A Linear Algebra Primer

Standard texts on Linear Algebra and Algebra are [1, 8].

1 Preliminaries

1.1 Vectors and matrices

We shall use $\mathbb{R}$ to denote the set of real numbers and $\mathbb{C}$ to denote the set of complex numbers. For any $c = a + bi \in \mathbb{C}$, the *complex conjugate* of c, denoted by $\bar{c}$ is defined to be $\bar{c} = a - bi$. The *modulus* of c, denoted by $|c|$, is $\sqrt{a^2 + b^2}$. It is easy to see that $|c|^2 = c\bar{c}$.

If we mention the word "vector" alone, it is understood to be a column vector. An n-dimensional vector x has n entries in some field of numbers, such as $\mathbb{R}$ or $\mathbb{C}$:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

The set of all n-dimensional vectors over $\mathbb{R}$ (respectively $\mathbb{C}$) is denoted by $\mathbb{R}^n$ (respectively $\mathbb{C}^n$). They are also called *real vectors* and *complex vectors*, respectively.

Similar to vectors, matrices need an underlying field. We thus have complex matrices and real matrices just as in the case of vectors. In fact, an n-dimensional vector is nothing but an $n \times 1$ matrix. In the discussion that follows, the concepts of complex conjugates, transposes, and conjugate transposes also apply to vectors in this sense.

Given an $m \times n$ matrix $A = (a_{ij})$, the *complex conjugate* $\bar{A}$ of A is a matrix obtained from A by replacing each entry a_{ij} of A by the corresponding complex conjugate $\bar{a}_{ij}$. The *transpose* A^T of A is the matrix obtained from A by turning its rows into columns and vice versa. For example,

$$A = \begin{bmatrix} 0 & 3 & 1 \\ -2 & 0 & 1 \end{bmatrix}, \text{ and } A^T = \begin{bmatrix} 0 & -2 \\ 3 & 0 \\ 1 & 1 \end{bmatrix}.$$

The *conjugate transpose* A^* of A is defined to be $(\bar{A})^T$. A square matrix A is *symmetric* iff $A = A^T$, and is *Hermitian* iff $A = A^*$.

Given a real vector $x \in \mathbb{R}^n$, the *length* $\|x\|$ of x is

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \tag{1}$$

Notice that $\|x\|^2 = xx^T$. When x is a complex vector, we use x^* instead of x^T. Hence, in general we define $\|x\| = \sqrt{xx^*} = \sqrt{x^*x}$. (You should check that $xx^* = x^*x$, and that it is a real number so that the square root makes sense.)

The length $\|x\|$ is also referred to as the L_2-*norm* of vector x, denoted by $\|x\|_2$. In general, the L_p-*norm* of an n-dimensional vector x, denoted by $\|x\|_p$, where $p = 1, 2, \dots,$ is defined to be

$$\|x\|_p := (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}}, \tag{2}$$

and

$$\|x\|_\infty := \max_{i=1..n} |x_i|. \tag{3}$$

The following identities are easy to show, yet of great importance. Given a $p \times q$ matrix A and a $q \times r$ matrix B, we have

$$(AB)^T = B^T A^T \tag{4}$$

$$(AB)^* = B^* A^* \tag{5}$$

(Question: what are the dimensions of the matrices $(AB)^T$ and $(AB)^*$?)

A square matrix A is said to be *singular* if there is no unique solution to the equation $Ax = b$. For A to be singular, it does not matter what b is. The uniqueness of a solution to $Ax = b$ is an intrinsic property of A alone. If there is one and only one x such that $Ax = b$, then A is said to be *non-singular*.

1.2 Determinant and trace

Given a square matrix $A = (a_{ij})$ of order n, the equation $Ax = 0$ has a unique solution if and only if $\det A \neq 0$, where $\det A$ denotes the *determinant* of A, which is defined by

$$\det A = \sum_{\pi \in S_n} (-1)^{I(\pi)} \prod_{i=1}^{n} a_{i\pi(i)} = \sum_{\pi \in S_n} \text{sign}(\pi) \prod_{i=1}^{n} a_{i\pi(i)}. \tag{6}$$

Here, S_n denotes the set of all permutations on the set $[n] = \{1, \dots, n\}$. (S_n is more often referred to as the *symmetric group* of order n.) Given a permutation $\pi \in S_n$, we use $I(\pi)$ to denote the number of *inversions* of π, which is the number of pairs $(\pi(i), \pi(j))$ for which $i < j$ and $\pi(i) > \pi(j)$. The *sign* of a permutation π, denoted by sign(π), is defined to be sign$(\pi) = (-1)^{I(\pi)}$.

Exercise 1.1. Find an involution for S_n to show that, for $n \geq 2$, there are as many permutations with negative sign as permutations with positive sign.

Let us take an example for $n = 3$. In this case S_n consists of 6 permutations:

$$S_n = \{123, 132, 213, 231, 312, 321\}.$$

Notationally, we write $\pi = 132$ to mean a permutation where $\pi(1) = 1$, $\pi(2) = 3$, and $\pi(3) = 2$. Thus, when $\pi = 132$ we have sign$(\pi) = -1$ since there is only one "out-of-order" pair $(3, 2)$. To be more precise, sign$(123) = 1$, sign$(132) = -1$, sign$(312) = 1$, sign$(213) = -1$, sign$(231) = 1$, sign$(321) = -1$.

Consequently, for

$$A = \begin{bmatrix} 0 & 3 & 1 \\ -2 & 0 & 1 \\ -1 & 2 & 2 \end{bmatrix}$$

we have

$$\begin{aligned} \det A &= a_{11}a_{22}a_{33} + (-1)a_{11}a_{23}a_{32} + (-1)a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + \\ &\quad a_{13}a_{21}a_{32} + (-1)a_{13}a_{22}a_{31} \\ &= 0 \cdot 0 \cdot 2 + (-1) \cdot 0 \cdot 1 \cdot 2 + (-1) \cdot 3 \cdot (-2) \cdot 2 + 3 \cdot 1 \cdot (-1) + \\ &\quad 1 \cdot (-2) \cdot 2 + (-1) \cdot 1 \cdot 0 \cdot (-1) \\ &= 5 \end{aligned}$$

The *trace* of a square matrix A, denoted by tr A is the sum of its diagonal entries. The matrix A above has

$$\text{tr } A = 0 + 0 + 2 = 2.$$

1.3 Combinations of vectors and vector spaces

A vector w is a *linear combination* of m vectors $v_1, \dots, v_m$ if w can be written as

$$w = a_1 v_1 + a_2 v_2 + \dots a_m v_m. \tag{7}$$

The number a_j is called the *coefficient* of the vector v_j in this linear combination. Note that, as usual, we have to fix the underlying field such as $\mathbb{R}$ or $\mathbb{C}$. If, additionally, we also have $a_1 + a_2 + \cdots + a_m = 1$, then w is called an *affine combination* of the v_i.

A *canonical combination* is a linear combination in which $a_j \geq 0, \forall j$; and a *convex combination* is an affine combination which is also canonical. The *linear (affine, canonical, convex) hull* of $\{v_1, \dots, v_m\}$ is the set of all linear (affine, canonical, convex) combinations of the v_j. Note that in the above definitions, m could be infinite. The convex hull of a finite set of vectors is called a *cone*, or more specifically a *convex polyhedral cone*.

A *real vector space* is a set V of real vectors so that a linear combination of any subset of vectors in V is also in V. In other words, vector spaces have to be *closed* under taking linear combinations. Technically speaking, this is an incomplete definition, but it is sufficient for our purposes. One can also replace the word "real" by "complex". A *subspace* of a vector space V is a subset of V which is closed under taking linear combinations.

Given a set $V = \{v_1, \dots, v_m\}$ of vectors, the set of all linear combinations of the v_j forms a vector space, denoted by span $\{(V)\}$, or span $\{(v_1, \dots, v_m)\}$. The *column space* of a matrix A is the span of its column vectors. The *row space* of A is the span of A's rows. Note that equation $Ax = b$ (with A not necessarily a square matrix) has a solution if and only if b lies in the column space of A. The coordinates of x form the coefficients of the column vectors of A in a linear combination to form b.

A set $V = \{v_1, \dots, v_m\}$ of (real, complex) vectors is said to be *linearly independent* if

$$a_1 v_1 + a_2 v_2 + \dots a_m v_m = 0 \text{ only happens when } a_1 = a_2 = \dots a_m = 0.$$

Otherwise, the vectors in V are said to be (linearly) *dependent*.

The *dimension* of a vector space is the maximum number of linearly independent vectors in the space. The *basis* of a vector space V is a subset $\{v_1, \dots, v_m\}$ of V which is linearly independent and span $\{(v_1, \dots, v_m)\} = V$. It is easy to show that m is actually the dimension of V. A vector space typically has infinitely many bases. All bases of a vector space V have the same size, which is also the dimension of V. The sets $\mathbb{R}^n$ and $\mathbb{C}^n$ are vector spaces by themselves.

In an n-dimensional vector space, a set of $m > n$ vectors must be linearly dependent.

The dimensions of a matrix A's column space and row space are equal, and is referred to as the *rank* of A. This fact is not very easy to show, but not too difficult either. Gaussian elimination is of great use here.

Exercise 1.2. Show that for any basis B of a vector space V and some vector $v \in V$, there is exactly one way to write v as a linear combination of vectors in B.

1.4 Inverses

We use diag$(a_1, \dots, a_n)$ to denote the matrix $A = (a_{ij})$ where $a_{ij} = 0$ for $i \neq j$ and $a_{ii} = a_i, \forall i$. The *identity matrix*, often denoted by I, is defined to be diag$(1, \dots, 1)$.

Given a square matrix A, the *inverse* of A, denoted by A^{-1} is a matrix B such that

$$AB = BA = I, \text{ or } AA^{-1} = A^{-1}A = I.$$

Exercise 1.3. Show that, if A and B both have inverses, then the inverse of AB can be calculated easily by

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{8}$$

Similarly, the same rule holds for 3 or more matrices. For example,

$$(ABCD)^{-1} = D^{-1}C^{-1}B^{-1}A^{-1}.$$

If A has an inverse, it is said to be *invertible*. Not all matrices are invertible. There are many conditions to test if a matrix has an inverse, including: non-singularity, non-zero determinant, non-zero eigenvalues (to be defined), linearly independent column vectors, linearly independent row vectors.

2 Eigenvalues and eigenvectors

In this section, we shall be concerned with square matrices only, unless stated otherwise.

The *eigenvalues* of a matrix A are the numbers λ such that the equation $Ax = \lambda x$, or $(\lambda I - A)x = 0$, has a non-zero solution vector, in which case the solution vector x is called a λ-*eigenvector*.

The *characteristic polynomial* $p_A(\lambda)$ of a matrix A is defined to be

$$p_A(\lambda) := \det(\lambda I - A).$$

Since the all-0 vector, denoted by $\vec{0}$, is always a solution to $(\lambda I - A)x = 0$, it would be the only solution if $\det(\lambda I - A) \neq 0$. Hence, the eigenvalues are solutions to the equation $p_A(\lambda) = 0$. For example, if

$$A = \begin{bmatrix} 2 & 1 \\ -2 & 3 \end{bmatrix},$$

then,

$$p_A(\lambda) = \det \begin{bmatrix} \lambda - 2 & -1 \\ +2 & \lambda - 3 \end{bmatrix} = (\lambda - 2)(\lambda - 3) + 2 = \lambda^2 - 5\lambda + 8.$$

Hence, the eigenvalues of A are $(5/2 \pm i\sqrt{7}/2)$.

If we work on the complex numbers, then equation $p_A(\lambda) = 0$ always has n roots (up to multiplicities). However, we shall be concerned greatly with matrices which have real eigenvalues. We shall establish sufficient conditions for a matrix to have real eigenvalues, as shall be seen in later sections.

Theorem 2.1. *Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of an $n \times n$ complex matrix A, then*

(i) $\lambda_1 + \cdots + \lambda_n =$ *tr* A.

(ii) $\lambda_1 \dots \lambda_n = \det A$.

