest eigenvalue of the normalized Laplacian. Recall that the normalized Laplacian is given by $\mathscr{L} = D^{-1/2} L_G D^{-1/2}$, where

$$D^{-1/2} = \begin{pmatrix} \frac{1}{\sqrt{d(1)}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{d(2)}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{d(n)}} \end{pmatrix},$$

and $d(i)$ is the degree of vertex i. When $S \subseteq V$, we define $\delta(S)$ as the set of edges with exactly one endpoint in S, and $\mathrm{vol}(S) = \sum_{i \in S} d(i)$. The conductance of S is defined as

$$\phi(S) = \frac{|\delta(S)|}{\min(\mathrm{vol}(S), \mathrm{vol}(V-S))},$$

and the conductance of G is defined as $\phi(G) = \min_{S \subseteq V} \phi(S)$. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ denote the eigenvalues of $\mathscr{L}$.

Denote x_2 to be the eigenvector associated with λ_2. Its Raleigh quotient $R(x_2) = \frac{x_2^\top \mathscr{L} x_2}{x_2^\top x_2}$ is simply λ_2. Recall from last time we define $y = (x_2)_+$ meaning $y(i) = \max(0, x_2(i))$ for each i. The support of y, $\mathrm{supp}(y) := \{i \mid y(i) > 0\}$, has cardinality less than or equal to $\frac{n}{2}$, by assuming (without loss of generality) x_2 satisfying $|\mathrm{supp}^+(x_2)| \leq |\mathrm{supp}^-(x_2)|$. The support of y, $\mathrm{supp}(y)$, is also nonempty, as x_2 has to be perpendicular to $D^{\frac{1}{2}} e$ where e is the all one vector.

1 Cheeger's Inequality

Let us now restate the upper bound of Cheeger's inequality.

Theorem 1 (Cheeger's inequality, upper bound) *We have $\phi(G) \leq \sqrt{2\lambda_2}$.*

Recall we are only dealing with d-regular graph in the proof. We have shown last time that $R(y) \leq R(x_2) = \lambda_2$ (Claim 3 in Lecture 9) and it is then enough for us to find an $S \subset \mathrm{supp}(y)$ such that $\frac{|\delta(S)|}{d|S|} \leq \sqrt{2R(y)}$. We state this as a lemma below.

[0]This lecture is derived from Lau's 2012 notes, Week 2, `http://appsrv.cse.cuhk.edu.hk/~chi/csc5160/notes/L02.pdf` and Lau's 2015 notes, Lecture 4, `https://cs.uwaterloo.ca/~lapchi/cs798/notes/L04.pdf`.

Lemma 2 *Given any nonzero $y \in \mathbb{R}^n$, if the graph is d-regular, then there exists an $S \subset supp(y)$ such that*

$$\frac{|\delta(S)|}{d|S|} \leq \sqrt{2R(y)}.$$

Proof: To start, we may assume without loss of generality that $y(i) \in [-1, 1]$ for each i as we can divide y by the largest entry (in magnitude) of it without affecting the Raleigh quotient $R(y)$ and the support of y.

We shall construct the S randomly. Let $S(t) := \{i \mid |y(i)|^2 > t\}$, where t is picked uniformly random from $[0, 1]$. Now the expectation of $|\delta(S(t))|$ is

$$\begin{aligned}
\mathbb{E}(|\delta(S(t))|) &= \sum_{(i,j)\in E} \mathbb{P}(\{i \in S(t),\, j \in V - S(t)\} \cup \{i \in V - S(t),\, j \in S(t)\}) \\
&\overset{(a)}{=} \sum_{(i,j)\in E} \mathbb{P}(|y(i)|^2 \leq t \leq |y(j)|^2 \text{ or } |y(j)|^2 \leq t \leq |y(i)|^2) \\
&= \sum_{(i,j)\in E} ||y(j)|^2 - |y(i)|^2| \\
&= \sum_{(i,j)\in E} |y(i) - y(j)||y(i) + y(j)| \\
&\overset{(b)}{\leq} \sqrt{\sum_{(i,j)\in E} (y(i) - y(j))^2} \sqrt{\sum_{(i,j)\in E} (y(i) + y(j))^2} \\
&\overset{(c)}{\leq} \sqrt{\sum_{(i,j)\in E} (y(i) - y(j))^2} \sqrt{2 \sum_{(i,j)\in E} (y(i)^2 + y(j)^2)} \\
&\overset{(d)}{=} \sqrt{\sum_{(i,j)\in E} (y(i) - y(j))^2} \sqrt{2d \sum_{i=1}^{n} y(i)^2}
\end{aligned} \tag{1}$$

The equality (a) is due to the distribution of t. The inequality (b) uses Cauchy-Schwarz. The inequality (c) uses the fact that $(a + b)^2 \leq 2a^2 + 2b^2$. The last equality (d) is due to that the graph is d-regular.

