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 Self-adjoint endomorphisms of finite-dimensional real vector spaces

This section reviews and generalizes some basic facts about the eigenvalues and eigenvectors of real symmetric matrices. We will be applying these properties in the setting of finite-dimensional inner product spaces over the real numbers. The appropriate generalization of a symmetric matrix to this setting is called a *self-adjoint endomorphism.* We will provide the relevant definitions and theorems about self-adjoint endomorphisms of finite-dimensional real vector spaces below.

Definition 1.1 (inner product)**.** If V is a vector space over the real numbers, an inner product on V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ that satisfies:

1. **[symmetry]** $\langle x, y \rangle = \langle y, x \rangle_V$.
2. **[bilinearity]**
$$\left\langle \sum_{i=1}^{n} a_i x_i, \sum_{j=1}^{m} b_j y_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \langle x_i, y_j \rangle$$

The inner product is called *positive definite* if, in addition, it satisfies $\langle x, x \rangle > 0$ for all $x \neq 0$. For a positive definite inner product, we will use the notation $\|x\|$ to denote $\langle x, x \rangle^{1/2}$.

The most important example of a positive definite inner product is the *standard inner product* on $\mathbb{R}^n$, also known as the dot product.

$$\langle x, y \rangle = x^{\mathsf{T}} y = \sum_{i=1}^{n} x_i y_i.$$

Generalizing this example, if D is an n-by-n diagonal matrix with diagonal entries $D_{ii} = d_i > 0$, then the D-weighted inner product

$$\langle x, y \rangle_D = x^{\mathsf{T}} D y = \sum_{i=1}^{n} d_i x_i y_i$$

is positive definite. When D is a diagonal matrix whose entries are the degrees of the vertices of a finite graph, this D-weighted inner product plays an important role in spectral graph theory.

Definition 1.2. If V is a vector space with positive definite inner product $\langle \cdot, \cdot \rangle$, two vectors $x, y \in V$ are called *orthogonal* if $\langle x, y \rangle = 0$. For a linear subspace W, the orthogonal complement of W is defined to be the set of all vectors that are orthogonal to every vector in W:

$$W^{\perp} = \{x \mid \langle w, x \rangle = 0 \, \forall w \in W\}.$$

A set of vectors $\{v_1, \ldots, v_m\} \subset V$ is called *orthonormal* if they are pairwise orthogonal and each has norm 1, i.e.

$$\forall i, j \in \{1, \ldots, m\} \quad \langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

An important property of orthonormal sets in finite-dimensional positive definite inner product spaces is that any orthonormal set can be extended to an orthonormal basis. The procedure for doing so is called the *Gram-Schmidt procedure.*

Lemma 1.3. *If V is a finite-dimensional positive definite inner product space and $S = \{v_1, \ldots, v_m\}$ is an orthonormal set in V, then there exists a basis $B \supseteq S$ that is an orthonormal set.*

Proof. First, observe that the vectors in any orthonormal set are linearly independent. In fact, for any linear combination $w = a_1 v_1 + \cdots + a_m v_m$ we have

$$\langle w, w \rangle = \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \langle v_i, v_j \rangle = \sum_{i=1}^{m} a_i^2,$$

so if $w = 0$ then $a_i = 0$ for all i.

Given this observation, we will denote the dimension of V by n and we'll prove the lemma by induction on $n - m$. In the base case when $n - m = 0$, the given orthonormal set S is a basis of V and there is nothing to prove. In the induction step when $m < n$, we just need to find one vector $v_{m+1} \notin S$ such that $S \cup \{v_{m+1}\}$ is orthonormal, then we can apply the induction hypothesis to extend $S \cup \{w\}$ to an orthonormal basis.

