Geometry of Polynomials

Lecturer : **Satyaki M.**, Scribes : **Chia Gim Siang**, **Clarence Chew**

July 2024

1 Chapter 1

This section will first introduce basic linear algebra knowledge required in this lecture series, largely adapted from [1]. A graph $G = (V, E)$ is a the set of vertices or nodes V, and a set of edges E which are unordered tuples from $V \times V$. We let $n = |V|$ and $m = |E|$. The Adjacency matrix A associated with G is defined as follows:

$$A_{i,j} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

A is thus a $n \times n$ matrix.

The Degree matrix D associated with G is defined as follows:

$$D_{i,j} \stackrel{\text{def}}{=} \begin{cases} d_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

where d_i is the number of neighbours of vertex i. Note that D is a $n \times n$ diagonal matrix with positive entries. Also of note is that A and D are real symmetric matrices, a fact we will heavily use going forward.

A vertex cut is a partition of V into two disjoint subsets. That is, for a vertex $S \subseteq V$, we can express $(S, V \setminus S)$ as a cut of the graph, where $|S| \leq |V \setminus S|$.

A flow is another perspective of a cut. To broadly explain flow without going to details, we could consider the graph as a water pipe system where water flows from a source vertex the sink vertex, with the edges acting as pipes for the water to flow through. The maximum flow problem aims to find the maximum amount of flow that can be sent from the source to the sink without violating capacity constraints. One thing that relates the cut to the flow is the max-flow min-cut theorem, which states that the maximum flow of a network is equal to the capacity of the maximum cut. This theorem has many applications, such as the projection selection problem where given a number of projects and a number of machines, we determine which projects to select and which machines to be purchased to maximize profit.

The thing that we are interested in is however not quite the minimum cut but rather the minimum cut in respective to total number of vertices that is being separated. This problem arises quite naturally out of flows following different laws e.g. the heat equation. Or as we will see later this quantity is related to the rate of mixing for a random walk. Other often commonly cited examples include traffic flows, clusters in a network etc. Some of the rationale behind these connections we will see later.

Either way we need a good linear algebraic way to represent cuts. One way is to represent a cut $(S, V \setminus S)$ by a vector $\boldsymbol{x} \in \mathbb{R}^n$, where

$$x_i \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases},$$

Then the number of edges across the cut is $\sum_{\{u,v\} \in E} (x_u - x_v)^2$. Since every homogeneous polynomial of degree 2 can be realized as $\boldsymbol{x}^T M \boldsymbol{x}$. In studying cuts in a graph $G = (V, E)$, we will want to choose a matrix L such that

$$\boldsymbol{x}^T L \boldsymbol{x} = \sum_{\{u,v\} \in E} (x_u - x_v)^2$$

where the boolean vector $\boldsymbol{x} \in \{0, 1\}^V$ represents a cut in the graph. Then the right-hand side would represent the number of edges that cross the cut.

A matrix that satisfies this property would be the Laplacian matrix of G, defined by $D - A$. Hence, we will obtain the following expression:

$$\boldsymbol{x}^T (D - A) \boldsymbol{x} = \sum_{\{u,v\} \in E} (x_u - x_v)^2.$$

This could be verified by showing that both the left-hand side and right-hand side are equal to

$$\sum_v d_v x_v^2 - 2 \sum_{\{u,v\} \in E} x_u x_v.$$

The quantity we are interested in is called the **edge expansion** (also known as the Cheeger constant or the isoperimetric constant), given by

$$\phi(G) \stackrel{\text{def}}{=} \min_{S \subset V; |S| \leq n/2} \frac{|E(S, V \setminus S)|}{|E(S)|}.$$

Continuing in the vein of the previous linear algebraic definition then for our vector x, we can define the edge expansion as :

$$\Phi(G) \stackrel{\text{def}}{=} \min_{\boldsymbol{x}: x_i \in \{0,1\}; |x|_1 \leq n/2} \frac{\boldsymbol{x}^\top L \boldsymbol{x}}{\boldsymbol{x}^\top D \boldsymbol{x}}$$

