Approximation Algorithms **Fall 2019**

Lecture 1: Introduction to Spectral Graph Theory

Scribe: Robert Wang Sept 20

Linear Algebra

It will be taken for granted that you are familiar with basic linear algebra concepts like vectors, matrices/linear transformations, vector spaces, inner product spaces, and so on. We will begin our exploration of Spectral Graph Theory by reviewing the spectral theorem. Hopefully, you have encountered this at least once before. Recall, that for a matrix $M \in \mathbb{R}^{n\times n}$, a vector $v \in \mathbb{R}^n$ is an eigenvector if:

$$Av = \lambda v$$

For some $\lambda \in \mathbb{R}$. The λ, in this case is an eigenvalue of M. A square matrix M is said to be diagonalizable, if it has n linearly independent eigenvectors. Note that if M has n distinct, eigenvalues, then it must be diagonalizable, as distinct eigenvalues must correspond to linearly independent eigenvectors. The converse, however, is not necessarily, true. Consider the case where $M = I$.

Theorem 1.1. *(**Real Spectral Theorem**) If a matrix $M \in \mathbb{R}^{n\times n}$ is symmetrical, then it is diagonalizable with n orthogonormal eigenvectors.*

Note that I say real because it's impossible to study the spectral theorem without talking about complex numbers, but since this is not a linear algebra reading group, I won't mention it here.

Spectral Decomposition

So why is it so important that M has orthonormal eigenvectors? Recall that n orthonormal vectors form a basis of $\mathbb{R}^n$. This means that any vector $x \in \mathbb{R}^n$ can be expressed as a linear combination of the eigenvectors of M.

$$x = \sum_{i=1}^{n} < x, v_i > v_i$$

If we think of M as a linear transformation, we can decompose M as follows:

$$\begin{aligned} Mx &= M(\sum_{i=1}^{n} < x, v_i > v_i) \\ &= \sum_{i=1}^{n} \lambda_i < x, v_i > v_i \\ M &= \sum_{i=1}^{n} \lambda_i v_i v_i^\top \end{aligned}$$

Thus, we can think of M as simply stretching the axis of x corresponding to v_i by a λ_i amount. We can express the summation of rank one matrices above in a more compact way as follows:

$$M = PDP^\top$$

Where D is the matrix with λ_i's down the diagonal, and P is the matrix with columns $v_1, ...v_n$. Note that $P^\top P = I$.

Positive Semi-definite (PSD) Matrices

A symmetric matrix M is said to be PSD, if $\forall x \in \mathbb{R}$, $x^\top Mx \geq 0$ and $x^\top Mx = 0$ iff $x = 0$. There are two other conditions that are equivalent to M being PSD:

- all eigenvalues of M are non-negative
- $M = UU^\top$ for some matrix U.

Proof. Suppose $\forall x \in \mathbb{R}^d$, $x^\top Ax = 0$. If A has a negative eigenvalue, that is, $\exists v \in \mathbb{R}^d$ such that $Av = \lambda v$, $\lambda < 0$, then $v^\top Av = \lambda ||v||_2^2 < 0$, which is a contradiciton. Thus, all eigenvalues of A are non-negative. Now suppose all eigenvalues of A are non-negative. This means $A = PDP^\top$ and $D^{1/2}$ exists, as each diagonal entry of D has a real square root. Let $U = PD^{1/2}$. Clearly, $A = UU^\top$. And finally, $\forall x \in \mathbb{R}$, we have $x^\top Ax = x^\top UU^\top x = ||Ux||_2^2 \geq 0$. $\square$

Note that if M is symmetric and PSD, then the operation:

$$< x, y >_M = x^\top My$$

is an inner product. Verification of this is left as an exercise to the reader.