To find v_{m+1} we start by taking an arbitrary vector $w \in V$ that is not a linear combination of elements of S; such a w must exist because S has too few elements to constitute a basis of V. Now let $x = w - \sum_{i=1}^{m} \langle w, v_i \rangle \cdot v_i$, and observe that x is orthogonal to each element of S because

$$\langle x, v_j \rangle = \langle w, v_j \rangle - \sum_{i=1}^{m} \langle w, v_i \rangle \cdot \langle v_i, v_j \rangle = \langle w, v_j \rangle - \langle w, v_j \rangle = 0.$$

Furthermore, $x \neq 0$ by our choice of w, so $\|x\| > 0$. (This is the one and only step of the proof where we use the fact that the inner product is positive definite.) The vector $v_{m+1} = x/\|x\|$ has norm 1 and is orthogonal to each element of S, which completes the induction step. $\square$

It follows from Lemma 1.3 that all positive-definite inner product spaces of a given (finite) dimension are isomorphic.

Lemma 1.4. *For $n \in \mathbb{N}$, every n-dimensional positive-definite inner product space is isomorphic to $\mathbb{R}^n$ with the standard inner product.*

Proof. If V is an n-dimensional positive-definite inner product space, we can apply Lemma 1.3 with $S = \emptyset$ to conclude that V has an orthonormal basis $B = \{v_1, \dots, v_n\}$. Now consider the linear transformation $F : \mathbb{R}^n \to V$ defined by $F(a_1, \dots, a_n) = \sum_{i=1}^n a_i v_i$. Since B is a basis, we know that F is a bijection between $\mathbb{R}^n$ and V, so we only need to verify that it preserves the inner product structure. For any $a, b \in \mathbb{R}^n$, we have

$$\begin{aligned}
\langle F(a),\, F(b)\rangle &= \left\langle \sum_{i=1}^n a_i v_i,\, \sum_{j=1}^n b_j v_j \right\rangle \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i b_j \langle v_i,\, v_j\rangle \\
&= \sum_{i=1}^n a_i b_i
\end{aligned}$$

which confirms that the bijection F preserves inner products. □

Lemma 1.5. *If V is a finite-dimensional positive definite inner product space and $W \subseteq V$ is a linear subspace, then $(W^\perp)^\perp = W$.*

Proof. If an orthonormal basis of V is partitioned into two subsets, B_1 and B_2, and if W_1 and W_2 denote the linear spans of B_1 and B_2, respectively, then $W_1^\perp = W_2$ and $W_2^\perp = W_1$. To see that $W_1^\perp = W_2$, note that every vector $x \in V$ can be uniquely represented as $x = \sum_{i=1}^n a_i v_i$, where $\{v_1, \dots, v_n\}$ are the elements of the orthonormal basis, and that $x \in W_1$ if and only if $a_i = 0$ for all $v_i \notin B_1$, whereas $x \in W_1^\perp$ if and only if $a_i = 0$ for all $v_i \in B_1$. The proof that $W_2^\perp = W_1$ is the same, with the roles of B_1 and B_2 reversed.

Now, if $W \subseteq V$ is a linear subspace, apply Lemma 1.3 to the inner product space W with starting set $S = \emptyset$ to obtain an orthonormal basis B_1 for W. Then apply Lemma 1.3 again to the inner product space V with starting set $S = B_1$ to obtain an orthonormal basis $B_1 \cup B_2$ for V. Then apply the observation in the first paragraph to derive the lemma's conclusion. □

Definition 1.6 (self-adjoint)**.** If V is a finite-dimensional vector space with inner product $\langle \cdot, \cdot \rangle_V$, an endomorphism of a V is a linear transformation $T : V \to V$. An endomorphism T is called *self-adjoint* if it satisfies the identity

$$\langle x,\, Ty\rangle_V = \langle Tx,\, y\rangle_V$$

for all $x, y \in V$.

