CMSE 890-001: Spectral Graph Theory and Related Topics, MSU, Spring 2021

Lecture 03: Eigenvalues, Optimization, and Connectivity

January 26, 2021

Lecturer: Matthew Hirn

6 Eigenvalues and optimization: The Courant-Fischer Theorem

The *Rayleigh quotient* of a vector $\boldsymbol{x}$ with respect to a matrix $\boldsymbol{M}$ is

$$\text{Rayleigh quotient} = \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} .$$

An important fact is that the Rayleigh quotient is that the Rayleigh quotient of an eigenvector is its eigenvalue. That is, if $\boldsymbol{M}\boldsymbol{\psi} = \mu\boldsymbol{\psi}$, then

$$\frac{\boldsymbol{\psi}^T \boldsymbol{M} \boldsymbol{\psi}}{\boldsymbol{\psi}^T \boldsymbol{\psi}} = \mu .$$

Imagine now that you want to maximize the Rayleigh quotient of some symmetric matrix $\boldsymbol{M}$. The Courant-Fischer Theorem tells us that the maximum will be the largest eigenvalue of $\boldsymbol{M}$, and the vector that achieves this maximum will be the corresponding eigenvector. In fact, it will characterize every eigenvalue of $\boldsymbol{M}$.

Theorem 3 (Courant-Fischer Theorem)**.** *Let $\boldsymbol{M}$ be an $n \times n$ symmetric, real valued matrix with eigenvalues $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$. Then,*

$$\mu_k = \max_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S) = k}} \min_{\substack{\boldsymbol{x} \in S \\ \boldsymbol{x} \neq \boldsymbol{0}}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \min_{\substack{T \subseteq \mathbb{R}^n \\ \dim(T) = n-k+1}} \max_{\substack{\boldsymbol{x} \in T \\ \boldsymbol{x} \neq \boldsymbol{0}}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} ,$$

where the outer max *and* min *are over subspaces S and T of $\mathbb{R}^n$.*

We will need the following lemma to prove Theorem 3.

Lemma 4. *Let $\boldsymbol{M}$ be an $n \times n$ symmetric, real valued matrix with eigenvalues $\mu_1, \ldots, \mu_n$ and corresponding eigenvectors $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n$. Then for any $\boldsymbol{x} \in \mathbb{R}^n$,*

$$\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} = \sum_{i=1}^{n} \mu_i |\langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle|^2 .$$

Proof. Since $\boldsymbol{M}$ is a real valued, symmetric matrix we can apply the Spectral Theorem (Theorem 1). Let $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n$ be the orthonormal eigenvectors of $\boldsymbol{M}$ with corresponding eigenvalues $\mu_1, \ldots, \mu_n$. Since we have n such eigenvectors, they form an orthonormal basis (ONB) for $\mathbb{R}^n$, meaning that we may write

$$\boldsymbol{x} = \sum_{i=1}^{n} \langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle \boldsymbol{\psi}_i \quad \text{and} \quad \|\boldsymbol{x}\|^2 = \sum_{i=1}^{n} |\langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle|^2 \,. \tag{6}$$

Equation (6) is a standard fact from Linear Algebra that we will use a lot in this course; if you do not remember it and/or do not remember why it is true, please go find your favorite Linear Algebra book and look it up ☺.

The lemma will follow from (6) and the following calculation:

$$\begin{aligned}
\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} &= \langle \boldsymbol{x}, \boldsymbol{M}\boldsymbol{x} \rangle \\
&= \left\langle \sum_{i=1}^{n} \langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle \boldsymbol{\psi}_i, \boldsymbol{M} \sum_{j=1}^{n} \langle \boldsymbol{x}, \boldsymbol{\psi}_j \rangle \boldsymbol{\psi}_j \right\rangle \\
&= \left\langle \sum_{i=1}^{n} \langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle \boldsymbol{\psi}_i, \sum_{j=1}^{n} \langle \boldsymbol{x}, \boldsymbol{\psi}_j \rangle \boldsymbol{M} \boldsymbol{\psi}_j \right\rangle \\
&= \left\langle \sum_{i=1}^{n} \langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle \boldsymbol{\psi}_i, \sum_{j=1}^{n} \langle \boldsymbol{x}, \boldsymbol{\psi}_j \rangle \mu_j \boldsymbol{\psi}_j \right\rangle \\
&= \sum_{i,j=1}^{n} \mu_j \langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle \langle \boldsymbol{x}, \boldsymbol{\psi}_j \rangle \underbrace{\langle \boldsymbol{\psi}_i, \boldsymbol{\psi}_j \rangle}_{\delta(i-j)} \\
&= \sum_{i=1}^{n} \mu_i |\langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle|^2 \,.
\end{aligned}$$

