Bachelors with Mathematics as Major
3rd Semester
MMT322J: Mathematics/Applied Mathematics: Theory of Matrices

Credits: 4 THEORY + 2 TUTORIAL **Theory: 60 Hours & Tutorial: 30 Hours**

Course Objectives: (i) To understand matrix theory as a tool to solve various real life problems.
(ii) To make students aware about the properties and applications of matrices.

Course Outcome: After the completion of this course, students shall be able to (i) apply techniques of matrix theory to solve real life problems (ii) use matrix techniques in coding theory and cryptography (iii) use eigenvalues to find the stability of various systems.

Theory: 4 Credits

Unit –I

Generalization of reversal law of transpose, Hermitian and skew- Hermitian matrices, Representation of a square matrix as P + iQ, where P and Q are both Hermition. Adjoint of a matrix, For a square matrix A, A(adjA) = (adjA)A=|A|I. Commutative and associative laws in matrix operations. Necessary and sufficient condition for a square matrix to be invertible. Generalization of reversal law for the inverse of matrices under multiplication.

Unit –II

The operation of transposing and inverting are commutative, trace of a matrix, trace of AB= trace of BA and its generalization. Partitioning of matrices, Matrix polynomials and Characteristic equation of a square matrix. Cayley-Hamilton theorem, Eigen values and Eigen vector, minimal equation of a matrix.

Unit –III

Rank of a matrix. Elementary row (Column) transformations of a matrix do not alter its rank, rank of a matrix by elementary transformations, reduction of a matrix to the normal form, Elementary matrices. Every non- singular matrix is a product of elementary matrices, employment of only row (column) transformations. Rank of product of two matrices. Linear combination, linear dependence and linear independence of Row (Column) vectors, the columns of a matrix A are linearly dependent iff there exists a vector $X \neq 0$ such that AX=0. The columns of a matrix A of order $m \times n$ are linearly dependent iff rank of A< n. The matrix A has rank r if and only if it has r linearly independent columns and any s-columns (s > r) are linearly dependent (analogous results for rows).

Unit –IV

Linear homogeneous and non- homogeneous equations, the equation AX=0 has a non-zero solution if and only if rank of A < n, the number of its columns, the number of linearly independent solutions of the equation AX=0 is n-r, where r is the rank of matrix A of order $m \times n$, the equation AX=B is consistent if and only if two matrices A and [A: B] are of the same rank. Inner product of two vectors, length of a vector, normal vectors, Orthogonal and Unitary matrices, A matrix P is orthogonal (Unitary) if and only it its column vectors are normal and orthogonal in pairs.

Tutorial: 2 Credits

Unit –V

Problems based on Hermition and skew-Hermitian and inverse of matrices. Problems on characteristic roots and characteristic polynomials. Applications of Cayley Hamilton theorem for the inverse of a matrix.

Unit –VI

System of equations and their solutions, examples of determination of orthogonal matrices and examples of system of homogenous and non-homogenous equation having unique, infinite and no solution.

Books Recommended:

1. Shanti Narayan, A textbook of Matrices, Schaum S. Chand and Company, 1957.
2. A. Aziz and NA Rather and BA Zargar, Elementary Matrix theory, Kapoor Book Depot, Srinagar, 2007.
3. K. Hoffman and R. Kunze, Linear Algebra, Pearson Education, 2018.
4. S. Lipschutz & M. Lipson, Linear Algebra, Schaum"s outline series, Tata McGraw-Hill, 4th Edition 2009.

——THEORY OF MATRICES - (MMT322J)—–

Author: DR. AIJAZ

ASSISTANT PROFESSOR OF MATHEMATICS

JK HIGHER EDUCATION DEPARTMENT

All stuyd materials from UG Mathematics, PG Mathematics, NET-JRF/SET/GATE and other enntrance exams are avalaible at

WEBSITE: AIJAZ.VERCEL.APP

Chapter 1

Generalizations of Operatoins on Matrices

1.1 Special Matrix Properties and Decompositions

1.1.1 Generalization of the Reversal Law of Transpose

Theorem 1.1.1 (Reversal Law for Two Matrices)**.** *If the product AB is defined, then $(AB)^T = B^T A^T$.*

Proof. To prove this, we compare the (i, j)-th element of each side. The (i, j)-th element of $(AB)^T$ is the (j, i)-th element of AB, which is $\sum_k a_{jk} b_{ki}$. The (i, j)-th element of $B^T A^T$ is the dot product of the i-th row of B^T and the j-th column of A^T. The i-th row of B^T is the i-th column of B, and the j-th column of A^T is the j-th row of A. This gives the product $\sum_k (B^T)_{ik}(A^T)_{kj} = \sum_k b_{ki} a_{jk}$. Since the resulting sums are identical, the theorem holds. ■

This rule can be generalized to the product of any finite number of matrices.

Theorem 1.1.2 (Generalization of Reversal Law)**.** *If the product $A_1 A_2 \cdots A_n$ is well-defined, then:*

$$(A_1 A_2 \cdots A_n)^T = A_n^T \cdots A_2^T A_1^T$$

Proof. We will prove this theorem by mathematical induction on n, the number of matrices.
For $n = 2$, the theorem states $(A_1 A_2)^T = A_2^T A_1^T$. This is the standard reversal law for two matrices, which we take as established.
Assume the formula is true for a product of k matrices, where $k \geq 2$. That is, we assume:

$$(A_1 A_2 \cdots A_k)^T = A_k^T \cdots A_2^T A_1^T$$

We must now prove that the formula holds for a product of $k+1$ matrices. We can cleverly group the matrices as follows:

$$(A_1 A_2 \cdots A_k A_{k+1})^T = ((A_1 A_2 \cdots A_k) A_{k+1})^T$$

Let the product of the first k matrices be a single matrix, $B = (A_1 A_2 \cdots A_k)$. The expression becomes $(BA_{k+1})^T$. Applying the base case for two matrices gives:

$$(BA_{k+1})^T = A_{k+1}^T B^T$$

Now, we substitute back the expression for B:

$$A_{k+1}^T B^T = A_{k+1}^T (A_1 A_2 \cdots A_k)^T$$

By our inductive hypothesis, we can replace $(A_1 A_2 \cdots A_k)^T$ with its expanded form:

$$A_{k+1}^T (A_k^T \cdots A_2^T A_1^T) = A_{k+1}^T A_k^T \cdots A_2^T A_1^T$$

Thus, we have shown that if the law holds for k matrices, it also holds for $k+1$ matrices. By the principle of mathematical induction, the theorem holds for all integers $n \geq 2$. ■

Example 1.1.3 (Generalization of Reversal Law of Transpose)**.** Verify the generalized reversal law $(ABC)^T = C^T B^T A^T$ for the matrices:

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 0 \\ 2 & 3 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

Solution. First, we compute the left-hand side, $(ABC)^T$. We calculate the product AB:

$$AB = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 2 & 3 \end{pmatrix}$$

Next, we find the full product ABC:

$$(AB)C = \begin{pmatrix} 3 & 6 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 12 \\ 2 & 6 \end{pmatrix}$$

The transpose of this product is the left-hand side:

$$(ABC)^T = \begin{pmatrix} 3 & 2 \\ 12 & 6 \end{pmatrix}$$

Now, we compute the right-hand side, $C^T B^T A^T$. We first find the individual transposes:

$$A^T = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \quad B^T = \begin{pmatrix} -1 & 2 \\ 0 & 3 \end{pmatrix}, \quad C^T = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

We compute the product in reverse order, starting with $C^T B^T$:

$$C^T B^T = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ 0 & 6 \end{pmatrix}$$

Finally, we complete the product for the right-hand side:

$$(C^T B^T)A^T = \begin{pmatrix} -1 & 2 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 12 & 6 \end{pmatrix}$$

Since the left-hand side and right-hand side are identical, the generalized reversal law is verified for this example. ■

1.1.2 Hermitian and Skew-Hermitian Matrices

These matrices are the complex analogues of real symmetric and skew-symmetric matrices. Their definitions rely on the **conjugate transpose** (or Hermitian conjugate) of a matrix A, denoted A^*, which is defined as $A^* = (\overline{A})^T$.

Definition 1.1.4 (Hermitian Matrix)**.** A square matrix A is **Hermitian** if it is equal to its own conjugate transpose.

$$A^* = A$$

A key property of a Hermitian matrix is that its main diagonal entries must be real numbers, since for any diagonal entry a_{ii}, the condition $a_{ii} = \overline{a_{ii}}$ must hold.

Definition 1.1.5 (Skew-Hermitian Matrix)**.** A square matrix A is **skew-Hermitian** if it is equal to the negative of its conjugate transpose.

$$A^* = -A$$

A key property of a skew-Hermitian matrix is that its main diagonal entries must be purely imaginary or zero. This is because the condition $a_{ii} = -\overline{a_{ii}}$ implies that if $a_{ii} = x + iy$, then $x + iy = -(x - iy) = -x + iy$, which requires $2x = 0$, so $x = 0$.

Example 1.1.6 (Identifying Hermitian and Skew-Hermitian Matrices)**.** Determine if the matrices $A = \begin{pmatrix} 3 & 2-i \\ 2+i & -1 \end{pmatrix}$ and $B = \begin{pmatrix} i & 2+i \\ -2+i & 0 \end{pmatrix}$ are Hermitian, skew-Hermitian, or neither.

Solution. For matrix A, we compute its conjugate transpose $A^* = (\overline{A})^T$.

$$\overline{A} = \begin{pmatrix} 3 & 2+i \\ 2-i & -1 \end{pmatrix} \implies A^* = (\overline{A})^T = \begin{pmatrix} 3 & 2-i \\ 2+i & -1 \end{pmatrix}$$

Since $A^* = A$, the matrix A is **Hermitian**.

For matrix B, we compute its conjugate transpose $B^* = (\overline{B})^T$.

$$\overline{B} = \begin{pmatrix} -i & 2-i \\ -2-i & 0 \end{pmatrix} \implies B^* = (\overline{B})^T = \begin{pmatrix} -i & -2-i \\ 2-i & 0 \end{pmatrix}$$

Now we compare B^* with $-B$:

$$-B = -\begin{pmatrix} i & 2+i \\ -2+i & 0 \end{pmatrix} = \begin{pmatrix} -i & -2-i \\ 2-i & 0 \end{pmatrix}$$

Since $B^* = -B$, the matrix B is **skew-Hermitian**. ■

1.1.3 Representation of a Square Matrix as P + iQ

A fundamental result states that any square matrix with complex entries can be uniquely decomposed into a combination of two Hermitian matrices.

Theorem 1.1.7. *Every square matrix M can be uniquely expressed in the form $M = P + iQ$, where P and Q are both Hermitian matrices.*

Proof. The proof is constructive. We define the matrices P and Q as follows:

$$P = \frac{1}{2}(M + M^*) \quad \text{and} \quad Q = \frac{1}{2i}(M - M^*)$$

We must show that P and Q are Hermitian. For P, we take the conjugate transpose:

$$P^* = \left(\frac{1}{2}(M + M^*)\right)^* = \frac{1}{2}(M^* + (M^*)^*) = \frac{1}{2}(M^* + M) = P$$

Thus, P is Hermitian. For Q, we proceed similarly, noting that $(i)^* = \bar{i} = -i$:

$$Q^* = \left(\frac{1}{2i}(M - M^*)\right)^* = \frac{1}{-2i}(M^* - (M^*)^*) = \frac{1}{-2i}(M^* - M) = \frac{1}{2i}(M - M^*) = Q$$

Thus, Q is also Hermitian. To confirm the decomposition, we can substitute P and Q back:

$$P + iQ = \frac{1}{2}(M + M^*) + i\left(\frac{1}{2i}(M - M^*)\right) = \frac{1}{2}(M + M^*) + \frac{1}{2}(M - M^*) = \frac{1}{2}(2M) = M$$

The decomposition is valid. The uniqueness of this representation can also be proven. ■

Example 1.1.8 (Representation of a Square Matrix as P + iQ)**.** Express the matrix $M = \begin{pmatrix} 2+i & 4 \\ 2i & 3-2i \end{pmatrix}$ in the form $P + iQ$, where P and Q are both Hermitian matrices.

Solution. We use the formulas $P = \frac{1}{2}(M + M^*)$ and $Q = \frac{1}{2i}(M - M^*)$. First, we find the conjugate transpose of M:

$$M^* = (\overline{M})^T = \begin{pmatrix} 2-i & -2i \\ 4 & 3+2i \end{pmatrix}$$

Now we compute the sum $M + M^*$ to find P:

$$M + M^* = \begin{pmatrix} (2+i) + (2-i) & 4+4 \\ 2i + (-2i) & (3-2i) + (3+2i) \end{pmatrix} = \begin{pmatrix} 4 & 8 \\ 0 & 6 \end{pmatrix}$$

$$P = \frac{1}{2}(M + M^*) = \begin{pmatrix} 2 & 4 \\ 0 & 3 \end{pmatrix}$$

Next, we compute the difference $M - M^*$ to find Q:

$$M - M^* = \begin{pmatrix} (2+i) - (2-i) & 4-4 \\ 2i - (-2i) & (3-2i) - (3+2i) \end{pmatrix} = \begin{pmatrix} 2i & 0 \\ 4i & -4i \end{pmatrix}$$

$$Q = \frac{1}{2i}(M - M^*) = \frac{1}{2i}\begin{pmatrix} 2i & 0 \\ 4i & -4i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & -2 \end{pmatrix}$$

We can quickly verify that both P and Q are real symmetric matrices, and thus are Hermitian. The required decomposition is:

$$M = \begin{pmatrix} 2 & 4 \\ 0 & 3 \end{pmatrix} + i\begin{pmatrix} 1 & 0 \\ 2 & -2 \end{pmatrix}$$

■

1.2 The Adjoint Matrix and Fundamental Laws of Operations

1.2.1 The Adjoint of a Matrix

The adjoint (or adjugate) of a square matrix is a fundamental concept used in the theoretical definition of the inverse of a matrix. It is constructed from the matrix of cofactors.

