Cornell University, Fall 2013	**CS 6820: Algorithms**
Lecture notes on spectral methods in algorithm design	

Studying the eigenvalues and eigenvectors of matrices has powerful consequences for at least three areas of algorithm design: graph partitioning, analysis of high-dimensional data, and analysis of Markov chains. Collectively, these techniques are known as *spectral methods* in algorithm design. These lecture notes present the fundamentals of spectral methods.

1 Review: symmetric matrices, their eigenvalues and eigenvectors

This section reviews some basic facts about real symmetric matrices. If $A = (a_{ij})$ is an $n-\times-n$ square symmetric matrix, then $\mathbb{R}^n$ has a basis consisting of eigenvectors of A, these vectors are mutually orthogonal, and all of the eigenvalues are real numbers. Furthermore, the eigenvectors and eigenvalues can be characterized as solutions of natural maximization or minimization problems involving *Rayleigh quotients*.

Definition 1.1. If x is a nonzero vector in $\mathbb{R}^n$ and A is an $n - \times - n$ matrix, then the Rayleigh quotient of x with respect to A is the ratio

$$RQ_A(x) = \frac{x^\mathsf{T} Ax}{x^\mathsf{T} x}.$$

Definition 1.2. If A is an $n - \times - n$ matrix, then a linear subspace $V \subseteq \mathbb{R}^n$ is called an *invariant subspace* of A if it satisfies $Ax \in V$ for all $x \in V$.

Lemma 1.3. *If A is a real symmetric matrix and V is an invariant subspace of A, then there is some $x \in V$ such that $RQ_A(x) = \inf\{RQ_A(y) \mid y \in V\}$. Any $x \in V$ that minimizes $RQ_A(x)$ is an eigenvector of A, and the value $RQ_A(x)$ is the corresponding eigenvalue.*

Proof. If x is a vector and r is a nonzero scalar, then $RQ_A(x) = RQ_A(rx)$, hence every value attained in V by the function RQ_A is attained on the unit sphere $S(V) = \{x \in V \mid x^\mathsf{T} x = 1\}$. The function RQ_A is a continuous function on $S(V)$, and $S(V)$ is compact (closed and bounded) so by basic real analysis we know that RQ_A attains its minimum value at some unit vector $x \in S(V)$. Using the quotient rule we can compute the gradient

$$\nabla RQ_A(x) = \frac{2Ax - 2(x^\mathsf{T} Ax)x}{(x^\mathsf{T} x)^2}. \tag{1}$$

At the vector $x \in S(V)$ where RQ_A attains its minimum value in V, this gradient vector must be orthogonal to V; otherwise, the value of RQ_A would decrease as we move away from x in the direction of any $y \in V$ that has negative dot product with $\nabla RQ_A(x)$. On the other hand, our assumption that V is an invariant subspace of A implies that the right side of (1) belongs to V. The only way that $\nabla RQ_A(x)$ could be orthogonal to V while also belonging to V is if it is the zero vector, hence $Ax = \lambda x$ where $\lambda = x^\mathsf{T} Ax = RQ_A(x)$. □

Lemma 1.4. *If A is a real symmetric matrix and V is an invariant subspace of A, then $V^\perp = \{x \mid x^\mathsf{T} y = 0 \; \forall y \in V\}$ is also an invariant subspace of A.*

Proof. If V is an invariant subspace of A and $x \in V^\perp$, then for all $y \in V$ we have

$$(Ax)^\mathsf{T} y = x^\mathsf{T} A^\mathsf{T} y = x^\mathsf{T} A y = 0,$$

hence $Ax \in V^\perp$. □

Combining these two lemmas, we obtain a recipe for extracting all of the eigenvectors of A, with their eigenvalues arranged in increasing order.