The objective we are minimising above is closely related to the Rayleigh quotient. In particular for a symmetric matrix $M \in \mathbb{R}^{n \times n}$, The Rayleigh quotient of $\boldsymbol{x}$ with respect to M is defined as follows:

$$R_M(\boldsymbol{x}) = \frac{\boldsymbol{x}^\top M \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}$$

The Rayleigh quotient is useful in min-max theorems to get the exact values of all eigenvalues. Clearly then for a d regular graph our optimization problem is the same as optimizing the Rayleigh quotient with respect to L under certain constraints on the vector x. In fact in general using the fact that D is invertible, we have that

$$\Phi(G) \stackrel{\text{def}}{=} \min_{\boldsymbol{x}:x_i\in\{0,1\};|x|_1\le n/2} \frac{\boldsymbol{x}^\top L\boldsymbol{x}}{\boldsymbol{x}^\top D\boldsymbol{x}} = \min_{\boldsymbol{x}:x_i\in\{0,1\};|x|_1\le n/2} \frac{\boldsymbol{x}^\top DLD^{-1/2}\boldsymbol{x}}{\boldsymbol{x}^\top D\boldsymbol{x}}.$$

An obvious relaxation then would be to consider the naive relaxation of the problem to

$$\min_{\boldsymbol{x}} \frac{\boldsymbol{x}^\top L\boldsymbol{x}}{\boldsymbol{x}^\top D\boldsymbol{x}}.$$

We then get that

$$\Phi(G) = \min_{\boldsymbol{x}} \frac{\boldsymbol{x}^\top L\boldsymbol{x}}{\boldsymbol{x}^\top D\boldsymbol{x}} = \min_{\boldsymbol{x}} \frac{\boldsymbol{x}^\top D^{-1/2}LD^{-1/2}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{x}}.$$

This relaxation then reduces our problem to minimizing the Rayleigh quotient with respect to $D^{-1/2}LD^{-1/2}$. We define or denote $D^{-1/2}LD^{-1/2}$ as a normalized Laplacian $\tilde{L}$. Unfortunately however by itself this relaxation fails for the simple reason that the all 1 vector causes this quantity to become 0. However if we put the additional constraint that $x \perp 1$, then things do kind of work out. This result is called the Cheeger's inequality which we will explore later.

Before we reach there we should remind ourselves of some fundamentals of Linear Algebra.

Theorem 1.1 (Spectral Theorem)**.** *Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix with real-valued entries, then there are n real numbers, which may not be distinct, $\lambda_1, ..., \lambda_n$ and n orthonormal real vectors $\boldsymbol{x}_1, ..., \boldsymbol{x}_n, \boldsymbol{x}_i \in \mathbb{R}^n$ such that $\boldsymbol{x}_i$ is an eigenvector of λ_i.*

Proof sketch. The fundamental theorem of algebra states that every polynomial has at least one complex root. Using the fact that M is a real symmetric matrix, we can then conclude that the eigenvalues of M are real. Hence, M must have a real eigenvalue λ_1 with real eigenvector $\boldsymbol{v}_1$. We can then show that M maps vectors that are orthogonal to $\boldsymbol{v}_1$ to vectors that are orthogonal to $\boldsymbol{v}_1$. We could then use induction to expand this to all n. □

For undirected graphs, the Adjacency matrix is symmetric. Hence we could apply the Spectral Theorem.

Now we will introduce a variational characterization of eigenvalues for real symmetric matrices.