Variational characterization of eigenvalues

For a symmetrical matrix, M, we define the **Rayleigh quotient** of a vector x, $R_M(x)$ as follows:

$$R_M(x) = \frac{x^\top Mx}{x^\top x}$$

We now reach one of the most useful results: Suppose we sort the eigenvalues of M so that $\lambda_1 \leq \lambda_2 ... \lambda_n$

$$\lambda_k = \min_{U:\, dimU=k} \max_{x \in U} \frac{x^\top Mx}{x^\top x}$$

Consider the dimension k subspace, $span\{v_1, v_2, ...v_k\}$. First, I claim that the maximum value of the Rayleigh quotient in this subspace is λ_k

Proof. Let $x \in span\{v_1, ...v_k\}$. We have:

$$\begin{aligned}
x^\top Mx &= x(\sum_{i=1}^{k} < v_i, x_i > v_i) \\
&= (\sum_{j=1}^{k} < v_j, x > v_j)\top(\sum_{i=1}^{k} \lambda_i < v_i, x_i > v_i) \\
&= \sum_{i,j \leq k} \lambda_i \lambda_j < v_i, x >< v_j, x >< v_i, v_j > \\
&= \sum_{i=1}^{k} \lambda_i < v_i, x >^2 \\
&\leq \lambda_k \sum_{i=1}^{k} < v_i, x >^2 \\
&= \lambda_k < x, x >
\end{aligned}$$

Thus, it follows that λ_k is the best we can do in this subspace. □

Next, I claim that all $k-$dimensional subspaces contain a vector with Rayleigh quotient at least λ_k.

Proof. Suppose by contradiction, that there is some k dimensional subspace U such that the Rayleigh quotients of all vectors is less than λ_k. This means, it cannot contain $v_n, v_{n-1}, ...v_k$. However, this means that there is at most $k-1$ linearly independent vectors in U, which is a contradiction. □

An immediate result of this is that

$$\lambda_k = \max_{U:dimU=n-k+1} \min_{x \in U} x^\top Mx$$

Which follows from that fact that $-\lambda_k$ is the $n-k+1^{th}$ largest eigenvalue of $-M$ and so

$$-\lambda_k = \min_{U:dimU=n-k+1} \max_{x \in U} x^\top Mx$$

Graphs and Matrices

Let $G = (V, E)$ be an undirected graph with n vertices and m edges. First, we define the adjacency matrix, A, as an $n \times n$ matrix with rows and columns corresponding to vertices such that:

$$A_{u,v} = \begin{cases} 1 & (u, v) \in E \\ 0 & \text{otherwise} \end{cases}$$

Notice A is symmetrical, which means that, it is orthogonally diagonalizable. For now, however, it is more convienient for us to work with another symmetrical matrix, called the **Laplacian** matrix. Let D be a diagonal matrix such that $D_{v,v} = deg(v)$. We define the Laplacian, L, as follows:

$$L = D - A$$

One of the reasons the Laplacian matrix is important is that it is very useful for solving graph partitioning problems. Suppose, we wanted a concise way of representing a cut, $S \subseteq V$, in a graph. Let $x \in \mathbb{R}^n$ be a vector such that:

$$x_v = \begin{cases} 1 & v \in S \\ 0 & \text{otherwise} \end{cases}$$

The numerator of the Rayleigh quotient of x with respect to the Laplacian is exactly equalled to the number of edges across the cut $S|V-S$. I will denote the number of edges across this cut as $\partial(S, V-S)$:

Proof. We can express the number of edges across the cut as $\sum_{u,v \in E}(x_u - x_v)^2$, since $(x_u - x_v)^2 = 1$ iff u and v are in the same different partitions. Thus, we have:

$$\begin{aligned}
\partial(V, V-S) &= \sum_{(u,v)\in E} (x_u - x_v)^2 \\
&= \sum_{(u,v)\in E} x_u^2 - 2x_u x_v + x_v^2 \\
&= \sum_{v \in V} deg(v) x_v^2 - \sum_{(u,v)\in E} 2x_u x_v \\
&= \sum_{v \in V} deg(v) x_v^2 - \sum_u \sum_v x_u x_v * 1\{(u,v) \in E\} \\
&= x^\top D x - x^\top A x \\
&= x^\top L x
\end{aligned}$$

□

Since for all $x \in \mathbb{R}^n$ $x^\top L x = \sum_{(u,v)\in E}(x_u - x_v)^2 \geq 0$, L is PSD. Another result of this is that $\vec{1}$ is always in the nullspace of L, as the partition that in includes every vertex obviously doesn't have any cut edges. Thus, we know that the smallest eigenvalue of L is 0, with eigenvector $\vec{1}$. Moreover, we have the following result:

Theorem 1.2. *If $\lambda_1 \leq \lambda_2 \leq ...\lambda_n$ are the eigenvalues of L, the $\lambda_k = 0$ iff G has at least k connected components.*