Definition 1.2.1 (Cofactor and Adjoint)**.** Let A be an $n \times n$ square matrix.

1. The **minor** M_{ij} of the element a_{ij} is the determinant of the $(n-1)\times(n-1)$ matrix obtained by deleting the i-th row and j-th column of A.
2. The **cofactor** C_{ij} of the element a_{ij} is given by $C_{ij} = (-1)^{i+j}M_{ij}$.
3. The **matrix of cofactors** is the matrix $C = [C_{ij}]$.
4. The **adjoint** of A, denoted $\text{adj}(A)$, is the transpose of the matrix of cofactors.

$$\text{adj}(A) = C^T$$

For a simple 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the cofactors are $C_{11} = d, C_{12} = -c, C_{21} = -b, C_{22} = a$. The cofactor matrix is $\begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$. The adjoint is therefore $\text{adj}(A) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

1.2.2 The Adjoint-Inverse Identity

The most important property of the adjoint is its relationship with the original matrix and its determinant.

Theorem 1.2.2. *For any $n \times n$ square matrix A, the following identity holds:*

$$A(\textit{adj } A) = (\textit{adj } A)A = \det(A)I$$

where I is the $n \times n$ identity matrix.

Proof. Let's consider the product $A(\text{adj } A)$. The element in the i-th row and j-th column of this product is the dot product of the i-th row of A and the j-th column of $\text{adj}(A)$. The j-th column of $\text{adj}(A)$ is the j-th row of the cofactor matrix C, which consists of the cofactors $(C_{j1}, C_{j2}, \ldots, C_{jn})$. So, the (i,j)-th element of the product is:

$$(A(\text{adj } A))_{ij} = a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn} = \sum_{k=1}^{n} a_{ik}C_{jk}$$

We consider two cases for this sum, based on Laplace's expansion of a determinant.

- **Case 1:** $i = j$ **(Diagonal Elements)**. The sum becomes $\sum_{k=1}^{n} a_{ik}C_{ik}$. This is precisely the definition of the determinant of A expanded along the i-th row. Thus, all diagonal elements of the product are equal to $\det(A)$.
- **Case 2:** $i \neq j$ **(Off-Diagonal Elements)**. The sum $\sum_{k=1}^{n} a_{ik}C_{jk}$ represents the determinant of a matrix formed by replacing the j-th row of A with a copy of the i-th row. A matrix with two identical rows has a determinant of zero. Thus, all off-diagonal elements of the product are zero.

Combining these two cases, the resulting matrix has $\det(A)$ on its main diagonal and zeros everywhere else. This is exactly the matrix $\det(A)I$. The proof for $(\text{adj } A)A$ is analogous, using column expansions. ■

1.2.3 Commutative and Associative Laws in Matrix Operations

Not all laws of real number arithmetic apply directly to matrices. It is essential to know which laws hold.

Matrix Addition Matrix addition is both commutative and associative, provided the matrices have the same dimensions.

- **Commutative Law:** $A + B = B + A$
- **Associative Law:** $(A + B) + C = A + (B + C)$

Matrix Multiplication Matrix multiplication is associative, but it is **not** commutative in general.

- **Associative Law:** $A(BC) = (AB)C$ (provided the products are defined).
- **Non-Commutative Law:** In general, $AB \neq BA$.

  For a direct counterexample, let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.

$$AB = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

$$BA = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

  Clearly, $AB \neq BA$.

Scalar Multiplication Scalar multiplication is associative and commutative with respect to scalars and matrices. For any scalar c and matrices A, B:

$$c(AB) = (cA)B = A(cB)$$

1.3 Invertibility and the Reversal Law for Inverses

1.3.1 Necessary and Sufficient Condition for Invertibility

The invertibility of a square matrix is one of its most important properties. There are many equivalent conditions for a matrix to be invertible, but the most fundamental one relates to its determinant.

Theorem 1.3.1. *A square matrix A is invertible (or non-singular) if and only if its determinant is non-zero, i.e.,* $\det(A) \neq 0$.

Proof. We prove both directions of the "if and only if" statement.

(⇒) Assume A is invertible. By definition, there exists a matrix A^{-1} such that $AA^{-1} = I$, where I is the identity matrix. Taking the determinant of both sides, we get $\det(AA^{-1}) = \det(I)$. Using the property that the determinant of a product is the product of the determinants, we have $\det(A)\det(A^{-1}) = 1$. Since the product of these two numbers is 1, neither can be zero. Therefore, $\det(A) \neq 0$.

(⇐) Assume $\det(A) \neq 0$. We know from the properties of the adjoint matrix that $A(\text{adj } A) = \det(A)I$. Since $\det(A)$ is a non-zero scalar, we can divide the entire equation by it:

$$A\left(\frac{1}{\det(A)}\text{adj } A\right) = I$$

This equation is of the form $AB = I$, where $B = \frac{1}{\det(A)}\text{adj}(A)$. This shows that A has a right inverse. A similar argument using $(\text{adj } A)A = \det(A)I$ shows it has a left inverse. Therefore, A is invertible, and its inverse is given by the formula $A^{-1} = \frac{1}{\det(A)}\text{adj}(A)$. ■

Example 1.3.2 (Condition for Invertibility)**.** Determine if the following matrix A is invertible.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 2 & 1 & 1 \end{pmatrix}$$

Solution. A square matrix is invertible if and only if its determinant is non-zero. We calculate the determinant of A by expanding along the first column:

$$\begin{aligned}\det(A) &= 1 \cdot \det\begin{pmatrix} 1 & 4 \\ 1 & 1 \end{pmatrix} - 0 \cdot \det\begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix} + 2 \cdot \det\begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix} \\ &= 1(1-4) - 0 + 2(8-3) \\ &= -3 + 2(5) = -3 + 10 = 7\end{aligned}$$

Since $\det(A) = 7 \neq 0$, the matrix A is invertible. ■

1.3.2 Generalization of the Reversal Law for Inverses

Similar to the transpose, the inverse of a product of matrices is the product of their inverses in the reverse order.

Theorem 1.3.3 (Reversal Law for Two Matrices)**.** *If A and B are invertible matrices of the same size, then their product AB is also invertible, and its inverse is given by:*

$$(AB)^{-1} = B^{-1}A^{-1}$$

Proof. To prove that $B^{-1}A^{-1}$ is the inverse of AB, we must show that their product is the identity matrix. We multiply (AB) by $(B^{-1}A^{-1})$:

$$\begin{aligned}(AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} \quad \text{(by associativity)} \\ &= A(I)A^{-1} = AA^{-1} = I\end{aligned}$$

Since their product is the identity matrix, $B^{-1}A^{-1}$ is indeed the inverse of AB. ■

This law can be generalized to the product of any finite number of invertible matrices.

Theorem 1.3.4 (Generalization of Reversal Law)**.** *If $A_1, A_2, \ldots, A_n$ are invertible matrices of the same size, then their product is also invertible, and:*

$$(A_1A_2\cdots A_n)^{-1} = A_n^{-1}\cdots A_2^{-1}A_1^{-1}$$

Proof. We prove this by mathematical induction on n.

Base Case: For $n = 2$, the theorem is $(A_1A_2)^{-1} = A_2^{-1}A_1^{-1}$, which we have just proven.

Inductive Hypothesis: Assume the formula holds for a product of k invertible matrices.

$$(A_1A_2\cdots A_k)^{-1} = A_k^{-1}\cdots A_2^{-1}A_1^{-1}$$

Inductive Step: We prove the formula for $k+1$ matrices. We group the product as $(A_1\cdots A_k)A_{k+1}$ and apply the base case for two matrices:

$$((A_1\cdots A_k)A_{k+1})^{-1} = A_{k+1}^{-1}(A_1\cdots A_k)^{-1}$$

Now, using our inductive hypothesis on the term $(A_1\cdots A_k)^{-1}$, we get:

$$A_{k+1}^{-1}(A_k^{-1}\cdots A_2^{-1}A_1^{-1}) = A_{k+1}^{-1}A_k^{-1}\cdots A_2^{-1}A_1^{-1}$$

Thus, the formula holds for $k+1$ matrices. By the principle of mathematical induction, the theorem is true for all integers $n \geq 2$. ■

Example 1.3.5 (Reversal Law for Inverses)**.** Verify the reversal law $(AB)^{-1} = B^{-1}A^{-1}$ for the matrices:

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

Solution. Left-Hand Side (LHS). First, we compute the product AB:

$$AB = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 3 \\ 5 & 4 \end{pmatrix}$$

The determinant is $\det(AB) = (4)(4) - (3)(5) = 16 - 15 = 1$. The inverse is:

$$(AB)^{-1} = \frac{1}{1}\begin{pmatrix} 4 & -3 \\ -5 & 4 \end{pmatrix} = \begin{pmatrix} 4 & -3 \\ -5 & 4 \end{pmatrix}$$

Right-Hand Side (RHS). First, we find the individual inverses of A and B.

$$\det(A) = 3 - 2 = 1 \implies A^{-1} = \frac{1}{1}\begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix}$$

$$\det(B) = 2 - 1 = 1 \implies B^{-1} = \frac{1}{1}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

Now, we compute the product $B^{-1}A^{-1}$:

$$B^{-1}A^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 3+1 & -2-1 \\ -3-2 & 2+2 \end{pmatrix} = \begin{pmatrix} 4 & -3 \\ -5 & 4 \end{pmatrix}$$

Conclusion. Since the LHS and RHS are identical, the reversal law is verified. ■

Chapter 2

Matrix Operations and Eigenvalues

2.1 Matrix Operation of Transposing and Inverting are Commutative

For an invertible matrix, the order in which you perform the transpose and inverse operations does not matter.

Theorem 2.1.1. *If A is an invertible square matrix, then its transpose A^T is also invertible, and the transpose of the inverse is the inverse of the transpose.*

$$(A^{-1})^T = (A^T)^{-1}$$

Proof. To prove this, we must show that $(A^{-1})^T$ is the inverse of A^T. We can do this by showing that their product is the identity matrix. Using the reversal law for transposes, $(XY)^T = Y^T X^T$, we have:

$$A^T(A^{-1})^T = (A^{-1}A)^T$$

Since $A^{-1}A = I$, the expression becomes:

$$(I)^T = I$$

Because $A^T(A^{-1})^T = I$, it follows by the definition of an inverse that $(A^{-1})^T$ is the inverse of A^T. ■

Example 2.1.2 (Commutativity of Transpose and Inverse)**.** Verify that $(A^{-1})^T = (A^T)^{-1}$ for the matrix $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$.

Solution. Left-Hand Side (LHS). First, we find the inverse of A. Since $\det(A) = 1$, the inverse is:

$$A^{-1} = \frac{1}{1}\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$$

Now, we take the transpose of A^{-1}:

$$(A^{-1})^T = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$$

Right-Hand Side (RHS). First, we find the transpose of A:

$$A^T = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

Now, we find the inverse of A^T. Since $\det(A^T) = 1$, the inverse is:

$$(A^T)^{-1} = \frac{1}{1}\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$$

Conclusion. The left-hand side and right-hand side are identical, verifying the theorem. ■

2.1.1 Trace of a Matrix

The trace is a simple but useful operator that maps a square matrix to a scalar value.

Definition 2.1.3 (Trace)**.** The **trace** of an $n \times n$ square matrix A, denoted $\mathrm{tr}(A)$, is the sum of the elements on its main diagonal.

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}$$

Properties of the Trace: The trace is a linear operator, meaning for any square matrices A, B and scalar c:

- $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$
- $\mathrm{tr}(cA) = c \cdot \mathrm{tr}(A)$

2.1.2 The Cyclic Property of the Trace

While matrix multiplication is not commutative ($AB \neq BA$), the trace of the product is surprisingly immune to the order of multiplication.

Theorem 2.1.4. *If A is an $m \times n$ matrix and B is an $n \times m$ matrix, then the trace of their product is commutative.*

$$tr(AB) = tr(BA)$$

Proof. By the definition of matrix multiplication, the (i, i)-th element of the product AB is $(AB)_{ii} = \sum_{j=1}^{n} a_{ij}b_{ji}$. The trace of AB is the sum of these diagonal elements:

$$\text{tr}(AB) = \sum_{i=1}^{m}(AB)_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ji}$$

Similarly, the (j, j)-th element of the product BA is $(BA)_{jj} = \sum_{i=1}^{m} b_{ji}a_{ij}$. The trace of BA is:

$$\text{tr}(BA) = \sum_{j=1}^{n}(BA)_{jj} = \sum_{j=1}^{n}\sum_{i=1}^{m} b_{ji}a_{ij}$$

Since the summation of scalars is commutative, we can reorder the terms and the summations. The two expressions are equal. ■

This property can be generalized for any number of matrices in what is known as the **cyclic property of the trace**.