Theorem 1.2. *Let $M \in \mathbb{R}^{n\times n}$ be a real symmetric matrix, and $\lambda_1 \le \lambda_2 \le ... \le \lambda_n$ be the eigenvalues of M in non-increasing order. Then,*

$$\lambda_k = \min_{k-dim\ V} \max_{\boldsymbol{x}\in V-\{0\}} \frac{\boldsymbol{x}^T M\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}$$

where V is a k-dimensional subspace of $\mathbb{R}^n$

Proof. Let $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ be orthonormal eigenvectors associated with the eigenvalues $\lambda_1, ..., \lambda_n$ as guaranteed by the Spectral Theorem. For every $\boldsymbol{x} = \sum_{i=1}^{k} a_i \boldsymbol{v_i}$ in such a space, the numerator of the Rayleigh quotient is given by:

$$\begin{aligned}\sum_{i,j} a_i a_j \boldsymbol{v}_i^T M \boldsymbol{v}_j &= \sum_{i,j} a_i a_j \lambda_j \boldsymbol{v}_i^T \boldsymbol{v}_j \\ &= \sum_{i=1}^{k} a_i^2 \lambda_i \\ &\leq \lambda_k \cdot \sum_{i=1}^{k} a_i^2\end{aligned}$$

By the same argument, the denominator is $\sum_{i=1}^{k} a_i^2$. Hence, $R_M(\boldsymbol{x}) \leq \lambda_k$. Thus,

$$\lambda_k \geq \min_{k-\dim\ V} \max_{\boldsymbol{x} \in V - \{0\}} \frac{\boldsymbol{x}^T M \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}$$

To prove the other direction, we first let V be any k-dimensional subspace. We want to show that V must contain a vector of Rayleigh quotient $\geq \lambda_k$. Let S be the span of $\boldsymbol{v_k}, ..., \boldsymbol{v_n}$. Since S has dimension of $n-k+1$ and V has dimension k, they will have some non-zero vector $\boldsymbol{x}$ in common. Similarly, we can write $\boldsymbol{x} = \sum_{i=k}^{n} a_i \boldsymbol{v}_i$. The numerator of the Rayleigh quotient is then

$$\sum_{i=k}^{n} \lambda_i a_i^2 \geq \lambda_k \sum_i a_i^2$$

Since the denominator is $\sum_i a_i^2$, $R_M(\boldsymbol{x}) \geq \lambda_k$. ☐

This theorem gives us the following consequences:

Corollary 1.3. *If M is a real symmetric matrix and $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ be its eigenvalues then :*

$$\begin{aligned}\lambda_1 &= \min_{\boldsymbol{x} \neq \boldsymbol{0}} R_M(\boldsymbol{x}) \\ \lambda_2 &= \min_{\boldsymbol{x} \neq \boldsymbol{0}, \boldsymbol{x} \perp \boldsymbol{x}_1} R_M(\boldsymbol{x}) \\ \lambda_n &= \max_{\boldsymbol{x} \neq \boldsymbol{0}} R_M(\boldsymbol{x})\end{aligned}$$

We now state the relation of the eigenvalues of $\tilde{L}$ with the graph G.

Theorem 1.4. *Let G be an undirected graph, D and A be the degree and adjacncecy matrices of G respectively, and $\tilde{L} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ be the normalized Laplacian matrix of G. Let $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ be real eigenvalues of $\tilde{L}$. Then*

1. $\lambda_1 = 0$ *and* $\lambda_n \leq 2$

2. $\lambda_k = 0$ *if and only if* G *has at least* k *connected components.*

3. $\lambda_n = 2$ *if and only if at least one of the connected components of* G *is bipartite.*

