Spectral Graph Theory Lecture 2

Essential spectral theory, Hall's spectral graph drawing, the Fiedler value

Daniel A. Spielman August 31, 2018

2.1 Eigenvalues and Optimization

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

$$\frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

At the end of the last class, I gave the following characterization of the largest eigenvalue of a symmetric matrix in terms of the Rayleigh quotient.

Theorem 2.1.1. *Let $\boldsymbol{M}$ be a symmetric matrix and let $\boldsymbol{x}$ be a non-zero vector that maximizes the Rayleigh quotient with respect to $\boldsymbol{M}$:*

$$\frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

Then, $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{M}$ with eigenvalue equal to the Rayleigh quotient. Moreover, this eigenvalue is the largest eigenvalue of $\boldsymbol{M}$.

Proof. We first observe that the maximum is achieved: As the Rayleigh quotient is homogeneous, it suffices to consider unit vectors $\boldsymbol{x}$. As the set of unit vectors is a closed and compact set, the maximum is achieved on this set.

Now, let $\boldsymbol{x}$ be a non-zero vector that maximizes the Rayleigh quotient. We recall that the gradient of a function at its maximum must be the zero vector. Let's compute that gradient.

We have

$$\nabla \boldsymbol{x}^T \boldsymbol{x} = 2\boldsymbol{x},$$

and

$$\nabla \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} = 2\boldsymbol{M}\boldsymbol{x}.$$

So,

$$\nabla \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{(\boldsymbol{x}^T \boldsymbol{x})(2\boldsymbol{M}\boldsymbol{x}) - (\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x})(2\boldsymbol{x})}{(\boldsymbol{x}^T \boldsymbol{x})^2}.$$

In order for this to be zero, we must have

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

That is, if and only if $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{M}$ with eigenvalue equal to its Rayleigh quotient. $\square$

Of course, the analogous characterization holds for the smallest eigenvalue. A substantial generalization of these characterizations is given by the Courant-Fischer Theorem. We will state it for the Laplacian, as that is the case we will consider for the rest of the lecture.

Theorem 2.1.2 (Courant-Fischer Theorem)**.** *Let $\boldsymbol{L}$ be a symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then,*

$$\lambda_k = \min_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S)=k}} \max_{\boldsymbol{x} \in S} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \max_{\substack{T \subseteq \mathbb{R}^n \\ \dim(T)=n-k+1}} \min_{\boldsymbol{x} \in T} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

For example, consier the case $k = 1$. In this case, S is just the span of $\boldsymbol{\psi}_1$ and T is all of $\mathbb{R}^n$. For general k, the proof reveals that the optimum is achieved when S is the span of $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k$ and when T is the span of $\boldsymbol{\psi}_k, \ldots, \boldsymbol{\psi}_n$.

As many proofs in Spectral Graph Theory begin by expanding a vector in the eigenbasis of a matrix, we being by carefully stating a key property of these expansions.

Lemma 2.1.3. *Let $\boldsymbol{M}$ be a symmetric matrix with eigenvalues $\mu_1, \ldots, \mu_n$ and a corresponding orthnormal basis of eigenvectors $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n$. Let $\boldsymbol{x}$ be a vector and expand $\boldsymbol{x}$ in the eigenbasis as*

$$\boldsymbol{x} = \sum_{i=1}^{n} c_i \boldsymbol{\psi}_i.$$

Then,

$$\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} = \sum_{i=1}^{n} c_i^2 \lambda_i.$$

You should check for yourself (or recall) that $c_i = \boldsymbol{x}^T \boldsymbol{\psi}_i$ (this is obvious if you consider the standard coordinate basis).

Proof. Compute:

$$\begin{aligned}
\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} &= \left(\sum_i c_i \boldsymbol{\psi}_i \right)^T \boldsymbol{M} \left(\sum_j c_j \boldsymbol{\psi}_j \right) \\
&= \left(\sum_i c_i \boldsymbol{\psi}_i \right)^T \left(\sum_j c_j \lambda_j \boldsymbol{\psi}_j \right) \\
&= \sum_{i,j} c_i c_j \lambda_j \boldsymbol{\psi}_i^T \boldsymbol{\psi}_j \\
&= \sum_i c_i^2 \lambda_i,
\end{aligned}$$

as $\boldsymbol{\psi}_i^T \boldsymbol{\psi}_j = 0$ for $i \neq j$. $\square$

Proof of 2.1.2. Let $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n$ be an orthonormal set of eigenvectors of $\boldsymbol{L}$ corresponding to $\lambda_1, \ldots, \lambda_n$. We will just verify the first characterization of λ_k. The other is similar.

