CS 860 Spectral graph theory, Spring 2019, Waterloo.

Lecture 5 : Improved Cheeger's inequality

We will analyze the performance guarantee of Cheeger's rounding using higher eigenvalues.

This provides a better explanation of its success in practice.

Analysis of spectral partitioning

Cheeger's inequality states that $\frac{\lambda_2}{2} \le \phi(G) \le \sqrt{2\lambda_2}$, and the proof shows that the simple "sweep" algorithm on the second eigenvector will find a set $S \subseteq V$ with $\phi(S) \le \sqrt{2\lambda_2}$ with $|S| \le |V|/2$.

This implies that $\phi(S) \le 2\sqrt{\phi(G)}$ (as $\lambda_2 \le 2\phi(G)$), and thus the spectral partitioning algorithm is a $\frac{1}{\sqrt{\phi(G)}}$-approximation algorithm for the graph conductance problem.

But $\phi(G)$ could be as small as $\frac{1}{n^2}$ for a simple graph, and so the worst case approximation ratio could be as big as $\Omega(n)$.

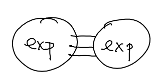
This does not give a satisfactory explanation of its success in practical applications.

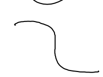
It would be good if we can identify a property, which is usually satisfied in a practical instance, and under which we can prove a stronger performance guarantee of the spectral partitioning algorithm.

In image segmentation, the instances usually have only a few outstanding objects that are easily recognized by humans, and we just want the compute to find them automatically.

If the graph has only one "outstanding" sparse cut, can we prove that the spectral partitioning algorithm perform better? How to formulate this assumption?

Suppose $\phi_2(G)$ is small but $\phi_3(G)$ is large. Then there is a good way to cut the graph into two pieces but there is no good way to cut the graph into three pieces. In this case, the graph should look like a sparse cut separating two expanders.



Then, intuitively, the second eigenvector should look more like a binary solution , with vertices on one side having similar values.

Can we prove that spectral partitioning works better when $\phi_3(G)$ is large?

The higher order Cheeger's inequality says that $\phi_3(G)$ is large if and only if λ_3 is large.

Can we prove that spectral partitioning works better when λ_3 is large?

<u>Theorem</u> [KLLOT 13] For any $k \ge 2$, the spectral partitioning algorithm outputs a set S with $\phi(S) = O\left(\frac{k\lambda_2}{\sqrt{\lambda_k}}\right)$.

Theorem [KLLOT 13] For any $k \geq 2$, the spectral partitioning algorithm outputs a set S with $\phi(S) = O\left(\frac{k\lambda_2}{\sqrt{\lambda_k}}\right)$.

Theorem [KLL 16] For any $k \geq 2$, the spectral partitioning algorithm outputs a set S with $\phi(S) = O\left(\frac{k\lambda_2}{\phi_k}\right)$.

So, when λ_k or ϕ_k is a constant for some constant k, then the spectral partitioning algorithm is a constant factor approximation algorithm for graph conductance.

This assumption is usually satisfied in practical instances of image segmentation and data clustering, and so this provides a better explanation why it works so well in practice.

Note that there are graphs in which the inequality is tight for all k, e.g. cycles.

Proof outline

We will prove the first result only.

We follow the intuition and prove that if λ_k is large, then the second eigenvector looks like a k-step function (that takes k distinct values), and that would allow us to do better rounding.

Lemma 1 (2k-step approximation) For any vector $x \in \mathbb{R}^n$ with $\|x\|=1$, there is a vector $y \in \mathbb{R}^n$ with only $2k$ distinct values such that $\|x-y\|^2 \leq O\left(\frac{R(x)}{\lambda_k}\right)$.

Lemma 2 (2k-step rounding) Let $x \in \mathbb{R}^n$ and $\|x\|=1$ and $y \in \mathbb{R}^n$ be a vector with only $2k$ distinct values. The spectral partitioning algorithm applied on x will return a set S with $\phi(S) \leq O(kR(x) + k\sqrt{R(x)}\,\|x-y\|)$.