Proof. In the complex domain, $p_A(\lambda)$ has n complex roots since it is a polynomial of degree n. The eigenvalues $\lambda_1, \dots, \lambda_n$ are the roots of $p_A(\lambda)$. Hence, we can write

$$p_A(\lambda) = \prod_i (\lambda - \lambda_i) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0.$$

It is evident that

$$\begin{aligned} c_{n-1} &= -(\lambda_1 + \cdots + \lambda_n) \\ c_0 &= (-1)^n \lambda_1 \dots \lambda_n. \end{aligned}$$

On the other hand, by definition we have

$$p_A(\lambda) = \det \begin{bmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \dots & \dots & \vdots \\ -a_{n1} & -a_{n2} & \dots & \lambda - a_{nn} \end{bmatrix}.$$

Expanding $p_A(\lambda)$ in this way, the coefficient of λ^{n-1} (which is c_{n-1}) is precisely $-(a_{11}+a_{22}+\cdots+a_{nn})$; and the coefficient of λ^0 (which is c_0) is $(-1)^n \det A$ (think carefully about this statement!). □

2.1 The diagonal form

Proposition 2.2. *Suppose the $n \times n$ matrix A has n linearly independent eigenvectors $\mathbf{x_1}, \dots, \mathbf{x_n}$, where $\mathbf{x_i}$ is a λ_i-eigenvector. Let S be the matrix whose columns are the vectors $\mathbf{x_i}$, then $S^{-1}AS = \Lambda$, where $\Lambda =$ diag $(\lambda_1, \dots, \lambda_n)$.*

Proof. Note that since the column vectors of S are independent, S is invertible and writing S^{-1} makes sense. We want to show $S^{-1}AS = \Lambda$, which is the same as showing $AS = S\Lambda$. Since $A\mathbf{x}_i = \mathbf{x}_i\lambda_i$, it follows that

$$AS = A \begin{bmatrix} | & \dots & | \\ \mathbf{x}_1 & \dots & \mathbf{x}_n \\ | & \dots & | \end{bmatrix} = \begin{bmatrix} | & \dots & | \\ A\mathbf{x}_1 & \dots & A\mathbf{x}_n \\ | & \dots & | \end{bmatrix} = \begin{bmatrix} | & \dots & | \\ \lambda_1\mathbf{x}_1 & \dots & \lambda_n\mathbf{x}_n \\ | & \dots & | \end{bmatrix} = S\Lambda.$$

□

In general, if a matrix S satisfies the property that $S^{-1}AS$ is a diagonal matrix, then S is said to *diagonalize* A, and A is said to be *diagonalizable*. It is easy to see from the above proof that if A is diagonalizable by S, then the columns of S are eigenvectors of A; moreover, since S is invertible by definition, the columns of S must be linearly independent. In other words, we just proved

Theorem 2.3. *A matrix is diagonalizable if and only if it has n independent eigenvectors.*

Proposition 2.4. *If $x_1, \dots x_k$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, \dots \lambda_k$, then $x_1, \dots x_k$ are linearly independent.*

Proof. When $k = 2$, suppose $c_1x_1 + c_2x_2 = 0$. Multiplying by A gives $c_1\lambda_1x_1 + c_2\lambda_2x_2 = 0$. Subtracting λ_2 times the previous equation we get

$$c_1(\lambda_1 - \lambda_2)x_1 = 0.$$

Hence, $c_1 = 0$ since $\lambda_1 \neq \lambda_2$ and $x_1 \neq 0$. The general case follows trivially by induction. □

Exercise 2.5. If $\lambda_1, \dots \lambda_n$ are eigenvalues of A, then $\lambda_1^k, \dots \lambda_n^k$ are eigenvalues of A^k. If S diagonalizes A, i.e. $S^{-1}AS = \Lambda$, then $S^{-1}A^kS = \Lambda^k$

2.2 Symmetric and Hermitian matrices

For any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, the *inner product* of $\mathbf{x}$ and $\mathbf{y}$ is defined to be

$$\mathbf{x}^*\mathbf{y} = \bar{\mathbf{x}}^T\mathbf{y} = \bar{x}_1y_1 + \cdots + \bar{x}_ny_n$$

Two vectors are *orthogonal* to one another if their inner product is 0. The vector $\vec{\mathbf{0}}$ is orthogonal to all vectors. Two orthogonal non-zero vectors must be linearly independent. For, if $\mathbf{x}^*\mathbf{y} = 0$ and $a\mathbf{x}+b\mathbf{y} = 0$, then $0 = a\mathbf{x}^*\mathbf{x} + b\mathbf{x}^*\mathbf{y} = a\mathbf{x}^*\mathbf{x}$. This implies $a = 0$, which in turns implies $b = 0$ also. With the same reasoning, one easily shows that a set of pairwise orthogonal non-zero vectors must be linearly independent.

If A is any complex matrix, recall that the *Hermitian transpose* A^* of A is defined to be $\bar{A}^T$, and that A is said to be *Hermitian* if $A = A^*$. A real matrix is Hermitian if and only if it is symmetric. Also notice that the diagonal entries of a Hermitian matrix must be real, because they are equal to their respective complex conjugates. The next lemma lists several useful properties of a Hermitian matrix.

Lemma 2.6. *Let A be a Hermitian matrix, then*

(i) for all $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x}^*A\mathbf{x}$ is real.

(ii) every eigenvalue of A is real.

(iii) the eigenvectors of A, if correspond to distinct eigenvalues, are orthogonal to one another.

Proof. It is straightforward that

(i) $(\mathbf{x}^*A\mathbf{x})^* = \mathbf{x}^*A^*\mathbf{x}^{**} = \mathbf{x}^*A\mathbf{x}$.

(ii) $A\mathbf{x} = \lambda\mathbf{x}$ implies $\lambda = \frac{\mathbf{x}^*A\mathbf{x}}{\mathbf{x}^*\mathbf{x}}$.

(iii) Suppose $A\mathbf{x} = \lambda_1\mathbf{x}$, $A\mathbf{y} = \lambda_2\mathbf{y}$, and $\lambda_1 \neq \lambda_2$, then

$$(\lambda_1\mathbf{x})^*\mathbf{y} = (A\mathbf{x})^*\mathbf{y} = \mathbf{x}^*A\mathbf{y} = \mathbf{x}^*(\lambda_2\mathbf{y}).$$

Hence, $(\lambda_1 - \lambda_2)\mathbf{x}^*\mathbf{y} = 0$, implying $\mathbf{x}^*\mathbf{y} = 0$.

□

2.3 Orthonormal and unitary matrices

A real matrix Q is said to be *orthogonal* if $Q^TQ = I$. A complex matrix U is *unitary* if $U^*U = I$. In other words, the columns of U (and Q) are *orthonormal*. Obviously being orthogonal is a special case of being unitary. We state without proof a simple proposition.

Proposition 2.7. *Let U be a unitary matrix, then*

(i) $(U\mathbf{x})^*(U\mathbf{y}) = \mathbf{x}^*\mathbf{y}$, and $\|U\mathbf{x}\|^2 = \|\mathbf{x}\|^2$.

(ii) Every eigenvalue λ of U has modulus 1 (i.e. $|\lambda| = \lambda^*\lambda = 1$).

(iii) Eigenvectors corresponding to distinct eigenvalues of U are orthogonal.

(iv) If U' is another unitary matrix, then UU' is unitary.

3 The Spectral Theorem and the Jordan canonical form

Two matrices A and B are said to be *similar* iff there is an invertible matrix M such that $M^{-1}AM = B$. Thus, a matrix is diagonalizable iff it is similar to a diagonal matrix. Similarity is obviously an equivalence relation. The following proposition shows what is common among matrices in the same similarity equivalent class.

Proposition 3.1. *If $B = M^{-1}AM$, then A and B have the same eigenvalues. Moreover, an eigenvector $\mathbf{x}$ of A corresponds to an eigenvector $M^{-1}\mathbf{x}$ of B.*

Proof. $A\mathbf{x} = \lambda\mathbf{x}$ implies $(M^{-1}A)\mathbf{x} = \lambda M^{-1}\mathbf{x}$, or $(BM^{-1})\mathbf{x} = \lambda(M^{-1}\mathbf{x})$. □

An eigenvector corresponding to an eigenvalue λ is called a λ-*eigenvector*. The vector space spanned by all λ-eigenvectors is called the λ-*eigenspace*. We shall often use V_λ to denote this space.

Corollary 3.2. *If A and B are similar, then the corresponding eigenspaces of A and B have the same dimension.*

Proof. Suppose $B = M^{-1}AM$, then the mapping $\phi : x \to M^{-1}x$ is an invertible linear transformation from one eigenspace of A to the corresponding eigenspace of B.[1] □

If two matrices A and B are similar, then we can say a lot about A if we know B. Hence, we would like to find B similar to A where B is as "simple" as possible. The first "simple" form is the upper-triangular form, as shown by the following Lemma, which is sometime referred to as the Jacobi Theorem.

Lemma 3.3 (Schur's lemma). *For any $n \times n$ matrix A, there is a unitary matrix U such that $B = U^{-1}AU$ is upper triangular. Hence, the eigenvalues of A are on the diagonal of B.*

Proof. We show this by induction on n. The lemma holds when $n = 1$. When $n > 1$, over $\mathbb{C}$ A must have at least one eigenvalue λ_1. Let $\mathbf{x}'_1$ be a corresponding eigenvector. Use the *Gram-Schmidt* process to extend x'_1 to an orthonormal basis $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ of $\mathbb{C}^n$. Let U_1 be the matrix whose columns are these vectors in order. From the fact that $U_1^{-1} = U_1^*$, it is easy to see that

$$U_1^{-1}AU_1 = \begin{bmatrix} \lambda_1 & * & * & \dots & * \\ 0 & * & * & \dots & * \\ 0 & * & * & \dots & * \\ \dots & \dots & \dots & \dots & \dots \\ 0 & * & * & \dots & * \end{bmatrix}.$$

Now, let $A' = (U_1^{-1}AU_1)_{11}$ (crossing off row 1 and column 1 of $U_1^{-1}AU_1$). Then, by induction there exists an $(n-1) \times (n-1)$ unitary matrix M such that $M^{-1}A'M$ is upper triangular. Let U_2 be the $n \times n$ matrix obtained by adding a new row and new column to M with all new entries equal 0 except $(U_2)_{11} = 1$. Clearly U_2 is unitary and $U_2^{-1}(U_1^{-1}AU_1)U_2$ is upper triangular. Letting $U = U_1U_2$ completes the proof. □

The following theorem is one of the most important theorems in elementary linear algebra, beside the Jordan form.

Theorem 3.4 (Spectral theorem). *Every real symmetric matrix can be diagonalized by an orthogonal matrix, and every Hermitian matrix can be diagonalized by a unitary matrix:*

$$\textit{(real case)} \quad Q^{-1}AQ = \Lambda, \quad \textit{(complex case)} \quad U^{-1}AU = \Lambda$$

Moreover, in both cases all the eigenvalues are real.

Proof. The real case follows from the complex case. Firstly, by Schur's lemma there is a unitary matrix U such that $U^{-1}AU$ is upper triangular. Moreover,

$$(U^{-1}AU)^* = U^*A^*(U^{-1})^* = U^{-1}AU,$$

i.e. $U^{-1}AU$ is also Hermitian. But an upper triangular Hermitian matrix must be diagonal. The realness of the eigenvalues follow from Lemma 2.6. □

[1] I have not define linear transformation yet. The thing to remember is that if there is an invertible linear transformation from one vector space to another, then the two vector spaces have the same dimension. Invertible linear transformations are like isomorphisms or bijections, in some sense. A curious student should try to prove this fact directly without using the term linear transformation.