Theorem 2.1.5 (Generalization)**.** *For any three matrices A, B, C whose product ABC is a square matrix, the trace is invariant under cyclic permutations of the matrices.*

$$tr(ABC) = tr(BCA) = tr(CAB)$$

Proof. We can prove this by cleverly grouping the matrices and applying the base case $\text{tr}(XY) = \text{tr}(YX)$. Let $X = AB$ and $Y = C$. Then:

$$\text{tr}(ABC) = \text{tr}((AB)C) = \text{tr}(XY) = \text{tr}(YX) = \text{tr}(C(AB)) = \text{tr}(CAB)$$

Now let $X = A$ and $Y = BC$. Then:

$$\text{tr}(ABC) = \text{tr}(A(BC)) = \text{tr}(XY) = \text{tr}(YX) = \text{tr}((BC)A) = \text{tr}(BCA)$$

Thus, all three expressions are equal. ■

Example 2.1.6 (Trace of a Matrix)**.** Find the trace of the matrix $A = \begin{pmatrix} 5 & -1 & 0 \\ 2 & 3 & 7 \\ 1 & 6 & -4 \end{pmatrix}$.

Solution. The trace of a square matrix is the sum of its main diagonal elements.

$$\text{tr}(A) = 5 + 3 + (-4) = 4$$

The trace of the matrix is 4. ■

Example 2.1.7 (Cyclic Property of the Trace: Two Matrices)**.** Verify that $\text{tr}(AB) = \text{tr}(BA)$ for the non-square matrices:

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 3 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 1 \end{pmatrix}$$

Solution. First, we compute the 2×2 product AB and its trace.

$$AB = \begin{pmatrix} 1 & 2 & 0 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1+4+0 & 1+4+0 \\ 3+0+0 & 3+0+1 \end{pmatrix} = \begin{pmatrix} 5 & 5 \\ 3 & 4 \end{pmatrix}$$

$$\text{tr}(AB) = 5 + 4 = 9$$

Next, we compute the 3×3 product BA and its trace.

$$BA = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 3 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1+3 & 2+0 & 0+1 \\ 2+6 & 4+0 & 0+2 \\ 0+3 & 0+0 & 0+1 \end{pmatrix} = \begin{pmatrix} 4 & 2 & 1 \\ 8 & 4 & 2 \\ 3 & 0 & 1 \end{pmatrix}$$

$$\text{tr}(BA) = 4 + 4 + 1 = 9$$

The traces are equal, verifying the property. ■

Example 2.1.8 (Cyclic Property of the Trace: Three Matrices)**.** Verify that $\mathrm{tr}(ABC) = \mathrm{tr}(BCA)$ for the matrices:

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

Solution. First, we calculate $\mathrm{tr}(ABC)$.

$$AB = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix}$$

$$(AB)C = \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$

$$\mathrm{tr}(ABC) = 2 + 2 = 4$$

Next, we calculate $\mathrm{tr}(BCA)$.

$$BC = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix}$$

$$(BC)A = \begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 0 & 0 \end{pmatrix}$$

$$\mathrm{tr}(BCA) = 4 + 0 = 4$$

The traces are equal, verifying the cyclic property for this set of matrices. ■

2.2 Partitioning of Matrices

Partitioning (or blocking) a matrix is the process of dividing it into smaller rectangular submatrices, called **blocks** or **submatrices**.

Definition 2.2.1 (Partitioned Matrix)**.** An $m \times n$ matrix A can be partitioned by drawing horizontal and vertical lines between its rows and columns. For example, a matrix can be represented as a 2×2 partitioned matrix:

$$A = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right)$$

Here, A_{11}, A_{12}, A_{21}, and A_{22} are the blocks (submatrices) of A.

Operations on Partitioned Matrices

Addition: If two matrices A and B have the same dimensions and are partitioned in exactly the same way, their sum is found by simply adding the corresponding blocks:

$$A + B = \left(\begin{array}{c|c} A_{11} + B_{11} & A_{12} + B_{12} \\ \hline A_{21} + B_{21} & A_{22} + B_{22} \end{array}\right)$$

Block Multiplication: If two matrices A and B are partitioned such that the number of columns in each block of A matches the number of rows in the corresponding block of B (i.e., the column partitions of A conform to the row partitions of B), their product can be computed by treating the blocks as if they were individual elements. If $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ and $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$, their product is:

$$AB = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}$$

Example 2.2.2 (Block Multiplication)**.** Calculate the product AB using block multiplication, where:

$$A = \left(\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & 1 & 1 \\ \hline -1 & 0 & 3 \end{array}\right) \quad \text{and} \quad B = \left(\begin{array}{c|c} 1 & 0 \\ 0 & 2 \\ \hline 3 & 1 \end{array}\right)$$

Solution. We define the partitions as follows:

$$A_{11} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad A_{12} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad A_{21} = \begin{pmatrix} -1 & 0 \end{pmatrix}, \quad A_{22} = \begin{pmatrix} 3 \end{pmatrix}$$

$$B_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad B_{12} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad B_{21} = \begin{pmatrix} 3 & 1 \end{pmatrix}, \quad B_{22} = \begin{pmatrix} 1 \end{pmatrix}$$

The column partition of A (2 columns, 1 column) matches the row partition of B (2 rows, 1 row), so block multiplication is possible. We compute the blocks of the product $C = AB$:

Top-Left Block (C_{11}):

$$C_{11} = A_{11}B_{11} + A_{12}B_{21} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 3 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 4 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 3 & 3 \end{pmatrix}$$

Top-Right Block (C_{12}):

$$C_{12} = A_{11}B_{12} + A_{12}B_{22} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

Bottom-Left Block (C_{21}):

$$C_{21} = A_{21}B_{11} + A_{22}B_{21} = \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 3 \end{pmatrix} \begin{pmatrix} 3 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} -1 & 0 \end{pmatrix} + \begin{pmatrix} 9 & 3 \end{pmatrix} = \begin{pmatrix} 8 & 3 \end{pmatrix}$$

Bottom-Right Block (C_{22}):

$$C_{22} = A_{21}B_{12} + A_{22}B_{22} = \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \end{pmatrix} \begin{pmatrix} 1 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix} + \begin{pmatrix} 3 \end{pmatrix} = \begin{pmatrix} 3 \end{pmatrix}$$

Combining these blocks gives the final product:

$$AB = \left(\begin{array}{cc|c} 1 & 4 & 4 \\ 3 & 3 & 3 \\ \hline 8 & 3 & 3 \end{array}\right) = \begin{pmatrix} 1 & 4 & 4 \\ 3 & 3 & 3 \\ 8 & 3 & 3 \end{pmatrix}$$

This result is identical to the one obtained by standard matrix multiplication. ■

2.3 Eigenvalues, Eigenvectors, and Associated Polynomials

Definition 2.3.1 (Eigenvalue and Eigenvector)**.** For an $n \times n$ matrix A, a non-zero vector $\mathbf{v}$ is an **eigenvector** if there exists a scalar λ such that:

$$A\mathbf{v} = \lambda\mathbf{v}$$

The scalar λ is the **eigenvalue** corresponding to $\mathbf{v}$. It represents the factor by which the eigenvector $\mathbf{v}$ is stretched or shrunk.

To find the eigenvalues, we rearrange the defining equation:

$$A\mathbf{v} - \lambda\mathbf{v} = \mathbf{0} \implies (A - \lambda I)\mathbf{v} = \mathbf{0}$$

This is a homogeneous system of linear equations. Since we require the eigenvector $\mathbf{v}$ to be non-zero, we are looking for non-trivial solutions. A non-trivial solution exists if and only if the matrix $(A - \lambda I)$ is singular (i.e., not invertible). This gives us a direct method for finding eigenvalues.

2.3.1 The Characteristic Equation

A matrix is singular if and only if its determinant is zero. This principle leads us to the characteristic equation.

Definition 2.3.2 (Characteristic Polynomial and Equation)**.** The **characteristic polynomial** of an $n \times n$ matrix A is the polynomial $p(\lambda) = \det(A - \lambda I)$. The **characteristic equation** is $p(\lambda) = 0$.

From the reasoning above, the eigenvalues of a matrix A are precisely the roots of its characteristic equation.

Example 2.3.3 (Finding Eigenvalues and Eigenvectors)**.** Find the eigenvalues and corresponding eigenvectors of the matrix $A = \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}$.

Solution. First, we find the eigenvalues by solving the characteristic equation $\det(A - \lambda I) = 0$.

$$\det \begin{pmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{pmatrix} = (4-\lambda)(3-\lambda) - 2 = \lambda^2 - 7\lambda + 12 - 2 = \lambda^2 - 7\lambda + 10 = 0$$

Factoring the polynomial gives $(\lambda - 5)(\lambda - 2) = 0$. The eigenvalues are $\lambda_1 = 5$ and $\lambda_2 = 2$.

Now, we find the eigenvector for each eigenvalue.

For $\lambda_1 = 5$: We solve the system $(A - 5I)\mathbf{v} = \mathbf{0}$.

$$\begin{pmatrix} 4-5 & 1 \\ 2 & 3-5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This gives the equation $-x + y = 0$, or $x = y$. A simple non-zero solution is $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

For $\lambda_2 = 2$: We solve the system $(A - 2I)\mathbf{v} = \mathbf{0}$.

$$\begin{pmatrix} 4-2 & 1 \\ 2 & 3-2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This gives the equation $2x + y = 0$, or $y = -2x$. A simple non-zero solution is $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$. ∎

Example 2.3.4 (Evaluating a Matrix Polynomial)**.** Given the matrix $C = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and the polynomial $p(x) = x^2 - 2x + 3$, find the matrix $p(C)$.

Solution. We need to compute $p(C) = C^2 - 2C + 3I$. First, we calculate the matrix C^2:

$$C^2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}$$

Now we substitute this into the polynomial expression:

$$\begin{aligned} p(C) &= C^2 - 2C + 3I \\ &= \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix} - 2\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} + 3\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 2 & 4 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 1-2+3 & 4-4+0 \\ 0-0+0 & 1-2+3 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \end{aligned}$$

Thus, $p(C) = 2I$. ∎

2.3.2 The Cayley-Hamilton Theorem

This remarkable theorem states that a matrix satisfies its own characteristic equation. Before stating it, we must define a matrix polynomial.

Definition 2.3.5 (Matrix Polynomial)**.** If $p(x) = c_k x^k + \cdots + c_1 x + c_0$ is a scalar polynomial, then for a square matrix A, the matrix polynomial $p(A)$ is:

$$p(A) = c_k A^k + \cdots + c_1 A + c_0 I$$

Note that the constant term is multiplied by the identity matrix I.

Theorem 2.3.6 (Cayley-Hamilton Theorem)**.** *If $p(\lambda)$ is the characteristic polynomial of a square matrix A, then $p(A) = 0$, where 0 is the zero matrix.*

Proof. Let A be an $n \times n$ square matrix. Its characteristic polynomial is $p(\lambda) = \det(A - \lambda I)$. This is a polynomial of degree n in λ, which can be written as:

$$p(\lambda) = c_n \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1 \lambda + c_0$$

Our goal is to show that $p(A) = c_n A^n + c_{n-1} A^{n-1} + \cdots + c_1 A + c_0 I = 0$.

The proof relies on the fundamental property of the adjugate (or adjoint) matrix. For any $n \times n$ matrix M, we have the identity:

$$M(\text{adj } M) = (\det M) I$$

Let us apply this identity to the matrix $M = A - \lambda I$.

$$(A - \lambda I) \cdot \text{adj}(A - \lambda I) = \det(A - \lambda I) \cdot I$$

$$(A - \lambda I) \cdot \text{adj}(A - \lambda I) = p(\lambda) I$$

The adjugate of $(A - \lambda I)$ is the transpose of its cofactor matrix. The entries of $(A - \lambda I)$ are linear polynomials in λ. The cofactors are determinants of $(n-1) \times (n-1)$ submatrices, so they are polynomials in λ of degree at most $n - 1$.

Therefore, the adjugate matrix, $\text{adj}(A - \lambda I)$, can be written as a matrix polynomial in λ of degree at most $n - 1$. Let us denote it as $B(\lambda)$:

$$\text{adj}(A - \lambda I) = B(\lambda) = B_{n-1}\lambda^{n-1} + B_{n-2}\lambda^{n-2} + \cdots + B_1\lambda + B_0$$

where the coefficients B_k are $n \times n$ matrices with constant entries.