Combining the two lemmas with a vector x with $R(x) = O(\lambda_2)$ will prove the theorem, because

$$\phi(S) \leq O(k\lambda_2 + k\sqrt{\lambda_2}\,\|x-y\|) \leq O\left(k\lambda_2 + k\sqrt{\lambda_2}\cdot\sqrt{\frac{\lambda_2}{\lambda_k}}\right) = O\left(\frac{k\lambda_2}{\sqrt{\lambda_k}}\right).$$

As in L03, we will assume that x is a non-negative vector with $R(x) \leq \lambda_2$ and $|\mathrm{supp}(x)| \leq n/2$, using the truncation argument.

2k-step approximation

We need to prove that if λ_k is large, then the second eigenvector looks like a $2k$-step function.

We will prove the contrapositive: if the second eigenvector is far from any $2k$-step function, then λ_k is small.

Intuitively, if the second eigenvector x is far from any $2k$-step function, then it looks like a "smooth" function  .

To prove that λ_k is small, we will use the "smooth" function x to construct $2k$ disjointly supported functions $\psi_1, \psi_2, \dots, \psi_{2k} \in \mathbb{R}^n$ such that $R(\psi_i)$ is not much bigger than $R(x)$.

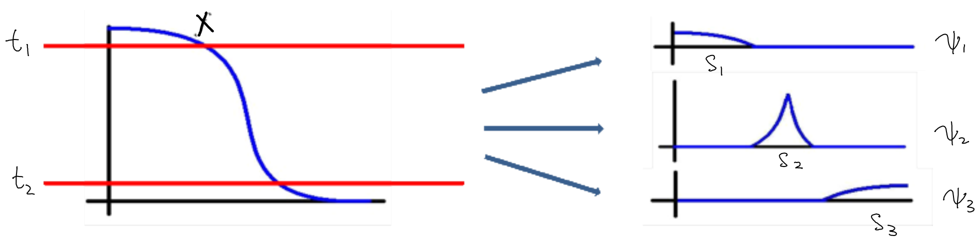
The following claim is a generalization of the easy direction of higher order Cheeger's inequality, whose proof is left as a homework problem.

<u>Claim</u> If $\psi_1, \psi_2, \dots, \psi_k \in \mathbb{R}^n$ are vectors with disjoint support, then $\lambda_k \le 2 \max_{1 \le i \le k} R(\psi_i)$.

Given a vector x that is "smooth", we will construct from it k disjointly supported vectors $\psi_1, \dots, \psi_k$, each with $R(\psi_i)$ not much larger than $R(x)$, and use the claim to bound λ_k.

To get the idea, let's consider the case when $k=3$.

We will pick two threshold values $t_1 \ge t_2$ and partition the vertices into three groups.



Let $S_1 = \{ i \mid x_i \ge t_1 \}$, $S_2 = \{ i \mid t_1 \ge x \ge t_2 \}$, and $S_3 = \{ i \mid t_2 \ge x \ge 0 \}$.

Each ψ_ℓ is a vector with support in S_ℓ defined as follows: $\psi_\ell(i) = \begin{cases} \min\{|x_i - t_1|, |x_i - t_2|\} & \text{if } i \in S_\ell. \\ 0 & \text{otherwise.} \end{cases}$

We want to show that if x is "smooth", then $R(\psi_\ell)$ is not much bigger than $R(x)$.

We will use a term-by-term analysis to compare $R(\psi_\ell)$ and $R(x)$.

By our construction of ψ_ℓ, it is clear that $|\psi_\ell(i) - \psi_\ell(j)| \le |x_i - x_j| \;\; \forall i \sim j$, and so the numerator of $R(\psi_\ell)$ is not larger than that of $R(x)$.

For the denominator, the idea is that if x is "smooth", then we can choose t_1, t_2 such that the denominator of $R(\psi_\ell)$ is not too small compared to the denominator of $R(x)$.

To do this, we choose t_1, t_2 such that the denominators of ψ_ℓ are the same, i.e. $\|\psi_1\|^2 = \|\psi_2\|^2 = \|\psi_3\|^2 = C$.

