Spectral Graph Theory and Algorithms, Summer 2017	20 June 2017
Lecture 2	
Lecturer: David P. Williamson	

1 Rayleigh Quotients

Recall that last time we stated the following theorem.

Theorem 1 *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be its eigenvalues and $x_1, x_2, \cdots, x_n$ be orthonormal vectors (e.g. $\|x_i\|^2 = 1$, $\langle x_i, x_j\rangle = 0 \ \ \forall i \neq j$) such that $Ax_i = \lambda_i x_i$ for $i = 1, 2, \cdots, n$. Then, for all $0 \leq k \leq n-1$,*

$$\lambda_{k+1} = \min_{x\in\mathbb{R}^n : x \perp span(x_1,\cdots,x_k)} \frac{x^TAx}{x^Tx}$$

and any minimizer is the associated eigenvector.

The expression $\frac{x^TAx}{x^Tx}$ is called the *Rayleigh quotient*. One reason this theorem is very useful is that it allows us to get an upper bound on λ_{k+1}: the Rayleigh quotient of x for any $x \perp \text{span}(x_1, \cdots, x_k)$ yields an upper bound. We will be using this technique for bounding eigenvalues *ad nauseum.*

At the end of last time we proved the following lemma.

Lemma 2 *Let $A \in \mathbb{R}^{n\times n}$ be symmetric and $k \leq n-1$. Let $x_1, \cdots, x_k$ be orthogonal eigenvectors of A. Then there exists an eigenvector x_{k+1} orthogonal to $x_1, \cdots, x_k$.*

Now we can use it to prove our main theorem.

Proof of Theorem 1: Given eigenvalues $\lambda_1, \cdots, \lambda_k$ and associated vectors $x_1, \cdots, x_k$, we can use Lemma 2 repeatedly to find orthonormal eigenvectors $x_{k+1}, \cdots, x_n$. We sort the remaining eigenvalues so that $\lambda_{k+1} \leq \cdots \leq \lambda_n$. Note that

$$\frac{x_{k+1}^T A x_{k+1}}{x_{k+1}^T x_{k+1}} = \frac{\lambda_{k+1}(x_{k+1}^T x_{k+1})}{x_{k+1}^T x_{k+1}} = \lambda_{k+1}.$$

Consider any other feasible solution x. Let V be the subspace containing all $y \in \mathbb{R}^n$ such that $y \perp \text{span}(x_1, x_2, \cdots, x_k)$. Then $x_{k+1}, \cdots, x_n$ is a basis of V. Assume that $x = \alpha_{k+1}x_{k+1} + \cdots + \alpha_n x_n$. We have

$$\begin{aligned}
\frac{x^TAx}{x^Tx} &= \frac{x^T(\alpha_{k+1}\lambda_{k+1}x_{k+1} + \cdots + \alpha_n\lambda_n x_n)}{x^Tx} \\
&= \frac{(\alpha_{k+1}x_{k+1} + \cdots + \alpha_n x_n)^T(\alpha_{k+1}\lambda_{k+1}x_{k+1} + \cdots + \alpha_n\lambda_n x_n)}{(\alpha_{k+1}x_{k+1} + \cdots + \alpha_n x_n)^T(\alpha_{k+1}x_{k+1} + \cdots + \alpha_n x_n)} \\
&= \frac{\alpha_{k+1}^2\lambda_{k+1} + \cdots + \alpha_n^2\lambda_n}{\alpha_{k+1}^2 + \cdots + \alpha_n^2} \\
&\geq \lambda_{k+1}\frac{\alpha_{k+1}^2 + \cdots + \alpha_n^2}{\alpha_{k+1}^2 + \cdots + \alpha_n^2} \\
&= \lambda_{k+1}.
\end{aligned}$$

[0]This lecture was drawn from Trevisan, *Lecture Notes on Expansion, Sparsest Cut, and Spectral Graph Theory*, Chapter 1, Lau's 2015 lecture notes, Lecture 1:`https://cs.uwaterloo.ca/~lapchi/cs798/notes/L01.pdf`, and Spielman's 2012 lecture notes, Lecture 3: `http://www.cs.yale.edu/homes/spielman/561/2012/lect03-12.pdf` Notes were scribed by Qinru Shi and Wilson Yoo.