Observation 1.7. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ represented by a matrix A is self-adjoint with respect to the standard inner product if and only if A is a symmetric matrix. To see this, first observe that if A is symmetric then

$$\langle x,\, Ay\rangle = x^{\mathsf{T}}(Ay) = (x^{\mathsf{T}}A)y = (A^{\mathsf{T}}x)^{\mathsf{T}}y = (Ax)^{\mathsf{T}}y = \langle Ax,\, y\rangle.$$

Conversely, if $\mathbf{e}_i$ denotes the i^{th} standard basis vector of $\mathbb{R}^n$ then

$$\langle \mathbf{e}_i,\, A\mathbf{e}_j\rangle = a_{ij}, \qquad \langle A\mathbf{e}_i,\, \mathbf{e}_j\rangle = a_{ji}$$

hence self-adjointness of the endomorphism represented by A implies that A is symmetric.

A basic fact about self-adjoint endomorphisms is that all of their eigenvalues are real, their eigenvectors are mutually orthogonal, and they can be characterized by an iterative process of optimization over unit vectors. One proof of this fact is based on *invariant subspaces*, which we now define.

Definition 1.8. If A is an endomorphism of an inner product space V, then a subspace $W \subseteq V$ is called an *invariant subspace* of A if it satisfies $Ax \in W$ for all $x \in W$.

Lemma 1.9. *If A is a self-adjoint endomorphism of V and $W \subseteq V$ is A-invariant then $W^\perp$ is also A-invariant.*

Proof. If $x \in W^\perp$ then for any $y \in W$, we have $\langle Ax,\, y\rangle = \langle x,\, Ay\rangle = 0$, since $x \in W^\perp$ and $Ay \in W$. The fact that $\langle Ax,\, y\rangle = 0$ for all $y \in W$ means that $Ax \in W^\perp$, hence $W^\perp$ is A-invariant as claimed. □

Lemma 1.10. *If A is a self-adjoint endomorphism of a finite-dimensional positive definite inner product space V, consider the quadratic function $Q_A(y) = \langle y,\, Ay\rangle$. Let $S(V) = \{y \in V \mid \|y\| = 1\}$ denote the unit sphere of V. The value $\lambda_{\min}(A) = \inf\{Q_A(y) \mid y \in S(V)\}$ is finite, and the set of points $x \in S(V)$ where $Q_A(x) = \lambda_{\min}(A)$ is non-empty. Any such x is an eigenvector of A, with eigenvalue $\lambda_{\min}(A)$. Furthermore, $\lambda_{\min}(A)$ is the minimum eigenvalue of A.*

Proof. The isomorphism between V and $\mathbb{R}^n$ defines a homeomorphism between $S(V)$ and the unit sphere of $\mathbb{R}^n$, so $S(V)$ is a compact topological space. Therefore, the minimum value of the continuous function Q_A on $S(V)$ is finite, and it is attained at a non-empty set of points.

Now consider any $x \in S(V)$ where Q_A attains its minimum, and consider any $y \in S(V)$ orthogonal to x. We can define a one-parameter family of vectors $x(t)$, parameterized by $t \in \mathbb{R}$, using the formula $x(t) = (\cos t)x + (\sin t)y$. Each of these vectors belongs to $S(V)$ because

$$\langle x(t),\, x(t)\rangle = (\cos^2 t)\langle x,\, x\rangle \;+\; (2\sin t\cos t)\langle x,\, y\rangle \;+\; (\sin^2 t)\langle y,\, y\rangle = \cos^2 t + \sin^2 t = 1.$$

The function $f(t) = Q_A(x(t))$ attains its minimum at $t = 0$, hence $f'(0) = 0$. Now

$$\begin{aligned}
f'(t) &= \left\langle \tfrac{dx(t)}{dt},\, Ax(t) \right\rangle + \left\langle x(t),\, \tfrac{dAx(t)}{dt} \right\rangle && \textit{[product rule]} \\
&= \left\langle \tfrac{dx(t)}{dt},\, Ax(t) \right\rangle + \left\langle x(t),\, A\left(\tfrac{dx(t)}{dt}\right) \right\rangle && \textit{[chain rule]} \\
&= \left\langle \tfrac{dx(t)}{dt},\, Ax(t) \right\rangle + \left\langle Ax(t),\, \tfrac{dx(t)}{dt} \right\rangle && \textit{[self-adjointness]} \\
&= 2\left\langle \tfrac{dx(t)}{dt},\, Ax(t) \right\rangle && \textit{[symmetry]} \\
&= 2\langle -(\sin t)x + (\cos t)y,\, (\cos t)Ax + (\sin t)Ay \rangle
\end{aligned}$$