First, let's verify that λ_k is achievable. Let S_k be the span of $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k$. We can expand every $\boldsymbol{x} \in S_k$ as

$$\boldsymbol{x} = \sum_{i=1}^{k} c_i \boldsymbol{\psi}_i.$$

Applying Lemma 2.1.3 we obtain

$$\frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{\sum_{i=1}^{k} \lambda_i c_i^2}{\sum_{i=1}^{k} c_i^2} \leq \frac{\sum_{i=1}^{k} \lambda_k c_i^2}{\sum_{i=1}^{k} c_i^2} = \lambda_k.$$

To show that this is in fact the maximum, we will prove that for all subspaces S of dimension k,

$$\max_{\boldsymbol{x} \in S} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \lambda_k.$$

Let T_k be the span of $\boldsymbol{\psi}_k, \ldots, \boldsymbol{\psi}_n$. As T_k has dimension $n - k + 1$, every S of dimension k has an intersection with T_k of dimension at least 1. So,

$$\max_{\boldsymbol{x} \in S} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \max_{\boldsymbol{x} \in S \cap T_k} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

Any such $\boldsymbol{x}$ may be expressed as

$$\boldsymbol{x} = \sum_{i=k}^{n} c_i \boldsymbol{\psi}_i,$$

and so

$$\frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{\sum_{i=k}^{n} \lambda_i c_i^2}{\sum_{i=k}^{n} c_i^2} \geq \frac{\sum_{i=k}^{n} \lambda_k c_i^2}{\sum_{i=k}^{n} c_i^2} = \lambda_k.$$

□

We give one last characterization of the eigenvalues and eigenvectors of a symmetric matrix. Its proof is similar, so we will save it for an exercise.

Theorem 2.1.4. *Let $\boldsymbol{L}$ be an $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ with corresponding eigenvectors $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n$. Then,*

$$\lambda_i = \min_{\boldsymbol{x} \perp \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{i-1}} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}},$$

and the eigenvectors satisfy

$$\boldsymbol{\psi}_i = \arg \min_{\boldsymbol{x} \perp \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{i-1}} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

2.2 Drawing with Laplacian Eigenvalues

I will now explain the motivation for the pictures of graphs that I drew last lecture using the Laplacian eigenvalues. Well, the real motivation was just to convince you that eigenvectors are cool. The following is the technical motivation. It should come with the caveat that it does not produce nice pictures of all graphs. In fact, it produces bad pictures of most graphs. But, it is still the first thing I always try when I encounter a new graph that I want to understand.

This approch to using eigenvectors to draw graphs was suggested by Hall [Hal70] in 1970.

To explain Hall's approach, I'll begin by describing the problem of drawing a graph on a line. That is, mapping each vertex to a real number. It isn't easy to see what a graph looks like when you do this, as all of the edges sit on top of one another. One can fix this either by drawing the edges of the graph as curves, or by wrapping the line around a circle.

Let $\boldsymbol{x} \in \mathbb{R}^V$ be the vector that describes the assignment of a real number to each vertex. We would like most pairs of vertices that are neighbors to be close to one another. So, Hall suggested that we choose an $\boldsymbol{x}$ minimizing

$$\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}. \tag{2.1}$$

Unless we place restrictions on $\boldsymbol{x}$, the solution will be degenerate. For example, all of the vertices could map to 0. To avoid this, and to fix the scale of the embedding overall, we require

$$\sum_{a \in V} \boldsymbol{x}(a)^2 = \|\boldsymbol{x}\|^2 = 1. \tag{2.2}$$

Even with this restriction, another degenerate solution is possible: it could be that every vertex maps to $1/\sqrt{n}$. To prevent this from happening, we add the additional restriction that

$$\sum_{a} \boldsymbol{x}(a) = \mathbf{1}^T \boldsymbol{x} = 0. \tag{2.3}$$

On its own, this restriction fixes the shift of the embedding along the line. When combined with (2.2), it guarantees that we get something interesting.