□

Proof of Theorem 3. We will prove the second formulation, as the proof of the first formulation can be found in [1, Chapter 2]. To start, let

$$T = \operatorname{span}\{\boldsymbol{\psi}_k, \ldots, \boldsymbol{\psi}_n\} \,.$$

We can expand any $\boldsymbol{x} \in T$ as

$$\boldsymbol{x} = \sum_{i=k}^{n} \langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle \boldsymbol{\psi}_i \,.$$

Furthermore, using Lemma 4 and (6),

$$\frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{\sum_{i=k}^{n} \mu_i |\langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle|^2}{\sum_{j=k}^{n} |\langle \boldsymbol{x}, \boldsymbol{\psi}_j \rangle|^2} \leq \frac{\mu_k \sum_{i=k}^{n} |\langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle|^2}{\sum_{j=k}^{n} |\langle \boldsymbol{x}, \boldsymbol{\psi}_j \rangle|^2} = \mu_k \,.$$

Therefore,

$$\max_{\substack{\boldsymbol{x}\in T\\ \boldsymbol{x}\neq \boldsymbol{0}}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \leq \mu_k \,.$$

Thus we have an upper bound.

To prove equality, we will show that for all subspaces T of dimension $n-k+1$ we have

$$\max_{\substack{\boldsymbol{x}\in T\\ \boldsymbol{x}\neq \boldsymbol{0}}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \mu_k \,.$$

To do so, let

$$S = \operatorname{span}\{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k\} \,.$$

Since S has dimension k, any T of dimension $n-k+1$ has an intersection with S of dimension at least one, i.e., $\dim(S \cap T) \geq 1$. Therefore

$$\max_{\substack{\boldsymbol{x}\in T\\ \boldsymbol{x}\neq \boldsymbol{0}}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \max_{\substack{\boldsymbol{x}\in S\cap T\\ \boldsymbol{x}\neq \boldsymbol{0}}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \min_{\substack{\boldsymbol{x}\in S\cap T\\ \boldsymbol{x}\neq \boldsymbol{0}}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \min_{\substack{\boldsymbol{x}\in S\\ \boldsymbol{x}\neq \boldsymbol{0}}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \,.$$

Since any $\boldsymbol{x} \in S$ can be written as

$$\boldsymbol{x} = \sum_{i=1}^{k} \langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle \boldsymbol{\psi}_i \,,$$

we have, using Lemma 4 and (6) again, for any $\boldsymbol{x} \in S$,

$$\frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{\sum_{i=1}^{k} \mu_i |\langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle|^2}{\sum_{j=1}^{k} |\langle \boldsymbol{x}, \boldsymbol{\psi}_j \rangle|^2} \geq \frac{\mu_k \sum_{i=1}^{k} |\langle \boldsymbol{x}, \boldsymbol{\psi}_i \rangle|^2}{\sum_{j=1}^{k} |\langle \boldsymbol{x}, \boldsymbol{\psi}_j \rangle|^2} = \mu_k \,.$$

□

7 The Laplacian and connectivity

Recall the graph Laplacian is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{M}$. Let us think of $\boldsymbol{L}$ as an operator and compute $\boldsymbol{L}\boldsymbol{x}(a)$ at a vertex $a \in V$:

$$\begin{aligned}
\boldsymbol{L}\boldsymbol{x}(a) &= \boldsymbol{D}\boldsymbol{x}(a) - \boldsymbol{M}\boldsymbol{x}(a) \\
&= \boldsymbol{d}(a)\boldsymbol{x}(a) - \sum_{b\in N(a)} w(a,b)\boldsymbol{x}(b) \\
&= \left(\sum_{b\in N(a)} w(a,b) \right) \boldsymbol{x}(a) - \sum_{b\in N(a)} w(a,b)\boldsymbol{x}(b) \\
&= \sum_{b\in N(a)} w(a,b)(\boldsymbol{x}(a) - \boldsymbol{x}(b)) \,. \qquad (7)
\end{aligned}$$

Equation (7) gives us an alternate way of writing $\boldsymbol{Lx}(a)$. We immediately see from (7) that

$$\boldsymbol{L}\mathbf{1} = \mathbf{0}\,,$$

which verifies this fact that we had used in the proof of Theorem 2. Thus $\lambda_1 = 0$ is always the smallest eigenvalue of $\boldsymbol{L}$ and it has corresponding eigenvector $\mathbf{1}$. The next theorem shows that whether λ_2 is zero or not corresponds to whether G is connected or not; it is our first theorem relating spectral properties of $\boldsymbol{L}$ to the structure of G.