Substituting $t = 0$ we find that

$$0 = f'(0) = 2\langle y,\, Ax \rangle.$$

We have shown that every vector orthogonal to x is also orthogonal to Ax. In other words, letting W denote the 1-dimensional linear subspace generated by x, we have $Ax \in (W^\perp)^\perp$. By Lemma 1.5, this means Ax is a scalar multiple of x, i.e. x is an eigenvector of A. Writing $Ax = \lambda x$, we can determine the value of λ using

$$\lambda = \lambda\langle x,\, x \rangle = \langle x,\, \lambda x \rangle = \langle x,\, Ax \rangle = Q_A(x) = \lambda_{\min}(A). \tag{1}$$

Finally, the lemma asserts that $\lambda_{\min}(A)$ is the minimum eigenvalue of A. This is true, because if $\tilde{\lambda}$ is any other eigenvalue of A and $\tilde{x} \in S(V)$ is a corresponding eigenvector, then we can repeat the calculation in (1) to deduce that $\tilde{\lambda} = Q_A(\tilde{x}) \geq Q_A(x)$. □

Combining Lemmas 1.9 and 1.10, we obtain a recipe for extracting all of the eigenvectors of A, with their eigenvalues arranged in increasing order.

Theorem 1.11. *Let A be a self-adjoint endomorphism of an n-dimensional positive definite inner product space V, and let us inductively define sequences*

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

by specifying that

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

Then $x_1, \ldots, x_n$ is an orthonormal basis of V consisting of eigenvectors of A, and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the corresponding eigenvalues.

Proof. The proof is by induction on i. The induction hypothesis is that $\{x_1, \ldots, x_i\}$ is an orthonormal basis of V_i consisting of eigenvectors of A, and that $\lambda_1 \leq \cdots \leq \lambda_i$ 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 1.9 its orthogonal complement W_{i-1} is also A-invariant and this implies, by Lemma 1.10 that x_i is an eigenvector of A and λ_i is its corresponding eigenvalue. Finally, $\lambda_i \geq \lambda_{i-1}$ because $\lambda_{i-1} = \min\{Q_A(x) \mid x \in W_{i-2}, \|x\| = 1\}$, while $\lambda_i = Q_A(x_i) \in \{Q_A(x) \mid x \in W_{i-2}, \|x\| = 1\}$. □

An easy corollary of Theorem 1.11 is the *Courant-Fischer Theorem.*

Theorem 1.12 (Courant-Fischer)**.** *The eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ of a self-adjoint endomorphism of an n-dimensional positive definite inner product space satisfy*

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

Proof. The vector space W_{k-1} constructed in the proof of Theorem 1.11 has dimension $n - k + 1$, and by construction it satisfies $\min_{x \in S(W_{k-1})} Q_A(x) = \lambda_k$. Therefore

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

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 $\dim(W) + \dim(V_k) > n$. Assuming $\|x\| = 1$ without loss of generality. Since $V_k = \operatorname{span}(x_1, \ldots, x_k)$ we can write $x = a_1 x_1 + \cdots + a_k x_k$. Then, using the fact that $x_1, \ldots, x_k$ are orthonormal eigenvectors of A, we obtain

$$Q_A(x) = \lambda_1 a_1^2 + \cdots + \lambda_k a_k^2 \leq \lambda_k,$$

where the last inequality follows because the coefficients $a_1^2, a_2^2, \ldots, a_k^2$ sum up to $\|x\|^2 = 1$, so the expression $\lambda_1 a_1^2 + \cdots + \lambda_k a_k^2$ represents a weighted average of $\lambda_1, \ldots, \lambda_k$. Starting from an arbitrary $(n - k + 1)$-dimensional subspace W we have constructed $x \in S(W)$ such that $Q_A(x) \leq \lambda_k$, so we can conclude that $\max_{\dim(W)=n-k+1} \left(\min_{x \in S(W)} Q_A(x) \right) \leq \lambda_k$. Combining this with the inequality derived in the preceding paragraph, we obtain $\max_{\dim(W)=n-k+1} \min_{x \in S(W)} Q_A(x) = \lambda_k$. Replacing A with $-A$, and k with $n - k + 1$, we obtain $\min_{\dim(V)=k} \left(\max_{x \in S(V)} Q_A(x) \right) = \lambda_k$. □

