Spectral Graph Theory

Vignesh M Pai

Contents

1 Introduction **1**
- 1.1 Eigenvalues and Optimization . . . 1
- 1.2 The Laplacian and Graph Drawing . . . 2
- 1.3 Adjacency Matrices . . . 3
- 1.4 Comparing Graphs . . . 4

2 Eigenvalues and Eigevectors of Some Graphs **5**
- 2.1 Quotient Graphs . . . 5
- 2.2 Cycle Graph . . . 6
- 2.3 Path Graph . . . 6

3 Random Graphs **7**
- 3.1 Introduction . . . 7
- 3.2 Moments of Eigenvalues . . . 7
- 3.3 Expectation of The Trace . . . 8
- 3.4 The Number of Walks . . . 8
- 3.5 Putting It All Together . . . 9

4 Random Walks on Graphs **9**
- 4.1 Convergence . . . 10
- 4.2 Rate of Convergence . . . 11

5 Appendix **11**
- 5.1 Linear Algebraic Results . . . 11
- 5.2 Cauchy's Interlacing Theorem for Rank One Updates . . . 12
  - 5.2.1 The First Proof . . . 12
  - 5.2.2 The Second Proof . . . 13
- 5.3 Probability . . . 14

1 Introduction

1.1 Eigenvalues and Optimization

Let M be a n dimensional symmetric matrix.

Definition 1.1.1: Rayleigh Quotient

The Rayleigh quotient for a matrix M is defined as

$$R(\boldsymbol{x}, M) := \frac{\boldsymbol{x}^T M \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}$$

the matrix is ommitted if obvious from context.

Theorem 1.1.1

Let

$$\boldsymbol{x} \in \arg \max_{\boldsymbol{x} \in \mathbb{R}^n - \{0\}} R(\boldsymbol{x})$$

Such an $\boldsymbol{x}$ exists and is an eigenvector of M with the maximum eigenvalue μ_1. We can write a similar statement for the minimum eigenvalue by minimizing R.

Theorem 1.1.2: Spectral Theorem for Symmetric Matrices

There exist numbers $\mu_1, ..., \mu_n$ and orthonormal vectors $\boldsymbol{\psi}_1, ..., \boldsymbol{\psi}_n$ such that $M\boldsymbol{\psi}_i = \mu_i \boldsymbol{\psi}_i$ iff for $1 \leq i \leq n$

$$\boldsymbol{\psi}_i \in \arg \max_{\substack{\|\boldsymbol{x}\|=1 \\ \boldsymbol{x}^T \boldsymbol{\psi}_j = 0,\ j<i}} R(\boldsymbol{x})$$

or equivalently

$$\boldsymbol{\psi}_i \in \arg \min_{\substack{\|\boldsymbol{x}\|=1 \\ \boldsymbol{x}^T \boldsymbol{\psi}_j = 0,\ j>i}} R(\boldsymbol{x})$$

Theorem 1.1.3: Courant-Fischer Theorem

Let M have eigenvalues $\mu_1 \geq \mu_2 \geq ... \geq \mu_n$, then

$$\mu_k = \max_{\substack{S \subset \mathbb{R}^n \\ \dim(S)=k}} \min_{\substack{\boldsymbol{x} \in S \\ \boldsymbol{x} \neq 0}} R(\boldsymbol{x}) = \min_{\substack{T \subset \mathbb{R}^n \\ \dim(T)=n-k+1}} \max_{\substack{\boldsymbol{x} \in T \\ \boldsymbol{x} \neq 0}} R(\boldsymbol{x})$$

where S, T are subspaces of $\mathbb{R}^n$.

Theorem 1.1.4: Cauchy's Interlacing Theorem

Let A be a symmetric real matrix of dimension n. Let B be obtained by deleting the same row and column of A (N is a principal submatrix of dimension $n-1$). Let $\alpha_1 \geq ... \geq \alpha_n$ be the eigenvalues of A and $\beta_1 \geq ... \geq \beta_{n-1}$ be the eigenvalues of B. Then for $1 \leq k \leq n-1$

$$\alpha_k \geq \beta_k \geq \alpha_{k+1}$$

1.2 The Laplacian and Graph Drawing

Vectors on graphs are functions $V \to \mathbb{R}$, the vector $\mathbf{1}$ denotes the function $\mathbf{1}(a) = 1$. The degree function $\boldsymbol{d}$ is also a vector.

Matrices on graphs are functions $V \times V \to \mathbb{R}$ or can be viewed as linear operators on the space of vectors on graphs. Let G be a graph, then we use M_G to denote the adjacency matrix, D_G to denote the diagonal matrix of vertex degrees, and $L_G = D_G - M_G$ to denote the Laplacian matrix. Observe that $M_G \mathbf{1} = \boldsymbol{d}$ and $L_G \mathbf{1} = \mathbf{0}$.

The Laplacian is also a natural quadratic form on a graph

$$\boldsymbol{x}^T L_G \boldsymbol{x} = \sum_{(a,b) \in E(G)} w_{a,b} (\boldsymbol{x}(a) - \boldsymbol{x}(b))^2$$

Let $\lambda_1 = 0 \leq \lambda_2 \leq ... \leq \lambda_n$ be the eigenvalues of L_G with eigenvectors $\boldsymbol{\psi}_1, ..., \boldsymbol{\psi}_n$.