Now, we substitute this back into our identity:

$$(A - \lambda I)(B_{n-1}\lambda^{n-1} + \cdots + B_0) = (c_n\lambda^n + \cdots + c_0)I$$

We expand both sides of this equation. The left-hand side (LHS) becomes:

$$A(B_{n-1}\lambda^{n-1} + \cdots + B_0) - \lambda I(B_{n-1}\lambda^{n-1} + \cdots + B_0)$$

$$(AB_{n-1}\lambda^{n-1} + \cdots + AB_0) - (B_{n-1}\lambda^n + \cdots + B_0\lambda)$$

The right-hand side (RHS) is:

$$c_n I\lambda^n + c_{n-1}I\lambda^{n-1} + \cdots + c_1 I\lambda + c_0 I$$

This equation is an identity between two matrix polynomials in λ. For this to be true, the matrix coefficients of each power of λ must be equal. We now equate the coefficients for each power of λ from λ^n down to λ^0.

$$\begin{aligned}
\text{coeff of } \lambda^n: &\quad -B_{n-1} = c_n I \\
\text{coeff of } \lambda^{n-1}: &\quad AB_{n-1} - B_{n-2} = c_{n-1}I \\
\text{coeff of } \lambda^{n-2}: &\quad AB_{n-2} - B_{n-3} = c_{n-2}I \\
&\quad \vdots \\
\text{coeff of } \lambda^1: &\quad AB_1 - B_0 = c_1 I \\
\text{coeff of } \lambda^0: &\quad AB_0 = c_0 I
\end{aligned}$$

To arrive at the final result, we will multiply each of these equations on the left by the corresponding power of A (from A^n down to I) and then sum them together.

$$\begin{aligned}
A^n(-B_{n-1}) &= c_n A^n I \\
A^{n-1}(AB_{n-1} - B_{n-2}) &= c_{n-1}A^{n-1}I \\
A^{n-2}(AB_{n-2} - B_{n-3}) &= c_{n-2}A^{n-2}I \\
&\vdots \\
A(AB_1 - B_0) &= c_1 AI \\
I(AB_0) &= c_0 I
\end{aligned}$$

Adding all these equations vertically, we observe that the sum of the right-hand sides is:

$$\text{RHS Sum} = c_n A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I = p(A)$$

Now let's examine the sum of the left-hand sides:

$$\text{LHS Sum} = (-A^n B_{n-1}) + (A^n B_{n-1} - A^{n-1}B_{n-2}) + \cdots + (A^2 B_1 - AB_0) + (AB_0)$$

This is a telescoping sum where each positive term cancels with the negative term from the line above it. For instance, $A^n B_{n-1}$ cancels with $-A^n B_{n-1}$, and $A^{n-1}B_{n-2}$ will cancel with the term from the next line, and so on, until AB_0 cancels with $-AB_0$. The entire sum on the left-hand side collapses to the zero matrix. Therefore, we are left with:

$$0 = p(A)$$

This completes the proof. ■

This theorem is incredibly useful for finding the inverse of a matrix. If $p(\lambda) = c_n\lambda^n + \cdots + c_1\lambda + c_0$, then $p(A) = c_n A^n + \cdots + c_1 A + c_0 I = 0$. If A is invertible, then its determinant, c_0, is non-zero. We can then multiply by A^{-1} and solve for it in terms of powers of A.

Example 2.3.7 (Application of the Cayley-Hamilton Theorem)**.** Use the Cayley-Hamilton theorem to find the inverse of the matrix $A = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}$.

Solution. First, we find the characteristic equation of A.

$$\det(A - \lambda I) = \det \begin{pmatrix} 1-\lambda & 2 \\ 3 & 5-\lambda \end{pmatrix} = (1-\lambda)(5-\lambda) - 6 = 5 - 6\lambda + \lambda^2 - 6 = \lambda^2 - 6\lambda - 1 = 0$$

By the Cayley-Hamilton theorem, the matrix A satisfies this equation:

$$A^2 - 6A - I = 0$$

Since $\det(A) = -1 \neq 0$, the inverse exists. We can multiply the equation by A^{-1}:

$$A^{-1}(A^2 - 6A - I) = 0 \implies A - 6I - A^{-1} = 0$$

Solving for A^{-1} gives:

$$A^{-1} = A - 6I = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix} - 6\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1-6 & 2-0 \\ 3-0 & 5-6 \end{pmatrix} = \begin{pmatrix} -5 & 2 \\ 3 & -1 \end{pmatrix}$$

■

2.3.3 The Minimal Polynomial of a Matrix

A matrix satisfies its characteristic polynomial, but is that the polynomial of the *smallest* degree that it satisfies? The answer is not always yes, which leads to the concept of the minimal polynomial.

Definition 2.3.8 (Minimal Polynomial)**.** The **minimal polynomial** of a square matrix A is the unique monic polynomial (leading coefficient is 1) $m(\lambda)$ of least degree such that $m(A) = 0$.

Theorem 2.3.9. *The minimal polynomial $m(\lambda)$ of a matrix A always divides the characteristic polynomial $p(\lambda)$. Furthermore, every eigenvalue of A (i.e., every root of $p(\lambda)$) is also a root of $m(\lambda)$.*

This means that the minimal polynomial and the characteristic polynomial share the same roots, but the characteristic polynomial may have these roots with higher multiplicities. For example, if $p(\lambda) = (\lambda-2)^3(\lambda-5)$, the minimal polynomial could be $(\lambda-2)(\lambda-5)$, $(\lambda-2)^2(\lambda-5)$, or the characteristic polynomial itself.

Proof. We will prove the two parts of the theorem separately.

Part 1: The minimal polynomial divides the characteristic polynomial. By the Division Algorithm for polynomials, we can divide the characteristic polynomial $p(\lambda)$ by the minimal polynomial $m(\lambda)$ to get a quotient $q(\lambda)$ and a remainder $r(\lambda)$:

$$p(\lambda) = q(\lambda)m(\lambda) + r(\lambda)$$

where the degree of the remainder, $\deg(r(\lambda))$, is strictly less than the degree of the minimal polynomial, $\deg(m(\lambda))$.

This is an identity for scalar polynomials, so we can evaluate it with the matrix A:

$$p(A) = q(A)m(A) + r(A)$$

By the Cayley-Hamilton theorem, we know that $p(A) = 0$. By the definition of the minimal polynomial, we know that $m(A) = 0$. Substituting these into the equation gives:

$$0 = q(A) \cdot 0 + r(A)$$

$$0 = r(A)$$

This result states that the remainder polynomial $r(\lambda)$, when evaluated at A, gives the zero matrix. However, the minimal polynomial $m(\lambda)$ is defined as the monic polynomial of *least* degree for which this is true. Since we know that $\deg(r(\lambda)) < \deg(m(\lambda))$, this would be a contradiction unless the remainder polynomial $r(\lambda)$ is itself the zero polynomial. Therefore, $r(\lambda) = 0$. This means our original division equation has no remainder:

$$p(\lambda) = q(\lambda)m(\lambda)$$

This is precisely the definition of $m(\lambda)$ dividing $p(\lambda)$.

Part 2: Every eigenvalue is a root of the minimal polynomial. Let λ_0 be an eigenvalue of the matrix A. By definition, this means there exists a corresponding non-zero eigenvector $\mathbf{v}$ such that:

$$A\mathbf{v} = \lambda_0 \mathbf{v}$$

Applying the matrix A repeatedly to this equation, we can see a pattern:

$$A^2\mathbf{v} = A(A\mathbf{v}) = A(\lambda_0\mathbf{v}) = \lambda_0(A\mathbf{v}) = \lambda_0(\lambda_0\mathbf{v}) = \lambda_0^2\mathbf{v}$$

By induction, for any non-negative integer k, we have $A^k\mathbf{v} = \lambda_0^k\mathbf{v}$.

Now, let the minimal polynomial be $m(\lambda) = c_k\lambda^k + \cdots + c_1\lambda + c_0$. By its definition, we know that the matrix polynomial $m(A)$ is the zero matrix:

$$m(A) = c_kA^k + \cdots + c_1A + c_0I = 0$$

Let us apply this matrix polynomial to the eigenvector $\mathbf{v}$:

$$m(A)\mathbf{v} = (c_kA^k + \cdots + c_1A + c_0I)\mathbf{v}$$

$$= c_k(A^k\mathbf{v}) + \cdots + c_1(A\mathbf{v}) + c_0(I\mathbf{v})$$

Using the property we just established, we can substitute for each power of A:

$$= c_k(\lambda_0^k\mathbf{v}) + \cdots + c_1(\lambda_0\mathbf{v}) + c_0(\mathbf{v})$$

Factoring out the common vector $\mathbf{v}$:

$$= (c_k\lambda_0^k + \cdots + c_1\lambda_0 + c_0)\mathbf{v}$$

The expression in the parentheses is simply the scalar polynomial $m(\lambda)$ evaluated at the scalar λ_0. So, we have:

$$m(A)\mathbf{v} = m(\lambda_0)\mathbf{v}$$

Since $m(A)$ is the zero matrix, the left side of this equation is the zero vector, $\mathbf{0}$.

$$m(\lambda_0)\mathbf{v} = \mathbf{0}$$

We have a scalar, $m(\lambda_0)$, multiplying a non-zero vector, $\mathbf{v}$, resulting in the zero vector. This is only possible if the scalar itself is zero. Therefore, $m(\lambda_0) = 0$. This shows that any eigenvalue λ_0 of A must also be a root of the minimal polynomial $m(\lambda)$. ∎

Example 2.3.10 (Finding the Minimal Polynomial)**.** Find the minimal polynomial of the matrix $D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$.

Solution. First, we find the characteristic polynomial $p(\lambda)$.

$$p(\lambda) = \det(D - \lambda I) = \det \begin{pmatrix} 3-\lambda & 0 & 0 \\ 0 & 3-\lambda & 0 \\ 0 & 0 & 5-\lambda \end{pmatrix} = (3-\lambda)^2(5-\lambda) = -(\lambda-3)^2(\lambda-5)$$

The characteristic polynomial is $p(\lambda) = (\lambda-3)^2(\lambda-5)$. The minimal polynomial $m(\lambda)$ must have the same roots (3 and 5) and must divide $p(\lambda)$. The candidates for the minimal polynomial are:

1. $m_1(\lambda) = (\lambda-3)(\lambda-5) = \lambda^2 - 8\lambda + 15$
2. $m_2(\lambda) = (\lambda-3)^2(\lambda-5) = p(\lambda)$

We test the polynomial of the smallest degree first. We check if $m_1(D) = 0$.

$$m_1(D) = D^2 - 8D + 15I$$

$$D^2 = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 25 \end{pmatrix}$$

$$m_1(D) = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 25 \end{pmatrix} - 8\begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} + 15\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 9-24+15 & 0 & 0 \\ 0 & 9-24+15 & 0 \\ 0 & 0 & 25-40+15 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Since $m_1(D) = 0$ and $m_1(\lambda)$ is a monic polynomial of the least possible degree that annihilates D, the minimal polynomial of D is $m(\lambda) = (\lambda-3)(\lambda-5) = \lambda^2 - 8\lambda + 15$. ∎

Example 2.3.11. For the matrix A:

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

Find (a) its eigenvalues and corresponding eigenvectors, (b) its minimal polynomial, and (c) its inverse using the Cayley-Hamilton theorem.

Solution. (a) Eigenvalues and Eigenvectors Since A is an upper triangular matrix, its eigenvalues are the entries on its main diagonal. The eigenvalues are $\lambda_1 = 3$ and $\lambda_2 = 2$ (with an algebraic multiplicity of 2).

Now, we find the eigenvectors for each distinct eigenvalue by solving the system $(A - \lambda I)\mathbf{v} = \mathbf{0}$.

For $\lambda_1 = 3$: We solve the system $(A - 3I)\mathbf{v} = \mathbf{0}$.

$$\begin{pmatrix} 2-3 & 1 & 0 \\ 0 & 2-3 & 0 \\ 0 & 0 & 3-3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

This gives the equations $-x + y = 0$ and $-y = 0$. From the second equation, $y = 0$. Substituting into the first equation gives $x = 0$. The variable z is free. A simple non-zero choice is $z = 1$. The eigenvector corresponding to $\lambda_1 = 3$ is $\mathbf{v}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

For $\lambda_2 = 2$: We solve the system $(A - 2I)\mathbf{v} = \mathbf{0}$.

$$\begin{pmatrix} 2-2 & 1 & 0 \\ 0 & 2-2 & 0 \\ 0 & 0 & 3-2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

This gives the equations $y = 0$ and $z = 0$. The variable x is free. A simple non-zero choice is $x = 1$. The eigenvector corresponding to $\lambda_2 = 2$ is $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

Note: Even though $\lambda = 2$ is an eigenvalue with algebraic multiplicity 2, we can only find one linearly independent eigenvector for it. This means the matrix is not diagonalizable.

(b) Minimal Polynomial The characteristic polynomial $p(\lambda)$ is determined by the eigenvalues:

$$p(\lambda) = (\lambda - 3)(\lambda - 2)^2 = (\lambda - 3)(\lambda^2 - 4\lambda + 4) = \lambda^3 - 7\lambda^2 + 16\lambda - 12$$

The minimal polynomial $m(\lambda)$ must have the same roots (3 and 2) and must divide $p(\lambda)$. The candidates for the minimal polynomial are:

1. $m_1(\lambda) = (\lambda - 3)(\lambda - 2) = \lambda^2 - 5\lambda + 6$
2. $m_2(\lambda) = (\lambda - 3)(\lambda - 2)^2 = p(\lambda)$

We test the polynomial of the smallest degree, $m_1(\lambda)$. We check if $m_1(A) = A^2 - 5A + 6I$ is the zero matrix. First, we compute A^2:

$$A^2 = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

Now we evaluate $m_1(A)$:

$$A^2 - 5A + 6I = \begin{pmatrix} 4 & 4 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{pmatrix} - 5 \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} + 6 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 4-10+6 & 4-5+0 & 0 \\ 0 & 4-10+6 & 0 \\ 0 & 0 & 9-15+6 \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \neq 0$$

Since $m_1(A) \neq 0$, the minimal polynomial must be the next candidate, which is the characteristic polynomial itself. The minimal polynomial is $m(\lambda) = (\lambda - 3)(\lambda - 2)^2 = \lambda^3 - 7\lambda^2 + 16\lambda - 12$.

(c) Inverse using the Cayley-Hamilton Theorem By the Cayley-Hamilton theorem, $p(A) = 0$, so:

$$A^3 - 7A^2 + 16A - 12I = 0$$

The determinant is the constant term of the characteristic polynomial evaluated at $\lambda = 0$, so $\det(A) = -(-12) = 12$. Since the determinant is non-zero, the inverse exists. Multiply the equation by A^{-1}:

$$A^2 - 7A + 16I - 12A^{-1} = 0$$

Solving for A^{-1} gives:

$$12A^{-1} = A^2 - 7A + 16I$$

Using our previously calculated A^2:

$$12A^{-1} = \begin{pmatrix} 4 & 4 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{pmatrix} - 7 \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} + 16 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$12A^{-1} = \begin{pmatrix} 4 - 14 + 16 & 4 - 7 + 0 & 0 \\ 0 & 4 - 14 + 16 & 0 \\ 0 & 0 & 9 - 21 + 16 \end{pmatrix} = \begin{pmatrix} 6 & -3 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

Therefore, the inverse is:

$$A^{-1} = \frac{1}{12} \begin{pmatrix} 6 & -3 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 4 \end{pmatrix} = \begin{pmatrix} 1/2 & -1/4 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{pmatrix}$$

■

→ A matrix is **non-derogatory** if its minimal polynomial is identical to its characteristic polynomial.