Hence,

$$\lambda_{k+1} = \min_{x\in\mathbb{R}^n : x\perp \text{span}(x_1,\cdots,x_k)} \frac{x^TAx}{x^Tx}.$$

□

In fact, we can further extend Theorem 1 and reach the following conclusion.

Theorem 3 *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Let $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ be its eigenvalues and $x_1, x_2, \cdots, x_n$ be associated orthonormal eigenvectors. Then,*

$$\begin{aligned}
\lambda_k &= \min_{x\in\mathbb{R}^n : x\perp span(x_1,\cdots,x_{k-1})} \frac{x^TAx}{x^Tx} \\
&= \min_{x\in\mathbb{R}^n : x\in span(x_k,\cdots,x_n)} \frac{x^TAx}{x^Tx} \\
&= \max_{x\in\mathbb{R}^n : x\perp span(x_{k+1},\cdots,x_n)} \frac{x^TAx}{x^Tx} \\
&= \max_{x\in\mathbb{R}^n : x\in span(x_1,\cdots,x_k)} \frac{x^TAx}{x^Tx}.
\end{aligned}$$

We leave the proof of Theorem 3 as an exercise.

Some special cases of the theorem are

$$\lambda_n = \max_{x\in\mathbb{R}^n} \frac{x^TAx}{x^Tx}$$

and

$$\lambda_1 = \min_{x\in\mathbb{R}^n} \frac{x^TAx}{x^Tx}.$$

2 Some identities

We now give a sequence of identities that will be useful to us as we prove various statements about eigenvalues and graphs.

Since $x_1, \cdots, x_n$ are orthonormal, for any $x \in \mathbb{R}^n$, we can write x as

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n.$$

Then,

$$\langle x, x_i\rangle = \langle \alpha_1 x_1 + \cdots + \alpha_n x_n, x_i\rangle = \alpha_i \langle x_i, x_i\rangle = \alpha_i.$$

Therefore,

$$\begin{aligned}
x &= \langle x, x_1\rangle x_1 + \cdots \langle x, x_n\rangle x_n \\
&= x_1(x_1^Tx) + \cdots + x_n(x_n^Tx) \\
&= (x_1x_1^T + \cdots + x_nx_n^T)x
\end{aligned}$$

for all $x \in \mathbb{R}^n$. Hence,

$$x_1x_1^T + \cdots + x_nx_n^T = I. \tag{1}$$

By (1),

$$Ax = AIx = A(x_1x_1^T + \cdots + x_nx_n^T)x = (\lambda_1 x_1x_1^T + \cdots + \lambda_n x_nx_n^T)x,$$

so

$$A = \lambda_1 x_1x_1^T + \cdots + \lambda_n x_nx_n^T. \tag{2}$$

We know that A^{-1} exists iff all eigenvalues of A are non-zero. Also,

$$(\lambda_1 x_1 x_1^T + \cdots + \lambda_n x_n x_n^T)(\frac{1}{\lambda_1} x_1 x_1^T + \cdots + \frac{1}{\lambda_n} x_n x_n^T) = x_1 x_1^T + \cdots x_n x_n^T = I.$$

Thus,

$$A^{-1} = \frac{1}{\lambda_1} x_1 x_1^T + \cdots + \frac{1}{\lambda_n} x_n x_n^T.$$

When A is singular, we define the *pseudo-inverse* of A analogously:

$$A^{\dagger} \equiv \sum_{i:\lambda_i \neq 0} \frac{1}{\lambda_i} x_i x_i^T.$$

One of the reasons that spectral graph theory has become an intense area of study in theoretical computer science in the last few years is that researchers (starting with Spielman and Teng) have shown how to compute $A^{\dagger}b$ quickly for some cases of A and b. This has led to further research on how to solve this product quickly plus additional research on what can be done with a quick solver of this type. We will hear more about this later in the term.

e first present a few useful eigenvalue identities. Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix with real eigenvalues $\lambda_1 \geq \lambda_2 \geq ...\lambda_n$ with corresponding eigenvectors $x_1, x_2, \ldots, x_n$ such that the x_i are orthonormal.