Theorem 1.5. *Let A be an $n - \times - n$ real symmetric matrix and let us inductively define sequences*

$$\begin{aligned} x_1, \ldots, x_n &\in \mathbb{R}^n \\ \lambda_1, \ldots, \lambda_n &\in \mathbb{R} \\ \{0\} = V_0 \subseteq V_1 \subseteq \cdots \subseteq V_n &= \mathbb{R}^n \\ \mathbb{R}^n = W_0 \supseteq W_1 \supseteq \cdots \supseteq W_n &= \{0\} \end{aligned}$$

by specifying that

$$\begin{aligned} x_i &= \operatorname{argmin} \{RQ_A(x) \mid x \in W_{i-1}\} \\ \lambda_i &= RQ_A(x_i) \\ V_i &= \operatorname{span}(x_1, \ldots, x_i) \\ W_i &= V_i^\perp. \end{aligned}$$

Then $x_1, \ldots, x_n$ are mutually orthogonal eigenvectors of A, and $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ are the corresponding eigenvalues.

Proof. The proof is by induction on i. The induction hypothesis is that $\{x_1, \ldots, x_i\}$ is a set of mutually orthogonal eigenvectors of A constituting a basis of V_i, and $\lambda_1 \le \cdots \le \lambda_i$ are the corresponding eigenvectors. Given this induction hypothesis, and the preceding lemmas, the proof almost writes itself. Each time we select a new x_i, it is guaranteed to be orthogonal to the preceding ones because $x_i \in W_{i-1} = V_{i-1}^\perp$. The linear subspace V_{i-1} is A-invariant because it is spanned by eigenvectors of A; by Lemma 1.4 its orthogonal complement W_{i-1} is also A-invariant and this implies, by Lemma 1.3 that x_i is an eigenvector of A and λ_i is its corresponding eigenvalue. Finally, $\lambda_i \ge \lambda_{i-1}$ because $\lambda_{i-1} = \min\{RQ_A(x) \mid x \in W_{i-2}\}$, while $\lambda_i = RQ_A(x_i) \in \{RQ_A(x) \mid x \in W_{i-2}\}$. □

An easy corollary of Theorem 1.5 is the *Courant-Fischer Theorem.*

Theorem 1.6 (Courant-Fischer)**.** *The eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ of an $n - \times - n$ real symmetric matrix satisfy:*

$$\forall k \; \lambda_k = \min_{\dim(V)=k} \left(\max_{x \in V} RQ_A(x) \right) = \max_{\dim(W)=n-k+1} \left(\min_{x \in W} RQ_A(x) \right).$$

Proof. The vector space W_{k-1} constructed in the proof of Theorem 1.5 has dimension $n-k+1$, and by construction it satisfies $\min_{x \in W_{k-1}} RQ_A(x) = \lambda_k$. Therefore

$$\max_{\dim(W)=n-k+1} \left(\min_{x \in W} RQ_A(x) \right) \geq \lambda_k.$$

If $W \subseteq \mathbb{R}^n$ is any linear subspace of dimension $n - k + 1$ then $W \cap V_k$ contains a nonzero vector x, because $\dim(W) + \dim(V_k) > n$. Since $V_k = \text{span}(x_1, \ldots, x_k)$ we can write $x = a_1 x_1 + \cdots + a_k x_k$. Rescaling $x_1, \ldots, x_k$ if necessary, we can assume that they are all unit vectors. Then, using the fact that $x_1, \ldots, x_k$ are mutually orthogonal eigenvectors of A, we obtain

$$RQ_A(x) = \frac{\lambda_1 a_1 + \cdots + \lambda_k a_k}{a_1 + \cdots + a_k} \leq \lambda_k.$$

Therefore $\max_{\dim(W)=n-k+1} (\min_{x \in W} RQ_A(x)) \leq \lambda_k$. Combining this with the inequality derived in the preceding paragraph, we obtain $\max_{\dim(W)=n-k+1} \min_{x \in W} RQ_A(x) = \lambda_k$. Replacing A with $-A$, and k with $n-k+1$, we obtain $\min_{\dim(V)=k} (\max_{x \in V} RQ_A(x)) = \lambda_k$. ☐

2 The Graph Laplacian

Two symmetric matrices play a vital role in the theory of graph partitioning. These are the Laplacian and normalized Laplacian matrix of a graph G.

