Chapter 2

Linear Algebra

2.1 Eigenvalues and Eigenvectors

Definition 2.1 (Eigenvalues and Eigenvectors). *Let A be an $n \times n$ matrix. A nonzero vector v is called an eigenvector of A if $Av = \lambda v$ for some scalar λ. A scalar λ is called an eigenvalue of A if there exists an eigenvector v with $Av = \lambda v$.*

The multi-set of eigenvalues of A is given by the roots of the characteristic polynomial. This viewpoint will not be used often in the first part of the course, but it will be of central importance in the second part of the course.

Definition 2.2 (Characteristic Polynomial). *Let A be an $n \times n$ matrix. The characteristic polynomial of A is $p_A(x) := \det(xI - A)$.*

Two matrices are said to be similar if one is obtained from another by a change of basis.

Definition 2.3 (Similar Matrices). *A matrix X is similar to another matrix Y if there exists a non-singular matrix B so that $X = BYB^{-1}$.*

It is well known that similar matrices have the same spectrum.

Fact 2.4 (Spectrum of Similar Matrices). *If X is similar to Y, then the multi-set of eigenvalues of X and that of Y are the same.*

Proof. One way to see it is that they have the same characteristic polynomial, as

$$p_X(x) = \det(xI - X) = \det(xI - BYB^{-1}) = \det(B(xI - Y)B^{-1}) = \det(xI - Y) = p_Y(x),$$

where the second last equality is by Fact 2.27. □

Real Symmetric Matrices

In this course, we mostly work with real symmetric matrices, which have all eigenvalues being real numbers.

Theorem 2.5 (Spectral Theorem)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then all eigenvalues of A are real numbers. Furthermore, there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.*

See L01 of 2019 for a proof, which is based on the proof given in the book by Godsil and Royle [GR].

Remark 2.6 (Undirected and Directed Graphs)**.** *Theorem 2.5 applies to the adjacency/Laplacian matrices of undirected graphs, but not for directed graphs. This is the main reason that spectral graph theory is much more devleoped in undirected graphs. It has been an open direction to develop spectral graph theory for directed graphs.*

Diagonalization: Using the spectral theorem, real symmetric matrices can be written in the following form. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be an orthonormal basis of eigenvectors guaranteed by Theorem 2.5 with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Let V be the $n \times n$ matrix with the i-th column being v_i. Let D be the $n \times n$ diagonal matrix with the (i,i)-th entry being λ_i. Then the conditions $Av_i = \lambda_i v_i$ for $1 \leq i \leq n$ can be compactly written as $AV = VD$. Since the columns in V form an orthonormal basis, it follows that $V^T V = I$ and thus $V^{-1} = V^T$. So, we can rewrite $AV = VD$ as

$$A = VDV^{-1} = VDV^T.$$

Power of Matrices: Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. The diagonalization form $A = VDV^T$ is very convenient in computations. To compute A^k, observe that it is simply $A^k = (VDV^T)^k = VD^kV^T$ where D^k is readily computed as D is a diagonal matrix.

This is very useful in analyzing random walks, as P^t is the transition matrix of the random walks after t steps where P is the transition matrix in one step. We will use the eigenvalues of the transition matrix to bound the mixing time of random walks.

Eigen-Decomposition: If $v_1, \ldots, v_n$ form an orthonormal basis, then any $x \in \mathbb{R}^n$ can be written as a linear combination $c_1v_1 + \ldots + c_nv_n$. By orthonormality, for any $1 \leq i \leq n$,

$$\langle x, v_i\rangle = \langle c_1v_1 + \ldots + c_nv_n, v_i\rangle = \langle c_iv_i, v_i\rangle = c_i.$$

Therefore, for any $x \in \mathbb{R}^n$,

$$x = \langle x, v_1\rangle v_1 + \ldots + \langle x, v_n\rangle v_n = v_1v_1^Tx + \ldots v_nv_n^Tx = \big(v_1v_1^T + \ldots + v_nv_n^T\big)x.$$

Since this is true for all $x \in \mathbb{R}^n$, it follows that

$$v_1v_1^T + \ldots + v_nv_n^T = I_n.$$

Now, if $v_1, \ldots, v_n$ are also eigenvectors of a matrix $A \in \mathbb{R}^{n\times n}$, then for any $x \in \mathbb{R}^n$,

$$Ax = A(v_1v_1^T + \ldots + v_nv_n^T)x = \big(\lambda_1v_1v_1^T + \ldots + \lambda_nv_nv_n^T\big)x.$$

This implies that

$$A = \lambda_1v_1v_1^T + \ldots + \lambda_nv_nv_n^T.$$

Verify that we can also write the inverse using the eigen-decomposition as

$$A^{-1} = \frac{1}{\lambda_1}v_1v_1^T + \ldots + \frac{1}{\lambda_n}v_nv_n^T.$$

Later, this form will also be used to define the "psuedo-inverse" of a matrix A when A is not of full rank.