Lemma 4 *The eigenvectors of A^k are $x_1, \ldots, x_n$ with corresponding eigenvalues $\lambda_1^k, \ldots, \lambda_n^k$.*

Proof: Let

$$X = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}$$

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$$

Then,

$$AX = XD$$

since the columns of X are the eigenvectors. Interestingly, $X^T X = I$ as the columns are orthonormal, so $X^T = X^{-1}$. This implies $A = XDX^{-1}$ (right multiplying $AX = DX$ by X^{-1}). Then, $A^k = (XDX^{-1})^k = XD^kX^{-1}$, where D^k has the form

$$D^k = \begin{bmatrix} \lambda_1^k & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n^k \end{bmatrix}$$

It follows that $A^k X = XD^k$; therefore, the eigenvectors of A^k are $x_1, \ldots, x_n$ with corresponding eigenvalues $\lambda_1^k, \ldots, \lambda_n^k$. $\square$

For our next few identities, we need the following fact that we present without proof.

Fact 1 $\det(AB) = \det(A)\det(B)$

An easy corollary of this fact is $\det(A^{-1}) = \frac{1}{\det(A)}$ since:

$$\det(A)\det(A^{-1}) = \det(AA^{-1}) = \det(I) = 1$$

Lemma 5

$$\det(A) = \prod_{i=1}^{n} \lambda_i.$$

Proof:

$$\begin{aligned}\det(A) &= \det(XDX^{-1}) \\ &= \det(X)\det(D)\det(X^{-1}) \\ &= \det(D) \\ &= \prod_{i=1}^{n} \lambda_i.\end{aligned}$$

□

Recall that in the first lecture, we defined the *trace* of A to be $\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}$. We used the following without proof in the first lecture, and now we can prove it.

Lemma 6

$$\operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i.$$

Proof: Consider the characteristic polynomial of A, which we defined in the first lecture to be $\det(\lambda I - A)$; it is a degree n polynomial in λ. Then we can rewrite it as:

$$\begin{aligned}\det(\lambda I - A) &= \det(\lambda X X^T - XDX^T) \\ &= \det(X(\lambda I - D)X^T) \\ &= \det(X)\det(\lambda I - D)\det(X^T) \\ &= \det(\lambda I - D) \\ &= \prod_{i=1}^{n} (\lambda - \lambda_i).\end{aligned}$$

We see that indeed the eigenvalues are precisely the roots of the polynomial. Here, the coefficient of λ^{n-1} is exactly $-\sum_{i=1}^{n} \lambda_i$.

Now, recall that the determinant of a matrix Z with permutations S_n is:

$$\det(Z) = \sum_{\sigma \in S_n} sgn(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}.$$

The only permutation that can produce a λ^{n-1} term is the identity, so the only term in the sum with a λ^{n-1} term is $\prod_{i=1}^{n}(\lambda - a_{ii})$. The coefficient of λ^{n-1} here is $-\sum_{i=1}^{n} a_{ii}$. Then, we are done, with

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i.$$

□

3 The Perron-Frobenius Theorem

Recall for an undirected graph G, its *adjacency matrix* is defined as $A = (a_{ij})$ where

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

Before we begin, we need the following lemma.

Lemma 7 *Let A be the adjacency matrix of a connected graph G. Suppose $x \geq 0$, $x \neq 0$, is an eigenvector of A. Then $x > 0$.*

Proof: Suppose otherwise and there exists a vertex j such that $x(j) = 0$ (where $x(j)$ is the jth component of x). Then we claim that there must exist an edge $(i,k) \in E$ such that $x(i) = 0$ and $x(k) > 0$. To see this, since $x \neq 0$ and $x \geq 0$, there must exist some vertex l such that $x(l) > 0$. Because G is connected, there is a path in G between j and l. It cannot be the case that for every edge (i,k) on the path that $x(i) = x(k) = 0$ (in particular, since this is not true of l, the endpoint of the path) or that $x(i) > 0$ and $x(k) > 0$ (since this is not true of i, the starting point of the path). So such an edge must exist.