Definition 2.1. If G is an undirected graph with non-negative edge weights $w(u, v) \geq 0$, the *weighted degree* of a vertex u, denoted by $d(u)$, is the sum of the weights of all edges incident to u. The Laplacian matrix of G is the matrix L_G with entries

$$(L_G)_{uv} = \begin{cases} d(u) & \text{if } u = v \\ -w(u,v) & \text{if } u \neq v \text{ and } (u,v) \in E \\ 0 & \text{if } u \neq v \text{ and } (u,v) \notin E. \end{cases}$$

If D_G is the diagonal matrix whose (u, u)-entry is $d(u)$, and if G has no vertex of weighted degree 0, then the normalized Laplacian matrix of G is

$$\overline{L}_G = D_G^{-1/2} L_G D_G^{-1/2}.$$

The eigenvalues of L_G and $\overline{L}_G$ will be denoted in these notes by $\lambda_1(G) \leq \cdots \leq \lambda_n(G)$ and $\nu_1(G) \leq \cdots \leq \nu_n(G)$. When the graph G is clear from context, we will simply write these as $\lambda_1, \ldots, \lambda_n$ or $\nu_1, \ldots, \nu_n$.

The "meaning" of the Laplacian matrix is best explained by the following observation.

Observation 2.2. The Laplacian matrix L_G is the unique symmetric matrix satisfying the following relation for all vectors $x \in \mathbb{R}^V$.

$$x^{\mathsf{T}} L_G x = \sum_{(u,v) \in E} w(u,v)(x_u - x_v)^2. \tag{2}$$

The following lemma follows easily from Observation 2.2.

Lemma 2.3. *The Laplacian matrix of a graph G is a positive semidefinite matrix. Its minimum eigenvalue is 0. The multiplicity of this eigenvalue equals the number of connected components of G.*

Proof. The right side of (2) is always non-negative, hence L_G is positive semidefinite. The right side is zero if and only if x is constant on each connected component of G (i.e., it satisfies $x_u = x_v$ whenever u, v belong to the same component), hence the multiplicity of the eigenvalue 0 equals the number of connected components of G. □

The normalized Laplacian matrix has a more obscure graph-theoretic meaning than the Laplacian, but its eigenvalues and eigenvectors are actually more tightly connected to the structure of G. Accordingly, we will focus on normalized Laplacian eigenvalues and eigenvectors in these notes. The cost of doing so is that the matrix $\overline{L}_G$ is a bit more cumbersome to work with. For example, when G is connected the 0-eigenspace of L_G is spanned by the all-ones vector $\mathbf{1}$ whereas the 0-eigenspace of $\overline{L}_G$ is spanned by the vector $\mathbf{d}^{1/2} = D_G^{1/2}\mathbf{1}$.

3 Conductance and expansion

We will relate the eigenvalue $\nu_2(G)$ to two graph parameters called the *conductance* and *expansion* of G. Both of them measure the value of the "sparsest" cut, with respect to subtly differing notions of sparsity. For any set of vertices S, define

$$d(S) = \sum_{u \in S} d(u)$$

and define the edge boundary

$$\partial S = \{e = (u, v) \mid \text{exactly one of } u, v \text{ belongs to } S\}.$$

The *conductance* of G is

$$\phi(G) = \min_{(S,\overline{S})} \left\{ d(V) \cdot \frac{w(\partial S)}{d(S)d(\overline{S})} \right\}$$

and the *expansion* of G is

$$h(G) = \min_{(S,\overline{S})} \left\{ \frac{w(\partial S)}{\min\{d(S), d(\overline{S})\}} \right\},$$

where the minimum in both cases is over all vertex sets $S \neq \emptyset, V$. Note that for any such S,

$$\frac{d(V)}{d(S)d(\overline{S})} = \frac{d(V)}{\min\{d(S), d(\overline{S})\} \cdot \max\{d(S), d(\overline{S})\}} = \frac{1}{\min\{d(S), d(\overline{S})\}} \cdot \frac{d(V)}{\max\{d(S), d(\overline{S})\}}.$$