Positive Semidefinite Matrices

An important class of real symmetric matrices is the class of positive semidefinite matrices. A real symmetric matrix is called positive semidefinite if all of its eigenvalues are nonnegative. This can be seen as a matrix analog of a non-negative number. The following are some equivalent characterizations of a positive semidefinite matrix.

Fact 2.7 (Positive Semidefinite Matrix)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. The following statements are equivalent.*

1. *A is positive semidefinite, i.e. all eigenvalues of A are non-negative.*
2. *For any $x \in \mathbb{R}^n$, it holds that $x^T Ax \geq 0$, i.e. all quadratic forms are non-negative.*
3. *$A = U^T U$ for some matrix $U \in \mathbb{R}^{n\times n}$.*

The notation $A \succcurlyeq 0$ will be used to denote that A is a positive semidefinite matrix.

It is a good exercise to prove this fact; see L01 from 2019 for a proof. A matrix is called positive definite if all eigenvalues of A are positive. It is left as an exercise to find the equivalent characterizations for positive definite matrices as in Fact 2.7.

Check that the set of positive semidefinite matrices forms a convex set. Optimizing a linear function over the set of positive semidefinite matrices with linear constraints is called semidefinite programming. This is a very important class of convex optimization problems that can be solved in polynomial time. We will see it once in this course and we will explain more when we use it.

It is also a good exercise to prove the following useful fact.

Fact 2.8. *For any two positive semidefinite matrices $A, B \in \mathbb{R}^{n\times n}$,*

$$\langle A, B\rangle := \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}B_{ij} \geq 0.$$

Optimization Formulation for Eigenvalues

The main reason why eigenvalues are useful for optimization is through the following formulation, which is basically the quadratic form but normalized by the vector length.

Definition 2.9 (Rayleigh Quotient)**.** *The Rayleigh quotient of a vector $x \in \mathbb{R}^n$ with respect to a matrix $A \in \mathbb{R}^{n\times n}$ is defined to be*

$$R_A(x) := \frac{x^T Ax}{x^T x} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j}{\sum_{i=1}^{n} x_i^2}.$$

The largest eigenvalue is the maximum value of the Rayleigh quotient.

Lemma 2.10 (Optimization Formulation for α_1)**.** *Suppose $A \in \mathbb{R}^{n\times n}$ is a real symmetric matrix with eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$. Then*

$$\alpha_1 = \max_{x\in\mathbb{R}^n} \frac{x^T Ax}{x^T x}.$$

Proof. Let $v_1, v_2, \ldots, v_m$ be the corresponding orthonormal basis of eigenvectors guaranteed by Theorem 2.5. As $v_1, \ldots, v_n$ forms a basis of $\mathbb{R}^n$, any vector $x \in \mathbb{R}^n$ can be written as a linear combination $x = c_1v_1 + \cdots + c_nv_n$. Then, the numerator can be written as

$$x^TAx = (c_1v_1+\cdots+c_nv_n)^TA(c_1v_1+\cdots+c_nv_n) = (c_1v_1+\cdots+c_nv_n)^T(c_1\alpha_1v_1+\cdots+c_n\alpha_nv_n) = \sum_{i=1}^{n} c_i^2\alpha_i,$$

where the second equality is because $v_1, \cdots, v_n$ are eigenvectors and the last equality is because $v_1, \cdots, v_n$ are orthonormal. Similarly, the denominator can be written as

$$x^Tx = (c_1v_1 + \cdots + c_nv_n)^T(c_1v_1 + \cdots + c_nv_n) = \sum_{i=1}^{n} c_i^2.$$

So, the Rayleigh quotient of x is

$$\frac{x^TAx}{x^Tx} = \frac{\sum_{i=1}^n c_i^2\alpha_i}{\sum_{i=1}^n c_i^2} \leq \frac{\alpha_1\sum_{i=1}^n c_i^2}{\sum_{i=1}^n c_i^2} = \alpha_1.$$

On the other hand, note that v_1 attains the maximum, and the lemma follows. □

This can be extended to characterize other eigenvalues. In particular, we will use the following lemma for the second largest eigenvalue later.