Proof sketch. Note that the Rayleigh quotient of $D^{1/2}\boldsymbol{x}$ with respect to $\tilde{L}$ is $\frac{\boldsymbol{x}^T D^{1/2} \tilde{L} D^{1/2} \boldsymbol{x}}{\boldsymbol{x}^T D \boldsymbol{x}}$. Clearly, the denominator and numerator are both positive. Hence,
$\lambda_1 = \min_{\boldsymbol{x} \neq \boldsymbol{0}} R_{\tilde{L}}(D^{1/2}\boldsymbol{x}) \geq 0$
If we take $\boldsymbol{x} = \boldsymbol{1}$, then the Rayleigh quotient is 0. Hence 0 is the smallest eigenvalue of $\tilde{L}$. Thus $\lambda_1 = 0$
Using the variational characterization as described earlier, as well as the quadratic form of the numerator, we can express

$$\begin{aligned}
\lambda_k &= \min_{k-\text{dim } X} \max_{\boldsymbol{x} \in X - \{0\}} \frac{\boldsymbol{x}^T D^{1/2} \tilde{L} D^{1/2} \boldsymbol{x}}{\boldsymbol{x}^T D \boldsymbol{x}} \\
&= \min_{k-\text{dim } X} \max_{\boldsymbol{x} \in X - \{0\}} \frac{\boldsymbol{x}^T (D - A) \boldsymbol{x}}{\boldsymbol{x}^T D \boldsymbol{x}} \\
&= \min_{k-\text{dim } X} \max_{\boldsymbol{x} \in X - \{0\}} \frac{\sum_{(u,v) \in E} (x_u - x_v)^2}{\boldsymbol{x}^T D \boldsymbol{x}}
\end{aligned}$$

We can thus deduce that the multiplicity of zero is equal to the number of connected components by considering both directions.
Finally, we have

$$\begin{aligned}
\lambda_n &= \max_{\boldsymbol{x} \neq \boldsymbol{0}} R_{\tilde{L}}(\boldsymbol{x}) \\
&= \max_{\boldsymbol{x} \neq \boldsymbol{0}} R_{\tilde{L}}(D^{1/2}\boldsymbol{x}) \\
&= 2 - \min_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\sum_{(u,v) \in E} (x_u + x_v)^2}{\boldsymbol{x}^T D \boldsymbol{x}}
\end{aligned}$$

We observe that $2\boldsymbol{x}^T\boldsymbol{x} - \boldsymbol{x}^T L \boldsymbol{x} = \frac{1}{d} \sum_{(u,v) \in E} (x_u + x_v)^2$ in order to get from line 2 to line 3. Thus, $\lambda_n \leq 2$ and $\lambda_n = 2$ if and only if one of the connected components of G is bipartite. □

For simplicity sake, we will consider the $d-$regular case. If $G = (V, E)$ is an undirected $d-$regular graph, and $S \subseteq V$ is a set of vertices, then we call

$$\phi(S) \stackrel{\text{def}}{=} \frac{E(S, V \setminus S)}{d|S|}$$

the **edge expansion of** S. The quantity $\phi(S)$ is the average fraction of neighbours outside of S for a random element of S, and it compares the actual number of edges crossing the cut $(S, V \setminus S)$ with the trivial upper bound $d|S|$.

Algorithm 1: Fiedler's Algorithm

Input: graph $G(V, E)$, vector $\boldsymbol{x} \in \mathbb{R}^V$
Output: $S \subseteq V$ such that $\phi(S, V \setminus S) \leq 2\sqrt{\phi(G)}$

Sort the vertices according to the values x_v, and let $v_1, ..., v_n$ be the sorted order.

Find a k that minimizes $\phi(\{v_1, ..., v_k\}, \{v_{k+1}, ..., v_n\})$ and output such a cut.

We define the **edge expansion of a cut** $(S, V \setminus S)$ as

$$\phi(S, V \setminus S) \stackrel{\text{def}}{=} \max\{\phi(S), \phi(V \setminus S)\} = \frac{E(S, V \setminus S)}{d \cdot \min\{|S|, |V \setminus S|\}}$$

The **edge expansion of a graph** G is defined as

$$\phi(G) \stackrel{\text{def}}{=} \min_S \phi(S, V \setminus S) = \min_{S: 1 \leq |S| \leq \frac{|V|}{2}} \phi(S)$$

Finding cuts of small expansion is a problem of interest that have many applications. It is an open question whether there is a polynomial-time approximation with a constant-factor approximation ratio.