Lemma 1.2.1

G is connected iff $\lambda_2 \neq 0$.

We draw a graph in k dimensions by using the eigenvectors corresponding to $\lambda_2, ..., \lambda_{k+1}$ as the coordinates of vertices. These coordinates minimize the expression (excluding coordinates that lead to trivial drawings):

$$\sum_{(a,b)\in G} w_{a,b} \|x(a) - x(b)\|^2 = \sum_{i=1}^{k} \boldsymbol{x_i}^T L_G \boldsymbol{x_i}$$

where $x : V \to \mathbb{R}^k, x = (\boldsymbol{x_1}, ..., \boldsymbol{x_k})$ is the coordinate function.

Theorem 1.2.2: Hall's Drawing Theorem

Let $\boldsymbol{x_1}, ..., \boldsymbol{x_k}$ be orthonormal vectors that are orthogonal to $\mathbf{1}$, then

$$\sum_{i=1}^{k} \boldsymbol{x_i}^T L_G \boldsymbol{x_i} \geq \sum_{i=2}^{k+1} \lambda_1$$

where equality holds for $\boldsymbol{\psi}_j^T \boldsymbol{x_i} = 0$ for all i and $j > k + 1$.

1.3 Adjacency Matrices

Let the eigenvalues of M_G be $\mu_1 \geq ... \geq \mu_n$.

Lemma 1.3.1

Let d_{avg} and d_{max} be the average and maximum degrees respectively, then

$$d_{avg} \leq \mu_1 \leq d_{max}$$

further, if H be a subgraph of G, then

$$d_{avg}(H) \leq \mu_1$$

Lemma 1.3.2

If G is connected and $\mu_1 = d_{max}$ then G is d_{max}-regular.

Theorem 1.3.3: Wilf's Theorem

Let $\chi(G)$ be the chromatic number of G, then

$$\chi(G) \leq \lfloor \mu_1 \rfloor + 1$$

Lemma 1.3.4

Let G be a connected weighted graph and $\boldsymbol{\phi}$ be a non negative eigenvector of M_G, then $\boldsymbol{\phi}$ is strictly positive.

Theorem 1.3.5: Perron-Frobenius Theorem

Let G be connected, then

1. μ_1 has a strictly positive eigenvector
2. $\mu_1 \geq -\mu_n$
3. $\mu_1 > \mu_2$

Lemma 1.3.6

If G is bipartite, then the eigenvalues of M_G are symmetric about 0.

Theorem 1.3.7

Let G be connected, $\mu_1 = -\mu_n$ iff G is bipartite.

1.4 Comparing Graphs

We introduce the partial order $\succeq$ on matrices as

$$A \succeq B \iff \forall \boldsymbol{x}, \boldsymbol{x}^T A \boldsymbol{x} \geq \boldsymbol{x}^T B \boldsymbol{x}$$

In particular $A \succeq 0$ means A is positive semidefinite. For graphs G, H on the same set of vertices, we write $G \succeq H$ iff $L_G \succeq L_H$. If H is a subset of G, we have

$$G \succeq H$$

For a graph H, define $c \cdot H$ to be the graph H with each edge weight multiplied by c. Let $\lambda_k(H)$ denote the kth smallest eigenvalue of L_H.

Lemma 1.4.1

If G, H are graphs

$$G \succeq c \cdot H \implies \lambda_k(G) \geq c\lambda_k(H)$$

Let $G_{a,b}$ be the graph with only the edge (a, b). The proof of the following lemmas follow trivially from the Cauchy Schwarz inequality applied to the Laplacian.

Lemma 1.4.2

Let P_n be the path graph on n vertices between vertex 1 and n.

$$G_{1,n} \preceq (n-1)P_n$$

Lemma 1.4.3: Extension to Weighted Paths

Let $P_{n,w}$ be the weighted path graph on n vertices with w_i the weight on the edge $(i, i+1)$.

$$G_{1,n} \preceq \left(\sum_{i=1}^{n-1} \frac{1}{w_i} \right) P_{n,w}$$

We use these lemmas to prove bounds on the eigenvalues of graphs.

Let K_n be the complete graph on n vertices, it is easy to see that $\lambda_i(K_n) = n$ for $i \geq 2$. We can also write that

$$L_{K_n} = \sum_{a<b} L_{G_{a,b}}$$

this allows us to prove the following

$$\begin{aligned} K_n = \sum_{a<b} G_{a,b} \preceq \sum_{a<b} (b-a) P_{a,b} \preceq \sum_{a<b} (b-a) P_n \\ \implies \lambda_2(K_n) = n \leq \lambda_2(P_n) \sum_{a<b} (b-a) = \frac{n(n+1)(n-1)}{6} \lambda_2(P_n) \end{aligned}$$