Lemma 2.11 (Optimization Formulation for α_k)**.** *Suppose $A \in \mathbb{R}^{n\times n}$ is a real symmetric matrix with eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ and corresponding orthonormal eigenvectors $v_1, \ldots, v_n$. Let T_k be the set of vectors that are orthogonal to $v_1, v_2, \ldots, v_{k-1}$. Then*

$$\alpha_k = \max_{x\in T_k} \frac{x^TAx}{x^Tx}.$$

Proof. Let $x \in T_k$. Write $x = c_1v_1+\cdots+c_nv_n$. Recall that $c_i = \langle x, v_i\rangle$ from the eigen-decomposition subsubsection. Since $x \in T_k$, it follows that $c_1 = c_2 = \cdots = c_{k-1} = 0$. Using the same calculation as in Lemma 2.10,

$$\frac{x^TAx}{x^Tx} = \frac{\sum_{i=k}^n c_i^2\alpha_i}{\sum_{i=k}^n c_i^2} \leq \frac{\alpha_k\sum_{i=k}^n c_i^2}{\sum_{i=k}^n c_i^2} = \alpha_k.$$

On the other hand, $v_k \in T_k$ and $v_k^TAv_k/v_k^Tv_k = \alpha_k$, and the lemma follows. □

The above result gives a characterization of α_k, but it requires the knowledge of the previous eigenvectors. The Courant-Fischer theorem gives a characterization without knowing the eigenvectors, and is more useful in giving bounds on eigenvalues. In words, the Courant-Fischer theorem says that to prove a lower bound on α_k, one needs to show a k-dimensional subspace in which every vector has large Rayleigh quotient, and the best k-dimensional subspace gives the tight lower bound. And to prove an upper bound on α_k, one needs to show a $(n-k+1)$-dimensional subspace in which every vector has small Rayleigh quotient, and the best $(n-k+1)$-dimensional subspace gives the tight upper bound.

Theorem 2.12 (Courant-Fischer Theorem)**.** *Suppose $A \in \mathbb{R}^{n\times n}$ is a real symmetric matrix with eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$. Then*

$$\alpha_k = \max_{S\subseteq\mathbb{R}^n:\dim(S)=k}\ \min_{x\in S} \frac{x^TAx}{x^Tx} = \min_{S\subseteq\mathbb{R}^n:\dim(S)=n-k+1}\ \max_{x\in S} \frac{x^TAx}{x^Tx}.$$

Proof. We prove the max-min equality. The min-max equality is similar and is left as an exercise.

Let S_k be the k-dimensional subspace spanned by the first k orthonormal eigenvectors $v_1, \ldots, v_k$, i.e. $S_k = \{x \mid x = c_1 v_1 + \cdots + c_k v_k \text{ for some } c_1, \ldots, c_k \in \mathbb{R}\}$. Then, for any $x \in S_k$,

$$\frac{x^T A x}{x^T x} = \frac{(c_1 v_1 + \cdots + c_k v_k)^T A (c_1 v_1 + \cdots + c_k v_k)}{(c_1 v_1 + \cdots + c_k v_k)^T (c_1 v_1 + \cdots + c_k v_k)} = \frac{\sum_{i=1}^k c_i^2 \alpha_i}{\sum_{i=1}^k c_i^2} \geq \frac{\alpha_k \sum_{i=1}^k c_i^2}{\sum_{i=1}^k c_i^2} = \alpha_k.$$

Therefore,

$$\max_{S \subseteq \mathbb{R}^n : \dim(S) = k} \min_{x \in S} \frac{x^T A x}{x^T x} \geq \min_{x \in S_k} \frac{x^T A x}{x^T x} \geq \alpha_k.$$

To prove that the maximum cannot be greater than α_k, observe that any k-dimensional subspace must intersect the $(n-k+1)$-dimensional subspace T_k spanned by $\{v_k, v_{k+1}, \ldots, v_n\}$. For any $x \in T_k$,

$$\frac{x^T A x}{x^T x} = \frac{\sum_{i=k}^n c_i^2 \alpha_i}{\sum_{i=k}^n c_i^2} \leq \alpha_k.$$

Therefore,

$$\max_{S \subseteq \mathbb{R}^n : \dim(S) = k} \min_{x \in S} \frac{x^T A x}{x^T x} \leq \min_{x \in S \cap T_k} \frac{x^T A x}{x^T x} \leq \alpha_k.$$

□

One consequence of the Courant-Fischer theorem is the eigenvalue interlacing theorem, which will be useful in the second part of the course.

