

Lecture 07: Power Method for Top Eigenvalues

*Lecturer: Yuan Zhou**Scribe: Syed Mahbub Hafiz*

1 Review

In the last lecture, we have proved the harder side of the Cheeger's Inequality, i.e. the conductance of a graph is upper bounded by second smallest eigenvalue, in formula,

$$\Phi_G \leq \sqrt{2\lambda_2}.$$

To prove that we took any vector \vec{X} having the property $\vec{X} \perp \vec{1}$ and Rayleigh Quotient

$$R(L_G, \vec{X}) = \frac{\vec{X}^T L_G \vec{X}}{\vec{X}^T \vec{X}} = \lambda_2$$

to construct a set so that the conductance of the graph is upper bounded by the inequality.

2 Topics

Today's lecture we'll discuss the followings:

1. Generalization of Cheeger's Inequality to non-regular graphs (still undirected).
2. Computation of the largest (and second smallest) eigenvalue.

3 Generalization to non-regular graph

The inequality proved on the assumption of regular graph can be extended to non-regular graph as well. Now, each vertex has different number of neighbors, hence degree of a vertex is different compared to others. Firstly, we define the Laplacian Matrix after normalizing it by the degree of the vertices this way:

$$L_G = I_n - D^{-\frac{1}{2}} \cdot A_G \cdot D^{-\frac{1}{2}},$$

where D is the diagonal matrix (consisting of only non-zero elements (degrees of the vertices) along its diagonal), and A_G is the adjacency matrix of the graph. We can check whether the defined Laplacian Matrix is generalized or not by the special case when the graph is d -regular, then $D = d \cdot I_n$ and L_G turns into

$$L_G = I_n - d^{-\frac{1}{2}} I_n \cdot A_G \cdot d^{-\frac{1}{2}} I_n = I_n - \frac{1}{d} A_G = I_n - W_G,$$

where W_G is the random walk matrix. So, the regular case is the special case of our generalized definition.

Another important property Local Variance for the non-regular graph should be defined in this way:

$$\mathcal{E}(\vec{X}) = \vec{X}^T L_G \vec{X} = \sum_{(u,v) \in E} \left(\frac{\vec{X}(u)}{\sqrt{\deg(u)}} - \frac{\vec{X}(v)}{\sqrt{\deg(v)}} \right)^2 \geq 0.$$

The smallest eigenvalue of the non-regular graph version is zero also, i.e. $\lambda_1(L_G) = 0$. And the corresponding eigenvector is

$$\vec{\psi}_1(u) = \sqrt{\frac{\deg(u)}{2m}}.$$

It can be checked that the second eigenvalue is zero if and only if the graph is not connected ($\lambda_2 = 0$), and $\lambda_n = 2$ if and only if the graph is bipartite. Cheeger's Inequality also holds for connected non-regular graph as:

$$\frac{\lambda_2(L_G)}{2} \leq \Phi_G \leq \sqrt{2\lambda_2(L_G)}.$$

Thus, we can extend regular graph properties for more general non-regular graph.

4 Finding the second smallest eigenvalue (λ_2)

We first introduce an approximate algorithm to find the largest eigenvalue of a Positive Semi-Definite (PSD) Matrix, then modification of this approach will lead us to find the second smallest one from our intended Laplacian Matrix.

4.1 Finding the largest eigenvalue (λ_{max}) of a PSD Matrix

A real symmetric matrix is PSD if all of its eigenvalues are ≥ 0 . Here, we're given a PSD Matrix $M \in \mathcal{R}^{n \times n}$, now we will find an approximation \vec{X} of the eigenvector associated with

the largest eigenvalue of λ_{max} (with error ε) in such a way so that the eigenvalue (Rayleigh Quotient) is at least

$$\frac{\vec{X}^T L_G \vec{X}}{\vec{X}^T \vec{X}} \geq (1 - \varepsilon) \lambda_{max}.$$

We follow a power method for this approximation:

Pseudo-code of the Power Method:

1. Start with a random unit vector $\vec{X}^{(0)} \in \mathcal{R}^n$.
2. For $t \leftarrow 1$ to T do compute

$$\vec{X}^{(t)} = \frac{M \cdot \vec{X}^{(t-1)}}{\|M \cdot \vec{X}^{(t-1)}\|_2}.$$

3. Return $\vec{X}^{(T)}$.

In step 1, we have our first guess and then gradually update it by powering until T . In the step 2, the normalization, to make it unit vector, is important otherwise the vector could be unreasonably long or short. Here, an important question is to decide what should be a reasonable (enough) T until when $\vec{X}^{(T)}$ will be aligned (almost) with the demanded eigenvector. We have the following claim to fix the T :

Proposition 1. *When $T \geq \frac{3 \ln n}{2\varepsilon}$, then the largest eigenvalue λ_{max} can be approximated from the bound*

$$R(M, \vec{X}^{(T)}) \geq (1 - \varepsilon) \lambda_{max}.$$

Proof. According to the *Spectral Theorem*, the real symmetric matrix M has n mutually orthogonal vectors $\vec{\psi}_1, \vec{\psi}_2, \dots, \vec{\psi}_n$ associated with n eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, and M can be written as

$$M = \sum_{i=1}^n \lambda_i \vec{\psi}_i \vec{\psi}_i^T.$$

Here, we assume that the eigenvalues are sorted and the order is

$$0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n,$$

so we are approximating $\lambda_n = \lambda_{max}$ basically. In addition, we fix T in a way so that the Rayleigh quotient of the returned vector is very close to λ_{max} .