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 $c(u,v) \geq 0$, 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

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

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

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

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 $x \in \mathbb{R}^V$.

$$\langle x,\, L_G x \rangle = \sum_{(u,v) \in E} c(u,v)(x_u - x_v)^2. \tag{2}$$

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 the same nullspace as the Laplacian, so Lemma 2.3 applies to $\overline{L}_G$ without modification. There is also a nice interpretation of $\overline{L}_G$ as an endomorphism of $\mathbb{R}^V$, due to the observation that if $y = \overline{L}_G x$, then for every vertex v,

$$y_v = \frac{\sum_{u \neq v} c(u,v)(x_v - x_u)}{\sum_{u \neq v} c(u,v)} = x_v - \frac{\sum_{u \neq v} c(u,v) x_u}{\sum_{u \neq v} c(u,v)}.$$

Hence, the normalized Laplacian endormophism takes a vector $x \in \mathbb{R}^V$ and replaces each coordinate x_v with the difference between x_v and the weighted average of the neighboring values x_u, weighted by $c(u,v)$.

We will focus on normalized Laplacian eigenvalues and eigenvectors in these notes, because it turns out they enjoy a much tighter connection to the finding sparse cuts in G. The cost of working with the normalized Laplacian is that it is not a symmetric matrix. (More precisely, L_G is symmetric only when G is a regular graph, meaning that all of its vertices have the same degree.) Fortunately, the normalized Laplacian is self-adjoint with respect to an appropriate positive definite inner product. Define the degree-averaged inner product by

$$\langle x, y \rangle_d = \frac{1}{d(V)} \sum_{v \in V} d(v) x_v y_v \qquad \text{where} \qquad d(V) = \sum_{v \in V} d(v).$$

Lemma 2.4. *The normalized Laplacian $\overline{L}_G$ is self-adjoint with respect to the degree-averaged inner product.*

Proof. For all vectors x, y we have

$$d(V) \cdot \langle x, y \rangle_d = \sum_{v \in V} d(v) x_v y_v = x^\mathsf{T} D_G y$$

Hence,

$$\begin{aligned} d(V) \cdot \left\langle x, \overline{L}_G y \right\rangle_d &= x^\mathsf{T} D_G (D_G^{-1} L_G y) = x^\mathsf{T} L_G y \\ d(V) \cdot \left\langle \overline{L}_G x, y \right\rangle_d &= x^\mathsf{T} \overline{L}_G^\mathsf{T} D_G y = x^\mathsf{T} L_G^\mathsf{T} (D_G^{-1})^\mathsf{T} D_G y = x^\mathsf{T} L_G y \end{aligned}$$

which confirms that $\left\langle x, \overline{L}_G y \right\rangle_d = \left\langle \overline{L}_G x, y \right\rangle_d$ as required by the definition of self-adjointness. □

2.1 A probabilistic interpretation

The degree-averaged inner product and the normalized Laplacian have a useful probabilistic interpretation that will be useful later on.

Define a probability distribution π on vertices of G by specifying that the probability of vertex u is $\pi(u) = d(u)/d(V)$. Now consider the following two probability distributions on ordered pairs (u, v).

- The *product distribution* draws u and v independently from π; that is, it assigns probability $\pi(u)\pi(v)$ to the pair (u, v).

- The *edge distribution* samples a random edge of G with probability proportional to its weight and orients it randomly. In other words, it assigns probability $c(u, v)/d(V)$ to the pair (u, v).