Then

$$(Ax)(i) = \sum_{j:(i,j)\in E} x(j) \geq x(k) > 0.$$

However, since x is an eigenvector, it is also true that

$$(Ax)(i) = \lambda x(i) = 0,$$

which is a contradiction. Therefore $x > 0$. □

Theorem 8 (Perron-Frobenius) *Let G be a connected graph with adjacency matrix A, eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and corresponding eigenvectors $x_1, \ldots, x_n$. Then:*

(i) There exists an eigenvector $x_1 > 0$ (that is, every coordinate is strictly positive).

(ii) $\lambda_1 \geq -\lambda_n$;

(iii) $\lambda_1 > \lambda_2$;

Proof: First we will prove (i). Let $x_1, \ldots, x_n$ be the corresponding eigenvectors, and assume they are orthonormal. Recall that

$$\lambda_1 = \max_{x\in\mathbb{R}^n} \frac{x^T Ax}{x^T x} = x_1^T Ax_1.$$

Define a vector y such that $\forall i, y(i) = |x_1(i)|$; then $y^T y = x_1^T x_1 = 1$. We show that y is also an eigenvector corresponding to λ_1. To see this, we have

$$\begin{aligned}
\lambda_1 &= x_1^T Ax_1 \\
&= \sum_{ij} a_{ij} x_1(i) x_1(j) \\
&\leq \sum_{ij} a_{ij} |x_1(i)||x_1(j)| \\
&= \sum_{ij} a_{ij} y(i) y(j) \\
&= y^T Ay \\
&\leq \lambda_1.
\end{aligned}$$

The last inequality follows by the definition of λ_1, and by the fact that y has unit norm. Since $\lambda_1 = \lambda_1$, all the inequalities must be equalities, so y is an eigenvector of λ_1. Thus we have that $y \geq 0$ and $y \neq 0$, and so by Lemma 7, we know that $y > 0$. For the rest of the proof we assume that $x_1 > 0$.

Now we prove (ii). Let $\forall i$, $y(i) = |x_n(i)|$. Again, we have

$$y^T y = x_n^T x_n = 1$$

so the vector y has unit norm. Then

$$\begin{aligned}
|\lambda_n| &= |x_n^T A x_n| \\
&\leq \sum_{ij} a_{ij} |x_n(i)||x_n(j)| \\
&= \sum_{ij} a_{ij} y(i) y(j) \\
&= y^T A y \\
&\leq \lambda_1,
\end{aligned}$$

as desired.

For (iii), let $\forall i$, $y(i) = |x_2(i)|$. Then $y^y = x_2^T x_2 = 1$. Then we have that

$$\begin{aligned}
\lambda_2 &= x_2^T A x_2 \\
&\leq \sum_{ij} a_{ij} |x_2(i)||x_2(j)| \\
&= \sum_{ij} a_{ij} y(i) y(j) \\
&= y^T A y \\
&\leq \lambda_1,
\end{aligned}$$

Now we show that somewhere along the way, the inequality is strict. Since $x_1 > 0$ and $\langle x_1, x_2 \rangle = 0$, and both are nonzero, some of the entries of x_2 are positive and some are negative. We split into two cases:

- **Case 1:** All of the entries of x_2 are nonzero. Then, since G is connected, $\exists (i, j) \in E$ such that $x_2(i) < 0, x_2(j) > 0$ (by an argument similar to the one we used in the proof of Lemma 7). Then, $x_2(i)x_2(j) < |x_2(i)||x_2(j)|$ which gives us the strict inequality that we wanted. Hence, $\lambda_2 < \lambda_1$.
- **Case 2:** $x_2(i) = 0$ for some i. If all inequalities are equalities, y is an eigenvector of λ_1 with $y \geq 0$. But by Lemma 7, when $y \geq 0$ and is an eigenvector corresponding to λ_1 and G is connected, $y > 0$, which contradicts $x_2(i) = 0$ for some i.

□

4 Bipartite Graphs

We now turn to showing how all the various identities we've proven over this lecture and the last can be applied to showing something about the structure of graphs. In particular, we show that the spectrum of the adjacency matrix tells us whether the graph is bipartite or not.

