Lecture 6: Cheeger-type Inequalities for λ_n

In which we state an analog of Cheeger's inequalities for the k-th smallest Laplacian eigenvalue, and we discuss the connection between this result and the analysis of spectral partitioning algorithms

1 Cheeger-type Inequalities for λ_k

Let $G = (V, E)$ be an undirected d-regular graph, A its adjacency matrix, $L = I - \frac{1}{d}A$ its normalized Laplacian matrix, and $0 = \lambda_1 \leq \cdots \leq \lambda_n \leq 2$ be the eigenvalues of L, counted with multiplicities and listed in non-decreasing order.

In Handout 2, we proved that $\lambda_k = 0$ if and only if G has at least k connected components, that is, if and only if there are k disjoint sets $S_1, \ldots, S_k$ such that $\phi(S_i) = 0$ for $i = 1, \ldots, k$. In this lecture and the next one we will prove a robust version of this fact.

First we introduce the notion of higher-order expansion. If $S_1, \ldots, S_k$ is a collection of disjoint sets, then their order-k expansion is defined as

$$\phi_k(S_1, \ldots, S_k) = \max_{i=1,\ldots,k} \phi(S_i)$$

and the order-k expansion of a graph G is

$$\phi_k(G) = \min_{S_1,\ldots,S_k \text{ disjoint}} \phi(S_1, \ldots, S_k)$$

If the edges of a graph represent a relation of similarity of affinity, a low-expansion collection of sets $S_1, \ldots, S_k$ represents an interesting notion of clustering, because the vertices in each set S_i are more related to each other than to the rest of the graph. (Additional properties are desirable in a good clustering, and we will discuss this later.)

We will prove the following higher-order Cheeger inequalities:

$$\frac{\lambda_k}{2} \leq \phi_k(G) \leq O(k^{3.5})\sqrt{\lambda_k}$$

Stronger upper bounds are known, but the bound above is easier to prove from scratch. It is known that $\phi_k(G) \leq O(k^2)\sqrt{\lambda_k}$ and that $\phi_k(G) \leq O_\epsilon(\sqrt{\log k}) \cdot \sqrt{\lambda_{(1+\epsilon)\cdot k}}$.

2 The Easy Direction

As usual, the direction $\frac{\lambda_k}{2} \leq \phi_k(G)$ is the easy one, and it comes from viewing λ_k as a sort of continuous relaxation of the problem of minimizing order-k expansion.

Recall that, in order to prove the easy direction of Cheeger's inequality for λ_2, we proved that if $\mathbf{x}$ and $\mathbf{y}$ are two orthogonal vectors, both of Rayleigh quotient at most ϵ, then the Rayleigh quotient of their sum is at most 2ϵ. A similar argument could be made to show that the Rayleigh quotient of the sum of k such vectors is at most $k\epsilon$. Such results hold for every positive semidefinite matrix.

In the special case of the Laplacian of a graph, and of vectors that are not just orthogonal but actually *disjointly supported*, then we can lose only a factor of 2 instead of a factor of k. (The *support* of a vector is the set of its non-zero coordinates; two vectors are disjointly supported if their supports are disjoint.)

Lemma 1 *Let $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$ be disjointly supported vectors. Then*

$$R_L\left(\sum_i \mathbf{x}^{(i)}\right) \leq 2 \cdot \max_{i=1,\ldots,k} R_L(\mathbf{x}^{(i)})$$

PROOF: We just have to prove that, for every edge $\{u, v\}$,

$$\left(\sum_i x_u^{(i)} - x_v^{(i)}\right)^2 \leq 2\sum_i (x_u^{(i)} - x_v^{(i)})^2$$

The support disjointness implies that there is an index j such that $x_u^{(i)} = 0$ for $i \neq j$, and an index k such that $x_v^{(i)} = 0$ for $i \neq k$. If $j = k$, then

$$\left(\sum_i x_u^{(i)} - x_v^{(i)}\right)^2 = (x_u^{(j)} - x_v^{(j)})^2 = \sum_i (x_u^{(i)} - x_v^{(i)})^2$$

and, if $j \neq k$, then

$$\left(\sum_i x_u^{(i)} - x_v^{(i)}\right)^2 = (x_u^{(j)} - x_v^{(k)})^2$$

$$\leq 2(x_u^{(j)})^2 + 2(x_v^{(k)})^2 = 2\sum_i (x_u^{(i)} - x_v^{(i)})^2$$

and now, using also the fact that disjoint support implies orthogonality, we have

$$R_L\left(\sum_i \mathbf{x}^{(i)}\right) = \frac{\sum_{\{u,v\}}\left(\sum_i x_u^{(i)} - x_v^{(i)}\right)^2}{\|\sum_i \mathbf{x}^{(i)}\|^2}$$