Algorithm 1 is an algorithm that was proposed by Fiedler, and it works well in practices when $\boldsymbol{x}$ is the eigenvector of λ_2. Fiedler's algorithm could be implemented in $O(|E| + |V| \log |V|)$ time, because it takes $O(|V| \log |V|)$ time to sort the vertices, and the cut of minimal expansion can be found in $O(E)$ time. To see why we could output a cut with edge expansion upper bounded by $2\sqrt{\phi(G)}$, we will introduce the Cheeger's Inequality.

Theorem 1.5 (Cheeger's Inequality)**.** *[1] Let G be an undirected regular graph and $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ be the eigenvalues of the normalized Laplacian $\tilde{L}$, with repetitions, then*

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$$

Furthermore, if $(S, V \setminus S)$ is the cut found by Fiedler's algorithm given the eigenvector of λ_2, then

$$\phi(S, V \setminus S) \leq \sqrt{2\lambda_2}$$

From Theorem 1.5, it follows that if $(S, V \setminus S)$ is the cut found by Fiedler's algorithm given an eigenvector of λ_2, then we have

$$\phi(S, V \setminus S) \leq 2\sqrt{\phi(G)}$$

We will break up the proof of Cheeger's Inequality. First, we start with the lower bound.

1.1 Proof of $\frac{\lambda_2}{2} \leq \phi(G)$

Let S be a set of vertices such that $\phi(S, V \setminus S) = \phi(G)$. For every set S, the expansion of S is the same as the Rayleigh quotient of the indicator vector $\mathbf{1}_S$. The indicator vector of a set S is the boolean vector where the v^{th} coordinate of $\mathbf{1}_S$ is 1 if and only if $v \in S$. Hence

$$R_{\tilde{L}}(\mathbf{1}_S) \leq \phi(G)$$
$$R_{\tilde{L}}(\mathbf{1}_{V \setminus S}) \leq \phi(G)$$

From the variational characterization of eigenvalues, we have

$$\lambda_2 = \min_{2-\text{dim } X} \max_{\boldsymbol{x} \in X - \{0\}} R_{\tilde{L}}(\boldsymbol{x})$$

We prove $\lambda_2 \leq 2\phi(G)$ by showing that all the vectors in the 2-dimensional space X of linear combinations of the orthogonal vectors $\mathbf{1}_S$, $\mathbf{1}_{V \setminus S}$ have Rayleigh quotient at most $2\phi(G)$. This is a consequence of the following Lemma.

Lemma 1.6. *Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two orthogonal vectors and let M be a positive semidefinite matrix. Then*

$$R_M(\boldsymbol{x} + \boldsymbol{y}) \leq 2 \cdot \max\{R_M(\boldsymbol{x}), R_M(\boldsymbol{y})\}$$

Proof. Let $0 \leq \lambda_1 \leq ... \leq \lambda_n$ be eigenvalues of M and $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ be the corresponding eigenvectors. Writing $\boldsymbol{x} = \sum_i a_i \boldsymbol{v}_i$ and $\boldsymbol{y} = \sum_i b_i \boldsymbol{v}_i$, the Rayleigh quotient of $\boldsymbol{x} + \boldsymbol{y}$ is

$$\begin{aligned}
\frac{\sum_i \lambda_i (a_i + b_i)^2}{||\boldsymbol{x} + \boldsymbol{y}||^2} &= \frac{\sum_i \lambda_i (a_i + b_i)^2}{||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2} \quad (||\boldsymbol{x} + \boldsymbol{y}||^2 = ||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2 \text{ by orthogonality}) \\
&\leq \frac{\sum_i 2\lambda_i (a_i^2 + b_i^2)}{||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2} \quad \text{(By Cauchy-Schwarz inequality)} \\
&= \frac{2R_M(\boldsymbol{x}) \cdot ||\boldsymbol{x}||^2 + 2R_M(\boldsymbol{y}) \cdot ||\boldsymbol{y}||^2}{||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2} \\
&\leq 2 \max\{R_M(\boldsymbol{x}), R_M(\boldsymbol{y})\}
\end{aligned}$$