→ A matrix is **derogatory** if its minimal polynomial is of a strictly smaller degree than its characteristic polynomial. This can only happen if the matrix has at least one repeated eigenvalue.

Example 2.3.12 (A Non-Derogatory Matrix)**.** Show that the following matrix A is non-derogatory.

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

Solution. Since A is an upper triangular matrix, its eigenvalues are its diagonal entries: $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3$. The characteristic polynomial $p(\lambda)$ is the product of the factors corresponding to these roots:

$$p(\lambda) = (\lambda - 1)(\lambda - 2)(\lambda - 3) = \lambda^3 - 6\lambda^2 + 11\lambda - 6$$

The degree of the characteristic polynomial is 3.

The minimal polynomial $m(\lambda)$ must have the same distinct roots as the characteristic polynomial. In this case, the roots are 1, 2, and 3. The monic polynomial of least degree containing these roots is:

$$m(\lambda) = (\lambda - 1)(\lambda - 2)(\lambda - 3)$$

The degree of this minimal polynomial is 3.

The minimal polynomial is identical to the characteristic polynomial (up to a sign, though both are typically taken as monic). Since the degree of the minimal polynomial (3) is equal to the degree of the characteristic polynomial (3), the matrix A is **non-derogatory**. This is always true for matrices with distinct eigenvalues. ■

Example 2.3.13 (A Derogatory Matrix)**.** Show that the following matrix B is derogatory.

$$B = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Solution. The matrix B is diagonal, so its eigenvalues are the diagonal entries: $\lambda_1 = 4$ (with algebraic multiplicity 2) and $\lambda_2 = 1$. The characteristic polynomial is:

$$p(\lambda) = (\lambda - 4)^2(\lambda - 1) = (\lambda^2 - 8\lambda + 16)(\lambda - 1) = \lambda^3 - 9\lambda^2 + 24\lambda - 16$$

The degree of the characteristic polynomial is 3.

The minimal polynomial $m(\lambda)$ must have the same distinct roots (4 and 1) and must divide $p(\lambda)$. The possible candidates for the minimal polynomial are:

1. $m_1(\lambda) = (\lambda - 4)(\lambda - 1) = \lambda^2 - 5\lambda + 4$

2. $m_2(\lambda) = (\lambda - 4)^2(\lambda - 1) = p(\lambda)$

We must test the candidate with the smallest degree first. Let's check if $m_1(B) = B^2 - 5B + 4I$ is the zero matrix. First, we compute B^2:

$$B^2 = \begin{pmatrix} 16 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Now, we evaluate the polynomial:

$$\begin{aligned} m_1(B) &= B^2 - 5B + 4I \\ &= \begin{pmatrix} 16 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 1 \end{pmatrix} - 5 \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix} + 4 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 16 - 20 + 4 & 0 & 0 \\ 0 & 16 - 20 + 4 & 0 \\ 0 & 0 & 1 - 5 + 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

Since $m_1(B) = 0$, the minimal polynomial of B is $m(\lambda) = \lambda^2 - 5\lambda + 4$.

The degree of the characteristic polynomial is 3, while the degree of the minimal polynomial is 2. Since the degree of the minimal polynomial is strictly less than the degree of the characteristic polynomial, the matrix B is **derogatory**. ■

Chapter 3

Rank and Elementary Transformations

3.1 Rank, Elementary Transformations, and Decomposition

3.1.1 Elementary Transformations

To simplify a matrix, we use three types of **elementary row operations**:

1. **Interchange:** Swapping two rows ($R_i \leftrightarrow R_j$).
2. **Scaling:** Multiplying a row by a non-zero scalar ($R_i \to cR_i$).
3. **Addition:** Adding a multiple of one row to another ($R_i \to R_i + cR_j$).

Elementary column operations are defined analogously.

3.1.2 Reduction to Echelon Form

By applying row operations, we can transform any matrix into a simpler structure where the rank is obvious.

Definition 3.1.1 (Row Echelon Form)**.** A matrix is in **row echelon form** if (1) all zero rows are at the bottom, and (2) the leading non-zero entry of any row is to the right of the leading entry of the row above it.

The rank of a matrix is the number of non-zero rows in its row echelon form.

Example 3.1.2 (Finding Rank via Echelon Form)**.** Find the rank of $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 3 \\ 1 & 3 & 2 \end{pmatrix}$.

Solution. We reduce the matrix to row echelon form.

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 3 \\ 1 & 3 & 2 \end{pmatrix} \xrightarrow[R_3 \to R_3 - R_1]{R_2 \to R_2 - 2R_1} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{R_3 \to R_3 - R_2} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

The echelon form has two non-zero rows. Therefore, the rank of A is 2. ∎

3.1.3 Reduction to Normal Form

By using both row and column operations, any matrix of rank r can be reduced to its **Normal Form**:

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

where I_r is the $r \times r$ identity matrix.

Example 3.1.3 (Finding the Normal Form)**.** Reduce $B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ to its normal form.

Solution.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \xrightarrow{R_2 \to R_2 - 3R_1} \begin{pmatrix} 1 & 2 \\ 0 & -2 \end{pmatrix} \xrightarrow{C_2 \to C_2 - 2C_1} \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix} \xrightarrow{C_2 \to -\frac{1}{2}C_2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The normal form is I_2, which confirms the rank is 2. ∎

Example 3.1.4 (Reduction of a Matrix to Normal Form)**.** Reduce the following matrix A to its normal form and determine its rank.

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 5 & 7 & 10 \\ 3 & 7 & 10 & 14 \\ 1 & 1 & 2 & 3 \end{pmatrix}$$

Solution. The goal is to use a combination of elementary row and column operations to transform the matrix into the form $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$. We will proceed by clearing the rows and columns around the pivot elements, starting from the top-left.

Clear the first row and first column The element in the (1,1) position is already a 1, which is convenient. First, we use this leading '1' to create zeros in the rest of the first column using row operations.

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 5 & 7 & 10 \\ 3 & 7 & 10 & 14 \\ 1 & 1 & 2 & 3 \end{pmatrix} \xrightarrow[R_4 \to R_4 - R_1]{R_2 \to R_2 - 2R_1,\ R_3 \to R_3 - 3R_1} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & -1 & -1 & -1 \end{pmatrix}$$

Next, we use the same leading '1' to create zeros in the rest of the first row using column operations.

$$\sim \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & -1 & -1 & -1 \end{pmatrix} \xrightarrow[C_4 \to C_4 - 4C_1]{C_2 \to C_2 - 2C_1,\ C_3 \to C_3 - 3C_1} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & -1 & -1 & -1 \end{pmatrix}$$

Clear the second row and second column Now we focus on the submatrix starting at the (2,2) position. The element at (2,2) is a 1. We use it to clear the entries below it and to its right. First, clear the rest of the second column using row operations.

$$\sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & -1 & -1 & -1 \end{pmatrix} \xrightarrow[R_4 \to R_4 + R_2]{R_3 \to R_3 - R_2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Next, clear the rest of the second row using column operations.

$$\sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow[C_4 \to C_4 - 2C_2]{C_3 \to C_3 - C_2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Final Arrangement The matrix is now a diagonal matrix, but not yet in the standard normal form. We need all the '1's to be in a contiguous block in the top-left. We can achieve this by swapping rows.

$$\sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_3 \leftrightarrow R_4} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Now, we swap columns to bring the final '1' into the (3,3) position.

$$\sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{C_3 \leftrightarrow C_4} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

This is the normal form of the matrix. We can write this in block form as:

$$\begin{pmatrix} I_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

The size of the identity matrix block is 3×3, which means $r = 3$. Therefore, the rank of the matrix A is 3. ■

3.1.4 Elementary Matrices

An **elementary matrix** is a matrix obtained by performing a single elementary operation on an identity matrix. Pre-multiplying by an elementary matrix performs a row operation; post-multiplying performs a column operation.

Solution. We apply the operation to the 3×3 identity matrix:

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_2 \to R_2 - 2R_1} E = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Applying this to A:

$$EA = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 7 & 11 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 3 & 7 & 11 \end{pmatrix}$$

This matches the first step of the row reduction in the first example. ■

3.1.5 Every Non-Singular Matrix is a Product of Elementary Matrices

This theorem connects invertibility to row operations.

Theorem 3.1.5. *Every non-singular matrix A can be expressed as a product of elementary matrices.*

Proof. Let A be an $n \times n$ non-singular matrix. By definition, a non-singular matrix has full rank (n), which means its reduced row echelon form is the identity matrix, I_n. This implies that there exists a finite sequence of elementary row operations that transforms A into I_n. Let the elementary matrices corresponding to these operations be $E_1, E_2, \ldots, E_k$. Applying these in order to A gives:

$$(E_k E_{k-1} \cdots E_2 E_1)A = I_n$$

This equation shows that the product of the elementary matrices is the inverse of A:

$$E_k E_{k-1} \cdots E_2 E_1 = A^{-1}$$

Since A is invertible, so is its inverse A^{-1}. We can take the inverse of both sides of this equation to solve for A:

$$A = (A^{-1})^{-1} = (E_k E_{k-1} \cdots E_2 E_1)^{-1}$$

Using the reversal law for the inverse of a product, we get:

$$A = E_1^{-1} E_2^{-1} \cdots E_{k-1}^{-1} E_k^{-1}$$

The inverse of any elementary matrix is also an elementary matrix. For example, the inverse of a row addition is a row subtraction, and the inverse of scaling by c is scaling by $1/c$. Therefore, we have successfully expressed A as a product of elementary matrices. ■

Example 3.1.6 (Decomposition into Elementary Matrices)**.** Express $C = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ as a product of elementary matrices.

Solution. We reduce C to the identity matrix. The only required operation is $R_1 \to R_1 - 2R_2$. The corresponding elementary matrix is $E = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$. Applying this operation gives $EC = I$. Therefore, $C = E^{-1}$. The inverse of the operation $R_1 \to R_1 - 2R_2$ is $R_1 \to R_1 + 2R_2$. The inverse elementary matrix is:

$$E^{-1} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

Thus, C is itself an elementary matrix. In this simple case, the "product" is just one matrix. ■

Example 3.1.7. Express the non-singular matrix A as a product of elementary matrices.

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

Solution. The process involves two main parts: first, we find the sequence of elementary matrices that reduces A to the identity matrix I. Second, we use the inverses of these elementary matrices to construct A.

Reduce A to the Identity Matrix We apply row operations sequentially, identifying the elementary matrix at each step.

Start with $A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$. The first operation is $R_3 \to R_3 - R_1$. The elementary matrix for this is $E_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$.

$$E_1 A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & -2 & 1 \end{pmatrix}$$

The second operation is $R_3 \to R_3 + 2R_2$. The elementary matrix is $E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}$.

$$E_2(E_1 A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & -2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 3 \end{pmatrix}$$

The third operation is $R_3 \to \frac{1}{3}R_3$. The elementary matrix is $E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/3 \end{pmatrix}$.

$$E_3(E_2E_1A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

The fourth operation is $R_2 \to R_2 - R_3$. The elementary matrix is $E_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$.

$$E_4(E_3E_2E_1A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The final operation is $R_1 \to R_1 - 2R_2$. The elementary matrix is $E_5 = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

$$E_5(E_4E_3E_2E_1A) = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I$$

The complete reduction is given by the equation $E_5E_4E_3E_2E_1A = I$. Solving for A gives:

$$A = (E_5E_4E_3E_2E_1)^{-1} = E_1^{-1}E_2^{-1}E_3^{-1}E_4^{-1}E_5^{-1}$$

Find the Inverse Elementary Matrices We find the inverse of each elementary matrix by finding the inverse of its corresponding operation.

- E_1^{-1} (inverse of $R_3 \to R_3 - R_1$) is $R_3 \to R_3 + R_1$: $E_1^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$.
- E_2^{-1} (inverse of $R_3 \to R_3 + 2R_2$) is $R_3 \to R_3 - 2R_2$: $E_2^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix}$.
- E_3^{-1} (inverse of $R_3 \to \frac{1}{3}R_3$) is $R_3 \to 3R_3$: $E_3^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$.
- E_4^{-1} (inverse of $R_2 \to R_2 - R_3$) is $R_2 \to R_2 + R_3$: $E_4^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$.
- E_5^{-1} (inverse of $R_1 \to R_1 - 2R_2$) is $R_1 \to R_1 + 2R_2$: $E_5^{-1} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Conclusion The matrix A can be written as the following product of elementary matrices, in the reverse order of their inverses:

$$A = E_1^{-1}E_2^{-1}E_3^{-1}E_4^{-1}E_5^{-1}$$

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

This product, when multiplied out, will yield the original matrix A. ■

3.2 The Rank of a Matrix and its Properties

3.2.1 Rank of a Matrix

The rank of a matrix is arguably its most important single numerical property. It provides deep insight into the structure of the matrix, the nature of its associated linear transformation, and the solvability of linear systems.

Definition 3.2.1 (Rank)**.** The rank of a matrix is the order r of the largest non-singular (i.e., non-zero determinant) square submatrix that can be formed from it. This is equivalent to saying the rank is the order of the highest-order non-vanishing minor.