The initial assumption can be represented by the eigenvectors of the matrix M as projections to the space, i.e.

$$\vec{X}^{(0)} = \sum_{i=1}^n y_i \cdot \vec{\psi}_i.$$

From the power method (successive exponentiation) above, it is clear that for each t ,

$$\vec{X}^{(t)} = \frac{M^t \cdot \vec{X}^{(0)}}{\|M \cdot \vec{X}^{(0)}\|_2}.$$

Therefore, we have

$$M^t \cdot \vec{X}^{(0)} = \sum_{i=1}^n y_i \cdot \vec{\psi}_i \cdot \lambda_i^t$$

and

$$\|M \cdot \vec{X}^{(0)}\|_2 = \sqrt{\sum_{i=1}^n y_i^2 \cdot \lambda_i^{2t}}.$$

Now, to get the eigenvalue- we have to evaluate the Rayleigh Quotient of $R(M, \vec{X}^{(t)})$. Here, by definition, $R(M, \vec{X}^{(t)}) = \frac{\vec{X}^{(t)T} M \vec{X}^{(t)}}{\vec{X}^{(t)T} \vec{X}^{(t)}}$. We already computed that $\vec{X}^{(t)} = \frac{M^t \cdot \vec{X}^{(0)}}{\|M \cdot \vec{X}^{(0)}\|_2}$. Hence, the numerator of the Rayleigh Quotient becomes

$$(M^t \cdot \vec{X}^{(0)})^T M (M^t \cdot \vec{X}^{(0)}) = (\vec{X}^{(0)})^T M^{2t+1} \vec{X}^{(0)} = \sum_{i=1}^n y_i^2 \cdot \lambda_i^{2t+1}.$$

The denominator can be easily figured out by squaring $\|M \cdot \vec{X}^{(0)}\|_2$ and thus $\sum_{i=1}^n y_i^2 \cdot \lambda_i^{2t}$.

Eventually,

$$R(M, \vec{X}^{(t)}) = \frac{\sum_{i=1}^n y_i^2 \cdot \lambda_i^{2t+1}}{\sum_{i=1}^n y_i^2 \cdot \lambda_i^{2t}}.$$

Now, In order to get the approximation, this Rayleigh Quotient should be at least $(1 - \varepsilon)\lambda_n$; i.e.

$$\begin{aligned} R(M, \vec{X}^{(t)}) = \frac{\sum_{i=1}^n y_i^2 \cdot \lambda_i^{2t+1}}{\sum_{i=1}^n y_i^2 \cdot \lambda_i^{2t}} &\geq (1 - \varepsilon)\lambda_n \Leftrightarrow \sum_{i=1}^n y_i^2 \cdot \lambda_i^{2t+1} \geq (1 - \varepsilon)\lambda_n \cdot \sum_{i=1}^n y_i^2 \cdot \lambda_i^{2t} \\ &\Leftrightarrow \sum_{i=1}^n y_i^2 \cdot \lambda_i^{2t} (\lambda_n - \lambda_i) \leq \varepsilon \cdot \lambda_n \cdot \sum_{i=1}^n y_i^2 \cdot \lambda_i^{2t}. \quad (1) \end{aligned}$$

Here, we do 2-case analysis, one is always above $(1 - \varepsilon)\lambda_n$, other is below it.

In the first case, let j be the smallest index (in the eigenvector λ_j) which is immediately above $(1 - \varepsilon)\lambda_n$. For each $i \geq j$, we have

$$\lambda_i \geq (1 - \varepsilon)\lambda_n \Rightarrow \lambda_n - \lambda_i \leq \varepsilon\lambda_n.$$

At this point in order to satisfy equation 1, when $i \geq j$, it is sufficient to prove that

$$\sum_{i=1}^{j-1} y_i^2 \cdot \lambda_i^{2t} (\lambda_n - \lambda_i) < \varepsilon\lambda_n \sum_{i=1}^{j-1} y_i^2 \lambda_i^{2t} + \varepsilon y_n^2 \lambda_n^{2t+1},$$