$$\leq 2\frac{\sum_i \sum_{\{u,v\}\in E}(x_u^{(i)} - x_v^{(i)})^2}{\sum_i ||\mathbf{x}^{(i)}||^2}$$

$$\leq 2 \max_{i=1,\ldots,k} \quad R_L(\mathbf{x}^{(i)})$$

□

To finish the proof of the easy direction, let $S_1, \ldots, S_k$ be sets such that $\phi(S_i) \leq \phi(G)$ for every i. Consider the k-dimensional space X of linear combinations of the indicator vectors $\mathbf{1}_{S_i}$ of such sets. The indicator vectors have Rayleigh quotient at most $\phi(G)$ and are disjointly supported, so all their linear combinations have Rayleigh quotient at most $2\phi(G)$. We have found a k-dimensional space of vectors all of Rayleigh quotient $\leq 2\phi(G)$, which proves $\lambda_k \leq 2\phi(G)$.

3 The Difficult Direction: Main Lemma

We will prove the following result

Lemma 2 (Main) *Let $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$ be orthonormal vectors. Then we can find disjointly supported non-negative vectors $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k)}$ such that for every $i = 1, \ldots, k$*

$$R_L(\mathbf{y}^{(i)}) \leq O(k^7) \cdot \max_{j=1,\ldots,k} R_L(\mathbf{x}^{(j)})$$

By applying the Main Lemma to the eigenvectors of $\lambda_1, \ldots, \lambda_k$, we get disjointly supported vectors $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k)}$ all of Rayleigh quotient at most $O(k^7) \cdot \lambda_k$. In a past lecture, we proved that for every non-negative vector $\mathbf{y}$ there is a subset S of its support such that $\phi(S) \leq \sqrt{2R_L(\mathbf{y})}$, and applying this fact to the vectors $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k)}$ we find k disjoint sets all of expansion at most $O(k^{3.5}) \cdot \sqrt{\lambda_k}$, proving

$$\phi_k(G) \leq O(k^{3.5}) \cdot \sqrt{\lambda_k}$$

It is possible, with a more involved proof, to improve the $O(k^7)$ factor in the conclusion of the Main Lemma to $O(k^6)$, implying that $\phi_k(G) \leq O(k^3) \cdot \sqrt{\lambda_k}$. A different approach, which we will not discuss, is used to show that, given k orthonormal vectors, one can find k disjoint sets $S_1, \ldots, S_k$ such that, for all i,

$$\phi(S_i) \leq O(k^2) \cdot \sqrt{\max_{j=1,\ldots,k} R_L(\mathbf{x}^{(j)})}$$

implying $\phi_k(G) \leq O(k^2) \cdot \sqrt{\lambda_k}$, which is the best known bound.

Note that, in all the known arguments, the bounds still hold if one replaces λ_k by the (possibly smaller) quantity

$$\inf_{\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(k)} \text{ orthonormal}} \max_{i=1,\ldots,k} R_L(\mathbf{x}^{(i)}) \tag{1}$$

There are graphs, however, in which

$$\phi_k(G) \geq \Omega(\sqrt{k}) \cdot \sqrt{\inf_{\mathbf{x}^{(k)},\ldots,\mathbf{x}^{(k)} \text{ orthonormal}} \max_{i=1,\ldots,k} R_L(\mathbf{x}^{(i)})}$$

so, if a bound of the form $\phi_k(G) \leq (\log k)^{O(1)} \cdot \sqrt{\lambda_k}$ is true, then, in order to prove it, we need to develop new techniques that distinguish between λ_k and the quantity (1).

4 The Spectral Embedding

Given orthonormal vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$ as in the premise of the Main Lemma, we define the mapping $F : V \to \mathbb{R}^k$

$$F(v) := (x_v^{(1)}, \ldots, x_v^{(k)}) \tag{2}$$

If $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$ are the eigenvectors of the k smallest Laplacian eigenvalues of L, then $F(\cdot)$ is called the *spectral embedding* of G into $\mathbb{R}^k$. *Spectral clustering* algorithms compute such an embedding, and then find clusters of nodes by clustering the points $\{F(v) : v \in V\}$ using geometric clustering algorithms, such as k-means, according either to Euclidian distance, or to the normalized distance function

$$dist(u, v) := \left\| \frac{F(u)}{||F(u)||} - \frac{F(v)}{||F(v)||} \right\| \tag{3}$$

Our construction of disjointly supported vectors with small Rayleigh quotient will proceed similarly, by working only with the points $\{v : F(v)\}$ and forgetting the edge structure of the graph, and by making use of the above distance function.