Lemma 1.4.4: Bounding λ_2 of the Path Graph

$$\lambda_2(P_n) \geq \frac{6}{(n+1)(n-1)}$$

Let T_d be the complete binary tree of depth d, this will have $n = 2^{d+1} - 1$ vertices. Let $T_d^{a,b}$ be the shortest path between vertices a, b on T_d. Note that this path has size atmost $2d \leq 2\log_2 n$. Doing a similar comparision with K_n we get

$$K_n = \sum_{a<b} G_{a,b} \preceq \sum_{a<b} 2dT_d^{a,b} \preceq \sum_{a<b} 2\log_2 nT_d = \binom{n}{2} 2\log_2 nT_d$$
$$\implies \lambda_2(K_n) = n \leq \binom{n}{2} 2\log_2 n\lambda_2(T_d)$$

Lemma 1.4.5: Bounding λ_2 of Complete Binary Trees

$$\lambda_2(T_d) \geq \frac{1}{(n-1)\log_2 n}$$

2 Eigenvalues and Eigevectors of Some Graphs

2.1 Quotient Graphs

Definition 2.1.1: Equitable Partition

Given a graph G, let $C = (C_1, C_2, ..., C_k)$ be a partition of $V(G)$, the elements of the partition are called cells. Such a partition is said to equitable if the number of edges between a fixed vertex in C_i and C_j is independent of that fixed vertex, we denote this number by c_{ij}.

Lemma 2.1.1

A partition is equitable iff the induced subgraph of each cell is regular and the edges between two cells form a biregular graph.

As a particular example, the orbits in the action of the automorphism group of a graph form an equitable partition.

Definition 2.1.2: Quotient Graph

Given a partition C of G, the quotient G/C is defined to be the graph with vertices C and a cell C_i has a directed edge to C_j of weight c_{ij}.

In general the quotient graph has loops, non unit weights and cannot be represented as an undirect graph ($c_{ij} \neq c_{ji}$). We will show that the characteristic polynomial of $M_{G/C}$ divides that of M_G.

If $|G| = n, |C| = k$, the characteristic matrix $P(C)$ is a (n, k) matrix whose (i, j) entry is whether the ith vertex of G is contained in C_j. this gives us a computational way to represent C.

Lemma 2.1.2

If C is an equitable partition of G, then $M_G P(C) = P(C)M_{G/C}$. Conversely, if there exists a B such that $M_G P(C) = P(C)B$, then C is equitable.

This lemma can be rephrased as: C is equitable iff the column space of $P(C)$ is an invariant subspace of M_G. The column space of $P(C)$ is the vector space of functions on G that are constant on the cells of C.

Corollary 2.1.3

If C is equitable, then $L_G P(C) = P(C) L_{G/C}$. Conversely, if there exists a B such that $L_G P(C) = P(C)B$, then C is equitable.

The Laplacian of a graph is invariant under addition and removal of loops (this is not true for the adjacency matrix). Therefore, the above corollary is also true for G/C replaced by the same graph with its loops removed.

Theorem 2.1.4

Let C be an equitable partition of G, then for $A \in \{M, L\}$

1. $A_{G/C}\boldsymbol{x} = \lambda \boldsymbol{x} \implies A_G P\boldsymbol{x} = \lambda P\boldsymbol{x}$
2. $A_G\boldsymbol{y} = \lambda \boldsymbol{y} \implies \boldsymbol{y}^T P A_{G/C} = \lambda \boldsymbol{y}^T P$
3. The characteristic polynomial of $A_{G/C}$ divides the characteristic polynomial of A.

When $A_{G/C}$ is symmetric, the second condition can be rewritten as

$$A_G\boldsymbol{y} = \lambda y \implies A_{G/C} P^T \boldsymbol{y} = \lambda P^T \boldsymbol{y}$$

2.2 Cycle Graph

Represent the vertices of C_n as $\{0, 1, ..., n-1\}$.

Theorem 2.2.1

The cycle graph C_n has eigenvectors

$$\boldsymbol{x}_k(u) = \cos\left(\frac{2\pi k u}{n}\right)$$
$$\boldsymbol{y}_k(u) = \sin\left(\frac{2\pi k u}{n}\right)$$

for $0 \leq k \leq n/2$, ignoring $\boldsymbol{y}_0$ which is $\mathbf{0}$, with eigenvalue $2 - 2\cos\frac{2\pi k}{n}$.

2.3 Path Graph

Consider C_{2n} and the partition σ defined as $(\{0, 2n-1\}, \{1, 2n-2\}, ...\{n-1, n\})$. Clearly this partition is equitable and C_{2n}/σ is P_n with loops at the endpoint vertices. From the previous section, we can write

$$L_{C_{2n}} P = P L_{P_n}$$

Since L_{P_n} is symmetric, the the following must be eigenvectors of P_n

$$\begin{aligned}
P^T \boldsymbol{x}_k(u) &= \cos\left(\frac{\pi k u}{n}\right) + \cos\left(\frac{\pi k(2n-1-u)}{n}\right) \\
&= \cos\left(\frac{\pi k u}{n}\right) + \cos\left(\frac{\pi k(u+1)}{n}\right) \\
&= 2\cos\left(\frac{\pi k}{2n}\right)\cos\left(\frac{\pi k(u+\frac{1}{2})}{n}\right) \\
P^T \boldsymbol{y}_k(u) &= \sin\left(\frac{\pi k u}{n}\right) + \sin\left(\frac{\pi k(2n-1-u)}{n}\right) \\
&= \sin\left(\frac{\pi k u}{n}\right) - \sin\left(\frac{\pi k(u+1)}{n}\right) \\
&= -2\sin\left(\frac{\pi k}{2n}\right)\cos\left(\frac{\pi k(u+\frac{1}{2})}{n}\right)
\end{aligned}$$