Theorem 5. *Let $G = (V, E, w)$ be a weighted graph, and let $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of its graph Laplacian $\boldsymbol{L}$. Then, $\lambda_2 > 0$ if and only if G is connected.*

Proof. "If and only if" means we need to prove two things:

1. $\lambda_2 > 0$ implies G is connected;
2. G is connected implies $\lambda_2 > 0$.

Let us start with the first statement. First note that it is equivalent to "G is disconnected implies $\lambda_2 = 0$." Let us prove this statement. Since G is disconnected, it can be written as the union of two subgraphs, $G = G_1 \cup G_2$, where there are no edges going between G_1 and G_2; see also Figure 13. Order the vertices of G so that ones are from G_1 and the last ones

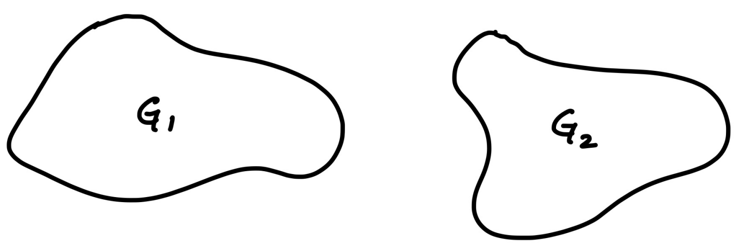

Figure 13: Since G is disconnected, it can be partitioned into two subgraphs G_1 and G_2 that have no edges between them.

are from G_2. Since there are no edges going between G_1 and G_2, we can write the graph Laplacian of G as:

$$\boldsymbol{L}_G = \begin{pmatrix} \boldsymbol{L}_{G_1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{L}_{G_2} \end{pmatrix},$$

where in the above equation $\mathbf{0}$ represents a sub-matrix of zeros ($\mathbf{0}$ will alternate between meaning a vectors of zeros and a matrix of zeros; the context should make clear which interpretation to use). Therefore $\boldsymbol{L}_G$ has two independent eigenvectors both with eigenvalue zero:

$$\boldsymbol{\psi}_1 = \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\psi}_2 = \begin{pmatrix} \mathbf{0} \\ \mathbf{1} \end{pmatrix},$$

where $\boldsymbol{\psi}_1$ is constant on G_1 and zero on G_2 and $\boldsymbol{\psi}_2$ is zero on G_1 and constant on G_2.

Now for the second statement. Let $\boldsymbol{\psi}$ be an eigenvector of $\boldsymbol{L}$ with eigenvalue zero,

$$\boldsymbol{L}\boldsymbol{\psi} = \mathbf{0}\,.$$

Therefore:

$$\boldsymbol{\psi}^T \boldsymbol{L}\boldsymbol{\psi} = \sum_{(a,b)\in E} w(a,b)(\boldsymbol{\psi}(a) - \boldsymbol{\psi}(b))^2 = 0\,.$$

Thus, for every $a, b \in V$ such that $(a,b) \in E$ (i.e., every pair of vertices connected by an edge), we have $\boldsymbol{\psi}(a) = \boldsymbol{\psi}(b)$. Now fix an $a \in V$ and let $b \in V$ be any vertex in G. Since G is connected there exists a path in G from a to b; that is, there is a sequence of vertices:

$$a = v_1, v_2, \ldots, v_m = b \quad \text{such that} \quad (v_i, v_{i+1}) \in E, \;\; \forall\, 1 \leq i \leq m-1\,.$$

Therefore:

$$\boldsymbol{\psi}(a) = \boldsymbol{\psi}(v_1) = \boldsymbol{\psi}(v_2) = \cdots = \boldsymbol{\psi}(v_m) = \boldsymbol{\psi}(b)\,.$$

Thus $\boldsymbol{\psi}$ is constant on G and we conclude that the only eigenvectors of $\boldsymbol{L}$ with eigenvalue zero are constant vectors. Thus there is only one independent eigenvector with eigenvalue zero, and so $\lambda_2 > 0$. □

Remark 7. In fact we can strengthen Theorem 5 and show that the multiplicity of the zero eigenvalue is equal to the number of connected components of G.

Using similar techniques we can also prove that an eigenvector $\boldsymbol{\psi}$ of zero eigenvalue must be constant on each of the connected components of G.