Theorem 3.5 (The Jordan canonical form). *If a matrix A has s linearly independent eigenvectors, then it is similar to a matrix which is in **Jordan form** with s square blocks on the diagonal:*

$$M^{-1}AM = \begin{bmatrix} B_1 & 0 & 0 & \dots & 0 \\ 0 & B_2 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & B_s \end{bmatrix}$$

Each block has exactly one 1-dimensional eigenspace, one eigenvalue, and 1's just above the diagonal:

$$B_j = \begin{bmatrix} \lambda_j & 1 & 0 & \dots & 0 \\ 0 & \lambda_j & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \dots & 0 \\ \dots & \dots & \dots & \dots & 1 \\ 0 & 0 & 0 & \dots & \lambda_j \end{bmatrix}$$

Proof. A proof could be read from Appendix B of [8]. Another proof is presented in [3], which has a nice combinatorial presentation in terms of digraphs. The fact that each Jordan block has exactly one 1-dimensional eigenspace is straightforward. The main statement is normally shown by induction in three steps. □

Corollary 3.6. *Let $n(\lambda)$ be the number of occurrences of λ on the diagonal of the Jordan form of A. The following hold*

1. $rank(A) = \sum_{\lambda_i \neq 0} n(\lambda_i) + n(0) - dim(V_0)$.
2. *If A is Hermitian, then the λ-eigenspace has dimension equal the multiplicity of λ as a solution to equation $p_A(x) = 0$.*
3. *In fact, in Hermitian case $\mathbb{C}^n = \bigoplus_i V_{\lambda_i}$ where V_{λ_i} denotes the λ_i-eigenspace.*

Proof. This follows directly from the *Jordan form* and our observation in Corollary 3.2. We are mostly concerned with the dimensions of eigenspaces, so we can think about Λ instead of A. Similar matrices have the same rank, so A and its Jordan form have the same rank. The Jordan form of A has rank equal the total number of non-zero eigenvalues on the diagonal plus the number of 1's in the Jordan blocks corresponding to the eigenvalue 0, which is exactly $n(0) - dim(V_0)$.

When A is Hermitian, it is diagonalizable. Every eigenvector corresponding to an *occurrence* of an eigenvalue λ is linearly independent from all others (including the eigenvector corresponding to another instance of the same λ). □

4 The Minimum Polynomial

I found the following very nice theorem stated without proof in a book called "Matrix Methods" by Richard Bronson. I'm sure we could find a proof in either [6] or [4], but I wasn't able to get them from the library. Here I present my little proof.

Theorem 4.1. *Suppose B_k is a Jordan block of size $(l+1)\times(l+1)$ corresponding to the eigenvalue λ_k of A, i.e.*

$$B_k = \begin{bmatrix} \lambda_k & 1 & 0 & \dots & 0 \\ 0 & \lambda_k & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \dots & \dots & \dots & \dots & 1 \\ 0 & 0 & 0 & \dots & \lambda_k \end{bmatrix}.$$

Then, for any polynomial $q(\lambda) \in \mathbb{C}[\lambda]$

$$q(B_k) = \begin{bmatrix} q(\lambda_k) & \frac{q'(\lambda_k)}{1!} & \frac{q''(\lambda_k)}{2!} & \dots & \frac{q^{(l)}(\lambda_k)}{l!} \\ 0 & q(\lambda_k) & \frac{q'(\lambda_k)}{1!} & \dots & \frac{q^{(l-1)}(\lambda_k)}{(l-1)!} \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \dots & \dots & \dots & \dots & \frac{q'(\lambda_k)}{1!} \\ 0 & 0 & 0 & \dots & q(\lambda_k) \end{bmatrix} \tag{9}$$

Proof. We only need to consider the case $q(x) = x^j, j \geq 0$, and then extend linearly into all polynomials. The case $j = 0$ is clear. Suppose equation (9) holds for $q(x) = x^{j-1}, j \geq 1$. Then, when $q(x) = x^j$ we have

$$\begin{aligned} q(B_k) &= B_k^{j-1} B_k \\ &= \begin{bmatrix} \lambda_k^{j-1} & \binom{j-1}{1}\lambda_k^{j-2} & \binom{j-1}{2}\lambda_k^{j-3} & \dots & \binom{j-1}{l}\lambda_k^{j-l-1} \\ 0 & \lambda_k^{j-1} & \binom{j-1}{1}\lambda_k^{j-2} & \dots & \binom{j-1}{l-1}\lambda_k^{j-l} \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \dots & \dots & \dots & \dots & \binom{j-1}{1}\lambda_k^{j-2} \\ 0 & 0 & 0 & 0 & \lambda_k^{j-1} \end{bmatrix} \begin{bmatrix} \lambda_k & 1 & 0 & \dots & 0 \\ 0 & \lambda_k & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \dots & \dots & \dots & \dots & 1 \\ 0 & 0 & 0 & \dots & \lambda_k \end{bmatrix} \\ &= \begin{bmatrix} \lambda_k^{j} & \binom{j}{1}\lambda_k^{j-1} & \binom{j}{2}\lambda_k^{j-2} & \dots & \binom{j}{l}\lambda_k^{j-l} \\ 0 & \lambda_k^{j} & \binom{j}{1}\lambda_k^{j-1} & \dots & \binom{j}{l-1}\lambda_k^{j-l+1} \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \dots & \dots & \dots & \dots & \binom{j}{1}\lambda_k^{j-1} \\ 0 & 0 & 0 & 0 & \lambda_k^{j} \end{bmatrix} \end{aligned}$$

□

The *minimum polynomial* $m_A(\lambda)$ of an $n \times n$ matrix A over the complex numbers is the monic polynomial of lowest degree such that $m_A(A) = 0$.

Lemma 4.2. *With the terminologies just stated, we have*

(i) $m_A(\lambda)$ *divides* $p_A(\lambda)$.

(ii) Every root of $p_A(\lambda)$ *is also a root of* $m_A(\lambda)$. *In other words, the eigenvalues of* A *are roots of* $m_A(\lambda)$.

(iii) A *is diagonalizable iff* $m_A(\lambda)$ *has no multiple roots.*

(iv) If $\{\lambda_i\}_{i=1}^s$ *are distinct eigenvalues of a Hermitian matrix* A, *then* $m_A(\lambda) = \prod_{i=1}^s(\lambda - \lambda_i)$.

Proof. (i) $m_A(\lambda)$ must divide every polynomial $q(\lambda)$ with $q(A) = 0$, since otherwise $q(\lambda) = h(\lambda)m_A(\lambda) + r(\lambda)$ implies $r(A) = 0$ while $r(\lambda)$ has smaller degree than $m_A(\lambda)$. On the other hand, by the Cayley-Hamilton Theorem (theorem 11.1), $p_A(A) = 0$.

(ii) Notice that $Ax = \lambda x$ implies $A^i x = \lambda^i x$. Thus, for any λ_k eigenvector x of A $\vec{0} = m_A(A)x = \sum_i c_i A^i x = \sum_i c_i \lambda_k^i x = m(\lambda_k)x$. This implies λ_k is a root of $m(\lambda)$.

(iii) $(\Rightarrow)$. Suppose $M^{-1}AM = \Lambda$ for some invertible matrix M, and $\lambda_1, \dots, \lambda_s$ are distinct eigenvalues of A. By (i) and (ii), we only need to show A is a root of $m_A(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)$. It is easy to see that for any polynomial $q(\lambda)$, $q(A) = Mq(\Lambda)M^{-1}$. In particular, $m_A(A) = M^{-1}m_A(\Lambda)M = 0$, since $m_A(\Lambda) = 0$.

$(\Leftarrow)$. Now we assume $m_A(\lambda)$ has no multiple root, which implies $m_A(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)$. By Proposition 2.2, we shall show that A has n linearly independent eigenvectors. Firstly, notice that if the Jordan form of A is

$$M^{-1}AM = \begin{bmatrix} B_1 & 0 & 0 & \dots & 0 \\ 0 & B_2 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & B_s \end{bmatrix}.$$

Then, for any $q(\lambda) \in \mathbb{C}[\lambda]$ we have

$$\begin{aligned} M^{-1}q(A)M &= q\left(\begin{bmatrix} B_1 & 0 & 0 & \dots & 0 \\ 0 & B_2 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & B_s \end{bmatrix}\right) \\ &= \begin{bmatrix} q(B_1) & 0 & 0 & \dots & 0 \\ 0 & q(B_2) & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & q(B_s) \end{bmatrix} \end{aligned}$$

So, $\prod_{i=1}^s (A - \lambda_i I) = 0$ implies $\prod_{i=1}^s (B_k - \lambda_i I) = 0$ for all $k = 1, \dots, s$. If A does not have n linearly independent eigenvectors, one of the blocks B_k must have size > 1. Applying Theorem 4.1 with $q(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)$, we see that $q(B_k)$ does not vanish since $q'(\lambda_i) \neq 0, \forall i \in [s]$. Contradiction!

(iv) Follows from (iii) since a Hermitian matrix is diagonalizable.

□

5 Positive definite matrices

The purpose of this section is to develop a necessary and sufficient conditions for a real symmetric matrix A (or Hermitian in general) to be *positive definite*. This is essentially the conditions for a *quadratic form* on $\mathbb{R}^n$ to have a minimum at some point.

5.1 Some analysis

Let us first recall two key theorems from real analysis, stated without proofs. For $f : \mathbb{R}^n \to \mathbb{R}$, if for some $a \in \mathbb{R}^n$ $\frac{\partial f}{\partial x_i}(a) = 0$, then a is called a *stationary point* of f.

Theorem 5.1 (The second derivative test). *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and its partial derivatives up to and including order 2 are continuous in a ball $B(a, r)$ (centered at $a \in \mathbb{R}^n$, radius r). Suppose that f has a stationary point at a. For $h = (h_1, \dots, h_n)$, define $\Delta f(a, h) = f(a + h) - f(a)$; also define*

$$Q(h) = \frac{1}{2!} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) h_i h_j$$

then,

1. *If $Q(h) > 0$ for $h \neq 0$, then f has a strict local minimum at a.*
2. *If $Q(h) < 0$ for $h \neq 0$, then f has a strict local maximum at a.*
3. *If $Q(h)$ has a positive maximum and a negative minimum, then $\Delta f(a, h)$ changes sign in any ball $B(a, \rho)$ such that $\rho < r$.*

Note. (3.) says that at any close neighborhood of a, there are some points b and c such that $f(b) > f(a)$ and $f(c) < f(a)$.

Example 5.2. Let us look at a quadratic form $F(x_1, \dots, x_n)$ with all real coefficients, i.e. every term of F has degree at most 2. Let A be the matrix defined by $a_{ij} = \frac{\partial^2 F}{\partial x_i \partial x_j}(a)$. Clearly, A is a real symmetric matrix. For any vector $h \in \mathbb{R}^n$,

$$\begin{aligned} h^T A h &= \begin{bmatrix} h_1 & h_2 & \dots & h_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix} \\ &= \sum_{i,j=1}^{n} a_{ij} h_i h_j \\ &= 2Q(h) \end{aligned}$$

So, $F(x_1, \dots, x_n)$ has a minimum at $(0, \dots, 0)$ (which is a stationary point of F) iff $h^T A h > 0$ for all $h \neq 0$.

Definition 5.3. A non-singular $n \times n$ Hermitian matrix A is said to be *positive definite* if $x^* A x > 0$ for all non zero vector $x \in \mathbb{C}^n$. A is *positive semidefinite* if we only require $x^* A x \geq 0$. The terms *negative definite* and *negative semidefinite* can be defined similarly.

Note. Continuing with our example, clearly $F(x_1, \dots, x_n)$ has a minimum at $(0, \dots, 0)$ iff A is positive definite. Also, since we already showed that if A is Hermitian, then $x^* A x$ is real, the definitions given above make sense.

A function f is in C^1 on some domain $D \subseteq \mathbb{R}^n$ if f and all its first order derivatives are continuous on D. For $a \in D$, $\bigtriangledown f(a) := (\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a))$.