Then, $R(\psi_\ell) = \dfrac{\sum_{i \sim j} |\psi_\ell(i) - \psi_\ell(j)|^2}{d \|\psi_\ell\|^2} \le \dfrac{\sum_{i \sim j} |x_i - x_j|^2}{dC}$ (by construction $|\psi_\ell(i) - \psi_\ell(j)| \le |x_i - x_j|$)

$= \dfrac{\lambda_2 \cdot d \|x\|^2}{dC} = \dfrac{\lambda_2}{C}$ (by assumption $\|x\| = 1$).

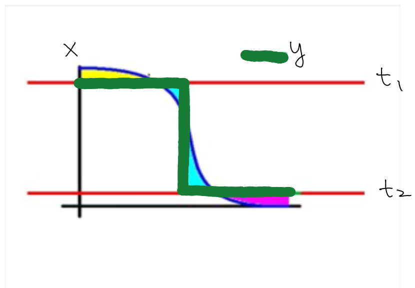
By the claim, this implies that $\frac{\lambda_3}{2} \le \max_{1\le \ell \le 3} R(\psi_\ell) \le \frac{\lambda_2}{C}$, and thus $C \le \frac{2\lambda_2}{\lambda_3}$.

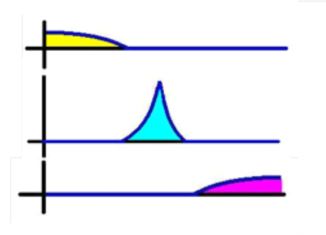
Now, notice that if we use t_1 and t_2 as a two step approximation y of x,

then $\|x-y\|^2 = \sum_{\ell=1}^{3} \|\psi_\ell\|^2$

$= 3C \le \frac{6\lambda_2}{\lambda_3}$,

proving Lemma 1 in this case.





If we follow the same argument for general k, then we use $k-1$ thresholds $t_1, \dots, t_{k-1}$ to partition the vertices into k groups, and define ψ_ℓ in the same way.

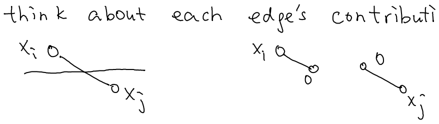
The same argument would imply that $C \le \frac{2\lambda_2}{\lambda_k}$, but this only implies that $\|x-y\|^2 = kC \le \frac{2k\lambda_2}{\lambda_k}$, with an additional factor k compared to Lemma 1.

To remove this extra factor k, we use a simple trick that is also useful in higher order Cheeger's inequality.

We use $2k-1$ thresholds $t_1 \ge t_2 \ge \dots \ge t_{2k-1}$ to divide the vertex set into $2k$ groups, and we choose these thresholds such that $\|\psi_\ell\|^2 = C$ for all $1 \le \ell \le 2k$.

Let us sort the ψ_ℓ by the numerator so that $\sum_{i\sim j} \|\psi_1(i)-\psi_1(j)\|^2 \le \sum_{i\sim j} \|\psi_2(i)-\psi_2(j)\|^2 \le \dots \le \sum_{i\sim j} \|\psi_{2k}(i)-\psi_{2k}(j)\|^2$.

The point is that $\sum_{\ell=1}^{2k} \sum_{i\sim j} \|\psi_\ell(i)-\psi_\ell(j)\|^2 \le \sum_{i\sim j} \|x_i - x_j\|^2$, (think about each edge's contribution



)

and thus $\sum_{i\sim j} \|\psi_\ell(i)-\psi_\ell(j)\|^2 \le \frac{1}{2k} \sum_{i\sim j} \|x_i - x_j\|^2 \quad \forall\, 1 \le \ell \le k$.

Therefore, for $1 \le \ell \le k$, $R(\psi_\ell) = \frac{\sum_{i\sim j} \|\psi_\ell(i)-\psi_\ell(j)\|^2}{d\,\|\psi_\ell\|^2} \le \frac{\frac{1}{2k}\sum_{i\sim j}\|x_i-x_j\|^2}{dC} = \frac{\frac{1}{2k}\lambda_2 d \|x\|^2}{dC} = \frac{\lambda_2}{2kC}$.