Theorem 3.2.2. *Elementary row (or column) transformations do not alter the rank of a matrix.*

Proof. The rank of a matrix is the dimension of its row space, which is the space spanned by its row vectors. We will show that each of the three elementary row operations preserves the row space, and therefore its dimension (the rank). Let the original row space be $S = \text{span}\{R_1, \ldots, R_m\}$.

1. **Interchange ($R_i \leftrightarrow R_j$):** This operation merely reorders the vectors in the spanning set. Since the set of vectors is identical, the space they span remains the same. Thus, the rank is unchanged.

2. **Scaling ($R_i \to cR_i$, where $c \neq 0$):** Let the new row space be $S' = \text{span}\{R_1, \ldots, cR_i, \ldots, R_m\}$. Any vector in S' is a linear combination of the new rows, and is therefore also a linear combination of the original rows. Conversely, any vector in S can be expressed as a combination of the new rows, since the original row R_i can be written as $R_i = \frac{1}{c}(cR_i)$. This is valid because $c \neq 0$. Since the row spaces are identical, the rank is preserved.

3. **Addition ($R_i \to R_i + cR_j$):** Let the new row space be $S' = \text{span}\{R_1, \ldots, R_i + cR_j, \ldots, R_m\}$. Any vector in S' is clearly a linear combination of the original rows. Conversely, the original row R_i can be recovered from the new set via $R_i = (R_i + cR_j) - cR_j$. This shows that any vector in the original space S is also in the new space S'. The row spaces are identical, so the rank is unchanged.

Since all three elementary row operations preserve the rank, any sequence of such operations will also preserve the rank. The proof for column operations is analogous. ■

In practice, the rank is most often computed by reducing a matrix to its **row echelon form** through elementary row operations. The rank is then simply the number of non-zero rows in the resulting echelon form matrix.

Example 3.2.3 (Finding Rank by Reduction to Echelon Form)**.** Find the rank of the following 3×4 matrix by reducing it to its row echelon form.

$$A = \begin{pmatrix} 1 & 2 & -1 & 3 \\ 2 & 4 & 1 & 2 \\ 3 & 6 & 0 & 5 \end{pmatrix}$$

Example 3.2.4 (Rank of a Matrix by Elementary Transformations)**.** Find the rank of the matrix A by reducing it to row echelon form.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 7 & 11 \end{pmatrix}$$

Solution. We apply elementary row operations to find the row echelon form.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 7 & 11 \end{pmatrix} \xrightarrow[R_3 \to R_3 - 3R_1]{R_2 \to R_2 - 2R_1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{pmatrix} \xrightarrow{R_3 \to R_3 - R_2} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

The resulting echelon form has two non-zero rows. Therefore, the rank of A is 2. ■

Solution. Our goal is to use elementary row operations to introduce zeros below the leading entries of each row, thereby transforming the matrix into row echelon form.

First, we use the leading '1' in the first row to create zeros in the first column of the second and third rows.

$$A = \begin{pmatrix} 1 & 2 & -1 & 3 \\ 2 & 4 & 1 & 2 \\ 3 & 6 & 0 & 5 \end{pmatrix} \xrightarrow[R_3 \to R_3 - 3R_1]{R_2 \to R_2 - 2R_1} \begin{pmatrix} 1 & 2 & -1 & 3 \\ 0 & 0 & 3 & -4 \\ 0 & 0 & 3 & -4 \end{pmatrix}$$

Now, the second and third rows are identical. We can use the second row to create a zero row in the third row. The leading entry of the second row is the '3' in the third column.

$$\sim \begin{pmatrix} 1 & 2 & -1 & 3 \\ 0 & 0 & 3 & -4 \\ 0 & 0 & 3 & -4 \end{pmatrix} \xrightarrow{R_3 \to R_3 - R_2} \begin{pmatrix} 1 & 2 & -1 & 3 \\ 0 & 0 & 3 & -4 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The matrix is now in row echelon form. To find the rank, we count the number of non-zero rows. There are two non-zero rows. Therefore, the rank of the matrix A is 2. ■

3.2.2 Rank of a Product of Two Matrices

The rank of a product of matrices is not, in general, the product of their ranks. Instead, it is constrained by the ranks of the individual factor matrices according to a fundamental inequality.

Theorem 3.2.5 (Sylvester's Rank Inequality)**.** *If A and B are matrices such that their product AB is defined, then the rank of the product is less than or equal to the rank of each individual matrix. That is,*

$$\mathit{rank}(AB) \le \min(\mathit{rank}(A), \mathit{rank}(B))$$

Proof. We will prove the two parts of the inequality separately: (i) $\text{rank}(AB) \le \text{rank}(A)$ and (ii) $\text{rank}(AB) \le \text{rank}(B)$.

Proof of rank$(AB) \le$ rank(A): The columns of the product matrix AB are linear combinations of the columns of A. Specifically, if $\mathbf{b}_j$ are the columns of B, then the columns of AB are of the form $A\mathbf{b}_j$. Any such vector is, by definition, an element of the column space of A. This implies that the entire column space of AB is a subspace of the column space of A. Since the dimension of a subspace cannot exceed the dimension of the space containing it, we have $\dim(\text{col}(AB)) \le \dim(\text{col}(A))$. By the definition of rank, this means $\text{rank}(AB) \le \text{rank}(A)$.

Proof of rank$(AB) \le$ rank(B): Similarly, the rows of the product matrix AB are linear combinations of the rows of B. If $\mathbf{a}_i$ are the rows of A, the rows of AB are of the form $\mathbf{a}_i B$. This implies that the row space of AB is a subspace of the row space of B. Comparing their dimensions gives $\dim(\text{row}(AB)) \le \dim(\text{row}(B))$, which means $\text{rank}(AB) \le \text{rank}(B)$.

Since $\text{rank}(AB)$ is less than or equal to both individual ranks, it must be less than or equal to their minimum. ∎

A very important consequence of this theorem is that multiplying by a non-singular matrix does not alter the rank.

Theorem 3.2.6. *If A is any matrix and P and Q are non-singular matrices of appropriate size, then:*

$$\mathit{rank}(PA) = \mathit{rank}(AQ) = \mathit{rank}(A)$$

Proof. Let's prove $\text{rank}(AQ) = \text{rank}(A)$. From Sylvester's inequality, we already know that $\text{rank}(AQ) \le \text{rank}(A)$. To show equality, we must prove the reverse inequality. We can write $A = (AQ)Q^{-1}$, which is valid since Q is non-singular. Now we apply Sylvester's inequality to this new product:

$$\text{rank}(A) = \text{rank}((AQ)Q^{-1}) \le \text{rank}(AQ)$$

Since we have both $\text{rank}(AQ) \le \text{rank}(A)$ and $\text{rank}(A) \le \text{rank}(AQ)$, we must conclude that the ranks are equal. The proof for $\text{rank}(PA)$ is analogous. ∎

Example 3.2.7 (Rank of a Product of Two Matrices)**.** Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$. Verify that $\text{rank}(AB) \le \min(\text{rank}(A), \text{rank}(B))$.

Solution. By inspection, matrix A has one non-zero row, so $\text{rank}(A) = 1$. Matrix B has one non-zero row, so $\text{rank}(B) = 1$. The minimum of the ranks is $\min(1, 1) = 1$. The product is:

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

The rank of the zero matrix is $\text{rank}(AB) = 0$. The inequality holds: $0 \le 1$. ∎

3.3 Linear Independence and its Consequences

3.3.1 Linear Combinations, Dependence, and Independence

These concepts form the bedrock for understanding the structure of a matrix and its associated vector spaces.

Definition 3.3.1 (Linear Combination)**.** A vector $\mathbf{w}$ is a **linear combination** of a set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ if it can be expressed as $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$ for some scalars c_i.

Definition 3.3.2 (Linear Dependence)**.** A set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is **linearly dependent** if there exist scalars $c_1, \dots, c_k$, not all zero, such that their linear combination equals the zero vector:

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

This means at least one vector in the set is redundant and can be written as a linear combination of the others.

Definition 3.3.3 (Linear Independence)**.** A set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is **linearly independent** if the only solution to the equation $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$ is the trivial solution where all scalars are zero ($c_1 = c_2 = \cdots = c_k = 0$). This means every vector is essential.

These definitions apply directly to the row vectors and column vectors of any matrix.

3.3.2 Linear Dependence and the Null Space

The linear dependence of a matrix's columns is directly equivalent to the existence of a non-trivial solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$.

Theorem 3.3.4. *The columns of a matrix A are linearly dependent if and only if there exists a vector $\mathbf{x} \neq \mathbf{0}$ such that $A\mathbf{x} = \mathbf{0}$.*

Proof. Let the columns of A be $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ and let $\mathbf{x} = [x_1, \dots, x_n]^T$. By definition, the product $A\mathbf{x}$ is the linear combination $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n$. First, assume the columns are linearly dependent. Then there exist scalars $c_1, \dots, c_n$, not all zero, such that $c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n = \mathbf{0}$. We can construct a non-zero vector $\mathbf{x} = [c_1, \dots, c_n]^T$. For this $\mathbf{x}$, we have $A\mathbf{x} = \mathbf{0}$. Conversely, assume there exists a non-zero vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$. This means $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$. Since $\mathbf{x} \neq \mathbf{0}$, at least one of the scalars x_i is non-zero. This is a non-trivial linear combination of the columns that equals the zero vector, which proves that the columns of A are linearly dependent. ■

Example 3.3.5 (Linear Combination of Vectors)**.** Express the vector $\mathbf{w} = (4, 5)$ as a linear combination of the vectors $\mathbf{v}_1 = (1, 1)$ and $\mathbf{v}_2 = (2, 1)$.

Solution. We need to find scalars c_1, c_2 such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{w}$.

$$c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$$

This gives the system $c_1 + 2c_2 = 4$ and $c_1 + c_2 = 5$. Subtracting the second equation from the first gives $c_2 = -1$. Substituting back gives $c_1 + (-1) = 5 \implies c_1 = 6$. The linear combination is $6\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{w}$. ■

Example 3.3.6 (Linear Dependence and Independence of Vectors)**.** Determine if the column vectors of $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ are linearly dependent or independent.

Solution. The matrix A is already in row echelon form. It has 3 non-zero rows, so its rank is 3. Since the rank (3) is equal to the number of columns (3), the column vectors are linearly independent. ■

Example 3.3.7 (Column Dependence and $A\mathbf{x} = \mathbf{0}$)**.** Show that the columns of $A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 4 \end{pmatrix}$ are linearly dependent by finding a non-zero vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$.

Solution. We solve the homogeneous system by row reducing A:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 4 \end{pmatrix} \xrightarrow{R_2 \to R_2 - R_1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \end{pmatrix}$$

This gives the system $x_1 + 2x_2 + 3x_3 = 0$ and $x_2 + x_3 = 0$. Let the free variable be $x_3 = t$. Then $x_2 = -t$. Substituting back gives $x_1 + 2(-t) + 3t = 0 \implies x_1 + t = 0 \implies x_1 = -t$. A non-trivial solution can be found by setting $t = 1$, which gives $\mathbf{x} = (-1, -1, 1)^T$. Since a non-zero vector $\mathbf{x}$ exists, the columns of A are linearly dependent. ■

3.3.3 Linear Dependence and Rank

The concept of rank provides a simple numerical test for the dependence of a matrix's columns.

Theorem 3.3.8. *The columns of a matrix A of order $m \times n$ are linearly dependent if and only if the rank of A is less than n.*

Proof. The rank of a matrix, $\text{rank}(A)$, is defined as the maximum number of linearly independent columns of A. Let this be r. First, assume the n columns of A are linearly dependent. This means that the entire set of n columns cannot form a linearly independent set. Therefore, the maximum number of linearly independent columns, r, must be strictly less than n. So, $\text{rank}(A) < n$. Conversely, assume $\text{rank}(A) < n$. Let the rank be r, where $r < n$. The column space of A has dimension r. We have n column vectors that all belong to this r-dimensional space. Since there are more vectors (n) than the dimension of the space they occupy (r), a fundamental theorem of linear algebra states that this set of n vectors must be linearly dependent. ■

Example 3.3.9 (Checking Column Vectors for Linear Dependence)**.** Determine if the column vectors of the following matrix A are linearly dependent or independent. If they are dependent, find a non-trivial linear combination that equals the zero vector.

$$A = \begin{pmatrix} 1 & 1 & 3 \\ 1 & 2 & 4 \\ -1 & 1 & -1 \end{pmatrix}$$

Solution. To determine the linear independence of the **columns**, we need to find the rank of the matrix and compare it to the number of columns. We find the rank by reducing the matrix to its row echelon form.

$$\begin{pmatrix} 1 & 1 & 3 \\ 1 & 2 & 4 \\ -1 & 1 & -1 \end{pmatrix} \xrightarrow[R_3 \to R_3 + R_1]{R_2 \to R_2 - R_1} \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 1 \\ 0 & 2 & 2 \end{pmatrix} \xrightarrow{R_3 \to R_3 - 2R_2} \begin{pmatrix} \mathbf{1} & 1 & 3 \\ 0 & \mathbf{1} & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

The echelon form has leading entries (pivots) in the first and second columns. The number of pivot columns is 2. Therefore, the rank of the matrix is $\text{rank}(A) = 2$. Since the rank (2) is less than the number of columns (3), the column vectors are **linearly dependent**.