Theorem 5.4 (Lagrange's multiplier rule). *Suppose that $f, \varphi_1, \ldots, \varphi_k$ are C^1 functions on an open set D in $\mathbb{R}^n$ containing a point a, that the vectors $\bigtriangledown\varphi_1(a), \ldots, \bigtriangledown\varphi_k(a)$ are linearly independent, and that f takes on its minimum among all points of D_0 at x_0, where D_0 is the subset of D so that for all $x \in D_0$,*

$$\varphi_i(x_1, \ldots, x_n) = 0,\ i = 1, \ldots, k$$

Then, if $F : \mathbb{R}^{n+k} \to \mathbb{R}$ is defined to be

$$F(x, \lambda) = f(x) - \sum_{i=1}^{k} \lambda_i \varphi_i(x)$$

then there exists $\lambda^0 \in \mathbb{R}^k$ such that

$$\begin{aligned} \frac{\partial F}{\partial x_i}(a, \lambda^0) &= 0 \quad i = 1, \ldots, n \\ \frac{\partial F}{\partial \lambda_j^0}(a, \lambda^0) &= 0 \quad i = 1, \ldots, k \end{aligned}$$

Note. This theorem essentially says that the maxima (or minima) of f subject to the *side conditions* $\varphi_1 = \cdots = \varphi_k = 0$ are among the maxima (or minima) of the function F without any constraints.

Example 5.5. To find the maximum of $f(x) = x_1 + 3x_2 - 2x_3$ *on* the sphere $14 - (x_1^2 + x_2^2 + x_3^2) = 0$, we let

$$F(x, \lambda) = x_1 + 3x_2 - 2x_3 + \lambda(x_1^2 + x_2^2 + x_3^2 - 14)$$

Then, $\frac{\partial F}{\partial x_1} = 1+2\lambda x_1$, $\frac{\partial F}{\partial x_2} = 3+2\lambda x_2$, $\frac{\partial F}{\partial x_3} = -2+2\lambda x_3$, and $\frac{\partial F}{\partial \lambda} = x_1^2+x_2^2+x_3^2-14$. Solving $\frac{\partial F}{\partial x_i} = 0$ we obtain two solutions $(x, \lambda) = (1, 3, -2, -1/2)$ and $(-1, -3, 2, 1/2)$. Which of these solutions give a maximum or a minimum ? We apply the second derivative test. All second derivatives of F are 0 except $\partial^2 F/\partial x_1^2 = \partial^2 F/\partial x_2^2 = \partial^2 F/\partial x_3^2 = 2\lambda$. $Q(h) = 2\lambda(h_1^2 + h_2^2 + h_3^2)$ has the same sign as λ. Hence, the first solution gives the maximum value of 14, the second solution gives the minimum value of -13.

5.2 Conditions for positive-definiteness

Now we are ready to specify the necessary and sufficient conditions for a Hermitian matrix to be positive definite, or positive semidefinite for that matter.

Theorem 5.6. *Each of the following tests is a necessary and sufficient condition for the real symmetric matrix A to be positive definite.*

(a) $x^T Ax > 0$ *for all non-zero vector x.*

(b) All the eigenvalues of A are positive.

(c) All the upper left submatrices A_k of A have positive determinants.

(d) If we apply Gaussian elimination on A without row exchanges, all the pivots satisfy $p_i > 0$.

Note. (a) and (b) hold for Hermitian matrices also.

Proof. $(a \Rightarrow b)$. Suppose x_i is a unit λ_i-eigenvector, then $0 < x_i^T Ax_i = x_i^T \lambda_i x_i = \lambda_i$.

$(b \Rightarrow a)$. Since A is real symmetric, it has a full set of orthonormal eigenvectors $\{x_1, \ldots, x_n\}$ by the Spectral theorem. For each non-zero vector $x \in \mathbb{R}^n$, suppose $x = c_1x_1 + \cdots + c_nx_n$, then

$$Ax = A(c_1x_1 + \cdots + c_nx_n) = c_1\lambda_1x_1 + \cdots + c_n\lambda_nx_n$$

Because the x_i are orthonormal, we get

$$\begin{aligned} x^T A x &= (c_1 x_1^T + \cdots + c_n x_n^T)(c_1 \lambda_1 x_1 + \cdots + c_n \lambda_n x_n) \\ &= \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2. \end{aligned}$$

Thus, every $\lambda_i > 0$ implies $x^T A x > 0$ whenever $x \neq 0$.

$(a \Rightarrow c)$. We know $\det A = \lambda_1 \ldots \lambda_n > 0$. To prove the same result for all A_k, we look at a non-zero vector x whose last $n - k$ components are 0, then

$$0 < x^T A x = \begin{bmatrix} x_k^T & 0 \end{bmatrix} \begin{bmatrix} A_k & * \\ * & * \end{bmatrix} \begin{bmatrix} x_k \\ 0 \end{bmatrix} = x_k^T A_k x^k.$$

Thus, $\det A_k > 0$ follows by induction.

$(c \Rightarrow d)$. Without row exchanges, the pivot p_k in Gaussian elimination is $\det A_k / \det A_{k-1}$. This can also be proved easily by induction.

$(d \Rightarrow a)$. Gaussian elimination gives us a LDU factorization of A where all diagonal entries of L and U are 1. Also, the diagonal entries d_i of D is exactly the i^{th} pivot p_i. The fact that A is symmetric implies $L = U^T$, hence $A = LDL^T$, which gives

$$x^T A x = (x^T L)(D)(L^T x) = d_1 (L^T x)_1^2 + d_2 (L^T x)_2^2 + \cdots + d_n (L^T x)_n^2$$

Since L is fully ranked, $L^T x \neq 0$ whenever $x \neq 0$. So the pivots $d_i > 0$ implies $x^T A x > 0$ for all non-zero vectors x. □

6 The Rayleigh's quotient and the variational characterizations

For a Hermitian matrix A, the following is known as the *Rayleigh's quotient* :

$$R(x) = \frac{x^* A x}{x^* x}.$$

Theorem 6.1 (Rayleigh-Ritz). *Suppose $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of a real symmetric matrix A. Then, the quotient $R(x)$ is maximized at any λ_1-eigenvector $x = x_1$ with maximum value λ_1. $R(x)$ is minimized at any λ_n-eigenvector $x = x_n$ with minimum value λ_n,*

Proof. Let Q be a matrix whose columns are a set of orthonormal eigenvectors $\{x_1, \ldots, x_n\}$ of A corresponding to $\lambda_1, \ldots, \lambda_n$, respectively. Writing x as a linear combination of columns of Q: $x = Qy$, then since $Q^T A Q = \Lambda$ we have

$$R(x) = \frac{x^T A x}{x^T x} = \frac{(Q^T y)^T A (Q^T y)}{(Q^T y)^T (Q^T y)} = \frac{y^T \Lambda x}{y^T y} = \frac{\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2}{y_1^2 + \cdots + y_n^2}$$

Hence,

$$\lambda_1 \geq R(x) = \frac{\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2}{y_1^2 + \cdots + y_n^2} \geq \lambda_n$$

Moreover, $R(x) = \lambda_1$ when $y_1 \neq 0$ and $y_i = 0, \ \forall i > 1$. This means $x = Qy$ is a λ_1-eigenvector. The case $R(x) = \lambda_n$ case is proved similarly. □

The theorem above is also referred to as the *Rayleigh principle*, which also holds when A is Hermitian. The proof is identical, except that we have to replace transposition (T) by Hermitian transposition (*). An equivalent statement of the principle is as follows.

Corollary 6.2. *Suppose $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of a real symmetric matrix A. Over all non-zero unit vectors $x \in \mathbb{R}^n$, $x^T Ax$ is maximized at a unit λ_1-eigenvector, with maximum value λ_1, and minimized at a unit λ_n-eigenvector, with minimum value λ_n.*

Rayleigh's principle essentially states that

$$\lambda_1 = \max_{x \in \mathbb{R}^n} R(x) \quad \text{and} \quad \lambda_n = \min_{x \in \mathbb{R}^n} R(x)$$

What about the rest of the eigenvalues ? Here is a simple answer, stated without proof. The proof is simple enough.

Theorem 6.3. *Suppose $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of a Hermitian matrix A, and $u_1, \ldots, u_n$ are the corresponding set of orthonormal eigenvectors. Then,*

$$\begin{aligned} \lambda_k &= \max_{\substack{0 \neq x \in \mathbb{C}^n \\ x \perp u_1, \ldots, u_{k-1}}} R_A(x) \\ \lambda_k &= \min_{\substack{0 \neq x \in \mathbb{C}^n \\ x \perp u_{k+1}, \ldots, u_n}} R_A(x) \end{aligned}$$

The theorem has a pitfall that sometime we don't know the eigenvectors. The following generalization of Rayleigh's principle, sometime referred to as the *minimax and maximin principles for eigenvalues*, fill the hole by not requiring us to know that eigenvectors.

Theorem 6.4 (Courant-Fisher). *Let V_k be the set of all k-dimensional subspaces of $\mathbb{C}^n$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of a Hermitian matrix A. Then,*

$$\begin{aligned} \lambda_k &= \max_{S \in V_k} \left[\min_{\substack{x \in S \\ x \neq 0}} R(x) \right] \\ &= \min_{S \in V_{n-k+1}} \left[\max_{\substack{x \in S \\ x \neq 0}} R(x) \right] \end{aligned}$$

Note. It should be noted that the previous two theorems are often referred to as the *variational characterization* of the eigenvalues.

Proof. Let $U = [u_1, u_2, \ldots, u_n]$ be the unitary matrix with unit eigenvectors $u_1, \ldots, u_n$ corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$. Let us first fix $S \in V_k$ and let S' be the image of S under the invertible linear transformation represented by U^*. Obviously, $dim(S') = k$. We have already known that $R(x)$ is bounded, so it is safe to say the following, with $x \neq 0, y \neq 0$ being implicit.

$$
\begin{aligned}
\inf_{x\in S} R(x) &= \inf_{x\in S} \frac{x^*Ax}{x^*x} \\
&= \inf_{x\in S} \frac{(U^*x)^*\Lambda(U^*x)}{(U^*x)^*(U^*x)} \\
&= \inf_{y\in S'} \frac{y^*\Lambda y}{y^*y} \\
&\le \inf_{\substack{y\in S' \\ y_1=\cdots=y_{k-1}=0}} \frac{y^*\Lambda y}{y^*y} \\
&= \inf_{\substack{y\in S' \\ y_1=\cdots=y_{k-1}=0}} \frac{\lambda_1 y_1^2+\cdots+\lambda_n y_n^2}{y_1^2+\cdots+y_n^2} \\
&= \inf_{\substack{y\in S' \\ y_1=\cdots=y_{k-1}=0}} \frac{\lambda_k y_k^2+\cdots+\lambda_n y_n^2}{y_k^2+\cdots+y_n^2} \\
&\le \lambda_k
\end{aligned}
$$

The inequality in line 4 is justified by the fact that there is a non zero vector $y \in S'$ such that $y_1 = \cdots = y_{k-1} = 0$. To get this vector, put k basis vectors of S' into the rows of a $k \times n$ matrix and do Gaussian elimination.

Now, S was chosen arbitrarily, so it is also true that

$$\sup_{S\in V_k} \inf_{x\in S} R(x) \le \lambda_k$$

Moreover, $R(u_k) = (U^*u_k)^*\Lambda(U^*u_k) = e_k\Lambda e_k = \lambda_k$. Thus, the infimum and supremum can be changed to minimum and maximum, and the inequality can be changed to equality. The other equality can be proven similarly. □

This theorem has a very important and beautiful corollary, called the *Interlacing of eigenvalues* to be presented in the next section. Let us introduce a simple corollary.

Corollary 6.5. *Let A be an $n \times n$ Hermitian matrix, let k be a given integer with $1 \le k \le n$, let $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ be the eigenvalues of A, and let S_k be a given k-dimensional subspace of $\mathbb{C}^n$. The following hold*

(a) If there exists c_1 such that $R(x) \le c_1$ for all $x \in S_k$, then $c_1 \ge \lambda_{n-k+1} \ge \cdots \ge \lambda_n$.

(b) If there exists c_2 such that $R(x) \ge c_2$ for all $x \in S_k$, then $\lambda_1 \ge \cdots \ge \lambda_k \ge c_2$.