Theorem 2.13 (Cauchy's Interlacing Theorem)**.** *Let $A \in \mathbb{R}^{n \times n}$ be a real symmetric matrix and B be a $(n-1) \times (n-1)$ principle submatrix of A. Then*

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \beta_2 \geq \ldots \geq \alpha_{n-1} \geq \beta_{n-1} \geq \alpha_n,$$

where $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ are the eigenvalues of A and $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_{n-1}$ are the eigenvalues of B.

Proof. Assume without loss of generality that B is in the top left corner of A, that is, the first $n-1$ coordinates.

It should be clear that $\alpha_k \geq \beta_k$ because the search space for α_k is larger than than for β_k. More precisely,

$$\alpha_k = \max_{S \subseteq \mathbb{R}^n : \dim(S) = k} \min_{x \in S} \frac{x^T A x}{x^T x} \geq \max_{S \subseteq \mathbb{R}^{n-1} : \dim(S) = k} \min_{x \in S} \frac{x^T A x}{x^T x} = \max_{S \subseteq \mathbb{R}^{n-1} : \dim(S) = k} \min_{x \in S} \frac{x^T B x}{x^T x} = \beta_k.$$

Next, we show $\beta_k \geq \alpha_{k+1}$. For any $S \subseteq \mathbb{R}^n$ with $\dim(S) = k+1$, its restriction to the first $n-1$ coordinates (i.e. $S \cap \mathbb{R}^{n-1}$) is of dimension at least k. So, if there is a good $(k+1)$-dimensional subspace for A, then there is a good k-dimensional subspace for B, and so β_k can do as well as α_{k+1}. More formally, let S_{k+1} be the $(k+1)$-dimensional subspace that attains maximum for α_{k+1},

$$\alpha_{k+1} = \min_{x \in S_{k+1}} \frac{x^T A x}{x^T x} \leq \min_{x \in S_{k+1} \cap \mathbb{R}^{n-1}} \frac{x^T A x}{x^T x} \leq \max_{S \subseteq \mathbb{R}^{n-1} : \dim(S) = k} \min_{x \in S} \frac{x^T A x}{x^T x} = \max_{S \subseteq \mathbb{R}^{n-1} : \dim(S) = k} \min_{x \in S} \frac{x^T B x}{x^T x} = \beta_k.$$

□

Perron-Frobenius Theorem

The Perron-Frobenius theorem is the most important result on the eigenvalues and eigenvectors on non-negative matrices. To state it, we need the definitions of an irreducible matrix and the spectral radius of a matrix.

Definition 2.14 (Irreducible Matrix)**.** *A matrix $A \in \mathbb{R}^{n\times n}$ is irreducible if its underlying directed graph $G = (V, E)$ is strongly connected, where the vertex set of G is $V = [n]$ and the edge set of G is $E = \{ij \mid A_{ij} \neq 0\}$.*

The spectral radius of a real symmetric matrix is simply the eigenvalue with largest absolute value. The following is the more general definition for matrices with complex eigenvalues.

Definition 2.15 (Spectral Radius)**.** *The spectral radius $\rho(A)$ of a matrix A is the maximum of the moduli of its eigenvalues.*

The Perron-Frobenius theorem is about the largest eigenvalue and its corresponding eigenvectors, which will be useful in the study of random walks. See chapter 8.8 in [GR] and chapter 8.4 in [HJ13] for more details and proofs.

Theorem 2.16 (Perron-Frobenius Theorem)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a non-negative irreducible matrix.*

1. *The spectral radius $\rho(A)$ is an eigenvalue of A with multiplicity one. In particular, for a real symmetric matrix, the largest eigenvalue is of multiplicity one and its absolute value is the largest.*
2. *If v is an eigenvector with eigenvalue $\rho(A)$, then all the entries of v are nonzero and they have the same sign.*

Matrix Norms

Definition 2.17 (Operator Norm)**.** *Let A be an $m \times n$ matrix. The operator norm $\|A\|_{\text{op}}$ of A is defined as*

$$\|A\|_{\text{op}} := \sup_{x\in\mathbb{R}^n, x\neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

This is also denoted by $\|A\|_2$ to denote that it is the ratio of the 2-norm of the vectors after and before the linear transformation. But sometimes it is confused with the Schatten 2-norm of A which is also denoted by $\|A\|_2$, so we use the notation $\|A\|_{\text{op}}$ that is also a common notation.