The second factor on the right side is between 1 and 2, and it easily follows that

$$h(G) \le \phi(G) \le 2h(G).$$

Thus, each of the parameters $h(G), \phi(G)$ is a 2-approximation to the other one. Unfortunately, it is not known how to compute a $O(1)$-approximation to either of these parameters in polynomial time. In fact, assuming the Unique Games Conjecture, it is NP-hard to compute an $O(1)$-approximation to either of them.

4 Cheeger's Inequality: Lower Bound on Conductance

There is a sense, however, in which $\nu_2(G)$ constitutes an approximation to $\phi(G)$. To see why, let us begin with the following characterization of $\nu_2(G)$ that comes directly from Courant-Fischer.

$$\nu_2(G) = \min\left\{ \frac{x^\top \overline{L}_G x}{x^\top x} \,\middle|\, x \neq 0,\, x^\top D_G^{1/2}\mathbf{1} = 0 \right\} = \min\left\{ \frac{y^\top L_G y}{y^\top D_G y} \,\middle|\, y \neq 0,\, y^\top D_G \mathbf{1} = 0 \right\}.$$

The latter equality is obtained by setting $x = D_G^{1/2} y$.

The following lemma allows us to rewrite the Rayleigh quotient $\frac{y^\top L_G y}{y^\top D_G y}$ in a useful form, when $y^\top D_G \mathbf{1} = 0$.

Lemma 4.1. *For any vector y we have*

$$y^\top D_G y \ge \frac{1}{2d(V)} \sum_{u \neq v} d(u)d(v)(y(u) - y(v))^2,$$

with equality if and only if $y^\top D_G \mathbf{1} = 0$.

Proof.

$$\begin{aligned}
\frac{1}{2}\sum_{u\neq v} d(u)d(v)(y(u)-y(v))^2 &= \frac{1}{2}\sum_{u\neq v} d(u)d(v)[y(u)^2 + y(v)^2] - \sum_{u\neq v} d(u)d(v)y(u)y(v) \\
&= \sum_{u\neq v} d(u)d(v)y(u)^2 - \sum_{u\neq v} d(u)d(v)y(u)y(v) \\
&= \sum_{u,v} d(u)d(v)y(u)^2 - \sum_{u,v} d(u)d(v)y(u)y(v) \\
&= d(V)\sum_{u} d(u)y(u)^2 - \left(\sum_{u} d(u)y(u)\right)^2 \\
&= d(V)y^\top D_G y - \left(y^\top D_G \mathbf{1}\right)^2.
\end{aligned}$$

□

A corollary of the lemma is the formula

$$\nu_2(G) = \inf \left\{ d(V) \frac{\sum_{(u,v)\in E(G)} w(u,v)(y(u)-y(v))^2}{\sum_{u<v} d(u)d(v)(y(u)-y(v))^2} \,\middle|\, \text{denominator is nonzero} \right\}, \quad (3)$$

where the summation over $u < v$ in the denominator is meant to indicate that each unordered pair $\{u, v\}$ of distinct vertices contributes exactly one term to the sum. The corollary is obtained by noticing that the numerator and denominator on the right side are invariant under adding a scalar multiple of $\mathbf{1}$ to y, and hence one of the vectors attaining the infimum is orthogonal to $D_G\mathbf{1}$.