Proof. It is almost straightforward from the Courant-Fisher theorem that

(a)
$$c_1 \ge \max_{0\ne x\in S_k} R(x) \ge \min_{dim(S)=n-(n-k+1)+1} \max_{0\ne x\in S} R(x) = \lambda_{n-k+1}$$

(b)
$$c_2 \le \min_{0\ne x\in S_k} R(x) \le \max_{dim(S)=k} \min_{0\ne x\in S} R(x) = \lambda_k$$

□

7 Other proofs of the variational characterizations

The presentation below follows that in [2]. Let A be a Hermitian matrix of order n with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Let $U = [u_1, \ldots, u_n]$ be the unitary matrix where u_i is the unit λ_i-eigenvector of A. Then,

$$A = \sum_j \lambda_j u_j u_j^*. \tag{10}$$

This equation is called the *spectral resolution* of A. From the equation we obtain

$$\langle x, Ax \rangle = \sum_j \lambda_j \left(u_j^* x\right)^* \left(u_j^* x\right) = \sum_j \lambda_j |u_j^* x|^2$$

When $|x| = 1$, we have

$$\sum_j |u_j^* x|^2 = \sum_j \left(u_j^* x\right)^* \left(u_j^* x\right) = x^* U U^* x = 1.$$

As we can always normalize x, we shall only consider unit vectors x in this section from here on.

Lemma 7.1. *Suppose* $1 \leq i < k \leq n$*, then*

$$\{\langle x, Ax \rangle \mid |x| = 1, x \in span\{u_i, \ldots, u_k\}\} = [\lambda_k, \lambda_i].$$

Additionally,

$$\langle u_i, Au_i \rangle = \lambda_i, \ \forall i$$

Proof. The fact that $\langle u_i, Au_i \rangle = \lambda_i$ is trivial. Assume x is a unit vector in $span\{u_i, \ldots, u_k\}$, then

$$\langle x, Ax \rangle = \sum_{i \leq j \leq k} \lambda_j |u_j^* x|^2,$$

which yields easily

$$\{\langle x, Ax \rangle \mid |x| = 1, x \in span\{u_i, \ldots, u_k\}\} \subseteq [\lambda_k, \lambda_i].$$

For the converse, let $\lambda \in [\lambda_k, \lambda_i]$. Let $y \in \mathbb{C}^n$ be a column vector all of whose components are 0, except that

$$\begin{aligned} y_i &= \sqrt{\frac{\lambda - \lambda_k}{\lambda_i - \lambda_k}} \\ y_k &= \sqrt{\frac{\lambda_i - \lambda}{\lambda_i - \lambda_k}}. \end{aligned}$$

Then, $x = Uy \in span\{u_i, u_k\} \subseteq span\{u_i, \ldots, u_k\}$, x is obviously a unit vector. Moreover,

$$\langle x, Ax \rangle = (Uy)^* A(Uy) = y^* \Lambda y = \lambda_i y_i^2 + \lambda_k y_k^2 = \lambda.$$

□

Lemma 7.1 gives an interesting proof of Theorem 6.4.

Another proof of Courant-Fisher Theorem. We shall show that

$$\lambda_k = \max_{S \in V_k} \left[\min_{\substack{x \in S \\ x \neq 0}} R(x) \right].$$

The other equality is obtained similarly. Fix $S \in V_k$, let $W = span\{u_k, \dots, u_n\}$, then W and S have total dimension $n+1$. Thus, there exists a unit vector $x \in W \cap S$. Lemma 7.1 implies $R(x) = \langle x, Ax \rangle \in [\lambda_n, \lambda_k]$. Consequently,

$$\min_{\substack{x \in S \\ x \neq 0}} R(x) \leq \lambda_k.$$

Equality is obtained by picking any k-dimensional subspace S containing u_k, and $x = u_k$. □

8 Applications of the variational characterizations and minimax principle

Throughout the rest of this section, we use M_n to denote the set of all $n \times n$ matrices over $\mathbb{C}$ (i.e. $M_n \approx \mathbb{C}^{n^2}$). The first application is a very important and beautiful theorem named the *Interlacing of Eigenvalues Theorem*.

Theorem 8.1 (Interlacing of eigenvalues). *Let A be a Hermitian matrix with eigenvalues $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$. Let B be the matrix obtained from A by removing row i and column i, for any $i \in [n]$. Suppose B has eigenvalues $\beta_1 \geq \cdots \geq \beta_{n-1}$, then*

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \beta_2 \geq \cdots \geq \beta_{n-1} \geq \alpha_n$$

Note. A proof of this theorem using Spectral Decomposition Theorem can also be given, but it is not very instructive.

Proof. We can safely assume that $i = n$ for the ease of presentation. We would like to show that as $1 \leq k \leq n-1$, $\alpha_k \geq \beta_k \geq \alpha_{k+1}$. Let $x = [y^T x_n]^T \in \mathbb{C}^n$ where $y \in \mathbb{C}^{n-1}$. Note that if $x_n = 0$ then $x^*Ax = y^*By$. We first use the maximin form of Courant-Fisher theorem to write

$$\begin{aligned}
\alpha_k &= \max_{\substack{S \subseteq \mathbb{C}^n \\ dim(S)=k}} \min_{0 \neq x \in S} \frac{x^*Ax}{x^*x} \\
&\geq \max_{\substack{S \subseteq \{e_n\}^\perp \\ dim(S)=k}} \min_{0 \neq x \in S} \frac{x^*Ax}{x^*x} \\
&= \max_{\substack{S \subseteq \{e_n\}^\perp \\ dim(S)=k}} \min_{\substack{0 \neq x \in S \\ x_n = 0}} \frac{x^*Ax}{x^*x} \\
&= \max_{\substack{S \subseteq \mathbb{C}^{n-1} \\ dim(S)=k}} \min_{0 \neq y \in S} \frac{y^*By}{y^*y} \\
&= \beta_k
\end{aligned}$$

Now, we use the minimax form of the theorem to obtain $\alpha_{k+1} \leq \beta_k$.

$$\begin{aligned}
\alpha_{k+1} &= \min_{\substack{S\subseteq\mathbb{C}^n\\ dim(S)=n-(k+1)+1}} \max_{0\neq x\in S} \frac{x^*Ax}{x^*x} \\
&\leq \min_{\substack{S\subseteq\{e_n\}^\perp\\ dim(S)=n-k}} \max_{0\neq x\in S} \frac{x^*Ax}{x^*x} \\
&= \min_{\substack{S\subseteq\{e_n\}^\perp\\ dim(S)=n-k}} \max_{\substack{0\neq x\in S\\ x_n=0}} \frac{x^*Ax}{x^*x} \\
&= \min_{\substack{S\subseteq\mathbb{C}^{n-1}\\ dim(S)=(n-1)-k+1}} \max_{0\neq y\in S} \frac{y^*By}{y^*y} \\
&= \beta_k
\end{aligned}$$

□

The converse of the Interlacing of Eigenvalues Theorem is also true.

Theorem 8.2. *Given real numbers*

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \beta_2 \geq \cdots \geq \beta_{n-1} \geq \alpha_n. \tag{11}$$

Let $B = diag[\beta_1, \ldots, \beta_{n-1}$. *Then, there exist a vector* $y \in \mathbb{R}^{n-1}$ *and a real number* a *such that the matrix*

$$A = \begin{bmatrix} B & y \\ y^T & a \end{bmatrix}$$

has eigenvalues $\alpha_1, \ldots, \alpha_n$.

Proof. Firstly, $a = \sum_{i=1}^n \alpha_i - \sum_{i=1}^{n-1} \beta_i$. To determine the vector $y = [y_1, \ldots, y_{n-1}]^T$, we evaluate $f(x) = \det(Ix - A)$.

$$\begin{aligned}
\det(Ix - A) &= \det \begin{bmatrix} x-\beta_1 & \ldots & 0 & y_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & x-\beta_{n-1} & y_{n-1} \\ y_1 & \ldots & y_{n-1} & x-a \end{bmatrix} \\
&= (x-\beta_1)\ldots(x-\beta_{n-1})\left(x - a - \frac{y_1^2}{x-\beta_1} - \cdots - \frac{y_{n-1}^2}{x-\beta_{n-1}}\right).
\end{aligned}$$

Let $g(x) = (x-\beta_1)\ldots(x-\beta_{n-1})$, then

$$f(x) = g(x)(x-a) + r(x),$$

where

$$r(x) = -y_1^2 \frac{g(x)}{x-\beta_1} - \cdots - y_{n-1}^2 \frac{g(x)}{x-\beta_{n-1}} \tag{12}$$

is a polynomial of degree $n-2$. We want $f(x) = (x-\alpha_1)\ldots(x-\alpha_n)$, which could be used to solve for the y_i.

If the β_i are all distinct, then $r(x)$ is determined at $n-1$ points: $r(\beta_i) = f(\beta_i)$. Lagrange interpolation gives:

$$r(x) = \sum_{i=1}^{n-1} f(\beta_i) \frac{g(x)}{g'(\beta_i)(x-\beta_i)} \tag{13}$$

Comparing (12) and (13), we conclude that if for all $i = 1, \dots, n-1$,

$$\frac{f(\beta_i)}{g'(\beta_i)} = \frac{(\beta_i - \alpha_1)\dots(\beta_i - \alpha_{i-1})(\beta_i - \alpha_i)(\beta_i - \alpha_{i+1})\dots(\beta_i - \alpha_n)}{(\beta_i - \beta_1)\dots(\beta_i - \beta_{i-1})(\beta_i - \beta_{i+1})\dots(\beta_i - \beta_n)} \le 0,$$

then we can solve for the y_i. The interlacing condition (11) implies that $(\beta_i - \alpha_j)$ and $(\beta_i - \beta_j)$ have the same sign except when $j = i$. Hence, $\frac{f(\beta_i)}{g'(\beta_i)} \le 0$ as desired.

If, say, $\beta_1 = \dots = \beta_k > \beta_{k+1} \ge \dots$, then the interlacing condition (11) forces $\beta_1 = \dots = \beta_k = \alpha_2 \dots = \alpha_k$. Hence, we can divide both sides of $f(x) = g(x)(x-a) + r(x)$ by $(x-\beta_1)^{k-1}$ to eliminate the multiple root β_1 of $g(x)$. After all multiple roots have been eliminated this way, we can proceed as before. □

Hermann Weyl (1912, [11]) derived a set of very interesting inequalities concerning the eigenvalues of three Hermitian matrices A, B, and C where $C = A + B$. We shall follow the notations used in [2]. For any matrix A, let $\lambda_j^{\downarrow}(A)$ and $\lambda_j^{\uparrow}(A)$ denote the jth eigenvalue of A when all eigenvalues are weakly ordered decreasingly and increasingly, respectively. When given three Hermitian matrices A, B, and C where $C = A + B$, implicitly we define $\alpha_j = \lambda_j^{\downarrow}(A)$, $\beta_j = \lambda_j^{\downarrow}(B)$, and $\gamma_j = \lambda_j^{\downarrow}(C)$, unless otherwise specified.

Theorem 8.3 (Weyl, 1912). *Given Hermitian matrices A, B, and C of order n such that $C = A + B$. Then,*

$$\gamma_{i+j-1} \le \alpha_i + \beta_j \text{ for } i + j - 1 \le n. \tag{14}$$

Proof. For $k = 1, \dots, n$, let u_k, v_k, and w_k be the unit α_k, β_k, and γ_k eigenvectors of A, B, and C, respectively. The three vector spaces spanned by $\{u_i, \dots, u_n\}$, $\{v_j, \dots, v_n\}$, and $\{w_1, \dots, w_{i+j-1}\}$ have total dimension $2n + 1$. Hence, they have a non-trivial intersection. Let x be a unit vector in the intersection, then by Lemma 7.1

$$\begin{aligned} \langle x, Ax \rangle &\in [\alpha_n, \alpha_i] \\ \langle x, Bx \rangle &\in [\beta_n, \beta_j] \\ \langle x, Cx \rangle &\in [\gamma_{i+j-1}, \beta_1]. \end{aligned}$$

Thus,

$$\gamma_{i+j-1} \le \langle x, Cx \rangle = \langle x, Ax \rangle + \langle x, Bx \rangle \le \alpha_i + \beta_j.$$

□

A few interesting consequences are summarized as follows.

Corollary 8.4. *(i) For all $k = 1, \dots, n$.*

$$\alpha_k + \beta_n \ge \gamma_k \ge \alpha_k + \beta_1,$$

(ii)

$$\begin{aligned} \gamma_1 &\le \alpha_1 + \beta_1 \\ \gamma_n &\ge \alpha_n + \beta_n \end{aligned}$$

Proof. (i) The second inequality is obtained by specializing $j = 1$, $i = k$ in Theorem 8.3. The first inequality follows from the first by noting that $-C = -A - B$.