Hence the eigenvector corresponding to eigenvalue $2 - 2\cos\frac{\pi k}{n}$ is

$$\boldsymbol{\phi}_k(u) = \cos\left(\frac{\pi k(u+\frac{1}{2})}{n}\right)$$

3 Random Graphs

3.1 Introduction

An Erdös-Rényi random graph is a graph in which each edge is present with probability p, independent of other edges. We discuss the eigenvalues of the adjacency matrix. The largest eigenvalue of such a graph is close to pn and all other eigenvalues are usually atmost of order $\sqrt{np}$. Our goal is to prove formal results regarding the eigenvalues.

We can write the adjacency matrix M of this graph as

$$M = \mathbb{E}(M) + R = p(J - I) + R$$

where J is the all ones matrix, I is the identity matrix and R is defined for off diagonal lower triangular entries as

$$R(a,b) = \begin{cases} 1-p & \text{with probability } p \\ -p & \text{with probability } 1-p \end{cases}$$

The upper triangular entries are determined from the symmetry of R. Clearly the expectation of R is the zero matrix. We will show that the eigenvalues of R are usually small and thus M is approximately $p(J-I)$. Note that the eigenvalues of $p(J-I)$ are $p(n-1)$ with multiplicity 1 and $-p$ with multiplicity $n-1$.

Let the eigenvalues of R be $\rho_1 \geq ... \geq \rho_n$. Then, the eigenvalues of $R - pI$ are $\rho_i - p$. Since pJ is rank one, the eigenvalues of $M = R - pI + pJ$ interlace the eigenvalues of $R - pI$ by the Cauchy interlacing theorem. Hence, the eigenvalues of R can give bounds on the eigenvalues of M.

3.2 Moments of Eigenvalues

We can get some quantitative results on the eigenvalues of R by calculating the moments of ρ_i. It is easy to verify that the kth moment is

$$\sum_{i=1}^{n} \rho_i^k = \mathrm{tr}(R^k)$$

We know that the eigenvalues of R^k are ρ_i^k. From the value of the trace, we know that for even k

$$\rho_1^k \leq Tr(R^k) \implies |\rho_1| \leq Tr(R^k)^{1/k}$$

If we calculate a bound: $\mathbb{E}(Tr(R^k)) < u$, we can write using Markov's inequality

$$\begin{aligned} Pr(|\rho_1| > (1+\epsilon)u^{1/k}) &= Pr(\rho_1^k > (1+\epsilon)^k u) \\ &\leq Pr(Tr(R^l) > (1+\epsilon)^k \mathbb{E}(Tr(R^k))) \\ &\leq (1+\epsilon)^{-k} \end{aligned}$$

The probability on the RHS will be small if $\epsilon > 1/k$. This will in particular be useful if for large k we can find small u.

3.3 Expectation of The Trace

Let M be a matrix, then we can write

$$M^l(a,b) = \sum_{v_1,\dots,v_{k-1}\in V} M(a,v_1)M(v_1,v_2)\dots M(v_{l-1},b)$$

Therefore, we can write

$$\mathbb{E}(R^l(a_0,a_0)) = \sum_{a_1,\dots,a_{l-1}\in V} \mathbb{E}(R(a_0,a_1)R(a_1,a_2)\dots R(a_l,a_0))$$

Note that distinct elements of R are independent, therefore the only dependence comes from terms that appear multiple times. Let S be the multiset $\{\{a_0,a_1\},\{a_1,a_2\},\dots,\{a_{l-1},a_0\}\}$ and S' denote the set of the same elements, then

$$\mathbb{E}(R(a_0,a_1)R(a_1,a_2)\dots R(a_l,a_0)) = \prod_{\{b_i,c_i\}\in S'} \mathbb{E}(R(b_i,c_i)^{d_i})$$

where d_i denotes the number of $\{b_i,c_i\}$ in S. The RHS is clearly zero if any d_i is 1. For $d \geq 2$, we can write

$$\mathbb{E}(R(b_i,c_i)^d) = (1-p)^d p + (-p)^d(1-p) = p(1-p)\left((1-p)^{d-1} - (-p)^{d-1}\right) \leq p(1-p)$$

this gives us

$$\mathbb{E}(R(a_0,a_1)R(a_1,a_2)\dots R(a_l,a_0)) \leq (p(1-p))^{|S'|}$$

We define a closed walk of length l as a sequence of vertices $a_0,a_1,\dots,a_{l-1},a_l$ such that $a_l = a_0$. A closed walk is called significant if the multiset $\{\{a_0,a_1\},\{a_1,a_2\},\dots\{a_{l-1},a_l\}\}$ has atleast two copies of each element.