To find the specific dependence relationship, we solve the homogeneous system $A\mathbf{c} = \mathbf{0}$. The echelon form gives us the system:

$$\begin{aligned} c_1 + c_2 + 3c_3 &= 0 \\ c_2 + c_3 &= 0 \end{aligned}$$

The third column is a non-pivot column, so c_3 is a free variable. Let $c_3 = 1$. The second equation gives $c_2 = -c_3 = -1$. The first equation gives $c_1 = -c_2 - 3c_3 = -(-1) - 3(1) = 1 - 3 = -2$. The non-trivial combination is $-2\mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_3 = \mathbf{0}$. ■

Example 3.3.10 (Checking Row Vectors for Linear Independence)**.** Determine if the row vectors of the following matrix B are linearly dependent or independent.

$$B = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

Solution. To determine the linear independence of the **rows**, we need to find the rank of the matrix and compare it to the number of rows. We find the rank by reducing the matrix to row echelon form.

$$\begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{R_2 \to R_2 - 2R_1} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -4 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{R_3 \to R_3 - R_2} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -4 \\ 0 & 0 & 5 \end{pmatrix}$$

The resulting echelon form has 3 non-zero rows. Therefore, the rank of the matrix is $\text{rank}(B) = 3$. Since the rank (3) is equal to the number of row vectors (3), the row vectors are **linearly independent**. ■

Theorem 3.3.11. *A matrix A has rank r if and only if it possesses a set of r linearly independent columns, and any set of s columns where $s > r$ is linearly dependent.*

Proof. This is a biconditional statement, so we must prove both directions of the implication. The proof relies on the definition of rank as the dimension of the column space of the matrix. Let $\text{col}(A)$ denote the column space of A.

Forward Direction ($\Rightarrow$): Assume that the matrix A has rank r. By definition, this means that the dimension of the column space of A is r. That is, $\dim(\text{col}(A)) = r$. From the definition of the dimension of a vector space, we know two things:

1. There exists a basis for the column space $\text{col}(A)$ that consists of exactly r vectors. Since a basis is a set of linearly independent vectors that spans the space, these r vectors are a set of r linearly independent columns from the matrix A. This satisfies the first part of the condition.

2. Any set of vectors in an r-dimensional space containing more than r vectors must be linearly dependent.

Now, consider any set of s columns of A, where $s > r$. Each of these s columns is a vector in the column space $\text{col}(A)$. Since $\text{col}(A)$ is an r-dimensional space and we have a set of s vectors from it with $s > r$, this set must be linearly dependent. This satisfies the second part of the condition. Thus, if the rank of A is r, the conditions of the theorem hold.

Reverse Direction ($\Leftarrow$): Now, assume that A has a set of r linearly independent columns, and that any set of s columns where $s > r$ is linearly dependent. We must show that the rank of A is r.

Let the maximum number of linearly independent columns in A be k. By definition, this number k is the rank of the matrix A, so $\text{rank}(A) = k$. From our assumption, we know there is at least one set of r linearly independent columns. This means that the maximum number of linearly independent columns must be at least r. So, $k \geq r$.

From our assumption, we also know that any set of $s > r$ columns is linearly dependent. This means that we cannot find a set of $r+1$ linearly independent columns. Therefore, the maximum number of linearly independent columns cannot be greater than r. This means $k \leq r$.

Since we have shown that $k \geq r$ and $k \leq r$, the only possibility is that $k = r$. Therefore, the rank of the matrix A is exactly r. ■

Example 3.3.12 (Column Dependence and Rank $< n$)**.** Show that the columns of the 3×4 matrix $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 7 & 9 \\ 3 & 6 & 10 & 13 \end{pmatrix}$ are linearly dependent.

Solution. The matrix A is of order 3×4. The rank of a matrix cannot exceed the number of its rows or columns. Therefore, the maximum possible rank of A is $\text{rank}(A) \leq \min(3, 4) = 3$. The number of columns is $n = 4$. Since the rank of A is at most 3, we have $\text{rank}(A) < 4$. Because the rank of the matrix is strictly less than the number of columns, the columns must be linearly dependent. ■

Example 3.3.13 (Rank and Number of Independent Columns)**.** Demonstrate that the matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{pmatrix}$ has rank 2, and show that it has 2 linearly independent columns but any 3 columns are linearly dependent.

Solution. First, we find the rank by row reduction:

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{pmatrix} \xrightarrow{R_3 \to R_3 - R_1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{R_3 \to R_3 - R_2} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

The rank is 2. This means the maximum number of linearly independent columns is 2. The first two columns, $\mathbf{c}_1 = (1, 0, 1)^T$ and $\mathbf{c}_2 = (2, 1, 3)^T$, are not multiples of each other, so they are a set of 2 linearly independent columns. The set of all 3 columns is linearly dependent because the rank of the matrix (2) is less than the number of columns (3), as required by the theorem. Specifically, we can see that $\mathbf{c}_3 = \mathbf{c}_1 + \mathbf{c}_2$. ■

Chapter 4

Systems of Linear Equations and Inner Products

4.1 Systems of Linear Equations

A general system of m linear equations in n unknowns $(x_1, x_2, \dots, x_n)$ can be written as:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

where the coefficients a_{ij} and the constants b_i are scalars.

A system of m linear equations in n variables can be compactly represented in matrix form as:

$$A\mathbf{x} = \mathbf{b}$$

where A is an $m \times n$ matrix of coefficients, $\mathbf{x}$ is an $n \times 1$ column vector of variables, and $\mathbf{b}$ is an $m \times 1$ column vector of constants. We categorize these systems into two types based on the vector $\mathbf{b}$.

4.1.1 The Non-Homogeneous System: $A\mathbf{x} = \mathbf{b}$

Definition 4.1.1 (Non-Homogeneous System)**.** A system of linear equations $A\mathbf{x} = \mathbf{b}$ is called **non-homogeneous** if the vector of constants $\mathbf{b}$ is not the zero vector (i.e., $\mathbf{b} \neq \mathbf{0}$).

The primary question for a non-homogeneous system is whether a solution exists at all. A system that has at least one solution is called **consistent**; otherwise, it is **inconsistent**. The rank of the matrix provides a definitive test for consistency.

Theorem 4.1.2 (Rouché-Capelli Theorem for Consistency)**.** *The system of equations $A\mathbf{x} = \mathbf{b}$ is consistent if and only if the rank of the coefficient matrix A is equal to the rank of the augmented matrix $[A \mid \mathbf{b}]$.*

$$\textit{rank}(A) = \textit{rank}([A \mid \mathbf{b}])$$

Proof. The system $A\mathbf{x} = \mathbf{b}$ has a solution if and only if the vector $\mathbf{b}$ can be expressed as a linear combination of the columns of A. This is the definition of $\mathbf{b}$ being in the column space of A. The column space of the augmented matrix $[A \mid \mathbf{b}]$ is spanned by the columns of A plus the additional vector $\mathbf{b}$. If the system is consistent, then $\mathbf{b}$ is already in the column space of A. Adding $\mathbf{b}$ to the set of column vectors of A does not introduce a new linearly independent vector, so the dimension of the column space does not increase. This implies $\mathrm{rank}(A) = \mathrm{rank}([A \mid \mathbf{b}])$. Conversely, if $\mathrm{rank}(A) = \mathrm{rank}([A \mid \mathbf{b}])$, the dimensions of the two column spaces are equal. Since the column space of A is a subspace of the column space of $[A \mid \mathbf{b}]$, their equality in dimension implies the spaces are identical. Therefore, $\mathbf{b}$ must be in the column space of A, which guarantees that a solution exists. ■

Once consistency is established, the rank can further tell us about the number of solutions.

Theorem 4.1.3 (Nature of Solutions)**.** *If a non-homogeneous system $A\mathbf{x} = \mathbf{b}$ is consistent and $\mathrm{rank}(A) = r$, where n is the number of variables (columns), then:*

1. *The system has a **unique solution** if $r = n$.*
2. *The system has **infinitely many solutions** if $r < n$. The number of free parameters in the solution will be $n - r$.*

Proof. If the system is consistent, let $\mathbf{x}_p$ be any particular solution, so $A\mathbf{x}_p = \mathbf{b}$. Any other solution $\mathbf{x}$ can be written as $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$, where $\mathbf{x}_h$ is a solution to the corresponding homogeneous system $A\mathbf{x}_h = \mathbf{0}$. If $r = n$, the n columns of A are linearly independent. This implies that the only solution to $A\mathbf{x}_h = \mathbf{0}$ is the trivial solution $\mathbf{x}_h = \mathbf{0}$. Therefore, the only solution is $\mathbf{x} = \mathbf{x}_p$, which is unique. If $r < n$, the n columns of A are linearly dependent, which means there are non-trivial solutions to $A\mathbf{x}_h = \mathbf{0}$. Any vector of the form $\mathbf{x} = \mathbf{x}_p + c\mathbf{x}_h$ for any scalar c will be a solution. Since there are infinitely many choices for c, there are infinitely many solutions. The number of linearly independent vectors $\mathbf{x}_h$ forming the basis of the solution space for the homogeneous system is $n - r$, which corresponds to the number of free variables. ∎

Example 4.1.4 (Consistency and Unique Solution)**.** Determine if the following system is consistent, and if so, find its solution.

$$\begin{aligned} x + y + z &= 3 \\ x + 2y + 2z &= 5 \\ 2x + 3y + 4z &= 9 \end{aligned}$$

Solution. We form the augmented matrix $[A \mid \mathbf{b}]$ and reduce it to row echelon form.

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & 2 & 2 & 5 \\ 2 & 3 & 4 & 9 \end{array}\right) \xrightarrow[R_3 \to R_3 - 2R_1]{R_2 \to R_2 - R_1} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 2 & 3 \end{array}\right) \xrightarrow{R_3 \to R_3 - R_2} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{array}\right)$$

From the echelon form, we see that the rank of the coefficient matrix A is 3, and the rank of the augmented matrix $[A \mid \mathbf{b}]$ is also 3. Since $\text{rank}(A) = \text{rank}([A \mid \mathbf{b}])$, the system is consistent. Furthermore, since the rank (3) is equal to the number of variables (3), the solution is unique. By back substitution, the last row gives $z = 1$. The second row gives $y + z = 2 \implies y + 1 = 2 \implies y = 1$. The first row gives $x + y + z = 3 \implies x + 1 + 1 = 3 \implies x = 1$. The unique solution is $\mathbf{x} = (1, 1, 1)^T$. ∎

Example 4.1.5 (Infinite Solutions)**.** Find the complete solution set for the system: $x - y + z = 2$, $x + y + 2z = 0$, $3x - y + 4z = 4$.

Solution. We reduce the augmented matrix for the system.

$$\left(\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 1 & 1 & 2 & 0 \\ 3 & -1 & 4 & 4 \end{array}\right) \xrightarrow[R_3 \to R_3 - 3R_1]{R_2 \to R_2 - R_1} \left(\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 0 & 2 & 1 & -2 \\ 0 & 2 & 1 & -2 \end{array}\right) \xrightarrow{R_3 \to R_3 - R_2} \left(\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 0 & 2 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

The rank of both the coefficient matrix and the augmented matrix is 2. Since this is less than the number of variables ($r = 2 < n = 3$), the system is consistent and has infinitely many solutions. The variable z is a free variable; let $z = t$. The second row gives $2y + z = -2 \implies 2y = -2 - t \implies y = -1 - \frac{1}{2}t$. The first row gives $x - y + z = 2 \implies x = 2 + y - z = 2 + (-1 - \frac{1}{2}t) - t = 1 - \frac{3}{2}t$. The solution set is $\mathbf{x} = (1 - \frac{3}{2}t, -1 - \frac{1}{2}t, t)^T$ for any $t \in \mathbb{R}$. ∎

Example 4.1.6 (Inconsistent System)**.** Show that the system $x + 2y - z = 1$, $2x + 4y - 2z = 3$ has no solution.

Solution. We form the augmented matrix and row-reduce.

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 2 & 4 & -2 & 3 \end{array}\right) \xrightarrow{R_2 \to R_2 - 2R_1} \left(\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

From the echelon form, the rank of the coefficient matrix A is 1, while the rank of the augmented matrix $[A \mid \mathbf{b}]$ is 2. Since $\text{rank}(A) \neq \text{rank}([A \mid \mathbf{b}])$, the system is inconsistent. The last row corresponds to the contradictory equation $0x + 0y + 0z = 1$. Therefore, no solution exists. ∎

4.1.2 The Homogeneous System: $A\mathbf{x} = \mathbf{0}$

Definition 4.1.7 (Homogeneous System)**.** A system of linear equations $A\mathbf{x} = \mathbf{b}$ is called **homogeneous** if the vector of constants is the zero vector, i.e., $A\mathbf{x} = \mathbf{0}$.

A homogeneous system is *always* consistent, because $\mathbf{x} = \mathbf{0}$ (the **trivial solution**) is always a solution. The critical question is whether other, non-zero solutions exist. These are called **non-trivial solutions**.

Theorem 4.1.8 (Existence of Non-Trivial Solutions)**.** *The homogeneous system $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix, has a non-trivial solution if and only if the rank of A is less than the number of variables n.*

$$\mathit{rank}(A) < n$$

Proof. A non-trivial solution exists if and only if there is a non-zero vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$. The product $A\mathbf{x}$ is a linear combination of the columns of A, with the components of $\mathbf{x}$ as coefficients. The existence of a non-zero $\mathbf{x}$ making this combination zero is, by definition, the condition for the columns of A to be linearly dependent. A set of n column vectors is linearly dependent if and only if the dimension of the space they span (the rank) is strictly less than the number of vectors, n. ∎

The set of all solutions to a homogeneous system forms a vector space, known as the **null space** of the matrix A. The dimension of this space is given by the Rank-Nullity Theorem.