Proof. Suppose that G has k connected components, $C_1, C_2, ...C_k \subseteq V$. Recall that

$$\lambda_k = \min_{U: dim U = k} \max_{x \in U} \frac{x^\top L x}{x^\top x}$$

Since L is PSD, if there exists any k-dimensional subspace where every vector has a Rayleigh quotient of 0, then λ_k must be 0. How do we find such a subspace? Simple. Define $c_1, ...c_k$ such that $c_i[v] = 1$ if $v \in C_i$ and 0 otherwise. Each c_i corresponds to a partition with no cut edges, since one connected component is on one side of the partition and the rest of the graph is on the other side. Also, $c_1, ...c_k$ are clearly linearly independent, as non of them are non-zero in the same entry. Thus, if $\sum_{i=1}^k a_i c_i = 0$, it must be that $a_i = 0$ for all i.

Now, suppose that $\lambda_k = 0$. This means that there exists a k-dimensional subspace, U, where every vector in U has Rayleigh quotient 0. This means for any $x \in U$, we must have:

$$x^\top L x = \sum_{(u,v)\in E} (x_u - x_v)^2 = \sum_u \sum_v (x_u - x_v)^2 \vec{1}\{(u,v) \in E\} = 0$$

This means that for $u, v \in V \times V$, either $x_u = x_v$ or $(u, v) \notin E$. If u and v are in the same connected component, then it must be that $x_u = x_v$. Thus, $U = \{x \in \mathbb{R}^n : \ u = v$ if they are in the same connected component$\}$. Since $dimU = k$ there must be at least k connected components. $\square$

Based on the previous theorem, we clearly see that 1) every Laplacian has at least one vector in it's nullspace, namely the all 1's vector, and 2) the dimension of the nullspace of the Laplacian is equalled to the number of connected components.

Normalized Laplacian

Let $G = (V, E)$ be a d-regular graph. That is, each vertex has degree d. Define the normalized Laplacian

$$L_N = \frac{1}{d} L$$

.

Theorem 1.3. *As if I haven't abused notation enough, let $\lambda_1 \leq, \ldots \lambda_n$ be the eigenvalues of L_N. $\lambda_N = 2$ iff G has at least one bipartite connected component.*

Proof. First, we express the Rayleigh Quotient in terms of L_N as follows:

$$\begin{aligned}
\frac{x^\top L_N x}{x^\top x} &= x^\top (I - \frac{1}{d} A) x \\
&= x^\top x - \frac{1}{d} x^\top A x \\
&= x^\top x - \frac{1}{d} \sum_u \sum_v x_u x_v 1\{(u, v) \in E\} \\
&= x^\top x - \frac{1}{d} \sum_{(u,v) \in E} 2 x_u x_v \\
&= x^\top x - \frac{1}{d} \sum_{(u,v) \in E} (x_u + x_v)^2 - x_u^2 - x_v^2 \\
&= x^\top x + \frac{1}{d} (\sum_{v \in V} deg(v) x_v^2) - \frac{1}{d} \sum_{(u,v) \in E} (x_u + x_v)^2 \\
&= 2x^\top x - \frac{1}{d} \sum_{(u,v) \in E} (x_u + x_v)^2
\end{aligned}$$

Thus, we have:

$$\begin{aligned}
\lambda_n &= \max_x \frac{x^\top L_N x}{x^\top x} \\
&= \max_x \frac{1}{x^\top x} (2x^\top x - \frac{1}{d} \sum_{(u,v) \in E} (x_u + x_v)^2) \\
&= 2 - \min_x \frac{\sum_{(u,v) \in E} (x_u + x_v)^2}{x^\top x}
\end{aligned}$$

The term $\sum_{(u,v)\in E}(x_u+x_v)^2$ goes to 0 iff $\forall(u,v)\in E$, $x_u=-x_v$ or $x_u,x_v=0$. Since, we do not allow $x=0$, let $A=\{v:\ x_v>0\}$ and $B=\{v:\ x_v<0\}$. If there is an edge (u,v) with only one endpoint in $A\cup B$, or both endpoints in the same set, it must be that $(x_u+x_v)^2>0$. Thus, we see that in order for $\lambda_n=2$, $A\cup B$ is bipartite, and vise versa. □