(ii) Applying Theorem 8.3 with $i = j = 1$ yields the first inequality. The second follows by the $-C = -A - B$ argument.

□

The following is a trivial consequence of the minimax principle.

Corollary 8.5 (Monotonicity principle). *Define a partial order of all Hermitian matrices as follows.*

$$A \leq B \text{ iff } \langle x, Ax \rangle \leq \langle x, Bx \rangle \; \forall x. \tag{15}$$

Then, for all $j = 1, \ldots, n$ we have $\lambda_j(A) \leq \lambda_j(B)$ whenever $A \leq B$.

Equivalently, if A and B are Hermitian with B being positive semidefinite, then

$$\lambda_k^{\downarrow}(A) \leq \lambda_k^{\downarrow}(A + B)$$

9 Sylvester's law of inertia

Material in this section follows closely that in the book *Matrix Analysis* by Horn and Johnson [5].

Definition 9.1. Let $A, B \in M_n$ be given. If there exists a non-singular matrix S such that

- $B = SAS^*$, then B is said to be $*$-congruent to A.
- $B = SAS^T$, then B is said to be T-congruent to A.

Note. These two notion of congruence must be closely related; they are the same if S is a real matrix. When it is not important to distinguish between the two, we use the term *congruence* without a prefix. Since S was required to be non-singular, congruent matrices have the same rank. Also note that, if A is Hermitian then so is SAS^*; if A is symmetric, then SAS^T is also symmetric.

Proposition 9.2. *Both $*$-congruence and T-congruence are equivalent relations.*

Proof. It is easy to verify that the relations are reflexive, symmetric and transitive. Only need to notice that S is non-singular. □

The set M_n, therefore, is partitioned into equivalence classes by congruence. As an abstract problem, we may seek a canonical representative of each equivalence class under each type of congruence. *Sylvester's law of inertia* gives us the affirmative answer for the $*$-congruence case, and thus also gives the answer for the set of real symmetric matrices.

Definition 9.3. Let $A \in M_n$ be a Hermitian matrix. The *inertia* of A is the ordered triple

$$i(A) = (i_+(A), i_-(A), i_0(A))$$

where $i_+(A)$ is the number of positive eigenvalues of A, $i_-(A)$ is the number of negative eigenvalues of A, and $i_0(A)$ is the number of zero eigenvalues of A, all counting multiplicity. The *signature* of A is the quantity $i_+(A) - i_-(A)$.

Note. Since $rank(A) = i_+(A) + i_-(A)$, the signature and the rank of A uniquely identify the inertia of A.

Theorem 9.4 (Sylvester's law of inertia). *Let $A, B \in M_n$ be Hermitian matrices. A and B are $*$-congruent if and only if A and B have the same inertia.*

Proof. $(\Rightarrow)$. Firstly, for any Hermitian matrix $C \in M_n$, C can be diagonalized by a unitary matrix U, i.e. $C = U\Lambda U^*$, with Λ being diagonal containing all eigenvalues of C. By multiplying U with a permutation matrix, we can safely assume that down the main diagonal of Λ, all positive eigenvalues go first: $\lambda_1, \ldots, \lambda_{i_+}$, then the negatives: $\lambda_{i_++1}, \ldots, \lambda_{i_++i_-}$, and the rest are 0's. Thus, if we set $D = diag(\sqrt{|\lambda_1|}, \ldots, \sqrt{|\lambda_{i_++i_-}|}, 0, \ldots, 0)$, then

$$\Lambda = D \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{bmatrix} D = DI_C D$$

with the entries not shown being 0. I_C is called the *inertia matrix* of C. Consequently, letting $S = UD$ (S is clearly non-singular) we get

$$C = U\Lambda U^* = UDI_C DU^* = SI_C S^* \tag{16}$$

Hence, if A and B have the same inertia, then they could be written in the form (16) with possibly a different S for each, but $I_A = I_B$. Since $*$-congruence is transitive, A is $*$-congruent to B as they are both congruent to I_A.

$(\Leftarrow)$. Now, assume $A = SBS^*$ for some non-singular matrix S. A and B have the same rank, so $i_0(A) = i_0(B)$. We are left to show that $i_+(A) = i_+(B)$. For convenience, let $a = i_+(A)$ and $b = i_+(B)$. Let $u_1, \ldots, u_a$ be the orthonormal eigenvectors for the positive eigenvalues of A, so that $dim(Span\{u_1, \ldots, u_a\}) = a$. If $x = c_1u_1 + \cdots + c_au_a$, then $x^*Ax = \lambda_1|c_1|^2 + \cdots + \lambda_a|c_a|^2 > 0$. But then

$$x^*Ax = x^*SBS^*x = (S^*x)^*B(S^*x) > 0$$

so $y^*By > 0$ for all vector y in $Span\{S^*u_1, \ldots, S^*u_a\}$, which also has dimension a. By Corollary 6.5, $b \geq a$. A similar argument shows that $a \geq b$, which completes the proof. □

Corollary 9.5. *Given $x \in \mathbb{R}^n$, if x^TAx can be written as the sum of m products involving two linear factors, that is*

$$x^TAx = \sum_{k=1}^{m}(\sum_{i \in S_k} b_{i,k}x_i)(\sum_{j \in T_k} c_{j,k}x_j)$$

Further assume that A has p positive eigenvalues and q positive eigenvalues (counting multiplicities), then $m \geq \max(p, q)$.

Proof. I have not been able to see why this corollary follows from Sylvester's law yet. A proof of the corollary can be given, but that's not the point. □

10 Majorizations

Several results in linear algebra are best presented through the concept of majorization. We first need a few definitions.

Let $v_1, \dots, v_m$ be vectors in $\mathbb{R}^n$. The vector

$$v = \lambda_1 v_1 + \cdots + \lambda_m v_m$$

is called the *linear combination* of the v_j; when $\sum \lambda_j = 1$, we get an *affine combination*; a *canonical combination* is a linear combination in which $\lambda_j \geq 0, \forall j$; and a *convex combination* is an affine combination which is also canonical. The *linear (affine, canonical, convex) hull* of $\{v_1, \dots, v_m\}$ is the set of all linear (affine, canonical, convex) combinations of the v_j. Note that in the above definitions, m could be infinite. The convex hull of a finite set of vectors is called a *(convex polyhedral) cone*.

Let $x = (x_1, \dots, x_n)$ be a vector in $\mathbb{R}^n$. The vector $x^{\downarrow} = (x_1^{\downarrow}, \dots, x_n^{\downarrow})$ is obtained from x by rearranging all coordinates in weakly decreasing order. The vector $x^{\uparrow}$ can be similarly defined.

Suppose $x, y \in \mathbb{R}^n$. We say x is *weakly majorized* by y and write $x \prec_w y$ if

$$\sum_{j=1}^{k} x_j^{\downarrow} \leq \sum_{j=1}^{k} y_j^{\downarrow}, \; \forall k = 1, \dots, n. \tag{17}$$

Additionally, if

$$\sum_{j=1}^{n} x_j^{\downarrow} = \sum_{j=1}^{n} y_j^{\downarrow}, \tag{18}$$

then x is said to be *majorized* by y and we write $x \prec y$.

The concept of majorization is very important in the theory of inequalities, as well as in linear algebra. We develop here a few essential properties of majorization.

Theorem 10.1. *Let $A \in M_n$ be Hermitian. Let a be the vector of diagonal entries of A, and $\alpha = \lambda(A)$ the vector of all eigenvalues of A. Then, $a \prec \alpha$.*

Proof. When $n = 1$, there is nothing to show. In general, let $B \in M_{n-1}$ be a Hermitian matrix obtained from A by removing the row and the column corresponding to a smallest diagonal entry of A. Let $\beta_1, \dots, \beta_{n-1}$ be the eigenvalues of B. Then, $\alpha_1 \geq \beta_1 \geq \cdots \geq \beta_{n-1} \geq \alpha_n$. Moreover, induction hypothesis yields

$$\sum_{i=1}^{k} a_i^{\downarrow} \leq \sum_{i=1}^{k} \beta_i, \;\; 1 \leq k \leq n-1.$$

Hence,

$$\sum_{i=1}^{k} a_i^{\downarrow} \leq \sum_{i=1}^{k} \alpha_i, \;\; 1 \leq k \leq n-1.$$

Lastly, $tr(A) = \sum a_i = \sum \alpha_i$ finishes the proof. □

It turns out of the converse also holds. Before showing the converse, we need a technical lemma.

Lemma 10.2. *Let $x, y \in \mathbb{R}^n$ such that $x \succ y$. Then, there exists a vector $z \in \mathbb{R}^{n-1}$ such that*

$$x_1^{\downarrow} \geq z_1 \geq x_2^{\downarrow} \geq z_2 \cdots \geq z_{n-1} \geq x_n^{\downarrow},$$

and $z \succ [y_1^{\downarrow}, \dots, y_{n-1}^{\downarrow}]^T$.

Proof. When $n = 2$, we must have $x_1^{\downarrow} \geq y_1^{\downarrow} \geq y_2^{\downarrow} \geq x_2^{\downarrow}$. Hence, picking $z_1 = y_1^{\downarrow}$ suffices.

Suppose $n \geq 3$. Let $D \subseteq \mathbb{R}^{n-1}$ be defined by

$$D = \left\{ v \in \mathbb{R}^{n-1} \mid x_1^{\downarrow} \geq v_1 \geq \cdots \geq v_{n-1} \geq x_n^{\downarrow}, \text{ and } \sum_{i=1}^{k} v_i \geq \sum_{i=1}^{k} y_k^{\downarrow}, \; 1 \leq k \leq n-2 \right\}.$$

Then, the existence of a point $z \in D$ for which $\sum_{i=1}^{n-1} z_i = \sum_{i=1}^{n-1} y_i^{\downarrow} = c$ would complete the proof.

Notice that as $[x_1^{\downarrow}, \ldots, x_{n-1}^{\downarrow}]^T \in D$, D is not empty. Define a continuous function $f : D \to \mathbb{R}$ by $f(v) = v_1 + \cdots + v_{n-1}$. Then, $f([x_1^{\downarrow}, \ldots, x_{n-1}^{\downarrow}]^T) \geq c$. Since D is a connected domain, if we could find $v \in D$ for which $f(v) \leq c$, then there must exist the vector z for which $f(z) = c$. Let $\hat{v} \in D$ be a vector such that $\hat{v} = \min\{f(z) \mid z \in D\}$. If $f(\hat{v}) \leq c$, then we are done. Suppose $f(\hat{v}) > c$, we shall show that $f(\hat{v}) \geq c$ to reach a contradiction. We have

$$\sum_{i=1}^{k} \hat{v}_i \geq \sum_{i=1}^{k} y_k^{\downarrow}, \ 1 \leq k \leq n-1 \tag{19}$$

$$\hat{v}_k \geq x_{k+1}^{\downarrow}, \ 1 \leq k \leq n-1. \tag{20}$$

Suppose first that all inequalities (19) are strict. Then, it must be the case that $\hat{v}_k = x_{k+1}^{\downarrow}, \ k = 1, \ldots, n-1$. Otherwise, one could reduce some $\hat{v}_k$ to make $f(\hat{v}_k)$ smaller. Consequently, $f(\hat{v}) = f([x_2^{\downarrow}, \ldots, x_n^{\downarrow}]^T) \leq c$.