Theorem 4.1.9 (Dimension of the Solution Space). *For an $m \times n$ matrix A with rank$(A) = r$, the number of linearly independent solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$ is $n - r$.*

Proof. By reducing the matrix A to its reduced row echelon form, we do not change the solution set. In this form, there are r pivot variables and $n - r$ free variables. We can construct a basis for the null space by creating $n - r$ solution vectors, where for each vector we set one free variable to 1 and the others to 0. This set of $n - r$ vectors is guaranteed to be linearly independent and can be shown to span the entire solution space. Therefore, the dimension of the solution space is $n - r$. ■

Corollary 4.1.10 (Dimension of the Solution Space). *For an $m \times n$ matrix A with rank$(A) = r$, the number of linearly independent solutions to the non-homogeneous system $A\mathbf{x} = \mathbf{b}$ is $n - r + 1$.*

Proof. Assuming the system being consistent, let s be a solution. Then $A\mathbf{s} = b$ or equivalently, $A(\mathbf{s} - \mathbf{b}) = \mathbf{0}$ is a homogeneous system which has $n - r$ linearly independent solutions. Hence the total number of linearly independent solution to the system $A\mathbf{x} = \mathbf{b}$ is $n - r + 1$. ■

Example 4.1.11 (Homogeneous System with Non-Trivial Solutions). Find the solution space for the homogeneous system $A\mathbf{x} = \mathbf{0}$, where $A = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 2 \\ 3 & -1 & 4 \end{pmatrix}$.

Solution. We reduce the coefficient matrix A to its echelon form.

$$\begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 2 \\ 3 & -1 & 4 \end{pmatrix} \xrightarrow[R_3 \to R_3 - 3R_1]{R_2 \to R_2 - R_1} \begin{pmatrix} 1 & -1 & 1 \\ 0 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \xrightarrow{R_3 \to R_3 - R_2} \begin{pmatrix} 1 & -1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

The rank of A is 2. Since the rank is less than the number of variables ($r = 2 < n = 3$), there are non-trivial solutions. The number of linearly independent solutions is $n - r = 3 - 2 = 1$. Let the free variable be $z = t$. The second row gives $2y + z = 0 \implies y = -\frac{1}{2}t$. The first row gives $x - y + z = 0 \implies x = y - z = -\frac{1}{2}t - t = -\frac{3}{2}t$. The solution space is the set of all vectors of the form $\mathbf{x} = (-\frac{3}{2}t, -\frac{1}{2}t, t)^T$, which is spanned by the basis vector $(-3, -1, 2)^T$. ■

Example 4.1.12 (Homogeneous System with Only the Trivial Solution). Solve the system $x+y = 0$, $y+z = 0$, $x+z = 0$.

Solution. The coefficient matrix is $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$. We find its rank.

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \xrightarrow{R_3 \to R_3 - R_1} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{pmatrix} \xrightarrow{R_3 \to R_3 + R_2} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

The rank of the matrix is 3. Since the rank is equal to the number of variables ($r = n = 3$), the homogeneous system has only the trivial solution. Thus, the only solution is $\mathbf{x} = (0, 0, 0)^T$. ■

4.2 Inner Product Spaces and Special Matrices

We now introduce geometric structure into vector spaces through the inner product, which allows us to define concepts like length and orthogonality. This leads to the study of special matrices that preserve this structure.

4.2.1 Inner Product, Length, and Normal Vectors

Definition 4.2.1 (Standard Inner Product). Let $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$ be two vectors in $\mathbb{R}^n$. The **standard inner product** (or dot product) of $\mathbf{u}$ and $\mathbf{v}$, denoted $\langle \mathbf{u}, \mathbf{v} \rangle$, is the scalar:

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

The inner product satisfies key properties such as symmetry, linearity, and positive definiteness. The most important of these for defining length is positive definiteness: $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

Definition 4.2.2 (Length or Norm). The **length** (or **norm**) of a vector $\mathbf{v} \in \mathbb{R}^n$, denoted $\|\mathbf{v}\|$, is the non-negative square root of the inner product of $\mathbf{v}$ with itself:

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

In many applications, only the direction of a vector is important, not its magnitude. This leads to the concept of a unit vector.

Definition 4.2.3 (Normal or Unit Vector)**.** A vector $\mathbf{u} \in \mathbb{R}^n$ is a **unit vector** (or a **normal vector**) if its norm is 1. That is, $\|\mathbf{u}\| = 1$.

Any non-zero vector can be converted into a unit vector through a process called normalization.

Theorem 4.2.4 (Normalization)**.** *If* $\mathbf{v}$ *is any non-zero vector, then the vector* $\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$ *is a unit vector in the same direction as* $\mathbf{v}$*.*

Proof. To prove $\mathbf{u}$ is a unit vector, we compute its norm. Let the scalar be $c = 1/\|\mathbf{v}\|$. Since $\mathbf{v} \neq \mathbf{0}$, c is a positive real number.

$$\|\mathbf{u}\| = \left\| \frac{1}{\|\mathbf{v}\|}\mathbf{v} \right\| = \left| \frac{1}{\|\mathbf{v}\|} \right| \|\mathbf{v}\| = \frac{1}{\|\mathbf{v}\|}\|\mathbf{v}\| = 1$$

Thus, $\mathbf{u}$ is a unit vector. ■

4.2.2 Orthogonal and Unitary Matrices

These are special square matrices that preserve lengths and inner products under transformation. Orthogonal matrices operate on real vector spaces, while unitary matrices are their generalization to complex vector spaces.

Definition 4.2.5 (Orthogonal Matrix)**.** A square matrix P with real entries is **orthogonal** if its transpose is equal to its inverse:

$$P^T = P^{-1} \quad \text{or equivalently,} \quad P^T P = I$$

Definition 4.2.6 (Unitary Matrix)**.** A square matrix U with complex entries is **unitary** if its conjugate transpose (or Hermitian conjugate), $U^* = (\overline{U})^T$, is equal to its inverse:

$$U^* = U^{-1} \quad \text{or equivalently,} \quad U^* U = I$$

The most intuitive way to understand these matrices is by examining their column vectors.

Theorem 4.2.7. *A square $n \times n$ matrix P is orthogonal if and only if its column vectors form an orthonormal set. Similarly, a square $n \times n$ matrix U is unitary if and only if its column vectors form an orthonormal set with respect to the complex inner product.*

Proof. We will prove the theorem for the orthogonal case. The unitary case is analogous, with the transpose replaced by the conjugate transpose. Let the matrix P have column vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$. The transpose, P^T, will have these vectors as its rows.

$$P = \begin{pmatrix} | & | & & | \\ \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \\ | & | & & | \end{pmatrix}, \quad P^T = \begin{pmatrix} - & \mathbf{p}_1^T & - \\ - & \mathbf{p}_2^T & - \\ & \vdots & \\ - & \mathbf{p}_n^T & - \end{pmatrix}$$

Consider the product $P^T P$. The entry in the i-th row and j-th column of this product is the inner product of the i-th row of P^T and the j-th column of P.

$$(P^T P)_{ij} = \mathbf{p}_i^T \mathbf{p}_j = \langle \mathbf{p}_i, \mathbf{p}_j \rangle$$

So, the product matrix is a matrix of inner products of the column vectors of P.

First, assume P is orthogonal. By definition, $P^T P = I$. This means the matrix of inner products must be the identity matrix. The entries of the identity matrix are given by the Kronecker delta, δ_{ij}. Equating the entries gives:

$$\langle \mathbf{p}_i, \mathbf{p}_j \rangle = \delta_{ij}$$

This implies that if $i = j$, $\langle \mathbf{p}_i, \mathbf{p}_i \rangle = \|\mathbf{p}_i\|^2 = 1$, so the vectors are normal. If $i \neq j$, $\langle \mathbf{p}_i, \mathbf{p}_j \rangle = 0$, so the vectors are orthogonal. Thus, the columns form an orthonormal set.

Conversely, assume the columns of P form an orthonormal set. By this definition, their inner products satisfy $\langle \mathbf{p}_i, \mathbf{p}_j \rangle = \delta_{ij}$. The (i, j)-th entry of the product $P^T P$ is precisely this inner product. Therefore, the resulting matrix is the identity matrix, $P^T P = I$. This proves that P is an orthogonal matrix. ■

Example 4.2.8 (Inner Product and Length of a Vector)**.** Let $\mathbf{u} = (1, -2, 3)$ and $\mathbf{v} = (4, 0, -1)$ be vectors in $\mathbb{R}^3$. Calculate their inner product $\langle \mathbf{u}, \mathbf{v} \rangle$ and their respective lengths, $\|\mathbf{u}\|$ and $\|\mathbf{v}\|$.

Solution. The inner product is found by summing the products of the corresponding components:

$$\langle \mathbf{u}, \mathbf{v} \rangle = (1)(4) + (-2)(0) + (3)(-1) = 4 + 0 - 3 = 1$$

The length (or norm) of each vector is the square root of the inner product of the vector with itself:

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = \sqrt{1^2 + (-2)^2 + 3^2} = \sqrt{1 + 4 + 9} = \sqrt{14}$$

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{4^2 + 0^2 + (-1)^2} = \sqrt{16 + 0 + 1} = \sqrt{17}$$

■

Example 4.2.9 (Normalization of a Vector)**.** Find the unit vector (normal vector) $\mathbf{u}$ that is in the same direction as the vector $\mathbf{v} = (2, -1, 2)$.

Solution. To normalize the vector $\mathbf{v}$, we must divide it by its length, $\|\mathbf{v}\|$. First, we calculate the length of $\mathbf{v}$:

$$\|\mathbf{v}\| = \sqrt{2^2 + (-1)^2 + 2^2} = \sqrt{4+1+4} = \sqrt{9} = 3$$

Now, we divide each component of $\mathbf{v}$ by this length to get the unit vector $\mathbf{u}$:

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v} = \frac{1}{3}(2, -1, 2) = \left(\frac{2}{3}, -\frac{1}{3}, \frac{2}{3}\right)$$

We can verify that $\|\mathbf{u}\| = \sqrt{(2/3)^2 + (-1/3)^2 + (2/3)^2} = \sqrt{4/9 + 1/9 + 4/9} = \sqrt{9/9} = 1$. ∎

Example 4.2.10 (Verifying an Orthogonal Matrix)**.** Show that the matrix $P = \frac{1}{3}\begin{pmatrix} 2 & 1 & 2 \\ 1 & 2 & -2 \\ 2 & -2 & -1 \end{pmatrix}$ is orthogonal.

Solution. We must show that the column vectors of P form an orthonormal set. Let the columns be $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$.

$$\mathbf{p}_1 = \frac{1}{3}\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, \quad \mathbf{p}_2 = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}, \quad \mathbf{p}_3 = \frac{1}{3}\begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix}$$

First, we check that each vector has a norm of 1.

$$\|\mathbf{p}_1\|^2 = \frac{1}{9}(2^2 + 1^2 + 2^2) = \frac{1}{9}(4+1+4) = \frac{9}{9} = 1$$

$$\|\mathbf{p}_2\|^2 = \frac{1}{9}(1^2 + 2^2 + (-2)^2) = \frac{1}{9}(1+4+4) = \frac{9}{9} = 1$$

$$\|\mathbf{p}_3\|^2 = \frac{1}{9}(2^2 + (-2)^2 + (-1)^2) = \frac{1}{9}(4+4+1) = \frac{9}{9} = 1$$

The columns are normal. Next, we check for mutual orthogonality by computing their inner products.

$$\langle \mathbf{p}_1, \mathbf{p}_2 \rangle = \frac{1}{9}((2)(1) + (1)(2) + (2)(-2)) = \frac{1}{9}(2+2-4) = 0$$

$$\langle \mathbf{p}_1, \mathbf{p}_3 \rangle = \frac{1}{9}((2)(2) + (1)(-2) + (2)(-1)) = \frac{1}{9}(4-2-2) = 0$$

$$\langle \mathbf{p}_2, \mathbf{p}_3 \rangle = \frac{1}{9}((1)(2) + (2)(-2) + (-2)(-1)) = \frac{1}{9}(2-4+2) = 0$$

Since the columns are normal and mutually orthogonal, they form an orthonormal set. Therefore, the matrix P is orthogonal. ∎

Example 4.2.11 (Verifying a Unitary Matrix)**.** Show that the matrix $U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$ is unitary.

Solution. We must show that the column vectors form an orthonormal set with respect to the complex inner product $\langle \mathbf{u}, \mathbf{v} \rangle = u_1\overline{v_1} + u_2\overline{v_2}$. Let the columns be $\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}$ and $\mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} i \\ 1 \end{pmatrix}$. First, we check the norms.

$$\|\mathbf{u}_1\|^2 = \langle \mathbf{u}_1, \mathbf{u}_1 \rangle = \frac{1}{2}((1)(\bar{1}) + (i)(\bar{i})) = \frac{1}{2}(1 + (i)(-i)) = \frac{1}{2}(1 - (-1)) = \frac{2}{2} = 1$$

$$\|\mathbf{u}_2\|^2 = \langle \mathbf{u}_2, \mathbf{u}_2 \rangle = \frac{1}{2}((i)(\bar{i}) + (1)(\bar{1})) = \frac{1}{2}((i)(-i) + 1) = \frac{1}{2}(1+1) = \frac{2}{2} = 1$$

The columns are normal. Now, we check for orthogonality.

$$\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \frac{1}{2}((1)(\bar{i}) + (i)(\bar{1})) = \frac{1}{2}((1)(-i) + (i)(1)) = \frac{1}{2}(-i+i) = 0$$

Since the columns form a complex orthonormal set, the matrix U is unitary. ∎