To develop some intuition about the spectral mapping, we introduce a notion of Laplacian Rayleigh quotient for a mapping $f : V \to \mathbb{R}^k$, defined, by formally replacing absolute values with norms, as

$$R_L(f) := \frac{\sum_{\{u,v\}} ||f(u) - f(v)||^2}{d \sum_v ||f(v)||^2}$$

For a mapping $F : V \to \mathbb{R}^k$ defined in terms of k orthonormal vectors $\mathbf{x}^{(i)}$ as in (2), we have

$$\begin{aligned}
R_L(F) &= \frac{\sum_{\{u,v\}} \sum_i (x_u^{(i)} - x_v^{(i)})^2}{d \sum_v \sum_i (x_v^{(i)})^2} \\
&= \frac{\sum_i \sum_{\{u,v\}} (x_u^{(i)} - x_v^{(i)})^2}{dk} \\
&= \frac{1}{k} \sum_i \frac{\sum_{\{u,v\}} (x_u^{(i)} - x_v^{(i)})^2}{d} \\
&= \frac{1}{k} \sum_i R_L(\mathbf{x}^{(i)}) \\
&\leq \max_{i=1,\ldots,k} \; R_L(\mathbf{x}^{(i)})
\end{aligned}$$

In particular, if $\mathbf{x}^{(i)}$ are the eigenvectors of the k smallest Laplacian eigenvalues, then $R_L(F) \leq \lambda_k$.

Let us use this setup to prove again that if $\lambda_k = 0$ then G has at least k connected components. If $\lambda_k = 0$, and we construct $F(\cdot)$ using the eigenvectors of the smallest Laplacian eigenvalues, then $R_L(F) = 0$, which means that $F(u) = F(v)$ for every edge $\{u, v\}$, and so $F(u) = F(v)$ for every u and v which are in the same connected component. Equivalently, if $F(u) \neq F(v)$, then u and v are in different connected component. For every point in the range $\{F(v) : v \in V\}$ in the range of $F(\cdot)$, let us consider its pre-image, and let $S_1, \ldots, S_t$ be the sets constructed in this way. Clearly, every set has expansion zero.

How many sets do we have? We claim that the range of $F(\cdot)$ must contain at least k distinct points, and so $t \geq k$ and G has at least k connected component. To prove the claim, consider the matrix X whose rows are the vectors $\mathbf{x}^{(i)}$; since the rows of X are linearly independent, X has full rank k; but if the range of $F(\cdot)$ contained $\leq k-1$ distinct points, then X would have $\leq k - 1$ distinct columns, and so its rank would be $\leq k - 1$.

Our proof of the higher-order Cheeger inequality will be somewhat analogous to the previous argument: we will use the fact that, if the Rayleigh quotient of $F(\cdot)$ is small, then the endpoints of edges $\{u, v\}$ are typically close, in the sense that the distance defined in (3) between u and v will typically be small; we will also use the fact that, because the $\mathbf{x}^{(i)}$ are orthonormal, $F(\cdot)$ tends to "spread out" vertices across $\mathbb{R}^k$, so that we can find k regions of $\mathbb{R}^k$ each containing a large (in a certain weighted sense) number of vertices, and such that the regions are well-separated according to the distance (3), implying that there are few edges crossing from one region to the other, so that the vertices in each region are a non-expanding set. (This is an imprecise description of the argument, but it conveys the basic intuition.)

5 Overview of the Proof of the Main Lemma

We will break up the proof of the Main Lemma into the following two Lemmas.

Lemma 3 (Well-Separated Sets) *Given a function $F : V \to \mathbb{R}^k$ defined in terms of k orthonormal vectors as in (2), we can find k disjoint subsets of vertices $A_1, \ldots, A_k$ such that*

- *For every $i = 1, \ldots, k$, $\sum_{v \in A_i} ||F(v)||^2 \geq \frac{1}{4}$*
- *For every u and v belonging to different sets, $dist(u, v) \geq \Omega(k^{-3})$*

Lemma 4 (Localization) *Given a function $F : V \to \mathbb{R}^k$ defined in terms of k orthonormal vectors as in (2), and t sets $A_1, \ldots, A_t$ such that, for every $i = 1, \ldots, t$, $\sum_{v \in A_i} ||F(v)||^2 \geq \frac{1}{4}$ and,for every u, v in different sets $dist(u, v) \geq \delta$, we can construct t disjointly supported vectors $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(t)}$ such that for every $i = 1, \ldots, t$, we have*

$$R_L(\mathbf{y}^{(t)}) \leq O(k \cdot \delta^{-2}) \cdot R_L(F)$$