Definition 3.3.1

Let $W_{n,l,k}$ denote the number of significant closed walks of length l in a graph with n vertices such that the above multiset has k distinct elements.

Lemma 3.3.1

$$\mathbb{E}(\mathrm{tr}(R^l)) \leq \sum_{k=1}^{l/2} W_{n,l,k}(p(1-p))^k$$

3.4 The Number of Walks

Lemma 3.4.1

$$W_{n,l,k} \leq n^{k+1}2^l l^{4(l-2k)}$$

3.5 Putting It All Together

Using the bound on the number of walks, we can write

$$
\begin{aligned}
\mathbb{E}(\operatorname{tr}(R^l)) &\le \sum_{k=1}^{l/2} n^{k+1} 2^l l^{4(l-2k)} (p(1-p))^k \\
&= n 2^l l^{4l} \sum_{k=1}^{l/2} (np(1-p)l^{-8})^k \\
&= n(4np(1-p))^{l/2} \sum_{k=1}^{l/2} (np(1-p)l^{-8})^{1-k}
\end{aligned}
$$

for $l^8 \le np(1-p)$, we get

$$
\begin{aligned}
\mathbb{E}(\operatorname{tr}(R^l)) &\le n(4np(1-p))^{l/2} \sum_{k=1}^{l/2} 2^{1-k} \\
&\le 2n(4np(1-p))^{l/2}
\end{aligned}
$$

Substituting back in the markov identity, we get

$$
Pr(|\rho_1| \ge (1+\epsilon)(2n)^{1/l}(4np(1-p))^{1/2}) \le (1+\epsilon)^{-l}
$$

Note that $(2n)^{1/l}$ can be written as

$$
(2n)^{1/l} = 2^{\log_2 2n/l}
$$

We get the required result, that the eigenvalues are of order $\sqrt{np}$, if l satisfies

$$
1 + \log_2 n < l < (np(1-p))^{1/8}
$$

Such an l can be found for sufficiently large n, for $p = 0.5$, n roughly needs to satisfy

$$
n^{1/8} > 1 + \log_2 n
$$

This is not satisfied for $n = 2^{40} \approx 10^{12}$. Therefore while asymptotically useful, this result may not be practically useful.

4 Random Walks on Graphs

A random walk on a weighted undirected graph G is a walk that randomly moves to an adjacent vertex with probability proportional to the weight of the edge. The probability density at time $t+1$ can be written as

$$
\begin{aligned}
\boldsymbol{p}_{t+1}(a) &= \sum_{b:(a,b)\in E(G)} \frac{w(a,b)}{\boldsymbol{d}(b)} \boldsymbol{p}_t(b) \\
\implies \boldsymbol{p}_{t+1} &= MD^{-1}\boldsymbol{p}_t
\end{aligned}
$$

The walk matrix W is defined as MD^{-1}. The lazy walk matrix is defined as

$$
\widetilde{W} = \frac{1}{2}I + \frac{1}{2}W
$$

The lazy walk remains at the current vertex with probability half, this is more well behaved than the walk matrix (in terms of convergence of $\boldsymbol{p}_t$).

4.1 Convergence

The walk matrix is similar to the normalized adjacency matrix which is symmetric and thus has nice properties of eigenvectors

$$\widetilde{M} = D^{-\frac{1}{2}} W D^{\frac{1}{2}} = D^{-\frac{1}{2}} M D^{-\frac{1}{2}}$$

Lemma 4.1.1

$\boldsymbol{\psi}$ is an eigenvector of $\widetilde{M}$ iff $D^{\frac{1}{2}}\boldsymbol{\psi}$ is an eigenvector of W, further they have same eigenvalue and the eigenvectors of W are the same as $\widetilde{W}$.

Note that $\boldsymbol{d}$ is a positive eigenvector of W with eigenvalue 1. By Perron-Frobenius theorem, the eigenvalues of W must lie between -1 and 1. The eigenvalues of $\widetilde{W}$ are simple of the form $(1+\lambda_i)/2$ where λ_i are eigenvectors of W. Hence, the eigenvalue of $\widetilde{W}$ lie between 0 and 1.

Clearly for a probability distribution to be stable with respect to the lazy random walk, it must be an eigenvector of $\widetilde{W}$ of eigenvalue 1. This distribution is

$$\boldsymbol{\pi} = \frac{\boldsymbol{d}}{\boldsymbol{1}^T \boldsymbol{d}}$$

Now we will prove that the probability distribution of the lazy random walk converges to $\boldsymbol{\pi}$. Let $\boldsymbol{\psi}_i$ be the orthonormal eigenvectors of $\widetilde{M}$ corresponding to eigenvalues ω_i of $\widetilde{W}$ $(1 = \omega_1 > \omega_2 \geq \ldots \geq 0)$. The eigenvalues of $\widetilde{M}$ are then $2\omega_i - 1$. Note that

$$\boldsymbol{\psi}_1 \propto D^{-\frac{1}{2}} \boldsymbol{d} = \sqrt{\boldsymbol{d}} \implies \boldsymbol{\psi}_1 = \frac{\sqrt{\boldsymbol{d}}}{\|\sqrt{\boldsymbol{d}}\|}$$

and we can write $\boldsymbol{p}_0$ as

$$D^{-\frac{1}{2}} \boldsymbol{p}_0 = \sum_i c_i \boldsymbol{\psi}_i$$

and in particular c_1 is

$$c_1 = \boldsymbol{\psi}_1^T D^{-\frac{1}{2}} \boldsymbol{p}_0 = \frac{\sqrt{\boldsymbol{d}}^T}{\|\sqrt{\boldsymbol{d}}\|} (D^{-\frac{1}{2}} \boldsymbol{p}_0) = \frac{\boldsymbol{1}^T \boldsymbol{p}_0}{\|\sqrt{\boldsymbol{d}}\|} = \frac{1}{\|\sqrt{\boldsymbol{d}}\|}$$