Here, the last term came from the fact that when $i = n$ (basically by evaluating the right hand side of the equation 1 at $i = n$). Now it suffices to prove that

$$\begin{aligned} \sum_{i=1}^{j-1} y_i^2 \cdot \lambda_i^{2t} (\lambda_n - \lambda_i) < \varepsilon y_n^2 \lambda_n^{2t+1} &\Leftrightarrow \sum_{i=1}^{j-1} \frac{y_i^2 \cdot \lambda_i^{2t} (\lambda_n - \lambda_i)}{\lambda_n^{2t+1}} < \varepsilon y_n^2 \\ &\Leftrightarrow \sum_{i=1}^{j-1} y_i^2 \cdot \left(\frac{\lambda_i}{\lambda_n}\right)^{2t} \left(1 - \frac{\lambda_i}{\lambda_n}\right) < \varepsilon y_n^2. \quad (2) \end{aligned}$$

To solve the equation 2, we first solve the following similar function by evaluating first derivative;

$$\begin{aligned} f(x) &= x^{2t}(1 - x) = x^{2t} - x^{2t+1} \\ \Rightarrow f'(x) &= 2tx^{2t-1} - (2t + 1)x^{2t} = x^{2t-1}(2t - (2t + 1)x) = x^{2t-1}(2t - 2tx - x). \end{aligned}$$

Now, to find the critical (optimal) point of x we put the first derivative to zero:

$$f'(x) = 0 \Rightarrow 2t - 2tx - x = 0 \Rightarrow x = \frac{2t}{2t + 1} = 1 - \frac{1}{2t + 1}.$$

If we observe the behavior of $f(x)$ against different x values, when x is small ($x < 1 - \frac{1}{2t+1}$), the function increasing slowly; when $x = 1 - \frac{1}{2t+1}$ then the function has its maximum (optimal) value as $f(1 - \frac{1}{2t+1})$, then with bigger x it goes down.

For $i < j$, we have that $\frac{\lambda_i}{\lambda_n}$ is upper bounded by $1 - \varepsilon$, therefore the function $\left(\frac{\lambda_i}{\lambda_n}\right)^{2t} \left(1 - \frac{\lambda_i}{\lambda_n}\right)$ should be smaller than $\varepsilon(1 - \varepsilon)^{2t}$ when $1 - \varepsilon$ is below the critical point $1 - \frac{1}{2t+1}$ (when $t > \frac{1}{\varepsilon}$ this is guaranteed). Therefore, to prove equation 2, we only need:

$$\sum_{i < j} y_i^2 \cdot (1 - \varepsilon)^{2t} \geq y_n^2$$

In the homework, we will prove that

$$\Pr \left[y_n^2 < \frac{1}{n^2} \right] = o(1).$$

When $y_n^2 \geq \frac{1}{n^2}$ we only need:

$$\sum_{i < j} (1 - \varepsilon)^{2t} < \frac{1}{n^2} \Leftrightarrow (1 - \varepsilon)^{2t} < \frac{1}{n^3} \Leftrightarrow 2t > \frac{3 \ln n}{\varepsilon} \Leftrightarrow t > \frac{3 \ln n}{2\varepsilon}.$$

To summarize, when $T > \frac{3 \ln n}{2\varepsilon}$, we have $R(M, \vec{X}^{(T)}) \geq (1 - \varepsilon)\lambda_{max}$ with probability $1 - o(1)$. □

4.2 Finding second smallest eigenvalue (λ_2) of the normalized Laplacian Matrix L_G

Recently we have a mechanism to find the largest eigenvalue, so definitely we can employ this idea to find the smallest eigenvalue of a symmetric matrix. Having smallest eigenvalue as 0 for L_G , here we can find the second smallest one (i.e. λ_2).

Hints: To find the second smallest eigenvalue we set

$$L' = 2I_n - L_G - 2\vec{\psi}_1\vec{\psi}_1^T,$$

where largest eigenvalue of the PSD matrix L' is the second smallest eigenvalue of L_G , because the smallest eigenvector $2\vec{\psi}_1\vec{\psi}_1^T$ is also subtracted. Furthermore,

$$L' = 2I_n - L_G - 2\vec{\psi}_1\vec{\psi}_1^T = I_n + W_G - 2n\vec{\pi}\vec{\pi}^T,$$

where $\vec{\pi}$ is the stationary distribution (which is uniform for regular graphs). Afterwards, the same mechanism discussed above can be applied for the PSD matrix L' .

4.3 Time complexity

If the graph is sparse and has m number of edges, then the time needed $O(\frac{m \log n}{\varepsilon})$, which is almost linear (only if ε is small). If the graph is d -regular, then $L'\vec{X}$ can be computed in $O(nd)$ time, which implies that the overall time is $O(\frac{nd \log n}{\varepsilon})$.

Additionally, we can analyze the time complexity of the exponentiation of $M^T \vec{X}^{(0)}$ by repeated squaring technique. Then number of matrix multiplication required $O(\log T)$, if each

matrix multiplication costs $O(n^\alpha)$, current $\alpha = 2.23$, then computation of $\vec{X}^{(t)}$ is bounded by

$$O(n^\alpha (\log \frac{1}{\varepsilon} + \log \log n)).$$

Hence, this is a polynomial-time algorithm.