If not all of the inequalities (19) are strict, then let r be the largest index for which

$$\sum_{i=1}^{r} \hat{v}_i = \sum_{i=1}^{r} y_i^{\downarrow} \tag{21}$$

$$\sum_{i=1}^{k} \hat{v}_i > \sum_{i=1}^{k} y_i^{\downarrow}, \ \ r < k \leq n-1 \tag{22}$$

(Notice the fact that $r \leq n-2$, since we assumed $f(\hat{v}) > c$.) By the same reasoning as before, we must have $\hat{v}_k = x_{k+1}^{\downarrow}$ for $k = r+1, \ldots, n-1$. Thus,

$$\begin{aligned}
f(\hat{v}) - c &= \sum_{i=1}^{n-1} \hat{v}_i - \sum_{i=1}^{n-1} y_i^{\downarrow} \\
&= \sum_{i=1}^{r} \hat{v}_i + \sum_{i=r+1}^{n-1} \hat{v}_i - \sum_{i=1}^{n-1} y_i^{\downarrow} \\
&= \sum_{i=1}^{r} y_i^{\downarrow} + \sum_{i=r+2}^{n} x_i^{\downarrow} - \sum_{i=1}^{n-1} y_i^{\downarrow} \\
&\leq \sum_{i=1}^{r} y_i^{\downarrow} + + \sum_{i=r+2}^{n} y_i^{\downarrow} - \sum_{i=1}^{n-1} y_i^{\downarrow} \\
&= \sum_{i=r+2}^{n} y_i^{\downarrow} - \sum_{i=r+1}^{n-1} y_i^{\downarrow} \\
&= \sum_{i=r+2}^{n} \left(y_i^{\downarrow} - y_{i-1}^{\downarrow} \right) \\
&\leq 0
\end{aligned}$$

□

We are now ready to show the converse of Theorem 10.1.

Theorem 10.3. *Let a and α be two vectors in $\mathbb{R}^n$. If $a \prec \alpha$, then there exists a real symmetric matrix $A \in M_n$ which has diagonal entries a_i, and $\lambda(A) = \alpha$.*

Proof. The case $n = 1$ is trivial. In general, assume without loss of generality that $a = a^{\downarrow}$, and $\alpha = \alpha^{\downarrow}$. Also, let $b = [a_1, \ldots, a_{n-1}]$. Then, Lemma 10.2 implies the existence of a vector $\beta \in \mathbb{R}^{n-1}$ such that

$$\alpha_1 \geq \beta_1 \geq \cdots \geq \beta_{n-1} \geq \alpha_n,$$

and that $\beta \succ b$. The induction hypothesis ensures the existence of a real symmetric matrix B which has diagonal entries b and eigenvalues β. Now, Theorem 8.2 allows us to extend B into a real symmetric matrix $A' \in M_n$:

$$A' = \begin{bmatrix} \Lambda & y \\ y^T & b \end{bmatrix},$$

where $\Lambda = diag(\beta_1, \ldots, \beta_{n-1})$, and A' has eigenvalues α. One more step needs to be done to turn A' into matrix A we are looking for. We know that there exists a orthonormal matrix $Q \in M_{n-1}$ for which $B = Q\Lambda Q^T$. Hence, letting

$$A = \begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Lambda & y \\ y^T & b \end{bmatrix} \begin{bmatrix} Q^T & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} Q\Lambda Q^T & Qy \\ (Qy)^T & a \end{bmatrix} = \begin{bmatrix} B & Qy \\ (Qy)^T & a \end{bmatrix}$$

finishes the proof. □

For any $\pi \in S_n$ and $y \in \mathbb{R}^n$, let $y_\pi := (y_{\pi(1)}, \ldots, y_{\pi(n)})$.

Theorem 10.4. *Given vectors $x, y \in \mathbb{R}^n$, the following three statements are equivalent.*

(i) $x \succ y$.

(ii) There exists a doubly stochastic matrix M for which $x = My$.

(iii) x is in the convex hull of all $n!$ points y_π, $\pi \in S_n$.

Proof. We shall show $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)$.

Firstly, assume $x \prec y$. By Theorem 10.3 there is a Hermitian matrix $A \in M_n$ with diagonal entries y and $\lambda(A) = x$. There thus must exist a unitary matrix $U = (u_{ij})$ for which $A = U diag(y_1, \ldots, y_n) U^*$, which implies

$$x_i = a_{ii} = \sum_{j=1}^{n} y_j |u_{ij}|^2.$$

Hence, taking $M = (|u_{ij}|^2)$ completes the proof.

Secondly, suppose $x = My$ where M is a doubly stochastic matrix. Birkhoff Theorem implies that there are non-negative real numbers $c_\pi, \pi \in S_n$ such that

$$\sum_{\pi \in S_n} c_\pi = 1$$
$$M = \sum_{\pi \in S_n} c_\pi P_\pi,$$

where P_π is the permutation matrix corresponding to π. Consequently,

$$x = My = \sum_{\pi \in S_n} c_\pi P_\pi y = \sum_{\pi \in S_n} c_\pi y_\pi.$$

Lastly, suppose there are non-negative real numbers $c_\pi, \pi \in S_n$ such that

$$\sum_{\pi \in S_n} c_\pi = 1$$
$$x = \sum_{\pi \in S_n} c_\pi y_\pi.$$

Without loss of generality, we assume $y = y^{\downarrow}$. We can write x_i is the following form:

$$x_i = \sum_{j=1}^{n} y_j \left(\sum_{\substack{\pi \in S_n \\ \pi(i)=j}} c_\pi \right) = \sum_{j=1}^{n} y_j d_{ij}.$$

Note that $\sum_i d_{ij} = \sum_j d_{ij} = 1$. (This is rather like having $(iii) \Rightarrow (ii)$ first, and then we show $(ii) \Rightarrow (i)$.) The following is straightforward:

$$\begin{aligned}
\sum_{i=1}^{k} x_i &= \sum_{i=1}^{k}\sum_{j=1}^{n} y_j d_{ij} \\
&= \sum_{j=1}^{n} y_j \sum_{i=1}^{k} d_{ij} \\
&\leq y_1 \sum_{i=1}^{k} d_{i1} + \cdots + y_k \sum_{i=1}^{k} d_{ik} + y_{k+1}\left(\sum_{j=k+1}^{n}\sum_{i=1}^{k} d_{ij} \right) \\
&= y_1 + \cdots + y_k - y_1\left(1 - \sum_{i=1}^{k} d_{i1}\right) - \cdots - y_k\left(1 - \sum_{i=1}^{k} d_{ik}\right) + y_{k+1}\left(\sum_{j=k+1}^{n}\sum_{i=1}^{k} d_{ij} \right) \\
&\leq y_1 + \cdots + y_k - y_{k+1}\left(k - \sum_{j=1}^{k}\sum_{i=1}^{k} d_{ij}\right) + y_{k+1}\left(\sum_{j=k+1}^{n}\sum_{i=1}^{k} d_{ij} \right) \\
&= y_1 + \cdots + y_k - y_{k+1}\left(k - \sum_{j=1}^{n}\sum_{i=1}^{k} d_{ij}\right) \\
&= y_1 + \cdots + y_k - y_{k+1}\left(k - \sum_{i=1}^{k}\sum_{j=1}^{n} d_{ij}\right) \\
&= y_1 + \cdots + y_k.
\end{aligned}$$

□

11 Two Examples of Linear Algebra in Combinatorics

11.1 The statements

We examine two elegant theorems which illustrate beautifully the inter-relations between Combinatorics, Algebra, and Graph Theory. These two theorems are presented not only for the purpose of demonstrating the relationships, but they will also be used to develop some of our later materials on Algebraic Graph Theory.

Theorem 11.1 (Cayley-Hamilton). *Let A be an $n \times n$ matrix over any field. Let $p_A(x) := \det(xI - A)$ be the characteristic polynomial of A. Then $p_A(A) = 0$.*

I will give a proof of this theorem combinatorially, following the presentation in [7]. A typical algebraic proof of this theorem would first shows that a weak version where A is diagonal holds, then extend to all matrices over $\mathbb{C}$. To show the most general version we stated, the Fundamental Theorem of Algebra is used. (FTA says $\mathbb{C}$ is algebraically closed, or any $p \in \mathbb{C}[x]$ has roots in $\mathbb{C}$).

Theorem 11.2 (Matrix-Tree). *Let G be a labeled graph on $[n] := \{1, \ldots, n\}$. Let A be the adjacency matrix of G and $d_i := deg(i)$ be the degree of vertex i. Then the number of spanning trees of G is any cofactor of L, where $L = D - A$, D is diagonal with diagonal entries $d_{ii} = d_i$,*

The matrix L is often referred to as the *Laplacian* of G. A cofactor of a square matrix L is $(-1)^{i+j} \det L_{ij}$ where L_{ij} is the matrix obtained by crossing off row i and column j of L. This theorem also has a beautiful combinatorial proof. See [7] for details. I will present the typical proof of this theorem which uses the Cauchy-Binet theorem on matrix expansion. This proof is also very elegant and helps us develope a bit of linear algebra. Actually, for weighted graphs, a minimum spanning tree can be shown to be a tree which minimizes certain determinant.

11.2 The proofs

Combinatorial proof of Cayley-Hamilton Theorem. (by Straubing 1983 [9]).

$$p_A(x) := \det(xI - A) := \sum_{\pi \in S_n} sgn(\pi) \prod_{i=1}^{n} (xI - A)_{i\pi(i)}$$

Let the set fixed points of a permutation π be denoted by $fp(\pi) := \{i \in [n] \mid \pi(i) = i\}$. Each $i \in fp(\pi)$ contributes either x or $-a_{ii}$ to a term. Each $i \notin fp(\pi)$ contributes $-a_{i\pi(i)}$. Hence, thinking of F as the set of fixed points contributing x, we get

$$\begin{aligned} p_A(x) &= \sum_{\pi \in S_n} sgn(\pi) \sum_{F \subseteq fp(\pi)} (-1)^{n-|F|} x^{|F|} \prod_{i \notin F} a_{i\pi(i)} \\ &= \sum_{\pi \in S_n} sgn(\pi) \sum_{\substack{S \subseteq [n], \\ [n]-S \subseteq fp(\pi)}} (-1)^{|S|} x^{n-|S|} \prod_{i \in S} a_{i\pi(i)}. \end{aligned}$$

Now we exchange the summation indices by first fixing a particular choice of S. The π will be the ones with $[n] - S \subseteq fp(\pi)$, i.e. the permutations which fix everything not in S. Let $P(S)$ be the set of permutations on S, then

$$p_A(x) = \sum_{k=0}^{n} x^{n-k} \sum_{S \in \binom{[n]}{k}} \sum_{\pi \in P(S)} sgn(\pi)(-1)^k \prod_{i \in S} a_{i\pi(i)}.$$

Let $c(\pi)$ be the number of cycles of π, it is easy to see that for $\pi \in P(S)$ with $|S| = k$, $sgn(\pi)(-1)^k = (-1)^{c(\pi)}$. Thus,

$$p_A(x) = \sum_{k=0}^{n} x^{n-k} \sum_{S \in \binom{[n]}{k}} \sum_{\pi \in P(S)} (-1)^{c(\pi)} \prod_{i \in S} a_{i\pi(i)}$$

Our objective is to show $p_A(A) = 0$. We'll do so by showing $(p_A(A))_{ij} = 0$, $\forall i, j \in [n]$. Firstly,

$$(p_A(A))_{ij} = \sum_{k=0}^{n} (A^{n-k})_{ij} \sum_{S \in \binom{[n]}{k}} \sum_{\pi \in P(S)} (-1)^{c(\pi)} \prod_{l \in S} a_{l\pi(l)}$$

Let $\mathcal{P}_{ij}^k$ be the set of all directed walks of length k from i to j in K_n - the complete directed graph on n vertices. Let an edge $e = (i, j) \in E(K_n)$ be weighted by $w(e) = a_{ij}$. For any $P \in \mathcal{P}_{ij}^k$, let $w(P) = \prod_{e \in P} w(e)$. It follows that

$$(A^{n-k})_{ij} = \sum_{P \in \mathcal{P}_{ij}^{n-k}} w(P)$$

To this end, let (S, π, P) be a triple satisfying (a) $S \subseteq [n]$; (b) $\pi \in P(S)$; and (c) $P \in \mathcal{P}_{ij}^{n-|S|}$. Define $w(S, \pi, P) := w(P)w(\pi)$, where $w(\pi) = \prod_{t \in S} a_{t\pi(t)}$. Let $sgn(S, \pi, P) := (-1)^{c(\pi)}$, then

$$(p_A(A))_{ij} = \sum_{(S,\pi,P)} w(S, \pi, P) sgn(S, \pi, P)$$

To show $(p_A(A))_{ij} = 0$, we seek a sign-reversing, weight-preserving involution ϕ on the set of triples (S, π, P). Let v be the first vertex in P along the walk such that either (i) $v \in S$, or (ii) v completes a cycle in P. Clearly,

- (i) and (ii) are mutually exclusive, since if v completes a cycle in P and $v \in S$ then v was in S before completing the cycle.
- One of (i) and (ii) must hold, since if no v satisfy (i) then P induces a graph on $n - |S|$ vertices with $n - |S|$ edges. P must have a cycle.

