CSE 594: Combinatorial and Graph Algorithms
SUNY at Buffalo, Spring 2004

Lecturer: Hung Q. Ngo
Scribe: Hung Q. Ngo

Linear Algebra and Examples

Standard texts on Linear Algebra and Algebra are [1, 6].

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

$$\begin{aligned} (AB)^T &= B^T A^T && (4) \\ (AB)^* &= B^* A^* && (5) \end{aligned}$$

(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(π) $= (-1)^{I(\pi)}$.

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(π) $= -1$ 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:*

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

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 [6]. Another proof is presented in [2], 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 [4] or [3], 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.*

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

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

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

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

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

□

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 5.1), $p_A(A) = 0$.

(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

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

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

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

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 Two Motivating Theorems

5.1 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 5.1 (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 [5]. 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 5.2 (Matrix-Tree). *Let G be a labeled graph on $[n] := \{1, \dots, n\}$. 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 [5] 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.

5.2 The proofs

Combinatorial proof of Cayley-Hamilton Theorem. (by Straubing 1983 [7]).

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

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 5.3 (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 $e_1^{t_1} e_2^{t_2} \dots e_m^{t_m}$, 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 $e_1^{t_1} e_2^{t_2} \dots e_m^{t_m}$ 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 $e_1^{t_1} \dots e_m^{t_m}$ is 0 unless it is a product of r distinct indeterminates, i.e. $\exists S \in \binom{[m]}{r}$ s.t. $e_1^{t_1} \dots e_m^{t_m} = \prod_{i \in S} e_i$.

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 5.4. *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 5.5 (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 5.4, $rank(N_S) = n - 1$ 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 5.5 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 [8]. The Lemma completes the proof because L is a matrix whose columns sum to the 0-vector. □

Lemma 5.6. *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

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

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

Corollary 5.7 (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).

$$
\begin{aligned}
\det(nI-J) &= \det\begin{bmatrix} n-1 & -1 & -1 & \dots & -1 \\ -1 & n-1 & -1 & \dots & -1 \\ -1 & -1 & n-1 & \dots & -1 \\ \hdotsfor{5} \\ -1 & -1 & -1 & \dots & n-1 \end{bmatrix} \\
&= \det\begin{bmatrix} n-1 & -1 & -1 & \dots & -1 \\ 0 & \frac{n(n-2)}{n-1} & \frac{-n}{n-1} & \dots & \frac{-n}{n-1} \\ 0 & \frac{-n}{n-1} & \frac{n(n-2)}{n-1} & \dots & \frac{-n}{n-1} \\ \hdotsfor{5} \\ 0 & \frac{-n}{n-1} & \frac{-n}{n-1} & \dots & \frac{n(n-2)}{n-1} \end{bmatrix} \\
&= \det\begin{bmatrix} n-1 & -1 & -1 & \dots & -1 \\ 0 & \frac{n(n-2)}{n-1} & \frac{-n}{n-1} & \dots & \frac{-n}{n-1} \\ 0 & 0 & \frac{n(n-3)}{n-2} & \dots & \frac{-n}{n-2} \\ \hdotsfor{5} \\ 0 & 0 & \frac{-n}{n-2} & \dots & \frac{n(n-3)}{n-2} \end{bmatrix} \\
&= \dots \\
&= \det\begin{bmatrix} n-1 & -1 & -1 & \dots & -1 \\ 0 & \frac{n(n-2)}{n-1} & \frac{-n}{n-1} & \dots & \frac{-n}{n-1} \\ 0 & 0 & \frac{n(n-3)}{n-2} & \dots & \frac{-n}{n-1} \\ \hdotsfor{5} \\ 0 & 0 & 0 & \dots & \frac{n(n-(n-1))}{n-(n-2)} \end{bmatrix} \\
&= n^{n-2}
\end{aligned}
$$

□

References

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

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

[3] 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.

[4] P. Lancaster and M. Tismenetsky, *The theory of matrices*, Academic Press Inc., Orlando, Fla., second ed., 1985.

[5] D. Stanton and D. White, *Constructive combinatorics*, Springer-Verlag, New York, 1986.

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

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

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