Lemma 9 *If G is bipartite, and λ is an eigenvalue of adjacency matrix A, then so is $-\lambda$.*

Proof: If G is bipartite, we can re-index the nodes such that

$$A = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix}.$$

Then let $v = \begin{bmatrix} x \\ y \end{bmatrix}$ be an eigenvector of A with eigenvalue λ. Then, we have

$$A = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$$

Hence we have $By = \lambda x$ and $B^T x = \lambda y$ So from this,

$$\begin{aligned}
\begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \begin{bmatrix} x \\ -y \end{bmatrix} &= \begin{bmatrix} -By \\ B^T x \end{bmatrix} \\
&= \begin{bmatrix} -\lambda x \\ \lambda y \end{bmatrix} \\
&= -\lambda \begin{bmatrix} x \\ -y \end{bmatrix}
\end{aligned}$$

So, $-\lambda$ is an eigenvalue corresponding to the eigenvector $[x - y]^T$. □

We can now show that this is statement can be made an "if and only if": that is the graph G is bipartite if and only if for each eigenvalue λ there is another eigenvalue $-\lambda$.

Theorem 10 *If for each eigenvalue $\lambda \neq 0$ there is another eigenvalue $\lambda' = -\lambda$, then G is bipartite.*

Proof: Let k be any odd positive integer. Then by hypothesis,

$$\mathrm{tr}(A^k) = \sum_{i=1}^{n} \lambda_i^k = 0$$

It can be shown by induction that $(A^k)_{ij}$ is the number of walks from i to j of length exactly k (recall from the first lecture that we used that $(A^2)_{ij} = \sum_k a_{ik} a_{kj}$ is the number of walks of length exactly two, using an edge from i to k then k to j). Notice that if there is an odd cycle of length k, then it must be the case that $(A^k)_{ii} > 0$, so that $\mathrm{tr}(A^k) > 0$. But since $Tr(A^k) = 0$, there are no odd cycles of length k. Since this is true for any odd positive integer k, there are no odd cycles in G, which implies that G is bipartite. □

Now we can show something even stronger than the previous statement: we only need to look at the smallest and largest eigenvalue to know whether or not the graph is bipartite.

Theorem 11 *Suppose G is connected. Then, $\lambda_n = -\lambda_1$ if and only if G is bipartite.*

Proof: By Perron-Frobenius, $\lambda_1 \geq -\lambda_n$, and by the previous theorem, the graph being bipartite implies that $\lambda_1 = -\lambda_n$.

For the other direction, Let x_n be the eigenvector corresponding to λ_n with $x_n^T x_n$. Let $y(i) = |x_n(i)|$ for all i. Again, we have $y^T y = x_n^T x_n = 1$. Also,

$$\begin{aligned}
|\lambda_n| &= |x_n^T A x_n| \\
&\leq \sum_{ij} a_{ij} |x_n(i)||x_n(j)| \\
&= \sum_{ij} a_{ij} y_n(i) y_n(j) \\
&= y^T A y \\
&\leq \lambda_1.
\end{aligned}$$

The assumption $\lambda_n = -\lambda_1$ implies that all the inequalities are equalities. This implies that y is an eigenvector corresponding to λ_1, with $y \geq 0$. By Lemma 7, since $y \geq 0$, we have $y > 0$ and this implies that $x_n(i) \neq 0$ for all i.

If all the inequalities are equalities, the product $x_n(i)x_n(j)$ has the same sign whenever $a_{ij} > 0$. Since $\lambda_n = x_n^T A x_n < 0$, all of these products must be negative. This implies that for any edge in the graph, either $x_n(i) > 0, x_n(j) < 0$ or $x_n(i) < 0, x_n(j) > 0$. This induces the bipartition

$$V_1 = \{i : x_n(i) < 0\},$$
$$V_2 = \{i : x_n(i) < 0\}.$$

$\square$

This brings about an interesting research question. What happens when λ_1 is close to $-\lambda_n$? Does this mean that the graph is "almost bipartite", in the sense that if we remove some of the edges, it would become bipartite? Possibly the answer to this question is already well known.