Note that the Cauchy-Schwarz inequality we used is that $(a+b)^2 \leq 2a^2 + 2b^2$. □

Since $\mathbf{1}$ is an eigenvector for 0, which is the smallest eigenvalue of L of G, from the variational characterization of eigenvalues, we have

$$\lambda_2 = \min_{\boldsymbol{x} \perp \mathbf{1}} \frac{\sum_{\{u,v\} \in E} (x_u - x_v)^2}{d \sum_v x_v^2}$$

and any eigenvector $\boldsymbol{x}$ of λ_2 is a minimizer of the above expression. We will prove the next part of Cheeger's Inequality by showing the following stronger result:

Lemma 1.7. *Let $\boldsymbol{x}$ be a vector orthogonal to $\mathbf{1}$ and let $(S, V \setminus S)$ be the cut found by Fiedler's algorithm given $\boldsymbol{x}$. Then*

$$\phi(S, V \setminus S) \leq \sqrt{2R_{\tilde{L}}(\boldsymbol{x})}$$

This stronger result is useful as one often runs Fiedler's algorithm on an approximate eigenvector, and this Lemma does not require $\boldsymbol{x}$ to be an eigenvector, as long as its Rayleigh quotient is small.

To prove Lemma 1.7, we introduce the next 2 lemmas.

Lemma 1.8. *Let $\boldsymbol{x} \in \mathbb{R}^V$ be orthogonal to $\mathbf{1}$. Then there is a vector $\boldsymbol{y} \in \mathbb{R}^V_{\geq 0}$ with at most $|V|/2$ non-zero entries such that*

$$R_{\tilde{L}}(\boldsymbol{y}) \leq R_{\tilde{L}}(\boldsymbol{x})$$

Futhermore, for every $0 < t \leq \max_v(y_v)$, the cut $(\{v : y_v \geq t\}, \{v : y_v < t\})$ is one of the cuts considered by Fiedler's algorithm on input $\boldsymbol{x}$.

Proof. We first observe that for every constant c,

$$R_{\tilde{L}}(\boldsymbol{x} + c\mathbf{1}) \leq R_{\tilde{L}}(\boldsymbol{x})$$

because the numerator of $R_{\tilde{L}}(\boldsymbol{x}+c\mathbf{1})$ and the numerator of $R_{\tilde{L}}(\boldsymbol{x})$ are the same, and the denominator of $R_{\tilde{L}}(\boldsymbol{x} + c\mathbf{1})$ is $||\boldsymbol{x} + c\mathbf{1}||^2 = ||\boldsymbol{x}||^2 + ||c\mathbf{1}||^2 \geq ||\boldsymbol{x}||^2$.

Let m be the median value of the entries of $\boldsymbol{x}$, and let $\boldsymbol{x}' \stackrel{\text{def}}{=} \boldsymbol{x} - m\mathbf{1}$. Then $R_{\tilde{L}}(\boldsymbol{x}') \leq R_{\tilde{L}}(\boldsymbol{x})$, and the median of the entries of $\boldsymbol{x}'$ is zero. This means that $\boldsymbol{x}'$ has at most $|V|/2$ positive entries and at most $|V|/2$ negative entries.