Now, we can write $\boldsymbol{p}_t$ as

$$\begin{aligned}
\boldsymbol{p}_t &= \widetilde{W}^t \boldsymbol{p}_0 \\
&= D^{\frac{1}{2}} \left(D^{-\frac{1}{2}} \widetilde{W} D^{\frac{1}{2}} \right)^t D^{-\frac{1}{2}} \boldsymbol{p}_0 \\
&= D^{\frac{1}{2}} \left(\frac{1}{2} I + \frac{1}{2} \widetilde{M} \right)^t \sum_i c_i \boldsymbol{\psi}_i \\
&= D^{\frac{1}{2}} \sum_i c_i \omega_i^t \boldsymbol{\psi}_i \\
&= D^{\frac{1}{2}} c_1 \boldsymbol{\psi}_1 + D^{\frac{1}{2}} \sum_{i \geq 2} c_i \omega_i^t \boldsymbol{\psi}_i
\end{aligned}$$

Note that for sufficiently large t, the second term goes to zero as $\omega_i < 1$ for $i \geq 2$. Evaluating the first term, we get

$$\lim_{t \to \infty} \boldsymbol{p}_t = D^{\frac{1}{2}} c_1 \boldsymbol{\psi}_1 = D^{\frac{1}{2}} \frac{\sqrt{\boldsymbol{d}}}{\|\sqrt{\boldsymbol{d}}\|^2} = \frac{\boldsymbol{d}}{\boldsymbol{1}^T \boldsymbol{d}} = \boldsymbol{\pi}$$

4.2 Rate of Convergence

We measure the rate of convergence pointwise, by calculating upper bounds on $|\boldsymbol{p}_t(b) - \boldsymbol{\pi}(b)|$. Assuming that $\boldsymbol{p}_0 = \boldsymbol{\delta}_a$

$$\begin{aligned}
\boldsymbol{p}_t &= \boldsymbol{\pi} + D^{\frac{1}{2}} \sum_{i\geq 2} c_i \omega_i^t \boldsymbol{\psi}_i \\
\implies \boldsymbol{p}_t(b) - \boldsymbol{\pi}(b) &= \boldsymbol{\delta}_b^T D^{\frac{1}{2}} \sum_{i\geq 2} \left(\boldsymbol{\psi}_i^T D^{-\frac{1}{2}} \boldsymbol{\delta}_a\right) \omega_i^t \boldsymbol{\psi}_i \\
&= \sqrt{\boldsymbol{d}(b)} \boldsymbol{\delta}_b^T \sum_{i\geq 2} \frac{(\boldsymbol{\psi}_i^T \boldsymbol{\delta}_a)}{\sqrt{d(a)}} \omega_i^t \boldsymbol{\psi}_i \\
&= \sqrt{\frac{\boldsymbol{d}(b)}{\boldsymbol{d}(a)}} \sum_{i\geq 2} \omega_i^t (\boldsymbol{\psi}_i^T \boldsymbol{\delta}_a)(\boldsymbol{\delta}_b^T \boldsymbol{\psi}_i)
\end{aligned}$$

We can bound the term on the right like

$$\begin{aligned}
\left| \sum_{i\geq 2} \omega_i^t (\boldsymbol{\psi}_i^T \boldsymbol{\delta}_a)(\boldsymbol{\delta}_b^T \boldsymbol{\psi}_i) \right| &\leq \sum_{i\geq 2} \omega_i^t |\boldsymbol{\psi}_i^T \boldsymbol{\delta}_a| |\boldsymbol{\delta}_b^T \boldsymbol{\psi}_i| \\
&\leq \omega_2^t \sum_{i\geq 1} |\boldsymbol{\psi}_i^T \boldsymbol{\delta}_a| |\boldsymbol{\delta}_b^T \boldsymbol{\psi}_i| \\
&\leq \omega_2^t \|\boldsymbol{\delta}_a\| \|\boldsymbol{\delta}_b\| \\
&\leq \omega_2^t
\end{aligned}$$

Theorem 4.2.1

Given $p_0 = \boldsymbol{\delta}_a$, we can write

$$|\boldsymbol{p}_t(b) - \boldsymbol{\pi}(b)| \leq \sqrt{\frac{\boldsymbol{d}(b)}{\boldsymbol{d}(a)}} \omega_2^t$$

5 Appendix

5.1 Linear Algebraic Results

Theorem 5.1.1

Let A, B be commuting symmetric matrices on a vector space V, then they can be simultaneously diagonalized by the same unitary transformation.