Let us evaluate the quotient on the right side of (3) when y is the characteristic vector of a cut $(S, \overline{S})$, defined by

$$y(u) = \begin{cases} 1 & \text{if } u \in S \\ 0 & \text{if } u \in \overline{S}. \end{cases}$$

In that case,

$$\sum_{(u,v)\in E(G)} w(u,v)(y(u)-y(v))^2 = \sum_{(u,v)\in\partial S} w(u,v) = w(\partial S)$$

while

$$\sum_{u<v} d(u)d(v)(y(u)-y(v))^2 = \sum_{u\in S}\sum_{v\in\overline{S}} d(u)d(v) = d(S)d(\overline{S}).$$

Hence,

$$\nu_2(G) \le d(V)\frac{w(\partial S)}{d(S)d(\overline{S})},$$

and taking the minimum over all $(S, \overline{S})$ we obtain

$$\nu_2(G) \le \phi(G).$$

5 Cheeger's Inequality: Upper Bound on Conductance

The inequality $\nu_2(G) \le \phi(G)$ is the easy half of Cheeger's Inequality; the more difficult half asserts that there is also an upper bound on $\phi(G)$ of the form

$$\phi(G) \le \sqrt{8\nu_2(G)}.$$

Owing to the inequality $\phi(G) \le 2h(G)$, it suffices to prove that

$$h(G) \le \sqrt{2\nu_2(G)}$$

and that is, in fact, the next thing we will prove.

For any vector y that is not a scalar multiple of $\mathbf{1}$, define

$$Q(y) = d(V)\frac{\sum_{(u,v)\in E(G)} w(u,v)(y(u)-y(v))^2}{\sum_{u<v} d(u)d(v)(y(u)-y(v))^2}.$$

Given any such y, we will find a cut $(S,\overline{S})$ such that $\frac{w(\partial S)}{\min\{d(S),d(\overline{S})\}} \leq \sqrt{2Q(y)}$; the upper bound $h(G) \leq \sqrt{2\nu_2(G)}$ follows immediately by choosing y to be a vector minimizing $Q(y)$. In fact, if we number the vertices of G as $v_1, v_2, \ldots, v_n$ such that $y_1 \leq y_2 \leq \cdots \leq y_n$, we will show that it suffices to take S to be one of the sets $\{y_1, \ldots, y_k\}$ for $1 \leq k < n$.

Note that $Q(y)$ is unchanged when we add a scalar multiple of $\mathbf{1}$ to y. Accordingly, we can assume without loss of generality that

$$\sum_{y_i<0} d(v_i) \leq \sum_{y_i\geq 0} d(v_i)$$
$$\sum_{y_i\leq 0} d(v_i) \geq \sum_{y_i>0} d(v_i)$$

For d-regular graphs, this essentially means that we're setting the median of the components of y to be zero. For irregular graphs, it essentially says that we're balancing the total degree of the vertices with positive $y(u)$ and those with negative $y(u)$.

Now here comes the most unmotivated part of the proof. Define a vector z by

$$z_i = \begin{cases} -y_i^2 & \text{if } y_i < 0 \\ y_i^2 & \text{if } y_i \geq 0. \end{cases}$$

Note also that $Q(y)$ is unchanged when we multiply y by a nonzero scalar. Accordingly, we can assume that $z_n - z_1 = 1$. Now choose a threshold value t uniformly at random from the interval $[z_1, z_n]$ and let

$$S = \{v_i \mid z_i < t\}.$$

We will prove that

$$\frac{\mathbb{E}[w(\partial S)]}{\mathbb{E}[\min\{d(S), d(\overline{S})\}]} \leq \sqrt{2Q(y)}$$

from which it follows that

$$\mathbb{E}[w(\partial S)] \leq \sqrt{2Q(y)} \cdot \mathbb{E}[\min\{d(S), d(\overline{S})\}]$$

and consequently that there is at least one S in the support of our distribution such that

$$w(\partial S) \leq \sqrt{2Q(y)} \cdot \min\{d(S), d(\overline{S})\}.$$

It is surprisingly easy to evaluate $\mathbb{E}[\min\{d(S), d(\overline{S})\}]$. Each vertex v_i contributes $d(v_i)$ to the expression inside the expectation operator when it belongs to the smaller side of the cut, which happens if and only if t lands between 0 and z_i, an event with probability $|z_i|$. Consequently,

$$\mathbb{E}[\min\{d(S), d(\overline{S})\}] = \sum_u d(u)|z(u)| = \sum_u d(u)y(u)^2 = y^\mathsf{T} D_G y.$$