Defining the following:

$$x_v^+ \stackrel{\text{def}}{=} \begin{cases} x_v' & \text{if } x_v' > 0 \\ 0 & \text{otherwise} \end{cases}, \quad x_v^- \stackrel{\text{def}}{=} \begin{cases} -x_v' & \text{if } x_v' < 0 \\ 0 & \text{otherwise} \end{cases}$$

We then have

$$\boldsymbol{x}' = \boldsymbol{x}^+ + \boldsymbol{x}^-$$

Note that $\boldsymbol{x}^+$ and $\boldsymbol{x}^-$ are orthogonal, non-negative and each of them has at most $|V|/2$ nonzero entries. Also for every t, the cuts defined by the set $\{v : x_v^+ \geq t\}$ is one of the cuts considered by Fiedler's algorithm on input $\boldsymbol{x}$, because it is the cut

$$(\{v : x_v < t + m\}, \{v : x_v \geq t + m\})$$

Similarly, for every t, the cut defined by the set $\{v : x_v^- \geq t\}$ is also one of the cuts considered, because it is the cut

$$(\{v : x_v \leq m - t\}, \{v : x_v > m - t\})$$

It remains to show that at least one of $\boldsymbol{x}^+$ or $\boldsymbol{x}^-$ has Rayleigh quotient smaller than or equal to the Rayleigh quotient of $\boldsymbol{x}'$, and hence smaller than equal to the Rayleigh quotient of $\boldsymbol{x}$. We claim

$$\begin{aligned} R_{\tilde{L}}(\boldsymbol{x}') &= \frac{\sum_{\{u,v\}}(x_u - x_v)^2}{||\boldsymbol{x}'||^2} \\ &= \frac{\sum_{\{u,v\}}((x_u^+ - x_v^+) - (x_u^- - x_v^-))^2}{||\boldsymbol{x}^+||^2 + ||\boldsymbol{x}^-||^2} \\ &\geq \frac{\sum_{\{u,v\}}(x_u^+ - x_v^+)^2 + (x_u^- - x_v^-)^2}{||\boldsymbol{x}^+||^2 + ||\boldsymbol{x}^-||^2} \\ &= \frac{R_{\tilde{L}}(\boldsymbol{x}^+)\cdot||\boldsymbol{x}^+||^2 + R_{\tilde{L}}(\boldsymbol{x}^-)\cdot||\boldsymbol{x}^-||^2}{||\boldsymbol{x}^+||^2 + ||\boldsymbol{x}^-||^2} \\ &\geq \min\{R_{\tilde{L}}(\boldsymbol{x}^+), R_{\tilde{L}}(\boldsymbol{x}^-)\} \end{aligned}$$

We just need to justify that for every edge $\{u, v\}$ we have

$$((x_u^+ - x_v^+) - (x_u^- - x_v^-))^2 \geq (x_u^+ - x_v^+)^2 + (x_u^- - x_v^-)^2$$

If $\{u, v\}$ is an edge between two non-positive vertices, or between two non-negative vertices, then the expression is equal. If it is an edge between a positive vertex u and a negative vertex v, then the left-hand side equation is equal to $(x_u^+ + x_v^-)^2$ and the right-hand side is equal to $(x_u^+)^2 + (x_v^-)^2$. □

Lemma 1.9. *Let $\boldsymbol{y} \in \mathbb{R}^V_{\geq 0}$ be a vector with non-negative entries. Then there is a $0 < t \leq \max_v\{y_v\}$ such that*

$$\phi(\{v : y_v \geq t\}) \leq \sqrt{2R_{\tilde{L}}(\boldsymbol{y})}$$

Proof. We will provide a probabilistic proof for this. Since the Rayleigh quotient is scalar-invariant, without loss of generality we can assume $\max_v y_v = 1$. We consider the probabilistic process where we pick $t > 0$ such that t^2 is uniformly distributed in $[0, 1]$. Defining the non-empty subset $S_t = \{v : y_v \geq t\}$. We claim that

$$\frac{\mathbb{E}(E(S_t, V \setminus S_t))}{\mathbb{E}(d|S_t|)} \leq \sqrt{2R_{\tilde{L}}(\boldsymbol{y})}$$