Lastly, given the observations above we can describe ϕ as follows. Take the first $v \in [n]$ satisfying (i) or (ii). If $v \in S$ then let C be the cycle of π containing v. Let P' be P with C added right after v. $S' = S - C$ and π' be π with the cycle C removed. The image of $\phi(S, \pi, P)$ is then (S', π', P'). Case (ii) v completes a cycle in P before touching S is treated in the exact opposite fashion, i.e. we add the cycle into π, and remove it from P. □

To prove the Matrix-Tree Theorem, we first need to show a sequence of lemmas. The first (Cauchy-Binet Theorem) is commonly stated with $D = I$.

Lemma 11.3 (Cauchy-Binet Theorem). *Let A and B be, respectively, $r \times m$ and $m \times r$ matrices. Let D be an $m \times m$ diagonal matrix with diagonal entries e_i, $i \in [m]$. For any r-subset S of $[m]$, let A_S and B^S denote, respectively, the $r \times r$ submatrices of A and B consisting of the columns of A, or the rows of B, indexed by S. Then*

$$\det(ADB) = \sum_{S \in \binom{[m]}{r}} \det A_S \det B^S \prod_{i \in S} e_i.$$

Proof. We will prove this assuming that $e_1, \ldots, e_m$ are indeterminates. With this assumption in mind, since $(ADB)_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj} e_k$, it is easy to see that $\det(ADB)$ is a homogeneous polynomial in $e_1, \ldots, e_m$ with degree r.

Consider a monomial $e_1^{t_1} e_2^{t_2} \ldots e_m^{t_m}$, where the number of *distinct* variables that occur is $< r$, i.e. $|\{i \mid t_i > 0\}| < r$. Substitute 0 for all other indeterminates then $e_1^{t_1} e_2^{t_2} \ldots e_m^{t_m}$ and its coefficient are unchanged. But, after this substitution, $rank(D) < r$, which implies $rank(ADB) < r$, making $\det(ADB) = 0$. So the coefficient of our monomial is 0.

Put it another way, the coefficient of a monomial $e_1^{t_1} \ldots e_m^{t_m}$ is 0 unless it is a product of r distinct indeterminates, i.e. $\exists S \in \binom{[m]}{r}$ s.t. $e_1^{t_1} \ldots e_m^{t_m} = \prod_{i \in S} e_i$.

The coefficient of $\prod_{i \in S} e_i$ can be calculated by setting $e_i = 1$ for all $i \in S$ and $e_j = 0$ for all $j \notin S$. It is not hard to see that the coefficient is $\det A_S \det B^S$. □

Lemma 11.4. *Given a directed graph H with incident matrix N. Let $C(H)$ be the set of connected component of H, then*

$$rank(N) = |V(H)| - |C(H)|$$

Proof. Recall that N is defined to be a matrix whose rows are indexed by $V(H)$, whose columns are indexed by $E(H)$, and

$$N_{i,e} = \begin{cases} 0 & \text{if } i \text{ is not incident to } e \text{ or } e \text{ is a loop} \\ 1 & \text{if } e = j \to i, j \neq i \\ -1 & \text{if } e = i \to j, j \neq i \end{cases}$$

To show $rank(N) = |V(H)| - |C(H)|$ we only need to show that $dim(col(N)^{\perp}) = |C(H)|$. For any row vector $g \in \mathbb{R}^{|V(H)|}$, $g \in col(N)^{\perp}$ iff $gN = 0$, i.e. for any edge $e = x \to y \in E(H)$ we must have $g(x) = g(y)$. Consequently, $g \in col(N)^{\perp}$ iff g is constant on the coordinates corresponding to any connected component of H. It is thus clear that $dim(col(N)^{\perp}) = |C(H)|$. □

Lemma 11.5 (Poincaré, 1901). *Let M be a square matrix with at most two non-zero entries in each column, at most one 1 and at most one -1, then $\det M = 0, \pm 1$.*

Proof. This can be done easily by induction. If every column has exactly a 1 and a -1, then the sum of all row vectors of M is $\vec{0}$, making $\det M = 0$. Otherwise, expand the determinant of M along the column with at most one ± 1 and use the induction hypothesis. □

Proof the Matrix-Tree Theorem. We will first show that the Theorem holds for the ii-cofactors for all $i \in [n]$. Then, we shall show that the ij-cofactors are all equal for all $j \in [n]$, which completes the proof. We can safely assume $m \geq n - 1$, since otherwise there is no spanning tree and at the same time $\det(NN^T) = 0$.

Step 1. If G' is any orientation of G, and N is the incident matrix of G', then $L = NN^T$. (Recall that L is the Laplacian of G.) For any $i \neq j \in [n]$, if i is adjacent to j then clearly $(NN^T)_{ij} = -1$. On the other hand, $(NN^T)_{ii}$ is obviously the number of edges incident to i.

Step 2. If B is an $(n-1) \times (n-1)$ submatrix of N, then $\det B = 0$ if the corresponding $n - 1$ edges contain a cycle, and $\det B = \pm 1$ if they form a spanning tree of G. Clearly, B is obtained by removing a row of N_S for some $(n-1)$-subset S of $E(H)$. By Lemma 11.4, $rank(N_S) = n - 1$ iff the edges corresponding to S form a spanning tree. Moreover, since the sum of all rows of N_S is the 0-vector, $rank(B) = rank(N_S)$. Hence, $\det B \neq 0$ iff S form a spanning tree. When S does not form a spanning tree, Lemma 11.5 implies $\det B = \pm 1$.

Step 3. Calculating $\det L_{ii}$, i.e. the ii-cofactor of L. Let $m = |E(G)|$. Let M be the matrix obtained from N by deleting row i of N, then $L_{ii} = MM^T$. Applying Cauchy-Binet theorem with $e_i = 1$, $\forall i$, we get

$$\begin{aligned} \det(MM^T) &= \sum_{S \in \binom{[m]}{n-1}} \det M_S \det(M^T)^S \\ &= \sum_{S \in \binom{[m]}{n-1}} (\det M_S)^2 \\ &= \text{\# of spanning trees of } G \end{aligned}$$

The following Lemma is my solution to exercise 2.2.18 in [10]. The Lemma completes the proof because L is a matrix whose columns sum to the 0-vector. □

Lemma 11.6. *Given an $n \times n$ matrix $A = (a_{ij})$ whose columns sum to the 0-vector. Let $b_{ij} = (-1)^{i+j} \det A_{ij}$, then for a fixed i, we have $b_{ij} = b_{ij'}$, $\forall j, j'$.*

Proof. Let $B = (b_{ij})^T = (b_{ji})$, then

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{jk}$$

Obviously, $(AB)_{ij} = \delta_{ij} \det A$ where δ_{ij} is the Kronecker delta. To see this, imagine replacing row j of A by row i of A and expand $\det A$ along row j, we get exactly the expression above. In other words, $AB = (\det A)I$.

Let $\vec{a_i}$ denote column i of A, then by assumption $\sum_i \vec{a_i} = \vec{0}$. Hence, $\det A = 0$ and $dim(col(A)) \leq n-1$. If $dim(col(A)) < n-1$ then $rank(A_{ij}) < n-1$, making $b_{ij} = 0$. Otherwise, if $dim(col(A)) = n-1$ then $n-1$ vectors $\vec{a_j} - \vec{a_1}$, $2 \leq j \leq n$ are linearly independent. Moreover, $AB = (\det A)I = 0$ and $\sum_i \vec{a_i} = \vec{0}$ implies that for all i

$$(b_{i2} - b_{i1})(\vec{a_2} - \vec{a_1}) + (b_{i3} - b_{i1})(\vec{a_3} - \vec{a_1}) + \dots (b_{in} - b_{i1})(\vec{a_n} - \vec{a_1}) = \vec{0}$$

So, $b_{ij} - b_{i1} = 0$, $\forall j \geq 2$. □

Corollary 11.7 (Cayley Formula). *The number of labeled trees on* $[n]$ *is* n^{n-2}.

Proof. Cayley formula is usually proved by using Prufer correspondence. Here I use the Matrix-Tree theorem to give us a different proof. Clearly the number of labeled trees on $[n]$ is the number of spanning trees of K_n. Hence, by the Matrix-Tree theorem, it is $\det(nI - J)$ where J is the all 1's matrix, and I and J are matrices of order $n-1$ (we are taking the 11-cofactor).

$$\begin{aligned}
\det(nI - J) &= \det \begin{bmatrix} n-1 & -1 & -1 & \dots & -1 \\ -1 & n-1 & -1 & \dots & -1 \\ -1 & -1 & n-1 & \dots & -1 \\ \dots & \dots & \dots & \dots & \dots \\ -1 & -1 & -1 & \dots & n-1 \end{bmatrix} \\
&= \det \begin{bmatrix} n-1 & -1 & -1 & \dots & -1 \\ 0 & \frac{n(n-2)}{n-1} & \frac{-n}{n-1} & \dots & \frac{-n}{n-1} \\ 0 & \frac{-n}{n-1} & \frac{n(n-2)}{n-1} & \dots & \frac{-n}{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \frac{-n}{n-1} & \frac{-n}{n-1} & \dots & \frac{n(n-2)}{n-1} \end{bmatrix} \\
&= \det \begin{bmatrix} n-1 & -1 & -1 & \dots & -1 \\ 0 & \frac{n(n-2)}{n-1} & \frac{-n}{n-1} & \dots & \frac{-n}{n-1} \\ 0 & 0 & \frac{n(n-3)}{n-2} & \dots & \frac{-n}{n-2} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \frac{-n}{n-2} & \dots & \frac{n(n-3)}{n-2} \end{bmatrix} \\
&= \dots \\
&= \det \begin{bmatrix} n-1 & -1 & -1 & \dots & -1 \\ 0 & \frac{n(n-2)}{n-1} & \frac{-n}{n-1} & \dots & \frac{-n}{n-1} \\ 0 & 0 & \frac{n(n-3)}{n-2} & \dots & \frac{-n}{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{n(n-(n-1))}{n-(n-2)} \end{bmatrix} \\
&= n^{n-2}
\end{aligned}$$

□

References

[1] M. ARTIN, *Algebra*, Prentice-Hall Inc., Englewood Cliffs, NJ, 1991.

[2] R. BHATIA, *Linear algebra to quantum cohomology: the story of Alfred Horn's inequalities*, Amer. Math. Monthly, 108 (2001), pp. 289–318.

[3] R. A. BRUALDI, *The Jordan canonical form: an old proof*, Amer. Math. Monthly, 94 (1987), pp. 257–267.

[4] F. R. GANTMACHER, *The theory of matrices. Vol. 1*, AMS Chelsea Publishing, Providence, RI, 1998. Translated from the Russian by K. A. Hirsch, Reprint of the 1959 translation.

[5] R. A. HORN AND C. R. JOHNSON, *Matrix analysis*, Cambridge University Press, Cambridge, 1985.

[6] P. LANCASTER AND M. TISMENETSKY, *The theory of matrices*, Academic Press Inc., Orlando, Fla., second ed., 1985.

[7] D. STANTON AND D. WHITE, *Constructive combinatorics*, Springer-Verlag, New York, 1986.

[8] G. STRANG, *Linear algebra and its applications*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, second ed., 1980.

[9] H. STRAUBING, *A combinatorial proof of theCayley-Hamilton theorem*, Discrete Math., 43 (1983), pp. 273–279.

[10] D. B. WEST, *Introduction to graph theory*, Prentice Hall Inc., Upper Saddle River, NJ, 1996.

[11] H. WEYL, *Das asymptotische verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen*, Math. Ann., 71 (1912), pp. 441–479.