Meanwhile, to bound the numerator $\mathbb{E}[w(\partial S)]$, observe that an edge (u, v) contributes $w(u, v)$ to the numerator if and only if it is cut, an event having probability $|z(u) - z(v)|$. A bit of case analysis reveals that

$$\forall u, v \;\; |z(u) - z(v)| \leq |y(u) - y(v)| \cdot (|y(u)| + |y(v)|),$$

since the left and right sides are equal when $y(u), y(v)$ have the same sign, and otherwise the left side equals $y(u)^2 + y(v)^2$ while the right side equals $(|y(u)| + |y(v)|)^2$. Combining this estimate of the numerator with Cauchy-Schwartz, we find that

$$\begin{aligned}
\mathbb{E}[w(\partial S)] &\leq \sum_{(u,v)\in E(G)} w(u,v)|y(u) - y(v)|(|y(u)| + |y(v)|) \\
&\leq \left(\sum_{(u,v)\in E(G)} w(u,v)(y(u) - y(v))^2 \right)^{1/2} \left(\sum_{(u,v)\in E(G)} w(u,v)(|y(u)| + |y(v)|)^2 \right)^{1/2} \\
&\leq \left(\frac{Q(y)}{d(V)} \sum_{u<v} d(u)d(v)(y(u) - y(v))^2 \right)^{1/2} \left(\sum_{(u,v)\in E(G)} w(u,v)(2y(u)^2 + 2y(v)^2) \right)^{1/2} \\
&\leq \left(Q(y) y^\mathsf{T} D_G y \right)^{1/2} \left(2 \sum_u d(u)y(u)^2 \right)^{1/2} \\
&= (2Q(y))^{1/2} y^\mathsf{T} D_G y.
\end{aligned}$$

6 Laplacian eigenvalues and spectral partitioning

We've seen a connection between sparse cuts and eigenvectors of the *normalized* Laplacian matrix. However, in some contexts it is easier to work with eigenvalues and eigenvectors of the unnormalized Laplacian, L_G. One can use eigenvectors of L_G for spectral partitioning, provided one is willing to tolerate weaker bounds for graphs with unbalanced degree sequences. For example, if y is an eigenvector of L_G satisfying $L_G y = \lambda_2 y$ then we can express $Q(y)$ as follows:

$$Q(y) = d(V) \frac{y^\mathsf{T} L_G y}{\sum_{u<v} d(u)d(v)(y(u) - y(v))^2} = \frac{\lambda_2 \|y\|^2 d(V)}{\sum_{u<v} d(u)d(v)(y(u) - y(v))^2}.$$

To estimate the denominator, let $d_{\min}$ and d_{avg} denote the minimum and the average degree of G, respectively. We have

$$\begin{aligned}\sum_{u<v} d(u)d(v)(y(u)-y(v))^2 &= \frac{1}{2}\sum_{u\neq v} d(u)d(v)(y(u)-y(v))^2 \\ &\geq \frac{1}{2}d_{\min}^2 \sum_{u\neq v}(y(u)-y(v))^2 \\ &= nd_{\min}^2 \sum_u y(u)^2 = \frac{d(V)}{d_{\text{avg}}} d_{\min}^2 \|y\|^2.\end{aligned}$$

Hence

$$Q(y) \leq \frac{d_{\text{avg}}}{d(V)} \frac{\lambda_2 \|y\|^2 d(V)}{d_{\min}^2 \|y\|^2} = \left(\frac{d_{\text{avg}}}{d_{\min}^2}\right) \lambda_2.$$