As $\mathbf{1}$ is the eigenvector of the 0 eigenvalue of the Laplacian, the nonzero vectors that minimize (2.1) subject to (2.2) and (2.3) are the unit eigenvectors of the Laplacian of eigenvalue λ_2.

Of course, we really want to draw a graph in two dimensions. So, we will assign two coordinates to each vertex given by $\boldsymbol{x}$ and $\boldsymbol{y}$. As opposed to minimizing (2.1), we will minimize

$$\sum_{(a,b) \in E} \left\| \begin{pmatrix} \boldsymbol{x}(a) \\ \boldsymbol{y}(a) \end{pmatrix} - \begin{pmatrix} \boldsymbol{x}(b) \\ \boldsymbol{y}(b) \end{pmatrix} \right\|^2.$$

This turns out not to be so different from minimizing (2.1), as it equals

$$\sum_{(a,b) \in E} (\boldsymbol{x}(a) - \boldsymbol{x}(b))^2 + (\boldsymbol{y}(a) - \boldsymbol{y}(b))^2 = \boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x} + \boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y}.$$

As before, we impose the scale conditions

$$\|\boldsymbol{x}\|^2 = 1 \quad \text{and} \quad \|\boldsymbol{y}\|^2 = 1,$$

and the centering constraints

$$\mathbf{1}^T \boldsymbol{x} = 0 \quad \text{and} \quad \mathbf{1}^T \boldsymbol{y} = 0.$$

However, this still leaves us with the degnerate solution $\boldsymbol{x} = \boldsymbol{y} = \boldsymbol{\psi}_2$. To ensure that the two coordinates are different, Hall introduced the restriction that $\boldsymbol{x}$ be orthogonal to $\boldsymbol{y}$. One can use the spectral theorem to prove that the solution is then given by setting $\boldsymbol{x} = \boldsymbol{\psi}_2$ and $\boldsymbol{y} = \boldsymbol{\psi}_3$, or by taking a rotation of this solution (this is a problem on the first problem set).

2.3 Isoperimetry and λ_2

Computer Scientists are often interested in cutting, partitioning, and clustering graphs. Their motivations range from algorithm design to data analysis. We will see that the second-smallest eigenvalue of the Laplacian is intimately related to the problem of dividing a graph into two pieces without cutting too many edges.

Let S be a subset of the vertices of a graph. One way of measuring how well S can be separated from the graph is to count the number of edges connecting S to the rest of the graph. These edges are called the *boundary* of S, which we formally define by

$$\partial(S) \stackrel{\text{def}}{=} \{(a, b) \in E : a \in S, b \notin S\}.$$

We are less interested in the total number of edges on the boundary than in the ratio of this number to the size of S itself. For now, we will measure this in the most natural way–by the number of vertices in S. We will call this ratio the *isoperimetric ratio* of S, and define it by

$$\theta(S) \stackrel{\text{def}}{=} \frac{|\partial(S)|}{|S|}.$$

The *isoperimetric number* of a graph is the minimum isoperimetric number over all sets of at most half the vertices:

$$\theta_G \stackrel{\text{def}}{=} \min_{|S| \leq n/2} \theta(S).$$

We will now derive a lower bound on θ_G in terms of λ_2. We will present an upper bound, known as Cheeger's Inequality, in a later lecture.

Theorem 2.3.1. *For every $S \subset V$*

$$\theta(S) \geq \lambda_2(1 - s),$$

where $s = |S| / |V|$. In particular,

$$\theta_G \geq \lambda_2/2.$$

Proof. As

$$\lambda_2 = \min_{\boldsymbol{x}:\boldsymbol{x}^T\mathbf{1}=0} \frac{\boldsymbol{x}^T \boldsymbol{L}_G \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}},$$

for every non-zero $\boldsymbol{x}$ orthogonal to $\mathbf{1}$ we know that

$$\boldsymbol{x}^T \boldsymbol{L}_G \boldsymbol{x} \geq \lambda_2 \boldsymbol{x}^T \boldsymbol{x}.$$

To exploit this inequality, we need a vector related to the set S. A natural choice is $\boldsymbol{\chi}_S$, the characteristic vector of S,

$$\boldsymbol{\chi}_S(a) = \begin{cases} 1 & \text{if } a \in S \\ 0 & \text{otherwise.} \end{cases}$$