It turns out that this condition holds even when G is not d-regular. However, for non-regular graph, we must define the normalized Laplacian, differently.:

$$L_N = D^{-1/2}LD^{-1/2}$$

This is valid if we assume there are no lone vertices, which is reasonable, as we can simply remove them. Now, consider the new Rayleigh quotient, defined as follows:

$$R^*_{L_N}(x) = R_{L_N}(D^{1/2}x) = \frac{x^\top Lx}{x^\top Dx} = \frac{\sum_{(u,v)\in E}(x_u-x_v)^2}{\sum_v deg(v)x_v}$$

Since $D^{1/2}$ is invertible, if $\{x_1,x_2,...x_k\}$ is the basis of some k-dimensional subspace, then $\{D^{1/2}x_1,...D^{1/2}x_k\}$ is also k-dimensional. Thus, it still have that:

$$\lambda_k = \min_{U:dimU=k}\max_{x\in U} R^*_{L_N}(x)$$

This means we can apply the same argument as before and show that:

$$\begin{aligned}
\lambda_n &= \max_x \frac{x^\top Lx}{x^\top Dx} \\
&= \max_x \frac{1}{x^\top Dx}(x^\top Dx - x^\top Ax) \\
&= \max_x \frac{1}{x^\top Dx}(2x^\top Dx - \sum_{(u,v)\in E}(x_u+x_v)^2) \\
&= 2 - \min_x \frac{\sum_{(u,v)\in E}(x_u+x_v)^2}{x^\top x}
\end{aligned}$$

Spectral Graph Drawing

Consider the problem of drawing a graph in 2 or 3 dimensions in a way that we can visualize nicely. Ideally, we would want to stretch the edges as little as possible. Moreover, vertices that are well connected should be placed closed to each-other. One way to achieve this is to treat each edge as a spring that connects two vertices. The amount of energy it takes to stretch the string is the square of the distance between the vertices. Suppose we want to draw our graph in $\mathbb{R}^2$. This means we must map each vertex, v to a point in $f(v)\in\mathbb{R}^2$. To minimize the total stretch over the edges, we want:

$$\begin{aligned}
\min_f \sum_{u,v\in E} ||f(u)-f(v)||_2^2 &= \min_{x,y\in\mathbb{R}^n}\sum_{u,v\in E}(x(u)-x(v))^2+(y(u)-y(v))^2 \\
&= \min_{x,y} x^\top Lx + y^\top Ly
\end{aligned}$$

To avoid mapping everything to 0, we resitrct x and y such that $||x||^2=||y||^2=1$. To avoid mapping everything to $\frac{1}{\sqrt{n}}\vec{1}$, or mapping x and y to the same points (in which case everything would lie along a line), we further add the restrictions that $x,y\ \perp\ \vec{1}$ and $y\ \perp\ x$. This means that x and y are the second and third eigenvectors of the normalized Laplacian. Thus, the eigenvectors of the Laplacian can give use nice ways of representing graphs visually.

Computing Eigenvalues and Eigenvectors

In general, computing eigenvalues exactly requires finding the roots of degree-n polynomials, which is a hard task. However, eigenvalues and eigenvectors for PSD matrices can be accurately approximated through a very simple method, called the power method. Suppose M is PSD, and we want to compute λ_n, the maximum eigenvalue of M.

1. pick a random vector $x \in \{-1, 1\}^n$
2. compute Mx
3. normalize x, and repeat

The output, x_k, after k iterations, will have a Rayleigh quotient close to λ_n. Specifically, we have the following:

Theorem 1.4. *With probability at least 3/16,* $x_k^\top Mx \geq \frac{\lambda_n(1-\epsilon)}{1+4n(1-\epsilon)^{2k}}$

This means that for $k = \log_{1/(1-\epsilon)}(n/\epsilon)$, $x_k^\top Mx = \lambda_n(1 - O(\epsilon))$. To understand this from an inuitive level, consider the fact that any $x \in \mathbb{R}^n$ can be expressed as:

$$\sum_i < v_i, x > v_i$$

which means:

$$M^k x = \sum_i \lambda_i^k < v_i, x > v_i$$

Clearly, as k increases, the largest term, $\lambda_n^k < v_i, x > v_i$ will dominate, and the resulting vector will point closer and closer in the v_n direction. Once we can v_n, we can then approximate v_{n-1} by applying the power method on $x - < x, v_n > v_n$.