Proof. We will prove by induction on the dimension of V that A, B have the same eigenspaces, orthonormal basis of these eigenspaces will then provide the unitary transformation.

Clearly the statement is true for $\dim V = 1$, let $\dim V = n > 1$ and assume the statement is true for all V such that $\dim V < n$. Let $\alpha_1, ..., \alpha_k$ be the distinct eigenvalues of A. If $k = 1$, then $A = \lambda_1 I$ and thus we are done. So assume $k \geq 2$, let Λ_i be the eigenspace corresponding to the eigenvalue α_i. Then, Λ_i is an invariant subspace of B since for any $\boldsymbol{v} \in \Lambda_i$

$$A(B\boldsymbol{v}) = BA\boldsymbol{v} = \lambda_i B\boldsymbol{v} \implies B\boldsymbol{v} \in \Lambda_i$$

Thus, we can restrict A, B to λ_i and apply the induction hypothesis (since $k \geq 2 \implies \dim \Lambda_i < n$) to get an eigenbasis of Λ_i that simultaneously diagonalizes A, B over Λ_i. The union of these eigenbases over all Λ_i gives us a basis for V that simultaneously diagonalizes both A, B. □

The above theorem is helpful in calculating the eigenvalues of $A + B$.

Theorem 5.1.2

Let M be a symmetric matrix with eigenvalues λ_i, then

$$\sum_i \lambda_i = Tr(M)$$

Proof. Let D be the diagonal matrix of M diagonalized by U, then

$$M = UDU^T \implies Tr(M) = Tr((UD)U^T) = Tr(U^T(UD)) = Tr(D) = \sum_i \lambda_i$$

□

5.2 Cauchy's Interlacing Theorem for Rank One Updates

The result that we want to prove is

Theorem 5.2.1: Cauchy Interlacing Theorem for Rank One Updates

Let A be a symmetric matrix with eigenvalues $\alpha_1 \geq ... \geq \alpha_n$. Let $B = A + \boldsymbol{x}\boldsymbol{x}^T$ for some vector $\boldsymbol{x}$, let B have eigenvalues $\beta_1 \geq ... \geq \beta_n$, then

$$\beta_i \geq \alpha_i \geq \beta_{i+1}$$

We will do this in two ways. In the first proof, we show that A is a principal submatrix of a matrix with 'similar' eigenvalues to B, this will allow us to apply Cauchy's interlacing theorem for principal submatrices. In the second proof, we will relate the characteristic polynomials and their roots.

5.2.1 The First Proof

We first prove a couple of lemmas that are known as 'Sylvester's theorem'.

Lemma 5.2.2

Let A be an $m \times n$ matrix and B be an $n \times m$ matrix, then we can write

$$x^n \det(xI_m - AB) = x^m \det(xI_n - BA)$$

Proof. Define the following matrices

$$C = \begin{bmatrix} xI_m & A \\ B & I_n \end{bmatrix}, \quad D = \begin{bmatrix} I_m & 0 \\ -B & xI_n \end{bmatrix}$$

we can see that

$$CD = \begin{bmatrix} xI_m - AB & xA \\ 0 & xI_n \end{bmatrix}, \quad DC = \begin{bmatrix} xI_m & A \\ 0 & xI_n - BA \end{bmatrix}$$

using $\det(CD) = \det(DC)$, we get the required result. □

Lemma 5.2.3

Let A be an $m \times n$ matrix and B be an $n \times m$ matrix, then AB and BA have the same set of non zero eigenvalues.

Proof. This follows trivially from the previous lemma or from the following proof. Let $\boldsymbol{x}$ be an eigenvector of AB with eigenvalue $\lambda \neq 0$, it is simple to check that $\boldsymbol{y} = B\boldsymbol{x}$ is an eigenvector of BA with eigenvalue λ. □

Lemma 5.2.4

For positive definite symmetric matrices A, a matrix $\sqrt{A}$ such that $\sqrt{A}^2 = A$ exists.

Proof. Existence follows since the diagonalized matrix of A has a square root. □

We are now ready to prove the theorem.

Proof. We will prove the result for positive definite symmetric matrices A by adding $(1-\alpha_n)I$ to A.
Let $\sqrt{A}$ be a square root of A, now define the $n \times (n+1)$ matrix C as

$$C := \begin{bmatrix} \sqrt{A} & \boldsymbol{x} \end{bmatrix}$$

then we can write

$$CC^T = B, \quad C^TC = \begin{bmatrix} A & \sqrt{A}x \\ x^T\sqrt{A} & x^Tx \end{bmatrix}$$

By previous lemma, C^TC and CC^T share the same set of nonnegative eigenvalues. The eigenvalues of A interlace the eigenvalues of C^TC since it is a principal submatrix. The eigenvalues of C^TC are thus all positive except the smallest (which is 0). Since the eigenvalues of $CC^T = B$ are positive, the result follows. □

5.2.2 The Second Proof

Definition 5.2.1: Cyclic Vectors

Let A be a linear operator on a vector space V. $\boldsymbol{x} \in V$ is called cyclic if the set of finite linear combinations of $\{A^n\boldsymbol{x} \mid n \in \mathbb{N}_0\}$ equals V.