In both distributions, the marginal distribution of u and the marginal distribution of v are both equal to π.

The two distributions give us two ways of averaging a function f defined on ordered pairs (u, v). We can take its average under the product distribution, obtaining

$$\mathbb{E}_\pi[f] = \sum_{u \in V} \sum_{v \in V} \pi(u)\pi(v) f(u, v)$$

or we can take its adverage under the edge distribution, obtaining

$$\mathbb{E}_G[f] = \sum_{u \in V} \sum_{v \neq u} \frac{c(u, v)}{d(V)} f(u, v).$$

Lemma 2.5. *For a vector $x \in \mathbb{R}^V$ we have the following identities.*

1. $\mathbb{E}_\pi[x_u] = \mathbb{E}_\pi[x_v] = \langle x, \mathbf{1} \rangle_d$.
2. $\mathbb{E}_\pi[x_u^2] = \mathbb{E}_\pi[x_v^2] = \langle x, x \rangle_d$.
3. $\frac{1}{2}\mathbb{E}_\pi[(x_u - x_v)^2] = \langle x, x \rangle_d - \left(\langle x, \mathbf{1} \rangle_d\right)^2$.
4. $\frac{1}{2}\mathbb{E}_G[(x_u - x_v)^2] = \left\langle x, \overline{L}_G x \right\rangle_d$.

Proof. Equations (1) and (2) follow by direct calculation.

$$\begin{aligned}
\langle x, \mathbf{1} \rangle_d &= \tfrac{1}{d(V)} \sum_{v \in V} d(v) x_v = \sum_{v \in V} \pi(v) x_v = \mathbb{E}_\pi[x_v] \\
\langle x, x \rangle_d &= \tfrac{1}{d(V)} \sum_{v \in V} d(v) x_v^2 = \sum_{v \in V} \pi(v) x_v^2 = \mathbb{E}_\pi[x_v^2].
\end{aligned}$$

Equation (3) follows from the fact that the variance of the sum of two independent random variables is the sum of of their variances. In this case, the two independent random variables are x_u and $-x_v$, the expected value of their sum is

$$\mathbb{E}_\pi[x_u - x_v] = \mathbb{E}_\pi[x_u] - \mathbb{E}_\pi[x_v] = 0$$

and the variance of their sum is

$$\mathrm{Var}(x_u - x_v) = \mathbb{E}_\pi[(x_u - x_v)^2] - \left(\mathbb{E}_\pi[x_u - x_v]\right)^2 = \mathbb{E}_\pi[(x_u - x_v)^2].$$

Hence,

$$\begin{aligned}
\frac{1}{2}\mathbb{E}_\pi[(x_u - x_v)^2] &= \frac{1}{2}\,\mathrm{Var}_\pi(x_u - x_v) = \frac{1}{2}\left[\mathrm{Var}_\pi(x_u) + \mathrm{Var}_\pi(-x_v)\right] \\
&= \frac{1}{2}\left[\mathrm{Var}_\pi(x_u) + \mathrm{Var}_\pi(x_v)\right] = \frac{1}{2}\left[2 \cdot \mathrm{Var}_\pi(x_u)\right] \\
&= \mathbb{E}_\pi[x_u^2] - \left(\mathbb{E}_\pi[x_u]\right)^2 = \langle x, x \rangle_d - \left(\langle x, \mathbf{1} \rangle_d\right)^2.
\end{aligned}$$