Exercise 2.18 (Operator Norm of a Real Symmetric Matrix)**.** *Show that $\|A\|_{\text{op}}$ is equal to the largest eigenvalue of A when A is a real symmetric matrix.*

The following are some simple properties that will be useful.

Fact 2.19 (Properties of Operator Norm)**.** *Let $A \in \mathbb{R}^{m\times n}$.*

1. *$\|A\|_{\text{op}} \geq 0$ and $\|A\|_{\text{op}} = 0$ if and only if $A = 0$.*
2. *$\|cA\|_{\text{op}} = |c|\|A\|_{\text{op}}$ for every scalar c.*

3. $\|A+B\|_{\text{op}} \leq \|A\|_{\text{op}} + \|B\|_{\text{op}}$.
4. $\|Ax\|_2 \leq \|A\|_{\text{op}} \|x\|_2$ *for every* $x \in \mathbb{R}^n$.
5. $\|BA\|_{\text{op}} \leq \|B\|_{\text{op}} \|A\|_{\text{op}}$

2.2 Formulas and Inequalities

We record some useful formulas and inequalities in this section. A general reference is the book by Horn and Johnson [HJ13].

Inverse

The following formulas are for updating the inverse of a matrix. See wiki for proofs.

Fact 2.20 (Sherman-Morrison Formula)**.** *Suppose* $A \in \mathbb{R}^{n\times n}$ *is an invertible square matrix and* $u, v \in \mathbb{R}^n$ *are column vectors. Then* $A + uv^T$ *is invertible if and only if* $1 + v^T A^{-1} u \neq 0$*. In this case,*

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1 + v^TA^{-1}u}$$

Fact 2.21 (Woodbury Formula)**.** *Given a square invertible* $n \times n$ *matrix* A*, a* $n \times k$ *matrix* U*, and a* $k \times n$ *matrix* V*, assuming* $(I_k + VA^{-1}U)$ *is invertible, then*

$$(A + UV)^{-1} = A^{-1} - A^{-1}U\big(I_k + VA^{-1}U\big)^{-1}VA^{-1}.$$

The following formula is for inverting a block matrix, and the Schur complement is a useful definition.

Fact 2.22 (Block Matrix Inversion)**.** *Let* A *and* D *be square matrices and*

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

If A *and the Schur complement* $D - CA^{-1}B$ *are invertible, then*

$$M^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B\big(D - CA^{-1}B\big)^{-1}CA^{-1} & -A^{-1}B\big(D - CA^{-1}B\big)^{-1} \\ -\big(D - CA^{-1}B\big)^{-1}CA^{-1} & \big(D - CA^{-1}B\big)^{-1} \end{pmatrix}.$$

When a matrix A is not invertible, we can work with the Moore-Penrose pseudoinverse of A.

Definition 2.23 (Pseudoinverse)**.** *Let* $A \in \mathbb{R}^{n\times n}$ *be a real symmetric matrix with eigen-decomposition* $A = \sum_{i=1}^n \lambda_i v_i v_i^T$*. The pseudoinverse of* A*, denoted by* $A^\dagger$*, is defined as*

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

Check the following properties of pseudoinverse.

Fact 2.24 (Properties)**.** *Let* A *be a real symmetric matrix and* $A^\dagger$ *be its pseudoinverse. Then*

$$AA^\dagger A = A \quad \textit{and} \quad A^\dagger A A^\dagger = A^\dagger \quad \textit{and} \quad \big(A^\dagger\big)^\dagger = A.$$

Determinant

These formulas about determinants will be used in the second part of the course.

Fact 2.25 (Laplace Co-Factor Expansion)**.** *Let A be a $n \times n$ matrix. For every $1 \leq i \leq n$,*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{i,j} \det\left(A_{[n]\setminus i,[n]\setminus j}\right),$$

where $A_{S,T}$ is the submatrix with rows in $S \subseteq [n]$ and columns in $T \subseteq [n]$.

Applying Laplace expansion recursively gives the Leibniz formula.

Fact 2.26 (Leibniz Formula)**.** *Let A be a $n \times n$ matrix. Then*

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{\sigma(i),i},$$

where S_n is the set of permutations of the set $[n] = \{1, \ldots, n\}$ and $\operatorname{sgn}(\sigma)$ is the sign function of permutation σ which returns $+1$ and -1 for even and odd permutations.