Theorem 6. *Let $G = (V, E, w)$ be a weighted graph and let $\boldsymbol{\psi}$ be an eigenvector of its graph Laplacian with eigenvalue 0, i.e., $\boldsymbol{L}\boldsymbol{\psi} = 0$. Then $\boldsymbol{\psi}$ must be constant on each of the connected components of G.*

Proof. As in the proof of Theorem 5, since $\boldsymbol{L}\boldsymbol{\psi} = \mathbf{0}$ we have

$$0 = \boldsymbol{\psi}^T \boldsymbol{L}\boldsymbol{\psi} = \sum_{(a,b)\in E} w(a,b)(\boldsymbol{\psi}(a) - \boldsymbol{\psi}(b))^2\,,$$

and it follows that $\boldsymbol{\psi}(a) = \boldsymbol{\psi}(b)$ for all $(a,b) \in E$. Now let $a, b \in V$ and suppose that a and b are in the same connected component of G. Since they are in the same connected component, there is a path from a to b, i.e.,

$$a = v_1, v_2, \ldots, v_m = b \quad \text{such that} \quad (v_i, v_{i+1}) \in E, \;\; \forall\, 1 \leq i \leq m-1\,.$$

Therefore:

$$\boldsymbol{\psi}(a) = \boldsymbol{\psi}(v_1) = \boldsymbol{\psi}(v_2) = \cdots = \boldsymbol{\psi}(v_m) = \boldsymbol{\psi}(b)\,,$$

and it follows that $\boldsymbol{\psi}$ is constant on the connected component. □

Remark 8. If we want to cluster our data in terms of connected components, Remark 7 and Theorem 6 tells us how to do so. We can compute the graph Laplacian $\boldsymbol{L}$ and then compute its eigenvalues and eigenvectors. We then check the multiplicity of the zero eigenvalue, which gives us the number of connected components. If the multiplicity of the zero eigenvalue is k, we can use $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k$ to find the connected components of G, although it might not be so easy that each eigenvector corresponds to the indicator function on a connected component. Nevertheless, since each $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k$ has to be constant on each connected component, and since the eigenvectors can be taken to be orthogonal, we can combine the information from each eigenvector to figure out the clusters.

Here is an example with four clusters. Suppose our graph G has 20 vertices and 4 connected components G_1, G_2, G_3, and G_4, all of which have 5 vertices. Then the following four eigenvectors are orthogonal and all have eigenvalue zero:

$$\boldsymbol{\psi}_1 = \mathbf{1}\,, \quad \boldsymbol{\psi}_2 = \mathbf{1}_{G_1 \cup G_2} - \mathbf{1}_{G_3 \cup G_4}\,, \quad \boldsymbol{\psi}_3 = \mathbf{1}_{G_1} - \mathbf{1}_{G_2}\,, \quad \boldsymbol{\psi}_4 = \mathbf{1}_{G_3} - \mathbf{1}_{G_4}\,.$$

Clearly $\boldsymbol{\psi}_1$ tells us nothing. The second eigenvector, $\boldsymbol{\psi}_2$, is useful and separates $G_1 \cup G_2$ from $G_3 \cup G_4$, but does not separate G_1 from G_2 nor G_3 from G_4. The third eigenvector, though, separates G_1 from G_2, but still tells us nothing for separating G_3 from G_4. However, the last eigenvector, $\boldsymbol{\psi}_4$, allows us to separate G_3 from G_4.

A way to automate the example is the following. Suppose the multiplicity of the zero eigenvalue is k. Then we will use $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k$ to to embed each vertex into $\mathbb{R}^k$ via:

$$a \mapsto (\boldsymbol{\psi}_1(a), \ldots, \boldsymbol{\psi}_k(a)) \in \mathbb{R}^k\,.$$

Those vertices with the same k-tuples are in the same cluster, and those with different k-dimensional embeddings are in different clusters. In the above example we have:

$$\begin{aligned} a \in G_1 &\mapsto (1, 1, 1, 0) \\ a \in G_2 &\mapsto (1, 1, -1, 0) \\ a \in G_3 &\mapsto (1, -1, 0, 1) \\ a \in G_4 &\mapsto (1, -1, 0, -1) \end{aligned}$$

Remark 9. Later on in the course we will show, quantitatively, that even if G is connected, the size of $\lambda_2 > 0$ will indicate how well connected G is! We can use this idea to cluster G even when it has one connected component.

References

[1] Daniel A. Spielman. Spectral and algebraic graph theory. Book draft, available at: http://cs-www.cs.yale.edu/homes/spielman/sagt/, 2019.

[2] Michael Perlmutter, Feng Gao, Guy Wolf, and Matthew Hirn. Geometric scattering networks on compact Riemannian manifolds. In *Proceedings of The First Mathematical and Scientific Machine Learning Conference, Proceedings of Machine Learning Research*, volume 107, pages 570–604, 2020.