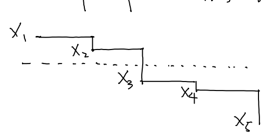
Hence, $\frac{\lambda_k}{2} \le \max_{1\le\ell\le k} R(\psi_\ell) \le \frac{\lambda_2}{2kC}$, and thus $C \le \frac{\lambda_2}{k\lambda_k}$.

It follows that $\|x-y\|^2 = \sum_{\ell=1}^{2k} \|\psi_\ell\|^2 = 2kC \le \frac{2\lambda_2}{\lambda_k}$, proving Lemma 1.

2k-step rounding

Ideal case

It is instructive to work out the ideal case when x is an exact $2k$-step function, i.e. there are only $2k$ distinct values $x_1 \ge x_2 \ge \dots \ge x_{2k} > 0$ in x.



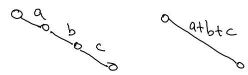
Naturally, we like to cut in places where $x_i - x_{i+1}$ is large, as the edges crossing must be long, and so there can't be too many such edges.

Let $S_\ell := \{ i \mid x_i \geq x_\ell \}$.

We output S_ℓ with probability $(x_\ell - x_{\ell+1})^2$, assuming $\sum_{\ell=1}^{2k} (x_\ell - x_{\ell+1})^2 = 1$ by scaling.

We will show that $\frac{E[|\delta(S)|]}{E[d|S|]} \leq k R(x)$ and conclude that there exists ℓ with $\phi(S_\ell) \leq k R(x)$.

$$E[|\delta(S)|] = \sum_{i \sim j} \Pr(ij \text{ is cut}) = \sum_{i \sim j} \sum_{\ell = i}^{j-1} (x_\ell - x_{\ell+1})^2 \leq \sum_{i \sim j} (x_i - x_j)^2$$



$a^2+b^2+c^2 \leq (a+b+c)^2$

$$E[d|S|] = \sum_{i \in V} d \Pr(i \text{ in } S) = \sum_{i \in V} d \sum_{\ell = i}^{2k} (x_\ell - x_{\ell+1})^2$$

$$\geq d \sum_{i \in V} \frac{\left((x_\ell - x_{\ell+1}) + (x_{\ell+1} - x_{\ell+2}) + \ldots + (x_{2k} - 0) \right)^2}{2k}$$

(Cauchy Schwarz says $\sum_{i=1}^{k} a_i^2 \geq \left(\sum_{i=1}^{k} a_i\right)^2 / k$)

$$= d \sum_{i \in V} \frac{x_i^2}{2k}.$$

Therefore, $\frac{E[|\delta(S)|]}{E[d|S|]} \leq \frac{\sum_{i \sim j} (x_i - x_j)^2}{d \sum_{i \in V} \frac{x_i^2}{2k}} = 2k R(x)$, proving Lemma 2 in this ideal case.

<u>General case</u>

In the general case, we are only given a vector x that is "close" to a $2k$-step vector y.

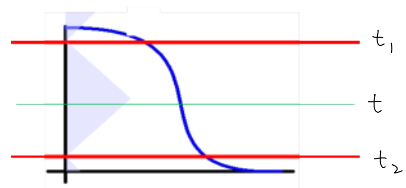
There is a nice way to generalize the above argument.

We pick t with probability proportional to $\min_{1 \leq \ell \leq 2k} \{ |t - t_\ell| \}$

(in words, the distance to the closest threshold), and

output $S = \{ i \mid x_i \geq t \}$.

probability distribution



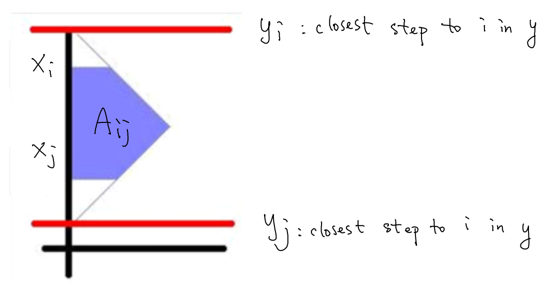
Without loss, we assume these probabilities sum to one by scaling.