The following is a simple fact.

Fact 2.27 (Product)**.**

$$\det(AB) = \det(A)\det(B).$$

The following result, sometimes known as Sylvester's determinant identity, can be used to deduce that the nonzero eigenvalues of AB and BA are the same (with multiplicity).

Fact 2.28 (Weinstein-Aronszajn Identity)**.**

$$\det(I + AB) = \det(I + BA).$$

The matrix determinant formula keeps track of how the determinant changes after a rank-one update.

Fact 2.29 (Matrix Determinant Formula)**.**

$$\det\left(M - uv^T\right) = \det(M)\left(1 - v^T M^{-1} u\right).$$

The Cauchy-Binet formula will be very useful. One application is to compute the number of spanning trees of a graph.

Fact 2.30 (Cauchy-Binet Formula)**.** *Let A be an $m \times n$ matrix and B be an $n \times m$ matrix. Then*

$$\det(AB) = \sum_{S \in \binom{[n]}{m}} \det\left(A_{[m],S}\right) \det\left(B_{S,[m]}\right)$$

The following formula gives the coefficients of the characteristic polynomials.

Fact 2.31 (Characteristic Polynomial)**.** *Let A be an $n \times n$ matrix.*

$$\det(\lambda I_n - A) = \sum_{k=0}^{n} \lambda^{n-k}(-1)^k \sum_{S \in \binom{[n]}{k}} \det(A_{S,S}).$$

Using Fact 2.31 and Cauchy-Binet formula in Fact 2.30, we have the following identity for the characteristic polynomial of a sum of outer products.

Fact 2.32 (Characteristic Polynomial of Sum of Outer Products)**.** *Let $u_1, \ldots, u_m \in \mathbb{R}^n$.*

$$\det\left(xI - \sum_{i=1}^{m} u_i u_i^T\right) = \sum_{k=0}^{n} \lambda^{n-k}(-1)^k \sum_{S \in \binom{[m]}{k}} \det_k\left(\sum_{i \in S} u_i u_i^T\right),$$

where $\det_k(A) = \sum_{S \in \binom{[n]}{k}} \det(A_{S,S})$.

Trace

Definition 2.33 (Trace)**.** *The trace of a matrix $A \in \mathbb{R}^{n \times n}$, denoted by $\mathrm{Tr}(A)$, is defined as the sum of the diagonal entries of A.*

Fact 2.34 (Cyclic Property of Trace)**.** *For two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times m}$,*

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA).$$

By looking at the coefficient of x^{n-1} of the characteristic polynomial $\det(xI - A)$ of $A \in \mathbb{R}^{n \times n}$ in two ways, one can obtain the following useful fact (see L01 from 2019 for a proof).

Fact 2.35 (Trace is Sum of Eigenvalues)**.** *Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $A \in \mathbb{R}^{n \times n}$. Then*

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i.$$

The following are two advanced results about traces; see [Bha97]. We may not need them explicitly in this course.

Fact 2.36 (Golden-Thompson Inequality)**.**

$$\mathrm{Tr}(e^{A+B}) \leq \mathrm{Tr}(e^A) \cdot \mathrm{Tr}(e^B)$$

Fact 2.37 (Lieb-Thirring Inequality)**.** *Let A and B be positive definite matrices and $q \geq 1$. Then*

$$\mathrm{Tr}\left((BAB)^q\right) \leq \mathrm{Tr}(B^q A^q B^q).$$

Matrix Calculus

The formula for differenting the inverse is obtained by differentiating the identity $AA^{-1} = I$.

Fact 2.38 (Inverse)**.**

$$d(A^{-1}) = -A^{-1}(dA)A^{-1}.$$

Jacobi's formula is obtained by differentiating the cofactor expansion in Fact 2.25.

Fact 2.39 (Jacobi's Formula)**.**

$$\frac{d}{dt}\det A(t) = \mathrm{Tr}\left(\mathrm{adj}(A(t)) \cdot \frac{dA(t)}{dt}\right) = \left(\det A(t)\right) \cdot \mathrm{Tr}\left(A(t)^{-1} \cdot \frac{dA(t)}{dt}\right).$$

2.3 References

[Bha97] Rajendra Bhatia. *Matrix Analysis*, volume 169. Springer, 1997. 15

[GR] Christopher D. Godsil and Gordon F. Royle. *Algebraic Graph Theory*. 8, 12

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge; New York, 2nd edition, 2013. 12, 13