The expectation of $|S(t)|$ is

$$\mathbb{E}|S(t)| = \sum_{i=1}^{n} \mathbb{P}(i \in V) = \sum_{i=1}^{n} \mathbb{P}(|y(i)|^2 \geq t) = \sum_{i=1}^{n} y(i)^2$$

Recall that the Raleigh quotient of y is

$$R(y) = \frac{y^\top \mathscr{L} y}{y^\top y} = \frac{y^\top L_G y}{d y^\top y} = \frac{\sum_{(i,j)\in E} (y(i) - y(j))^2}{d \sum_{i=1}^{n} y(i)^2}.$$

Combining pieces, we find that

$$\mathbb{E}[|\delta(S(t))| - \sqrt{2R(y)}|S(t)|d] \leq 0.$$

By considering the assumption y is not zero, there must be some t_0 such that $|S(t_0)| \neq 0$ and

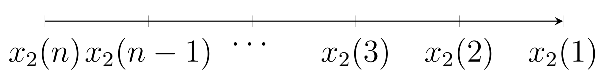
$$|\delta(S(t_0))| - \sqrt{2R(y)}|S(t_0)|d \leq 0.$$

Rearranging the terms yields the desired inequality. Note that we can find the desired t simply by trying all $t = y(i)^2$ for all $i \in V$. □

With this lemma, and consider the y constructed from x_2 with $\mathrm{supp}(y) \leq \frac{n}{2}$, we see the Cheeger's inequality for the upper bound is proved.

Last time, we mentioned **spectral partitioning** (Algorithm 1 in Lecture 9): Sort entries of x_2 and relabel them and the corresponding vertices so that $x_2(1) \geq x_2(2) \geq \cdots \geq x_2(n)$, take the **sweep cuts** for $i = 1, \ldots, n-1$, $S_i = \{1, \ldots, i\}$. Find $\min_{i=1,\ldots,n} \phi(S_i)$. The construction of the set $S(t_0)$ for $y = (x_2)_+$ in Lemma 2 shows that there is some i_0, t_0 such that $S(t_0) = V - S_{i_0}$ and

$$\min_{i=1,\ldots,n} \phi(S_i) \leq \phi(S_{i_0}) = \phi(S(t_0)) \leq \sqrt{2R(y)} \leq \sqrt{2R(x_2)} = \sqrt{2\lambda_2}.$$



2 Bounds on largest eigenvalue

We now turn to analyzing the largest eigenvalues λ_n of the normalized Laplacian. Note that

$$\lambda_n = \max_{x \in \mathbb{R}^n} \frac{x^\top \mathscr{L} x}{x^\top x} = \max_{x \in \mathbb{R}^n} \frac{x^\top D^{-1/2} L_G D^{-1/2} x}{x^\top x} = \max_{y \in \mathbb{R}^n} \frac{y^\top L_G y}{y^\top D y},$$

where we take $y = D^{-1/2}x$. Recall from last time, we have shown $\lambda_n \leq 2$. We also claim the following

Claim 3 $\lambda_n = 2$ *if and only if* G *has a bipartite component.*

We can easily show the if direction. If G has a bipartite component S with sides L, R, define a vector $y \in \mathbb{R}^n$ as $y(i) = 1$ if $i \in L$, $y(i) = -1$ if $i \in R$ and $y(i) = 0$ otherwise.

If $\delta(A, B)$ denotes the set of edges with one endpoint in A and another in B, we have

$$\frac{y^\top L_G y}{y^\top D y} = \frac{\sum_{(i,j) \in E} (y(i) - y(j))^2}{\sum_{i \in V} d(i) y(i)^2} = \frac{4\delta(L, R)}{\mathrm{vol}(S)} = \frac{2\,\mathrm{vol}(S)}{\mathrm{vol}(S)} = 2.$$

Now we'll show a statement stronger than the converse: G has a bipartite component when $\lambda_n = 2$, and has an "almost" bipartite component when λ_n is close to 2. To make this more precise, consider the following quantity

$$\beta(G) = \min_{\substack{S \subseteq V \\ S = L \cup R \\ L \cap R = \emptyset}} \frac{2|E(L)| + 2|E(R)| + |\delta(S)|}{\mathrm{vol}(S)},$$

for any $S \subset V$, where $E(X)$ denotes the set of edges with both endpoints in X. Note that

$$\frac{2|E(L)| + 2|E(R)| + |\delta(S)|}{\mathrm{vol}(S)} = \frac{\mathrm{vol}(S) - 2|\delta(L,R)|}{\mathrm{vol}(S)}.$$

Alternatively,

$$\beta(G) = \min_{y \in \{-1,0,1\}^n} \frac{\sum_{(i,j)\in E} |y(i) + y(j)|}{\sum_{i \in V} d(i)|y(i)|},$$

by taking $L = \{i : y(i) = 1\}, R = \{i : y(i) = -1\}$ and $S = L \cup R$.