The above definition is equivalent to saying that $\boldsymbol{x} \in V$ is cyclic if

$$V = \{p(A)\boldsymbol{x} \mid p \in \mathbb{R}[t]\}$$

Lemma 5.2.5

A symmetric linear operator A on a vector space V has a cyclic vector iff A has no repeated eigenvalues.

Proof. Suppose that $\boldsymbol{x}$ is a cyclic vector, then

$$V = \{p(A)\boldsymbol{x} \mid p \in \mathbb{R}[t]\}$$

Fix an orthonormal basis on V, since A is symmetric, it is diagonalizable: $A = UDU^T$ where D is diagonal and U is orthogonal. We can then rewrite the above condition as

$$V = \{p(D)\boldsymbol{y} \mid p \in \mathbb{R}[t]\}, \quad (\boldsymbol{y} = U^T\boldsymbol{x})$$

Note that $\{p(D) \mid p \in R[t]\}$ is a vector space with dimension equal to the number of distinct eigenvalues of A. Therefore, $\{p(D)\boldsymbol{y} \mid p \in \mathbb{R}[t]\}$ has the same dimension and we cannot have repeated eigenvalues.
Conversely, suppose A has no repeated eigenvalues, then $\boldsymbol{x} = U\boldsymbol{y}$ corresponding to $\boldsymbol{y} = \mathbf{1}$ is a cyclic vector (it is simple to show that each eigenvector can be written in the form $p(D)\boldsymbol{y}$ for some p, and thus $V = \{p(D)\boldsymbol{y} \mid p \in \mathbb{R}[t]\}$). □

Definition 5.2.2

Let A be a matrix, then $\chi(A)$ represents the characteristic polynomial of A given by

$$\chi(A)(z) = \det(zI - A)$$

Lemma 5.2.6

Let A be a symmetric linear operator on a vector space V, let W be a subspace of V. Then the eigenvalues of A are the same as the union of the eigenvalues of A restricted to W and $W^{\perp}$.

We are now ready to prove the theorem.

Proof. Define $W := \{p(A)\boldsymbol{x} \mid p \in \mathbb{R}[t]\}$ and let $W^{\perp}$ denote the orthogonal vector space. Since $\boldsymbol{x} \in W$, $\boldsymbol{x}^T\boldsymbol{v} = 0$ for all $\boldsymbol{v} \in W^{\perp}$, this means that $A = B$ on $W^{\perp}$. Now, let us retrict A, B to W, let $k = \dim W$. Since W is cyclic by definition, A has k distinct eigenvalues $\lambda_1 > ... > \lambda_k$ with eigenvectors $\boldsymbol{\phi}_i$. Let z not be an eigenvalue of A, then

$$\begin{aligned}\chi(B)(z) &= \det\bigl(zI - A - \boldsymbol{x}\boldsymbol{x}^T\bigr) \\ &= \det(zI - A)\det\bigl(I - (zI - A)^{-1}\boldsymbol{x}\boldsymbol{x}^T\bigr)\end{aligned}$$

$zI - A$ is invertible since z is not an eigenvalue of A. We can now use Sylvester's theorem with $x = 1$ to write

$$\begin{aligned}\chi(B)(z) &= \chi(A)(z)(1 - \boldsymbol{x}^T(zI - A)^{-1}\boldsymbol{x}) \\ &= \chi(A)(z)\left(1 - \sum_{i=1}^{k}\frac{(\boldsymbol{x}^T\boldsymbol{\phi}_i^T)^2}{z - \lambda_i}\right)\end{aligned}$$

Define the function

$$G(z) = \sum_{i=1}^{k}\frac{(\boldsymbol{x}^T\boldsymbol{\phi}_i)^2}{z - \lambda_i}$$

Note the following properties of G

1. Solutions to $G(z) = 1$ are eigenvalues of B.
2. Since $\phi_i \in W$, we must have $\boldsymbol{x}^T\boldsymbol{\phi}_i \neq 0$.
3. There is exactly one solution (μ_{i+1}) to $G(z) = 1$ for $z \in (\lambda_{i+1}, \lambda_i)$, $1 \leq i \leq k - 1$.
4. There is exactly one solution (μ_1) to $G(z) = 1$ for $z > \lambda_1$.

It is now obvious that μ_i are exactly the eigenvalues of B and they interlace the eigenvalues of A.

$$\mu_i \geq \lambda_i \geq \mu_{i+1}$$

Since $A = B$ on $W^{\perp}$, the interlacing property is not affected by including the eigenvalues from this vector space. □

5.3 Probability

Theorem 5.3.1: Markov's Inequality

Let X be a non negative random variable with $a > 0$, then

$$P(X > aE(X)) \leq a^{-1}$$

Proof.

$$\begin{aligned}E(X) = P(X \leq aE(X))E(X|X \leq aE(X)) + P(X > aE(X))E(X|X > aE(X)) \\ \implies P(X > aE(X)) \leq \frac{E(X)}{E(X|X > aE(X))} \leq \frac{E(X)}{aE(X)} = a^{-1}\end{aligned}$$

□