Finally,

$$\begin{aligned}
\frac{1}{2}\mathbb{E}_G[(x_u - x_v)^2] &= \frac{1}{2}\sum_{u\in V}\sum_{v\neq u}\frac{c(u,v)}{d(V)}\,(x_u - x_v)^2 \\
&= \frac{1}{2d(V)}\sum_{u\in V}\sum_{v\neq u} c(u,v)\,(x_u - x_v)x_u \;-\; \frac{1}{2d(V)}\sum_{u\in V}\sum_{v\neq u} c(u,v)\,(x_u - x_v)x_v \\
&= \frac{1}{2d(V)}\sum_{u\in V}\sum_{v\neq u} c(u,v)\,(x_u - x_v)x_u \;+\; \frac{1}{2d(V)}\sum_{u\in V}\sum_{v\neq u} c(u,v)\,(x_v - x_u)x_v \\
&= \frac{1}{d(V)}\sum_{u\in V}\sum_{v\neq u} c(u,v)\,(x_u - x_v)x_u \\
&= \frac{1}{d(V)}\sum_{u\in V} d(u)x_u\left(\frac{1}{d(u)}\cdot\sum_{v\neq u} c(u,v)(x_u - x_v)\right) \\
&= \frac{1}{d(V)}\sum_{u\in V} d(u)x_u(D_G^{-1}L_G x)_u = \left\langle x,\, \overline{L}_G x\right\rangle_d .
\end{aligned}$$

□

3 Sparsity and expansion

We will relate the eigenvalue $\nu_2(G)$ to two graph parameters called the *expansion* and the *sparsity* 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 *sparsity* of a vertex set S is

$$\phi(S) = \frac{c(\partial S)\cdot d(V)}{d(S)\cdot d(V-S)},$$

and the sparsity of G is $\phi(G) = \min\{\phi(S) : \emptyset \subsetneq S \subsetneq V\}$. The *expansion* of S is

$$h(S) = \frac{c(\partial S)}{\min\{d(S), d(V-S)\}},$$

and the expansion of G is $h(G) = \min\{h(s) : \emptyset \subsetneq S \subsetneq V\}$. Note that for any S,

$$\phi(S) = h(S)\cdot\frac{d(V)}{\max\{d(S), d(V-S)\}}.$$

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

$$h(S) \le \phi(S) \le 2 \cdot h(S),$$

and that the same relation holds between $h(G)$ and $\phi(G)$, although the sets that attain the minimum in the definitions of $h(G)$ and $\phi(G)$ may not be identical to one another.

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.

$$\begin{aligned}\nu_2(G) &= \min\left\{\left\langle x,\, \overline{L}_G x\right\rangle_d \;\middle|\; \|x\| = 1,\, \langle x,\, \mathbf{1}\rangle_d = 0\right\} \\ &= \min\left\{\frac{\left\langle y,\, \overline{L}_G y\right\rangle_d}{\langle y,\, y\rangle_d} \;\middle|\; y \neq 0,\, \langle y,\, \mathbf{1}\rangle_d = 0\right\}.\end{aligned}$$

Lemma 2.5 provides a helpful probabilistic interpretation of the quotient on the last line.

$$\nu_2(G) = \min\left\{\frac{\mathbb{E}_G[(y_u - y_v)^2]}{\mathbb{E}_\pi[(y_u - y_v)^2]} \;\middle|\; y \neq 0,\, \langle y,\, \mathbf{1}\rangle_d = 0\right\} = \min\left\{\frac{\mathbb{E}_G[(y_u - y_v)^2]}{\mathbb{E}_\pi[(y_u - y_v)^2]} \;\middle|\; y \nparallel \mathbf{1}\right\}.$$

The second equation follows from the fact that the numerator and denominator are unaffected by adding a scalar multiple of $\mathbf{1}$ to y.

Let us evaluate the quotient on the right side of (**??**) 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,

$$\mathbb{E}_G[(y_u - y_v)^2] = \frac{2c(\partial S)}{d(V)}$$

while

$$\mathbb{E}_\pi[(y_u - y_v)^2] = \Pr_\pi(u \in S,\, v \notin S) + \Pr_\pi(u \notin S,\, v \in S) = 2\pi(S)\pi(V - S).$$

Observe that $\pi(S) = \frac{d(S)}{d(V)}$ while $\pi(V - S) = \frac{d(V-S)}{d(V)}$. Hence,

$$\frac{\mathbb{E}_G[(y_u - y_v)^2]}{\mathbb{E}_\pi[(y_u - y_v)^2]} = \frac{c(\partial S)}{d(V)\pi(S)\pi(V - S)} \frac{c(\partial S) \cdot d(V)}{d(S) \cdot d(V - S)} = \phi(S).$$