Since λ_n is the largest eigenvalue of $\mathscr{L}$, $\beta_n = 2 - \lambda_n$ is the smallest eigenvalue of $2I - \mathscr{L} = 2I - (I - \mathscr{A}) = I + \mathscr{A}$. Hence

$$\beta_n = \min_{x \in \mathbb{R}^n} \frac{x^\top (I + \mathscr{A})x}{x^\top x} = \min_{x \in \mathbb{R}^n} \frac{x^\top D^{-1/2}(D + \mathscr{A})D^{-1/2}x}{x^\top x} = \min_{y \in \mathbb{R}^n} \frac{y^\top (D + A)y}{y^\top D y};$$

that is,

$$\beta_n = \min_{y \in \mathbb{R}^n} \frac{\sum_{(i,j)\in E} (y(i) + y(j))^2}{\sum_{i \in V} d(i) y(i)^2}.$$

Trevisan proves the following very nice analogy to the Cheeger inequality.

Theorem 4 (Trevisan 2009)

$$\frac{1}{2}\beta_n \le \beta(G) \le \sqrt{2\beta_n}.$$

Note when $\lambda_n = 2$, then $\beta_n = 2 - \lambda_n$ is zero and hence $\beta(G) = 0$ by the theorem. This means there is some $S, L, R \subset V$ such that $L \cap R = \emptyset, S = L \cup R$, and $\mathrm{vol}(S) = 2\delta(L,R)$. This equality simply means S is a bipartite component.

Proof: For the first inequality, simply note that

$$\begin{aligned}
\beta_n = \min_{y \in \mathbb{R}^n} \frac{\sum_{(i,j)\in E} (y(i) + y(j))^2}{\sum_{i \in V} d(i) y(i)^2} &\le \min_{y \in \{-1,0,1\}^n} \frac{\sum_{(i,j)\in E} (y(i) + y(j))^2}{\sum_{i \in V} d(i) y(i)^2} \\
&\le \min_{y \in \{-1,0,1\}^n} \frac{\sum_{(i,j)\in E} 2|y(i) + y(j)|}{\sum_{i \in V} d(i) y(i)^2} = 2\beta(G),
\end{aligned}$$

by noticing that $(y(i) + y(j))^2 \leq 2|y(i) + y(j)|$ for $y(i), y(j) \in \{-1, 0, +1\}$.

For the second inequality, pick $y \in \mathbb{R}^n$ satisfying $\beta_n = \frac{y^\top (D+A) y}{y^\top y}$ and assume that $\max_i y^2(i) = 1$ (if this is not true, scale y accordingly). Choose $t \in [0, 1]$ uniformly at random, and set $x(i) = 1$ if $x(i) \geq \sqrt{t}$, $x(i) = -1$ if $x(i) \leq -\sqrt{t}$ and $x(i) = 0$ otherwise. Next time we will show that

$$\mathbb{E}[\sum_{(i,j)\in E} |x(i) + x(j)| - \sqrt{2\beta_n} \sum_{i \in V} d(i)|x(i)|] \leq 0.$$

Then if we set $L_t = \{i \in V : x(i) = -1\}$, and $R_t = \{i \in V : x(i) = 1\}$, and $S_t = L_t \cup R_t$, we get that

$$\mathbb{E}[2|E(L_t)| + 2|E(R_t)| + |\delta(S_t)| - \sqrt{2\beta_n}\, \mathrm{vol}(S_t)] \leq 0,$$

implying that there exists a t such that

$$\frac{2|E(L_t)| + 2|E(R_t)| + |\delta(S_t)|}{\mathrm{vol}(S_t)} \leq \sqrt{2\beta_n},$$

or

$$\beta(G) \leq \sqrt{2\beta_n}.$$

Again, we can find t efficiently by trying all n values where $t = y(i)^2$. Next time we will prove the inequality and use it to get an approximation algorithm for the MAX CUT problem. $\square$