Lemma 1.9 follows from this claim because of Lemma 1.10.
On the denominator, we see that

$$\mathbb{E}\left[d|S_t|\right] = d \cdot \sum_{v\in V} \mathbb{P}\left[v \in S_t\right] = d\sum_v y_v^2$$

because

$$\mathbb{P}\left[v \in S_t\right] = \mathbb{P}\left[y_v \geq t\right] = \mathbb{P}\left[y_v^2 \geq t^2\right] = y_v^2$$

Now to bound the numerator, we say that an edge is cut by S_t if one endpoint is in S_t and another is not. We have

$$\begin{aligned}\mathbb{E}[E(S_t, V \setminus S)] &= \sum_{\{u,v\}\in E} \mathbb{P}[\{u,v\} \text{ is cut}] \\ &= \sum_{\{u,v\}\in E} |y_v^2 - y_u^2| \\ &= \sum_{\{u,v\}\in E} |y_v - y_u| \cdot (y_u + y_v)\end{aligned}$$

Applying Cauchy-Schwarz inequality,

$$\mathbb{E}[E(S_t, V \setminus S)] \le \sqrt{\sum_{\{u,v\}\in E} (y_v - y_u)^2} \cdot \sqrt{\sum_{\{u,v\}\in E} (y_v + y_u)^2}$$

Applying Cauchy-Schwarz inequality again in the form of $(a+b)^2 \le 2a^2 + 2b^2$, we have

$$\sum_{\{u,v\}\in E} (y_v + y_u)^2 \le \sum_{\{u,v\}\in E} 2y_v^2 + 2y_u^2 = 2d \sum_v y_v^2$$

Combining everything together, we have

$$\frac{\mathbb{E}(E(S_t, V \setminus S_t))}{\mathbb{E}(d|S_t|)} \le \sqrt{2\frac{\sum_{\{u,v\}\in E}(y_v - y_u)^2}{d\sum_v y_v^2}}$$

□

Lemma 1.10. *Let X and Y be random variables such that $\mathbb{P}[Y > 0] = 1$. Then*

$$\mathbb{P}\left[\frac{X}{Y} < \frac{\mathbb{E}(X)}{\mathbb{E}(Y)}\right] > 0$$

Proof sketch. Let $r := \frac{\mathbb{E}(X)}{\mathbb{E}(Y)}$. We can then use the linearity of expectation to prove this. □

Finally, to prove Lemma 1.7, we have to use Lemma 1.8 and Lemma 1.9.

Proof of Lemma 1.7. Let $\boldsymbol{x}$ be orthogonal to $\mathbf{1}$, let $(S_F, V \setminus S_F)$ be the cut found by Fiedler's algorithm given $\boldsymbol{x}$. Let $\boldsymbol{y}$ be the non-negative vector with at most $|V|/2$ positive such that $R_{\tilde{L}}(\boldsymbol{y}) \le R_{\tilde{L}}(\boldsymbol{x})$ as seen in Lemma 1.8. Then, by Lemma 1.9, we can let $0 < t \le max_v\{y_v\}$ be a threshold such that

$$\phi(\{v : y_v \ge t\}) \le \sqrt{2R_{\tilde{L}}(\boldsymbol{y})} \le \sqrt{2R_{\tilde{L}}(\boldsymbol{x})}$$

The set $S_t := \phi(\{v : y_v \ge t\})$ contains at most $|V|/2$ vertices, and the cut $(S_t, V \setminus S_t)$ is one of the cuts considered by Fiedler's algorithm on input $\boldsymbol{x}$, so

$$\phi(S_F, V \setminus S_F) \le \phi(S_t, V \setminus S_t) = \phi(S_t) \le \sqrt{2R_L(\boldsymbol{x})}$$

□

References

[1] Luca Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. *University of California, Berkeley, https://people. eecs. berkeley. edu/luca/books/expanders-2016. pdf*, 2017.