We find

$$\boldsymbol{\chi}_S^T \boldsymbol{L}_G \boldsymbol{\chi}_S = \sum_{(a,b)\in E} (\boldsymbol{\chi}_S(a) - \boldsymbol{\chi}_S(b))^2 = \left|\partial(S)\right|.$$

However, $\boldsymbol{\chi}_S$ is not orthogonal to $\mathbf{1}$. To fix this, use

$$\boldsymbol{x} = \boldsymbol{\chi}_S - s\mathbf{1},$$

so

$$\boldsymbol{x}(a) = \begin{cases} 1 - s & \text{for } a \in S \text{, and} \\ -s & \text{otherwise.} \end{cases}$$

We have $\boldsymbol{x}^T\mathbf{1} = 0$, and

$$\boldsymbol{x}^T \boldsymbol{L}_G \boldsymbol{x} = \sum_{(a,b)\in E} ((\boldsymbol{\chi}_S(a) - s) - (\boldsymbol{\chi}_S(b) - s))^2 = \left|\partial(S)\right|.$$

To finish the proof, we compute

$$\boldsymbol{x}^T \boldsymbol{x} = |S|\,(1-s)^2 + (|V| - |S|)s^2 = |S|\,(1 - 2s + s^2) + |S|\,s - |S|\,s^2 = |S|\,(1-s).$$

This gives

$$\lambda_2 \leq \frac{\boldsymbol{\chi}_S^T \boldsymbol{L}_G \boldsymbol{\chi}_S}{\boldsymbol{\chi}_S^T \boldsymbol{\chi}_S} = \frac{|\partial(S)|}{|(S)|\,(1-s)}.$$

□

This theorem says that if λ_2 is big, then G is very well connected: the boundary of every small set of vertices is at least λ_2 times something just slightly smaller than the number of vertices in the set.

We will use the computation in the last line of that proof often, so we will make it a claim.

Claim 2.3.2. *Let $S \subseteq V$ have size $s\,|V|$. Then*

$$\left\|\chi_S - s\mathbf{1}\right\|^2 = s(1-s)\,|V|.$$

2.4 Exercises

The following exercises are for your own practice. They are intended as a review of fundamental linear algebra. I will put the solutions in a separate file that you can find on Canvas. I recommend that you try to solve all of these before you look at the solutions, so that you can get back in practice at doing linear algebra.

1. Characterizing Eigenvalues.

Prove Theorem 2.1.4.

2. Traces.

Recall that the trace of a matrix $\boldsymbol{A}$, written $\mathrm{Tr}(\boldsymbol{A})$, is the sum of the diagonal entries of $\boldsymbol{A}$. Prove that for two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,

$$\mathrm{Tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{Tr}(\boldsymbol{B}\boldsymbol{A}).$$

Note that the matrices **do not** need to be square for this to be true. They can be rectangular matrices of dimensions $n \times m$ and $m \times n$.

Use this fact and the previous exercise to prove that

$$\mathrm{Tr}(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i,$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\boldsymbol{A}$. You are probably familiar with this fact about the trace, or it may have been the definition you were given. This is why I want you to remember how to prove it.

3. The Characteristic Polynomial

Let $\boldsymbol{M}$ be a symmetric matrix. Recall that the eigenvalues of $\boldsymbol{M}$ are the roots of the characteristic polynomial of $\boldsymbol{M}$:

$$p(x) \stackrel{\text{def}}{=} \det(x\boldsymbol{I} - \boldsymbol{M}) = \prod_{i=1}^{n}(x - \mu_i).$$

Write

$$p(x) = \sum_{k=0}^{n} x^{n-k} c_k (-1)^k.$$

Prove that

$$c_k = \sum_{S \subseteq [n], |S|=k} \det(\boldsymbol{M}(S,S)).$$

Here, we write $[n]$ to denote the set $\{1, \ldots, n\}$, and $\boldsymbol{M}(S,S)$ to denote the submatrix of $\boldsymbol{M}$ with rows and columns indexed by S.

4. Reversing products.

Let $\boldsymbol{M}$ be a d-by-n matrix. Prove that the multiset of nonzero eigenvalues of $\boldsymbol{M}\boldsymbol{M}^T$ is the same as the multiset of nonzero eigenvalues of $\boldsymbol{M}^T\boldsymbol{M}$.

References

[Hal70] K. M. Hall. An r-dimensional quadratic placement algorithm. *Management Science*, 17:219–229, 1970.