Let's analyse $E[|\delta(S)|]$.

$$E[|\delta(S)|] = \sum_{i \sim j} \Pr(ij \text{ is cut})$$

$$\leq \sum_{i \sim j} \text{area}(A_{ij})$$

(this is worst possible: if there is another threshold in the middle then the area is)



big triangle   the two small triangles on the sides

$$= \sum_{i \sim j} \frac{1}{4} \left(|x_i - y_i| + |x_i - x_j| + |x_j - y_j| \right)^2 - \frac{1}{2}(x_i - y_i)^2 - \frac{1}{2}(x_j - y_j)^2$$

$$\leq \sum_{i \sim j} \frac{1}{4} \left[|x_i - x_j|^2 + 2 |x_i - x_j| \left(|x_i - y_i| + |x_j - y_j| \right) \right]$$

$$= \frac{1}{4} R(x) \cdot d \|x\|^2 + \frac{1}{2} \sum_{i \sim j} |x_i - x_j| \left(|x_i - y_i| + |x_j - y_j| \right)$$

$$= \frac{1}{4} R(x)\, d \|x\|^2 + \frac{1}{2} \sqrt{\sum_{i \sim j} (x_i - x_j)^2} \sqrt{\sum_{i \sim j} \left(|x_i - y_i| + |x_j - y_j| \right)^2}$$

Cauchy-Schwarz

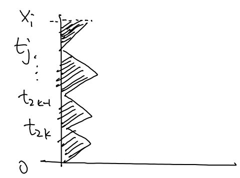
$$= \frac{1}{4} R(x)\, d \|x\|^2 + \frac{1}{2} \sqrt{R(x) \cdot d \|x\|^2} \sqrt{\sum_{i \sim j} \left(2(x_i - y_i)^2 + 2(x_j - y_j)^2 \right)}$$

$(a+b)^2 \leq 2a^2 + 2b^2$

$$= \frac{1}{4} R(x)\, d \|x\|^2 + \frac{1}{2}\sqrt{R(x)\cdot d\|x\|^2}\,\sqrt{\sum_{i\sim j}\left(2(x_i-y_i)^2 + 2(x_j-y_j)^2\right)} \qquad (a+b)^2 \le 2a^2+2b^2$$

$$= \frac{1}{4} R(x)\, d\|x\|^2 + \frac{1}{2}\sqrt{R(x)\cdot d\|x\|^2}\,\sqrt{\sum_{i\in V} 2d\,(x_i-y_i)^2}$$

$$= \frac{d}{4} R(x) + \frac{d}{\sqrt{2}}\sqrt{R(x)}\,\|x-y\|$$



$$E[|S|] = \sum_{i\in V} \Pr(i \text{ in } S)$$

$$= \sum_{i\in V}\left[\text{area below } x_i\right]$$

$$\ge \sum_{i\in V} \frac{1}{4}(x_i - t_j)^2 + \sum_{l=j}^{2k}\frac{1}{4}(t_l - t_{l+1})^2$$

$$\ge \sum_{i\in V}\frac{1}{4}\left((x_i-t_j)+(t_j-t_{j+1})+\dots+(t_{2k-1}-t_{2k})+(t_{2k}-0)\right)^2 / 2k \qquad \text{by Cauchy Schwarz } \sum_{i=1}^{k} a_i^2 \ge \left(\sum_{i=1}^{k} a_i\right)^2/k$$

$$= \sum_{i\in V}\frac{x_i^2}{8k} = \frac{1}{8k} \qquad \text{by our assumption } \|x\|^2 = 1.$$

Therefore, $\dfrac{E[|\delta(S)|]}{E[d|S|]} \le 2k\,R(x) + 4\sqrt{2}\,k\sqrt{R(x)}\,\|x-y\|$, proving Lemma 2.

References

[KLLOT13] Improved Cheeger's inequality: analysis of spectral partitioning through higher order spectral gap. by Kwok, Lau, Lee, Oveis Gharan, Trevisan, 2013.

[KLL16] Improved Cheeger's inequality and analysis of local graph partitioning using vertex expansion and expansion profile. by Kwok, Lau, Lee, 2016.