Taking the minimum over all nonempty proper subsets of V 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

$$R(y) = \frac{\mathbb{E}_G[(y_u - y_v)^2]}{\mathbb{E}_\pi[(y_u - y_v)^2]}.$$

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

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

$$\sum_{y_i < 0} d(v_i) \le \sum_{y_i \ge 0} d(v_i)$$
$$\sum_{y_i \le 0} d(v_i) \ge \sum_{y_i > 0} d(v_i)$$

For 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

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

Note also that $R(y)$ 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

$$\mathbb{E}[c(\partial S)] \le \sqrt{2R(y)} \cdot \mathbb{E}[\min\{d(S), d(\overline{S})\}]$$

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

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

It's 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,

$$\mathbb{E}[\min\{d(S), d(\overline{S})\}] = \sum_u d(u)|z_u| = \sum_u d(u)y_u^2 = d(V)\langle y,\, y\rangle_d.$$

Meanwhile, to bound the numerator $\mathbb{E}[c(\partial S)]$, observe that an edge (u,v) contributes $c(u,v)$ to the numerator if and only if it is cut, an event having probability $|z_u - x_v|$. 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_u 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

$$\begin{aligned}
\mathbb{E}[c(\partial S)] &= \frac{1}{2} \sum_{(u,v)\in V^2} c(u,v) \cdot \Pr(\{u,v\} \in \partial S) = \frac{1}{2} \sum_{(u,v)\in V^2} c(u,v)|z_u - z_V| \\
&\leq \frac{1}{2} \sum_{(u,v)\in V^2} c(u,v)|y_u - y_v|(|y_u| + |y_v|) \\
&\leq \frac{1}{2} \left(\sum_{(u,v)\in V^2} c(u,v)(y_u - y_v)^2 \right)^{1/2} \left(\sum_{(u,v)\in V^2} c(u,v)(|y_u| + |y_v|)^2 \right)^{1/2} \\
&= \frac{1}{2} \left(d(V)\mathbb{E}_G[(y_u - y_v)^2] \right)^{1/2} \left(\sum_{(u,v)\in V^2} c(u,v)(|y_u| + |y_v|)^2 \right)^{1/2}.
\end{aligned}$$

We can bound the second factor on the right side using the inequality $(a+b)^2 \leq 2a^2 + 2b^2$

which is valid for all $a, b \in \mathbb{R}$.

$$\begin{aligned}
\mathbb{E}[c(\partial S)] &\leq \frac{1}{2}\left(d(V)\, R(y)\, \mathbb{E}_{\pi}[(y_u - y_v)^2]\right)^{1/2} \left(\sum_{(u,v)\in V^2} c(u,v)(2y_u^2 + 2y_v^2) \right)^{1/2} \\
&\leq \frac{1}{2}\left(2d(V)\, R(y)\, \langle y,\, y\rangle_d\right)^{1/2} \left(4 \sum_{u\in V} d(u) y_u^2 \right)^{1/2} \\
&= \frac{1}{2}\left(2d(V)\, R(y)\, \langle y,\, y\rangle_d\right)^{1/2} \left(4\, d(V)\, \langle y,\, y\rangle_d\right)^{1/2} \\
&= \sqrt{2R(y)}\, d(V)\, \langle y,\, y\rangle_d \\
&= \sqrt{2R(y)} \cdot \mathbb{E}[\min\{d(S), d(\overline{S})\}].
\end{aligned}$$

It follows there is at least one S in the support of our distribution such that

$$c(\partial S) \leq \sqrt{2R(y)} \cdot \min\{d(S), d(\overline{S})\}$$

and the right side is not zero. For this vertex set S, we find that

$$h(G) \leq h(S) = \frac{c(\partial S)}{\min\{d(S), d(\overline{S})\}} \leq \sqrt{2R(y)} = \sqrt{2\nu_2(G)},$$

which completes the proof of Cheeger's Inequality.